Distortion Bounds for Broadcasting With Bandwidth Expansion

Zvi Reznic, *Member, IEEE*, Meir Feder, *Fellow, IEEE*, and Ram Zamir, *Senior Member, IEEE*

***Abstract*—We consider the problem of broadcasting a single Gaussian source to two listeners over a Gaussian broadcast channel, with ρ channel uses per source sample, where $\rho > 1$. A distortion pair (D_1, D_2) is said to be achievable if one can simultaneously achieve a mean-squared error (MSE) D_1 at receiver 1 and D_2 at receiver 2. The main result of this correspondence is an outer bound for the set of all achievable distortion pairs. That is, we find necessary conditions under which (D_1, D_2) is achievable. We then apply this result to the problem of point-to-point transmission over a Gaussian channel with unknown signal-to-noise ratio (SNR) and $\rho > 1$. We show that if a system must be optimal at a certain SNR$_{\min}$, then, asymptotically, the system distortion cannot decay faster than $O(1/SNR)$. As for achievability, we show that a previously reported scheme, due to Mittal and Phamdo (2002), is optimal at high SNR. We introduce two new schemes for broadcasting with bandwidth expansion, combining digital and analog transmissions. We finally show how a system with a partial feedback, returning from the bad receiver to the transmitter and to the good receiver, achieves a distortion pair that lies on the outer bound derived here.**

***Index Terms*—Distortion region, joint source–channel coding, lossy broadcasting.**

I. Introduction

The *broadcast channel*, illustrated in Fig. 1, is a communication channel in which one sender transmits to two or more receivers [1]. Suppose that we are given an analog source and a fidelity criterion, and we want to convey the source to both receivers simultaneously. The problem of *joint source–channel coding* for the broadcast channel is to find the *distortion region* which is the set of all simultaneously achievable distortion pairs (D_1, D_2) at the two receivers. For a general source, broadcast channel, and distortion measure, this problem is still open [2].

We investigate below an important special case, of transmitting a band-limited white Gaussian source over a band-limited white Gaussian broadcast channel with squared-error distortion measure. Note that a Gaussian broadcast channel is a degraded broadcast channel [1], and we shall say that receiver 1 is connected to the *good channel* and receiver 2 is connected to the *bad channel*. Also note that this type of problem can be characterized by the parameter ρ. In continuous-time systems, we define $\rho \triangleq W_c/W_s$, where W_c is the channel bandwidth and W_s is the source bandwidth. In discrete-time systems, ρ is defined as the number of channel uses per source sample. Since band-limited continuous-time systems can be translated to discrete-time systems, we shall use the discrete-time representation. We shall focus on the bandwidth expansion scenario, in which $\rho > 1$.

Manuscript received January 22, 2005; revised December 12, 2005. The material in this correspondence was presented in part at the 40th Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, October 2002. The work of R. Zamir was supported by the Israel Academy of Science under Grant ISF 65/01.

The authors are with the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv 69978, Israel (e-mail: zvi.reznic@amimon.com; meir@eng.tau.ac.il; zamir@eng.tau.ac.il)

Communicated by Y. Steinberg, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2006.878178

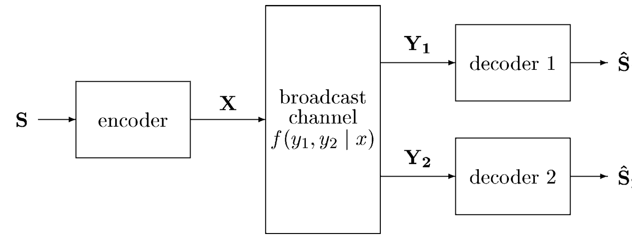

Fig. 1. Lossy transmission of a source through a broadcast channel.

Following Shannon's theory, a trivial Cartesian outer bound on the distortion region is given by $D_1 \geq R^{-1}(\rho C_1)$ and $D_2 \geq R^{-1}(\rho C_2)$, where

$$R(x) = \frac{1}{2} \log \frac{\sigma^2}{x} \tag{1}$$

is the rate-distortion function of a Gaussian source with variance σ^2 (in bits per source sample) [1], and C_1 and C_2 are the individual point-to-point capacities (in bits per channel use) of the good and bad channels, respectively. In the case of $\rho = 1$, the trivial outer bound is achieved by analog transmission, i.e., by sending the source *uncoded* [3]. This means that in this special case, there is no conflict between the needs of the two receivers, and both of them perform as if the needs of the other receiver could be ignored.

For the case of $\rho > 1$, Mittal and Phamdo [4] suggested a hybrid digital–analog scheme which achieves the distortion pair

$$(D_1, D_2) = (R^{-1}((\rho - 1)C_2 + C_1), R^{-1}(\rho C_2)). \tag{2}$$

Other schemes were developed for the case of $\rho > 1$, providing other achievable distortion pairs [3], [5], [6]. However, no nontrivial outer bound (converse) on the distortion region was ever derived. The main result of this correspondence is such an outer bound. For deriving the outer bound we use an auxiliary random variable, similar to the one used by Ozarow [7] for proving the converse for the Gaussian multiple description problem. It follows from our outer bound that the distortion pair (2) is optimal in the limit of high signal-to-noise ratio (SNR).

Regarding an inner bound for the distortion region, we develop a new coding scheme which combines elements from the Mittal–Phamdo scheme together with a Wyner–Ziv source encoding and a broadcast channel encoding. In addition, we outline a second scheme, that can be thought of as a multidimensional extension of Chen and Wornell's analog error-correction scheme [3], making further use of analog transmission.

A variant of the problem above is the problem of sending a Gaussian source over an additive white Gaussian noise (AWGN) channel, where the SNR is unknown except that SNR $\geq$ SNR$_{\min}$, where SNR$_{\min}$ is known. Using our outer bound on the distortion region for the broadcast channel, we prove that for any system, if SNR$_{\min}$ is high, and *if the system is tuned to be optimal at* SNR$_{\min}$, then, as the SNR improves (but the transmitter is held fixed), the distortion cannot decay faster than $1/\text{SNR}$ for all values of ρ. For comparison, we recall that the solution of $R(D') = \rho C$ is given by $D' = \sigma^2/(1 + \text{SNR})^\rho$, and hence, the mean-squared error (MSE) of a collection of systems, each optimally designed for a different (high) SNR, decays as $1/\text{SNR}^\rho$. We note that for the case where the system is optimal at SNR$_{\min}$, our result is stronger than a previous result by Ziv [8], who showed that asymptotically, the distortion cannot decay faster than $1/\text{SNR}^2$ for all values of ρ.

0018-9448/$20.00 © 2006 IEEE

The correspondence is organized as follow: In Section II, we introduce the outer bound on the distortion region. In Section III, we prove the theorem and corollaries of Section II. In Section IV, we apply our results to the case of point-to-point communication over a channel with unknown SNR. In Section V, we introduce a coding scheme for broadcasting with bandwidth expansion. In Section VI, we introduce the modulo-lattice modulation, which makes further use of analog transmission. In Section VII. we analyze the performance of a system with a feedback, and in Section VIII, we conclude the correspondence.

II. Outer Bound on the Distortion Region

In this section, we introduce the outer bound, which is the main result of the correspondence. Before doing so, we note that in the general case of lossy broadcasting the distortion region depends only on the marginal distributions of the channel (see Appendix I for proof). We recall that this is also the case for the *channel coding* problem of broadcast channels [1, p. 422].

We denote the source by $\boldsymbol{S} = (S_1, \ldots, S_m)$, and the decoders output by $\hat{\boldsymbol{S}}_{\boldsymbol{1}} = (\hat{S}_{1,1}, \ldots, \hat{S}_{1,m})$ and $\hat{\boldsymbol{S}}_{\boldsymbol{2}} = (\hat{S}_{2,1}, \ldots, \hat{S}_{2,m})$. We denote the channel input by $\boldsymbol{X} = (X_1, \ldots, X_n)$ and the channel outputs by $\boldsymbol{Y}_1 = (Y_{1,1}, \ldots, Y_{1,n})$ and $\boldsymbol{Y}_{\boldsymbol{2}} = (Y_{2,1}, \ldots, Y_{2,n})$. The bandwidth expansion ratio ρ is defined by

$$\rho = \frac{n}{m} \tag{3}$$

and we shall focus on the case where $\rho > 1$.

Definition 1: A Gaussian broadcast channel with input $\boldsymbol{X}$ and outputs $\boldsymbol{Y}_{\boldsymbol{1}}$ and $\boldsymbol{Y}_{\boldsymbol{2}}$, satisfies for $i = 1, 2$

$$\frac{1}{n}\sum_{t=1}^{n} E\left(X_t^2\right) \le P,$$
$$Y_{i,t} = X_t + Z_{i,t}, \;\; Z_{i,t} \sim \mathcal{N}(0, N_i), \;\; t = 1, \ldots, n \tag{4}$$

where $\boldsymbol{Z}_{\boldsymbol{1}} = (Z_{1,1}, \ldots, Z_{1,n})$ and $\boldsymbol{Z}_{\boldsymbol{2}} = (Z_{2,1}, \ldots, Z_{2,n})$ are memoryless and statistically independent of $\boldsymbol{X}$, and $N_2 \ge N_1$.

The capacities C_1 and C_2 of the good and bad channel, respectively, are given by

$$C_i = \frac{1}{2}\log\left(1 + \frac{P}{N_i}\right) \text{ bits per channel use}, \qquad i = 1, 2. \tag{5}$$

We denote the distortion measure by $d(\boldsymbol{S}, \hat{\boldsymbol{S}}_{\boldsymbol{i}})$ for $(i = 1, 2)$, and define the following.

Definition 2: (D_1, D_2) is an achievable distortion pair if, for any $\epsilon^* > 0$, there exist integers m and $n = \rho m$, an encoding function $\boldsymbol{X} = i_m^n(\boldsymbol{S})$ and reconstruction functions $\hat{\boldsymbol{S}}_{\boldsymbol{1}} = g_{1m}^n(\boldsymbol{Y}_{\boldsymbol{1}})$ and $\hat{\boldsymbol{S}}_{\boldsymbol{2}} = g_{2m}^n(\boldsymbol{Y}_{\boldsymbol{2}})$, such that

$$E(d(\boldsymbol{S}, \hat{\boldsymbol{S}}_{\boldsymbol{i}})) < D_i + \epsilon^*, \qquad \text{for } i = 1, 2. \tag{6}$$

The achievable distortion region is defined as the convex closure of the set of achievable distortion pairs.

