Wyner-Ziv Coding over Broadcast Channels: Digital Schemes

Jayanth Nayak, Ertem Tuncel, Deniz Gündüz

Abstract

This paper addresses lossy transmission of a common source over a broadcast channel when there is correlated side information at the receivers. The quadratic Gaussian and binary Hamming cases are especially targeted. Using ideas from the lossless version of the problem, i.e., Slepian-Wolf coding over broadcast channels, and dirty paper coding, several digital schemes are proposed and their single-letter distortion tradeoffs are characterized. These schemes use layered coding where the common layer information is intended for both receivers and the refinement information is destined only for the receiver that is chosen using an appropriately defined "combined" channel/side information quality measure. When this quality is constant at each receiver, all the new schemes converge and become optimal. When the source and the channel bandwidths are equal, it is shown that one of the proposed schemes outperforms all the others as well as separate coding. For the quadratic Gaussian problem, it is also shown that if the combined quality criterion chooses the worse channel as the refinement receiver, then the same scheme also outperforms uncoded transmission. Unlike its lossless counterpart, however, the problem eludes a complete characterization.

I. Introduction

Consider a sensor network where $K+1$ nodes take periodic measurements of a common phenomenon. We study the communication scenario where one of the sensors is required to transmit its measurements to the other K nodes over a broadcast channel. The receiver nodes are themselves equipped with side information unavailable to the sender, e.g., measurements correlated with the sender's data. This scenario, which is depicted in Figure 1, can be of interest either by itself or as a part of a larger scheme where all nodes are required to broadcast their measurements to all the other nodes. Finding the capacity of a broadcast channel is a longstanding open problem, and thus, limitations of using separate source and channel codes in this scenario may never be fully understood. In contrast, a very simple joint source-channel coding strategy is optimal for the special case of *lossless* coding [18]. More specifically, it was shown in [18] that in Slepian-Wolf coding over broadcast channels (SWBC), as the lossless case was referred to, for a given source X, side information $Y_1, \ldots, Y_K$, and a broadcast channel $p_{V_1 \ldots V_K|U}$, lossless transmission (in the Shannon sense) is possible with κ channel uses per source symbol if and only if there exists a channel input distribution U such that

$$H(X|Y_k) \leq \kappa I(U; V_k) \tag{1}$$

for $k = 1, \ldots, K$. This result exhibits some striking features which are worth summarizing here.

J. Nayak and E. Tuncel are with the University of California, Riverside, CA, E-mail: {jnayak,ertem}@ee.ucr.edu. D. Gündüz is with Princeton University and Stanford University, E-mail: dgunduz@princeton.edu

This work was presented in part at the Information Theory Workshop (ITW) 2008, Porto, Portugal.

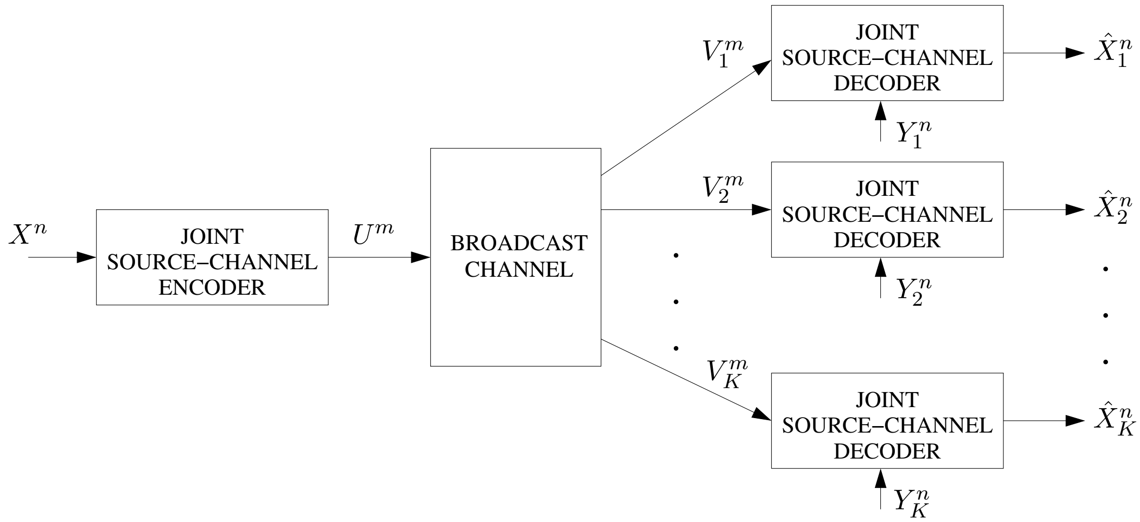

Fig. 1. Block diagram for Wyner-Ziv coding over broadcast channels.

(i) The optimal coding scheme is not separable in the classical sense, but consists of *separate components* that perform source and channel coding in a broader sense. This results in separation of source and channel variables as in (1).

(ii) If the broadcast channel is such that the same input distribution achieves capacity for all individual channels, then (1) implies that one can utilize all channels in *full capacity*. Binary symmetric channels and Gaussian channels are the widely known examples of this phenomenon.

(iii) The optimal coding scheme does not necessarily involve *binning*, which is commonly used in network information theory. Instead, with the simple coding strategy of [18], the channel performs the binning automatically in a virtual manner[1].

In this paper, we consider the general lossy coding problem in which the reconstruction of the source at the receivers need not be perfect. We shall refer to this problem setup as Wyner-Ziv coding over broadcast channels (WZBC). We present coding schemes and analyze their performances in the quadratic Gaussian and binary Hamming cases. These schemes use ideas from SWBC [18] and dirty paper coding (DPC) [3], [6] as a starting point. The SWBC scheme is modified to a) allow quantization of the source, and b) additionally handle channel state information (CSI) at the encoder by using DPC. These modifications are then employed in layered transmission schemes with $K = 2$, where there is common layer (CL) information destined for both receivers and refinement layer (RL) information meant for only one of the receivers. Varying the encoding and the decoding orders of these two layers, we initially obtain four different schemes. In each scheme, the channel codewords corresponding to the two layers are superposed, and to mitigate the resultant interference, we either use successive decoding or DPC. We then show that one of these four schemes prevail as the best for the quadratic Gaussian problem. The same is

[1] An alternative binning-based scheme using block Markov encoding and backward decoding can be found in [7].

observed experimentally for the binary Hamming problems, although an analytical proof is difficult to devise.

DPC is used in this work in a manner quite different from the way it was used in [2], which concentrated on sending private information to each receiver in a broadcast channel setting, where the information that forms the CSI and the information that is dirty paper coded are meant for different receivers. Therefore, although the DPC auxiliary codewords are decoded at one of the receivers, unlike in our schemes, this is of no use to that receiver. For our problem, this difference leads to an additional interplay in the choice of channel random variables. To the best of our knowledge, the DPC techniques in this work are most similar to those in [15], [19], where, as in our schemes, the CSI carries information about the source and hence decoding the DPC auxiliary codeword helps improve the performance. However, our results indicate a unique feature of DPC in the framework of WZBC. In particular, in our best layered scheme, the optimal Costa parameter for the quadratic Gaussian problem turns out to be either 0 or 1. When it is 0, there is effectively no DPC, and when it is 1, the auxiliary codeword is identical to the channel input corrupted by the CSI. To the best of our knowledge, although the latter choice is optimal for binary symmetric channels, it has never been optimal for a Gaussian channel in a scenario considered before.

When an appropriately defined "combined" channel and side information quality is constant at each receiver, the new schemes are shown to be optimal in the quadratic Gaussian case. We also derive conditions for the same phenomenon to occur in the binary Hamming case, albeit not as elegantly as in the quadratic Gaussian problem. Unlike in [18], however, the schemes that we derive are not always optimal. A simple alternative approach is to separate the source and channel coding. Both Gaussian and binary symmetric broadcast channels are degraded. Hence their capacity regions are known [4] and further, there is no loss of optimality in confining ourselves to two layer source coding schemes. The corresponding source and side information pairs are also degraded. Although a full characterization of the rate-distortion performance is available for the quadratic Gaussian case [16], only a partial characterization is available for the binary Hamming problem [14], [16]. In any case, we obtain the distortion tradeoff of separate source and channel coding by combining the known rate-distortion results with the capacity results. For the quadratic Gaussian problem, we show that one of our schemes always outperforms the others as well as separate coding. The same phenomenon is numerically observed for the binary Hamming case.

For the two examples we consider, a second alternative is uncoded transmission if there is no bandwidth expansion or compression. This scheme is optimal in the absence of side information at the receivers in both the quadratic Gaussian and binary Hamming cases. However, in the presence of side information, the optimality may break down. We show that, depending on the quality of the side information, our schemes can indeed outperform uncoded transmission as well. In particular, if the combined quality criterion chooses the worse channel as the refinement receiver (because it has much better side information), then the best layered scheme also outperforms uncoded transmission for the quadratic Gaussian problem.

The paper is organized as follows. In Section II, we formally define the problem and present relevant past work. Our main results are presented in Section III and Section IV: we develop extensions of the scheme in [18]. We then apply them to the quadratic Gaussian and binary Hamming cases in Sections V and VI, respectively. For these cases, we compare the derived schemes among themselves, with separate source and channel coding, and finally with uncoded transmission. Section VII concludes the paper by summarizing the results and pointing to the future work.

