Joint Source-Channel Coding of a Gaussian Mixture Source over the Gaussian Broadcast Channel*

Zvi Reznic Ram Zamir Meir Feder

Technical Report, Dept. of EE-Systems, Tel Aviv University, April 24, 2001

Abstract

Suppose that we want to send a description of a source to two listeners through a Gaussian broadcast channel, where the channel is used once per source sample. The problem of joint source-channel coding is to design a communication system to minimize the distortion D_1 at receiver 1 and at the same time minimize the distortion D_2 at receiver 2. If the source is Gaussian, the optimal solution is well known, and it is achieved by uncoded "analog" scheme. In this paper we consider a Gaussian mixture source. We derive inner and outer bounds for the distortion region of all (D_1, D_2) pairs that are simultaneously achievable. The outer bound is based on the entropy power inequality. The inner bound is attained by a digital-over-analog encoding scheme, which splits the source into a "discrete component" and a "continuous component", encodes the discrete component, and superimposes it over the continuous component which is sent uncoded. We also show that if the modes of the Gaussian mixture are highly separated, our bounds are tight, and hence, our scheme attains the entire distortion region. This optimal region exceeds the region attained by separating source and channel coding.

Key Words: Separation principle, joint source-channel coding, digital-over-analog, robust communication, distortion region.

*Zvi Reznic is with the Dept. of EE-Systems, Tel Aviv University, Israel, and with Texas Instruments - Cable Broadband Communications (Libit) *email: zvi@ti.com.* Ram Zamir and Meir Feder are with the Dept. of EE-Systems, Tel Aviv University, Israel. *email: zamir,meir@eng.tau.ac.il.* This paper was presented in part at the thirty-sixth Allerton conference on communication, control, and computing, Oct. 1998.

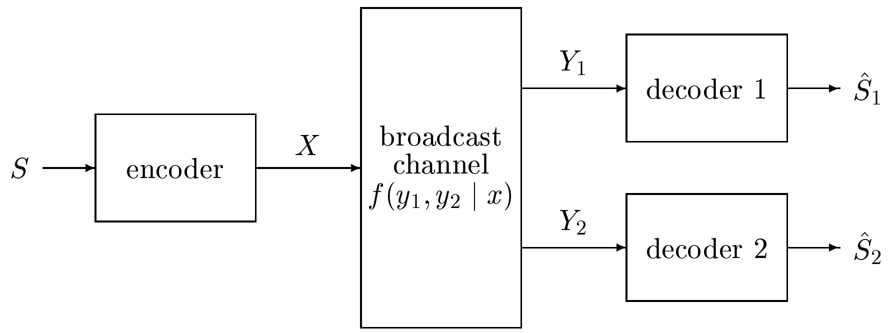


Figure 1: Lossy transmission of a source through a broadcast channel

1 Introduction

The *broadcast channel*, illustrated in Figure 1, is a communication channel in which one sender transmits to two or more receivers. In the usual formulation of the problem, the sender wishes to send two private messages, one to each receiver, and possibly a common massage, to both receivers. These messages need to be transmitted losslessly [1].

Suppose, however, that we are given a *single* source and a fidelity criterion, and we want to convey the source to both receivers simultaneously. Suppose further that the source entropy is large to the extent that it cannot be sent losslessly through the channel; there must be some distortion in the receivers' output. The problem of *joint source-channel coding* for the broadcast channel is to find the set of all achievable distortion pairs (D_1, D_2) at the two recievers. For a general source, broadcast channel and distortion measure, this problem is yet open [2]. We investigate below one example, and derive inner and outer bounds for the distortion region. These bounds become tight for a limiting case.

In our example the channel is a *degraded broadcast channel* in which receiver 1 (the "better" one) can decode all the information that receiver 2 can, and some additional information. Hence, information decoded at receiver 2 is actually common to both receivers. To minimize the distortion at both receivers, one might apply a two steps encoding. Step number one is *source coding.* Here, we create two messages: one that contains a coarse description of the source and another one which is a *refinement* message [3, 4]. The refinement message is an addendum to the first one, such that both messages together form a fine description of the source. We shall denote by D_c the distortion obtained with the coarse description, and by D_f the distortion obtained with the fine description, that is, $D_f \leq D_c$. Step number two is *channel coding.* We use a broadcast channel code (see [5], Theorem 14.6.2) to send the coarse description message to both receivers, and the refinement message to receiver 1 only. Hence, receiver 1 and receiver 2 obtain distortions D_f and D_c respectively.

The two steps encoding is based on *separation*; the source coding and channel coding are done separately [5]. Unfortunately, unlike for the case of point-to-point communication, the distortion pair obtained by the two steps approach is usually suboptimal. One simple example where separation is strictly suboptimal is the case of a Gaussian source sent over a Gaussian broadcast channel, with a squared error distortion measure. In fact, if the channel

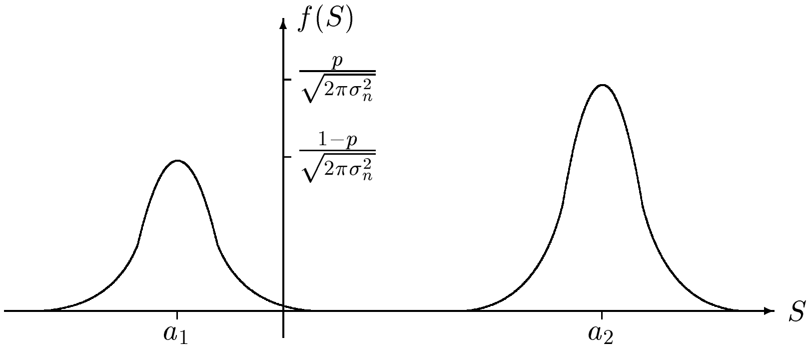


Figure 2: Example of a Gaussian mixture source with two modes

is used once per each source sample, then optimality is achieved in this case by analog transmission, i.e., by sending the source *uncoded* [8, 9].

In this paper, the source S, which needs to be transmitted over the broadcast channel, is the *Gaussian mixture* source. That is,

$$S = B + N, \tag{1}$$

where N is a zero mean Gaussian with variance σ_n^2 and $B \in \{a_1, \cdots, a_m\}$ is a discrete random variable which is statistically independent of N. One example of such a source, with $m = 2$, is illustrated in Figure 2. Our goal is to transmit this source over a Gaussian broadcast channel, with one channel use per source sample, and to minimize the mean squared error obtained by the receivers.

We derive inner and outer bounds for set of the achievable distortion pairs (D_1, D_2). The proof for the outer bound resembles Bergmans' proof of the converse theorem for the Gaussian broadcast channel (for lossless transmission) [10]. It is based on the entropy power inequality.

The inner bound is achieved by the *source-decomposition and digital-over-analog* scheme illustrated in Figure 3. With this scheme we split the source into a discrete part and a Gaussian part, and transmit the Gaussian part analogly, without any coding. At the same time we use source coding and channel coding to digitally transmit the discrete part of the source, considering the uncoded transmission as part of the channel noise. This scheme is simpler than separation based encoding, because part of the source is not encoded. Moreover it is asymptotically optimal; the source-decomposition digital-over-analog scheme attains the *outer* bound when the Gaussian "modes" of S are highly separated. Thus, our characterization of the distortion region is asymptotically tight.