Note that it follows from Definition 2 that ρ is a rational number. This does not limit the scope of the results in any practical way, since any nonrational value could be replaced by a rational value which is arbitrary close to it.

In this correspondence, the source is memoryless with $S_t \sim \mathcal{N}(0, \sigma^2)$, and the distortion measure is squared-error, that is,

$$D_i = Ed(\boldsymbol{S}, \hat{\boldsymbol{S}}_{\boldsymbol{i}}) = \frac{1}{m}\sum_{t=1}^{m} E(S_t - \hat{S}_{i,t})^2, \quad i = 1, 2, t = 1, \ldots, m. \tag{7}$$

In summary, we wish to send a memoryless Gaussian source over the Gaussian broadcast channel, with $\rho > 1$, minimizing the squared-error distortion. Our main result is the following.

Theorem 1 (Outer Bound): Let (D_1, D_2) be an achievable distortion pair, and let $\alpha \ge 1$ be defined by

$$D_2 = \alpha R^{-1}(\rho C_2) = \alpha \sigma^2 2^{-2\rho C_2}. \tag{8}$$

Then

$$D_1 \ge \sup_{\kappa > 0} \frac{\sigma^2}{f(\kappa)} \tag{9}$$

where

$$f(\kappa) \triangleq \frac{1}{\kappa}\left(\left\{\frac{N_2}{N_1}\left[\alpha + \left(\frac{P}{N_2} + 1\right)^{\rho}\kappa\right]^{1/\rho} - \left(\frac{N_2}{N_1} - 1\right)(1+\kappa)^{1/\rho}\right\}^{\rho} - 1\right). \tag{10}$$

We note that α is in fact an *excess distortion ratio*, which is the ratio between D_2 and the smallest possible distortion in receiver 2. We shall prove Theorem 1 and its corollaries in Section III. We refer the reader to Appendix II in which we show graphs of $f(\kappa)$ and outline the properties of this function.

An important special case is when we make no compromise in receiver 2 in favor of receiver 1. That is, we require that receiver 2 performs as if it were an optimal point-to-point scenario. In this case, there is no excess distortion, and $\alpha = 1$. Corollary 1 addresses this case.

Corollary 1 (Lower Bound on D_1 When D_2 is Optimal): Let (D_1, D_2) be an achievable distortion pair where

$$D_2 = R^{-1}(\rho C_2). \tag{11}$$

Then

$$D_1 \ge \sigma^2\left(\left(\frac{P + N_2}{N_2}\right)^{\rho - 1}\frac{P + N_2}{N_1} - \frac{N_2 - N_1}{N_1}\right)^{-1}. \tag{12}$$

We shall prove Corollary 1 in Section III. Note that in the matching-bandwidth case ($\rho = 1$), the bound in (12) reduces to the trivial joint source–channel bound

$$D_1 \ge \frac{\sigma^2 N_1}{P + N_1} = R^{-1}(C_1) \tag{13}$$

so at least for this case the bound is tight.

For comparison, Mittal and Phamdo [4] suggested a coding scheme which achieves the distortion pair

$$D_2' = R^{-1}(\rho C_2)$$

and

$$D_1' = R^{-1}((\rho - 1)C_2 + C_1) \tag{14}$$

$$= \sigma^2\left(\left(\frac{P + N_2}{N_2}\right)^{\rho - 1}\frac{P + N_1}{N_1}\right)^{-1}. \tag{15}$$

Comparing this with (11) and (12), we see that their scheme is asymptotically optimal in the limit of high SNR ($P/N_2 \to \infty$).

It can also be shown that in the limit of $N_2 \to \infty$, the lower bound of (12) becomes

$$D_1 \geq \sigma^2 \left(1 + \frac{\rho P}{N_1}\right). \tag{16}$$

In this case, one can actually achieve this bound by sending the source uncoded with power ρP at $1/\rho$ of the time (first m samples of a length-n block), and sending zeros at the rest of the time (last $n - m$ samples). Alternatively, if ρ is an integer, one can achieve this bound by repeating each source sample ρ times at a constant power P.

Corollary 2 addresses the special case in which we make no compromise in receiver 1 in favor of receiver 2.

Corollary 2 (Lower Bound on D_2 When D_1 is Optimal): Let (D_1, D_2) be an achievable distortion pair where

$$D_1 = R^{-1}(\rho C_1). \tag{17}$$

Then

$$D_2 \geq R^{-1}(C_2)\left[1 - \frac{N_1}{N_2} + \frac{N_1}{N_2}\left(\frac{N_1}{P+N_1}\right)^{\rho-1}\right]. \tag{18}$$

We shall prove Corollary 2 in Section III. For comparison, the scheme of Shamai, Verdú, and Zamir [5] (although not designed originally for broadcast channels) achieves the distortion pair

$$D_1 = R^{-1}(\rho C_1), \quad D_2 = R^{-1}(C_2).$$

Hence, their scheme is optimal in the limit of $N_1/N_2 \to 0$.

III. Proofs of Theorem 1, Corollary 1 and Corollary 2

Proof of Theorem 1: We introduce an auxiliary random variable $\boldsymbol{U}$, similar to the one used by Ozarow [7]. Specifically, let $\boldsymbol{U} = (U_1, \ldots, U_m)$ and $\boldsymbol{V} = (V_1, \ldots, V_m)$ be memoryless vectors such that

$$V_t \sim \mathcal{N}(0, \kappa\sigma^2) \quad \text{and} \quad U_t = S_t + V_t \quad (t = 1, \ldots, m) \tag{19}$$

where $\kappa > 0$. Hence, we have Markov chains $\boldsymbol{U} \leftrightarrow \boldsymbol{S} \leftrightarrow \boldsymbol{X} \leftrightarrow \boldsymbol{Y}_i \leftrightarrow \hat{\boldsymbol{S}}_i$ for $i = 1, 2$ (see Fig. 2). By the chain rule for mutual information we have for $(i = 1, 2)$

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y_1}) &= I(\boldsymbol{X};\boldsymbol{U}) + I(\boldsymbol{X};\boldsymbol{Y_1}\,|\,\boldsymbol{U}) - I(\boldsymbol{X};\boldsymbol{U}\,|\,Y_1) && (20)\\ &= I(\boldsymbol{X};\boldsymbol{U}) + h(\boldsymbol{Y_1}\,|\,\boldsymbol{U}) - h(\boldsymbol{Z_1}) - I(\boldsymbol{X};\boldsymbol{U}\,|\,\boldsymbol{Y_1}) && (21)\\ &= I(\boldsymbol{X};\boldsymbol{U}) + h(\boldsymbol{Y_1}\,|\,\boldsymbol{U}) - h(\boldsymbol{Z_1}) - h(\boldsymbol{U}\,|\,\boldsymbol{Y_1})\\ &\quad + h(\boldsymbol{U}\,|\,\boldsymbol{X},\boldsymbol{Y_1})\\ &= I(\boldsymbol{X};\boldsymbol{U}) + h(\boldsymbol{Y_1}\,|\,\boldsymbol{U}) - h(\boldsymbol{Z_1}) - h(\boldsymbol{U}\,|\,\boldsymbol{Y_1})\\ &\quad + h(\boldsymbol{U}\,|\,\boldsymbol{X}) && (22)\\ &= h(\boldsymbol{U}) - h(\boldsymbol{U}\,|\,\boldsymbol{Y_1}) + h(\boldsymbol{Y_1}\,|\,\boldsymbol{U}) - h(\boldsymbol{Z_1}) && (23) \end{aligned}$$

where (21) follows from (4), and in (22) we used the Markov chain relation to replace $h(\boldsymbol{U}\,|\,\boldsymbol{X},\boldsymbol{Y_1})$ with $h(\boldsymbol{U}\,|\,\boldsymbol{X})$.

In (23), $I(\boldsymbol{X};\boldsymbol{Y_1})$ is expressed as a sum of four terms. We shall now upper-bound $I(\boldsymbol{X};\boldsymbol{Y_1})$ by bounding those terms. First, we note that $\boldsymbol{U}$ and $\boldsymbol{Z_1}$ are Gaussian memoryless vectors, where $\boldsymbol{U}$ has variance $(\kappa+1)\sigma^2$ and length m, and $\boldsymbol{Z_1}$ has variance N_1 and length n. Hence, their differential entropies [1] are given by

$$h(\boldsymbol{U}) = \frac{m}{2}\log 2\pi e(\kappa+1)\sigma^2 \tag{24}$$

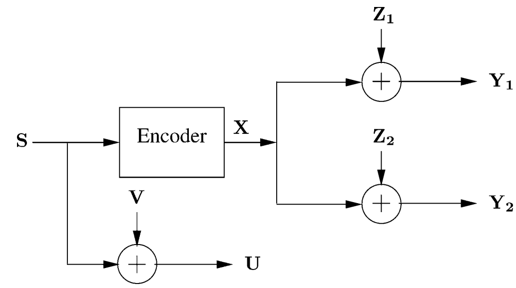

Fig. 2. The Gaussian broadcast channel with the auxiliary variable $\boldsymbol{U}$.