II. Background and Notation

Let $(X, Y_1, \ldots, Y_K) \in \mathcal{X} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_K$ be random variables denoting a source with independent and identically distributed (i.i.d.) realizations. The X sequence is to be transmitted over a memoryless broadcast channel defined by $p_{V_1 \cdots V_K|U}(v_1, \ldots, v_K|u), u \in \mathcal{U}, v_k \in \mathcal{V}_k, k = 1, \ldots K$. Decoder k has access to side information Y_k in addition to the channel output V_k. Let single-letter distortion measures $d_k : \mathcal{X} \times \hat{\mathcal{X}}_k \to [0, \infty)$ be defined at each receiver, i.e.,

$$d_k(x^n, \hat{x}_k^n) = \frac{1}{n} \sum_{j=1}^{n} d_k(x_j, \hat{x}_{kj})$$

for $k = 1, \ldots, K$. We denote the variance of a random variable A by $\mathbf{A}$. Also, all logarithms are base 2.

Definition 1: An $(m, n, f, g_1, \ldots, g_K)$ code consists of an encoder

$$f : \mathcal{X}^n \to \mathcal{U}^m$$

and decoders at each receiver

$$g_k : \mathcal{V}_k^m \times \mathcal{Y}_k^n \to \hat{\mathcal{X}}_k^n \ .$$

The rate of the code is $\kappa = \frac{m}{n}$ channel uses per source symbol.

Definition 2: A distortion tuple $(D_1, \ldots, D_K)$ is said to be achievable at a rational rate κ if for every $\epsilon > 0$, there exists n_0 such that for all integers $m > 0, n > n_0$ with $\frac{m}{n} = \kappa$, there exists an $(m, n, f, g_1, \ldots, g_K)$ code satisfying

$$\frac{1}{n} \mathrm{E} \left[d_k(X^n, \hat{X}_k^n) \right] \le D_k + \epsilon$$

where $\hat{X}_k^n = g_k(V_k^m, Y_k^n)$ and V_k^m denotes the channel output corresponding to $f(U_k^m)$.

In this paper, we present some general WZBC techniques and derive the corresponding achievable distortion regions. We study the performance of these techniques for the following cases.

- *Quadratic Gaussian*: All source and channel variables are real-valued. The source and side information are jointly Gaussian and the channels are additive white Gaussian, i.e., $V_k = U + W_k$ where W_k is Gaussian and $W_k \perp U$. There is an input power constraint on the channel:

$$\frac{1}{m} \mathrm{E} \left[\sum_{j=1}^{m} U_j^2 \right] \le P$$

where $U^m = f(X^n)$. Without loss of generality, we assume that $\mathbf{X} = \mathbf{Y}_1 = \cdots = \mathbf{Y}_K = 1$ and $Y_k = \rho_k X + N_k$ with $N_k \perp X$ and $\rho_k > 0$. Thus, $\mathbf{N}_k = 1 - \rho_k^2$, denotes the mean squared-error in estimating Y_k from X, or equivalently, X from Y_k. Reconstruction quality is measured by squared-error distance: $d_k(x, \hat{x}_k) = (x - \hat{x}_k)^2$.

- *Binary Hamming*: All source and channel alphabets are binary. The source is $\mathrm{Ber}(\frac{1}{2})$, where $\mathrm{Ber}(\epsilon)$ denotes the Bernoulli distribution with $P[1] = \epsilon$. The channels are binary symmetric with transition probabilities p_k, i.e., $V_k = U_k \oplus W_k$ where $W_k \sim \mathrm{Ber}(p_k)$ and $W_k \perp U_k$ with $\oplus$ denoting modulo 2 addition (or the XOR operation). The side information sequences at the receivers are also noisy versions of the source corrupted by passage through virtual binary symmetric channels. Therefore $Y_k = X_k \oplus N_k$ with $N_k \sim \mathrm{Ber}(\beta_k)$ and $N_k \perp X_k$. Reconstruction quality is measured by Hamming distance: $d_k(x, \hat{x}_k) = x \oplus \hat{x}_k$.

The problems considered in [9], [12], [18] can all be seen as special cases of the WZBC problem. However, the quadratic Gaussian and the binary Hamming cases with non-trivial side information have never, to our knowledge, been analyzed before. Nevertheless, separate source and channel coding and uncoded transmission are obvious strategies and we shall present numerical comparisons of the new schemes with those.

A. Wyner-Ziv Coding over Point-to-point Channels

Before analyzing the WZBC problem in depth, we shall briefly discuss known results for Wyner-Ziv coding over a point-to-point channel, i.e., the case $K = 1$. The Wyner-Ziv rate-distortion performance is characterized in [21] as

$$R^{WZ}(D) \triangleq \min_{\substack{Z, g \,:\, Y - X - Z \\ \mathrm{E}[d(X, g(Z,Y))] \leq D}} I(X; Z|Y) \tag{2}$$

where $Z \in \mathcal{Z}$ is an auxiliary random variable, and the capacity of the channel $p_{V|U}$ is well-known (cf. [4]) to be

$$C = \max_U I(U; V) \ .$$

It is then straightforward to conclude that combining separate source and channel codes yields the distortion

$$D = D^{WZ}(\kappa C) \tag{3}$$

where D^{WZ}, the distortion-rate function, is the inverse of R^{WZ}, i.e.,

$$D^{WZ}(R) \triangleq \min_{\substack{Z, g \,:\, Y - X - Z \\ I(X; Z|Y) \leq R}} \mathrm{E}[d(X, g(Z,Y))] \ .$$

On the other hand, a converse result in [13] shows that even by using joint source-channel codes, one cannot improve the distortion performance further than (3).

We are further interested in the evaluation of $D^{WZ}(R)$, as well as the test channels achieving it, for the quadratic Gaussian and binary Hamming cases. We will use similar test channels in our WZBC schemes.

1) Quadratic Gaussian: It was shown in [20] that the optimal backward test channel is given by

$$X = Z + S$$

where Z and S are independent Gaussians. The rate then becomes

$$I(X;Z|Y) = \frac{1}{2}\log\left(1 - \mathbf{N} + \frac{\mathbf{N}}{\mathbf{S}}\right) . \tag{4}$$

The optimal reconstruction is a linear estimate $g(Z,Y) = a_1 Z + a_2 Y$, which yields the distortion

$$\mathrm{E}[d(X, g(Z,Y))] = \frac{\mathbf{N}}{1 - \mathbf{N} + \frac{\mathbf{N}}{\mathbf{S}}} \tag{5}$$

and therefore,

$$D^{WZ}(R) = \mathbf{N}2^{-2R} . \tag{6}$$

2) Binary Hamming: It was implicitly shown in [21] that the optimal auxiliary random variable $Z \in \mathcal{Z} = \{0, 1, \lambda\}$ is given by

$$Z = E \circ (X \oplus S)$$

where X, E, S are all independent, E and S are Ber(q) and Ber(α) with $0 \le q \le 1$ and $0 \le \alpha \le \frac{1}{2}$, respectively, and $\circ$ is an erasure operator, i.e.,

$$a \circ b = \begin{cases} \lambda & a = 0 \\ b & a = 1 \end{cases} .$$

This choice results in

$$I(X;Z|Y) = qr(\alpha, \beta) \tag{7}$$

where

$$r(\alpha, \beta) = H_2(\alpha \star \beta) - H_2(\alpha)$$

with $\star$ and H_2 denoting the binary convolution, i.e., $a \star b = (1-a)b + a(1-b)$, and the binary entropy function, i.e.,

$$H_2(p) = -p\log p - (1-p)\log(1-p)$$

respectively. It is easy to show that when $0 \le \alpha, \beta \le \frac{1}{2}$, $r(\alpha, \beta)$ is increasing in β and decreasing in α.

Since $\mathrm{E}[d(X, g(Z,Y))] = \Pr[X \ne g(Z,Y))]$ and $X \sim$Ber($\frac{1}{2}$), the corresponding optimal reconstruction function g boils down to a maximum likelihood estimator given by

$$\begin{aligned} g(z,y) &= \arg\max_x p_{YZ|X}(y,z|x) \\ &= \arg\max_x p_{Z|X}(z|x)p_{Y|X}(y|x) \\ &= \begin{cases} y & z = \lambda \text{ or } z = y \\ z & z \ne \lambda, z \ne y \text{ and } \beta > \alpha \\ y & z \ne \lambda, z \ne y \text{ and } \beta \le \alpha \end{cases} . \end{aligned}$$

The resultant distortion is given by

$$E[d(X, g(Z, Y))] = q\min\{\alpha, \beta\} + (1-q)\beta \tag{8}$$

implying together with (7) that

$$D^{WZ}(R) = \min_{\substack{0 \le q \le 1, 0 \le \alpha \le \beta : \\ q\, r(\alpha,\beta) \le R}} \Big[q\alpha + (1-q)\beta \Big] \tag{9}$$

where the extra constraint $\alpha \le \beta$ is imposed because $\alpha > \beta$ is a provably suboptimal choice. It also follows from the discussion in [21] that there exists a critical rate $R_0(\beta)$ above which the optimal test channel assumes $q = 1$ and $0 \le \alpha \le \alpha_0(\beta) \le \beta$, and below which it assumes $\alpha = \alpha_0(\beta)$ and $0 \le q < 1$. The reason why we discussed other values of (q, α) above is because we will use the test channel in its most general form in all WZBC schemes.

