Distortion Bounds for Broadcasting with Bandwidth Expansion *

Zvi Reznic, Meir Feder and Ram Zamir

Dept. of EE-Systems, Tel-Aviv University, Israel
e-mails: zvi@amimon.com, meir,zamir@eng.tau.ac.il

January 22, 2005

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 paper 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 (high) SNR_{min}, then as the SNR improves, the distortion cannot decay faster than $1/SNR$. 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. Additionally, we show how a partial feedback, returning from the bad receiver to the transmitter and to the good receiver, can improve the performance beyond that of the proposed schemes. Interestingly, the distortion pair achieved with this feedback lies on the outer bound derived here.

Index terms - distortion region, joint source-channel coding, lossy broadcasting

*The material in this paper was presented in part at the 40th Annual Allerton Conference on Communication, Control and Computing, Oct. 2002.

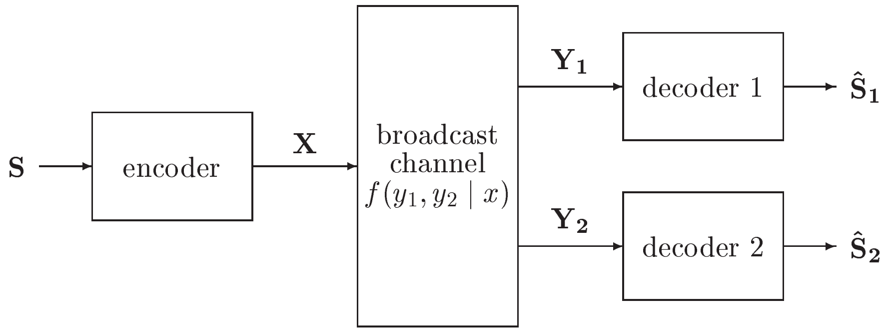

Figure 1: Lossy transmission of a source through a broadcast channel

1 Introduction

The *broadcast channel*, illustrated in Figure 1, is a communication channel in which one sender transmits to two or more receivers. 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 yet open [1]. We recall that in the *channel coding* problem for broadcast channels, the capacity region depends only on the marginal distributions of the channel [2, page 422]. We shall show in Appendix I that the same is true for the distortion region.

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, 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 a 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 $\rho > 1$.

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) [2], 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

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

Other schemes were developed for the case of $\rho > 1$, providing other achievable distortion pairs [3, 5, 6]. However, no non-trivial outer bound (converse) on the distortion region was ever derived. The main result of this paper 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 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, whose concept resembles that of Chen and Wornell [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, the distortion cannot decay faster than $1/SNR$ for all values of ρ. For comparison, we recall that the solution of $R(D') = \rho C$ is given by $D' = \sigma^2/(1 + SNR)^\rho$, and hence, the MSE of a collection of systems, each optimally designed for a different (high) SNR decays as $1/SNR^{\rho}$. We note that our result is stronger than a previous result by Ziv [8], who showed that asymptotically, the distortion cannot decay faster than $1/SNR^2$ for all values of ρ.

2 Outer bound on the distortion region

We shall now derive the outer bound. We denote the source by $\mathbf{S} = (S_1, \ldots, S_m)$, and the decoders output by $\hat{\mathbf{S}}_\mathbf{1} = (\hat{S}_{1,1}, \ldots, \hat{S}_{1,m})$ and $\hat{\mathbf{S}}_\mathbf{2} = (\hat{S}_{2,1}, \ldots, \hat{S}_{2,m})$. We denote the channel input by $\mathbf{X} = (X_1, \ldots, X_n)$ and the channel outputs by $\mathbf{Y}_1 = (Y_{1,1}, \ldots, Y_{1,n})$ and $\mathbf{Y}_2 = (Y_{2,1}, \ldots, Y_{2,n})$. 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* $\mathbf{X} = (X_1, \ldots, X_n)$ *and outputs* $\mathbf{Y}_1 = (Y_{1,1}, \ldots, Y_{1,n})$ *and* $\mathbf{Y}_2 = (Y_{2,1}, \ldots, Y_{2,n})$*, satisfies for* $i = 1, 2:$

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

where $Z_{1,t}, Z_{2,t}$ *are memoryless and statistically independent of* X_t*, and* $N_2 \geq N_1$*.*

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

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

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

Definition 2 (D_1, D_2) *is an* <u>*achievable distortion pair*</u> *if, for any* $\epsilon^* > 0$*, there exist integers* m *and* $n = \rho m$*, an encoding function* $\mathbf{X} = i_m^n(\mathbf{S})$ *and reconstruction functions* $\hat{\mathbf{S}}_\mathbf{1} = g_{1m}^n(\mathbf{Y_1})$ *and* $\hat{\mathbf{S}}_\mathbf{2} = g_{2m}^n(\mathbf{Y_2})$*, such that*

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

The <u>*achievable distortion region*</u> *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 non-rational value could be replaced by a rational value which is arbitrary close to it.

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

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

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 $\alpha \geq 1$ be defined by*

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

Then

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

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 in section 3 and we shall now outline the properties of the function $f(\kappa)$, in order to shed some light on the RHS of (9). Examples of $f(\kappa)$ are illustrated in Figures 2-4.

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

Property 2 *If $\alpha > 1$ then*

$$\lim_{\kappa\to 0} f(\kappa) = \infty. \tag{11}$$

Property 3 *If $\alpha = 1$ then*

$$\lim_{\kappa \to 0} f(\kappa) = \left(\frac{P + N_2}{N_2}\right)^{\rho - 1} \frac{P + N_2}{N_1} - \frac{N_2 - N_1}{N_1}. \tag{12}$$

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

$$\lim_{\kappa \to \infty} f(\kappa) = \left(1 + \frac{P}{N_1}\right)^{\rho} = 2^{2\rho C_1}. \tag{13}$$

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

$$\frac{\partial f(\kappa)}{\partial \kappa} = \frac{g(\kappa)}{\kappa^2}, \tag{14}$$

where

$$g(\kappa) = \frac{h_1(\kappa) h_2(\kappa)}{N_1^{\rho}} + 1, \tag{15}$$

where

$$h_1(\kappa) = \left\{ N_2 \left[\frac{\alpha}{\kappa} + \left(1 + \frac{P}{N_2}\right)^{\rho} \right]^{1/\rho} - (N_2 - N_1) \left(\frac{1}{\kappa} + 1\right)^{1/\rho} \right\}^{\rho - 1}, \tag{16}$$

and

$$h_2(\kappa) = N_2(-\alpha) \left[\frac{\alpha}{\kappa} + \left(1 + \frac{P}{N_2}\right)^{\rho} \right]^{1/\rho - 1} + (N_2 - N_1) \left(\frac{1}{\kappa} + 1\right)^{1/\rho - 1}. \tag{17}$$

Property 6 *If follows from Property 5 that*

$$\lim_{\kappa \to \infty} \frac{\partial f(\kappa)}{\partial \kappa} = 0. \tag{18}$$

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

$$\alpha > \alpha_{th} \triangleq \left(1 + \frac{P}{N_2}\right)^{\rho - 1} \left[\frac{N_1}{N_2} \left(\frac{N_1}{P + N_1}\right)^{\rho - 1} + \frac{N_2 - N_1}{N_2} \right], \tag{19}$$

where $g(\kappa)$ was defined in (15).