and

$$h(\boldsymbol{Z_1}) = \frac{n}{2}\log 2\pi e N_1. \tag{25}$$

We shall now derive a lower bound the second term in (23), which is $h(\boldsymbol{U}\,|\,\boldsymbol{Y_1})$. By the conditional form of the entropy power inequality [9], and since $\boldsymbol{U}$ is the independent sum of $\boldsymbol{S}$ and $\boldsymbol{V}$ we have that

$$\begin{aligned} 2^{\frac{2}{m}h(\boldsymbol{U}\,|\,\boldsymbol{Y_1})} &\geq 2^{\frac{2}{m}h(\boldsymbol{S}\,|\,\boldsymbol{Y_1})} + 2^{\frac{2}{m}h(\boldsymbol{V}\,|\,\boldsymbol{Y_1})} && (26)\\ &= 2^{\frac{2}{m}h(\boldsymbol{S}\,|\,\boldsymbol{Y_1})} + 2^{\frac{2}{m}h(\boldsymbol{V})} && (27)\\ &= 2^{\frac{2}{m}h(\boldsymbol{S}\,|\,\boldsymbol{Y_1})} + 2\pi e\kappa\sigma^2 && (28) \end{aligned}$$

where (28) follows since $\boldsymbol{V}$ is Gaussian [1]. The term $h(\boldsymbol{S}\,|\,\boldsymbol{Y_1})$ in (28) can be further bounded as follows:

$$\begin{aligned} h(\boldsymbol{S}\,|\,\boldsymbol{Y_1}) &= h(\boldsymbol{S}) - I(\boldsymbol{S};\boldsymbol{Y_1}) && (29)\\ &= \frac{m}{2}\log 2\pi e\sigma^2 - I(\boldsymbol{S};\boldsymbol{Y_1}) && (30)\\ &\geq \frac{m}{2}\log 2\pi e\sigma^2 - I(\boldsymbol{X};\boldsymbol{Y_1}) && (31) \end{aligned}$$

where (30) is since $\boldsymbol{S}$ is Gaussian and (31) is by the data processing inequality. Combining (28) and (31) with the fact that $\rho = n/m$ yields

$$h(\boldsymbol{U}\,|\,\boldsymbol{Y_1}) \geq \frac{m}{2}\log\left(2\pi e\sigma^2\left(2^{-\frac{2\rho}{n}I(\boldsymbol{X};\boldsymbol{Y_1})} + \kappa\right)\right). \tag{32}$$

We shall now derive an upper bound the third term in (23), which is $h(\boldsymbol{Y_1}\,|\,\boldsymbol{U})$. We note that $\boldsymbol{Y_2}$ is the sum of $\boldsymbol{Y_1}$ and a noise with variance $N_2 - N_1$. Hence, using the conditional form of the entropy power inequality [9], we can show (see Appendix III) that

$$2^{\frac{2}{n}h(\boldsymbol{Y_2}\,|\,\boldsymbol{U})} \geq 2^{\frac{2}{n}h(\boldsymbol{Y_1}\,|\,\boldsymbol{U})} + 2^{\log(2\pi e(N_2-N_1))}. \tag{33}$$

(Note that a similar derivation was done in [10].) The left—hand side of (33) can be expressed as follows:

$$\begin{aligned} 2^{\frac{2}{n}h(\boldsymbol{Y_2}\,|\,U)} &= 2^{\frac{2}{n}(h(\boldsymbol{Y_2}) - I(\boldsymbol{Y_2};U))}\\ &\leq 2\pi e(P+N_2)2^{-\frac{2}{n}I(\boldsymbol{Y_2};U)} && (34) \end{aligned}$$

where we used the fact that the variance of $\boldsymbol{Y_2}$ is $P + N_2$, and hence its differential entropy cannot exceed $\frac{n}{2}\log(2\pi e(P+N_2))$ [1, p.262]. Note that the combination of (33) and (34) can serve as an upper bound for $h(\boldsymbol{Y_1}\,|\,\boldsymbol{U})$ in terms of $I(\boldsymbol{Y}_2;\boldsymbol{U})$. We shall now use rate distortion theory to derive a lower bound on $I(\boldsymbol{Y_2};\boldsymbol{U})$. Using (5), we can rewrite (8) as

$$D_2 = \alpha\sigma^2 2^{-2\rho C_2} = \frac{\alpha\sigma^2}{(1+P/N_2)^\rho} = \alpha\sigma^2\left(\frac{N_2}{P+N_2}\right)^\rho. \tag{35}$$

We have

$$E(d(\hat{\boldsymbol{S}}_2, \boldsymbol{U})) = E\left(\frac{1}{m}\sum_{t=1}^{m}(\hat{S}_{2,t} - U_t)^2\right) \quad (36)$$

$$= E\left(\frac{1}{m}\sum_{t=1}^{m}(\hat{S}_{2,t} - S_t + S_t - U_t)^2\right) \quad (37)$$

$$= E\left(\frac{1}{m}\sum_{t=1}^{m}(\hat{S}_{2,t} - S_t)^2\right) + E\left(\frac{1}{m}\sum_{t=1}^{m}(S_t - U_t)^2\right) \quad (38)$$

$$= D_2 + E\left(\frac{1}{m}\sum_{t=1}^{m}V_t^2\right) \quad (39)$$

$$= \alpha\sigma^2\left(\frac{N_2}{P+N_2}\right)^{\rho} + \kappa\sigma^2 \quad (40)$$

where (38) follows since $S_t - U_t = V_t$ is independent of $\hat{S}_{2,t} - S_t$, (39) follows from (7) and (19), and (40) follows from (19) and (35). We now have

$$\frac{1}{n}I(\boldsymbol{Y}_2; \boldsymbol{U}) \geq \frac{1}{n}I(\hat{\boldsymbol{S}}_2; \boldsymbol{U}) \quad (41)$$

$$\geq \frac{1}{n}mR(Ed(\hat{\boldsymbol{S}}_2; \boldsymbol{U})) \quad (42)$$

$$\geq \frac{1}{2\rho}\log\frac{(\kappa+1)\sigma^2}{\alpha\sigma^2\left(\frac{N_2}{P+N_2}\right)^{\rho} + \kappa\sigma^2} \quad (43)$$

where (41) is by the data processing inequality, (42) is by rate-distortion theory, and (43) follows since $\boldsymbol{U}$ is Gaussian with variance $(\kappa+1)\sigma^2$, and by (1) and (40). Combining (33), (34), and (43) yields

$$h(\boldsymbol{Y_1} \,|\, \boldsymbol{U}) \leq \frac{n}{2}\log\left(2\pi e(P+N_2) \times \left(\frac{\alpha\left(\frac{N_2}{P+N_2}\right)^{\rho} + \kappa}{\kappa+1}\right)^{1/\rho} - 2\pi e(N_2 - N_1)\right). \quad (44)$$

Hence, we have bounded all four terms in (23). Combining these terms, that is, combining (23), (24), (25), (32), and (44) yields (45) at the bottom of the page, for all $\kappa > 0$. Algebraic manipulation of (45) yields

$$\frac{1}{n}I(\boldsymbol{X}; \boldsymbol{Y_1}) \leq \frac{1}{2\rho}\log f(\kappa) \quad (46)$$

for all $\kappa > 0$, where $f(\kappa)$ is defined in (10).

By rate distortion theory, by the data processing inequality, and by (46) we have that if (D_1, D_2) is achievable than

$$\frac{1}{\rho}R(D_1) \leq \frac{1}{n}I(\boldsymbol{X}; \boldsymbol{Y_1}) \leq \frac{1}{2\rho}\log f(\kappa) \quad (47)$$

for all $\kappa > 0$. Combining this with the rate distortion function (1) and taking the supermum over all $\kappa > 0$ proves the theorem. □

Proof of Corrolary 1: By Theorem 1 we have that $D_1 \geq \frac{\sigma^2}{f(\kappa)}$ for all $\kappa > 0$, and in particular for $\kappa \to 0$ (from above). By (11) and (8) we have that $\alpha = 1$. Combining this with Property 3 of $f(\kappa)$, which is described in Appendix II, proves the theorem. □

Proof of Corrolary 2: By Property 4 of $f(\kappa)$, which is described in Appendix II, we have that

$$\lim_{\kappa\to\infty}\frac{1}{2}\log f(\kappa) = \rho C_1. \quad (48)$$

Hence, the requirement set by (17) can be written as

$$R(D_1) = \lim_{\kappa\to\infty}\frac{1}{2}\log f(\kappa). \quad (49)$$

Using (1), we can write (49) as

$$D_1 = \lim_{\kappa\to\infty}\frac{\sigma^2}{f(\kappa)}. \quad (50)$$

Combining this with Theorem 1 yield that (D_1, D_2) may only be achievable if $f(\kappa_1) \geq \lim_{\kappa\to\infty} f(\kappa)$ for all $\kappa_1 > 0$ (otherwise, there would be a lower bound on D_1 that contradicts (17)). By Properties 7 and 5 this may only happen if $\alpha \geq \alpha_{th}$. This means that α_{th} is in fact a lower bound on the excess distortion ratio which is possible when receiver 1 is optimal. Combining the definition of α_{th}, (92) with (1), (5), and (8) proves the corollary. □

IV. Transmission Over Channels With Unknown SNR

We now turn to the issue of lossy transmission over a point-to-point channel with unknown SNR. Corollary 1 sets a lower bound on the distortion D_1, achieved at SNR of P/N_1, given that the transmitter is optimal at SNR of P/N_2. Hence, by defining $\mathrm{SNR}_{\min} \triangleq P/N_2$ and $\mathrm{SNR} \triangleq P/N_1$ and by (12) we prove the following corollary.

Corollary 3: For every $\rho > 1$, if a transmitter is designed to be optimal at signal-to-noise ratio $\mathrm{SNR}_{\min}$ and the actual signal-to-noise ratio is SNR, where $\mathrm{SNR} > \mathrm{SNR}_{\min}$, then, the resulting distortion $D(\mathrm{SNR})$ must satisfy

$$D(\mathrm{SNR}) \geq \Phi \cdot \frac{\sigma^2}{\mathrm{SNR}} \cdot (1 - o(1))$$

where Φ is independent of the actual SNR and is given by

$$\Phi = \left(\frac{1}{\mathrm{SNR}_{\min}}\right)^{\rho-1}$$

and $o(1) \to 0$ as $\mathrm{SNR}_{\min} \to \infty$.

Fig. 3 illustrates the results of Corollary 3 in the case of high $\mathrm{SNR}_{\min}$. The bold dots represent the distortion achieved by systems which were designed for specific SNRs (e.g., by separating source coding from channel coding). The dotted line, which connects the bold points, represents the solution for D of the equation

$$R(D) = \rho C(\mathrm{SNR}).$$

The slope of the dotted line (at the limit of high SNR), on a log-log scale is $-\rho$. It follows from Corollary 3 that no scheme can achieve the dotted line for more than one value of SNR. In fact, the solid line, whose slope (at the limit of high SNR) is -1, represents the lower bound of

$$\frac{1}{n}I(\boldsymbol{X}; \boldsymbol{Y_1}) \leq \frac{1}{2}\log\frac{(P+N_2)\left(\frac{\alpha}{\kappa}\left(\frac{N_2}{P+N_2}\right)^{\rho} + 1\right)^{1/\rho} - (N_2 - N_1)\left(\frac{\kappa+1}{\kappa}\right)^{1/\rho}}{N_1\left(\frac{1}{\kappa}2^{-2\frac{\rho}{n}I(\boldsymbol{X};\boldsymbol{Y_1})} + 1\right)^{1/\rho}} \quad (45)$$

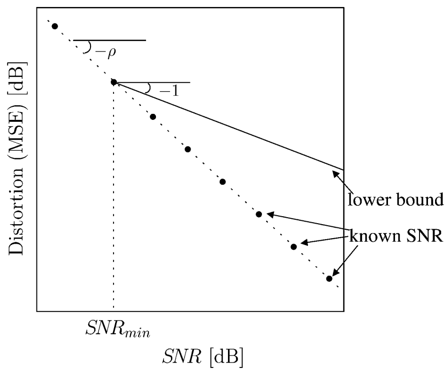

Fig. 3. MSE versus SNR. Solid line: the lower bound of Corollary 3. Dotted line: the solution of $R(D) = \rho C(\mathrm{SNR})$.