B. A Trivial Converse for the WZBC Problem

At each terminal, no WZBC scheme can achieve a distortion less than the best distortion achievable by ignoring the presence of the other terminal. Thus,

$$D_k \ge D_k^{WZ}(\kappa C_k) \tag{10}$$

where C_k is the capacity of channel k. For the source-channel pairs we consider, (10) can be further specialized. For the quadratic Gaussian case, we obtain using (6) and

$$C_k = \frac{1}{2}\log\left(1 + \frac{P}{\mathbf{W}_k}\right)$$

that

$$D_k \ge \frac{\mathbf{N}_k}{(1 + \frac{P}{\mathbf{W}_k})^{\kappa}} \, . \tag{11}$$

For the binary Hamming case, using (9) and $C_k = 1 - H_2(p_k)$, the converse becomes

$$D_k \ge \min_{\substack{0 \le q \le 1, 0 \le \alpha \le \beta_k : \\ q\, r(\alpha,\beta) \le \kappa[1 - H_2(p_k)]}} q\alpha + (1-q)\beta_k$$

C. Separate Source and Channel Coding

For a general source and channel pair, the source and channel coding problems are extremely challenging. The set of all achievable rate triples (common and two private rates) for broadcast channels has not been characterized. The corresponding source coding problem has not been explicitly considered in previous work either. But there is considerable simplification in the quadratic Gaussian and binary Hamming cases since in both cases, the channel and the side information are degraded: we can assume that one of the two Markov chains, $U - V_1 - V_2$ or $U - V_2 - V_1$, holds (for arbitrary channel input U) for the channel, and similarly either $X - Y_1 - Y_2$ or $X - Y_2 - Y_1$ holds for the source. The capacity region for degraded broadcast channels is fully known. In fact, since any information sent to the weaker channel can be decoded by the stronger channel, we can assume that no private information is sent to the weaker channel. As a result, two layer source coding, which has been considered in [14], [16], [17], is sufficiently general.

For simplicity, we denote the random variables, rates, and distortion levels associated with the $g$ood channel by the subscript g and those associated with the $b$ad one by b, i.e., the channel variables always satisfy $U - V_g - V_b$ where g is either 1 or 2 and b takes the other value. A distortion pair (D_b, D_g) is achievable by separate source and channel coding with κ channel uses per source symbol if and only if

$$\mathcal{R}(D_b, D_g) \cap \kappa\mathcal{C} \neq \emptyset$$

where, as shown in [1], [5], $\kappa\mathcal{C}$ is the convex closure of all (R_b, R_g) such that there exist a channel input $U \in \mathcal{U}$ and an auxiliary random variable $U_b \in \mathcal{U}_b$ satisfying $U_b - U - V_g - V_b$, the power constraint (if any) $\mathrm{E}[U^2] \leq P$, and

$$R_b \leq \kappa I(U_b; V_b) \tag{12}$$

$$R_g \leq \kappa[I(U_b; V_b) + I(U; V_g|U_b)] \,. \tag{13}$$

Note that we use cumulative rates at the good receiver.

As for $\mathcal{R}(D_b, D_g)$, despite the simplification brought by degraded side information, there is no known complete single-letter characterization for all sources and distortion measures when $X - Y_b - Y_g$. Let $\mathcal{R}^*(D_b, D_g)$ be defined as the convex closure of all (R_b, R_g) such that there exist source auxiliary random variables $(Z_b, Z_g) \in \mathcal{Z}_b \times \mathcal{Z}_g$ with either $(Y_b, Y_g) - X - Z_b - Z_g$ or $(Y_b, Y_g) - X - Z_g - Z_b$, and reconstruction functions $g_k : \mathcal{Z}_k \times \mathcal{Y}_k \to \hat{\mathcal{X}}, k = b, g$ satisfying

$$\mathrm{E}[d_k(X, g_k(Z_k, Y_k))] \leq D_k \tag{14}$$

for $k = b, g$, and

$$R_b \geq I(X; Z_b|Y_b) \tag{15}$$

$$R_g \geq \begin{cases} I(X; Z_b|Y_b) + [I(X; Z_g|Y_g) - I(X; Z_b|Y_g)]^+ & \text{if } X - Y_g - Y_b \\ I(X; Z_g|Y_g) + [I(X; Z_b|Y_b) - I(X; Z_g|Y_b)]^+ & \text{if } X - Y_b - Y_g \end{cases} . \tag{16}$$

It was shown in [14] that $\mathcal{R}(D_b, D_g) = \mathcal{R}^*(D_b, D_g)$ when $X - Y_g - Y_b$. On the other hand, [16] showed that even when $X - Y_b - Y_g$, $\mathcal{R}(D_b, D_g) = \mathcal{R}^*(D_b, D_g)$ for the quadratic Gaussian problem. For all other sources and distortion measures, we only know $\mathcal{R}(D_b, D_g) \supset \mathcal{R}^*(D_b, D_g)$ in general when $X - Y_b - Y_g$.

We next specialize the discussion to the quadratic Gaussian and binary Hamming problems.

1) Quadratic Gaussian: For the Gaussian channel, $\kappa\mathcal{C}$ is achieved by Gaussian U_b and $U-U_b$ with $U_b \perp U-U_b$, using which (12) and (13) become (cf. [4])

$$R_b \leq \frac{\kappa}{2}\log\left(1+\frac{\nu P}{\bar{\nu}P+\mathbf{W}_b}\right) \quad (17)$$

$$R_g \leq \frac{\kappa}{2}\log\left(\left[1+\frac{\nu P}{\bar{\nu}P+\mathbf{W}_b}\right]\left[1+\frac{\bar{\nu}P}{\mathbf{W}_g}\right]\right) \quad (18)$$

where $0 \leq \nu \leq 1$ and $\bar{\nu} = 1-\nu$ control the power allocation between U_b and $U-U_b$. On the other hand, $\mathcal{R}(D_b, D_g) = \mathcal{R}^*(D_b, D_g)$ is also achieved by the test channel $X = Z_k + S_k$ with $Z_k \perp S_k$ for $k = b, g$. From (4) and (5), this implies that for any $k, k' \in \{b, g\}$

$$\begin{aligned} I(X; Z_k|Y_{k'}) &= \frac{1}{2}\log\left(1-\mathbf{N}_{k'}+\frac{\mathbf{N}_{k'}}{\mathbf{S}_k}\right) \\ &= \begin{cases} \frac{1}{2}\log\frac{\mathbf{N}_k}{D_k} & k=k' \\ \frac{1}{2}\log\left(\frac{\mathbf{N}_{k'}}{D_k}+1-\frac{\mathbf{N}_{k'}}{\mathbf{N}_k}\right) & k\neq k' \end{cases}. \end{aligned} \quad (19)$$

Combining (15), (17), and (19), we obtain

$$\frac{\mathbf{N}_b}{D_b} \leq \left(1+\frac{\nu P}{\bar{\nu}P+\mathbf{W}_b}\right)^{\kappa}. \quad (20)$$

Similarly, (16), (18), and (19) yields

$$\frac{\mathbf{N}_b^2\mathbf{N}_g}{D_g\left[\mathbf{N}_g\mathbf{N}_b+D_b\left(\mathbf{N}_b-\mathbf{N}_g\right)\right]} \leq \left(1+\frac{\nu P}{\bar{\nu}P+\mathbf{W}_b}\right)^{\kappa}\left(1+\frac{\bar{\nu}P}{\mathbf{W}_g}\right)^{\kappa} \quad (21)$$

when $X - Y_g - Y_b$, and

$$\frac{\mathbf{N}_g}{\min\left\{D_g, D_b+\frac{D_bD_g}{\mathbf{N}_b\mathbf{N}_g}\left(\mathbf{N}_g-\mathbf{N}_b\right)\right\}} \leq \left(1+\frac{\nu P}{\bar{\nu}P+\mathbf{W}_b}\right)^{\kappa}\left(1+\frac{\bar{\nu}P}{\mathbf{W}_g}\right)^{\kappa} \quad (22)$$

when $X - Y_b - Y_g$. The next lemma, whose proof we defer to Appendix A, characterizes the (D_b, D_g) tradeoff for both $X - Y_g - Y_b$ and $X - Y_b - Y_g$ for the important case $\kappa = 1$. This result will be useful when we compare our schemes to separate coding.

Lemma 1: For the quadratic Gaussian case with $\kappa = 1$, the distortion pair (D_b, D_g) with $D_b^{WZ}(C_b) \leq D_b \leq \mathbf{N}_b$ is achievable using separate coding if and only if $D_g \geq D_{\text{SEP}}(D_b)$ where $D_{\text{SEP}}(D_b)$ is the convex hull of

$$D_{\text{SEP}}^*(D_b) = \frac{\mathbf{N}_g\mathbf{N}_b^2\mathbf{W}_gD_b}{\Big(D_b\mathbf{N}_b+\mathbf{N}_g(\mathbf{N}_b-D_b)\Big)\Big((\mathbf{W}_g-\mathbf{W}_b)\mathbf{N}_b+(P+\mathbf{W}_b)D_b\Big)} \quad (23)$$

when $X - Y_g - Y_b$, and

$$D_{\text{SEP}}^*(D_b) = \frac{\mathbf{N}_g}{\Big((\mathbf{W}_g-\mathbf{W}_b)\mathbf{N}_b+(P+\mathbf{W}_b)D_b\Big)}\max\left\{\mathbf{W}_gD_b, \frac{\mathbf{N}_b\Big(\mathbf{N}_g\mathbf{W}_g-(P+\mathbf{W}_b)D_b-\mathbf{N}_b(\mathbf{W}_g-\mathbf{W}_b)\Big)}{\mathbf{N}_g-\mathbf{N}_b}\right\} \quad (24)$$

when $X - Y_b - Y_g$.