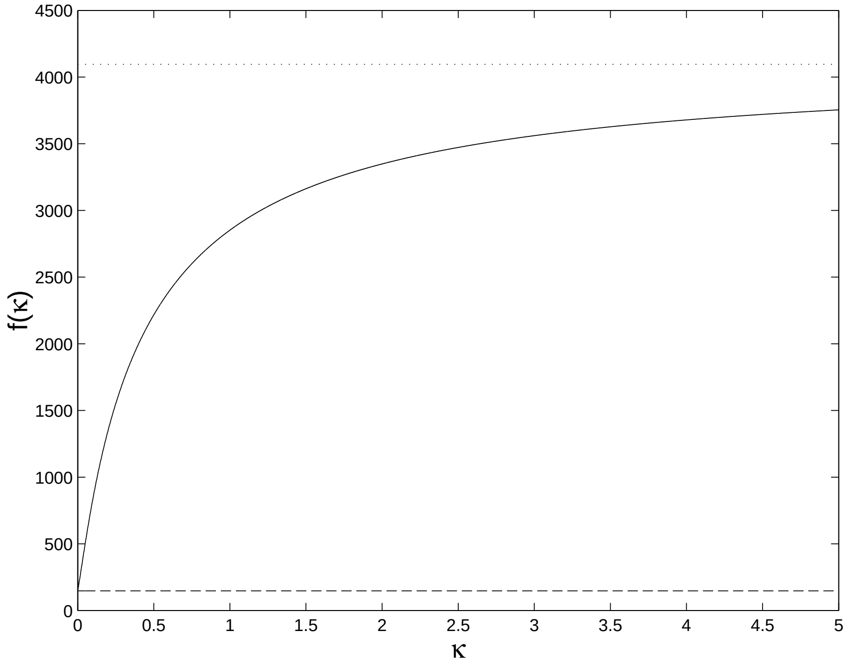

Figure 2: $f(\kappa)$ in the case of $\alpha = 1$. Solid: $f(\kappa)$, dotted: the limit of $f(\kappa)$ as $\kappa \to \infty$ according to Property 4, dashed: the limit of $f(\kappa)$ as $\kappa \to 0$ according to Property 3. Note that the slope of $f(\kappa)$ approaches zero as $\kappa \to \infty$ according to Property 6. (Parameters: $P = 0.15$, $N_1 = 0.01$, $N_2 = 0.1$, $\rho = 3$)

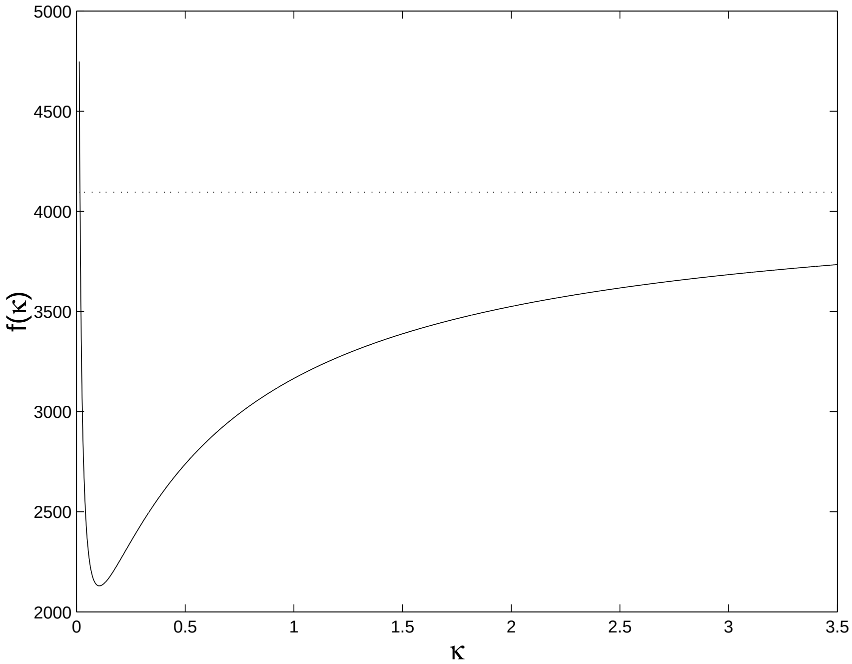

Figure 3: $f(\kappa)$ in the case of $1 < \alpha < \alpha_{th}$. Solid: $f(\kappa)$, dotted: the limit of $f(\kappa)$ as $\kappa \to \infty$. Note that $f(\kappa) \to \infty$ as $\kappa \to 0$ according to Property 2, and that the minimal value of $f(\kappa)$ is smaller than its asymptotic value at $\kappa \to \infty$, according to Properties 5-7. (Parameters: $P = 0.15$, $N_1 = 0.01$, $N_2 = 0.1$, $\rho = 3$, $\alpha = 2$, $\alpha_{th} = 5.63$)

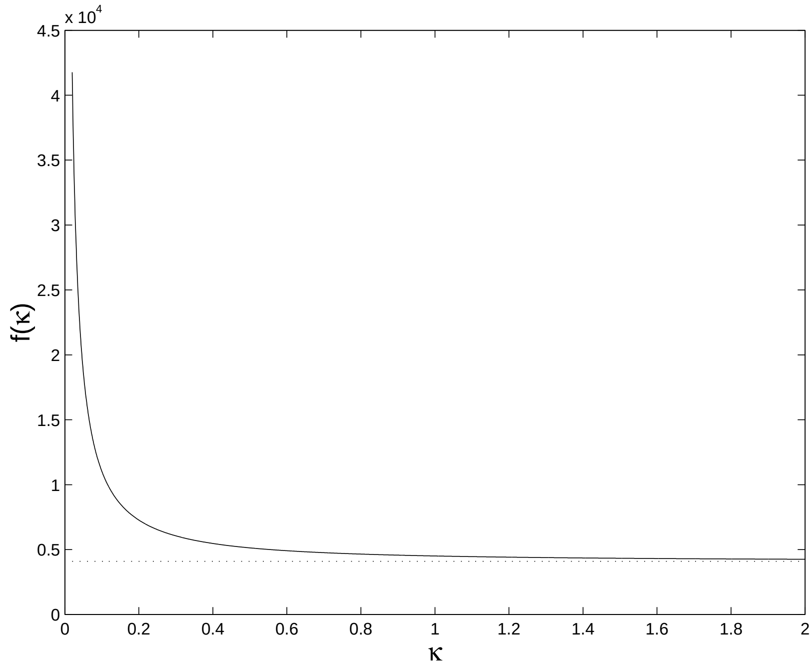

Figure 4: $f(\kappa)$ in the case of $\alpha > \alpha_{th}$. Solid: $f(\kappa)$, dotted: the limit of $f(\kappa)$ as $\kappa \to \infty$. Note that $f(\kappa)$ is always larger than its asymptotic value at $\kappa \to \infty$, according to Properties 5-7. (Parameters: $P = 0.15$, $N_1 = 0.01$, $N_2 = 0.1$, $\rho = 3$, $\alpha = 6$, $\alpha_{th} = 5.63$)

We shall show in the proof of Corollary 2 that α_{th} is in fact a lower bound on the excess distortion ratio which is possible when receiver 1 is optimal. Figures 2-4 demonstrate the properties of $f(\kappa)$.

