Since the tables in [4, Ch. 14] deal only with the cases where $\mu \leq 4$, let us examine in the next example the first two values of $K(n, 1, 5)$.

Example 1: Clearly, $K(4, 1, 5) = 16$ and hence $M_1^{(\leq 3)}(4) = 16$. Let then $n = 5$. The code

$$C = F_2^n \setminus \{00000, 11100, 00111, 11011\}$$

gives $K(5, 1, 5) \leq 28$. In fact, there cannot be a fivefold 1-covering with only 27 codewords. Assume, to the contrary, the existence of such a code C. We may assume without loss of generality that 00000 belongs to $C' = F_2^n \setminus C$. Words of weight less than three cannot therefore exist in C'. Since there can only be one word of weight four or five in C', at least three words of weight three are in C'. However, evidently at most two words of weight three can be in C', a contradiction. Thus, $K(5, 1, 5) = 28 = M_1^{(\leq 3)}(5)$.

It follows easily from [13, Theorem 8] (and Theorem 1) that

$$\frac{K(n, 1, \mu)}{\mu 2^n/(n+1)} \to 1$$

as $n \to \infty$ when μ is fixed. Hence, also

$$\frac{M_1^{(\leq l)}(n)}{(2l-1)2^n/(n+1)} \to 1$$

for a given l $(n \to \infty)$.

ACKNOWLEDGMENT

The author wishes to thank I. Honkala and the anonymous referee for valuable comments on this correspondence.

REFERENCES

[1] U. Blass, I. Honkala, and S. Litsyn, “Bounds on identifying codes,” *Discr. Math.*, to be published.

[2] ——, “On binary codes for identification,” *J. Combin. Des.*, vol. 8, pp. 151–156, 2000.

[3] I. Charon, I. Honkala, O. Hudry, and A. Lobstein, “General bounds for identifying codes in some infinite regular graphs,” *Electron. J. Combin.*, submitted for publication.

[4] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering Codes*. Amsterdam, The Netherlands: Elsevier, 1997.

[5] G. Cohen, I. Honkala, A. Lobstein, and G. Zémor, “On identifying codes,” in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Proc. DIMACS Workshop Codes and Association Schemes*, Nov. 9–12, 1999, pp. 97–109.

[6] G. Exoo, “Computational results on identifying t-codes,” preprint.

[7] I. Honkala, “On the identifying radius of codes,” in *Proc. 7th Nordic Combinatorial Conf.*, Turku, Finland, 1999, pp. 39–43.

[8] I. Honkala, T. Laihonen, and S. Ranto, “On codes identifying sets of vertices in Hamming spaces,” *Des. Codes Cryptogr.*, to be published.

[9] ——, “On strongly identifying codes,” *Discr. Math.*, to be published.

[10] ——, “Codes for strong identification,” *Electron. Notes Discr. Math.*, to be published.

[11] M. G. Karpovsky, K. Chakrabarty, and L. B. Levitin, “On a new class of codes for identifying vertices in graphs,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 599–611, Mar. 1998.

[12] T. Laihonen and S. Ranto, “Families of optimal codes for strong identification,” *Discr. Appl. Math.*, to be published.

[13] P. R. J. Östergård, “A coloring problem in Hamming spaces,” *Europ. J. Comb.*, vol. 18, pp. 303–309, 1997.

[14] S. Ranto, I. Honkala, and T. Laihonen, “Two families of optimal identifying codes in binary Hamming spaces,” *IEEE Trans. Inform. Theory*, submitted for publication.

Joint Source–Channel Coding of a Gaussian Mixture Source Over the Gaussian Broadcast Channel

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

***Abstract*—Suppose that we want to send a description of a single 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 an uncoded “analog” scheme. In this correspondence, 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, while the inner bound is attained by a digital-over-analog encoding scheme, which we present here. 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, although it does not contain the “ideal” point $(D_1, D_2) = (R^{-1}(C_1), R^{-1}(C_2))$.**

***Index Terms*—Digital-over-analog scheme, distortion region, joint source–channel coding, separation principle.**

I. INTRODUCTION

The *broadcast channel*, illustrated in Fig. 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 message, 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 receivers. For a general source, broadcast channel, and distortion measure, this problem is still 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*, which means that if the transmitted information is *digital*, then 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-step encoding. Step one is *source coding.* Here, we create two messages: one that contains a coarse description of the source and another 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

Manuscript received May 2, 2001; revised October 14, 2001. The material in this correspondence was presented in part at the 36th Allerton Conference on Communication, Control, and Computing, September 1998.

