

Hybrid Digital-Analog Codes for Source-Channel Broadcast of Gaussian Sources over Gaussian Channels

Vinod M. Prabhakaran, *Member, IEEE*, Rohit Puri,
and Kannan Ramchandran, *Fellow, IEEE*

Abstract

The problem of broadcasting a parallel Gaussian source over an additive white Gaussian noise broadcast channel under the mean-squared error distortion criterion is studied. A hybrid digital-analog coding strategy which combines source coding with side information, channel coding with side information, layered source coding, and superposition broadcast channel coding is presented. When specialized to the open problem of broadcasting a white Gaussian source over an additive white Gaussian noise broadcast channel with bandwidth mismatch which has been the subject of several previous investigations, this coding scheme strictly improves on the state-of-the-art.

Keywords: Source-channel coding, source broadcasting, parallel Gaussian source, broadcast channel, hybrid-digital-analog, Wyner-Ziv, Gel'fand-Pinsker, dirty-paper coding, MMSE estimation, bandwidth mismatch.

I. INTRODUCTION

In this paper, we study the problem of broadcasting the same source to a set of receivers over a common channel. The objective is to devise an encoding strategy (which utilizes a common power and channel bandwidth resource) to simultaneously deliver different qualities of service depending on the quality of the channel experienced by the receivers.

For the problem of transmitting a memoryless Gaussian source over a memoryless additive Gaussian noise point-to-point channel (*i.e.*, when there is only one receiver) operating at the same symbol rate (in other words, when the memoryless source and channel have the same bandwidth), Goblick recognized that the uncoded transmission strategy of transmitting the source samples scaled so as to meet the average encoder power constraint, followed by the optimal linear minimum mean-squared error estimation of the source samples from the channel observations at the receiver, results in the optimal delivered quality (measured in mean-squared error (MSE) distortion) [1]. Since the transmitter remains the same irrespective of the channel noise variance, the same strategy is optimal even when there are multiple receivers (*i.e.*, it is optimal for broadcasting over a memoryless Gaussian broadcast channel [2]). However, the problem remains open when the source and channel bandwidths are mismatched [3], [4], [5], or more generally, when the source has memory [6].

One obvious digital approach to the bandwidth mismatched problem is to use the classical separation method of scalable source coding [7] followed by a degraded-message-set broadcast channel coding [8]. In this approach, a coarse source layer is communicated as a common message intended for all users and a refinement layer is communicated only to some of the users. For the two-user problem, employing this simple scheme, we make the following observation (which is proved in the appendix I)

Part of this work was presented at Allerton Conference 2005 [6]. Vinod M. Prabhakaran is with the Coordinated Science Laboratory, University of Illinois, Urbana-Champaign, Urbana, IL 61801, USA, e-mail: vinodmp@uiuc.edu. Rohit Puri may be contacted at rpuri@eecs.berkeley.edu. Kannan Ramchandran is with the Wireless Foundations, Dept. of EECS, University of California, Berkeley, CA 94702, USA, e-mail: kannanr@eecs.berkeley.edu.

Proposition 1: For a memoryless Gaussian source, and a memoryless Gaussian broadcast channel, the gap in mean-squared error distortion (measured in dB) achieved by the optimal point-to-point source-channel coder and that achieved by the separation approach above can be upperbounded by a constant which depends only on the bandwidth mismatch. In particular,

$$\frac{1}{2} \log_2 \left(\frac{D_k}{D_k^{\text{optimal}}} \right) \leq \frac{\text{BW}_{\text{channel}}}{\text{BW}_{\text{source}}}, \quad k = 1, 2,$$

where D_1 and D_2 are the MSE distortions incurred by the receivers 1 and 2 respectively, D_k^{optimal} is the optimal point-to-point distortion to receiver k , $k = 1, 2$, and $\text{BW}_{\text{source}}$ and $\text{BW}_{\text{channel}}$ are the source and channel bandwidths respectively.

The above proposition upperbounds the gap between the trivial lowerbound on the distortion (namely, the optimal point-to-point distortion) and the distortion achieved by the separation scheme. The gap to optimality of the separation scheme for any number of users was studied recently in [9]. Better achievable strategies have been proposed by Mittal and Phamdo in [4] and for the bandwidth expansion case (where the bandwidth of the channel is larger than that of the source) in [3] and more recently by Reznic, Feder and Zamir in [5]. Also, an improved upperbound for the bandwidth expansion case is available in [5]. In this paper, we consider a slightly more general problem which allows for multiple independent source components and propose an improved achievable strategy. When our solution is specialized to the memoryless source and memoryless channel setting with bandwidth mismatch, it improves the methods proposed in [3], [4]. We would also like to point out that for the special case of bandwidth expansion, our scheme essentially matches the proposal of Reznic, Feder, and Zamir [5]. We obtain a slight improvement over [5] through a generalization overlooked there.

For the case of a memoryless Gaussian source communicated over an additive memoryless Gaussian noise broadcast channel operating at the same symbol rate, the observation of Goblick mentioned earlier gives a rather simple optimal scheme – transmit the source samples scaled so as to meet the average encoder power constraint, and the receivers perform the optimal linear minimum mean-squared error estimation of the source samples from their respective channel observations. This illustrates an interesting feature of analog methods – their ability to enable simultaneous enjoyment of the power and bandwidth resource by each of the broadcast receivers.

On the other hand, the obvious digital approach to the above problem is to use the classical separation method of scalable source coding [7] followed by degraded message-set broadcast channel coding [8]. In this approach, the coarse source layer is communicated as a common message intended for all users and the refinement layer is communicated only to some of the users. Consider a white Gaussian broadcast channel with two receivers; let us call the one with the lower noise variance *strong* and the other *weak*. While the common portion of the information, being limited by the weak receiver, is sub-optimal for the strong receiver, the refinement portion is completely unusable by the weak receiver and in fact acts as interference to it. Thus unlike the analog methods, the digital approach necessarily involves a “splitting” of the total system resource.

While the above discussion illustrates the power of analog methods, real-world sources are characterized by a high degree of memory, and thus they are far from the memoryless model. Parallel source models describe these sources more effectively than a memoryless model. For this case, in a point-to-point set-up, analog transmission is sub-optimal in general. For a parallel Gaussian source with m source components and an additive memoryless Gaussian noise channel model with an equal number of sub-channels, the loss in performance of the analog approach with respect to the digital approach for sufficiently large transmit powers can be shown to be

$$\frac{\text{Analog MSE Distortion}}{\text{Digital MSE Distortion}} = \left(\frac{(\sigma_1 + \sigma_2 + \dots + \sigma_m)/m}{(\sigma_1 \sigma_2 \dots \sigma_m)^{1/m}} \right)^2 \quad (1)$$

where σ_j^2 denotes the variance of the j -th source component [10], [11]. Thus, this gap grows with the memory of the source and can be arbitrarily large.

This motivates the main question posed in this paper: what is an efficient way to broadcast parallel Gaussian sources over memoryless two-user Gaussian broadcast channels? Our solution is driven by aiming to extract the best of both the analog and the digital worlds. We do this by invoking a hybrid uncoded-coded strategy, where the coded system uses a combination of the tools of successive refinement source coding [7], source coding with side-information or Wyner-Ziv (WZ) coding [12], super-position broadcast channel coding [2], and channel coding with side-information or Gel'fand-Pinsker (GP) coding [13] or dirty-paper coding [14]. We would like to point out that this remains an open problem in general and we present an achievable strategy which constitutes the state-of-the-art to the best of our knowledge.

In the next section we present the problem setup. We then proceed by first considering two special cases in section III: (i) when the weak user obtains its point-to-point optimal performance, and (ii) when the strong user obtains its point-to-point optimal performance. In the first case, we design a hybrid analog-digital scheme by modifying the ideas of successive refinement source coding and superposition channel coding. The idea which was presented in a conference version of this paper [6] was also independently explored in [15] where a point-to-point setting with a memoryless source and channel without any bandwidth mismatch was considered. In the second case, we present a hybrid scheme based on the ideas of source coding with side-information and channel coding with side-information. Section IV then considers a scheme which combines all these ideas to obtain a trade-off between the qualities of reproductions at the receivers. We take up the special case of memoryless Gaussian source and memoryless Gaussian broadcast channel with bandwidth mismatch in section V. We conclude with some comments on potential directions of research.

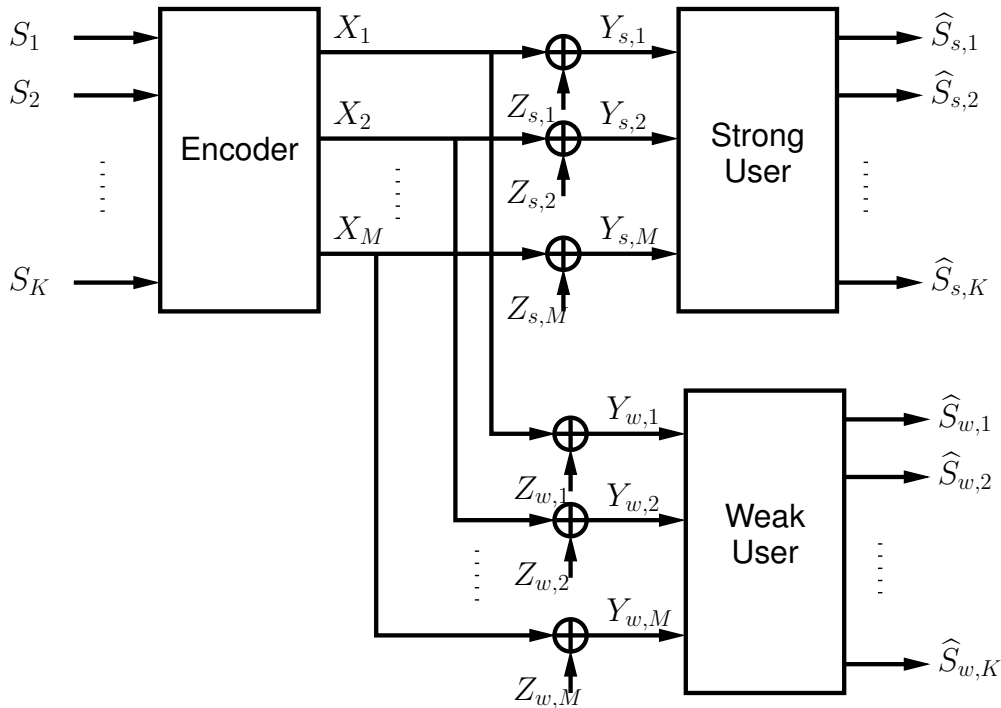


Fig. 1. Problem setup.

II. PROBLEM SETUP

The setup is shown in Fig. I. In this paper, we only consider the case where all the sub-channels to a particular user have the same statistics. The upshot of this assumption is a simplification which results from recognizing that there is an ordering of the users according to the noise variance of their channel. We will call the user with the smaller noise variance the *strong user*, and the user with the larger noise

variance the *weak user*¹. Note that the ideas presented in this paper also apply to the case of parallel broadcast channels where the different sub-channels to the same user may not have the same statistics, but we do not explore them here.