Hybrid analog-digital schemes were suggested for various joint source-channel coding settings, mainly for the case where the channel bandwidth is larger than the source bandwidth, e.g., [6, 9, 11, 13]. However, no proof of optimality was given, nor useful outer bounds were derived in order to evaluate these schemes for transmission over a broadcast channel.

In Section 2 we analyze the distortion attainable with the proposed scheme in a *point-to-point* communication. In Section 3, we return to the broadcast channel and derive the inner and outer bounds for the distortion region. Section 4 discusses shortly the case where the source components, B and N, are dependent, and Section 5 concludes the paper.

2 Source Decomposition and Digital-over-Analog Coding

In this section we present an encoding scheme to transmit a Gaussian mixture source over a Gaussian point-to-point channel. We analyze the performance of this scheme with respect to the squared error distortion measure

$$d(\mathbf{S}, \hat{\mathbf{S}}) = \frac{1}{n} \sum_{t=1}^{n} (S_t - \hat{S}_t)^2 \tag{2}$$

where $\mathbf{S} = (S_1, \cdots, S_n)$ is the source block and $\hat{\mathbf{S}} = (\hat{S}_1, \cdots, \hat{S}_n)$ is the reconstruction block.

Define

$$a \triangleq \frac{1}{2} \min_{i \neq j} (|a_i - a_j|)$$

and $l \triangleq a/\sigma_n$. That is, l denotes the separation level of the two closest modes of the Gaussian mixture. In Appendix I we show that the rate-distortion function of the memoryless Gaussian mixture source S defined in (1), for distortions $D \leq \sigma_n^2$, satisfies

$$R_S(D) = H(B) + \frac{1}{2} \log \left(\frac{\sigma_n^2}{D} \right) - \epsilon(l) \tag{3}$$

where $\epsilon(l) \geq 0$ and $\lim_{l \to \infty} \epsilon(l) = 0$.

Consider a point-to-point channel and denote its input and output by X and Y respectively. Suppose it is a memoryless Gaussian channel with average power constraint P, that is

$$\frac{1}{n} \sum_{t=1}^{n} E(X_t^2) \leq P$$

where n is the coding block length,

$$Y = X + Z,$$

and $Z \sim \mathcal{N}(0, \sigma_z^2)$ is statistically independent of X. The capacity of the memoryless Gaussian channel [5] is given by:

$$C = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_z^2} \right). \tag{4}$$

Assume $C \geq H(B)$. By the joint source-channel coding theorem, the minimum achievable distortion D_{min} (the "OPTA") is the solution of $R_S(D_{min}) = C$, implying by (3) that

$$D_{min} = \frac{\sigma_n^2}{1 + P/\sigma_z^2} 2^{2(H(B) - \epsilon(l))}, \tag{5}$$

where the assumption $C \geq H(B)$ guarantees that $D_{min} \leq \sigma_n^2$ so (3) holds.

Note that for a fixed l, D_{min} depends linearly on σ_n^2. This property appears throughout the paper in all the distortion expressions. Hence, the distortion can be normalized by σ_n^2.

The *source-decomposition and digital-over-analog* (or simply the digital-over-analog) encoder and decoder are illustrated in Figure 3. The source splitter splits each sample S of the source into a discrete variable $B' \in (a_1 \cdots a_m)$ and a continues variable N', such that $B'+N'=S$. Note that our goal is to transmit S so as long as $B'+N'=S$ it is not important whether B' equals B.

One possible implementation of the source splitter is by setting $B'=q(S)$, and $N'=S-B'$, where the quantization function $q(\cdot)$ is defined as

$$q(S) = \arg\left(\min_{a_i}(|S-a_i|)\right). \tag{6}$$

However, with this splitting method the p.d.f. of N' given B' is limited at least on one side, and therefore N' is not Gaussian. Since a Gaussian N' will simplify our derivation, we use another implementation as follows. For each input $S=s$ the source splitter randomly assigns a value to B' according to $\Pr(B'=a_i|S=s)=\Pr(B=a_i|S=s)$, for every i. It then sets $N'=S-B'$. By taking the expectation over S we observe that for every i, $\Pr(B'=a_i)=\Pr(B=a_i)$. In fact, since the conditional joint distribution of B' and N' given S is the same as that of B and N given S, we have $N' \sim \mathcal{N}(0,\sigma_n^2)$, and B' and N' are statistically independent.

Note that this implementation is suboptimal for small values of l, and is especially wasteful for $l=0$. Nevertheless, we use it because we will be focusing on large values of l, for which, as we show later, the scheme is optimal. Note that for both implementations above $B' \to B$ as $l \to \infty$ in probability.

Each sample N' is scaled by a scalar K_1 to produce X_N. A source-channel code i_n encodes the n-block $\mathbf{B}'=(B'_1,\ldots,B'_n)$ to produce the vector $\mathbf{X}_B=i_n(\mathbf{B}')$, which is then added to the vector $\mathbf{X}_N$ to produce the encoder output $\mathbf{X}$. We have

$$Y_t = X_{B,t} + (X_{N,t}+Z_t) \qquad t=1\ldots n \tag{7}$$

where for each time instant t, $(X_{N,t}+Z_t) \sim \mathcal{N}(0, P_G+\sigma_z^2)$ and $P_G \triangleq E\left(X_N^2\right)$. Define

$$P_B \triangleq \frac{1}{n}\sum_{t=1}^{n} EX_{B,t}^2. \tag{8}$$

Equations (7)-(8) specify an equivalent Gaussian channel over which we send B', and whose capacity is given by:

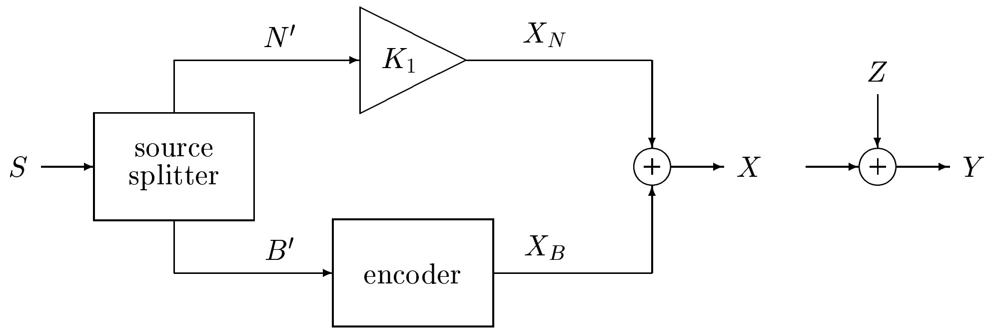
$$C_B = \frac{1}{2}\log\left(1+\frac{P_B}{\sigma_z^2+P_G}\right). \tag{9}$$