Z. Reznic is with the Department of Electrical Engineering–Systems, Tel- Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel, and with Texas Instruments–Cable Broadband Communications (Libit), Herzlia 46766, Israel (e-mail: zvi@eng.tau.ac.il).

R. Zamir and M. Feder are with the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel (e-mail: zamir@eng.tau.ac.il; meir@eng.tau.ac.il).

Communicated by P. Narayan, Associate Editor for Shannon Theory.

Publisher Item Identifier S 0018-9448(02)00856-8.

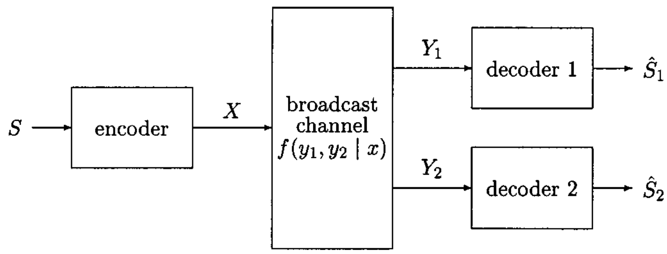

Fig. 1. Lossy transmission of a source through a broadcast channel.

denote by D_c and D_f the distortions obtained with the coarse and fine descriptions, respectively. Step 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-step encoding is based on *separation*; the source coding and channel coding are done separately. Unfortunately, unlike for the case of point-to-point communication, the distortion pair obtained by the two-step approach is usually suboptimal. This stems from the threshold effect of digital codes; although the worse receiver cannot decode the refinement message losslessly, it could still obtain some information about it. However, no technique exists to utilize this information in an optimal way [2], [7], [8]. Suboptimality of separation is known for other multiuser problems as well [5, p. 448]. 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 is used once per each source sample, then optimality is achieved in this case by analog transmission, i.e., by sending the source *uncoded* [6], [7]. Furthermore, the resulting distortion pair is "ideal" in the sense that there is no conflict between the two receivers. In the general case, the "ideal" point is not achievable.

In this correspondence, we consider a specific example of source and channel which enables us to combine the advantages of coded and uncoded transmission, and gain some mathematical insight. Our broadcast channel is Gaussian and the distortion measure is squared-error. 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 white Gaussian with variance σ_n^2 and $B \in \{a_1, \ldots, a_m\}$ is a discrete random variable which is statistically independent of N. One example of such a source, with $m = 2$, is illustrated in Fig. 2. We 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 Section II, we derive nontrivial inner and outer bounds for the set of the achievable distortion pairs (D_1, D_2). Specifically, we present the digital-over-analog scheme that is used to derive the inner bound in

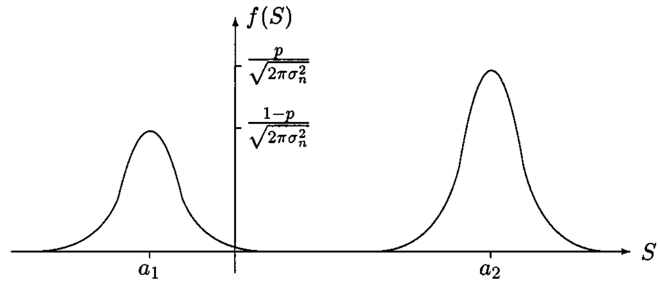

Fig. 2. Example of a Gaussian mixture source with two modes.

Theorem 2. 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., [9], [7], [8]. However, no proof of optimality was given for broadcasting.

The outer bound is given in Theorem 2, whose proof resembles Bergmans' proof of the converse theorem for the Gaussian broadcast channel (for lossless transmission) [10]. It is based on the entropy power inequality. In Theorem 3, we show that as the Gaussian mixture becomes highly separated ($l \to \infty$), the two bounds become tight, and hence, the entire distortion region is completely characterized. The asymptotic solution is nontrivial in the sense that, on one hand, it is strictly better than source–channel separation, and on the other hand, it is strictly worse than the ideal (nonconflict) distortion pair.