2) Binary Hamming: For the binary symmetric channel, $\kappa\mathcal{C}$ is achieved by $U_b \sim \text{Ber}(\frac{1}{2})$ and $U = U_b \oplus U_g$ with $U_g \sim \text{Ber}(\theta)$ and $U_g \perp U_b$. The parameter θ serves as a tradeoff between R_b and R_g. The conditions (12) and (13) then become (cf. [4])

$$R_b \leq \kappa[1 - H_2(\theta \star p_b)] \tag{25}$$

$$R_g \leq \kappa[H_2(\theta \star p_g) - H_2(p_g)] \; . \tag{26}$$

For the source coding part, we evaluate $\mathcal{R}^*(D_b, D_g)$ only with $\mathcal{Z}_b = \mathcal{Z}_g = \{0, 1, \lambda\}$, where the test channels are also confined to degraded versions of those that achieve $D^{WZ}(R)$, as shown in Figure 2 for the case $(Y_b, Y_g) - X - Z_g - Z_b$. More specifically,

$$\begin{aligned} Z_g &= E_g \circ (X \oplus S_g) \\ Z_b &= E_b \circ (X \oplus S_b) \end{aligned}$$

where E_g, E_b, S_g, and S_b are all Bernoulli random variables with parameters q_g, q_b, α_g, and α_b, respectively. To obtain a Markov relation $X - Z_g - Z_b$, it suffices to enforce $q_b \leq q_g$ and $\alpha_b \geq \alpha_g$. In that case, one can find $0 \leq q_b' \leq 1$ and $0 \leq \alpha_b' \leq \frac{1}{2}$ such that $q_b = q_g q_b'$ and $\alpha_b = \alpha_g \star \alpha_b'$, and Z_b can alternatively be written as

$$Z_b = \begin{cases} E_b' \circ (Z_g \oplus S_b') & Z_g \neq \lambda \\ \lambda & Z_g = \lambda \end{cases}$$

where E_b' and S_b' are $\text{Ber}(q_b')$ and $\text{Ber}(\alpha_b')$, respectively. Similarly, to obtain the other Markov chain, $X - Z_b - Z_g$, we need $q_b \geq q_g$ and $\alpha_b \leq \alpha_g$.

These simple choices may *potentially* result in degradation of the separate coding performance, as the bounds on the alphabet sizes for $\mathcal{Z}_b$ and $\mathcal{Z}_g$ in [14], [16], [17] are much larger. However, our limited choice of (Z_b, Z_g) can be justified in two ways: (i) to the best of our knowledge, there is no other choice known to achieve better rates, and (ii) to be fair, we use the same choice in our joint source-channel coding schemes.

As in the quadratic Gaussian case, we can write

$$I(X; Z_k | Y_{k'}) = q_k r(\alpha_k, \beta_{k'}) \tag{27}$$

for $k, k' \in \{b, g\}$. Combining (15), (25), and (27) yields

$$q_b r(\alpha_b, \beta_b) \leq \kappa[1 - H_2(\theta \star p_b)] \; . \tag{28}$$

Similarly, combining (16), (26), and (27), we obtain

$$q_b r(\alpha_b, \beta_b) + [q_g r(\alpha_g, \beta_g) - q_b r(\alpha_b, \beta_g)]^+ \leq \kappa[H_2(\theta \star p_g) - H_2(p_g)] \tag{29}$$

when $X - Y_g - Y_b$, and

$$q_g r(\alpha_g, \beta_g) + [q_b r(\alpha_b, \beta_b) - q_g r(\alpha_g, \beta_b)]^+ \leq \kappa[H_2(\theta \star p_g) - H_2(p_g)] \tag{30}$$

when $X - Y_b - Y_g$.

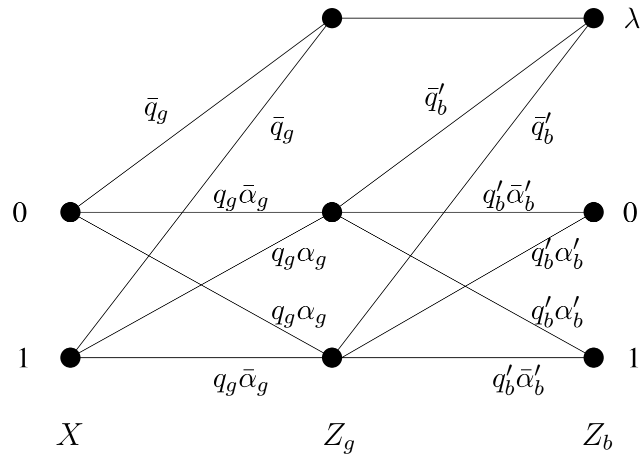

Fig. 2. Auxiliary random variables for binary source coding. The edge labels denotes transition probabilities. We also use the convention that $\bar{a} = 1 - a$.

D. Uncoded Transmission

If the rate $\kappa = 1$, and if the source and channel alphabets are compatible, uncoded transmission is a possible strategy. For the quadratic Gaussian case, the distortion achieved by uncoded transmission is given by

$$D_k = \frac{\mathbf{N}_k \mathbf{W}_k}{\mathbf{W}_k + \mathbf{N}_k P} \tag{31}$$

for $k = 1, 2$. This, in turn, is also because the channel is the same as the test channel up to a scaling factor. More specifically, when $\sqrt{P}X$ is transmitted and corrupted by noise W_k, one can write $X = Z_k + S_k$ with $S_k \perp Z_k$, where Z_k is an appropriately scaled version of the received signal $\sqrt{P}X + W_k$ and

$$\mathbf{S}_k = \frac{\mathbf{W}_k}{\mathbf{W}_k + P} \ .$$

Substituting this into (5) then yields (31). Comparing with (11), we note that (31) achieves $D_k^{WZ}(C_k)$ only when $\mathbf{N}_k = 1$ or when $\mathbf{W}_k \to \infty$, which, in turn, translate to trivial Y_k or zero C_k, respectively.

For the binary Hamming case, this strategy achieves the distortion pair

$$D_k = \min\{p_k, \beta_k\} \tag{32}$$

for $k = 1, 2$. That is because the channel is the same as the test channel that achieves $D^{WZ}(R)$ with $q = 1$. The distortion expression in (32) then follows using (8). One can also show that (32) coincides with $D_k^{WZ}(C_k)$ only when $\beta_k = \frac{1}{2}$ or $p_k = \frac{1}{2}$. Once again, these respectively correspond to trivial Y_k and zero C_k.

III. Basic WZBC Schemes

In this section, we present the basic coding schemes that we shall then develop into the schemes that form the main contributionS of this paper.

The first scheme, termed Scheme 0, is a basic extension of the scheme in [18] where the source is first quantized before transmission over the channel. Even though our layered schemes are constructed for the case of $K = 2$ receivers, Scheme 0 can be utilized for any $K \geq 2$.

Theorem 1: A distortion tuple $(D_1, \dots, D_K)$ is achievable at rate κ if there exist random variables $Z \in \mathcal{Z}$, $U \in \mathcal{U}$ and functions $g_k : \mathcal{Z} \times \mathcal{Y}_k \rightarrow \hat{\mathcal{X}}_k$ with $(Y_1, \dots, Y_K) - X - Z$ such that

$$I(X; Z|Y_k) < \kappa I(U; V_k) \tag{33}$$

$$\mathrm{E}[d_k(X, g_k(Z, Y_k))] \leq D_k \tag{34}$$

for $k = 1, \dots, K$.

Here and in what follows, we only present code constructions for discrete sources and channels. The constructions can be extended to the continuous case in the usual manner. Our coding arguments rely heavily on the notion of typicality. Given a random variable $X \sim P_X(x), x \in \mathcal{X}$ the typical set at block length n is defined as [10]

$$\mathcal{T}_\delta^n(X) \triangleq \left\{ x^n \in \mathcal{X}^n : \left| \frac{N(a|x^n)}{n} - P_X(a) \right| \leq \delta P_X(a), \forall a \in \mathcal{X} \right\}$$

where $N(a|x^n)$ denotes the number of times a appears in x^n.

Proof: Pick random variables satisfying (33) and (34). For a fixed $\delta, \delta', \delta'' > 0$, the coders will operate as follows. The encoder constructs a source codebook $\mathcal{C}_Z \triangleq \{z^n(i), i = 1, \dots, M\}$ by choosing sequences from $\mathcal{T}_\delta^n(Z)$ uniformly at random. The codebook size, M, is to be determined later. Similarly, the encoder constructs a channel codebook $\mathcal{C}_U \triangleq \{u^m(i), i = 1, \dots, M\}$ from $\mathcal{T}_\delta^m(U)$. Given a source sequence X^n, the encoder finds $i^* \in \{1, \dots, M\}$ such that $(X^n, z^n(i^*)) \in \mathcal{T}_{\delta'}^n(X, Z)$ and transmits $u^m(i^*)$. The encoder declares an error if it cannot find i^*. The decoder at terminal k finds the smallest $i \in \{1, \dots, M\}$ such that $(u^m(i), V_k^m) \in \mathcal{T}_{\delta''}^m(U, V_k)$ and simultaneously $(Y_k^n, z^n(i)) \in \mathcal{T}_{\delta''}^n(Y_k, Z)$. If no such i is found, the decoder sets $i = 1$. Once i is found, coordinate-wise reconstruction is performed using g_k with Y_k^n and $z^n(i)$.