To transmit B' reliably we choose P_G and P_B such that for some $\Delta>0$,

$$C_B - \Delta = H(B') = H(B). \tag{10}$$

Combining (9),(10) and the power constraint $P_G+P_B=P$ and solving for P_G yields for $\Delta \to 0$:

$$P_G = P_G^* \triangleq \frac{\sigma_z^2+P}{2^{2H(B)}} - \sigma_z^2 \quad \text{and} \quad K_1 = \sqrt{\frac{P_G^*}{\sigma_n^2}}\,, \tag{11}$$



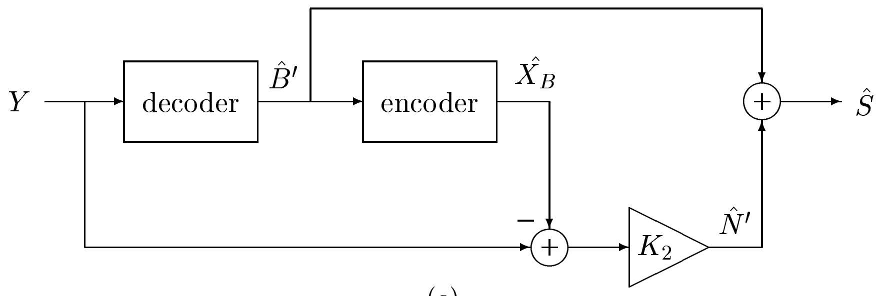


Figure 3: Digital over analog encoding scheme. (a) Encoder. (b) Channel. (c) Decoder.

where $P_G^* \geq 0$ since $C \geq H(B)$. The decoder first estimates $\hat{\mathbf{B}}' = g_n(\mathbf{Y})$, where $\mathbf{Y} = (Y_1, \ldots, Y_n)$. Then it generates $\hat{\mathbf{X}}_B$ from $\hat{\mathbf{B}}'$, in the same manner as the encoder generates $\mathbf{X_B}$ from $\mathbf{B}'$, i.e., $\hat{\mathbf{X}}_\mathbf{B} = i_n(\hat{\mathbf{B}}')$. The decoder subtracts $\hat{\mathbf{X}}_B$ from the channel output $\mathbf{Y}$ and estimates

$$\hat{\mathbf{N}}' = K_2(\mathbf{Y} - \hat{\mathbf{X}}_B) = K_2\left(K_1\mathbf{N}' + i_n(\mathbf{B}') + \mathbf{Z} - i_n(\hat{\mathbf{B}}')\right) \tag{12}$$

and

$$\hat{\mathbf{S}} = \hat{\mathbf{B}}' + \hat{\mathbf{N}}'. \tag{13}$$

The overall mean squared error of the coding scheme is

$$\begin{aligned} D_{split} &\triangleq E\left(\frac{1}{n}\sum_{t=1}^{n}(\hat{S}_t - S_t)^2\right) \qquad (14)\\ &= E\left(\frac{1}{n}\sum_{t=1}^{n}(\hat{B}'_t + \hat{N}'_t - B'_t - N'_t)^2\right)\\ &= E\left(\frac{1}{n}\sum_{t=1}^{n}\left(\hat{B}'_t + K_2(K_1N'_t + X_{B,t} + Z_t - \hat{X}_{B,t}) - B'_t - N'_t\right)^2\right)\\ &= E\left(\frac{1}{n}\sum_{t=1}^{n}\left((\hat{B}'_t - B'_t) + K_2(K_1N'_t + Z_t - \frac{1}{K_2}N'_t) + K_2(X_{B,t} - \hat{X}_{B,t})\right)^2\right). \end{aligned}$$

The capacity of a power constrained AWGN channel can be arbitrarily approached by a pick limited, power limited input, provided that the pick constraint is large enough. It then follows that for every $\Delta > 0$, if we allocate the power such that $C_B - \Delta = H(B')$, then there exists a bound $J < \infty$ and there exists a series of encoding functions i_n and reconstruction functions g_n with codebooks $\mathcal{C}(n)$ of rate $C_B - \Delta$, such that all the samples in $\mathcal{C}(n)$ are smaller than J, and

$$\lim_{n\to\infty} \Pr(\hat{\mathbf{B}}' \neq \mathbf{B}') = 0. \tag{15}$$

This implies also

$$\lim_{n\to\infty} \Pr(\hat{\mathbf{X}}_{\mathbf{B}} \neq \mathbf{X}_{\mathbf{B}}) = 0. \tag{16}$$

Since $\mathbf{X}_{\mathbf{B}}$ and $\mathbf{B}'$ are bounded, convergence in probability implies convergence in quadratic mean. Hence,

$$\lim_{n\to\infty} E\left((X_{B,t} - \hat{X}_{B,t})^2\right) = \lim_{n\to\infty} E\left((B'_t - \hat{B}'_t)^2\right) = 0. \tag{17}$$

Substituting (17) into (14), and using the Cauchy-Schwarz inequality, the first and third terms vanish, and we have for any fixed, finite a

$$\lim_{n\to\infty} D_{split} = E((K_2(K_1 N' + Z) - N')^2) \,. \tag{18}$$

Thus, as $\Delta \to 0$, the expected distortion is asymptotically the same as the average squared error for a Gaussian source with variance σ_n^2 transmitted uncoded with power P_G^* (as given in (11)) over a channel with additive white Gaussian noise with power σ_z^2, leading to (with $K_2 = P_G^*/(P_G^* + \sigma_z^2)$)

$$D^*_{split} \triangleq \lim_{\Delta\to 0} \lim_{n\to\infty} D_{split} = \frac{\sigma_n^2}{1 + P_G^*/\sigma_z^2}, \tag{19}$$

where P_G^* is defined in (11). Note that, unlike the minimum theoretic distortion (5), D^*_{split} is independent of the separation level l. However, with $l \to \infty$, the minimum theoretic distortion coincides with the RHS of (19). Therefore, we have proved the following theorem:

Theorem 1 *If $C \geq H(B)$ then the digital-over-analog coding scheme of Figure 3 with K_1, K_2, i_n and g_n as described above, is optimal in the sense that*

$$D^*_{split} = \lim_{l\to\infty} D_{min}. \tag{20}$$

Note that $l = a/\sigma_n$ can approach infinity either by fixing a and letting $\sigma_n \to 0$, or by fixing σ_n and letting $a \to \infty$. The former is less interesting, since the distortion approaches zero. We are interested in the latter since it represents the case where the distortion stays in the same range, as $l \to \infty$.

The result of Theorem 1 can be explained as follows: For a fixed σ_n and as $l \to \infty$, in order to achieve a finite distortion the digital information B must be transmitted without loss. The problem then becomes that of transmitting two sources, B and N, where B needs to be transmitted losslessly. This is done by the digital-over-analog scheme by means of overlaid communication and successive cancellation decoding [6].