II. Inner and Outer Bounds on the Distortion Region

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 $f(y_1, y_2|x)$ if for some n there exist an encoding function $\boldsymbol{X} = i_n(\boldsymbol{S})$ and two reconstruction functions $\hat{\boldsymbol{S}}_{\boldsymbol{1}} = g_{1n}(\boldsymbol{Y}_{\boldsymbol{1}})$ and $\hat{\boldsymbol{S}}_{\boldsymbol{2}} = g_{2n}(\boldsymbol{Y}_{\boldsymbol{2}})$ such that for $i = 1, 2$

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

where bold-face letters denote blocks of size n, i.e., $\boldsymbol{S} = (S_1, \ldots, S_n)$ is the source block, $\boldsymbol{X} = (X_1, \ldots, X_n)$ is the channel input block, $\boldsymbol{Y}_{\boldsymbol{i}} = (Y_{i,1}, \ldots, Y_{i,n})$ are the channel output blocks, and $\hat{\boldsymbol{S}}_{\boldsymbol{i}} = (\hat{S}_{i,1}, \ldots, \hat{S}_{i,n})$ 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 $f(y_1, y_2|x)$ satisfies for $i = 1, 2$

$$\frac{1}{n}\sum_{t=1}^{n} E(X_t^2) \le P, \quad Y_i = X + Z_i, \quad Z_i \sim \mathcal{N}(0, \sigma_i^2)$$

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

The Gaussian broadcast channel is a degraded broadcast channel [5, p. 379]. The individual capacities C_1 and C_2 of the good and bad channels, respectively, are given by

$$C_i = \tfrac{1}{2}\log(1 + P/\sigma_i^2), \qquad i = 1, 2. \tag{2}$$

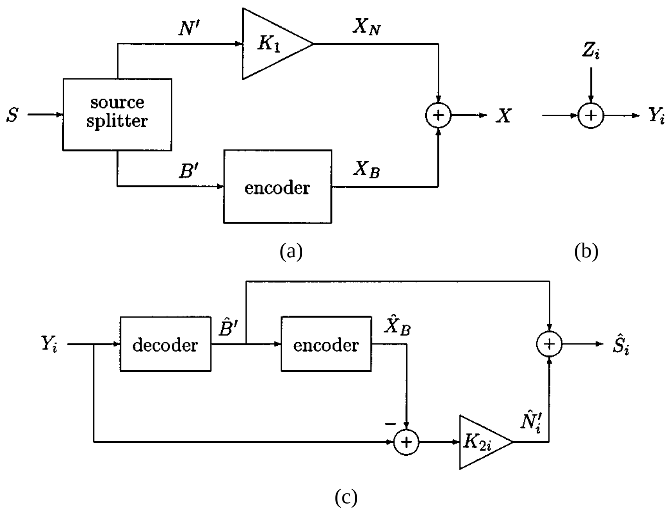

Fig. 3. Digital-over-analog encoding scheme. (a) Encoder. (b) Broadcast channel ($i = 1, 2$). (c) Decoder ($i = 1, 2$).

(Note that for *lossless* transmission, the rate pair (C_1, C_2) lies outside of the capacity region of the Gaussian broadcast channel [10].)

With the above definitions we state that our goal is to find the achievable distortion region for the Gaussian mixture source, the Gaussian-broadcast channel (with one channel use per source sample), and the squared error distortion measure

$$d(\boldsymbol{S}, \hat{\boldsymbol{S}}_{\boldsymbol{i}}) = \frac{1}{n}\sum_{t=1}^{n}(S_t - \hat{S}_{i,t})^2. \tag{3}$$

We start by calculating $R_S(D)$, the rate-distortion function of S. We note that for $D \le \sigma_n^2$ we can write S as the independent sum of a Gaussian with variance D and some random variable. Hence, by ([11, Theorem 4.3.1]), the Shannon lower bound is tight. That is,

$$\begin{aligned} R_S(D) &= h(S) - \frac{1}{2}\log(2\pi e D) && (4)\\ &= I(B; S) + h(S|B) - \frac{1}{2}\log(2\pi e D) && (5)\\ &= H(B) - \epsilon(l) + \frac{1}{2}\log\left(\frac{\sigma_n^2}{D}\right) && (6)\end{aligned}$$

where $\epsilon(l) \triangleq H(B|S)$, and we used

$$h(S|B) = h(N) = \frac{1}{2}\log(2\pi e\sigma_n^2).$$

Using Fano's inequality and the nonnegativity of $H(B|S)$, one can verify that

$$0 \le \epsilon(l) \le H_b(2Q(l)) + 2Q(l)\log(m-1) \tag{7}$$

where $H_b(\cdot)$ is the binary entropy function and Q is the error function. Clearly, $\epsilon(l) \to 0$ as $l \to \infty$.

The *source-decomposition and digital-over-analog* (or simply the digital-over-analog) encoder and decoders are illustrated in Fig. 3. We shall briefly analyze this scheme here. Similar analysis can be found in [12], while the full analysis is in [13]. 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.