Defining the error events

$$\begin{aligned}
\mathcal{E}_1 &= \left\langle \forall i, \;\; (X^n, z^n(i)) \notin \mathcal{T}_{\delta'}^n(X, Z) \right\rangle \\
\mathcal{E}_2(k) &= \left\langle (Y_k^n, z^n(i^*)) \notin \mathcal{T}_{\delta''}^n(Y_k, Z) \right\rangle \\
\mathcal{E}_3(k) &= \left\langle (u^m(i^*), V_k^m) \notin \mathcal{T}_{\delta''}^m(U, V_k) \right\rangle \\
\mathcal{E}_4(k) &= \left\langle \exists i \neq i^*, \;\; (Y_k^n, z^n(i)) \in \mathcal{T}_{\delta''}^n(Y_k, Z) \;\text{ and }\; (u^m(i), V_k^m) \in \mathcal{T}_{\delta''}^m(U, V_k) \right\rangle
\end{aligned}$$

it suffices to show that the probability of all the error events go to zero uniformly as $n \rightarrow \infty$. Using standard arguments, we have that $\Pr[\mathcal{E}_1] < \epsilon$ for any $\epsilon > 0$ and large enough n if

$$M \geq 2^{n[I(X;Z) + \epsilon_1(\delta, \delta', \delta'')]}$$

where $\epsilon_1(\delta,\delta',\delta'') \to 0$ as $\delta,\delta',\delta'' \to 0$. That $\Pr[\mathcal{E}_2(k)] < \epsilon$ and $\Pr[\mathcal{E}_3(k)] < \epsilon$ for large enough n also follow from standard arguments. As for $\Pr[\mathcal{E}_4(k)]$, it is easy to show using properties of typical sequences that

$$\begin{aligned}\Pr[\mathcal{E}_4(k)] &\leq M 2^{-n[I(Y_k;Z)-\epsilon_2(\delta,\delta',\delta'')]} 2^{-m[I(U;V_k)-\epsilon_2(\delta,\delta',\delta'')]} \\ &= M 2^{-n[I(Y_k;Z)+\kappa I(U;V_k)-(\kappa+1)\epsilon_2(\delta,\delta',\delta'')]}\end{aligned}$$

for large enough n, where $\epsilon_2(\delta,\delta',\delta'') \to 0$ as $\delta,\delta',\delta'' \to 0$. $\Pr[\mathcal{E}_4(k)]$ then vanishes if

$$M \leq 2^{n[I(X;Z)+2\epsilon_1(\delta,\delta',\delta'')]}$$

and

$$I(X;Z|Y_k) \leq \kappa I(U;V_k) - (\kappa+1)\epsilon_2(\delta,\delta',\delta'') - 2\epsilon_1(\delta,\delta',\delta'')]$$

which will be granted when $\delta,\delta',\delta'' \to 0$. It is also straightforward to show that when none of the error events occur,

$$\mathrm{E}[d_k(X^n, g_k(z^n(i^*), Y_k^n))] \leq D_k + \epsilon$$

for sufficiently high n as $\delta,\delta',\delta'' \to 0$. The proof is therefore complete. ■

Note that, as in the proof of the achievability part of Theorem 6 in [18], there is no explicit binning of the source codebook and an equivalent operation (i.e., virtual binning) is automatically performed by the channel.

Next, we give a dirty-paper version of Theorem 1 which will be useful in some of our achievable schemes. Suppose that there is CSI available solely at the encoder, i.e., the broadcast channel is defined by the transition probability $p_{V_1V_2|US}(v_1, v_2|u, s)$ and the CSI $S^m \in \mathcal{T}_\eta^m(S)$ with some $\eta > 0$, where S is some fixed distribution defined on the CSI alphabet $\mathcal{S}$, is available non-causally at the encoder. Given a source and side information at the decoders (X, Y_1, Y_2), codes (m, n, f, g_1, g_2) and achievability of distortion pairs is defined as in the WZBC scenario except that the encoder now takes the form $f : \mathcal{X}^n \times \mathcal{S}^m \to \mathcal{U}^m$.

Theorem 2: A distortion pair $(D_1, \ldots, D_K)$ is achievable at rate κ if there exist random variables $Z \in \mathcal{Z}$, $T \in \mathcal{T}, U \in \mathcal{U}$ and functions $g_k : \mathcal{Z} \times \mathcal{Y}_k \to \hat{\mathcal{X}}$ with $(Y_1, \ldots, Y_K) - X - Z$ and $T - (U, S) - (V_1, \ldots, V_K)$ such that

$$I(X;Z|Y_k) < \kappa\big[I(T;V_k) - I(T;S)\big] \quad (35)$$

$$\mathrm{E}[d_k(X, g_k(Z, Y_k))] \leq D_k \quad (36)$$

for $k = 1, \ldots, K$.

Proof: The code construction, which will be referred to as Scheme 0 with DPC, is as follows. As before, for fixed $\delta,\delta',\delta'' > 0$, a source codebook $\mathcal{C}_Z \triangleq \{z^n(i), i = 1, \ldots, M\}$ is chosen from $\mathcal{T}_\delta^n(Z)$. A set of M bins $\mathcal{C}_T(i) = \{t^m(i,j), j = 1, \ldots, M'\}$, where each $t^m(i,j)$ is chosen randomly at uniform from $\mathcal{T}_\delta^m(T)$, is also constructed. Given a source word X^n and CSI S^m, the encoder tries to find a pair (i^*, j^*) such that $(X^n, z^n(i^*)) \in$

$\mathcal{T}^n_{\delta'}(X,Z)$ and $(S^m, t^m(i^*,j^*)) \in \mathcal{T}^m_{\delta'}(S,T)$. If it is unsuccessful, it declares an error. If it is successful, the channel input is drawn from the distribution $\prod_{l=1}^m p_{U|TS}(u_l|t_l(i^*,j^*),S_l)$. At terminal k, the decoder goes through all pairs $(i,j) \in \{1,\ldots,M\} \times \{1,\ldots,M'\}$ until it finds the first pair satisfying $(Y_k^n, z^n(i)) \in \mathcal{T}^n_{\delta''}(Y_k,Z)$ and $(V_k^m, t^m(i,j)) \in \mathcal{T}^m_{\delta''}(V_k,T)$ simultaneously. If there is no such pair, the decoder sets $i=1, j=1$. Once (i,j) is decided, coordinate-wise reconstruction is performed using g_k with Y_k^n and $z^n(i)$ as in Scheme 0.

This time we define the error events as

$$\begin{aligned}
\mathcal{E}_1 &= \left\langle \forall(i,j),\ \text{ either } (X^n, z^n(i)) \notin \mathcal{T}^n_{\delta'}(X,Z) \ \text{ or } \ (S^m, t^m(i,j)) \notin \mathcal{T}^m_{\delta'}(S,T) \right\rangle \\
\mathcal{E}_2(k) &= \left\langle (Y_k^n, z^n(i^*)) \notin \mathcal{T}^n_{\delta''}(Y_k,Z) \right\rangle \\
\mathcal{E}_3(k) &= \left\langle (V_k^m, t^m(i^*,j^*)) \notin \mathcal{T}^m_{\delta''}(V_k,T) \right\rangle \\
\mathcal{E}_4(k) &= \left\langle \exists(i \neq i^*, j),\ \ (Y_k^n, z^n(i)) \in \mathcal{T}^n_{\delta''}(Y_k,Z) \ \text{ and } \ (V_k^m, t^m(i,j)) \in \mathcal{T}^m_{\delta''}(V_k,T) \right\rangle .
\end{aligned}$$

Again, using standard typicality arguments, it can be shown that for fixed $\delta, \delta', \delta''$, if

$$M \geq 2^{n[I(X;Z)+\epsilon_1(\delta,\delta',\delta'')]}$$

and

$$M' \geq 2^{m[I(S;T)+\epsilon_1(\delta,\delta',\delta'')]}$$

then $\Pr[\mathcal{E}_1] < \epsilon$, and that $\Pr[\mathcal{E}_2(k)] < \epsilon$ and $\Pr[\mathcal{E}_3(k)] < \epsilon$ for any $\epsilon > 0$ and large enough n. Similarly, it follows that if

$$M \leq 2^{n[I(X;Z)+2\epsilon_1(\delta,\delta',\delta'')]}$$

and

$$M' \leq 2^{m[I(S;T)+2\epsilon_1(\delta,\delta',\delta'')]}$$

then

$$\begin{aligned}
\Pr[\mathcal{E}_4(k)] &\leq M \cdot M' \cdot 2^{-n[I(Y_k;Z)-\epsilon_2(\delta,\delta',\delta'')]} 2^{-m[I(T;V_k)-\epsilon_2(\delta,\delta',\delta'')]} \\
&= M \cdot M' \cdot 2^{-n[I(Y_k;Z)+\kappa I(T;V_k)-(\kappa+1)\epsilon_2(\delta,\delta',\delta'')]} \\
&\leq 2^{n[I(X;Z|Y_k)-\kappa\{I(T;V_k)-I(S;T)\}+(\kappa+1)\epsilon_2(\delta,\delta',\delta'')+2(\kappa+1)\epsilon_1(\delta,\delta',\delta'')]} .
\end{aligned}$$

This probability also vanishes if $\delta, \delta', \delta'' \to \infty$ thanks to (35). This completes the proof. ■

Corollary 1: The coding scheme in the proof can also decode $t^m(i^*,j^*)$ successfully.