We let our source have K independent components. The k -th component is denoted by $S_k(i)$ and it is independent and identically distributed (i.i.d.) over the time index $i = 1, 2, \dots$. Our source is Gaussian, $S_k(i) \sim \mathcal{N}(0, \sigma_k^2)$. Without loss of generality, we will assume that $\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_K^2$. We will explicitly model the fact that the source and channel bandwidths do not necessarily match. We will assume that there are M parallel broadcast sub-channels of the same statistics. When $M = K$, the bandwidth of the source matches the bandwidth of the channel. The weak user observes $Y_{wm}(i) = X_m(i) + Z_{wm}(i)$, $m = 1, 2, \dots, M$, $i = 1, 2, \dots$, and the strong user observes $Y_{sm}(i) = X_m(i) + Z_{sm}(i)$, where $X_m(i)$ is the input to the m -th sub-channel. The noise processes are $Z_{wm}(i)$ and $Z_{sm}(i)$ independent i.i.d. (over m and i) Gaussians with variances N_w and N_s respectively, where $N_s < N_w$. The source-channel encoder $f^n : \mathbb{R}^{Kn} \rightarrow \mathbb{R}^{Mn}$ maps an n -length block of the source to an n -length block of the channel input. There is an average power constraint on the encoder so that $(\sum_{i=1}^n \sum_{m=1}^M X_m^2(i)) / (nM) \leq P$. The source-channel decoders $g_s^n : \mathbb{R}^{Mn} \rightarrow \mathbb{R}^{Kn}$ and $g_w^n : \mathbb{R}^{Mn} \rightarrow \mathbb{R}^{Kn}$ at the strong and weak user, respectively, reconstruct n -length blocks,

$$\begin{aligned} \left\{ \widehat{S}_{sk}(i), i = 1, \dots, n, k = 1, \dots, K \right\} &= g_s^n (\{Y_{sm}(i), i = 1, 2, \dots, n, m = 1, 2, \dots, M\}), \text{ and} \\ \left\{ \widehat{S}_{wk}(i), i = 1, \dots, n, k = 1, \dots, K \right\} &= g_w^n (\{Y_{wm}(i), i = 1, 2, \dots, n, m = 1, 2, \dots, M\}) \end{aligned}$$

of the source from n -length blocks of the channel outputs. Distortions are measured as the average of the mean-squared error distortion over all source components

$$D_j^n = \frac{1}{nK} \sum_{i=1}^n \sum_{k=1}^K \left(S_k(i) - \widehat{S}_{jk}(i) \right)^2, \quad j \in \{s, w\}.$$

A pair of distortions (D_s, D_w) will be said to be *achievable* if for any $\epsilon > 0$, for sufficiently large n , there is (f^n, g_s^n, g_w^n) such that $D_j^n \leq D_j + \epsilon$, $j \in \{s, w\}$. The problem is to characterize the region of all the achievable distortions (D_s, D_w) for a given transmit power P . This remains open. In the next sections, we present our inner bound to the region (*i.e.*, an achievable region).

III. AN ACHIEVABLE SOLUTION: THE EXTREME POINTS

In this section we present all the key ideas involved in our achievable strategy. We consider two extreme cases – when the weak user achieves its point-to-point optimal quality, and when the strong user achieves its point-to-point optimal quality – to illustrate two complementary strategies which together constitute the general achievable solution.

A. Weak-user-optimal case

From Shannon's separation theorem [16, pg. 216] we know that the optimal solution for the point-to-point source-channel problem can be obtained by the separation principle of first optimally source coding and then transmitting the resulting bit stream using an optimal channel code. Thus the lowest distortion D_w^* attainable by the weak user is given by "reverse water-filling" over the source components [16, pg. 348]

¹It is known that the performance of such a broadcast channel is identical to that of a *degraded* broadcast channel where the weak user receives the signal received by the strong user but further corrupted by an additive Gaussian noise independent of the additive noise corrupting the strong user's channel and which has a variance equal to the difference between the variances of the additive noises of the weak user and the strong user in the original channel. This fact is often expressed by saying that the original broadcast channel is *stochastically degraded*. A proof of the above stated equivalence for the case of channel coding appears in [16, pg. 422]. The same idea can be used to show an equivalence for the problem of interest here though we do not need to make use of this equivalence in our discussion here.

$$D_w^* = \frac{1}{K} \sum_{k=1}^K D_k, \text{ where } D_k = \begin{cases} \mu, & \text{if } \mu < \sigma_k^2, \\ \sigma_k^2, & \text{if } \mu \geq \sigma_k^2, \end{cases} \quad (2)$$

where μ is chosen such that the total rate $(1/2) \sum_{k=1}^K \log(\sigma_k^2/D_k)$ equals the capacity C_w of the weak user's channel. C_w is in turn given by $(1/2) \sum_{m=1}^M \log(1 + P_m/N_w)$ where we choose $P_1 = P_2 = \dots = P_M = P$.

If this separation strategy is followed for the broadcast case as well, the strong user also recovers the source at a distortion D_w^* . However, without compromising the quality of reproduction for the weak user, better quality can be delivered to the strong user. Before presenting our solution in full generality, it is useful to consider the special case of $K = M = 2$ and $\sigma_1^2 > \sigma_2^2$; see Fig. 3. Let us suppose that the optimal point-to-point reverse water-filling solution for the weak user allocates distortions D_1 and D_2 for the source components S_1 and S_2 , respectively. Also, let us denote the powers allocated to the sub-channels X_1 and X_2 by P_1 and P_2 respectively. $P_1 = P_2 = P$. Then

$$\frac{1}{2} \log \left(\frac{\sigma_1^2}{D_1} \frac{\sigma_2^2}{D_2} \right) = \frac{1}{2} \log \left(\frac{P_1 + N_w}{N_w} \frac{P_2 + N_w}{N_w} \right).$$

We can source code S_1 using a successive refinement strategy thereby producing two bit streams: a coarse description at distortion D'_1 and a refinement stream which refines from D'_1 to D_1 . Since Gaussian sources are successively refinable [7], this can be done without loss of optimality for the weak user. We choose D'_1 such that the bitrate of the refinement stream is equal to the rate at which the first sub-channel operates. *i.e.*,

$$\frac{1}{2} \log \left(\frac{D'_1}{D_1} \right) = \frac{1}{2} \log \left(1 + \frac{P_1}{N_w} \right).$$

Combining the two equations above gives

$$\frac{1}{2} \log \left(\frac{\sigma_1^2}{D'_1} \frac{\sigma_2^2}{D_2} \right) = \frac{1}{2} \log \left(1 + \frac{P_2}{N_w} \right).$$

In other words, without loss of optimality, we may send the coarse description for S_1 and the bit stream for S_2 over the second sub-channel, and the refinement bitstream for S_1 over the first sub-channel; see Fig. 2(a). This further suggests that instead of sending the refinement bitstream over the first sub-channel, we may send uncoded the quantization error resulting from the coarse quantization of S_1 appropriately scaled to satisfy the power constraint. The input to the first sub-channel will be $\sqrt{(P_1/D'_1)} (S_1(i) - \hat{S}'_1(i))$, where $\hat{S}'_1(i)$ is the i -th sample of the coarsely quantized version of S_1 . It is easy to see that this satisfies the power constraint on the first sub-channel and also results in no loss of optimality for the weak user. The second fact is analogous to the optimality of uncoded transmission for the point-to-point Gaussian source-channel problem. The strong user can achieve a lower distortion on S_1 because of the uncoded transmission of the quantization error. The strong user estimates the refinement component as $(P_1/(P_1 + N_s)) \sqrt{D'_1/P_1} Y_1$ and adds it to the coarse description to form its reproduction of S_1 . We note that the resulting distortion for the strong user on S_1 is $D'_1/(1 + P_1/N_s)$.

The performance of the strong user can be further improved. Without losing optimality for the weak user, we may send the coarse description of S_1 and the bit stream for S_2 using superposition coding over the second sub-channel. In particular, we send the coarse quantization bit stream of S_1 using power $P_2 - P'_2$ (defined below) and the bit stream for S_2 using power P'_2 such that the decoder can first decode the former bit stream assuming the latter as interference. The decoder then cancels the interference from the bit stream for S_1 and decodes the bit stream for S_2 . Thus P'_2 is given by

$$\frac{1}{2} \log \left(\frac{\sigma_1^2}{D'_1} \right) = \frac{1}{2} \log \left(1 + \frac{P_2 - P'_2}{P'_2 + N_w} \right).$$

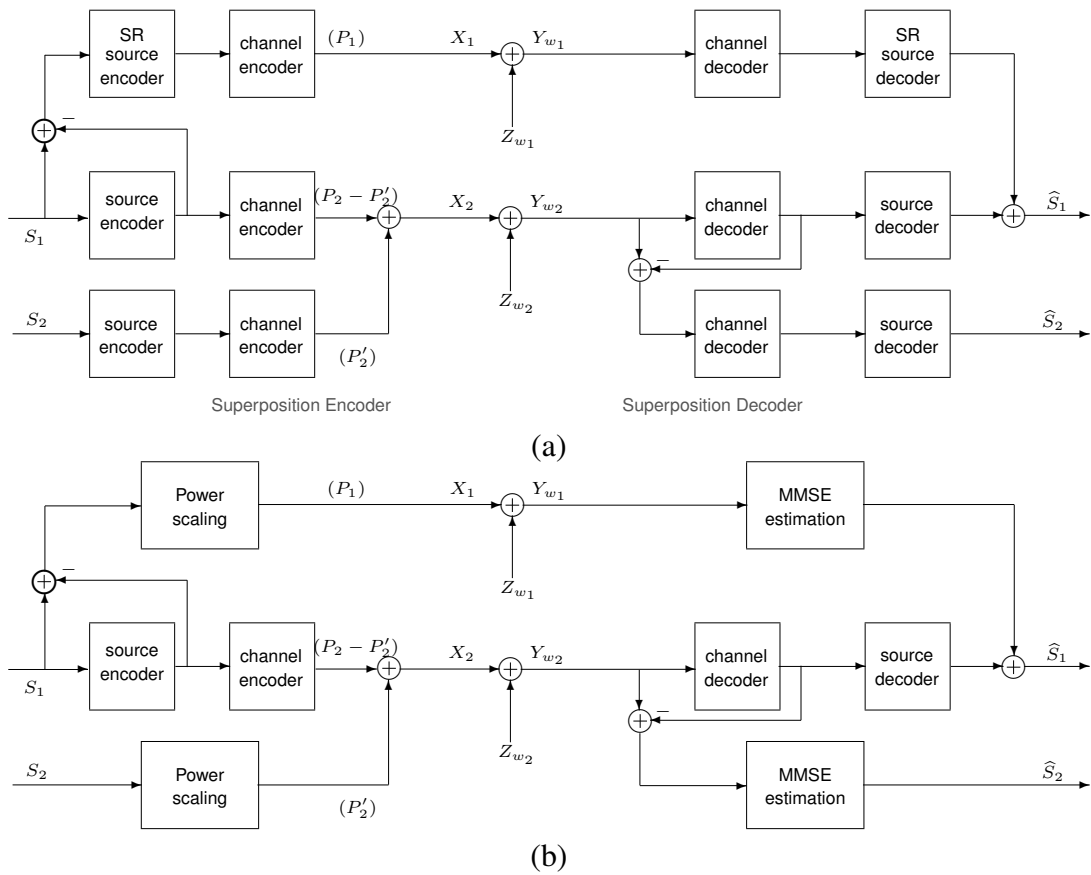


Fig. 2. Weak-user-optimal case: (a) separation scheme showing successive refinement (SR) and superposition coding, and (b) the hybrid digital-analog scheme.

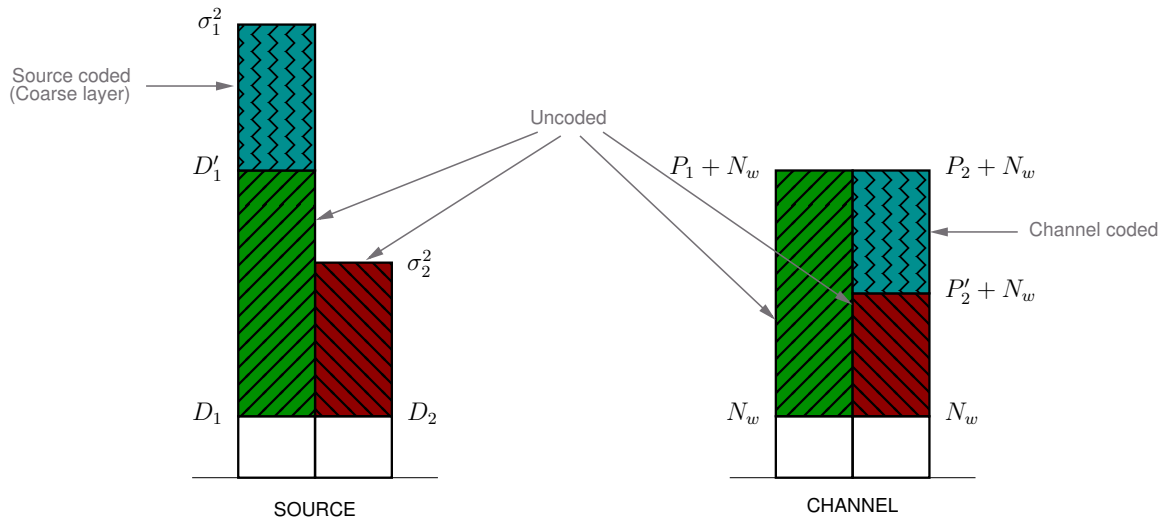


Fig. 3. Weak-user-optimal case: The schematic diagram shows the allocation for a $K = M = 2$ example. The optimal coded separation scheme for the weak-user may be thought of as sending source-coded bits about the second source component S_2 and the coarse-layer bits from a successive refinement source coding of the first source component S_1 over the second sub-channel using superposition coding, and the refinement-layer bits from the successive refinement coding of S_1 alone over the first sub-channel. An equivalent distortion performance can be achieved at the weak user while improving the strong user's performance by (i) sending the quantization error from the coarse quantization of S_1 uncoded (scaled) over the first sub-channel, and (ii) sending S_2 uncoded (scaled) over the second sub-channel and the coarse-layer bits on S_1 channel coded with sufficient power to be decoded (by the weak user) treating the uncoded transmission of S_2 as noise. The strong user benefits from both the uncoded transmissions since it can form better quality estimates than the weak user.