Each sample N' is scaled by a scalar K_1 to produce X_N. A source–channel code i_n encodes the n-block $\boldsymbol{B}' = (B_1', \ldots, B_n')$ to produce the vector $\boldsymbol{X}_B = i_n(\boldsymbol{B}')$, which is then added to the vector $\boldsymbol{X}_N$ to produce the encoder output $\boldsymbol{X}$. Hence, B' is transmitted over a virtual channel whose noise is the sum of X_N and the channel noise. To transmit B' reliably to the worse receiver (and hence also to the better receiver) we must have for some $\Delta > 0$

$$H(B') = H(B) = C_{B2} - \Delta = \frac{1}{2}\log\left(1 + \frac{P_B}{\sigma_2^2 + P_G}\right) - \Delta \tag{8}$$

where C_{B2} is the capacity of the worse (point-to-point) virtual channel, and P_B and P_G are the powers of $\boldsymbol{X_B}$ and $\boldsymbol{X_N}$, respectively. Combining (8) and the power constraint $P_G + P_B = P$ and solving for P_G yields for $\Delta \to 0$

$$P_G = P_G^* \triangleq \frac{\sigma_2^2 + P}{2^{2H(B)}} - \sigma_2^2 \quad \text{and} \quad K_1 = \sqrt{\frac{P_G^*}{\sigma_n^2}}. \tag{9}$$

Since by (8) $H(B') < C_{B2}$, both receivers can decode $\boldsymbol{B}'$ losslessly as $n \to \infty$, and subtract the associated codeword $\boldsymbol{X_B}$. 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^* over channels with additive white Gaussian noise with powers σ_1^2 and σ_2^2. This leads (using the minimum-mean-squared-error gain $K_{2i} = P_G^*/(P_G^* + \sigma_i^2)$), to the following theorem.

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

$$(D_1^*, D_2^*) = \left(\frac{\sigma_n^2}{1 + P_G^*/\sigma_1^2}, \frac{\sigma_n^2}{1 + P_G^*/\sigma_2^2}\right). \tag{10}$$

Note that by (2), (6), (9), and (10) we have

$$C_2 = R_S(D_2^*) + \epsilon(l) \tag{11}$$

where $\epsilon(l)$ is defined in (6). Hence, $\epsilon(l)$ is the "wasted rate" with respect to receiver 2, which vanishes for $l \to \infty$. When l is small, we could reduce the waste by allowing some probability of misdetecting B by receiver 2. This would allow to shift some of the power from the digital part to the analog part, up to the point where D_2 is minimized. However, even with this minimization, we do not claim that our scheme is optimal for finite l. Hence, we shall skip this optimization process and simplify our analysis.

Also note that since P_G in (9) was optimized for receiver 2, D_1^* is strictly larger than $R_S^{-1}(C_1)$ for any l. Still, we will prove that the scheme is optimal over all joint source–channel codes for $l \to \infty$.

Fig. 4 depicts the distortion region of a Gaussian mixture source with two modes ($m = 2$), 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 a source with $\sigma_n^2 = 1$, $H(B) = 1$, and various separation levels l. The "+" line is the distortion region achieved with separation. It is plotted for $l \to \infty$, although it hardly depends on l for $l > 3$. Note that separation is far from optimal for $l \le 3$ as well. (For example, for $l = 0$ see [7].) The solid line is the distortion region achieved with the

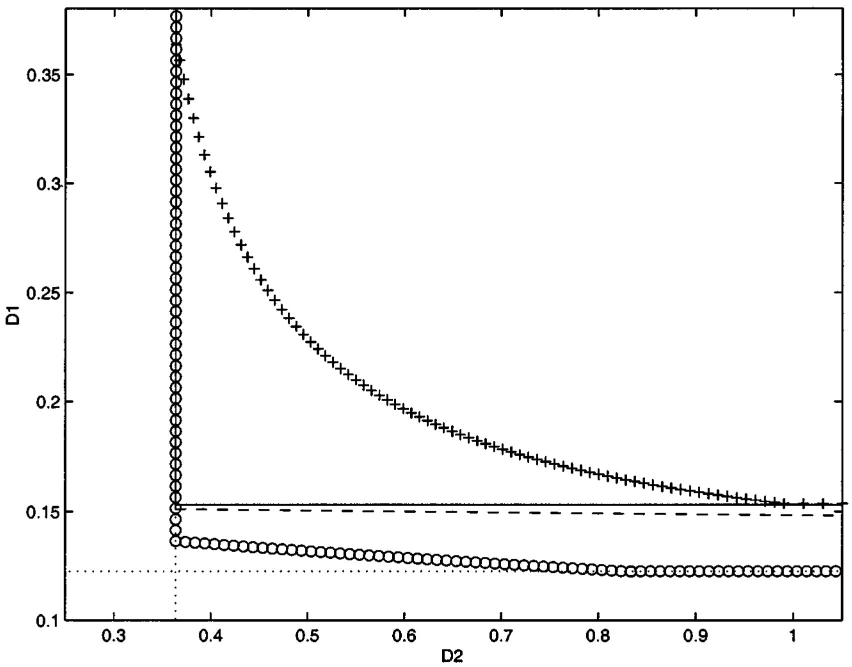

Fig. 4. Inner and outer bounds on the distortion region for a Gaussian mixture source with $l > 3$.

digital-over-analog scheme (independent of l). 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 for $l > 3$.) The "o" line and the dashed line are outer bounds for $l = 14$ and $l = 50$, which will be described later.