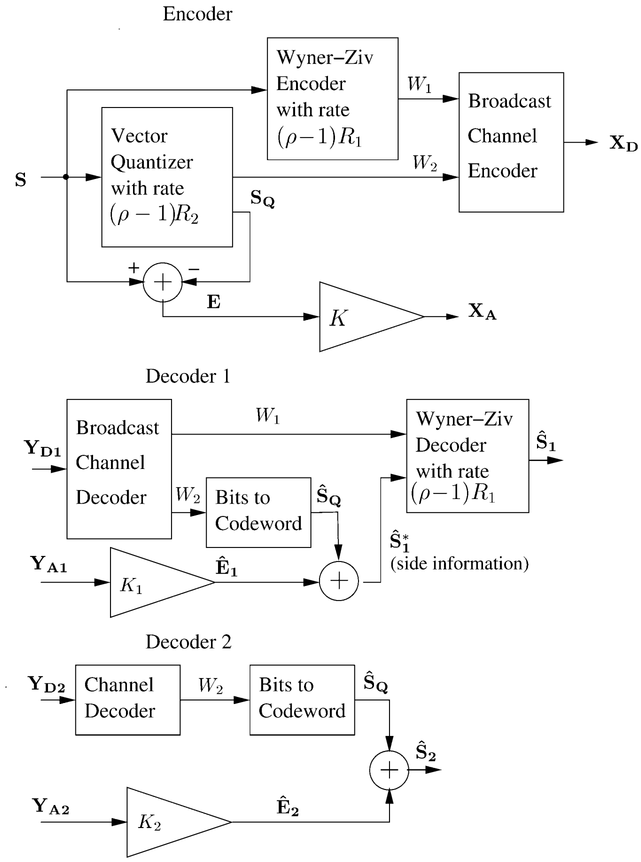

Fig. 4. A coding scheme for lossy transmission with bandwidth expansion.

Corollary 3. Thus, the MSE (SNR) behavior of any system, must be worse than what is represented by the solid line.

Note that in Corollary 3 we restricted the analysis to the case where the system is optimal at $\mathrm{SNR}_{\min}$, that is, when $\alpha = 1$. We conjecture that asymptotically, the distortion cannot decay faster than $1/\mathrm{SNR}$ also when the system is *suboptimal* at $\mathrm{SNR}_{\min}$, that is, when $\alpha > 1$.

It is interesting to compare these results to a previous result of Ziv who analyzed the same problem [8]. In the case where the system is optimal at $\mathrm{SNR}_{\min}$ our result is stronger than Ziv's result, since we showed that the distortion cannot decay faster than $1/\mathrm{SNR}$, while Ziv showed that it cannot decay faster than $1/\mathrm{SNR}^2$. (Although Ziv's result applies even if the system is not optimal at any SNR.) Additionally, we bounded the performance of *any* system, while Ziv restricted his result to a class of systems, which he called "practical."

V. Inner Bound on the Distortion Region

We shall now describe an encoding scheme for lossy transmission of a Gaussian source over a Gaussian broadcast channel with $\rho > 1$. We shall show that one of the Mittal–Phamdo schemes [4], as well as the scheme of Shamai, Verdú, and Zamir [5], are special cases of the scheme which we shall now describe. The encoder, and the two decoders are illustrated in Fig. 4. The transmission block $\boldsymbol{X}$ of length $n = \rho m$ is generated by concatenating (i.e., multiplexing in time) a "digital" block $\boldsymbol{X}_D$ of length $(\rho - 1)m$, and an "analog" block $\boldsymbol{X_A}$ of length m. The digital block is generated by a broadcast channel transmitter [1], such that a common message W_2 is losslessly sent to both receivers, and a private message W_1 is sent only to receiver 1. To allow lossless decoding, we set the rates R_1 and R_2 of W_1 and W_2, respectively (measured in bits per channel use), such that for some $0 \le \beta \le 1$ and some $\epsilon > 0$ (see [1, p. 380])

$$R_1 = \frac{1}{2}\log\left(1 + \frac{(1-\beta)P}{N_1}\right) - \epsilon \tag{51}$$

and

$$R_2 = \frac{1}{2}\log\left(1 + \frac{\beta P}{N_2 + (1-\beta)P}\right) - \epsilon. \tag{52}$$

Since we transmit (W_1, W_2) over a channel with $\rho - 1$ channel uses per source sample, the rates in the source domain are $((\rho - 1)R_1, (\rho - 1)R_2)$ bits per source sample.

We shall now describe the content of the messages and the analog signal, referring to Fig. 4. The source is quantized by a k-dimensional vector quantizer $Q(\,\cdot\,)$, with $2^{k(\rho-1)R_2}$ quantization points and average distortion D_Q. We fix $\epsilon_1 > 0$, choose k sufficiently large, and design the vector quantizer (VQ) such that it achieves

$$(\rho - 1)R_2 = R(D_Q) + \epsilon_1 \tag{53}$$

where $R(D_Q)$ is measured in bits per source sample. We denote the VQ output by $\boldsymbol{S_Q} = (S_{Q1}, \ldots, S_{Qm})$. That is,

$$S^{jk}_{Q(j-1)k+1} = Q(S^{jk}_{(j-1)k+1})$$

where

$$S^{jk}_{Q(j-1)k+1} = (S_{Q(j-1)k+1}, \ldots, S_{Qjk})$$

and

$$S^{jk}_{(j-1)k+1} = (S_{(j-1)k+1}, \ldots, S_{jk}).$$

(We assume that m/k is an integer). The quantization error $\boldsymbol{E} = (E_1, \ldots, E_m)$ is defined as $E_t = S_{Qt} - S_t$. Each sample in $\boldsymbol{E}$ is scaled by a scalar K to produce $\boldsymbol{X_A}$.

The message W_2 is an integer which uniquely describes the vector $\boldsymbol{S_Q}$. Since the length of $\boldsymbol{S_Q}$ is m, and its rate is $(\rho - 1)R_2$ bits per source sample, we have that $W_2 \in (1, \ldots, 2^{m(\rho-1)R_2})$.

Using broadcast channel decoders, both receivers will decode the message W_2 losslessly, and hence will be able to regenerate $\boldsymbol{S_Q}$ losslessly. Hence, the problem reduces to that of lossy transmission of $\boldsymbol{E}$, whose variance is D_Q.

Let (D_1', D_2') be the distortion pair which is achievable by our scheme. Referring again to Fig. 4, we denote by $\boldsymbol{Y_{D1}}$ and $\boldsymbol{Y_{D2}}$ the noisy outputs of the broadcast channel, in response to the input $\boldsymbol{X_D}$, and by $\boldsymbol{Y_{A1}}$ and $\boldsymbol{Y_{A2}}$ the noisy outputs of the broadcast channel, in response to the input $\boldsymbol{X_A}$. Receiver 2 estimates $\boldsymbol{E}$ by multiplying the input $\boldsymbol{Y}_{A2}$ by a gain factor K_2. By setting

$$K = \sqrt{\frac{P}{D_Q}} \quad \text{and} \quad K_2 = \frac{\sqrt{PD_Q}}{P + N_2}$$

and taking the limit as $\epsilon \to 0$ and $\epsilon_1 \to 0$ we have

$$D_2' = \frac{D_Q}{1+\frac{P}{N_2}} \tag{54}$$

$$= \frac{N_2}{P+N_2} R^{-1}((\rho-1)R_2) \tag{55}$$

$$= \frac{\sigma^2 N_2}{P+N_2} 2^{-2(\rho-1)R_2} \tag{56}$$

$$= \frac{\sigma^2 N_2}{P+N_2}\left(1+\frac{\beta P}{N_2+(1-\beta)P}\right)^{-(\rho-1)} \tag{57}$$

where (54) follows from standard MSE calculations (since the estimator of E_t is scalaric and linear, its performance depends only on the average of the variances of E_t), (55) is by (53), (56) is by (1), and (57) is by (52).

As for the good receiver, we note that we can make use of the private message W_1 to further reduce the distortion. However, as a temporary stage, suppose that receiver 1 would estimate the source while completely ignoring the private message. We shall denote this estimate by $\hat{\boldsymbol{S}}_1^*$. Let D_1^* be the average distortion between $\boldsymbol{S}$ and $\hat{\boldsymbol{S}}_1^*$. Repeating the steps that led to (57) one can verify that

$$D_1^* = \frac{\sigma^2 N_1}{P+N_1}\left(1+\frac{\beta P}{N_2+(1-\beta)P}\right)^{-(\rho-1)}. \tag{58}$$

Our problem with respect to decoder 1 reduces now to the following: the encoder needs to send a message W_1, (at rate $(\rho-1)R_1$ bits per source sample) to the decoder, describing the source $\boldsymbol{S}$, taking into account that the decoder already has side information $\hat{\boldsymbol{S}}_1^*$. This is in fact the Wyner–Ziv problem [11][12]. In Appendix IV, we prove a general upper bound on the quadratic Wyner–Ziv rate-distortion function in terms of the MSE between the source and the side information. In our case this bound asserts

$$R^{WZ}_{\boldsymbol{S}\mid\hat{\boldsymbol{S}}_1^*}(x) \le \frac{1}{2}\log\frac{\frac{1}{m}\sum_{t=1}^m E(S_t-\hat{S}^*_{1,t})^2}{x}$$
$$= \frac{1}{2}\log\frac{D_1^*}{x} \tag{59}$$

where $R^{WZ}_{\boldsymbol{S}\mid\hat{\boldsymbol{S}}_1^*}(\cdot)$ is the Wyner–Ziv rate-distortion function of $\boldsymbol{S}$ given side information $\boldsymbol{S}_1^*$. Therefore, it is possible to design Wyner–Ziv encoder and decoder with rate $(\rho-1)R_1$ bits per source sample that achieves (as $\epsilon \to 0$)

$$D_1' = D_1^* \cdot 2^{-2(\rho-1)R_1}$$
$$= \frac{\sigma^2 N_1}{P+N_1}\left\{\left(1+\frac{\beta P}{N_2+(1-\beta)P}\right)\left(1+\frac{(1-\beta)P}{N_1}\right)\right\}^{-(\rho-1)} \tag{60}$$

where (60) follows from (51) and (58).