For finite values of a and blocklength n, some error in decoding B' is unavoidable, and leads to some (finite) cost in distortion. In this case some excess rate $\Delta = C_B - H(B')$ must

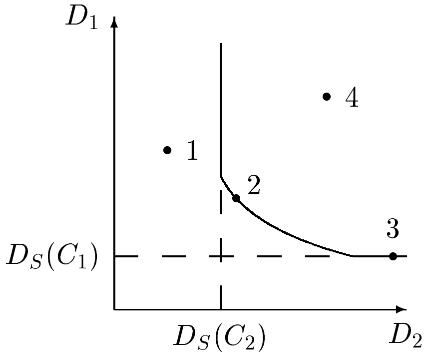


Figure 4: Achievable distortion region for the broadcast channel. Point 1 is not achievable. Points 2-4 are achievable.

be retained (by appropriate power allocation) to control the decoding error, and thus keep a good balance between the "digital distortion" and the "analog distortion".

The efficiency of the digital-over-analog scheme can be viewed in terms of information rate. When $a_i = a_j$ for every i and j, (which implies $l = 0$), there is no need to send the discrete part at all. Hence, the digital-over-analog scheme is wasting $H(B)$ bits per channel use. As l increases, the importance of the discrete part increases, and hence the waste becomes smaller, until it vanishes (exponentially) for $l \to \infty$.

An extension of the results, to the case where B and N are not independent, is described in Section 4.

In practice, the digital-over-analog scheme is simpler than the classical separation scheme (source encoder followed by channel encoder) because only a part of the source needs to be encoded and hence it requires less coding effort. This advantage becomes significant when $C \gg H(B)$. Theorem 1 shows that the digital-over-analog scheme is not only simple, it is also optimal for point to point communication and highly separated modes.

3 Distortion Region for the Gaussian Broadcast Channel

We turn to consider the broadcast channel. The goal of the encoder and decoders in Figure 1 is to simultaneously minimize the distortion D_1 at decoder 1 and D_2 at decoder 2. The combination of a certain source, broadcast channel and distortion measure implies a set of distortion pairs that are achievable. We would call this set the *achievable distortion region.* As mentioned in the Introduction this set is in general unknown. Figure 4 illustrates a typical shape of the distortion region, where $D_S(\cdot)$ denotes the distortion-rate function of the source, and C_1 and C_2 are the capacities of the good and bad channels, respectively.

In Section 1 we mentioned that analog transmission is optimal for sending a Gaussian source over a Gaussian broadcast channel of the same bandwidth; it achieves the "ideal" point $(D_1, D_2) = (D_S(C_1), D_S(C_2))$. Hence, the digital-over-analog encoding scheme, where the "Gaussian part" of the information is sent in an analog form (uncoded), seems appealing for sending the Gaussian mixture source. To apply the digital-over-analog encoder to the broadcast channel, we should choose the powers P_G and P_B so that receiver 2 will be able to decode B' losslessly. Since receiver 1 is better, it will also decode B' losslessly. As for the Gaussian part, each receiver will obtain a distortion in accordance with the noise level in its channel.

The distortion region achieved by the digital-over-analog scheme is plotted in Figure 5. For comparison, we show the boundary of the region achievable by separation of source coding and channel coding, as explained in the Introduction. Figure 5 also shows some outer bounds on the achievable distortion region, that will be derived here.

Theorem 2 below characterizes the distortion region which is achievable by the digital-over-analog scheme. This region does not depend on the value of l, and provides an inner bound for the distortion region. Theorem 3 sets an outer bound on the achievable distortion region, that depends on the value of l. Finally, Theorem 4 shows that as $l \to \infty$, the two bounds become tight, and hence, the entire distortion region is completely characterized, and is achieved by the digital-over-analog scheme.

Before we prove the Theorems, we introduce some formal definitions.

Definition 1 *(D_1, D_2) is an achievable distortion pair for the source S, the distortion measure $d(s, \hat{s})$ and the memoryless broadcast channel $f(y_1, y_2 \mid x)$ if for some n there exist an encoding function $\mathbf{X} = i_n(\mathbf{S})$ and two reconstruction functions $\hat{\mathbf{S}}_\mathbf{1} = g_{1n}(\mathbf{Y_1})$ and $\hat{\mathbf{S}}_\mathbf{2} = g_{2n}(\mathbf{Y_2})$ such that for $i = 1, 2$*

$$D_i = E\left(d(\mathbf{S}, \hat{\mathbf{S}}_\mathbf{i})\right),$$

where, as above, bold-face letters denote blocks of size n, i.e., $\mathbf{S} = (S_1, \cdots, S_n)$ is the source block, $\mathbf{X} = (X_1, \ldots, X_n)$ is the channel input block, $\mathbf{Y_i} = (Y_{i,1}, \ldots, Y_{i,n})$ are the channel output blocks, and $\hat{\mathbf{S}}_\mathbf{i} = (\hat{S}_{i,1}, \cdots, \hat{S}_{i,n})$ are the reconstruction blocks.

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

The achievable distortion region must be convex by a time sharing argument.

Definition 2 *A* Gaussian *broadcast channel $f(y_1, y_2 \mid x)$ satisfies for $i = 1, 2$:*

$$\frac{1}{n}\sum_{t=1}^{n} E(X_t^2) \le P,$$

$$Y_i = X + Z_i,$$

$$Z_i \sim \mathcal{N}(0, \sigma_i^2),$$

where Z_1, Z_2 and X are mutually statistically independent, and $\sigma_2^2 > \sigma_1^2$.