We shall now proceed to derive an outer bound. Suppose we are given some joint source–channel coding scheme, to transmit the source n-block $\boldsymbol{S}$. Let $\hat{\boldsymbol{S}}_1$ and $\hat{\boldsymbol{S}}_2$ 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 $\boldsymbol{B}$ from the reconstruction vector $\hat{\boldsymbol{S}}_2$ of the second (bad) receiver, i.e.,

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

where the expectation is over $\hat{\boldsymbol{S}}_2$. In Lemma 1 below, we shall derive an upper bound on P_{e2} in terms of D_2. In Lemma 2, we derive an upper bound on $I(\boldsymbol{S}; \hat{\boldsymbol{S}}_1)$ in terms of P_{e2}. In Theorem 2, we combine the two lemmas and the fact that $\frac{1}{n}I(\boldsymbol{S}; \hat{\boldsymbol{S}}_1)$ is lower-bounded by the rate distortion function at D_1, and derive a lower bound on D_1 in terms of D_2.

Lemma 1: Let P_{e2} and D_2 be defined as above. For any coding scheme

$$P_{e2} \le \frac{4}{l^2}\cdot\frac{D_2}{\sigma_n^2} + 2Q\left(\frac{l}{2}\right). \tag{12}$$

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

Proof: Define the nearest neighbor quantizer as

$$q(s) = \arg\min_{a_i} |s - a_i|. \tag{13}$$

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

$$\begin{aligned}
P_{e2} &\le \frac{1}{n}\sum_{t=1}^{n} \Pr\left(B_t \ne q\left(\hat{S}_{2,t}\right)\right)\\
&\le \frac{1}{n}\sum_{t=1}^{n}\left\{\Pr\left(|N_t| < \frac{a}{2}\right)\Pr\left(B_t \ne q\left(\hat{S}_{2,t}\right)\Big|\,|N_t| < \frac{a}{2}\right)\right.\\
&\qquad\left. + \Pr\left(|N_t| \ge \frac{a}{2}\right)\right\}\\
&\le \left\{\frac{1}{n}\sum_{t=1}^{n}\Pr\left(|N_t| < \frac{a}{2}\right)\right.\\
&\qquad\left.\cdot\Pr\left(\left(S_t - \hat{S}_{2,t}\right)^2 > \left(\frac{a}{2}\right)^2\Big|\,|N_t| < \frac{a}{2}\right)\right\}\\
&\quad + \Pr\left(|N_1| \ge \frac{a}{2}\right) &(14)\\
&\le \left\{\frac{1}{n}\sum_{t=1}^{n}\Pr\left(|N_t| < \frac{a}{2}\right)\frac{E\left(\left(S_t - \hat{S}_{2,t}\right)^2\Big||N_t| < \frac{a}{2}\right)}{(a/2)^2}\right\}\\
&\quad + \Pr\left(|N_1| \ge \frac{a}{2}\right) &(15)\\
&\le 4/a^2\cdot D_2 + \Pr\left(|N_1| \ge \frac{a}{2}\right) &(16)\\
&= \frac{4}{l^2}\cdot\frac{D_2}{\sigma_n^2} + 2Q\left(\frac{l}{2}\right) &(17)
\end{aligned}$$

where in (14) we concluded that if $|N_t| < \frac{a}{2}$, then an error implies that $|S_t - \hat{S}_{2,t}| \ge \frac{a}{2}$. Also, in (14) we used the fact that $\Pr(|N_t| > \frac{a}{2})$ does not depend on t. Finally, (15) follows by Chebyshev's inequality, and (16) since

$$D_2 = \frac{1}{n}\sum_t E((S_t - \hat{S}_{2,t})^2). \qquad \square$$

Note that P_{e2} can also be upper-bounded by $1 - \max_i(\Pr(B = a_i))$.