Proof: Define

$$\mathcal{E}_5(k) = \left\langle \exists j \neq j^*,\ \ (V_k^m, t^m(i^*,j)) \in \mathcal{T}^m_{\delta''}(V_k,T) \right\rangle .$$

It then suffices to show that $\Pr[\mathcal{E}_5(k)] < \epsilon$ for large enough n. Indeed, since $I(T;V_k) - I(S;T) > 0$,

$$\begin{aligned}\Pr[\mathcal{E}_5(k)] &\leq M' 2^{-m[I(T;V_k)-\epsilon_2(\delta,\delta',\delta'')]} \\ &\leq 2^{-m[I(T;V_k)-I(S;T)-\epsilon_2(\delta,\delta',\delta'')-2\epsilon_1(\delta,\delta',\delta'')]} \\ &\leq \epsilon \ .\end{aligned}$$

The assumption $I(T;V_k) - I(S;T) > 0$ is not restrictive at all, because otherwise no information can be delivered to terminal k to begin with. ■

The significance of this corollary is that decoding $t^m(i^*, j^*)$ provides information about the CSI S^m. This information, in turn, will be very useful in our layered WZBC schemes where the CSI is self-imposed and related to the source X^n itself.

Examining the proofs of Theorems 1 and 2, we notice an apparent separation between source and channel coding in that the source and channel codebooks are independently chosen. Furthermore, successful transmission is possible as long as the source coding rate for each terminal is less than the corresponding channel capacity for a common channel input. However, the decoding must be jointly performed and neither scheme can be split into separate stand-alone source and channel codes. Nevertheless, due to the quasi-independence of the source and channel codebooks we shall refer to source codes and channel codes separately when we discuss layered WZBC schemes. This quasi-separation was shown to be optimal for the SWBC problem and was termed *operational separation* in [18].

IV. Layered WZBC Schemes

In Scheme 0, the same information is conveyed to both receivers. However, since the side information and channel characteristics at the two receiving terminals can be very different, we might be able to improve the performance by layered coding, i.e., by not only transmitting a common layer (CL) to both receivers but also additionally transmitting a refinement layer (RL) to one of the two receivers. The resultant interference between the CL and RL can then be mitigated by successive decoding or by dirty paper encoding. This motivates us to initially consider four extensions of Scheme 0, one for each possible order of channel encoding and decoding (at the receiver which decodes both layers) the CL and the RL. We term our schemes CR-CR, CR-RC, RC-CR, and RC-RC, where the first and the second permutations of C and R respectively represent the encoding and the decoding orders. We then show that the effective capacity region of Scheme CR-RC is always inferior to that of Scheme CR-CR, and Scheme CR-CR, in turn, becomes a special case of Scheme RC-CR under the regime where the channel codeword in Scheme CR-CR is constrained to be the "addition" of the CL codeword and an independent RL codeword, which is the case in both the quadratic Gaussian and the binary Hamming problems.

Unless the better channel also has access to better side information, it is not straightforward to decide which receiver should receive only the CL and which should additionally receive the RL. For ease of exposition, we rename the source and channel random variables by replacing the subscripts 1 and 2 by c and r and defer the decision on which receiver is which. For the quadratic Gaussian problem, we will later develop an analytical decision tool. For all other sources and channels, one can combine the distortion regions resulting from the two choices, namely, $c = 1, r = 2$ and $c = 2, r = 1$.

As mentioned earlier, the inclusion of an RL codeword changes the effective channel observed while decoding the CL. It is on this modified channel that we send the CL using Scheme 0 or Scheme 0 with DPC and the capacity expressions in (33) and (35) must be modified in a manner that we describe in the following subsections where we also present the capacity of the effective channel for transmitting the RL.

The RL is transmitted by separate source and channel coding. In coding the source, we restrict our attention to systems where the communicated information satisfies $(Y_c, Y_r) - X - Z_r - Z_c$ where Z_c corresponds to the CL and Z_r is the RL. The source coding rate for the RL is therefore $I(X; Z_r | Z_c, Y_r)$ (cf. [16]). This has to be less than the RL capacity. Due to the separability of the source and channel variables in the required inequalities, for all four schemes, we can say that a distortion pair (D_c, D_r) is achievable if

$$\mathcal{R}_{\mathrm{WZBC}}(D_c, D_r) \cap \kappa \mathcal{C}_{\mathrm{WZBC}} \neq \emptyset \ .$$

Here, $\mathcal{R}_{\mathrm{WZBC}}(D_c, D_r)$ is the set of all triplets (R_{cc}, R_{cr}, R_{rr}) so that there exist (Z_c, Z_r) and reconstruction functions $g_c : \mathcal{Z}_c \times \mathcal{Y}_c \to \hat{\mathcal{X}}_c$ and $g_r : \mathcal{Z}_r \times \mathcal{Y}_r \to \hat{\mathcal{X}}_r$ satisfying $(Y_c, Y_r) - X - Z_r - Z_c$ and

$$I(X; Z_c | Y_c) \leq R_{cc} \tag{37}$$

$$I(X; Z_c | Y_r) \leq R_{cr} \tag{38}$$

$$I(X; Z_r | Z_c, Y_r) \leq R_{rr} \tag{39}$$

$$\mathrm{E}[d_c(X, g_c(Z_c, Y_c))] \leq D_c \tag{40}$$

$$\mathrm{E}[d_r(X, g_r(Z_r, Y_r))] \leq D_r \ . \tag{41}$$

The subscripts cc and cr are used to emphasize transmission of the CL to receivers c and r, respectively. Similarly, the subscript rr refers to transmission of RL to receiver r. The expressions for the source rates do not depend on the encoding and decoding orders.

Characterizing the effective capacity region $\mathcal{C}_{\mathrm{WZBC}}$ for the various schemes is the task of the rest of the section. We will only sketch the proofs of the theorems, as they rely only on Scheme 0, Scheme 0 with DPC, and other standard tools.

A. Scheme CR-CR

Consider the system shown in Figure 3, where RL is superposed over CL. This is the simplest extension of Scheme 0.

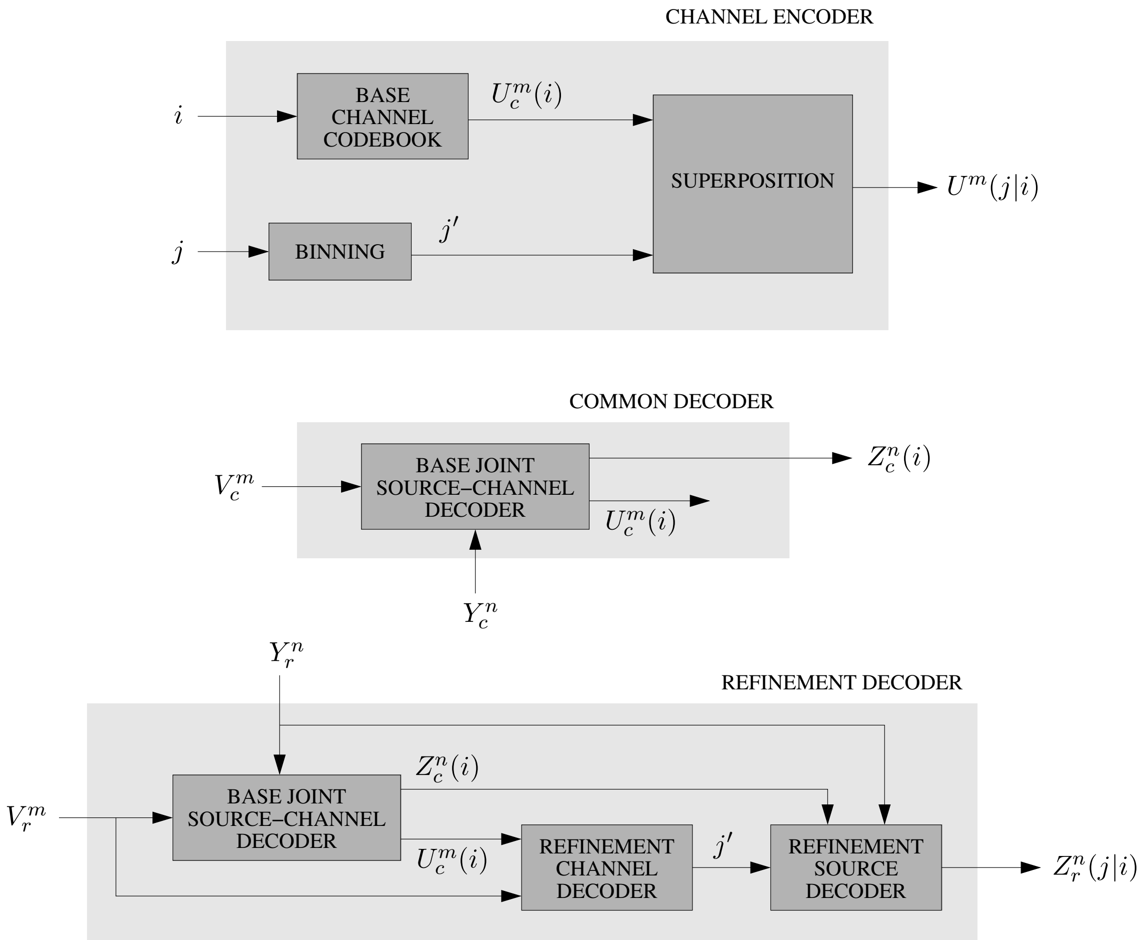

Fig. 3. Components of Scheme CR-CR.