This also gives the relation

$$\frac{1}{2} \log \left(\frac{\sigma_2^2}{D_2} \right) = \frac{1}{2} \log \left(1 + \frac{P'_2}{N_w} \right)$$

which indicates why decoding of the bit stream for S_2 after interference cancellation succeeds. This scheme suggests that we may send S_2 uncoded using power P'_2 instead of sending its quantized bits. Since, after canceling the interference from the S_1 bit stream, the channel to the weak user is an additive white Gaussian noise channel, the optimality of this scheme follows from the optimality of uncoded transmission for point-to-point Gaussian source-channel coding. The strong user can now reconstruct S_2 at a lower distortion, $D_2(1 + P'_2/N_w)/(1 + P'_2/N_s)$. The scheme is summarized in Fig. 2(b). The overall distortion achieved by the strong user is

$$D_s = \frac{1}{2} \left(\frac{1 + \frac{P_1}{N_w}}{1 + \frac{P_1}{N_s}} D_1 + \frac{1 + \frac{P'_2}{N_w}}{1 + \frac{P'_2}{N_s}} D_2 \right).$$

The extension to $K = M > 2$ is straightforward. Let us assume without loss of generality that under the point-to-point optimal inverse water-filling solution for the weak user, the first L source components satisfy $(1/2) \log(\sigma_k^2/D_k) \geq (1/2) \log(1 + P_k/N_w)$. For these L components we define D'_k such that $(1/2) \log(D'_k/D_k) = (1/2) \log(1 + P_k/N_w)$. The k -th such component ($k \leq L$) is source coded to a distortion of D'_k and the resulting error is sent uncoded (scaled by $\sqrt{P_k/D_k}$) over the k -th sub-channel. For sub-channels $m > L$, we define P'_m as $(1/2) \log(\sigma_m^2/D_m) = (1/2) \log(1 + P'_m/N_w)$. The m -th such component is sent uncoded over the m -th sub-channel scaled by $\sqrt{P'_m/\sigma_m^2}$. The rest of the power ($P_m - P'_m$) for these sub-channels $m > L$ are used to send the source coded bits from the first L components. On these sub-channels, the decoders first decode these bits, cancel the interference caused by them, and then estimate the source components. The first L source components are estimated directly from the corresponding sub-channel outputs. Thus, without compromising the quality of reproduction for the weak user, the strong user achieves a lower distortion.

$$D_s = \frac{1}{K} \left(\sum_{k=1}^L \frac{1 + \frac{P_k}{N_w}}{1 + \frac{P_k}{N_s}} D_k + \sum_{k=L+1}^K \frac{1 + \frac{P'_k}{N_w}}{1 + \frac{P'_k}{N_s}} D_k \right) < \frac{1}{K} \sum_{k=1}^K D_k.$$

This scheme directly extends to the $K \neq M$ case. When $K < M$ (bandwidth expansion), the extra sub-channels will be used to send additional coded bits. For the case of $K > M$ (bandwidth contraction), only at most M source components can be sent (wholly or partially) uncoded. In summary, we have the following

Theorem 2: For the source-channel problem in Section II (D_s, D_w^*) is achievable, where D_w^* is given by (2) and D_s is as defined below.

$$D_s = \frac{1}{K} \left(\sum_{k=1}^L \frac{1 + \frac{P_k}{N_w}}{1 + \frac{P_k}{N_s}} D_k + \sum_{k=L+1}^{K'} \frac{1 + \frac{P'_k}{N_w}}{1 + \frac{P'_k}{N_s}} D_k + \sum_{k=K'+1}^K D_k \right),$$

where D_k 's are given by (2), $P_k = P$,

$$L = \min \left\{ \left| \left\{ k : \frac{\sigma_k^2}{D_k} \geq 1 + \frac{P_k}{N_w} \right\} \right|, M \right\},$$

$$K' = \min \left\{ \left| \left\{ k : \mu \leq \sigma_k^2 \right\} \right|, K \right\},$$

and the P'_k 's are defined by

$$\frac{\sigma_k^2}{D_k} = 1 + \frac{P'_k}{N_w}, \quad k = L + 1, \dots, K'.$$

B. Strong-user-optimal case

If the point-to-point separation approach is used for providing optimal fidelity to the strong user, since the rate of transmission is greater than the channel capacity of the weak user, the weak user will not be able to get any useful information. However, in this subsection we show that we can provide useful information to the weak user without compromising the strong user's performance. As will become clear, in this case, the weak user's receiver will only involve scaling the signals received on different sub-channels, *i.e.*, it will be an analog receiver. An extension of the scheme in this subsection was shown in [11] to obtain the entire optimal distortion trade-off region for broadcasting a parallel Gaussian source over a parallel Gaussian broadcast channel to two receivers when one of the receivers is restricted to be a linear filter (but without no assumptions on the relative strengths of the channels to the receivers). The linear filter receiver is a model for a legacy analog receiver in a transitional broadcast system which supports digital and analog receivers.

Let the point-to-point optimal reverse water-filling solution for the strong user produce a total distortion D_s^* from a distortion allocation D_k , $k = 1, 2, \dots, K$ according to (2), where μ is now chosen so that the total rate equals the capacity $C_s = (1/2) \sum_{m=1}^M \log(1 + P_m/N_s)$, $P_m = P$ of the strong user's channel.

It is again helpful to consider the special case of $K = M = 2$ and $\sigma_1^2 > \sigma_2^2$ (see Fig. 5) before the general case. Note that this example was also used as a starting point for presenting the scheme in [11, Section III]. We summarize the discussion below for completeness. With $P_1 = P_2 = P$, we have

$$\frac{1}{2} \log \left(\frac{\sigma_1^2}{D_1} \frac{\sigma_2^2}{D_2} \right) = \frac{1}{2} \log \left(\frac{P_1 + N_s}{N_s} \frac{P_2 + N_s}{N_s} \right).$$

We source code S_1 using successive refinement (Fig. 4(a)) such that now the bitrate of the coarse description at distortion D_1'' is equal to the rate at which the first sub-channel operates.

$$\frac{1}{2} \log \left(\frac{\sigma_1^2}{D_1''} \right) = \frac{1}{2} \log \left(1 + \frac{P_1}{N_s} \right).$$

Note that this is different from the previous subsection where we set the bitrate of the refinement bit stream equal to the rate of the first sub-channel. Thus

$$\frac{1}{2} \log \left(\frac{D_1'' \sigma_2^2}{D_1 D_2} \right) = \frac{1}{2} \log \left(1 + \frac{P_2}{N_s} \right).$$

Instead of sending the coarse description over the first sub-channel and the refinement bit stream on the second, without losing optimality for the strong user, we may send S_1 uncoded (scaled by $\sqrt{P_1/\sigma_1^2}$) over the first sub-channel and on the second sub-channel a Wyner-Ziv bit stream (of rate equal to that of the refinement bit stream) which assumes that the corrupted version of S_1 from the first sub-channel will be available as side information at the decoder. The optimality of this scheme follows from the no rate-loss property of jointly Gaussian sources under Wyner-Ziv coding [12].

Thus the weak user can now form an estimate of S_1 from its output Y_1 of the first sub-channel. It is also possible to provide the weak user with an estimate of the second component without losing optimality for the strong user. Let us define P_2'' such that the Wyner-Ziv bit stream can be sent using this power, superposition coded with the bit stream for S_2 which uses the rest of the power $P_2 - P_2''$. The decoding order is first the bit stream for S_2 , followed by the Wyner-Ziv bit stream. Thus

$$\frac{1}{2} \log \left(\frac{D_1''}{D_1} \right) = \frac{1}{2} \log \left(1 + \frac{P_2''}{N_s} \right),$$

and

$$\frac{1}{2} \log \left(\frac{\sigma_2^2}{D_2} \right) = \frac{1}{2} \log \left(1 + \frac{P_2 - P_2''}{P_2'' + N_s} \right).$$

We can use dirty-paper coding of Gel'fand and Pinsker [13] and Costa [14] instead of superposition coding to achieve the same rates. Here the channel codeword for the bit stream of S_2 is treated as non-causal side

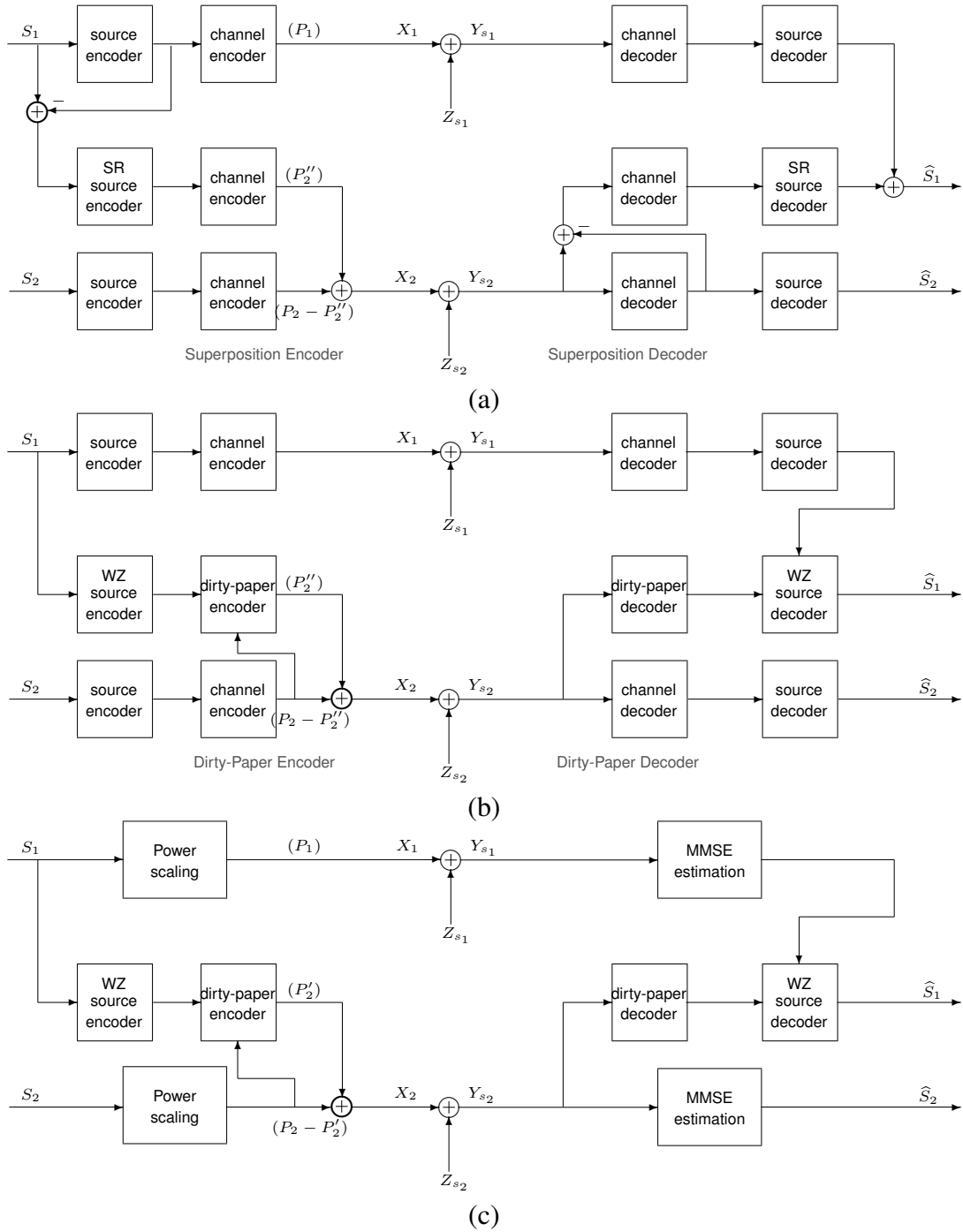


Fig. 4. Strong-user-optimal case (also see [11, Fig. 2]): (a) separation scheme showing successive refinement (SR) and superposition coding, (b) separation scheme with Wyner-Ziv (W-Z) code and dirty-paper coding (DPC), (c) the hybrid digital-analog scheme.

information available at the encoder when channel encoding the Wyner-Ziv bit stream. Fig. 4(b) shows this setup which uses a combination of dirty-paper coding and Wyner-Ziv coding. Note that the decoder does not need to decode the bit stream for S_2 in order to decode the Wyner-Ziv bit stream. This allows us to send S_2 uncoded (scaled by $\sqrt{(P_2 - P_2'')/\sigma_2^2}$) without impacting the optimality for the strong user (Fig. 4(c)). The weak user can now form an estimate of S_2 from Y_2 .