Lemma 2: Let P_{e2} be as defined above and let

$$\delta = H_b(P_{e2}) + P_{e2}\log(m-1) \quad \text{and} \quad P_\delta \triangleq \frac{\sigma_2^2 + P}{2^{2(H(B)-\delta)}} - \sigma_2^2. \tag{18}$$

Then, $I(\boldsymbol{S};\, \hat{\boldsymbol{S}}_1)$ satisfies

$$\frac{1}{n} I\left(\boldsymbol{S};\, \hat{\boldsymbol{S}}_1\right) \le H(B) + \frac{1}{2}\log\left(1 + \frac{P_\delta}{\sigma_1^2}\right). \tag{19}$$

Note that the tightest bound on $I(\boldsymbol{S};\, \hat{\boldsymbol{S}}_1)$ is achieved for $P_{e2} \to 0$.

Proof: Let $\boldsymbol{B}, \boldsymbol{N}, \boldsymbol{S}, \boldsymbol{X}, \boldsymbol{Y_1}, \boldsymbol{Y_2}, \hat{\boldsymbol{S}}_1, \hat{\boldsymbol{S}}_2$ be as above, and let $\boldsymbol{Z_1}, \boldsymbol{Z_2}$ be n-blocks of channel noise as described in Definition 2. We, therefore, have a Markov chain $\boldsymbol{B} \leftrightarrow \boldsymbol{S} \leftrightarrow \boldsymbol{X} \leftrightarrow \boldsymbol{Y}_2 \leftrightarrow \hat{\boldsymbol{S}}_2$ and by Fano's inequality we have

$$H(\boldsymbol{B}|\boldsymbol{S}) \le H(\boldsymbol{B}|\boldsymbol{X}) \le H(\boldsymbol{B}|\boldsymbol{Y_2}) \le H\left(\boldsymbol{B}|\hat{\boldsymbol{S}}_2\right) \le n\delta. \tag{20}$$

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

$$\begin{aligned} I(\boldsymbol{X};\, \boldsymbol{Y_i}) &= I(\boldsymbol{X};\, \boldsymbol{B}) + I(\boldsymbol{X};\, \boldsymbol{Y_i}|\boldsymbol{B}) - I(\boldsymbol{X};\, \boldsymbol{B}|\boldsymbol{Y_i}) \\ &= I(\boldsymbol{X};\, \boldsymbol{B}) + h(\boldsymbol{Y_i}|\boldsymbol{B}) - h(\boldsymbol{Z_i}) - I(\boldsymbol{X};\, \boldsymbol{B}|\boldsymbol{Y_i}). \end{aligned} \tag{21}$$

By the definition of capacity, and using (21) with $i = 2$, we upper-bound $h(\boldsymbol{Y_2}|\boldsymbol{B})$ in terms of C_2

$$\begin{aligned} nC_2 &\ge I(\boldsymbol{X};\, \boldsymbol{Y_2}) = I(\boldsymbol{X};\, \boldsymbol{B}) + h(\boldsymbol{Y_2}|\boldsymbol{B}) - h(\boldsymbol{Z_2}) - I(\boldsymbol{X};\, \boldsymbol{B}|\boldsymbol{Y_2}) \\ &= H(\boldsymbol{B}) - H(\boldsymbol{B}|\boldsymbol{X}) + h(\boldsymbol{Y_2}|\boldsymbol{B}) - h(\boldsymbol{Z_2}) - H(\boldsymbol{B}|\boldsymbol{Y_2}) \\ &\quad + H(\boldsymbol{B}|\boldsymbol{X}, \boldsymbol{Y_2}) \\ &\ge n\left(H(B) + \frac{1}{n} h(\boldsymbol{Y_2}|\boldsymbol{B}) - h(Z_2) - \delta\right) \end{aligned} \tag{22}$$

where we used (20) to upper-bound $\frac{1}{n} H(\boldsymbol{B}|\boldsymbol{Y_2})$ by δ, and we used the Markov chain relation to cancel $H(\boldsymbol{B}|\boldsymbol{X})$ with $H(\boldsymbol{B}|\boldsymbol{X}, \boldsymbol{Y_2})$. Using (21) with $i = 1$, we upper-bound $I(\boldsymbol{S};\, \hat{\boldsymbol{S}}_1)$ in terms of $h(\boldsymbol{Y_1}|\boldsymbol{B})$

$$\begin{aligned} I(\boldsymbol{S};\, \hat{\boldsymbol{S}}_1) &\le I(\boldsymbol{X};\, \boldsymbol{Y_1}) \\ &= I(\boldsymbol{X};\, \boldsymbol{B}) + h(\boldsymbol{Y_1}|\boldsymbol{B}) - h(\boldsymbol{Z_1}) - I(\boldsymbol{X};\, \boldsymbol{B}|\boldsymbol{Y_1}) \\ &\le I(\boldsymbol{X};\, \boldsymbol{B}) + h(\boldsymbol{Y_1}|\boldsymbol{B}) - h(\boldsymbol{Z_1}) \\ &\le n\left(H(B) + \frac{1}{n} h(\boldsymbol{Y_1}|\boldsymbol{B}) - h(Z_1)\right) \end{aligned} \tag{23}$$