Note that in the special case of $\beta = 1$ ($R_1 = 0$), this scheme is the same as one of the Mittal–Phamdo schemes [4]. On the other extreme, setting $\beta = 0$, ($R_2 = 0$) reduces this scheme to the one of Shamai, Verdú, and Zamir [5]. Rewriting (60) and (57) in terms of α of (8), leads to the following theorem.

Theorem 2 (Inner Bound): For sending a Gaussian source with variance σ^2 over the Gaussian broadcast channel, any distortion pair (D_1', D_2') of the form

$$D_2' = \alpha R^{-1}(\rho C_2) = \alpha\sigma^2\left(\frac{N_2}{P+N_2}\right)^{\rho} \tag{61}$$

and

$$D_1' \ge \alpha\sigma^2\left(\frac{N_2}{P+N_2}\right)^{\rho-1}\frac{N_1}{P+N_1}$$
$$\times\left(1+\frac{N_2}{N_1}\left(\alpha^{1/(\rho-1)}-1\right)\right)^{-(\rho-1)} \tag{62}$$

for some $\alpha > 1$, is achievable.

Fig. 5 shows the inner bound of Theorem 2 with the outer bound of Theorem 1. The graphs are shown for the case of $\rho = 2, \sigma^2 = 1, P = 1, N_1 = 0.001$ and $N_2 = 0.01$. For the outer bound we used a computer program to find the maximum of

$$\frac{\sigma^2}{f(\kappa)}$$

over all $\kappa > 0$. It can be seen from the graphs that the gap between the bounds is small. In [13], we compare the performance of the above scheme to the performance of the scheme of Mittal and Phamdo. The comparison is limited due to some mathematical difficulties.

VI. Inner Bound by Modulo-Lattice Modulation

In this section, we introduce the *modulo-lattice modulation* scheme. The scheme is designed for the case of $\rho = 2$ and $\alpha = 1$ (minimal D_2), although it could be generalized to other values of ρ. We shall only outline the concept of the scheme. A more detailed description and analysis can be found in [13].

Before proceeding, we refer back to Fig. 4 and point out that in the case of $\alpha = 1$, we have that $R_2 = C_2$ and $R_1 = 0$. Hence, the Wyner–Ziv encoder could be omitted, and the broadcast channel encoder reduces to a point-to-point channel encoder.

The new transmitter that we suggest is depicted in Fig. 6(a). It is similar to the one of Fig. 4 (with $\rho = 2$ and $\alpha = 1$), except that the message W_2 is not transmitted at all. Instead, we transmit the source $\boldsymbol{S}$ uncoded. (We denote $\boldsymbol{X'_A} = \tilde{K}_2\boldsymbol{S}$.) In addition, the vector quantizer is a lattice vector quantizer. Therefore, $\boldsymbol{E}$ can be expressed as $\boldsymbol{E} = \boldsymbol{S} \bmod \Lambda$, where Λ is the lattice. For this reason, we call this scheme the *modulo-lattice modulation* scheme. $\boldsymbol{E}$ is sent uncoded and we denote $\boldsymbol{X_A} = \tilde{K}_1\boldsymbol{E}$.

Receiver 1 and receiver 2, depicted in Fig. 6(b) are identical, except for different gain factors. The quantization-level decoder employs a modified nearest-neighbor algorithm which losslessly decodes $\hat{\boldsymbol{S}}_{\boldsymbol{Q}i}$. Hence, with high probability, $\hat{\boldsymbol{S}}_{\boldsymbol{Q}i} = \boldsymbol{S_Q}$ for $i = 1, 2$. We then add a scaled version of $\boldsymbol{Y_{Ai}}$ (a noisy version of $\boldsymbol{X_A}$) to $\hat{\boldsymbol{S}}_{\boldsymbol{Q}i}$ and generate an estimate $\hat{\boldsymbol{S}}_{\boldsymbol{i}}'$ of $\boldsymbol{S}$. The final estimate $\hat{\boldsymbol{S}}_{\boldsymbol{i}}$ is then generated by weighted averaging of $\boldsymbol{Y'_{Ai}}$ (a noisy version of $\boldsymbol{X'_A}$) and $\hat{\boldsymbol{S}}_{\boldsymbol{i}}'$.

In [13], we show that the modulo-lattice has the same performance as the hybrid digital–analog scheme described in Section V. Yet, we described it here because of the following reasons.

1. The modulo-lattice scheme is interesting since it allows correct “hard decision” in the receiver, although the transmitted signals are “soft.” “Soft” transmission has a potential for improved performance in broadcast scenarios, although we were not able to exploit this potential.
2. In light of the result of Section VII, we conjecture that small modification to the modulo-lattice scheme can result in optimal performance that meets the outer bound of Corollary 1.
3. The structure of the modulo-lattice scheme resembles the nested-lattice Wyner–Ziv encoding scheme of [14], if we view the channel noise as “quantization noise.” Hence, modulo-lattice modulation can also be interpreted as analog communication with side information, or as a joint Wyner–Ziv channel-coding scheme. This aspect will be explored in future work.

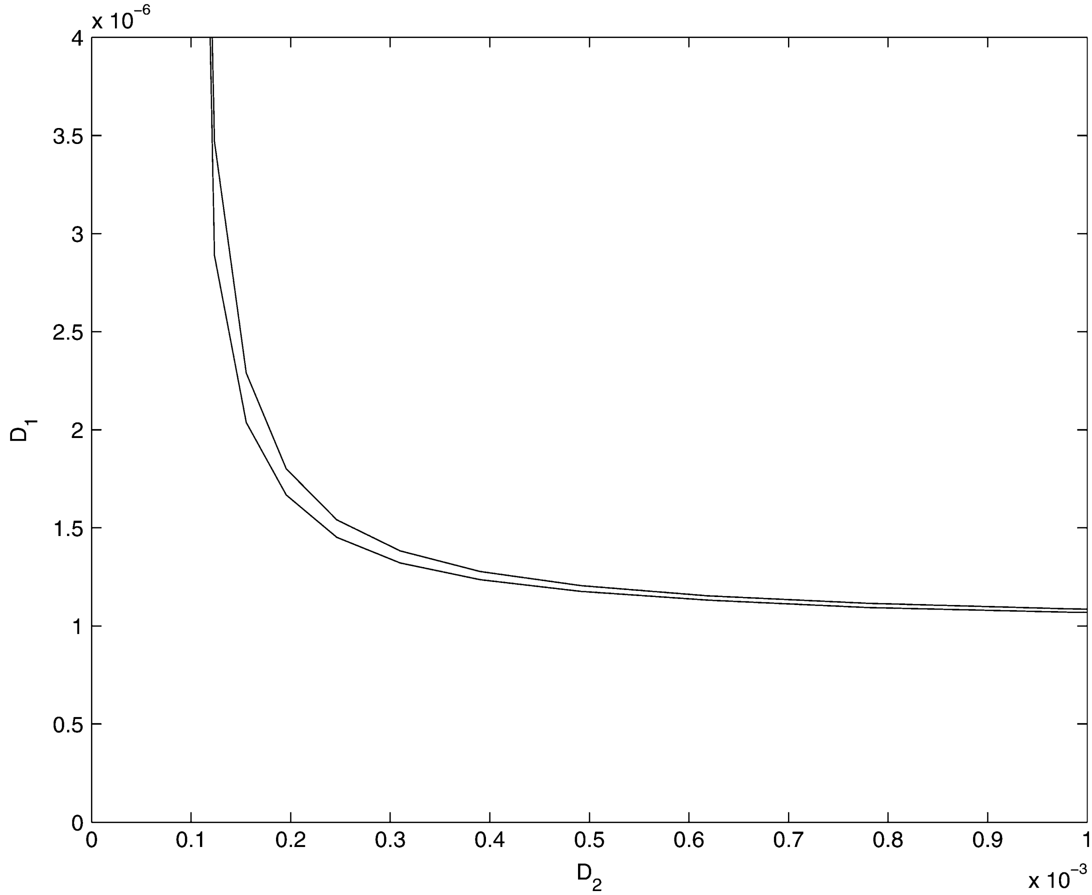

Fig. 5. Numerical analysis of the inner and outer bounds.

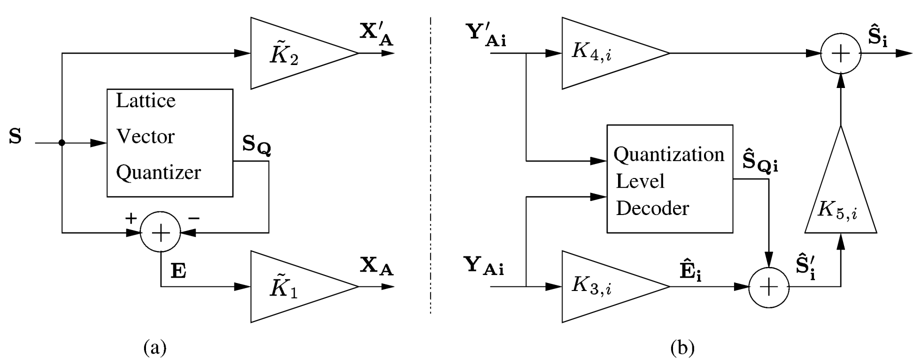

Fig. 6. Modulo-lattice modulation for lossy transmission with bandwidth expansion. (a) Transmitter. (b) Receivers $(i = 1, 2)$.

Note that this scheme is similar in concept to the analog error-correcting scheme suggested by Chen and Wornell [3]. The main difference is that in their scheme the lattice is one-dimensional, and their "tent-map" takes here the form of "sawtooth-map."