Again, we can easily extend the above intuition to $K = M > 2$. Let L be the number of source components and sub-channels such that $(1/2) \log(\sigma_k^2/D_k) > (1/2) \log(1 + P_k/N_s)$. Since σ_k^2 are monotonically decreasing, these will be the first L components. For these components, we define D_k'' such

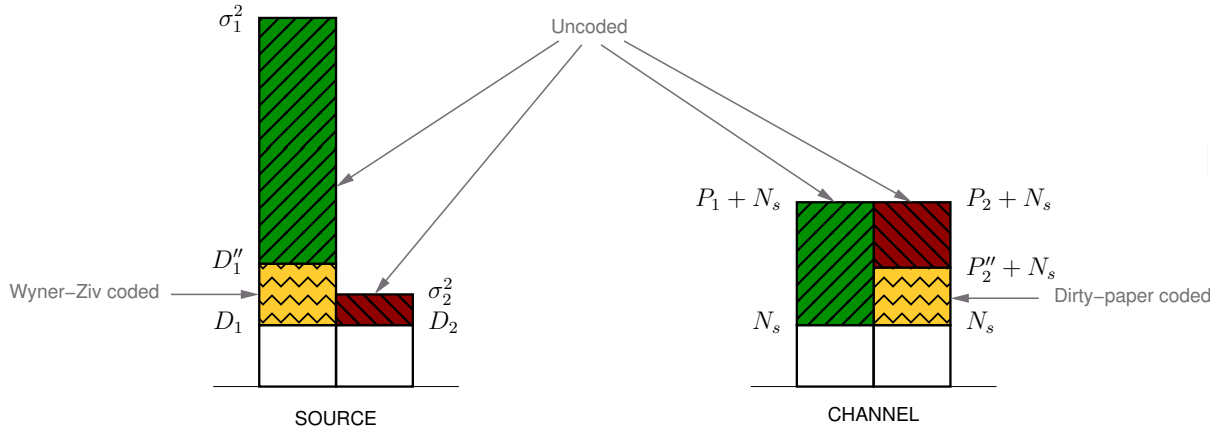


Fig. 5. Strong-user-optimal case: The schematic diagram showing the allocation for a $K = M = 2$ example. The optimal coded separation scheme for the strong user may be thought of as sending the coarse-layer bits from a successive refinement source coding of the first source component S_1 alone over the first sub-channel, and source coded bits of the second source component S_2 and the refinement bits of S_1 over the second sub-channel. However, none of these bits are decodable by the weak user. We may provide useful information without compromising the strong user's performance by (i) sending S_1 uncoded (scaled) over the first sub-channel and (ii) sending S_2 uncoded (scaled) over the second sub-channel and bits carrying refinement information on S_1 using Gel'fand and Pinsker's (dirty-paper) channel coding where the transmission of S_2 acts as Gaussian side-information at the transmitter. Dirty-paper coding ensures that the transmission of S_2 does not affect the rate of transmission of the refinement bits. The bits carrying the refinement information are produced using Wyner-Ziv coding where the noisy observation of S_1 over the first sub-channel acts as side-information at the decoder. While the weak user will be unable to decode the refinement information, it benefits from the two uncoded transmissions.

that $(1/2) \log(\sigma_k^2/D_k'') = (1/2) \log(1 + P_k/N_s)$. For the rest of the sub-channels ($m > L$) we define P_k'' by $(1/2) \log(\sigma_m^2/D_m) = (1/2) \log(1 + (P_m - P_m'')/(P_m'' + N_s))$. The first L source components are sent uncoded scaled by $\sqrt{P_k/\sigma_k^2}$ on their corresponding sub-channels, and rest of the source components are sent uncoded scaled by $\sqrt{(P_k - P_k'')/\sigma_k^2}$ on their corresponding sub-channels. The first L source components are Wyner-Ziv source coded at rates of $(1/2) \log(D_k''/D_k)$ assuming the availability at the decoder (strong user) of the noise corrupted versions sent over the corresponding sub-channels. These source coded bits are sent using dirty-paper coding over the rest of the sub-channels $m > L$. The resulting distortion for the weak user is

$$D_w = \frac{1}{K} \left(\sum_{k=1}^L \frac{\sigma_k^2}{1 + \frac{P_k}{N_w}} + \sum_{k=L+1}^K \frac{\sigma_k^2}{1 + \frac{P_k - P_k''}{N_w}} \right)$$

which is strictly less than $\frac{1}{K} \sum_{k=1}^L \sigma_k^2$, the distortion for the weak user in the separation approach under which no information is decodable by this user.

This scheme also directly extends to the bandwidth expansion ($K < M$) and bandwidth contraction ($K > M$) scenarios. To summarize, we can state the following

Theorem 3: The distortion pair (D_s^*, D_w) is achievable for the source-channel coding problem in section II, where the point-to-point optimal distortion D_s^* for the strong-user and D_w are as follows.

$$D_s^* = \frac{1}{K} \sum_{k=1}^K D_k, \text{ where } D_k = \begin{cases} \mu, & \text{if } \mu < \sigma_k^2, \\ \sigma_k^2, & \text{if } \mu \geq \sigma_k^2, \end{cases} \quad (3)$$

where μ is chosen such that the total rate $(1/2) \sum_{k=1}^K \log(\sigma_k^2/D_k)$ equals the capacity C_s of the strong user's channel. $C_s = \sum_{m=1}^M \log(1 + P_m/N_s)$ where $P_1 = P_2 = \dots = P_M = P$.

Let

$$L = \min \left\{ \left| \left\{ k : \frac{\sigma_k^2}{D_k} \geq 1 + \frac{P_k}{N_s} \right\} \right|, M \right\},$$

$$K' = \min \left\{ \left| \left\{ k : \mu \leq \sigma_k^2 \right\} \right|, K \right\},$$

and the P_k'' 's be defined by

$$\frac{\sigma_m^2}{D_m} = 1 + \frac{P_m - P_m''}{P_m'' + N_s}.$$

Then,

$$D_w = \frac{1}{K} \left(\sum_{k=1}^L \frac{\sigma_k^2}{1 + \frac{P_k}{N_w}} + \sum_{k=L+1}^{K'} \frac{\sigma_k^2}{1 + \frac{P_k - P_k''}{N_w}} + \sum_{k=K'+1}^K \sigma_k^2 \right)$$

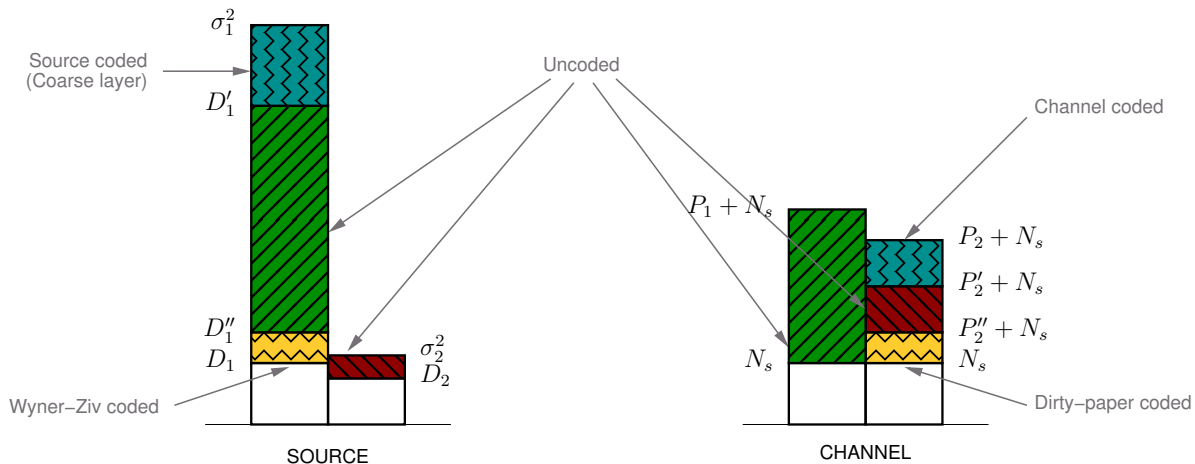


Fig. 6. An achievable trade-off: The schematic diagram shows an allocation for a $K = M = 2$ example. The first source component S_1 is sent in three different ways: (i) a coarse layer source codeword which will be decoded by both users, (ii) an uncoded version of the quantization error which the weak user will estimate from its noisy observation, and (iii) Wyner-Ziv bits on S_1 which only the strong user decodes. In decoding the Wyner-Ziv codeword, the strong user uses as side-information a linear estimate of S_1 using (i) and its noisy observation of (ii). The quantization error (ii) is sent uncoded (scaled) over the first sub-channel using all the power allocated to this sub-channel. Over the second sub-channel, the coarse layer bits from (i) above are sent using a Gaussian channel code using part of the power allocated to this sub-channel. This is meant to be decoded by both users treating the rest of the signals sent over this sub-channel as noise. Using part of the leftover power, the second source is sent uncoded (scaled). Since the codeword carrying bits from (i) is assumed to be successfully decoded by both users, they may estimate the second source component assuming only the rest of the power used in this sub-channel and the channel noise as the disturbance affecting this transmission. The leftover power in this sub-channel is used to send the Wyner-Ziv bits from (iii). This is done using Gel'fand-Pinsker's (dirty-paper) coding treating the scaled version of the second source component sent over this sub-channel as side-information (at the transmitter).

IV. AN ACHIEVABLE TRADE-OFF

We may also trade-off the quality of reproductions at the two users without being optimal for either. Clearly, time sharing between the two achievable extreme points is a possibility. It is often possible to do better. A natural strategy suggested by the above discussion is to combine the schemes for the weak- and strong-user-optimal cases.