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

$$2^{\frac{2}{n} h(\boldsymbol{Y_2}|\boldsymbol{B})} \ge 2^{\frac{2}{n} h(\boldsymbol{Y_1}|\boldsymbol{B})} + 2^{\log(2\pi e(\sigma_2^2 - \sigma_1^2))}. \tag{24}$$

Combining (22)–(24), and (2) yields the desired result. □

Note that the proof of Lemma 2 does not rely on a specific connection between B and S or a specific distortion measure.

Theorem 2 (Outer Bound): The distortion pair (D_1, D_2) for sending the Gaussian mixture source S of (1) over a Gaussian broadcast channel with $C_2 > H(B)$, may only be achievable if

1)

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

2)

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

3)

$$D_1 \ge \frac{\sigma_n^2}{1 + P'_G(D_2)/\sigma_1^2} \cdot \frac{1}{g(2Q(l))} \tag{27}$$

where

$$\begin{aligned} P'_G(D_2) &\triangleq \frac{\sigma_2^2 + P}{2^{2H(B)}} \cdot g(p'(D_2)) - \sigma_2^2 \\ p'(D_2) &= \min\left\{1,\, \frac{4}{l^2} \cdot \frac{D_2}{\sigma_n^2} + 2Q\left(\frac{l}{2}\right)\right\} \end{aligned}$$

and the function $g(\cdot)$ is defined as

$$g(x) \triangleq 2^{2(H_b(x) + x\log(m-1))}.$$

Note that $g(p'(D_2)),\, g(2Q(l)) \to 1$ as $l \to \infty$ for any $D_2 < \infty$.

Proof: By the definition of the rate-distortion function we have $R_S(D_1) \le \frac{1}{n} I(\boldsymbol{S};\, \hat{\boldsymbol{S}}_1)$. Combining this with (6), (7), (18), (19), and (12) proves condition 3) in the theorem. Conditions 1) and 2) follow from Shannon's joint source–channel coding theorem for point-to-point channels [14]. This proves the theorem. □

Returning to Fig. 4, we observe the outer bound in two cases: the "o" line is for $l = 14$ and the dashed line is for $l = 50$. We show these values since for smaller values of l, the outer bound is dominated by conditions 1) and 2) of Theorem 2. Note that for fixed σ_n and $l \to \infty$, $D_1 = R_S^{-1}(C_1)$ cannot be obtained for any finite D_2.

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 2 establishes an outer bound which depends on l. As $l \to \infty$, the outer bound of Theorem 2 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 3 (Tightness of Bounds For Highly Separated Modes): The distortion region for sending the Gaussian mixture source S of (1) in the limit as $l \to \infty$ with fixed σ_n, over a Gaussian broadcast channel with $C_2 > H(B)$, is composed of all pairs (D_1, D_2) satisfying $D_1 \ge D_1^*$ and $D_2 \ge D_2^*$, where D_1^* and D_2^* were defined in Theorem 1. Furthermore, the distortion pair (D_1^*, D_2^*) is achievable by the digital-over-analog scheme of Fig. 3.

Proof: The achievability of (D_1^*, D_2^*) follows from Theorem 1. As for the converse part, note that following the definitions of $p'(D_2)$, $g(\cdot)$, and $P'_G(D_2)$ in (27), we have

$$\lim_{l\to\infty} g(p'(D_2)) = \lim_{l\to\infty} g(2Q(l)) = 1$$

and $\lim_{l\to\infty} P'_G(D_2) = P_G^*$. Hence, by (27) and the definition of D_1^*, we have $\liminf_{l\to\infty} D_1 \ge D_1^*$. On the other hand, combining (11) with (26) yields $\liminf_{l\to\infty} D_2 \ge D_2^*$. Thus, asymptotically the distortions cannot be better than (D_1^*, D_2^*). □

Theorem 3 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.

III. Conclusion

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. In [13], we also discuss the case of dependent B and N.

We believe that some of the ideas we used can be developed further. For example, consider a general analog source, which is quantized by a vector quantizer. We can regard the quantizer output as the "digital" part of the source, and the quantization error as the "analog" part. We can then construct an encoding scheme, similar to the one presented here, and analyze it with similar tools [15].