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 was 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*

$$D_2 = R^{-1}(\rho C_2). \tag{20}$$

Then

$$D_1 \geq \sigma^2 \left(\left(\frac{P+N_2}{N_2} \right)^{\rho-1} \frac{P+N_2}{N_1} - \frac{N_2-N_1}{N_1} \right)^{-1}. \tag{21}$$

Proof: 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 (20) and (8) we have that $\alpha = 1$. Combining this with Property 3 proves the theorem. □

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

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

and

$$\begin{aligned} D_1' &= R^{-1}\left((\rho-1)C_2 + C_1\right) && (22) \\ &= \sigma^2 \left(\left(\frac{P+N_2}{N_2} \right)^{\rho-1} \frac{P+N_1}{N_1} \right)^{-1}. && (23) \end{aligned}$$

Comparing this with (20) and (21), we see that their scheme is asymptotically optimal in the case of high SNR ($P/N_2 \to \infty$). Additionally, it can be shown that their scheme is optimal in the limit of $N_2 \to \infty$, although this case is less interesting.

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*

$$D_1 = R^{-1}(\rho C_1). \tag{24}$$

Then

$$D_2 \geq R^{-1}(C_2)\left[1 - \frac{N_1}{N_2} + \frac{N_1}{N_2}\left(\frac{N_1}{P+N_1}\right)^{\rho-1}\right]. \tag{25}$$

Proof: By Property 4 we have that

$$\lim_{\kappa\to\infty} \frac{1}{2}\log f(\kappa) = \rho C_1 \tag{26}$$

Hence, the requirement set by (24) can be written as

$$R(D_1) = \lim_{\kappa\to\infty} \frac{1}{2}\log f(\kappa). \tag{27}$$

Using (1), we can write (27) as:

$$D_1 = \lim_{\kappa\to\infty} \frac{\sigma^2}{f(\kappa)}. \tag{28}$$

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 (24).) 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} (19) with (1), (5) and (8) proves the Corollary. □

For comparison, the scheme of Shamai, Verdú and Zamir (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 $N_1/N_2 \to 0$.

3 Proof of Theorem 1

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

$$V_t \sim \mathcal{N}(0, \kappa\sigma^2), \quad \text{and} \quad U_t = S_t + V_t \quad (t = 1, \ldots, m), \tag{29}$$

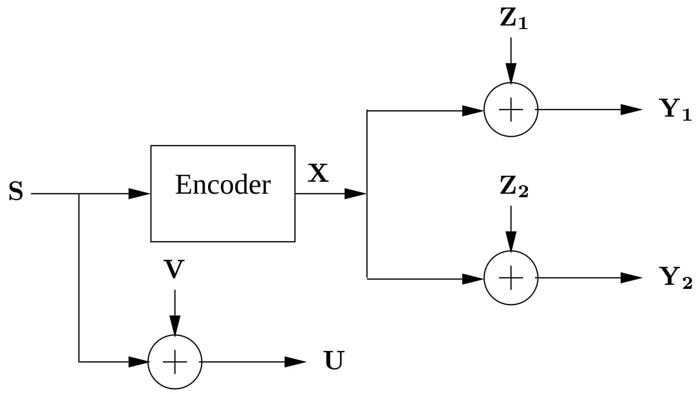

Figure 5: The Gaussian broadcast channel with the auxiliary variable $\mathbf{U}$.

where $\kappa > 0$. Hence we have Markov chains $\mathbf{U} \leftrightarrow \mathbf{S} \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y}_i \leftrightarrow \hat{\mathbf{S}}_i$ for $i = 1, 2$. (See figure 5). By the chain rule for mutual information we have for $(i = 1, 2)$:

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y_1}) &= I(\mathbf{X}; \mathbf{U}) + I(\mathbf{X}; \mathbf{Y_1}|\mathbf{U}) - I(\mathbf{X}; \mathbf{U}|\mathbf{Y_1}) && (30) \\ &= I(\mathbf{X}; \mathbf{U}) + h(\mathbf{Y_1}|\mathbf{U}) - h(\mathbf{Z_1}) - I(\mathbf{X}; \mathbf{U}|\mathbf{Y_1}) && (31) \\ &= I(\mathbf{X}; \mathbf{U}) + h(\mathbf{Y_1}|\mathbf{U}) - h(\mathbf{Z_1}) - h(\mathbf{U}|\mathbf{Y_1}) + h(\mathbf{U}|\mathbf{X}, \mathbf{Y_1}) \\ &= I(\mathbf{X}; \mathbf{U}) + h(\mathbf{Y_1}|\mathbf{U}) - h(\mathbf{Z_1}) - h(\mathbf{U}|\mathbf{Y_1}) + h(\mathbf{U}|\mathbf{X}) && (32) \\ &= h(\mathbf{U}) - h(\mathbf{U}|\mathbf{Y_1}) + h(\mathbf{Y_1}|\mathbf{U}) - h(\mathbf{Z_1}) && (33) \end{aligned}$$

where (31) follows from (4), and in (32) we used the Markov chain relation to replace $h(\mathbf{U}|\mathbf{X}, \mathbf{Y_1})$ with $h(\mathbf{U}|\mathbf{X})$.

In (33) $I(\mathbf{X}; \mathbf{Y_1})$ is expressed as a sum of four terms. We shall now upper bound $I(\mathbf{X}; \mathbf{Y_1})$ by bounding those terms. First, we note that $\mathbf{U}$ and $\mathbf{Z_1}$ are Gaussian memoryless vectors, where $\mathbf{U}$ has variance $(\kappa+1)\sigma^2$ and length m, and $\mathbf{Z_1}$ has variance N_1 and length n. Hence, their differential entropies [2] are given by :

$$\begin{aligned} h(\mathbf{U}) &= \frac{m}{2} \log 2\pi e(\kappa + 1)\sigma^2 && (34) \\ \text{and} \quad h(\mathbf{Z_1}) &= \frac{n}{2} \log 2\pi e N_1. && (35) \end{aligned}$$

We shall now lower bound the second term in (33), which is $h(\mathbf{U}|\mathbf{Y_1})$. By the conditional form of the entropy power inequality [9], and since $\mathbf{U}$ is the independent sum of $\mathbf{S}$ and $\mathbf{V}$ we have that:

$$\begin{aligned}
2^{\frac{2}{m}h(\mathbf{U}|\mathbf{Y_1})} &\geq 2^{\frac{2}{m}h(\mathbf{S}|\mathbf{Y_1})} + 2^{\frac{2}{m}h(\mathbf{V}|\mathbf{Y_1})} && (36)\\
&= 2^{\frac{2}{m}h(\mathbf{S}|\mathbf{Y_1})} + 2^{\frac{2}{m}h(\mathbf{V})} && (37)\\
&= 2^{\frac{2}{m}h(\mathbf{S}|\mathbf{Y_1})} + 2\pi e\kappa\sigma^2, && (38)
\end{aligned}$$

where (38) follows since $\mathbf{V}$ is Gaussian [2]. The term $h(\mathbf{S}|\mathbf{Y_1})$ in (38) can be further bounded as follow:

$$\begin{aligned}
h(\mathbf{S}|\mathbf{Y_1}) &= h(\mathbf{S}) - I(\mathbf{S};\mathbf{Y_1}) && (39)\\
&= \frac{m}{2}\log 2\pi e\sigma^2 - I(\mathbf{S};\mathbf{Y_1}) && (40)\\
&\geq \frac{m}{2}\log 2\pi e\sigma^2 - I(\mathbf{X};\mathbf{Y_1}), && (41)
\end{aligned}$$

where (40) is since $\mathbf{S}$ is Gaussian and (41) is by the data processing inequality. Combining (38) and (41) with the fact that $\rho = m/n$ yields:

$$h(\mathbf{U}|\mathbf{Y_1}) \geq \frac{m}{2}\log\left(2\pi e\sigma^2\left(2^{-\frac{2\rho}{n}I(\mathbf{X};\mathbf{Y_1})} + \kappa\right)\right). \tag{42}$$