VII. The Effect of Feedback

We shall show how a partial feedback can improve the performance relative to the schemes that were presented so far (and did not require a feedback). Moreover, we shall see that the resulting distortion pair meets the lower bound of Corollary 1 for $\rho = 2$. Although this does not imply optimality, the fact that the distortion pair achieved with feedback meets the lower bound is significant. This is since in many other communication problems, there exist schemes without feedback, that achieves the best possible performance of systems with feedback. We conjecture that this is also the case here. That is, we conjecture that there exist a system without feedback, that achieves the same performance as the system with feedback, and therefore we conjecture that the lower bound of Corollary 1 is tight for $\rho = 2$.

We recall that the distortion region of a stochastically degraded broadcast channel is the same as that of the corresponding physically degraded channel (see Appendix I). We shall now focus on the physically degraded channel and, as in Section VI, we shall only consider the case where the bad receiver is kept optimal and there are two channel uses per source sample $(\rho = 2)$. The encoder, the channel, the feedback, and the decoders are illustrated in Fig. 7. We concentrate on physically degraded channels since in all other cases, the feedback would give the good receiver an "unfair" advantage. This is since, in these cases, the feedback actually serves as a new

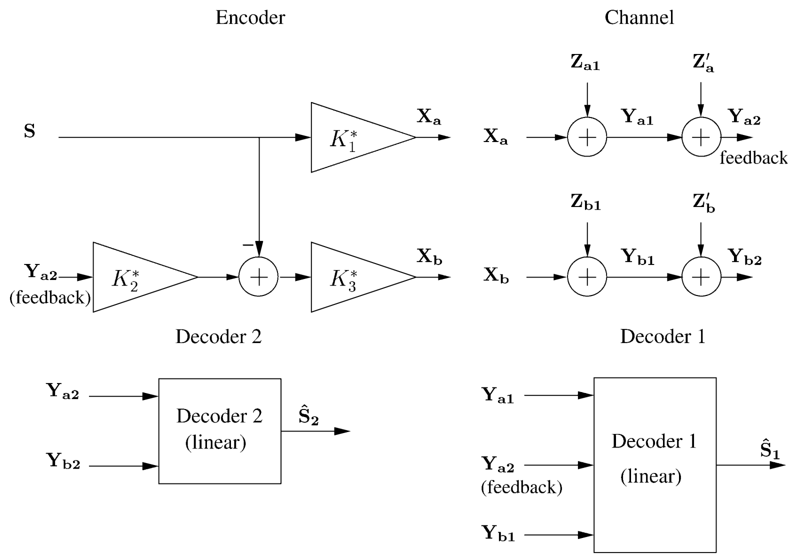

Fig. 7. Broadcasting with feedback.

observation of the source which is given to the good receiver. On the other hand, in physically degraded channels, the feedback conveys no new information about the source (only new information about the reception at the bad receiver).

The encoder output block $\boldsymbol{X}$ of length $n = 2m$ is a concatenation of two length-m blocks $\boldsymbol{X_a}$ and $\boldsymbol{X_b}$, where $\boldsymbol{X_a} = K_1^* \boldsymbol{S}$ and $K_1^* = \sqrt{P/\sigma^2}$. Alternatively, we can write

$$X_{a,t} = K_1^* S_t, \qquad t = 1, 2, \ldots, m. \tag{63}$$

The channel is a physically degraded channel and therefore [1]

$$Y_{a1,t} = X_{a,t} + Z_{a1,t} \tag{64}$$
$$Y_{a2,t} = X_{a,t} + Z_{a1,t} + Z'_{a,t}, \qquad t = 1, 2, \ldots, m \tag{65}$$

where $Z_{a1,t} \sim \mathcal{N}(0, N_1)$ and $Z'_{a,t} \sim \mathcal{N}(0, N_2 - N_1)$ and $\boldsymbol{Z}_{a1}$ and $\boldsymbol{Z}'_a$ are memoryless and independent of each other and of $\boldsymbol{X}$.

The noisy signal $Y_{a2,t}$ returns as a feedback to the transmitter and to receiver 1. The transmitter generates $X_{b,t}$ by

$$X_{b,t} = K_3^*(S - K_2^* Y_{a2,t}), \qquad t = 1, 2, \ldots, m \tag{66}$$

where $K_2^* = \sqrt{\frac{P\sigma^2}{P+N_2}}$ is the Wiener gain for receiver 2, and

$$K_3^* = \sqrt{\frac{(P + N_2)P}{N_2\sigma^2}}$$

is a gain factor that scales $X_{b,t}$ to have a power of P. As before, we have

$$Y_{b1,t} = X_{b,t} + Z_{b1,t} \tag{67}$$
$$Y_{b2,t} = X_{b,t} + Z_{b1,t} + Z'_{b,t}, \qquad t = 1, 2, \ldots, m \tag{68}$$

where $Z_{b1,t} \sim \mathcal{N}(0, N_1)$ and $Z'_{b,t} \sim \mathcal{N}(0, N_2 - N_1)$ and $\boldsymbol{Z}_{b1}$ and $\boldsymbol{Z}'_b$ are memoryless and independent of each other and of $\boldsymbol{X}$.

We shall now describe the operation of the two receivers. Let

$$\boldsymbol{Y_{2,t}} \triangleq \begin{bmatrix} Y_{a2,t} \\ Y_{b2,t} \end{bmatrix}$$

and

$$\boldsymbol{Y_{1,t}} \triangleq \begin{bmatrix} Y_{a1,t} \\ Y_{a2,t} \\ Y_{b1,t} \end{bmatrix}, \qquad t = 1, \ldots, m. \tag{69}$$

(Recall that $Y_{a2,t}$ is the feedback.) The two receivers employ the following optimal linear estimation of $\hat{S}_t$. Let

$$\boldsymbol{R_{y,i}} = E\left(\boldsymbol{Y^t_{i,t}} \cdot \boldsymbol{Y_{i,t}}\right)$$

and

$$\boldsymbol{r_{sy,i}} = E(S_t \boldsymbol{Y_{i,t}}). \tag{70}$$

Combining (63)–(70) yields

$$\boldsymbol{R_{y,2}} = \begin{bmatrix} P + N_2 & 0 \\ 0 & P + N_2 \end{bmatrix}$$
$$\boldsymbol{r_{sy,2}} = \begin{bmatrix} \sqrt{P\sigma^2} \\ \sqrt{\frac{N_2 P\sigma^2}{P+N_2}} \end{bmatrix} \tag{71}$$
$$\boldsymbol{R_{y,1}} = \begin{bmatrix} P + N_1 & P + N_1 & \frac{P(N_2 - N_1)}{\sqrt{N_2(P+N_2)}} \\ P + N_1 & P + N_2 & 0 \\ \frac{P(N_2 - N_1)}{\sqrt{N_2(P+N_2)}} & 0 & P + N_1 \end{bmatrix} \tag{72}$$

and

$$\boldsymbol{r_{sy,1}} = \begin{bmatrix} \sqrt{P\sigma^2} \\ \sqrt{P\sigma^2} \\ \sqrt{\frac{N_2 P\sigma^2}{P+N_2}} \end{bmatrix}. \tag{73}$$

The linear estimation is given by

$$\hat{S}_{i,t} = \boldsymbol{a_i^t} \cdot \boldsymbol{Y}_{i,t}, \tag{74}$$

where

$$\boldsymbol{a_i} = \boldsymbol{R_{y,i}^{-1}} \boldsymbol{r_{sy,i}}. \tag{75}$$

The resulting distortion is then given by

$$D_i = \sigma^2 - \boldsymbol{a_i^t} \cdot \boldsymbol{r}_{sy,i}. \tag{76}$$

Combining (71)–(76) yields

$$D_1 = \frac{\sigma^2 N_1 N_2}{P^2 + 2PN_2 + N_1 N_2}$$
$$\text{and}$$
$$D_2 = \frac{\sigma^2 N_2^2}{(P + N_2)^2}. \tag{77}$$

Using the rate distortion function of a Gaussian source (1) and the capacity of a Gaussian channel (5), one can verify that the distortion pair of (77) meets the lower bound of Corollary 1 for $\rho = 2$. We clarify that this does not imply optimality since the scheme assumed the existence of a feedback, whereas the lower bound did not assume any feedback. Yet, we shall now explain the potential we see.

Shannon showed that feedback does not improve the capacity of a point-to-point channel. There are other communication scenarios in which a feedback cannot improve the performance. We conjecture that in our case as well, there exists a scheme that does not require a feedback, and yields the same distortion pair as the one achieved with feedback. This conjecture, combined with the result above leads us to conjecture that the bound of Corollary 1 is tight for $\rho = 2$.

VIII. Conclusion

For lossy transmission of a Gaussian source over a Gaussian broadcast channel with bandwidth expansion, we have derived inner and outer bounds on the set of all achievable distortion pairs (D_1, D_2), and showed that one of the Mittal–Phamdo schemes is optimal at high SNR. The inner bound generalizes both the Mittal–Phamdo scheme and the Shamai–Verdú–Zamir scheme.

Although the distortion in point-to-point communications is given by $D = \sigma^2/(1 + \mathrm{SNR})^\rho$, we showed that if a system must be optimal at a certain $\mathrm{SNR}_{\min}$, then asymptotically the distortion cannot decay faster than $1/\mathrm{SNR}$.

Appendix I: The Distortion Depends Only on the Channel's Marginals

We shall now describe a general property of lossy broadcasting. We recall that in the *channel coding* problem for broadcast channels, the capacity region depends only on the marginal distributions of the channel [1, p. 422]. We shall show here that the same is true for the distortion region in lossy broadcasting. We start with a definition.

Definition 3: A broadcast channel consists of an input alphabet $\mathcal{X}$ and two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ and a probability transition function $f_{y_1,y_2\,|\,x}(\boldsymbol{y_1}, \boldsymbol{y_2}\,|\,\boldsymbol{x})$, where $\boldsymbol{x}, \boldsymbol{Y_1}$, and $\boldsymbol{Y_2}$ are of length n.