Acknowledgment

The authors wish to thank the anonymous reviewers for some useful comments and corrections. R. Zamir acknowledges an early discussion with E. Shlomot which inspired some of the ideas presented.

References

[1] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

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

[3] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 269–274, Mar. 1991.

[4] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Trans. Inform. Theory*, vol. 40, pp. 253–259, Jan. 1994.

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

[6] 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.

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

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

[9] S. Shamai (Shitz), S. Verdú, and R. Zamir, "Systematic lossy source-channel coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 564–579, Mar. 1998.

[10] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 279–280, Mar. 1974.

[11] T. Berger, *Rate-Distortion Theory: A Mathematical Basis for Data-Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[12] Z. Reznic, R. Zamir, and M. Feder, "Joint source-channel coding of a Gaussian mixture source over the Gaussian Broadcast channel," in *Proc. 36th Annu. Allerton Conf. Communication, Control and Computing*, 1998, pp. 800–809.

[13] ——, Joint source-channel coding of a Gaussian mixture source over the Gaussian broadcast channel. Tel-Aviv Univ., Tech. Rep. EES2001-1, Tel-Aviv, Israel. [Online]. Available: http://www.eng.tau.ac.il/Pages/Departments/systems.html

[14] C. E. Shannon, "A mathematical theory of communication," *Bell. Syst. Tech. Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[15] Z. Reznic and R. Zamir, "On the transmission of analog sources over channels with unknown SNR ," paper. Submitted to 2002 IEEE International Symposium on Information Theory, Lausanne, Switzerland, June 30–July 5, 2002.

Multiple Description Vector Quantization With a Coarse Lattice

Vivek K Goyal, *Member, IEEE*, Jonathan A. Kelner, and Jelena Kovačević, *Fellow, IEEE*

***Abstract*—A multiple description (MD) lattice vector quantization technique for two descriptions was recently introduced in which fine and coarse codebooks are both lattices. The encoding begins with quantization to the nearest point in the fine lattice. This encoding is an inherent optimization for the decoder that receives both descriptions; performance can be improved with little increase in complexity by considering all decoders in the initial encoding step. The altered encoding relies only on the symmetries of the coarse lattice. This allows us to further improve performance without a significant increase in complexity by replacing the fine lattice codebook with a nonlattice codebook that respects many of the symmetries of the coarse lattice. Examples constructed with the two-dimensional (2-D) hexagonal lattice demonstrate large improvement over time sharing between previously known quantizers.**

***Index Terms*—Codebook optimization, high-rate source coding, lattice vector quantization.**

I. Introduction

By using the additional structure of a lattice codebook, lattice vector quantizers can be implemented much more efficiently than their more general counterparts. By labeling the points of a lattice with ordered pairs of points in a sublattice, Servetto, Vaishampayan, and Sloane (SVS) create the two descriptions of a *multiple description* (MD) lattice vector quantizer that achieves similar performance gains over unconstrained MD vector quantizers [1], [2]. However, these quantizers turn out to be inherently optimized for the central decoder. This correspondence describes the result of modifying the encoding and decoding used by SVS to minimize a weighted combination of central and side distortions, while keeping the index assignments generated by their elegant theory. This generalization creates a continuum of quantizers for each SVS quantizer, improving the convex hull of operating points. It does this while retaining most of the computational advantages of lattice codebooks.

A. MD Coding

In an MD coding scenario, sequences of symbols are sent separately on two or more channels. Each sequence of channel symbols is called a *description*, and decoders are designed for each nonempty subset of the descriptions. Such a system with two channels is depicted in Fig. 1. This is a generalization of usual "single description" source coding.

We shall focus on the case in which the encoder receives an independent and identically distributed (i.i.d.) sequence of source symbols $\{X_k\}_{k=1}^{K}$ to communicate to three receivers over two noiseless (or error-corrected) channels. One decoder (the *central decoder*) receives information sent over both channels while the remaining two decoders

Manuscript received September 11, 2000; revised October 1, 2001. The material in this correspondence was presented in part at the IEEE Data Compression Conference, Snowbird, UT, March 2000. This work was initiated and largely completed while V. K Goyal and J. A. Kelner were with Bell Labs.

V. K Goyal is with Digital Fountain, Inc., Fremont, CA 94538 USA (e-mail: v.goyal@ieee.org).

J. A. Kelner is with Harvard University, Cambridge, MA 02138 USA (e-mail: kelner@fas.harvard.edu).

J. Kovačević is with Bell Labs, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: jelena@bell-labs.com).

Communicated by P. A. Chou, Associate Editor for Source Coding.

Publisher Item Identifier S 0018-9448(02)00857-X.