We shall now upper bound the third term in (33), which is $h(\mathbf{Y_1}|\mathbf{U})$. We note that $\mathbf{Y_2}$ is the sum of $\mathbf{Y_1}$ 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 II) that:

$$2^{\frac{2}{n}h(\mathbf{Y_2}|\mathbf{U})} \geq 2^{\frac{2}{n}h(\mathbf{Y_1}|\mathbf{U})} + 2^{\log(2\pi e(N_2-N_1))}. \tag{43}$$

(Note that a similar derivation was done in [10].) The LHS of (43) can be expressed as follow:

$$2^{\frac{2}{n}h(\mathbf{Y_2}|\mathbf{U})} = 2^{\frac{2}{n}(h(\mathbf{Y_2})-I(\mathbf{Y_2};\mathbf{U}))} \leq 2\pi e(P+N_2)2^{-\frac{2}{n}I(\mathbf{Y_2};\mathbf{U})}, \tag{44}$$

where we used the fact that the variance of $\mathbf{Y_2}$ is $P + N_2$, and hence its differential entropy cannot exceed $\frac{n}{2}\log(2\pi e(P+N_2))$ [2, page 262]. Note that the combination of (43) and (44) can serve as an upper bound for $h(\mathbf{Y_1}|\mathbf{U})$ in terms of $I(\mathbf{Y_2};\mathbf{U})$. We shall now use rate distortion theory to derive a lower bound on $I(\mathbf{Y_2};\mathbf{U})$. Using (5), we can rewrite (8) as:

$$D_2 = \alpha\sigma^2 2^{-2\rho C_2} = \frac{\alpha\sigma^2}{(1+P/N_2)^\rho} = \alpha\sigma^2\left(\frac{N_2}{P+N_2}\right)^\rho. \tag{45}$$

We have:

$$
\begin{aligned}
E\left(d(\hat{\mathbf{S}}_2, \mathbf{U})\right) &= E\left(\frac{1}{m}\sum_{t=1}^{m}(\hat{S}_{2,t} - U_t)^2\right) && (46)\\
&= E\left(\frac{1}{m}\sum_{t=1}^{m}(\hat{S}_{2,t} - S_t + S_t - U_t)^2\right) && (47)\\
&= E\left(\frac{1}{m}\sum_{t=1}^{m}(\hat{S}_{2,t} - S_t)^2\right) + E\left(\frac{1}{m}\sum_{t=1}^{m}(S_t - U_t)^2\right) && (48)\\
&= D_2 + E\left(\frac{1}{m}\sum_{t=1}^{m}V_t^2\right) && (49)\\
&= \alpha\sigma^2\left(\frac{N_2}{P+N_2}\right)^{\rho} + \kappa\sigma^2, && (50)
\end{aligned}
$$

where (48) follows since $S_t - U_t = V_t$ is independent of $\hat{S}_{2,t} - S_t$, (49) follows from (7) and (29), and (50) follows from (29) and (45). We now have:

$$
\begin{aligned}
\frac{1}{n}I(\mathbf{Y}_2; \mathbf{U}) &\geq \frac{1}{n}I(\hat{\mathbf{S}}_2; \mathbf{U}) && (51)\\
&\geq \frac{1}{n}mR(Ed(\hat{\mathbf{S}}_2; \mathbf{U})) && (52)\\
&\geq \frac{1}{2\rho}\log\frac{(\kappa+1)\sigma^2}{\alpha\sigma^2\left(\frac{N_2}{P+N_2}\right)^{\rho} + \kappa\sigma^2}, && (53)
\end{aligned}
$$

where (51) is by the data processing inequality, (52) is by rate-distortion theory, and (53) follows since $\mathbf{U}$ is Gaussian with variance $(\kappa+1)\sigma^2$, and by (1) and (50). Combining (43), (44) and (53) yields:

$$
h(\mathbf{Y_1}|\mathbf{U}) \leq \frac{n}{2}\log\left(2\pi e(P+N_2)\left(\frac{\alpha\left(\frac{N_2}{P+N_2}\right)^{\rho} + \kappa}{\kappa+1}\right)^{1/\rho} - 2\pi e(N_2 - N_1)\right). \tag{54}
$$

Hence, we have bounded all four terms in (33). Combining these terms, that is, combining (33), (34), (35), (42) and (54) yields:

$$
\frac{1}{n}I(\mathbf{X}; \mathbf{Y_1}) \leq \frac{1}{2}\log\frac{(P+N_2)\left(\frac{\alpha}{\kappa}\left(\frac{N_2}{P+N_2}\right)^{\rho} + 1\right)^{1/\rho} - (N_2 - N_1)\left(\frac{\kappa+1}{\kappa}\right)^{1/\rho}}{N_1\left(\frac{1}{\kappa}2^{-2\frac{\rho}{n}I(\mathbf{X};\mathbf{Y_1})} + 1\right)^{1/\rho}}, \tag{55}
$$

for all $\kappa > 0$. Algebraic manipulation of (55) yields:

$$\frac{1}{n} I(\mathbf{X}; \mathbf{Y_1}) \le \frac{1}{2\rho} \log f(\kappa) \tag{56}$$

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

By rate distortion theory, by the data processing inequality, and by (56) we have that if (D_1, D_2) is achievable than

$$\frac{1}{\rho} R(D_1) \le \frac{1}{n} I(\mathbf{X}; \mathbf{Y_1}) \le \frac{1}{2\rho} \log f(\kappa) \tag{57}$$

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

4 Transmission over Channels with Unknown SNR

We now turn to the issue of lossy transmission over a 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 $SNR_{min} \triangleq P/N_2$ and $SNR \triangleq P/N_1$ and by (21) we prove the following corollary:

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

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

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

$$\Phi = \left(\frac{1}{SNR_{min}}\right)^{\rho - 1},$$

and $o(1) \to 0$ as $SNR_{min} \to \infty$.