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

$$C_i = \frac{1}{2}\log(1 + P/\sigma_i^2) \quad i = 1, 2.$$

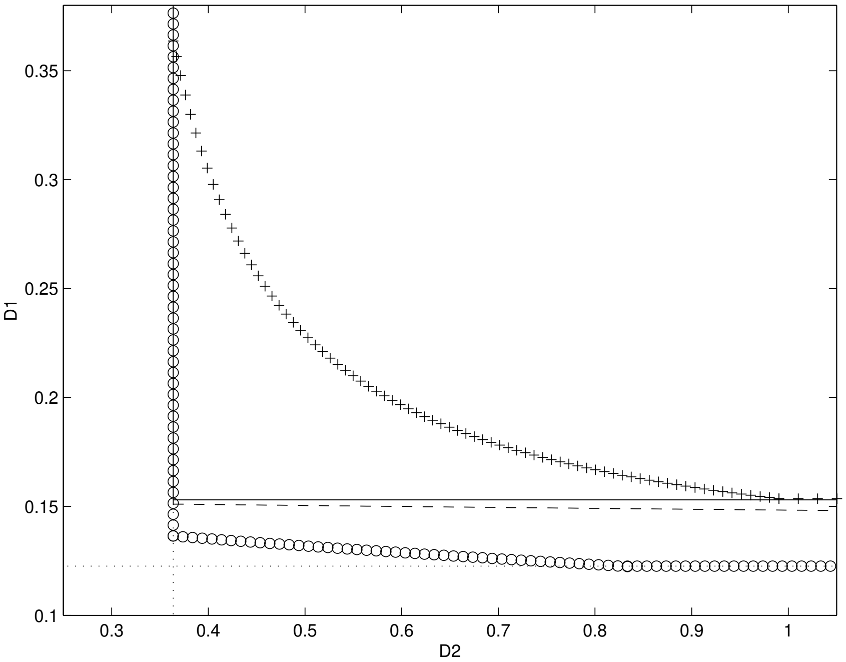


Figure 5: The distortion region of a Gaussian mixture source with two modes, transmitted over a Gaussian broadcast channel. Here, we show a channel with $P = 5$, $\sigma_1^2 = 0.158$ and $\sigma_2^2 = 0.5$, and two sources: one with $\sigma_n^2 = 1$ and $l = 14$, and another one with $\sigma_n^2 = 1$ and $l = 50$. The '+' line is the distortion region achieved with separation. It is plotted for $l \to \infty$, although it hardly depends on l, for $l > 3$. The solid line is the distortion region achieved with the digital-over-analog scheme (independent of l). The 'o' line and the dashed line are the outer bounds of Theorem 3 for $l = 14$, and $l = 50$, respectively. The dotted lines are the trivial outer bounds at $R_S(D_1) = C_1$ and $R_S(D_2) = C_2$. (They also hardly depend on l). Note that for fixed σ_n and $l \to \infty$, $D_1 = D(C_1)$ cannot be obtained for any finite D_2.

Theorem 2 (Inner bound): *The distortion pair* (D_1, D_2)*, for sending the Gaussian mixture source* S *of (1) over a Gaussian broadcast channel with* $C_2 \geq H(B)$*, is achievable if* $D_1 \geq D_1^*$ *and* $D_2 \geq D_2^*$*, where*

$$(D_1^*, D_2^*) = \left(\frac{\sigma_n^2}{1 + P_G/\sigma_1^2}, \frac{\sigma_n^2}{1 + P_G/\sigma_2^2} \right) \tag{21}$$

and

$$P_G = \frac{\sigma_2^2 + P}{2^{2H(B)}} - \sigma_2^2. \tag{22}$$

Furthermore, the pair (D_1^*, D_2^*) *is achievable by the digital-over-analog scheme of Figure 3.*

Remarks: Note that (D_1^*, D_2^*) are independent of the separation level l between the source modes. Note also that D_2^* in (21) is asymptotically optimal in the sense that it satisfies (see (3))

$$R_S(D_2^*) = C_2 - \epsilon(l). \tag{23}$$

However, since P_G above is not optimal for σ_1^2, we see from (3) that D_1^* is strictly worse than the point-to-point distortion for any l:

$$R_S(D_1^*) = \frac{1}{2} \log \left(2^{2C_1} - SNR_{LOSS} \right) - \epsilon(l) \tag{24}$$

where $SNR_{LOSS} = (\sigma_2^2/\sigma_1^2 - 1)(2^{2H(B)} - 1)$.

Proof: Suppose we use the digital-over-analog scheme, where we choose the power P_G of the Gaussian part such that the bad receiver (and obviously the good receiver) can decode B' losslessly, and we choose the gain K_2 of each receiver according to its own noise level. Then from (11) P_G is as given in (22), and from (19), as $n \to \infty$ and $\Delta \to 0$, the resulting distortions are as given in (21). $\square$

Theorem 3 (Outer bound): *The distortion pair* (D_1, D_2) *for sending the Gaussian mixture source* S *of (1) over a Gaussian broadcast channel with* $C_2 > H(B)$*, is achievable only if*

1.

$$R_S(D_1) \leq C_1 \tag{25}$$

2.

$$R_S(D_2) \leq C_2 \tag{26}$$

3.

$$D_1 \geq \frac{\sigma_n^2}{1 + P'_G/\sigma_1^2} \cdot \beta \tag{27}$$

where

$$P'_G \triangleq \frac{\sigma_2^2 + P}{2^{2H(B)}} \cdot \alpha - \sigma_2^2, \qquad \beta \triangleq 2^{-2\bar{\epsilon}(l)}, \qquad \alpha \triangleq 2^{2\bar{\delta}},$$

and where $\bar{\epsilon}(l) = H_b(2Q(l)) + 2Q(l)\log(m-1)$, $\quad \bar{\delta} = H_b(p') + p'\log(m-1)$, $\quad p'$ *is the minimum between 1/2 and* $\frac{4}{l^2}\frac{D_2}{\sigma_n^2} + 2Q(l/2)$*, and* $H_b(\cdot)$ *is the binary entropy function.*

Note that the tradeoff between D_1 and D_2 is via α, and that $\alpha, \beta \to 1$ as $l \to \infty$ for any $D_2 < \infty$. To prove Theorem 3 we present two lemmas. The first lemma provides a lower bound on D_1.

Lemma 1 *Suppose we are given some joint source-channel coding scheme, to transmit the source n-block* $\mathbf{S}$*. Let* $\hat{\mathbf{S}}_1$ *and* $\hat{\mathbf{S}}_2$ *be the reconstructions produced by that scheme, achieving distortions* D_1 *and* D_2 *respectively. Let* P_{e2} *be the minimum average probability of error in estimating the vector* $\mathbf{B}$ *from the reconstruction vector* $\hat{\mathbf{S}}_\mathbf{2}$ *of the second (bad) receiver,i.e.*

$$P_{e2} = \frac{1}{n}\sum_{t=1}^{n} E\left(1 - \max_{i=1,\ldots,m} \Pr\left(B_t = a_i | \, \hat{\mathbf{S}}_2\right)\right)$$

where the expectation is over $\hat{\mathbf{S}}_2$*, and let*

$$\delta = H_b(P_{e2}) + P_{e2}\log(m-1). \tag{28}$$

Then, D_1 *satisfies:*

$$R_S(D_1) \le H(B) + \frac{1}{2}\log\left(1 + \frac{P_\delta}{\sigma_1^2}\right) \tag{29}$$

where

$$P_\delta \triangleq \frac{\sigma_2^2 + P}{2^{2(H(B)-\delta)}} - \sigma_2^2.$$

Note that since $R_S(D)$ is monotonically non-increasing, (29) provides a lower bound for the distortion D_1 of the good receiver, in terms of the probability of error P_{e2} in detecting the source mode B by the bad receiver. The strongest bound is achieved for $P_{e2} \to 0$.