Theorem 3: Let $\mathcal{C}_{\mathrm{CRCR}}$ be the union of all (C_{cc}, C_{cr}, C_{rr}) for which there exist U_c in some auxiliary alphabet $\mathcal{U}_c$ and $U \in \mathcal{U}$ with $U_c - U - (V_c, V_r)$ such that

$$C_{cc} \leq I(U_c; V_c) \tag{42}$$

$$C_{cr} \leq I(U_c; V_r) \tag{43}$$

$$C_{rr} \leq I(U; V_r|U_c)\ . \tag{44}$$

Then $\mathcal{C}_{\mathrm{CRCR}} \subseteq \mathcal{C}_{\mathrm{WZBC}}$.

Remark 1: To see that Scheme 0 is indeed a special case of Scheme CR-CR, it suffices to use the trivial superposition $U = U_c$.

Proof: Given random variables U and U_c such that $U_c - U - (V_c, V_r)$ and (42)-(44) are satisfied, each $U_c^m(i)$ in the CL channel codebook is chosen uniformly and independently from $\mathcal{T}_{\delta}^m(U_c)$. Similarly, for each i, codewords $U^m(j'|i)$ to be transmitted over the channel are chosen uniformly and independently from $\mathcal{T}_{\delta'}^m(U|U_c)$. It then follows from Theorem 1 that (42) and (43) are sufficient for successful decoding of both $Z_c^n(i)$ and $U_c^m(i)$ simultaneously at both decoders. It also follows from standard arguments that (44) is sufficient for reliable transmission of additional

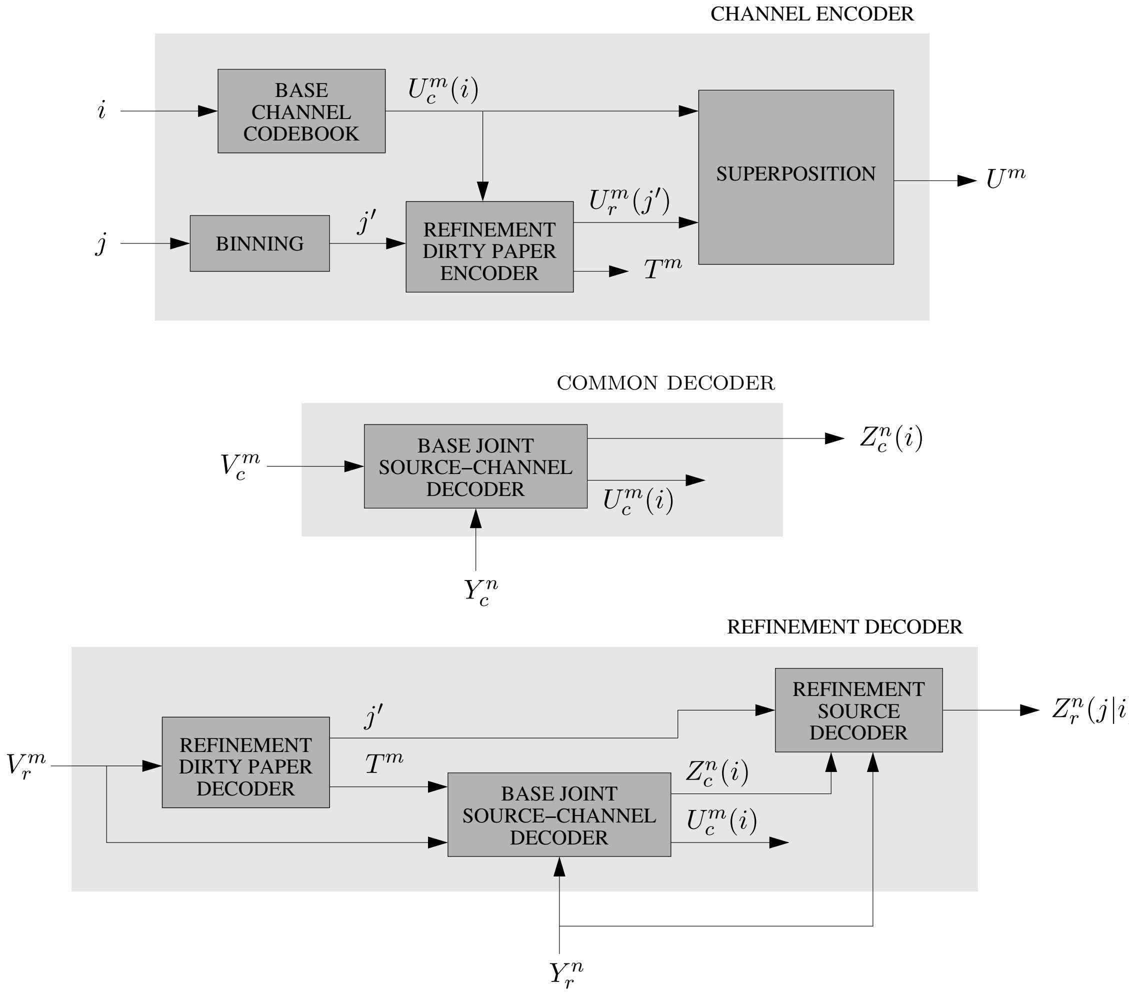

Fig. 4. Components of Scheme CR-RC.

information with rate C_{rr} to the refinement receiver. ■

B. Scheme CR-RC

This scheme is depicted in Figure 4. The CL is encoded as in Scheme CR-CR. The RL, however, is sent using dirty paper coding with the CL codeword as encoder CSI.

Theorem 4: Let $\mathcal{C}_{\text{CRRC}}$ be the union of all (C_{cc}, C_{cr}, C_{rr}) for which there exist $U_c \in \mathcal{U}_c$, $U_r \in \mathcal{U}_r$, and $T \in \mathcal{T}$ with $T - (U_r, U_c) - (V_r, V_c)$ and $(U_r, U_c) - U - (V_r, V_c)$ such that

$$C_{cc} \leq I(U_c; V_c) \tag{45}$$

$$C_{cr} \leq I(U_c; T, V_r) \tag{46}$$

$$C_{rr} \leq I(T; V_r) - I(T; U_c)\ . \tag{47}$$

Then $\mathcal{C}_{\text{CRRC}} \subseteq \mathcal{C}_{\text{WZBC}}$.

Remark 2: To specialize Scheme CR-RC to Scheme 0, one needs to set $U = U_c$ and trivially pick T as a random variable independent of U_c and U_r.

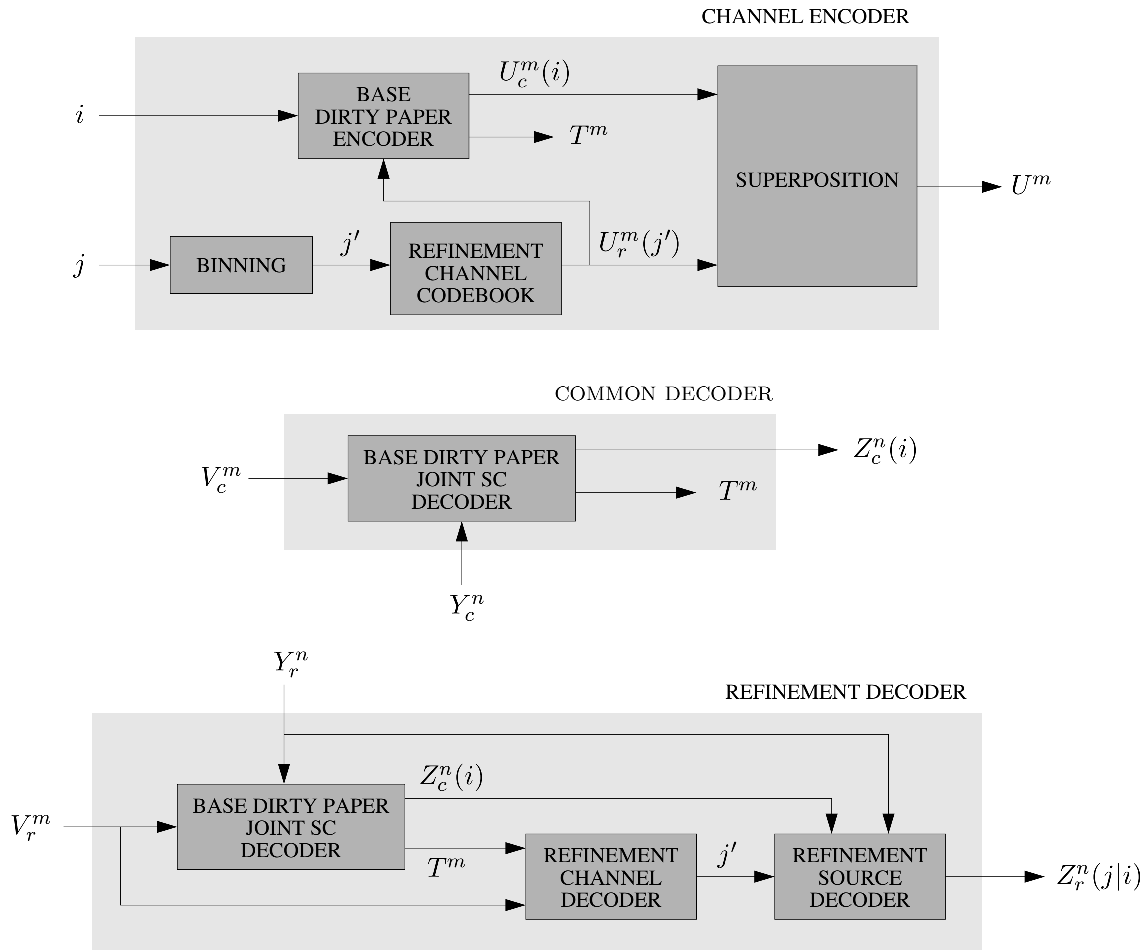

Fig. 5. Components of Scheme RC-CR.