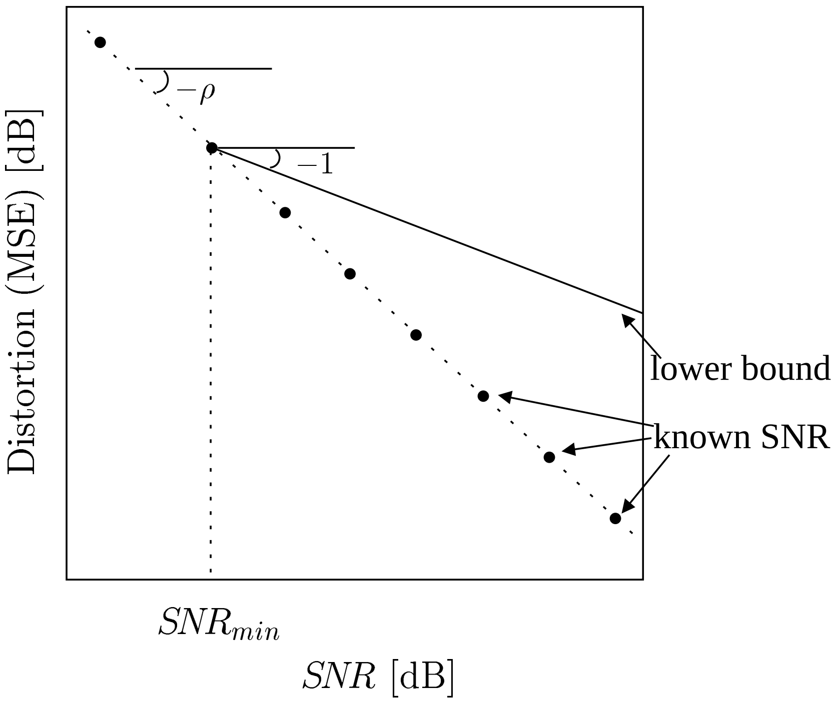

Figure 6: MSE vs. SNR. Solid: the lower bound of Corollary 3. Dotted: the solution of $R(D) = \rho C(SNR)$

Figure 6 illustrates the results of Corollary 3 in the case of high 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

$$R(D) = \rho C(SNR).$$

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

It is interesting to compare these results to a previous result of Ziv who analyzed the same problem [8]. Our result is stronger than Ziv's result, since we showed that the distortion cannot decay faster than $1/SNR$, while Ziv showed that it cannot decay faster than $1/SNR^2$. Additinally, we bounded the performance of *any* system, while Ziv restricted his result to a class of systems, which he called "practical".

5 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 Figure 7. The transmission block $\mathbf{X}$ of length $n = \rho m$ is generated by concatenating (i.e. multiplexing in time) a "digital" block $\mathbf{X_D}$ of length $(\rho-1)m$, and an "analog" block $\mathbf{X_A}$ of length m. The digital block is generated by a broadcast channel transmitter [2], such that a common message W_2 is losslessly sent to 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 [2, page 380]):

$$R_1 = \frac{1}{2}\log\left(1+\frac{(1-\beta)P}{N_1}\right) - \epsilon \tag{58}$$

$$\text{and} \quad R_2 = \frac{1}{2}\log\left(1+\frac{\beta P}{N_2+(1-\beta)P}\right) - \epsilon. \tag{59}$$

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 Figure 7. 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 VQ such that it achieves

$$(\rho-1)R_2 = R(D_Q) + \epsilon_1, \tag{60}$$

where $R(D_Q)$ is measured in bits per source sample. We denote the VQ output by $\mathbf{S_Q} = (S_{Q1}, \ldots, S_{Qm})$. 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 $\mathbf{E} = (E_1, \ldots, E_m)$ is defined as $E_t = S_{Qt} - S_t$. Each sample in $\mathbf{E}$ is scaled by a scalar K to produce $\mathbf{X_A}$.

The message W_2 is an integer which uniquely describes the vector $\mathbf{S_Q}$. Since the length of 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 $\mathbf{S_Q}$ losslessly. Hence, the problem reduces to that of lossy transmission of $\mathbf{E}$, whose variance is D_Q.

Let (D_1', D_2') be the distortion pair which is achievable by our scheme. Referring again to Figure 7, we denote by $\mathbf{Y_{D1}}$ and $\mathbf{Y_{D2}}$ the noisy outputs of the broadcast channel, in response to the input $\mathbf{X_D}$, and by $\mathbf{Y_{A1}}$ and $\mathbf{Y_{A2}}$ the noisy outputs of the broadcast channel, in response to the input $\mathbf{X_A}$. Receiver 2 estimates $\mathbf{E}$ by multiplying the input $\mathbf{Y_{A2}}$ by a gain factor K_2. By setting $K = \sqrt{\frac{P}{D_Q}}$ and $K_2 = \frac{\sqrt{PD_Q}}{P+N_2}$, and taking the limit as $\epsilon \to 0$ and $\epsilon_1 \to 0$ we have:

$$\begin{aligned} D_2' &= \frac{D_Q}{1+\frac{P}{N_2}} & (61) \\ &= \frac{N_2}{P+N_2} R^{-1}\left((\rho-1)R_2\right) & (62) \\ &= \frac{\sigma^2 N_2}{P+N_2} 2^{-2(\rho-1)R_2} & (63) \\ &= \frac{\sigma^2 N_2}{P+N_2}\left(1+\frac{\beta P}{N_2+(1-\beta)P}\right)^{-(\rho-1)}, & (64) \end{aligned}$$

where (61) follows from standard MSE calculations, (62) is by (60), (63) is by (1) and (64) is by (59).

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 $\hat{\mathbf{S}}_{\mathbf{1}}^*$. Let D_1^* be the average distortion between $\mathbf{S}$ and $\hat{\mathbf{S}}_{\mathbf{1}}^*$. Repeating the steps that led to (64) one can verify that

$$D_1^* = \frac{\sigma^2 N_1}{P+N_1}\left(1+\frac{\beta P}{N_2+(1-\beta)P}\right)^{-(\rho-1)}. \quad (65)$$