Proof: Let $\mathbf{B}, \mathbf{N}, \mathbf{S}, \mathbf{X}, \mathbf{Y_1}, \mathbf{Y_2}, \hat{\mathbf{S}}_\mathbf{1}, \hat{\mathbf{S}}_\mathbf{2}$ be as above, and let $\mathbf{Z_1}$, $\mathbf{Z_2}$ be n-blocks of channel noise as described in Definition 2. We therefore have a Markov chain $\mathbf{B} \leftrightarrow \mathbf{S} \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y}_2 \leftrightarrow \hat{\mathbf{S}}_2$ and by Fano's inequality we have

$$H(\mathbf{B}|\mathbf{S}) \le H(\mathbf{B}|\mathbf{X}) \le H(\mathbf{B}|\mathbf{Y_2}) \le H(\mathbf{B}|\hat{\mathbf{S}}_\mathbf{2}) \le n\delta. \tag{30}$$

By the chain rule for mutual information we have for $i = 1, 2$:

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y_i}) &= I(\mathbf{X}; \mathbf{B}) + I(\mathbf{X}; \mathbf{Y_i}|\mathbf{B}) - I(\mathbf{X}; \mathbf{B}|\mathbf{Y_i}) \\ &= I(\mathbf{X}; \mathbf{B}) + h(\mathbf{Y_i}|\mathbf{B}) - h(\mathbf{Z_i}) - I(\mathbf{X}; \mathbf{B}|\mathbf{Y_i}). \end{aligned} \tag{31}$$

By the definition of capacity, and using (31) with $i = 2$, we upper bound $h(\mathbf{Y_2}|\mathbf{B})$ in terms of C_2:

$$\begin{aligned} nC_2 &\geq I(\mathbf{X};\mathbf{Y_2}) = I(\mathbf{X};\mathbf{B}) + h(\mathbf{Y_2}|\mathbf{B}) - h(\mathbf{Z_2}) - I(\mathbf{X};\mathbf{B}|\mathbf{Y_2}) \\ &= H(\mathbf{B}) - H(\mathbf{B}|\mathbf{X}) + h(\mathbf{Y_2}|\mathbf{B}) - h(\mathbf{Z_2}) - H(\mathbf{B}|\mathbf{Y_2}) + H(\mathbf{B}|\mathbf{X},\mathbf{Y_2}) \\ &\geq n\left(H(B) + \frac{1}{n}h(\mathbf{Y_2}|\mathbf{B}) - h(Z_2) - \delta\right) \end{aligned} \tag{32}$$

where we used (30) to upper bound $\frac{1}{n}H(\mathbf{B}|\mathbf{Y_2})$ by δ, and we used the Markov chain relation to cancel $H(\mathbf{B}|\mathbf{X})$ with $H(\mathbf{B}|\mathbf{X},\mathbf{Y_2})$. By the definition of rate-distortion function, and using (31) with $i = 1$, we upper bound $R_S(D_1)$ in terms of $h(\mathbf{Y_1}|\mathbf{B})$:

$$\begin{aligned} nR_S(D_1) &\leq I(\mathbf{S};\hat{\mathbf{S}}_\mathbf{1}) \leq I(\mathbf{X};\mathbf{Y_1}) \qquad (33) \\ &= I(\mathbf{X};\mathbf{B}) + h(\mathbf{Y_1}|\mathbf{B}) - h(\mathbf{Z_1}) - I(\mathbf{X};\mathbf{B}|\mathbf{Y_1}) \\ &\leq I(\mathbf{X};\mathbf{B}) + h(\mathbf{Y_1}|\mathbf{B}) - h(\mathbf{Z_1}) \\ &\leq n\left(H(B) + \frac{1}{n}h(\mathbf{Y_1}|\mathbf{B}) - h(Z_1)\right). \end{aligned}$$

where we used $I(X;B) \leq H(B)$ and the nonnegativity of the mutual information. Finally, by the entropy power inequality [5, 10], and since $\mathbf{Y_2}$ is the independent sum of $\mathbf{Y_1}$ and white Gaussian vector with power $\sigma_2^2 - \sigma_1^2$, we upper bound $h(\mathbf{Y_1}|\mathbf{B})$ in terms of $h(\mathbf{Y_2}|\mathbf{B})$:

$$2^{\frac{2}{n}h(\mathbf{Y_2}|\mathbf{B})} \geq 2^{\frac{2}{n}h(\mathbf{Y_1}|\mathbf{B})} + 2^{\log(2\pi e(\sigma_2^2 - \sigma_1^2))}. \tag{34}$$

Combining (32), (33) and (34) and substituting $C_2 = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_2^2}\right)$ yields the desired result. □

Note that the proof of Lemma 1 does not rely on a specific source distribution or a specific distortion measure. The second lemma provides an upper bound on P_{e2}.

Lemma 2 *Let P_{e2} and D_2 be defined as in Lemma 1 above. For any coding scheme,*

$$P_{e2} \leq \frac{4}{l^2} \cdot \frac{D_2}{\sigma_n^2} + 2Q\left(\frac{l}{2}\right). \tag{35}$$

Remark: Note that this Lemma implies that for a fixed σ_n, $\lim_{l\to\infty} P_{e2} = 0$ for any $D_2 < \infty$.

Proof: The nearest neighbor estimate, that is, estimating each sample B_t by $q(\hat{S}_{2,t})$ (where $q(\cdot)$ was defined in (6)) cannot be better than the optimal estimate. Therefore:

$$\begin{aligned} P_{e2} &\leq \frac{1}{n}\sum_{t=1}^{n} \Pr(B_t \neq q(\hat{S}_{2,t})) \\ &\leq \frac{1}{n}\sum_{t=1}^{n} \left\{ \Pr\left(|N_t| < \frac{a}{2}\right) \Pr\left(B_t \neq q(\hat{S}_{2,t}) \,\middle|\, |N_t| < \frac{a}{2}\right) + \Pr\left(|N_t| > \frac{a}{2}\right) \right\} \end{aligned}$$

$$\leq \left\{\frac{1}{n}\sum_{t=1}^{n}\Pr\left(|N_t|<\frac{a}{2}\right)\Pr\left((S_t-\hat{S}_{2,t})^2>\left(\frac{a}{2}\right)^2\Big||N_t|<\frac{a}{2}\right)\right\}+\Pr\left(|N_1|>\frac{a}{2}\right) \quad (36)$$

$$\leq \left\{\frac{1}{n}\sum_{t=1}^{n}\Pr\left(|N_t|<\frac{a}{2}\right)\frac{E\left((S_t-\hat{S}_{2,t})^2\Big||N_t|<\frac{a}{2}\right)}{(a/2)^2}\right\}+\Pr\left(|N_1|>\frac{a}{2}\right) \quad (37)$$

$$\leq 4/a^2\cdot D_2+\Pr\left(|N_1|>\frac{a}{2}\right) \quad (38)$$

$$= \frac{4}{l^2}\cdot\frac{D_2}{\sigma_n^2}+2Q\left(\frac{l}{2}\right) \quad (39)$$

where in (36) we concluded that if $|N_t|<\frac{a}{2}$, then $|S_t-\hat{S}_{2,t}|$ must be greater than $\frac{a}{2}$ to cause an error (see the definitions of a and $q(\cdot)$). Also in (36) we used the fact that $\Pr\left(|N_t|>\frac{a}{2}\right)$ does not depend on t. Finally, (37) follows by Chebyshev's inequality, and (38) since $D_2=\frac{1}{n}\sum_t E\left((S_t-\hat{S}_{2,t})^2\right)$. □

Proof of Theorem 3

Condition 3 in the theorem is proved by combining (3), (28), (29) and (35). Conditions 1 and 2 follow from Shannon's joint source-channel coding theorem. This proves the Theorem. □

Note that the digital-over-analog scheme achieves the point (D_1^*,D_2^*) in the distortion region. This implies that there is a corner-shaped distortion region that is achievable. As mentioned before, this point does not depend on the value of l. On the other hand, Theorem 3 establishes a lower bound which depends on l. For any l for which $\epsilon(l)$ is negligible (typically $l>3$), the lower bound in the D_2 axis coincides with D_2^*, while the gap in the D_1 axis is at most the loss reflected by the term SNR_{LOSS} in (24). As $l\to\infty$, the lower bound of Theorem 3 becomes tight, and meets the corner-shaped distortion region which is achievable by the digital-over-analog scheme. Hence we have characterized the entire distortion region for this case. This is formally stated in the following Theorem.