Again, we will first consider a $K = M = 2$ example with $\sigma_1^2 > \sigma_2^2$. Let us suppose that the power allocation to the two sub-channels are P_1 and P_2 such that $P_1 + P_2 = 2P$. Consider Fig. 6. The first source component S_1 is sent in three different ways:

- (i) A (coarse layer) source codeword is formed which is meant to be decoded by both users. The source coded bits are sent over the second sub-channel using a Gaussian channel code utilizing part of the power P_2 allocated to this sub-channel. This is meant to be decoded by both users treating the rest of the signals sent over this sub-channel as noise. Let us denote the quantized version of S_1^n by S_1^n and MSE by D_1' . The rate of this coarse layer source codebook is then

$$\frac{1}{2} \log \left(\frac{\sigma_1^2}{D_1'} \right). \quad (4)$$

- (ii) An uncoded (scaled) version of the quantization error resulting from the source coding in (i) is transmitted. This transmission occurs over the first sub-channel using all the power P_1 allocated to this sub-channel, *i.e.*,

$$X_1^n = \sqrt{\frac{P_1}{D_1'}} (S_1^n - S_1^m).$$

The weak user estimates S_1^n from the codeword S_1^m in (i) and its noisy observation of the quantization error over the first sub-channel using a linear estimator

$$\widehat{S}_{w_1}^n = \frac{P_1}{P_1 + N_w} \left(\sqrt{\frac{D_1'}{P_1}} Y_1^n \right) + S_1^m.$$

The resulting MSE is

$$\frac{D_1'}{1 + \frac{P_1}{N_w}}. \quad (5)$$

- (iii) Wyner-Ziv bits on S_1^n are sent intended for the strong user alone. In decoding the Wyner-Ziv codeword, the strong user uses as side-information a linear estimate of S_1^n it forms using (i) and its noisy observation of (ii) in a manner similar to the weak user's estimate of S_1^n above. Let us denote the strong user's estimate of S_1^n by $S_1^{\prime\prime}$ and the MSE of this estimate by D_1'' . We have

$$\begin{aligned} S_1^{\prime\prime} &= \frac{P_1}{P_1 + N_s} \left(\sqrt{\frac{D_1'}{P_1}} Y_1^n \right) + S_1^m \\ &= \frac{P_1}{P_1 + N_s} S_1^n + \left(1 - \frac{P_1}{P_1 + N_s} \right) S_1^m + \frac{P_1}{P_1 + N_s} \sqrt{\frac{D_1'}{P_1}} Z_{s_1}^n, \quad \text{and} \end{aligned} \quad (6)$$

$$D_1'' = \frac{D_1'}{1 + \frac{P_1}{N_s}}. \quad (7)$$

Then, by Wyner-Ziv's theorem, the bit rate needed to achieve a distortion of D_1 on the source S_1^n at the strong user with $S_1^{\prime\prime}$ acting as side-information is²

$$\frac{1}{2} \log \frac{D_1''}{D_1}. \quad (8)$$

The transmission of these Wyner-Ziv bits occurs over the second sub-channel as described below. The transmission over the second sub-channel is meant to deliver the coarse layer source codeword bits about S_1 and an estimate of S_2 to the weak user, and in addition to these, the Wyner-Ziv bits about S_1 as well to the strong user. This is accomplished as follows:

- (a) The source codeword bits from (i) are transmitted using a Gaussian channel code utilizing power $P_2 - P_2'$ (a part of the total power P_2 allocated to this sub-channel). This is decoded by both users treating the rest of the signals sent over this sub-channel as noise. The decoding at the weak user presents the bottleneck to the rate at which bits may be delivered. Hence, to meet the rate required by (4), we must have D_1' , P_2 , and P_2' satisfy

$$\frac{1}{2} \log \left(\frac{\sigma_1^2}{D_1'} \right) = \frac{1}{2} \log \left(1 + \frac{P_2 - P_2'}{P_2' + N_w} \right). \quad (9)$$

²However, note that by (6), our side-information $S_1^{\prime\prime}$ is not S_1^n corrupted by a memoryless Gaussian disturbance as a classical statement of Wyner-Ziv's theorem would require. All we are guaranteed is that the side-information has a MSE of D_1'' with respect to the source. A simple extension of the achievability proof can handle this situation – we may invoke the achievability part with S_1^n and $S_1^{\prime\prime}$ acting as the (vector) symbols; see, for instance, [5, Appendix IV].

- (b) The second source component S_2^n is sent uncoded (scaled) using power $P_2' - P_2''$ (a part of the power P_2' leftover after (a)), *i.e.*, we send $\sqrt{(P_2' - P_2'')/\sigma_2^2}S_2^n$. Since the codeword from (a) is assumed to be successfully decoded by both users, they may strip it off their received signals $Y_{w_2}^n$ and $Y_{s_2}^n$ and estimate the second source component assuming only the rest of the power used in this sub-channel and the channel noise as the disturbance affecting this transmission. Both users employ linear estimators. Let us denote the signals after the codeword from (a) has been stripped off by $Y_{w_2}^m$ and $Y_{s_2}^m$ at the weak user and the strong user, respectively. Then, the estimates of S_2 are

$$\widehat{S}_{j_2}^m = \frac{P_2' - P_2''}{P_2' + N_j} \left(\sqrt{\frac{\sigma_2^2}{P_2' - P_2''}} Y_{j_2}^m \right), \quad j \in \{s, w\}.$$

The MSE D_2 on the second source component incurred by the strong user is given by

$$\frac{1}{2} \log \left(\frac{\sigma_2^2}{D_2} \right) = \frac{1}{2} \log \left(1 + \frac{P_2' - P_2''}{P_2'' + N_s} \right). \quad (10)$$

Similarly, the weak user incurs an average distortion on the second source component of

$$\frac{\sigma_2^2}{1 + \frac{P_2' - P_2''}{P_2'' + N_w}}.$$

- (c) Finally, the Wyner-Ziv bits from (iii) are transmitted intended for the strong user with the leftover power of P_2'' . Let us recall that we have assumed that the codeword from (a) is successfully decoded by the strong user and stripped off its received signal. The only other disturbances affecting the transmission of the Wyner-Ziv bits are the memoryless Gaussian noise in the channel and the scaled transmission of the memoryless Gaussian S_2^n in (b), both of which are independent of each other and the Wyner-Ziv bits being sent. Also, the disturbance S_2^n is known to the transmitter non-causally. This is precisely the setting of Gel'fand-Pinsker or dirty-paper coding by which a rate equal to the capacity of channel in which only the memoryless Gaussian noise is present can be achieved. This rate must be large enough to support the Wyner-Ziv bits whose rate is (8). Thus, P_2'' , D_1'' and D_1 must satisfy

$$\frac{1}{2} \log \left(\frac{D_1''}{D_1} \right) = \frac{1}{2} \log \left(1 + \frac{P_2''}{N_s} \right). \quad (11)$$

To summarize, the decoders can achieve distortions of

$$D_s = (D_1 + D_2)/2, \text{ and}$$

$$D_w = \frac{1}{2} \left(\frac{D_1'}{1 + \frac{P_1}{N_w}} + \frac{\sigma_2^2}{1 + \frac{P_2' - P_2''}{P_2'' + N_w}} \right),$$

for every choice of the non-negative power parameters P_1 , and $P_2 \geq P_2' \geq P_2''$ which satisfy the sum power constraint $P_1 + P_2 = 2P$, and non-negative distortion parameters $D_1 \leq D_1'' \leq D_1' \leq \sigma_1^2$, and $D_2 \leq \sigma_2^2$ provided they satisfy the conditions (5), (7), (9), (10), and (11).

Generalizing the above, in general, we have the following achievable trade-off

Theorem 4: Let $L \in \{0, 1, \dots, \min(K, M)\}$ and $K' \in \{L, L+1, \dots, \min(K, M)\}$. Also, let $P_1, P_2, \dots, P_M, P_{L+1}', P_{L+2}', \dots, P_M', P_{L+1}'', P_{L+2}'', \dots, P_{K'}'', D_1, D_2, \dots, D_M, D_1', D_2', \dots, D_L, D_{K'+1}', D_{K'+2}', \dots, D_K, D_1'', D_2'', \dots, D_L''$ be non-negative such that the following conditions are satisfied

$$\sum_{m=1}^M P_m \leq MP, \quad (12)$$

$$\begin{aligned} P'_m &\leq P_m, \quad m = L+1, L+2, \dots, M, \\ P''_m &\leq P'_m, \quad m = L+1, L+2, \dots, K', \\ D_k &\leq D''_k \leq D'_k \leq \sigma_k^2, \quad k = 1, 2, \dots, L, \\ D_k &\leq \sigma_k^2, \quad k = L+1, L+2, \dots, K', \\ D_k &\leq D'_k \leq \sigma_k^2, \quad k = K'+1, K'+2, \dots, K. \end{aligned} \quad (13)$$

The following (D_s, D_w) is achievable

$$\begin{aligned} D_s &= \frac{1}{K} \left(\sum_{k=1}^K D_k \right), \\ D_w &= \frac{1}{K} \left(\sum_{k=1}^L \frac{D'_k}{1 + \frac{P_k}{N_w}} + \sum_{k=L+1}^{K'} \frac{\sigma_k^2}{1 + \frac{P'_k - P''_k}{P''_k + N_w}} + \sum_{k=K'+1}^K D'_k \right), \end{aligned}$$

if the following conditions are satisfied

$$\begin{aligned} \frac{D'_k}{D''_k} &= 1 + \frac{P_k}{N_s}, \quad k = 1, 2, \dots, L, \\ \frac{\sigma_k^2}{D_k} &= 1 + \frac{P'_k - P''_k}{P''_k + N_s}, \quad k = L+1, L+2, \dots, K', \end{aligned}$$

$$\sum_{k=1}^L \log \frac{\sigma_k^2}{D'_k} + \sum_{k=K'+1}^K \log \frac{\sigma_k^2}{D'_k} \leq \sum_{m=L+1}^M \log \left(1 + \frac{P_m - P'_m}{P'_m + N_w} \right), \quad \text{and} \quad (14)$$

$$\sum_{k=1}^L \log \frac{D''_k}{D_k} + \sum_{k=K'+1}^K \log \frac{D'_k}{D_k} \leq \sum_{m=L+1}^{K'} \log \left(1 + \frac{P''_m}{N_s} \right) + \sum_{m=K'+1}^M \log \left(1 + \frac{P'_m}{N_s} \right). \quad (15)$$

The proof is relegated to appendix III.

In general, the above optimization problem appears to be computationally challenging for large values of K and M . However, simplification is possible for the important special case of memoryless sources and channels with bandwidth expansion and bandwidth contraction. This is taken up in the next section.

V. SPECIALIZATION TO MEMORYLESS SOURCES AND CHANNELS WITH BANDWIDTH MISMATCH

A special case of the problem is when the source is also memoryless, but has a bandwidth different from the channel. If we define the degree of mismatch by $\alpha = M/K$, we have

Theorem 5: For the special case of the problem in section II with $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_M^2 = \sigma_S^2$, the following (D_s, D_w) trade-off is achievable:

- For $\alpha < 1$ (bandwidth contraction)

$$\left\{ (D_s^{\text{BC}}(\lambda, \gamma), D_w^{\text{BC}}(\lambda, \gamma)) : 0 \leq \lambda \leq 1, 0 \leq \gamma \leq 1 \right\},$$

where

$$\begin{aligned} D_w^{\text{BC}}(\lambda, \gamma) &= \frac{\alpha \sigma_S^2}{\frac{\lambda P + N_w}{\lambda \gamma P + N_w}} + \frac{(1 - \alpha) \sigma_S^2}{\left(\frac{P + N_w}{\lambda P + N_w} \right)^{\frac{\alpha}{1 - \alpha}}}, \quad \text{and} \\ D_s^{\text{BC}}(\lambda, \gamma) &= \frac{\alpha \sigma_S^2}{\frac{\lambda P + N_s}{\lambda \gamma P + N_s}} + \frac{(1 - \alpha) \sigma_S^2}{\left(\frac{P + N_w}{\lambda P + N_w} \frac{\lambda \gamma P + N_s}{N_s} \right)^{\frac{\alpha}{1 - \alpha}}}. \end{aligned}$$

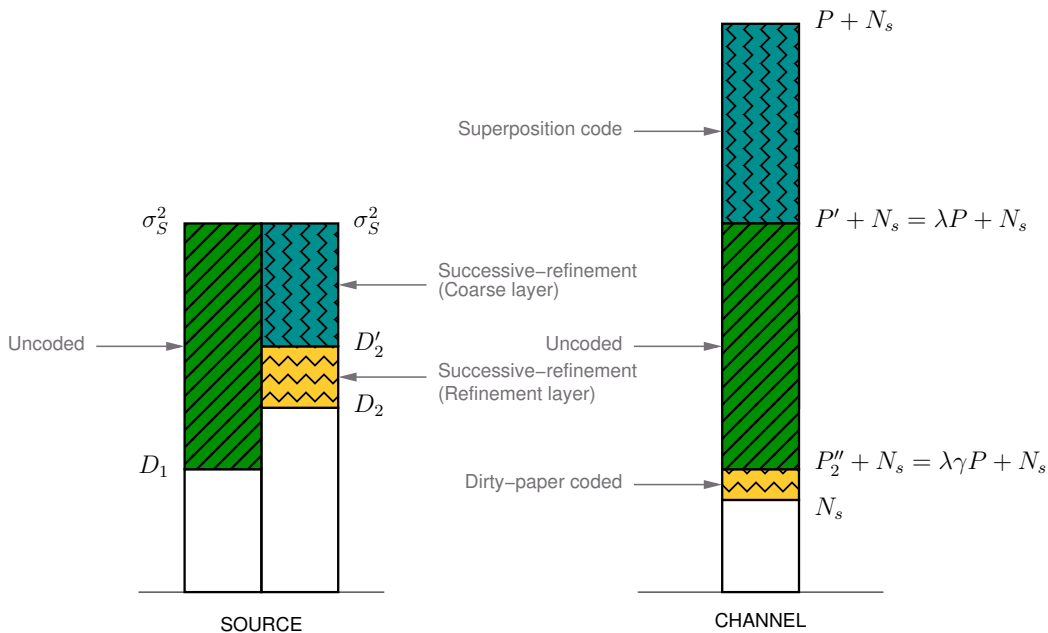


Fig. 7. Memoryless source over memoryless channel with bandwidth contraction: Schematic diagram showing an allocation for bandwidth expansion factor $\alpha = 1/2$. Note that we need not use Wyner-Ziv coding and may use successive refinement source coding as shown.