Now, suppose that we are given a source, a distortion measure, and two broadcast channels (with the same input and output alphabets), one with probability transition function $f_{y_1,y_2\,|\,x}(\boldsymbol{y_1}, \boldsymbol{y_2}\,|\,\boldsymbol{x})$ and one with probability transition function $f^*_{y_1,y_2\,|\,x}(\boldsymbol{y_1}, \boldsymbol{y_2}\,|\,\boldsymbol{x})$, such that

$$f_{y_1\,|\,x}(\boldsymbol{y_1}\,|\,\boldsymbol{x}) = f^*_{y_1\,|\,x}(\boldsymbol{y_1}\,|\,\boldsymbol{x}), \qquad \text{for all } \boldsymbol{Y_1} \in {\mathcal{Y}_1}^n \text{ and } \boldsymbol{x} \in \mathcal{X}^n \tag{78}$$

$$f_{y_2\,|\,x}(\boldsymbol{y_2}\,|\,\boldsymbol{x}) = f^*_{y_2\,|\,x}(\boldsymbol{y_2}\,|\,\boldsymbol{x}), \qquad \text{for all } \boldsymbol{Y_2} \in {\mathcal{Y}_2}^n \text{ and } \boldsymbol{x} \in \mathcal{X}^n \tag{79}$$

but

$$f_{y_1,y_2\,|\,x}(\boldsymbol{y_1}, \boldsymbol{y_2}\,|\,\boldsymbol{x}) \neq f^*_{y_1,y_2\,|\,x}(\boldsymbol{y_1}, \boldsymbol{y_2}\,|\,\boldsymbol{x}), \qquad \text{for some } (\boldsymbol{x}, \boldsymbol{y_1}, \boldsymbol{y_2}). \tag{80}$$

Now, using the notations of Definition 2, suppose that we arbitrarily choose an encoder $i_m(\boldsymbol{S})$ and decoders $g_{1m}(\boldsymbol{Y_1})$ and $g_{2m}(\boldsymbol{Y_2})$, and we calculate the average distortion that result from the use of these decoders. We denote by $D_i^f (i = 1, 2)$ the distortions in the case where the channel probability transition function is $f_{y_1,y_2\,|\,x}(\boldsymbol{y_1}, \boldsymbol{y_2}\,|\,\boldsymbol{x})$ and by $D_i^{f^*}$ $(i = 1, 2)$ the distortions in the case where the channel probability transition function is $f^*_{y_1,y_2\,|\,x}(\boldsymbol{y_1}, \boldsymbol{y_2}\,|\,\boldsymbol{x})$. Then, the distortions can be written for $i = 1, 2$ as follow:

$$D_i^f = \int_S \int_{Y_i} f(\boldsymbol{S}) \cdot f_{y_i\,|\,x}(\boldsymbol{Y_i}\,|\,i_m(\boldsymbol{S}))d(\boldsymbol{S}, g_{im}(\boldsymbol{Y_i}))d\boldsymbol{Y_i}d\boldsymbol{S} \tag{81}$$

and

$$D_i^{f^*} = \int_S \int_{Y_i} f(\boldsymbol{S}) \cdot f^*_{y_i\,|\,x}(\boldsymbol{Y_i}\,|\,i_m(\boldsymbol{S}))d(\boldsymbol{S}, g_{im}(\boldsymbol{Y_i}))d\boldsymbol{Y_i}d\boldsymbol{S}. \tag{82}$$

Combining (78), (79), (81), and (82) yields

$$\left(D_1^{f^*}, D_2^{f^*}\right) = \left(D_1^f, D_2^f\right). \tag{83}$$

It follows that any distortion pair that is achievable on $f_{y_1,y_2\,|\,x}(\boldsymbol{y_1}, \boldsymbol{y_2}\,|\,\boldsymbol{x})$ is also achievable on $f^*_{y_1,y_2\,|\,x}(\boldsymbol{y_1}, \boldsymbol{y_2}\,|\,\boldsymbol{x})$ and *vice versa*. We therefore proved the following lemma.

Lemma 1: The distortion region depends on the broadcast channel probability transition function $f_{y_1,y_2\,|\,x}(\boldsymbol{y_1}, \boldsymbol{y_2}\,|\,\boldsymbol{x})$ only through the marginal distributions $f_{y_1\,|\,x}(\boldsymbol{y_1}\,|\,\boldsymbol{x})$ and $f_{y_2\,|\,x}(\boldsymbol{y_2}\,|\,\boldsymbol{x})$.

An immediate conclusion from Lemma 1 is that the distortion region of a stochastically degraded broadcast channel is the same as that of the corresponding physically degraded broadcast channel.

Appendix II: Properties of the Function $f(\kappa)$

We shall now outline the properties of the function $f(\kappa)$ (note that $\kappa > 0$ by definition). Examples of $f(\kappa)$ are illustrated in Fig. 8.

Property 1: The function $f(\kappa)$ is continuous in κ.

Property 2: If $\alpha > 1$ then

$$\lim_{\kappa \to 0} f(\kappa) = \infty. \tag{84}$$

Property 3: If $\alpha = 1$ then

$$\lim_{\kappa \to 0} f(\kappa) = \left(\frac{P + N_2}{N_2}\right)^{\rho - 1} \frac{P + N_2}{N_1} - \frac{N_2 - N_1}{N_1}. \tag{85}$$

Property 4: In the limit of $\kappa \to \infty$, the function $f(\kappa)$ is independent of α and is given by

$$\lim_{\kappa \to \infty} f(\kappa) = \left(1 + \frac{P}{N_1}\right)^\rho = 2^{2\rho C_1}. \tag{86}$$

Property 5: The derivative of $f(\kappa)$ with respect to κ is given by

$$\frac{\partial f(\kappa)}{\partial \kappa} = \frac{g(\kappa)}{\kappa^2} \tag{87}$$

where

$$g(\kappa) = \frac{h_1(\kappa)h_2(\kappa)}{N_1^\rho} + 1 \tag{88}$$

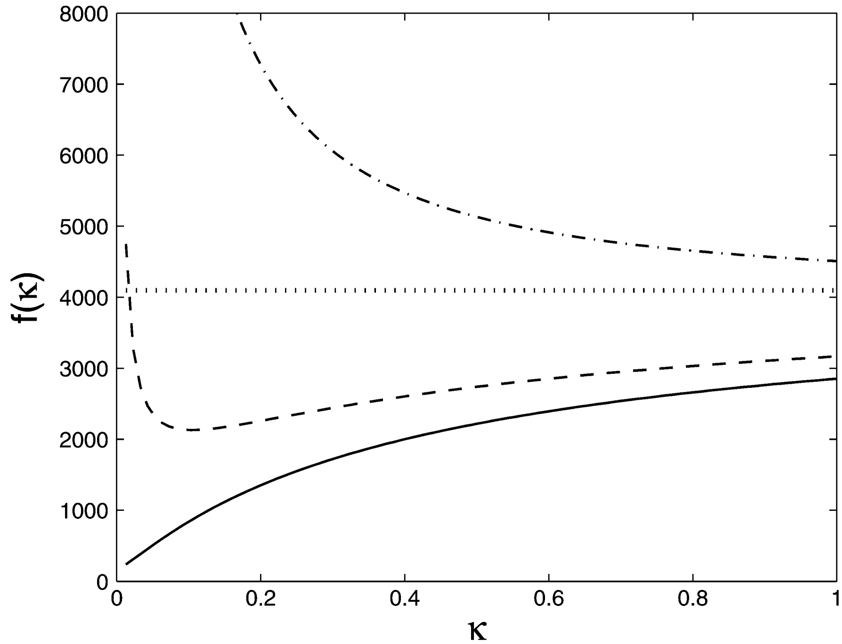

Fig. 8. $f(\kappa)$ for different values of α. Solid line: $\alpha = 1$, dashed line: $1 < \alpha < \alpha_{th}$ ($\alpha = 2$), dash-dot line: $\alpha > \alpha_{th}$ ($\alpha = 6$). The dotted line represents the limit of $f(\kappa)$ as $\kappa \to \infty$, which is independent of α. (Parameters: $P = 0.15, N_1 = 0.01, N_2 = 0.1, \rho = 3$, and, therefore, $\alpha_{th} = 5.63$).

where

$$h_1(\kappa) = \left\{ N_2 \left[\frac{\alpha}{\kappa} + \left(1 + \frac{P}{N_2}\right)^{\rho} \right]^{1/\rho} - (N_2 - N_1) \left(\frac{1}{\kappa} + 1 \right)^{1/\rho} \right\}^{\rho - 1} \tag{89}$$

and

$$h_2(\kappa) = N_2(-\alpha) \left[\frac{\alpha}{\kappa} + \left(1 + \frac{P}{N_2}\right)^{\rho} \right]^{1/\rho - 1} + (N_2 - N_1) \left(\frac{1}{\kappa} + 1 \right)^{1/\rho - 1}. \tag{90}$$

Property 6: If follows from Property 5 that

$$\lim_{\kappa \to \infty} \frac{\partial f(\kappa)}{\partial \kappa} = 0. \tag{91}$$

Property 7: $\lim_{\kappa \to \infty} g(\kappa) < 0$ if and only if

$$\alpha > \alpha\text{th} \triangleq \left(1 + \frac{P}{N_2}\right)^{\rho - 1} \left[\frac{N_1}{N_2} \left(\frac{N_1}{P + N_1} \right)^{\rho - 1} + \frac{N_2 - N_1}{N_2} \right] \tag{92}$$

where $g(\kappa)$ was defined in (88).

In the proof of Corollary 2 we show that α_{th} is in fact a lower bound on the excess distortion ratio in receiver 2 in the case that receiver 1 is optimal.