Theorem 4 (Tightness of bounds for highly separated modes): *The distortion region for sending the Gaussian mixture source S of (1) in the limit as $l\to\infty$, over a Gaussian broadcast channel with $C_2>H(B)$, is composed of all pairs (D_1,D_2) satisfying $D_1\geq D_1^*$ and $D_2\geq D_2^*$, where D_1^* and D_2^* were defined in Theorem 2. Furthermore, the distortion region is achievable by the digital-over-analog scheme of Figure 3.*

Proof: The achievability of (D_1^*,D_2^*) follows from Theorem 2. As for the converse part, note first that following the definitions of α, β and P_G' in (27), $\lim_{l\to\infty}\alpha=\lim_{l\to\infty}\beta=1$, and $\lim_{l\to\infty}P_G'=P_G$. Hence, by (27) and the definition of D_1^*, we have $\lim_{l\to\infty}D_1\geq D_1^*$. On the other hand, combining (23) with (26) yields $\lim_{l\to\infty}D_2\geq D_2^*$. Thus, asymptotically the distortions cannot be better than (D_1^*,D_2^*). □

Theorem 4 can be explained as follows: for a fixed σ_n and as $l\to\infty$, in order to achieve a finite distortion the digital information B must be transmitted without loss to the worse receiver. This digital code can also be decoded by the better receiver. Once both receivers have removed the digital component from the received signal, the problem becomes that of transmitting a Gaussian source over a Gaussian broadcast channel. For this problem, analog transmission is optimal.

4 Dependent B and N

Suppose that the source has only two modes ($m = 2$) and the variances of the two Gaussians are not the same. This implies the following dependency between B and N:

$$\mathrm{Var}(N) = \begin{cases} \sigma_{n0}^2 & \text{if } B = -1 \\ \sigma_{n1}^2 & \text{if } B = 1 \end{cases}$$

Without loss of generality we can assume that $\sigma_{n0}^2 < \sigma_{n1}^2$. Throughout this section we will assume that $l \to \infty$. To calculate the rate distortion function of the source, when B and N are dependent, we can use the same arguments that led to (3). This implies that:

$$R_S(D) = \begin{cases} H(B) + \frac{p_0}{2}\log\left(\frac{\sigma_{n0}^2}{D}\right) + \frac{p_1}{2}\log\left(\frac{\sigma_{n1}^2}{D}\right) & \text{if } D < \sigma_{n0}^2 \\ H(B) + \frac{p_1}{2}\log\left(\frac{p_1\sigma_{n1}^2}{D - p_0\sigma_{n0}^2}\right) & \text{if } \sigma_{n0}^2 < D < p_0\sigma_{n0}^2 + p_1\sigma_{n1}^2 \end{cases} \tag{40}$$

where $p_1 = \Pr(B = 1)$, and $p_0 = 1 - p_1$.

The digital-over-analog scheme can be modified so that it can still achieve $R_S(D) = C$ for point-to-point communication. The main modification is that the powers P_G and P_B allocated for transmitting N' and B' should not be constant for all symbols. Instead they should depend on the values of B'. This sets a decoding problem: To correctly decode B' from the channels output Y, the decoder needs to know the power allocation, but in order to know the power allocation it needs to know B', resulting in a frustrating endless loop. To overcome this, we suggest to use the *lookahead* method described in [12]: The encoder partitions X_N and X_B in blocks of length M. It then sends the first block of X_B without X_N. Hence, the first block of X_B can be properly decoded as $M \to \infty$. The encoder then sends the rest of the blocks overlaying block number j of X_N with block number $j+1$ of X_B. The decoder decodes the first block of X_B and by that learns the power allocation of the second block. Knowing the power allocation, the decoder can properly decode the second block and move on to the third block and so forth. For large number of blocks, the overhead of the first block becomes negligible.

We allocate the powers as follow: Let $P_{G,i}$ be the power for transmitting the Gaussian part (a sample from the j-th block of X_N) given that its variance is σ_{ni}^2 for $i = 0, 1$. Let $P_{B,i}$ be the power of the simultaneously transmitted sample from the $(j+1)$-th block of X_B. As before, the binary part B' should be sent losslessly, which implies:

$$\frac{p_0}{2}\log\left(1 + \frac{P_{B,0}}{P_{G,0} + \sigma_z^2}\right) + \frac{p_1}{2}\log\left(1 + \frac{P_{B,1}}{P_{G,1} + \sigma_z^2}\right) = H(B). \tag{41}$$

To maximize the mutual information between the channel input and output we require that the input $X = X_N + X_B$ will be identically distributed. This implies a constant power of X

regardless of the value of B, that is,

$$P_{B,j} + P_{G,j} = P \quad \text{for } j = 0, 1. \tag{42}$$

To minimize the average distortion we require a "source water pouring" condition:

$$D_{V,0} = \min(D_{V,1}, \sigma_{n0}^2), \tag{43}$$

where $D_{V,j}$ is the average distortion when $\mathrm{Var}(N) = \sigma_{nj}^2$. And finally, we require a maximal utilization of $P_{G,j}$:

$$\tfrac{1}{2}\log\left(\frac{\sigma_{nj}^2}{D_{V,j}}\right) = \tfrac{1}{2}\log\left(1 + \frac{P_{G,j}}{\sigma_z^2}\right) \quad \text{for } j = 0, 1. \tag{44}$$

We therefore have 6 conditions and 6 unknowns $(P_{G,0}, P_{B,0}, P_{G,1}, P_{B,1}, D_{V,0}, D_{V,1})$ which determine the power allocation. Define D' as the resulting distortion from this scheme, that is,

$$D' \triangleq p_0 D_{V,0} + p_1 D_{V,1}. \tag{45}$$

Combining (40)-(45) and (4) yields for $D' < p_0\sigma_{n0}^2 + p_1\sigma_{n1}^2$:

$$R_S(D') = C,$$

which means that the analog-over-digital encoding scheme, with the power allocation described here, is optimal.

Note that in some cases the solution yields $P_{B,1} < 0$. In such cases the digital-over-analog scheme is suboptimal, and equations (42)-(43) are not necessarily valid.