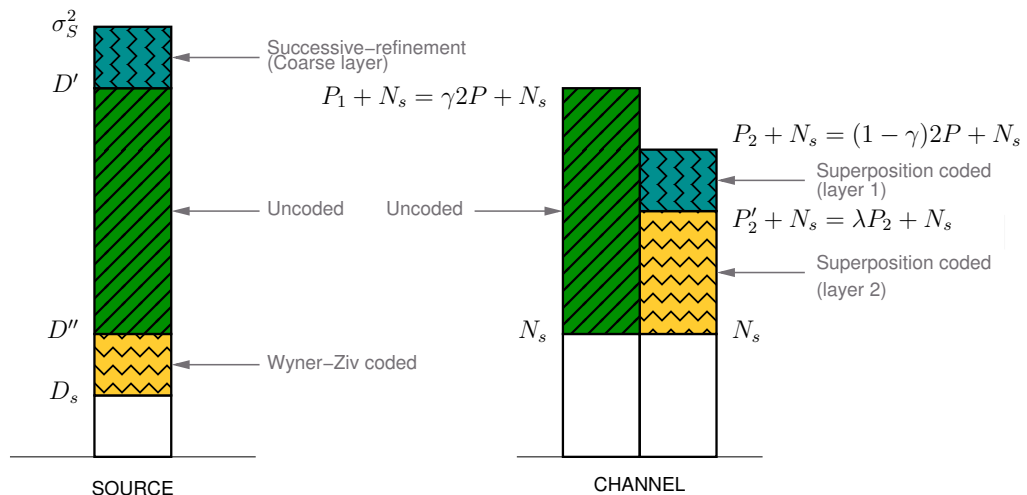


Fig. 8. Memoryless source over memoryless channel with bandwidth expansion: Schematic diagram showing an allocation for bandwidth expansion factor $\alpha = 2$. Note that we need not use Gel'fand-Pinsker's (dirty-paper) coding and may use superposition channel coding as shown.

- For $\alpha > 1$ (bandwidth expansion)

$$\left\{ (D_s^{\text{BE}}(\lambda, \gamma), D_w^{\text{BE}}(\lambda, \gamma)) : 0 \leq \lambda \leq 1, 0 \leq \gamma \leq 1 \right\},$$

where

$$D_w^{\text{BE}}(\lambda) = \frac{\sigma_S^2}{\left(\frac{\frac{\alpha(1-\gamma)}{\alpha-1} P + N_w}{\lambda \frac{\alpha(1-\gamma)}{\alpha-1} P + N_w} \right)^{\alpha-1} \left(\frac{\alpha\gamma P + N_w}{N_w} \right)}, \text{ and}$$

$$D_s^{\text{BE}}(\lambda) = \frac{\sigma_S^2}{\left(\frac{\frac{\alpha(1-\gamma)}{\alpha-1} P + N_w}{\lambda \frac{\alpha(1-\gamma)}{\alpha-1} P + N_w} \right)^{\alpha-1} \left(\frac{\alpha\gamma P + N_s}{N_s} \right) \left(\frac{\lambda \frac{\alpha(1-\gamma)}{\alpha-1} P + N_s}{N_s} \right)^{\alpha-1}}.$$

We prove this as a special case of Theorem 4 in appendix II.

For bandwidth contraction, we use only successive refinement source coding and Gel'fand-Pinsker channel coding; see Fig. 7. In the bandwidth expansion case, only Wyner-Ziv coding and superposition decoding is used (Fig. 8). As pointed out earlier, many researchers have investigated this special case. For the bandwidth expansion case, an almost identical scheme (using Wyner-Ziv coding and superposition coding) was presented in [5]. The trade-off expression above, but with $\gamma = 1/\alpha$ (corresponding to a flat power allocation) appears in [5, Theorem 2]. However, as shown in Fig. 9(b) and (c), the extra flexibility from non-flat power allocations can lead to slight gains. The bandwidth contraction case discussed above is new. The following remarks on the extreme points of these trade-offs are in order.

- At the weak-user-optimal points, our achievable schemes under bandwidth contraction and bandwidth expansion reduce to the schemes proposed by Mittal and Phamdo [4].
- At the strong-user-optimal point under bandwidth expansion, our scheme reduces to the systematic lossy source-channel codes of Shamai, Verdù, and Zamir [3].
- At the strong-user-optimal points, the achievable scheme is strictly better than the solution offered by Mittal and Phamdo in [4]. The gap can be computed explicitly to be

$$\begin{aligned} \text{(bandwidth contraction)} \quad & \frac{\alpha\sigma_S^2 N_s}{N_w + P} \left(1 - \frac{1}{\left(1 + \frac{P}{N_s}\right)^\alpha} \right), \quad \text{and} \\ \text{(bandwidth expansion)} \quad & \frac{\sigma_S^2}{\left(1 + \frac{P}{N_s}\right)^\alpha \left(1 + \frac{N_w}{P}\right)}. \end{aligned}$$

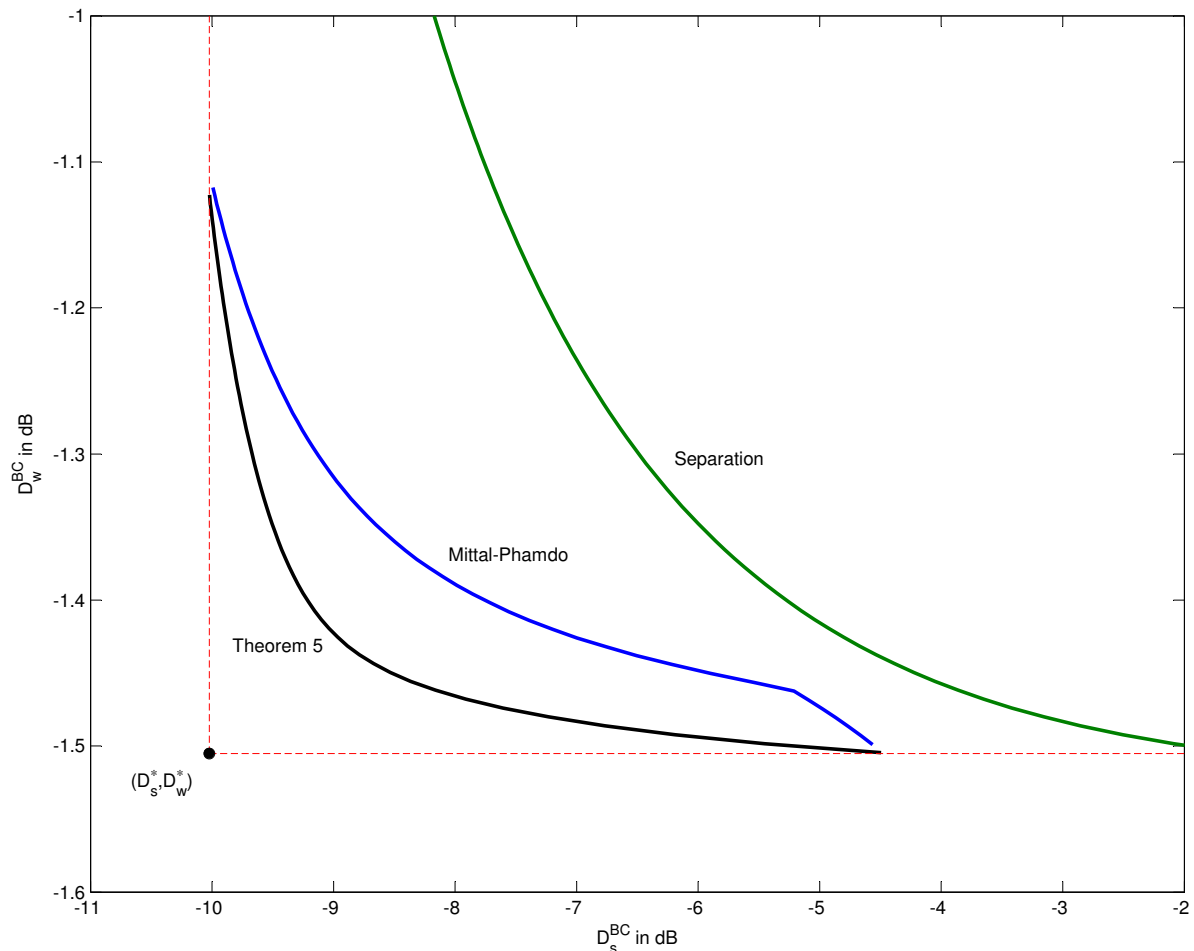
As pointed out above, at the weak-user-optimal point, the schemes coincide. For the boundary points in between the strong-user-optimal and weak-user-optimal points, an explicit computation is cumbersome, but numerical computation over a wide range of settings suggest that the achievable scheme strictly outperforms the schemes of Mittal and Phamdo. Fig. 9 shows a comparison of the trade-offs achieved by our scheme with those of Mittal and Phamdo [4] for a few examples.

Outer bounds: Only one non-trivial outer bound is available in the literature for this problem. It is due to Reznic, Feder, and Zamir [5] who developed it for the case of memoryless source and channel under bandwidth expansion. This bound, however, does not match the best available inner bound described above. The same bounding technique can be used to derive outer bounds for the parallel source-channel problem considered here. However, it does not always lead to a non-trivial bound. For instance, for memoryless source and channel with bandwidth contraction, the technique yields the trivial bound (resulting from considering the point-to-point source-channel problems involving either the weak user or the strong users alone).

VI. CONCLUSION

We have presented a hybrid digital-analog scheme for the problem of sending a parallel Gaussian source over a white Gaussian broadcast channel which potentially has a bandwidth mismatch with the source. We used the concepts of successive refinement and Wyner-Ziv source coding, and superposition and dirty paper channel coding to show that without compromising the point-to-point optimal performance of either the weak or strong user, we can strictly improve the performance of the other user over what the conventional separation approach offers. We also showed how to achieve better trade-offs when neither user is point-to-point optimal. While we do not have a converse for our scheme, the achievable points are the best available. When specialized to the case of memoryless sources and channels with bandwidth mismatch, our scheme matches or in some cases outperforms the best known schemes.

Only an achievable trade-off is available for two digital receivers with different channels. We believe that the limitation is primarily due to the lack of good outer bounds on the region of distortions that can be supported simultaneously. Even in simple cases like when the source and channel are memoryless, but