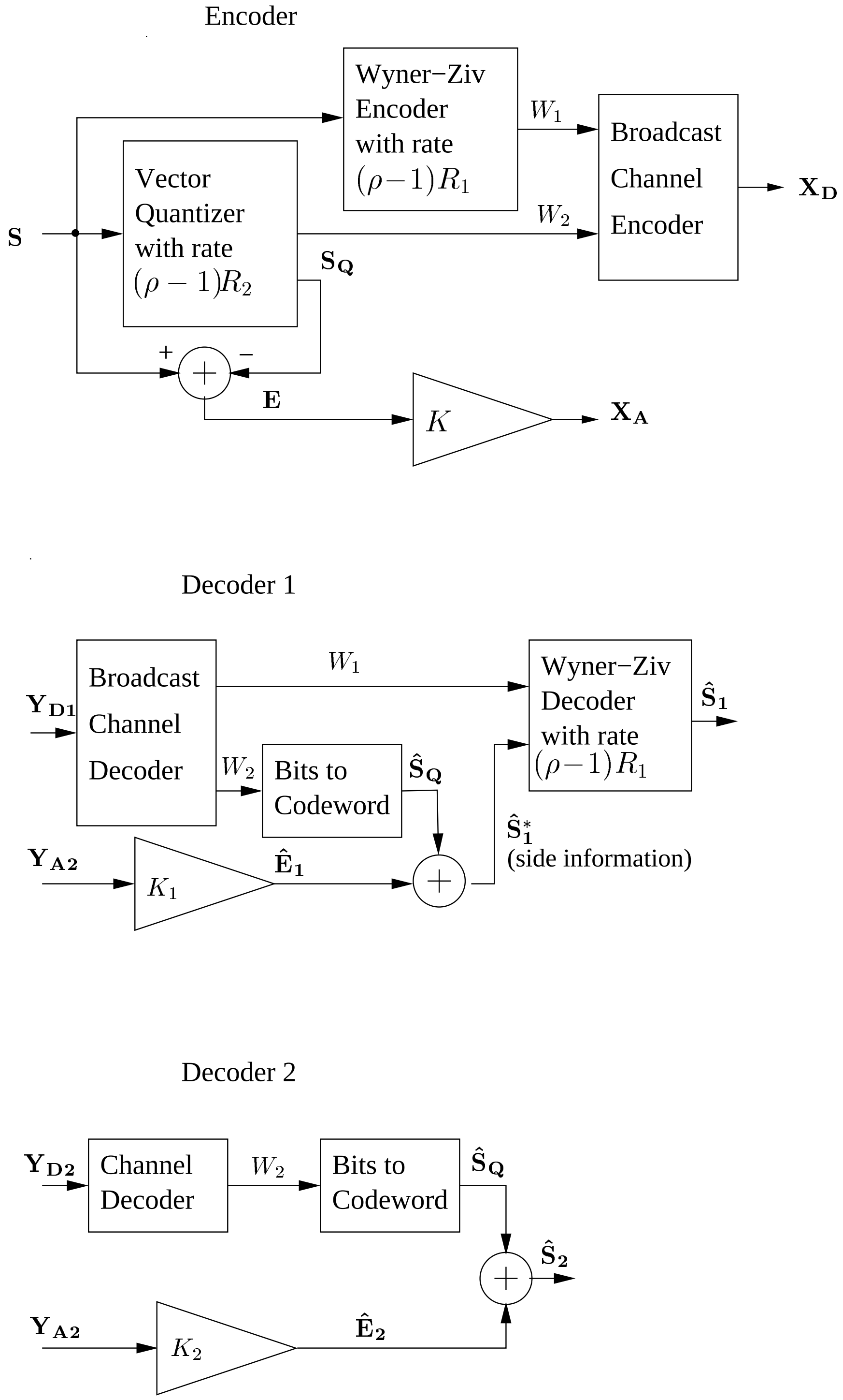

Figure 7: A coding scheme for lossy transmission with bandwidth expansion

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 $\mathbf{S}$, taking into account that the decoder already has side information $\hat{\mathbf{S}}_1^*$. This is in fact the Wyner-Ziv problem [11, 12]. Fortunately, since the source $\mathbf{S}$ is Gaussian, the Wyner-Ziv result ensures that we can achieve (as $\epsilon \to 0$):

$$\begin{aligned} D_1' &= D_1^* \cdot 2^{-2(\rho-1)R_1} \\ &= \frac{\sigma^2 N_1}{P+N_1} \left\{ \left(1 + \frac{\beta P}{N_2 + (1-\beta)P}\right) \left(1 + \frac{(1-\beta)P}{N_1}\right) \right\}^{-(\rho-1)}, \quad (66) \end{aligned}$$

where (66) follows from (58) and (65).

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]. Re-writing (66) and (64) 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:*

$$D_2' = \alpha R^{-1}(\rho C_2) = \alpha\sigma^2 \left(\frac{N_2}{P+N_2}\right)^{\rho} \tag{67}$$

and

$$D_1' \geq \alpha\sigma^2 \left(\frac{N_2}{P+N_2}\right)^{\rho-1} \frac{N_1}{P+N_1} \left(1 + \frac{N_2}{N_1}\left(\alpha^{1/(\rho-1)} - 1\right)\right)^{-(\rho-1)}, \tag{68}$$

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

Figure 8 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.

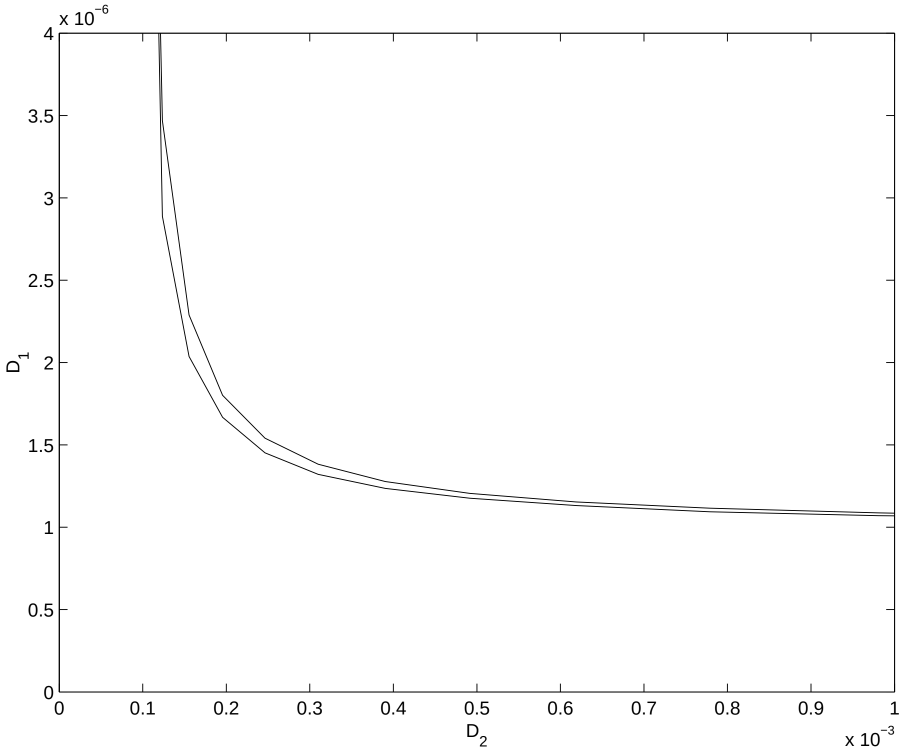

Figure 8: Numerical analysis of the inner and outer bounds.

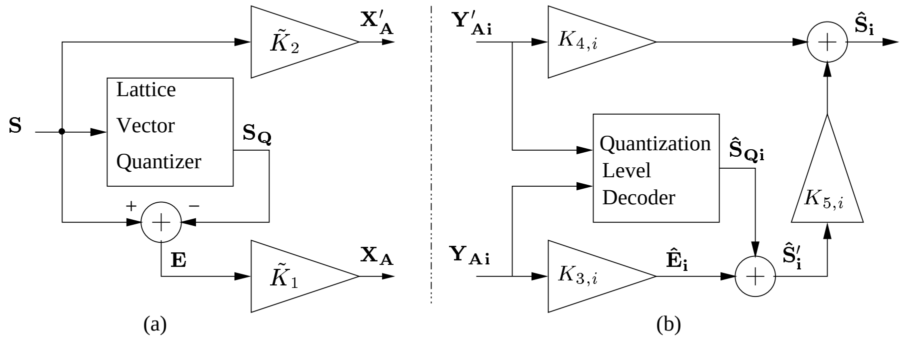

Figure 9: Modulo-Lattice modulation for lossy transmission with bandwidth expansion. (a) Transmitter. (b) Receivers ($i = 1, 2$)

6 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 Figure 7 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 Figure 9(a). It is similar to the one of Figure 7 (with $\rho = 2$ and $\alpha = 1$), except that the message W_2 is not transmitted at all. Instead we transmit the source $\mathbf{S}$ uncoded. (We denote $\mathbf{X'_A} = \tilde{K}_2\mathbf{S}$). In addition, the vector quantizer is a lattice vector quantizer. Therefore $\mathbf{E}$ can be expressed as $\mathbf{E} = \mathbf{S} \bmod \Lambda$, where Λ is the lattice. For this reason, we call this scheme the *modulo-lattice modulation* scheme. $\mathbf{E}$ is sent uncoded and we denote $\mathbf{X_A} = \tilde{K}_1\mathbf{E}$.