As for the broadcast channel, consider the special case in which the SNR in both channels is high, that is

$$\frac{P}{\sigma_i^2} \to \infty \quad \text{for } i = 1, 2. \tag{46}$$

In this case, the power allocation can be calculated by replacing σ_z^2 in (41) and (44) by σ_2^2 (the noise level in the bad channel). We can then prove that the digital-over-analog scheme is still optimal for the broadcast channel, when B and N are not independent.

Unfortunately, in the general case, when (46) does not hold, it is not evident whether the digital-over-analog scheme is optimal for the broadcast channel, and what should be the power allocation. Clearly, the binary part should be transmitted without error to both receivers, so in (41) we should replace σ_z^2 with σ_2^2. It is less clear what to use for σ_z^2 in (44). We can replace σ_z^2 with σ_1^2, and thus minimize the the distortion at the *good* receiver. Alternatively, we can replace σ_z^2 with σ_2^2 and thus minimize the distortion at the *bad* receiver (and actually achieve $R_S(D_2) = C_2$). If we do so, then our scheme implies a certain distortion $\tilde{D}_1$ at the good receiver. Theorem 3, if applied here, gives a lower bound $D_{1,L}$ for the distortion at the good receiver, when the bad receiver is optimal. Unfortunately $D_{1,L} < \tilde{D}_1$, so we cannot conclude from our derivation that the digital-over-analog scheme is optimal in this case.

The problems of finding the optimal distortion region and determining the optimal scheme for the case where B and N are not independent, are left for further study.

5 Conclusions

The digital-over-analog method splits the source into two parts, one is coded and the other is transmitted uncoded. For transmission over a Gaussian broadcast channel, the digital-over-analog scheme is better than traditional methods which are based on separating the source coding from the channel coding. Moreover, for a highly separated Gaussian mixture source, this method meets the theoretical bound on the achievable distortion region stated in Theorem 3, and hence it is optimal.

In the broadcast channel case, the transmitter needs to know only the noise level at the *bad* receiver, in order to guarantee reliable transmission of the discrete part. This implies that the digital-over-analog scheme is in fact *robust* (or "universal") with respect to the channel SNR, provided that the minimum possible SNR is known at the transmitter. For example, this scheme can be used for source-channel coding over a slowly varying fading channel, where the fading level is known (or estimated) at the receiver, if the transmitter knows the strongest possible fading level [2]. Here, each fading level corresponds to a "receiver" in the broadcast channel model.

The distortion region for joint source-channel coding over the broadcast channel is yet unknown. Here we derived inner and outer bounds for this region in one special case. These bounds are asymptotically tight. We believe that some of the ideas we used, e.g., optimum source decomposition, can be developed further and used in more general cases.

Appendix I Rate distortion function for the Gaussian mixture source

the rate-distortion function of the memoryless Gaussian mixture source S defined in (1), for distortions $D \leq \sigma_n^2$, satisfies

$$R_S(D) = h(S) - \frac{1}{2}\log(2\pi e D) \tag{47}$$

where h denotes differential entropy. This follows since by the Shannon lower bound $R_S(D) \geq h(S) - \frac{1}{2}\log(2\pi e D)$ for all D, and the lower bound is tight for $D \leq \sigma_n^2$ since we can write S as the independent sum of a Gaussian with variance D and some random variable; see [7], Theorem 4.3.1. Continuing from (47) we have by the chain rule and the properties of the entropy function [5]:

$$\begin{aligned}
R_S(D) &= I(B;S) + h(S|B) - \frac{1}{2}\log(2\pi e D) &(48)\\
&= H(B) - H(B|S) + \frac{1}{2}\log(2\pi e \sigma_n^2) - \frac{1}{2}\log(2\pi e D) &(49)\\
&\geq H(B) - H_b\Big(\Pr\left(B \neq q(S)\right)\Big) - \Pr\Big(B \neq q(S)\Big)\log(m-1) + \frac{1}{2}\log\left(\frac{\sigma_n^2}{D}\right) &\\
&\geq H(B) - H_b\left(2Q\left(\frac{a}{\sigma_n}\right)\right) - 2Q\left(\frac{a}{\sigma_n}\right)\log(m-1) + \frac{1}{2}\log\left(\frac{\sigma_n^2}{D}\right) &(50)
\end{aligned}$$

where the first inequality follows from the Fano inequality, and the quantization function $q(\cdot)$ is defined in (6). Hence, we can write:

$$R_S(D) = H(B) + \frac{1}{2}\log\left(\frac{\sigma_n^2}{D}\right) - \epsilon(l),$$

where $l = a/\sigma_n$, and, according to (49), $\epsilon(l) \geq 0$, and according to (50) $\epsilon(l) \leq H_b\left(2Q(l)\right) + 2Q\left(l\right)\log(m-1)$. Clearly, $\lim_{l\to\infty}\epsilon(l) = 0$.

Acknowledgment

The authors wish to thank the anonymous reviewers for some useful comments and corrections. Ram Zamir wishes to acknowledge an early discussion with Eyal Shlomot which inspired some of the ideas presented.

References

[1] T. M. Cover, "Broadcast channels", *IEEE Trans. Inform. Theory,* vol. IT-18, no. 1, pp. 2-14, Jan. 1972.

[2] M. D. Trott, "Unequal Error Protection Codes: Theory and Practice", *Proc. of IEEE IT-Workshop,* pp. 11, Haifa, Israel, June 1996.

[3] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory,* vol. IT-37, no. 2, pp. 269-274, Mar. 1991.

[4] B. Rimoldi, "Successive refinement of information: characterization of the achievable rates," *IEEE Trans. Inform. Theory,* vol. IT-40, no. 1, pp. 253-259, Jan. 1994.

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

[6] S. Shamai, S. Verdú and R. Zamir, "Systematic lossy source-channel coding", *IEEE Trans. Inform. Theory,* vol. IT-44, no. 2, pp. 564-579, March 1998.

[7] T. Berger, "Rate-distortion theory: a mathematical basis for data-compression", *Prentice-Hall, Englewood Cliffs, NJ,* 1971.

[8] T.J. Goblick, "Theoretical limitations on the transmission of data from analog sources", *IEEE Trans. Inform. Theory,* vol. IT-11, pp. 558-567, Oct. 1965.

[9] B. Chen and G. Wornell, "Analog error-correcting codes based on chaotic dynamical systems", *IEEE Trans. Communications,* vol. IT-46, pp. 881-890, July 1998.

[10] P. P. Bergmans " A simple converse for broadcast channels with additive white Gaussian noise" *IEEE Trans. Inform. Theory,* pp. 279-280, March 1974.

[11] J. Ziv, "The behavior of analog communication systems", *IEEE Trans. Inform. Theory,* vol. IT-16, pp. 587-594, Sept. 1970.

[12] S. Shamai and S. Verdú, "Capacity of channels with uncoded side information," *Europ. Trans. Telecommun.,* vol. 6, no. 5, pp 587-600, Sept.-Oct. 1995.

[13] U. Mittal and N. Phamdo "Joint source-channel codes for broadcasting and robust communication," submitted to *IEEE Trans. Inform. Theory.*