Appendix III
Proof of Equation (33)

We shall now prove (33). Let

$$\boldsymbol{Y}_2' \triangleq \boldsymbol{Y}_1 + \boldsymbol{Z}' \tag{93}$$

where $\boldsymbol{Z}' = Z_1', \ldots, Z_n'$ is memoryless with $Z_t' \sim \mathcal{N}(0, N_2 - N_1)$, and $\boldsymbol{Z}'$ is independent of $\boldsymbol{U}, \boldsymbol{X}$, and $\boldsymbol{Z_1}$. Define $\boldsymbol{Z}' = \boldsymbol{Z_1} + \boldsymbol{Z}'$. Hence, $\boldsymbol{Y}_2' = \boldsymbol{X} + \boldsymbol{Z}'$ where $\boldsymbol{Z}'$ is memoryless, zero mean, Gaussian, with variance N_2, and independent of $\boldsymbol{X}$. Additionally we have that

$$\boldsymbol{Y_2} = \boldsymbol{X} + \boldsymbol{Z_2} \tag{94}$$

where $\boldsymbol{Z_2}$ is also memoryless, zero mean, Gaussian, with variance N_2, and independent of $\boldsymbol{X}$. Now, since we have Markov chains $\boldsymbol{U} - \boldsymbol{X} - \boldsymbol{Y_2}$ and $\boldsymbol{U} - \boldsymbol{X} - \boldsymbol{Y}_2'$, we conclude that $f(\boldsymbol{y}_2' \,|\, \boldsymbol{u}) = f(\boldsymbol{y_2} \,|\, \boldsymbol{u})$ for all $(\boldsymbol{u}, \boldsymbol{y_2}, \boldsymbol{y}_2')$ and therefore,

$$h(\boldsymbol{Y_2} \,|\, \boldsymbol{U}) = h(\boldsymbol{Y}_2' \,|\, \boldsymbol{U}). \tag{95}$$

Now, by the conditional entropy power inequality [9], and since $\boldsymbol{Y}_2'$ is an independent sum of $\boldsymbol{Y_1}$ and $\boldsymbol{Z}'$, and $\boldsymbol{Z}'$ is Gaussian with variance $N_2 - N_1$, we have

$$2^{\frac{2}{n} h(\boldsymbol{Y}_2' \,|\, \boldsymbol{U})} \geq 2^{\frac{2}{n} h(\boldsymbol{Y_1} \,|\, \boldsymbol{U})} + 2^{\log(2\pi e (N_2 - N_1))}. \tag{96}$$

Combining (95) and (96) leads to (33).

Appendix IV
On the Quadratic Wyner–Ziv Rate Distortion Function of non-Gaussian Vectors

We shall now prove a general upper bound on the quadratic Wyner–Ziv rate-distortion function in terms of the MSE between the source and the side information. Consider a source–side information vector pair $(\boldsymbol{S}', \boldsymbol{U}')$ of length m, where $\boldsymbol{S}'$ and $\boldsymbol{U}'$ are not necessarily Gaussian and not necessarily memoryless. The quadratic distortion between $\boldsymbol{S}'$ and $\boldsymbol{U}'$ is defined as

$$d(\boldsymbol{S}', \boldsymbol{U}') \triangleq \frac{1}{m} \sum_{t=1}^{m} E(S_t' - U_t')^2.$$

We shall now show that for any such source–side information vector pair $(\boldsymbol{S}', \boldsymbol{U}')$ the quadratic Wyner–Ziv rate-distortion function satisfies

$$R^{WZ}_{\boldsymbol{S}'|\boldsymbol{U}'}(D) \le \frac{1}{2}\log\frac{d(\boldsymbol{S}',\boldsymbol{U}')}{D} \tag{97}$$

where D is the allowed distortion.

Proof: Let the pair $(\boldsymbol{S}^*, \boldsymbol{U}^*)$ be Gaussian with the same first- and second-order statistics as $(\boldsymbol{S}', \boldsymbol{U}')$, i.e., $(\boldsymbol{S}^*, \boldsymbol{U}^*) \sim \mathcal{N}(E(\boldsymbol{S}', \boldsymbol{U}'), \mathrm{Cov}(\boldsymbol{S}', \boldsymbol{U}'))$. Let $\boldsymbol{Z}^*$ be an independently distributed Gaussian vector satisfying for every t

$$\mathrm{Var}(S'_t \,|\, S'_t + Z_{t^*}, U'_t) = D. \tag{98}$$

We have

$$\begin{aligned}
\frac{1}{2}\log\frac{d(\boldsymbol{S}',\boldsymbol{U}')}{D} &\ge \frac{1}{2m}\sum_{t=1}^{m}\log\frac{E(S'_t - U'_t)^2}{D} && (99)\\
&\ge \min_a \frac{1}{2m}\sum_{t=1}^{m}\log\frac{E(S'_t - aU'_t)^2}{D} && (100)\\
&= \min_a \frac{1}{2m}\sum_{t=1}^{m}\log\frac{E(S^*_t - aU^*_t)^2}{D} && (101)\\
&= \frac{1}{2m}\sum_{t=1}^{m}\log\frac{\mathrm{Var}(S_{t^*}\,|\,U_{t^*})}{D} && (102)\\
&= \frac{1}{2m}\sum_{t=1}^{m}\log\frac{\mathrm{Var}(S_{t^*}\,|\,U_{t^*})}{\mathrm{Var}(S'_t\,|\,S'_t+Z_{t^*},U'_t)} && (103)\\
&\ge \frac{1}{2m}\sum_{t=1}^{m}\log\frac{\mathrm{Var}(S_{t^*}\,|\,U_{t^*})}{\mathrm{Var}(S^*_t\,|\,S^*_t+Z_{t^*},U^*_t)} && (104)\\
&= \frac{1}{m}\sum_{t=1}^{m} I(S_{t^*}; S^*_t + Z^*_t \,|\, U^*_t) && (105)\\
&\ge \frac{1}{m}\sum_{t=1}^{m} I(S'_t; S'_t + Z^*_t \,|\, U'_t) && (106)\\
&\ge \frac{1}{m} I(\boldsymbol{S}'; \boldsymbol{S}' + \boldsymbol{Z}^* \,|\, \boldsymbol{U}') && (107)\\
&\ge R^{WZ}_{\boldsymbol{S}'|\boldsymbol{U}'}(D). && (108)
\end{aligned}$$

The preceding sequence of inequalities and equalities is now explained. Equation (99) follows by Jensen's inequality, (100) follows as $a = 1$ gives a larger value than the minimal a, (101) and (102) follow since the optimal square error in linear estimation is the same for Gaussian and non-Gaussian variables with the same second moments and equals to the conditional variance, and (103) follows by the definition of $\boldsymbol{Z}^*$. As for (104), it follows since $\mathrm{Var}(S^*_t \,|\, S^*_t + Z_{t^*}, U^*_t)$, the Gaussian conditional variance, equals the MSE of the best *linear* estimator of S'_t given $S'_t + Z_{t^*}$ and U'_t, which is larger than the optimal MSE of any estimator given by the conditional variance of S'_t given $S'_t + Z_{t^*}$ and U'_t. Equation (105) follows directly from the expression of the mutual information of Gaussian variables. To see (106), we have $I(S'_{t^*}; S_{t^*} + Z_{t^*} \,|\, U_{t^*}) = h(S_{t^*} + Z_{t^*} \,|\, U_{t^*}) - h(Z_{t^*})$, but

$$\begin{aligned}
&h(S_{t^*} + Z_{t^*} \,|\, U_{t^*})\\
&\quad= \frac{1}{2}\log\left[2\pi e(\mathrm{Var}(S_{t^*}\,|\,U_{t^*}) + \mathrm{Var}(Z_{t^*}))\right] && (109)\\
&\quad\ge \frac{1}{2}\log\left[2\pi e(\mathrm{Var}(S'_t\,|\,U'_t) + \mathrm{Var}(Z_{t^*}))\right] && (110)\\
&\quad\ge \frac{1}{2}E\left\{\log\left[2\pi e(\mathrm{Var}(S'_t\,|\,U'_t = u') + \mathrm{Var}(Z_{t^*}))\right]\right\} && (111)\\
&\quad= h(S'_t + Z_{t^*} \,|\, U'_t) && (112)
\end{aligned}$$

where (109) follows from the definition of conditional entropy and the fact that Z_{t^*} is independent of S_{t^*}, U_{t^*}, (110) follows, as above, since the non-Gaussian conditional variance is smaller than the Gaussian conditional variance, (111) is by Jensen's inequality, and (112) comes from the definition of the conditional entropy, using the fact that Z_{t^*} is independent of S'_t, U'_t. As for (107) we have

$$I(S'_t; S'_t + Z_{t^*} \,|\, U'_t) = h(S'_t + Z_{t^*} \,|\, U'_t) - h(Z_{t^*})$$

but since conditioning reduces entropy we have

$$h(S'_t + Z_{t^*}|U'_t) \ge h(S'_t + Z_{t^*}|S'_1,\ldots,S'_{t-1},U'_1,\ldots,U'_t,\ldots,U'_m)$$

and so

$$\sum_{t=1}^{m} h(S'_t + Z_{t^*} \,|\, U'_t) \ge h(\boldsymbol{S}' + \boldsymbol{Z}^* \,|\, \boldsymbol{U}')$$

while $\sum_{t=1}^{m} h(Z_{t^*}) = h(\boldsymbol{Z}^*)$. Finally, (108) follows since (107) represents the mutual information of a specific test channel that satisfies the distortion constrains, while $R^{WZ}_{\boldsymbol{S}'|\boldsymbol{U}'}(D)$ is the *minimal* mutual information over all possible tests channels. □

Note that we can show by a straightforward extension of the proof of the direct part of the Wyner–Ziv coding theorem, that for any stationary and ergodic source, the vector form of the Wyner–Ziv function can be arbitrarily approached by a coding system operating on "super source symbols." Hence, it is possible to design an encoder–decoder pair with rate $1/2\log d(\boldsymbol{S}', \boldsymbol{U}')/D$ and distortion D.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[2] M. D. Trott, "Unequal error protection codes: Theory and practice," in *Proc. IEEE Information Theory Workshop*, Haifa, Israel, Jun. 1996, p. 11.

[3] B. Chen and G. Wornell, "Analog error-correcting codes based on chaotic dynamical systems," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 881–890, Jul. 1998.

[4] U. Mittal and N. Phamdo, "Joint source-channel codes for broadcasting and robust communication," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1082–1103, May 2002.

[5] S. Shamai (Shitz), S. Verdú, and R. Zamir, "Systematic lossy source-channel coding," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 564–579, Mar. 1998.

[6] V. A. Vaishampayan and S. I. R. Costa, "Curves on a sphere: Error control for continuous alphabet sources," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 376.

[7] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.

[8] J. Ziv, "The behavior of analog communication systems," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 5, pp. 587–594, Sep. 1970.

[9] N. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 2, pp. 267–271, Apr. 1965.

[10] Z. Reznic, R. Zamir, and M. Feder, "Joint source-channel coding of a Gaussian-mixture source over the Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 776–781, Mar. 2002.

[11] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[12] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder—II: General sources," *Inf. Contr.*, vol. 38, pp. 60–80, 1978.

[13] Z. Reznic, "Broadcasting analog sources over Gaussian channels," Ph.D. dissertation, Tel-Aviv Univ., Tel-Aviv, Israel, Aug. 2003.

[14] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattce codes for structural multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.