Receiver 1 and receiver 2, depicted in Figure 9(b) are identical, except for different gain factors. The quantization-level-decoder employs a modified-nearest-neighbor algorithm which losslessly decodes $\hat{\mathbf{S}}_{\mathbf{Qi}}$. Hence, with high probability, $\hat{\mathbf{S}}_{\mathbf{Qi}} = \mathbf{S_Q}$ for $i = 1, 2$. We then add a scaled version of $\mathbf{Y_{Ai}}$ (a noisy version of $\mathbf{X_A}$) to $\hat{\mathbf{S}}_{\mathbf{Qi}}$ and generate an estimate $\hat{\mathbf{S}}'_{\mathbf{i}}$ of $\mathbf{S}$. The final estimate $\hat{\mathbf{S}}_{\mathbf{i}}$ is then generated by weighted averaging of $\mathbf{Y'_{Ai}}$ (a noisy version of $\mathbf{X'_A}$) and $\hat{\mathbf{S}}'_{\mathbf{i}}$.

In [13] we show that the modulo-lattice has the same performance as the hybrid digital-analog scheme described in section 5. 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". (A similar concept appears in [3].) "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 7, 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 a future work.

7 Improved Performance Using Partial Feedback

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 6, we shall only consider the case where the bad receiver is kept optimal and there are two channel uses per source sample ($\rho = 2$). 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 $\rho = 2$. Note however, that this does not imply optimality since the scheme assumes the existence of a feedback, whereas the lower bound did not assume any feedback.

The encoder, the channel, the feedback and the decoders are illustrated in Figure 10. We concentrate on physically degraded channels since in all

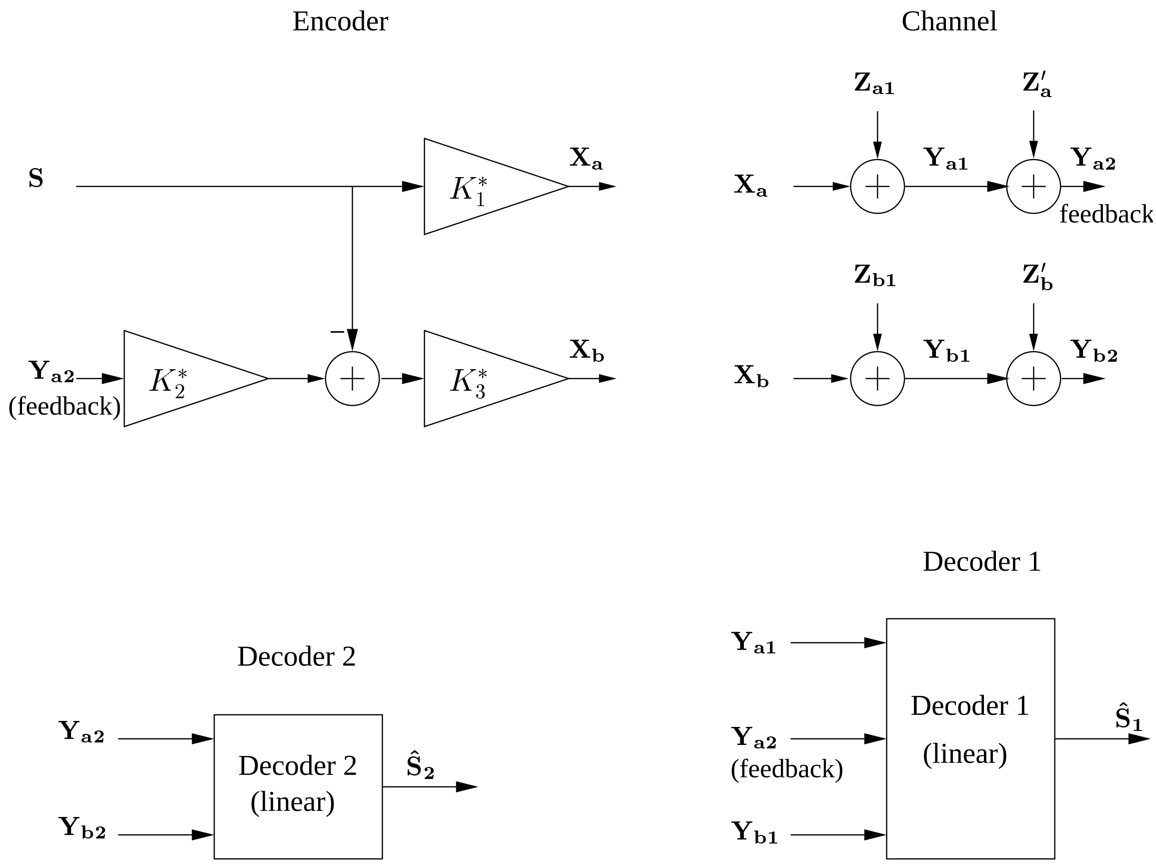

Figure 10: Broadcasting with feedback

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

$$X_{a,t} = K_1^* S_t, \qquad t = 1, 2, \cdots, m. \tag{69}$$

The channel is a physically degraded channel and therefore [2]:

$$\begin{aligned} Y_{a1,t} &= X_{a,t} + Z_{a1,t} && (70) \\ Y_{a2,t} &= X_{a,t} + Z_{a1,t} + Z'_{a,t} \quad t = 1, 2, \cdots, m, && (71) \end{aligned}$$

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

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

$$X_{b,t} = K_3^*(S - K_2^* Y_{a2,t}), \quad t = 1, 2, \cdots, m, \tag{72}$$

where $K_2^* = \sqrt{\frac{P\sigma^2}{P+N_2}}$ is the Wienner 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:

$$\begin{aligned} Y_{b1,t} &= X_{b,t} + Z_{b1,t} && (73) \\ Y_{b2,t} &= X_{b,t} + Z_{b1,t} + Z'_{b,t} \quad t = 1, 2, \cdots, m, && (74) \end{aligned}$$

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

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

$$\mathbf{Y_{2,t}} \triangleq \begin{bmatrix} Y_{a2,t} \\ Y_{b2,t} \end{bmatrix} \quad \text{and} \quad \mathbf{Y_{1,t}} \triangleq \begin{bmatrix} Y_{a1,t} \\ Y_{a2,t} \\ Y_{b1,t} \end{bmatrix}, \quad t = 1, \cdots, m. \tag{75}$$

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

$$\mathbf{R_{y,i}} = E\left(\mathbf{Y}_{\mathbf{i,t}}^t \cdot \mathbf{Y_{i,t}}\right) \quad \text{and} \quad \mathbf{r_{sy,i}} = E\left(S_t \mathbf{Y_{i,t}}\right). \tag{76}$$

Combining (69)-(76) yields:

$$\mathbf{R_{y,2}} = \begin{bmatrix} P+N_2 & 0 \\ 0 & P+N_2 \end{bmatrix}, \qquad \mathbf{r_{sy,2}} = \begin{bmatrix} \sqrt{P\sigma^2} \\ \sqrt{\frac{N_2 P\sigma^2}{P+N_2}} \end{bmatrix}, \tag{77}$$

$$\mathbf{R_{y,1}} = \begin{bmatrix} P+N_1 & P+N_1 & \frac{P(N_2-N_1)}{\sqrt{N_2(P+N_2)}} \\ P+N_1 & P+N_2 & 0 \\ \frac{P(N_2-N_1)}{\sqrt{N_2(P+N_2)}} & 0 & P+N_1 \end{bmatrix} \tag{78}$$

$$\text{and } \mathbf{r_{sy,1}} = \begin{bmatrix} \sqrt{P\sigma^2} \\ \sqrt{P\sigma^2} \\ \sqrt{\frac{N_2 P\sigma^2}{P+N_2}} \end{bmatrix}. \tag{79}$$