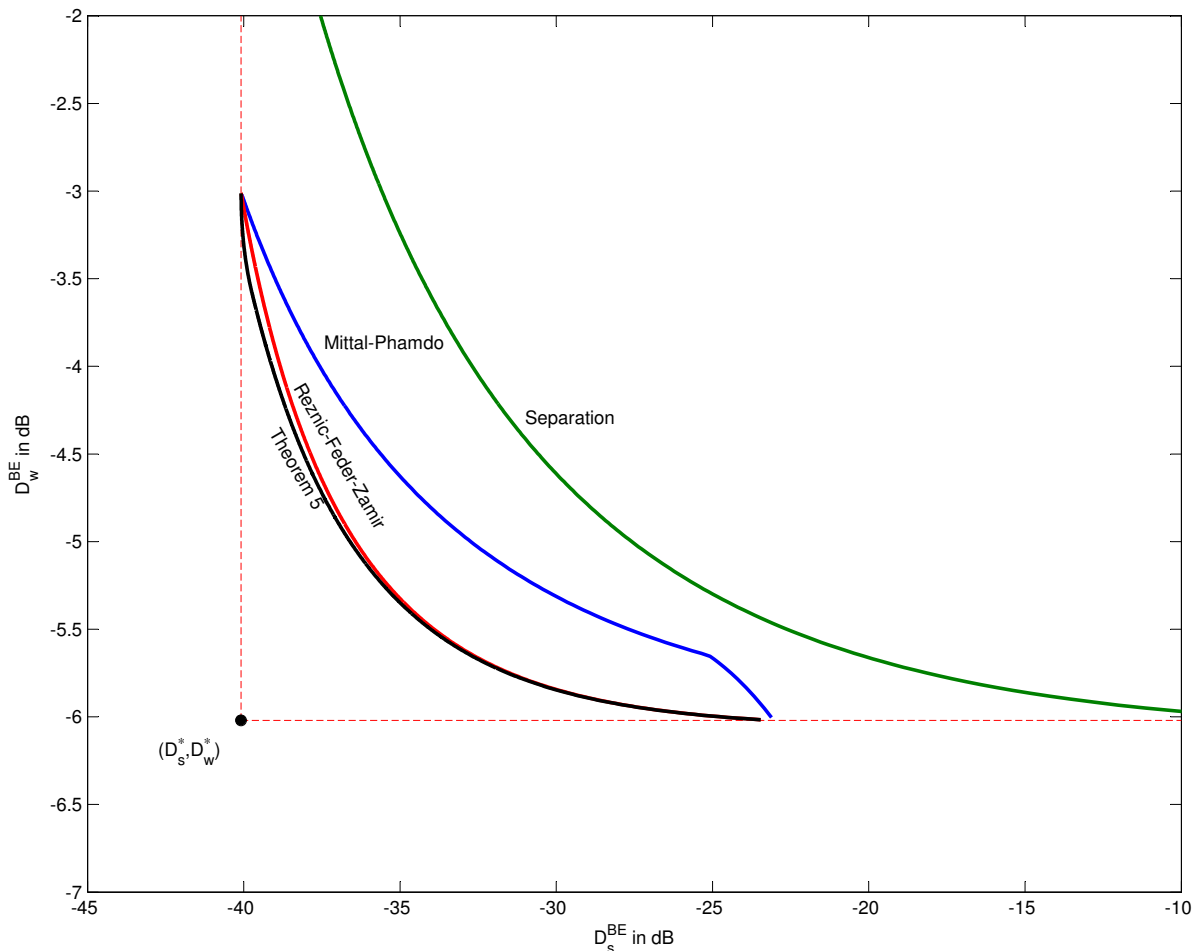
(a)

Fig. 9. Comparison of distortion trade-offs achieved by our scheme with that of Mittal and Phamdo: (a) Bandwidth contraction. $\sigma_S^2 = 1$, $P/N_s = 20\text{dB}$, $P/N_w = 0\text{dB}$, $\alpha = 0.5$. The best scheme suggested by Mittal and Phamdo was chosen for comparison [4, Fig. 14]. The dashed lines are drawn at the weak and strong user optimal distortions and thus give the trivial outer bound to the trade-off region. The strong-user-optimal points appear to coincide, but there is a small gap which is not visible at the scale of this plot. The weak user optimal points indeed coincide where both schemes reduce to the same scheme. The performance of the best separation-based scheme is also shown for comparison.

with mismatching bandwidths, a tight result is not available. In fact, the best available outer bound for the case where the source bandwidth is larger than the channel bandwidth is the trivial outer bound which considers the receivers separately. It is also not clear if codes with more structure can be used to obtain better trade-offs. These could be subjects of further investigation.

ACKNOWLEDGEMENT

The authors would like to thank the reviewers for their careful reading and their suggestions which have helped improve the quality of this paper.

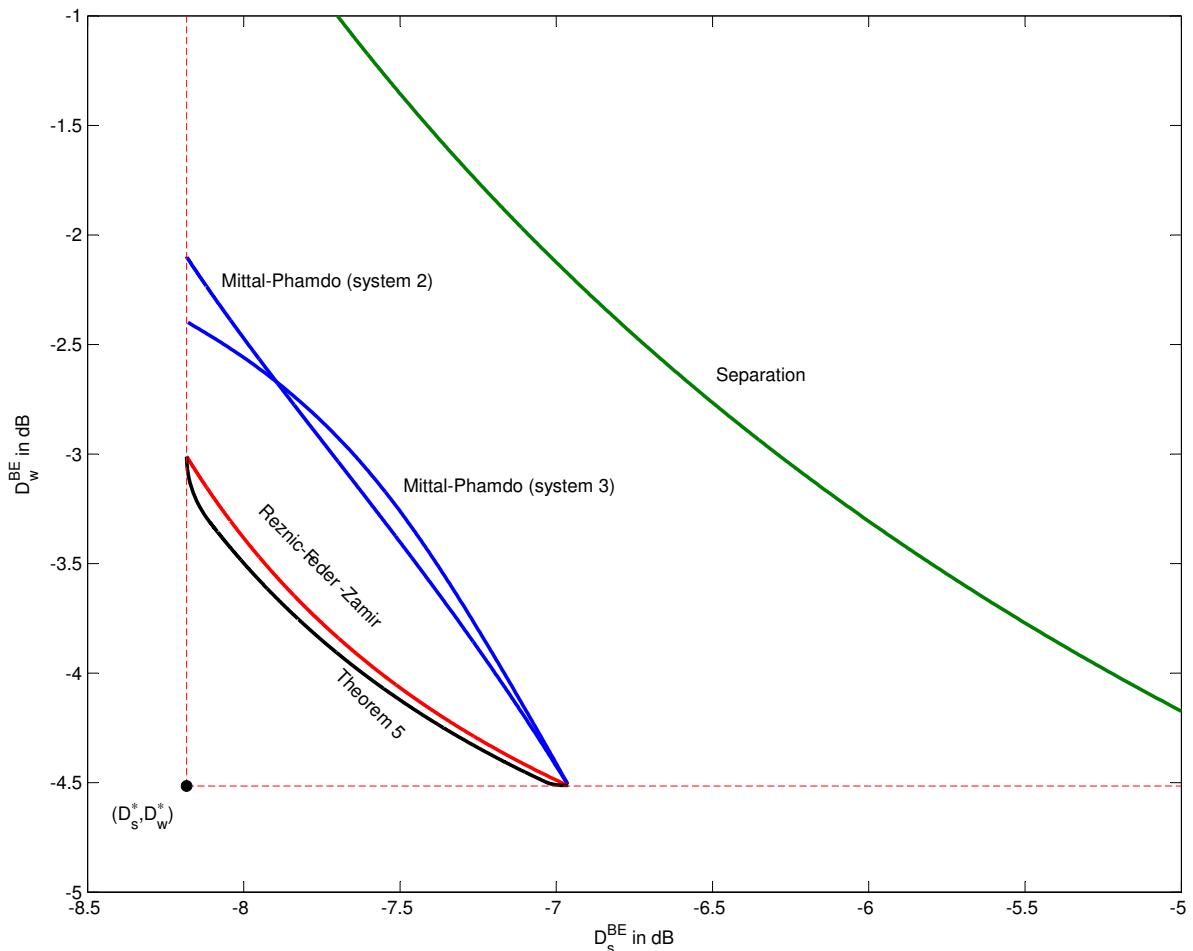


(b)

Fig. 9. Comparison of distortion trade-offs achieved by our scheme with that of Mittal and Phamdo: (b) Bandwidth expansion. $\sigma_S^2 = 1$, $P/N_s = 20\text{dB}$, $P/N_w = 0\text{dB}$, $\alpha = 2.0$. Again the best scheme of Mittal and Phamdo for this setting was chosen for comparison [4, Fig. 12]. The dashed lines are drawn at the weak and strong-user-optimal distortions and thus give the trivial outer bound to the trade-off region. The strong-user-optimal points appear to coincide, but there is a small gap which is not visible at the scale of this plot. The weak-user-optimal points indeed coincide where the schemes reduce to the same scheme. The slight improvement over [5] from allowing non-flat power allocation is visible. The performance of the best separation-based scheme is shown for comparison.

APPENDIX I PROOF OF PROPOSITION 1

We will use the notation introduced in section II with $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_K^2 = \sigma^2$. The proposition is a consequence of the following two facts: (i) the Gaussian source $\{S_k, k = 1, \dots, K\}$ is successively refinable [7], and (ii) there is an operating point in the degraded message set rate region of the white Gaussian broadcast channel in which each user's rate is within 1-bit of its point-to-point capacity. To see the second fact, let us recall that the boundary of the degraded message set rate region (for each sub-channel) of the Gaussian broadcast channel is given by the set of rates $\{(R_{\text{base}}(\beta), R_{\text{refine}}(\beta)), \beta \in [0, 1]\}$,



(c)

Fig. 9. Comparison of distortion trade-offs achieved by our scheme with those of Mittal and Phamdo: (b) Bandwidth expansion. $\sigma_S^2 = 1$, $P/N_s = 4\text{dB}$, $P/N_w = 0\text{dB}$, $\alpha = 1.5$. Two schemes of Mittal and Phamdo (systems 2 and 3) together give the best performance of all the new schemes proposed in [4, Fig. 13]. The dashed lines are drawn at the weak and strong user optimal distortions and thus give the trivial outer bound to the trade-off region. The gap between the strong-user-optimal points is visible in this plot. The weak-user-optimal points coincide. The slight improvement over [5] from allowing non-flat power allocation is again visible. The performance of the best separation-based scheme is shown for comparison.

where the rate of the message decoded by both users is (in bits per symbol)

$$R_{\text{base}}(\beta) = \frac{1}{2} \log_2 \left(1 + \frac{(1-\beta)P}{\beta P + N_w} \right),$$

and the rate of the message decoded only by the strong user is

$$R_{\text{refine}}(\beta) = \frac{1}{2} \log_2 \left(1 + \frac{\beta P}{N_s} \right).$$

And, for a given $\beta \in [0, 1]$, the overall rate delivered to the users is

$$\begin{aligned} R_w(\beta) &= R_{\text{base}}(\beta), \\ R_s(\beta) &= R_{\text{base}}(\beta) + R_{\text{refine}}(\beta). \end{aligned}$$

In order to show fact (ii), it is enough to show that there is a $\bar{\beta} \in [0, 1]$ such that

$$\begin{aligned} R_w(\bar{\beta}) + 1 &\geq R_w^* \stackrel{\text{def}}{=} R_w(0), \text{ and} \\ R_s(\bar{\beta}) + 1 &\geq R_s^* \stackrel{\text{def}}{=} R_s(1), \end{aligned}$$

where R_w^* and R_s^* are the point-to-point capacities of the weak and strong users respectively. Simplifying the above two conditions, we get

$$\frac{N_w}{P} \geq \bar{\beta} \geq \frac{1}{1/\left(\frac{N_w - N_s}{P}\right) + 2/\left(1 + \frac{N_s}{P}\right)} - \frac{N_s}{P}.$$

Observing that the left-hand side is always larger than the right-hand side and that the right-hand side is always less than 1, we can conclude that such a $\bar{\beta} \in [0, 1]$ always exists. By choosing this operating point $(R_{\text{base}}(\bar{\beta}), R_{\text{refine}}(\bar{\beta}))$ for the degraded message set broadcast channel code, and using the optimal layered (successive refinement) source code, the distortion pair (D_s, D_w) achieved by the user- j , $j \in \{s, w\}$ satisfies

$$\begin{aligned} \frac{K}{2} \log_2 \frac{\sigma^2}{D_j} &= MR_j(\bar{\beta}) \\ &\geq M(R_j^* - 1) \\ &= \frac{K}{2} \log_2 \frac{\sigma^2}{D_j^*} - M. \end{aligned}$$

Thus, since M/K is the ratio of the channel bandwidth to the source bandwidth, we have proved that

$$\frac{1}{2} \log_2 \frac{D_j}{D_j^*} \leq \frac{\text{BW}_{\text{channel}}}{\text{BW}_{\text{source}}}, \quad j \in \{s, w\}.$$

APPENDIX II PROOF OF THEOREM 5

The proof for the bandwidth contraction case ($\alpha = M/K < 1$) follows from the following choice of parameters in Theorem 4: $L = 0$, $K' = M$, and the power allocation is

$$\begin{aligned} P_m &= P, \quad m = 1, 2, \dots, M, \\ P'_m &= \lambda P, \quad m = 1, 2, \dots, M, \\ P''_m &= \lambda \gamma P, \quad m = 1, 2, \dots, M, \end{aligned}$$

where λ and γ are in $[0, 1]$. Also, we let

$$\begin{aligned} D_k &= D, \quad k = 1, 2, \dots, M, \\ D_k &= \tilde{D}, \quad k = M + 1, M + 2, \dots, K, \\ D'_k &= \tilde{D}', \quad k = M + 1, M + 2, \dots, K, \end{aligned}$$

where D , \tilde{D} , and \tilde{D}' are defined as below to satisfy the conditions of Theorem 4.

$$\begin{aligned}\frac{\sigma_S^2}{D} &= 1 + \frac{\lambda(1-\gamma)P}{\lambda\gamma P + N_s}, \\ (K-M) \log \frac{\sigma_S^2}{\tilde{D}'} &= M \log \frac{P + N_w}{\lambda P + N_w}, \\ (K-M) \log \frac{\tilde{D}'}{\tilde{D}} &= M \log \frac{\lambda\gamma P + N_s}{N_s}.\end{aligned}$$

Substituting these in the expression for the achievable (D_s, D_w) gives the result.

The choice of parameters for the bandwidth expansion case ($\alpha = M/K < 1$) is $L = K' = K$, and the power allocations are

$$\begin{aligned}P_m &= \frac{1}{K}(\gamma MP) = \alpha\gamma P, \quad m = 1, 2, \dots, K, \\ P_m &= \frac{1}{M-K}((1-\gamma)MP) = \frac{\alpha(1-\gamma)}{\alpha-1}P, \quad m = K+1, K+2, \dots, M,\end{aligned}$$

where $\gamma \in [0, 1]$. Clearly, we have $\sum_{m=1}^M P_m = MP$ as required. Also, we set

$$P'_m = (1-\lambda)P_m, \quad m = K+1, K+2, \dots, M,$$

where $\lambda \in [0, 1]$, and $D'_k = D'$, $D''_k = D''$, $D_k = D$, $k = 1, 2, \dots, K$, where D' , D'' , and D are chosen as follows to satisfy the conditions in Theorem 4.

$$\begin{aligned}K \log \frac{\sigma_S^2}{D'} &= (M-K) \log \frac{\frac{\alpha(1-\gamma)}{\alpha-1}P + N_w}{\lambda \frac{\alpha(1-\gamma)}{\alpha-1}P + N_w}, \\ \frac{D'}{D''} &= 1 + \frac{\alpha\gamma P}{N_s}, \\ K \log \frac{D''}{D} &= (M-K) \log \frac{\lambda \frac{\alpha(1-\gamma)}{\alpha-1}P + N_s}{N_s}.\end{aligned}$$

These choices give the achievability result for bandwidth expansion.

APPENDIX III PROOF OF THEOREM 4

The main ideas involved have already been described in section IV. We sketch the main steps of the proof which use the following results: successive refinement source coding [7], source coding with side-information or Wyner-Ziv (WZ) coding [12], super-position broadcast channel coding [2], and channel coding with side-information or Gel'fand-Pinsker (GP) coding [13] (in particular, as applied by Costa to the Gaussian case [14]).

The m -th sub-channel is allocated a power of P_m such that it satisfies the power constraint by (12). The coding will be performed as usual on block-length n sequences of sufficient length that the source codes invoked below have distortions close to optimal and the channel codes have low probabilities of errors. For clarity, we will suppress these small gaps as we did in the discussion of the $K = M = 2$ example in section IV.

a) *Source components 1 through L*: These source components are treated in a way similar to the first source component was in the $K = M =$ example (Fig. 6) of section IV in that they are sent in three different ways. An n -length block of the k -th such source component is source-coded (quantized) using an optimal source-code at distortion D'_k . This codeword will be made available to both the strong and the weak receivers. The rate required to do this is

$$\sum_{k=1}^L \frac{1}{2} \log \frac{\sigma_k^2}{D'_k}.$$

Let the decoder reconstruction of the i -th sample be $S'_k(i)$.

The quantization error of the k -th source component is transmitted over the k -th sub-channel using power P_k . In other words, the input to the k -th sub-channel is

$$X_k(i) = \sqrt{\frac{P_k}{D'_k}} (S_k(i) - S'_k(i)), \quad i = 1, 2, \dots, n.$$

To produce its reconstruction, the weak-user adds to $S'_k(i)$ the linear least-squared error estimate of $S_k(i) - S'_k(i)$ from $Y_{w_k}(i)$

$$\widehat{S}_{w_k}(i) = \frac{P_k}{P_k + N_w} \left(\sqrt{\frac{D'_k}{P_k}} Y_k(i) \right) + S'_k(i).$$

This estimate is at an MSE estimation error of $D'_k / \left(1 + \frac{P_k}{N_w}\right)$. This gives the first term in the expression for D_w in the theorem. The strong user also performs the same to get an intermediate reconstruction of the source component at MSE distortion of $D''_k = D'_k / \left(1 + \frac{P_k}{N_s}\right)$. This will act as a side-information available at the decoder for a Wyner-Ziv source coding of S_k (again footnote 2 applies). We would like to enhance this to a distortion of D_k in the expression for D_s in the theorem. Using an extension of Wyner and Ziv's result, the Wyner-Ziv bitrate needed to be delivered to the strong user is

$$\sum_{k=1}^L \frac{1}{2} \log \frac{D''_k}{D_k}.$$

b) *Source components L + 1 through K'*: These source components are similar to the second source component of the $K = M =$ example (Fig. 6) of section IV in that they are sent uncoded, but the sub-channels over which they are sent may have other bits sent using codewords. The k -th source component is sent uncoded on the k -th sub-channel using power $P'_k - P''_k$. The rest of the power spent on this sub-channel is utilized to send the coded parts as will be discussed below. However, we need to note the fact that both decoders will be able to decode a coded part sent using power $P_k - P'_k$ on this sub-channel and subtract it off before estimating S_k . The rest of the coded part which is sent at power P''_k will act as interference. Hence the MSE distortion for the k -th source component will be

$$D_{jk} = \frac{\sigma_k^2}{1 + \frac{P'_k - P''_k}{P''_k + N_j}}, \quad j \in \{s, w\}.$$

This gives the corresponding terms in the expressions for D_s and D_w in the theorem.

c) *Source components $K'+1$ through K* : These source components will have no uncoded component unlike the above two cases. Thus they are source coded using an optimal successive refinement code [7] which works at a coarse description distortion D'_k , and fine description distortion of D_k . These give the corresponding terms in the expressions for D_w and D_s respectively in the theorem. The bitrate for the coarse description which will be made available to both the users is

$$\sum_{k=K'+1}^K \frac{1}{2} \log \frac{\sigma_k^2}{D'_k},$$

and the refinement layer which will be made available only to the strong user has a bitrate of

$$\sum_{k=K'+1}^K \frac{1}{2} \log \frac{D'_k}{D_k}.$$

Thus over all we need to send bits at a rate of

$$\sum_{k=1}^L \frac{1}{2} \log \frac{\sigma_k^2}{D'_k} + \sum_{k=K'+1}^K \frac{1}{2} \log \frac{\sigma_k^2}{D'_k} \quad (16)$$

to both the users, and in addition bits at the rate of

$$\sum_{k=1}^L \frac{1}{2} \log \frac{D'_k}{D_k} + \sum_{k=K'+1}^K \frac{1}{2} \log \frac{D'_k}{D_k} \quad (17)$$

to the strong user. In the next two steps we show how this is accomplished.

d) *Sub-channels $L+1$ through K'* : As discussed above, on the m -th sub-channel a power of $P'_m - P''_m$ is used for uncoded transmission. The rest of the power is allocated as follows: $P'_m - P''_m$ is used for sending bits to both the receivers. The rest of the power P''_m is used to send bits which will be decoded only the strong receiver. When decoding the common bits, both receivers treat the rest of the power as interference. Hence the bitrate of the common part is limited by the weaker user resulting in a rate of

$$\sum_{m=L+1}^{K'} \frac{1}{2} \log \left(1 + \frac{P'_m - P''_m}{P'_m + N_w} \right).$$

As mentioned earlier, upon decoding, both users will subtract the codeword corresponding to the decoded common bits from their received signal. To send additional bits to the strong user, we use the concept of dirty-paper coding [14][13]. The uncoded transmission can be thought of as Gaussian side-information (or “dirt”) which is known at the encoder. From Costa [14, Sec. II], we know that a power allocation of P''_m can support a rate of

$$\sum_{m=L+1}^{K'} \frac{1}{2} \log \left(1 + \frac{P''_m}{N_s} \right)$$

to the strong user.

e) *Sub-channels $K'+1$ through M* : In these sub-channels no uncoded transmission is performed. We use the idea of superposition coding [2] to deliver a common bitstream to both the users and in addition a refinement bitstream to only the stronger user. With a power allocation of $P'_m - P''_m$ to the common bitstream and the rest P''_m to the refinement bitstream, we get the following bitrates for the common and refinement bit streams respectively

$$\sum_{m=K'+1}^M \frac{1}{2} \log \left(1 + \frac{P'_m - P''_m}{P'_m + N_w} \right)$$

$$\sum_{m=K'+1}^M \frac{1}{2} \log \left(1 + \frac{P''_m}{N_s} \right).$$

Thus the total rate available for sending a common bitstream to both the users is

$$\sum_{m=L+1}^M \frac{1}{2} \log \left(1 + \frac{P_m - P'_m}{P'_m + N_w} \right).$$

Condition (15) ensures that this is sufficient to handle the rate of the common bit stream in (16). Similarly, the total rate available for sending an enhancement bitstream to the stronger user is

$$\sum_{m=L+1}^{K'} \frac{1}{2} \log \left(1 + \frac{P''_m}{N_s} \right) + \sum_{m=K'+1}^M \frac{1}{2} \log \left(1 + \frac{P'_m}{N_s} \right)$$

which is larger than the rate of the enhancement bitstream in (17) by condition (15). This completes the proof.

REFERENCES

- [1] T. J. Goblick, Jr. Theoretical Limitations on the Transmission of Data from Analog Sources. *IEEE Transactions on Information Theory*, 11(4):558–567, October 1965.
- [2] T. M. Cover. Broadcast Channels. *IEEE Transactions on Information Theory*, 18(1):2–14, January 1972.
- [3] S. Shamai, S. Verdú, and R. Zamir. Systematic Lossy Source/Channel Coding. *IEEE Transactions on Information Theory*, 44(2):564–579, March 1998.
- [4] U. Mittal and N. Phamdo. Hybrid Digital-Analog (HDA) Joint Source-Channel Codes for Broadcasting and Robust Communications. *IEEE Transactions on Information Theory*, 48(5):1082–1102, May 2002.
- [5] Z. Reznic, M. Feder, and R. Zamir. Distortion Bounds for Broadcasting with Bandwidth Expansion. *IEEE Transactions on Information Theory*, 52(8):3778–3788, August 2006.
- [6] V. Prabhakaran, R. Puri, and K. Ramchandran, “Hybrid Analog-Digital Strategies for Source-Channel Broadcast,” in *Proc. 43rd Allerton Conference on Communication, Control and Computing*, Allerton, IL, September 2005.
- [7] W. H. R. Equitz and T. M. Cover. Successive Refinement of Information. *IEEE Transactions on Information Theory*, 37(2):269–275, March 1991.
- [8] J. Körner and K. Marton. General Broadcast Channels with Degraded Message Sets. *IEEE Transactions on Information Theory*, 23(1):60–64, January 1977.
- [9] C. Tian, S. Diggavi, and S. Shamai, “Approximate Characterizations for the Gaussian Source Broadcast Distortion Region,” in *IEEE International Symposium on Information Theory*, Seoul, Korea, June-July 2009.
- [10] T. Berger and D. W. Tufts, “Optimum Pulse Amplitude Modulation Part I: Transmitter-Receiver Design and Bounds from Information Theory.” *IEEE Transactions on Information Theory*, 13(2):196–208, April 1967.
- [11] V. Prabhakaran, R. Puri, and K. Ramchandran. Colored Gaussian Source-Channel Broadcast for Heterogeneous (Analog/Digital) Receivers. *IEEE Transactions on Information Theory*, 54(4):1807–1814, April 2008.
- [12] A. D. Wyner and J. Ziv. The Rate-Distortion Function for Source Coding with Side Information at the Decoder. *IEEE Transactions on Information Theory*, 22(1):1–10, January 1976.
- [13] S. I. Gel'fand and M. S. Pinsker. Coding for Channel with Random Parameters. *Problems of Control and Information Theory*, 9(1):19–31, 1980.
- [14] M. H. M Costa. Writing on Dirty Paper. *IEEE Transactions on Information Theory*, 29(3):439–441, May 1983.
- [15] S. Bross, A. Lapidoth, and S. Tinguely, “Superimposed Coded and Uncoded Transmissions of a Gaussian Source over the Gaussian Channel,” in *Proc. IEEE Int. Symp. on Inform. Theory*, 2153–2155, Seattle, WA, July 2006.
- [16] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.