The linear estimation is given by

$$\hat{S}_{i,t} = \mathbf{a_i}^t \cdot \mathbf{Y_{i,t}}, \tag{80}$$

where

$$\mathbf{a_i} = \mathbf{R}_{\mathbf{y,i}}^{-1} \mathbf{r_{sy,i}}. \tag{81}$$

The resulting distortion is then given by

$$D_i = \sigma^2 - \mathbf{a_i}^t \cdot \mathbf{r_{sy,i}}. \tag{82}$$

Combining (77) - (82) yields:

$$D_1 = \frac{\sigma^2 N_1 N_2}{P^2 + 2PN_2 + N_1 N_2} \quad \text{and} \quad D_2 = \frac{\sigma^2 N_2^2}{(P+N_2)^2}. \tag{83}$$

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 (83) meets the lower bound of Corollary 1 for $\rho = 2$. We emphasize again that this does not imply optimality since the scheme assumed the existence of a feedback, whereas the lower bound did not assume any feedback.

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. We also conjecture that the distortion pair described in Corollary 1 is achievable (and therefore optimal) for any $\rho > 1$.

8 Conclusions

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 $D = \sigma^2/(1 + SNR)^\rho$, we showed that if a system must be optimal at a certain SNR_{min}, then asymptotically the distortion cannot decay faster than $1/SNR$.

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 [2, page 422]. We shall now show 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 $f_{y_1,y_2|x}(\mathbf{y_1},\mathbf{y_2}|\mathbf{x})$, where $\mathbf{x}$, $\mathbf{y_1}$ and $\mathbf{y_2}$ are of length n.*

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

$$f_{y_1|x}(\mathbf{y_1}|\mathbf{x}) = f^*_{y_1|x}(\mathbf{y_1}|\mathbf{x}) \quad \text{for all} \quad \mathbf{y_1} \in {\mathcal{Y}_1}^n \quad \text{and} \quad \mathbf{x} \in \mathcal{X}^n \tag{84}$$

$$f_{y_2|x}(\mathbf{y_2}|\mathbf{x}) = f^*_{y_2|x}(\mathbf{y_2}|\mathbf{x}) \quad \text{for all} \quad \mathbf{y_2} \in {\mathcal{Y}_2}^n \quad \text{and} \quad \mathbf{x} \in \mathcal{X}^n \tag{85}$$

but

$$f_{y_1,y_2|x}(\mathbf{y_1},\mathbf{y_2}|\mathbf{x}) \neq f^*_{y_1,y_2|x}(\mathbf{y_1},\mathbf{y_2}|\mathbf{x}) \quad \text{for some} \quad (\mathbf{x},\mathbf{y_1},\mathbf{y_2}). \tag{86}$$

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

$$D_i^f = \int_{\mathbf{S}}\int_{\mathbf{Y_i}} f(\mathbf{S}) \cdot f_{y_i|x}(\mathbf{Y_i}|i_m(\mathbf{S})) \cdot d\left(\mathbf{S}, g_{im}(\mathbf{Y_i})\right) \; d\mathbf{Y_i}\, d\mathbf{S} \tag{87}$$

and

$$D_i^{f^*} = \int_{\mathbf{S}}\int_{\mathbf{Y_i}} f(\mathbf{S}) \cdot f^*_{y_i|x}(\mathbf{Y_i}|i_m(\mathbf{S})) \cdot d\left(\mathbf{S}, g_{im}(\mathbf{Y_i})\right) \; d\mathbf{Y_i}\, d\mathbf{S} \tag{88}$$

Combining (84), (85), (87) and (88) yields:

$$\left(D_1^{f^*}, D_2^{f^*}\right) = \left(D_1^f, D_2^f\right). \tag{89}$$

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

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

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 Proof on Equation (43)

We shall now prove equation (43). Let

$$\mathbf{Y_2'} \triangleq \mathbf{Y_1} + \mathbf{Z}', \tag{90}$$

where $\mathbf{Z}' = Z_1', \ldots, Z_n'$ is memoryless with $Z_t' \sim \mathcal{N}(0, N_2 - N_1)$, and $\mathbf{Z}'$ is independent of $\mathbf{U}$, $\mathbf{X}$ and $\mathbf{Z_1}$. Define $\mathbf{Z}'' = \mathbf{Z_1} + \mathbf{Z}'$. Hence, $\mathbf{Y_2'} = \mathbf{X} + \mathbf{Z}''$ where $\mathbf{Z}''$ is memoryless, zero mean, Gaussian, with variance N_2, and independent of $\mathbf{X}$. Additionally we have that

$$\mathbf{Y_2} = \mathbf{X} + \mathbf{Z_2}, \tag{91}$$

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

$$h(\mathbf{Y_2}|\mathbf{U}) = h(\mathbf{Y_2'}|\mathbf{U}). \tag{92}$$

Now, by the conditional entropy power inequality [9], and since $\mathbf{Y_2'}$ is an independent sum of $\mathbf{Y_1}$ and $\mathbf{Z}'$, and $\mathbf{Z}'$ is Gaussian with variance $N_2 - N_1$, we have:

$$2^{\frac{2}{n}h(\mathbf{Y_2'}|\mathbf{U})} \geq 2^{\frac{2}{n}h(\mathbf{Y_1}|\mathbf{U})} + 2^{\log(2\pi e(N_2 - N_1))}. \tag{93}$$

Combining (92) and (93) leads to (43).

References

[1] M. D. Trott, "Unequal error protection codes: Theory and practice," in *Proc. of IEEE IT-Workshop*, Haifa, Israel, June 1996.

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

[3] B. Chen and G. Wornell, "Analog error-correcting codes based on chaotic dynamical systems," *IEEE Transactions on Communication*, vol. 46, pp. 881–890, July 1998.

[4] U. Mittal and N. Phamdo, "Joint source-channel codes for broadcasting and robust communication," *IEEE Transactions on Information Theory*, vol. IT-48, pp. 1082–1103, May 2002.

[5] S. Shamai, S. Verdú, and R. Zamir, "Systematic lossy source-channel coding," *IEEE Transactions on Information Theory*, vol. IT-44, 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 International Symposium on Information Theory*, June 2002.

[7] L. Ozarow, "On a source-coding problem with two channels and three receivers," *The Bell System Technical Journal*, pp. 1909–1921, Dec. 1980.

[8] J. Ziv, "The behavior of analog communication systems," *IEEE Transactions on Information Theory*, vol. IT-16, pp. 587–594, Sept. 1970.

[9] N. Blachman, "The convolution inequality for entropy powers," *IEEE Transactions on Information Theory*, vol. 11, 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 Transactions on Information Theory*, vol. IT-48, 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 Transactions on Information Theory*, vol. IT-22, 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," *Information and Control*, vol. 38, pp. 60–80, 1978.

[13] Z. Reznic, *Broadcasting analog sources over Gaussian channels*, Ph.D. thesis, Tel-Aviv University, August 2003.

[14] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattce codes for structural multiterminal binning," *IEEE Transactions on Information Theory*, vol. IT-48, June 2002.