Proof: Since RL is to be sent by separate source and channel codes, the channel coding part can proceed as in standard dirty-paper coding (cf. [6]), if (47) is satisfied. Note that as in Corollary 1, the auxiliary codeword T^m can also be decoded in the process of decoding the RL. With high probability, this codeword is typical with the CL codeword U_c^m in addition to V_r^m. Subsequently, for decoding the CL, the channel output at the r decoder can be taken to be a pair (V_r^m, T^m). Therefore, as in Scheme CR-CR, $Z_c^n(i)$ can be successfully decoded given that (45) and (46) hold. ■

C. Scheme RC-CR

As shown in Figure 5, in this scheme, the CL codeword is now dirty paper coded with the RL codeword acting as CSI. The next theorem provides the effective capacity region for Scheme RC-CR.

Theorem 5: Let $\mathcal{C}_{\text{RCCR}}$ be the union of all (C_{cc}, C_{cr}, C_{rr}) for which there exist $U_c \in \mathcal{U}_c$, $U_r \in \mathcal{U}_r$, and $T \in \mathcal{T}$ with $T - (U_r, U_c) - (V_r, V_c)$ and $(U_r, U_c) - U - (V_r, V_c)$ such that

$$C_{cc} \leq I(T; V_c) - I(T; U_r) \tag{48}$$

$$C_{cr} \leq I(T; V_r) - I(T; U_r) \tag{49}$$

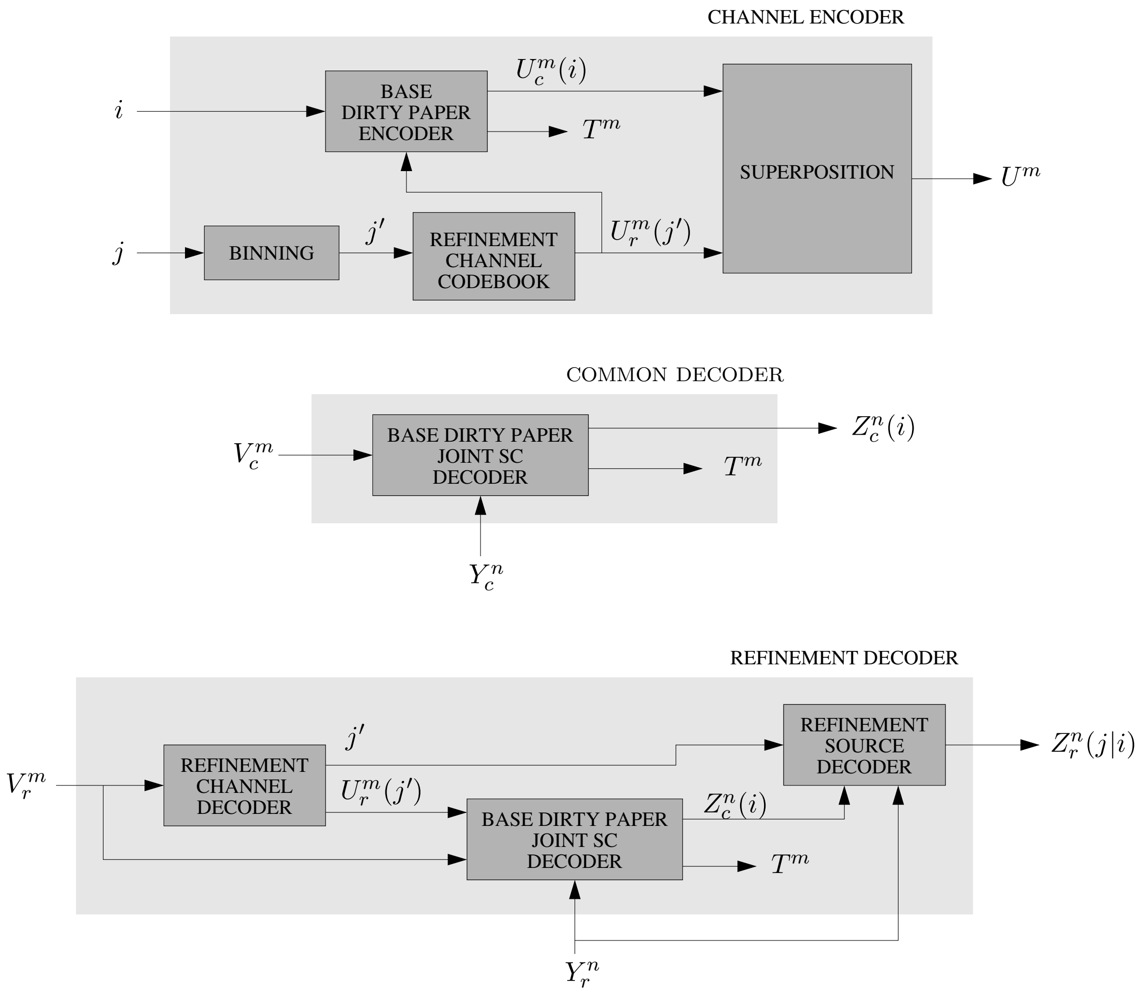

Fig. 6. Components of Scheme RC-RC.

$$C_{rr} \leq I(U_r; T, V_r) \ . \tag{50}$$

Then $\mathcal{C}_{\text{RCCR}} \subseteq \mathcal{C}_{\text{WZBC}}$.

Remark 3: In Scheme RC-CR, a trivial U_r together with $T = U$ achieves the same performance as Scheme 0.

Proof: We construct an RL codebook with elements from $\mathcal{T}_\delta^m(U_r)$. We then use the Scheme 0 with DPC construction with the chosen RL codeword acting as CSI. It follows from Theorem 2 that the CL information can be successfully decoded (together with the auxiliary codeword T^m) at both receivers if (48) and (49) are satisfied. This way, the effective communication system for transmission of RL becomes a channel with U_r^m as input and the pair T^m and V_r^m as output. For reliable transmission, (50) is then sufficient. ■

D. Scheme RC-RC

As shown in Figure 6, the encoding is performed as in Scheme RC-CR, but the decoding order is reversed. Since RL is decoded first at the r receiver, the CL codeword purely acts as noise. But the r decoder then has access to the RL codeword. So for that receiver, the CSI is also available at the decoder. The following theorem makes use of these observations.

Theorem 6: Let $\mathcal{C}_{\rm RCRC}$ be the union of all (C_{cc}, C_{cr}, C_{rr}) for which there exist $U_c \in \mathcal{U}_c$, $U_r \in \mathcal{U}_r$, and $T \in \mathcal{T}$ with $T - (U_r, U_c) - (V_r, V_c)$ and $(U_r, U_c) - U - (V_r, V_c)$ such that

$$C_{cc} \leq I(T; V_c) - I(T; U_r) \tag{51}$$

$$C_{cr} \leq I(T; V_r | U_r) \tag{52}$$

$$C_{rr} \leq I(U_r; V_r) \; . \tag{53}$$

Then $\mathcal{C}_{\rm RCRC} \subseteq \mathcal{C}_{\rm WZBC}$.

Remark 4: Similar to the Scheme RC-CR, choosing a trivial U_r and setting $T = U$ reduces this scheme to Scheme 0.

Proof: Since RL is both encoded and decoded first, (53) is necessary and sufficient for successful decoding of U_r^m. Once U_r^m is decoded, the channel between CL and receiver r reduces to one with input U_c^m, output (V_r^m, U_r^m), and CSI U_r^m. It then follows from Theorem 2 that (51) and (52) suffices for reliable transmission of Z_c^m. Note that the right-hand side of (52) is equivalent to $I(T; U_r, V_r) - I(T; U_r)$. ■

E. Comparison of Schemes

In this section, we compare the four schemes in highest generality possible. We first show that the performance of Scheme CR-CR is always superior to that of Scheme CR-RC, regardless of channel statistics. Thus, it will not be necessary to consider the latter in our performance comparisons. We also show that under the regime where $U = U_c + U_r$ where $+$ is an appropriately defined addition operation and $U_c \perp U_r$, Scheme CR-CR becomes a special case of Scheme RC-CR. Thus, for both the quadratic Gaussian and the binary Hamming cases, we need not discuss the performance of Scheme CR-CR either.

Lemma 2: $\mathcal{C}_{CRRC} \subseteq \mathcal{C}_{CRCR}$.

Proof: Let $(C_{cc}, C_{cr}, C_{rr}) \in \mathcal{C}_{CRRC}$. Then there must exist $U_c^{(1)}$, $U_r^{(1)}$, T, and U with $T - (U_c^{(1)}, U_r^{(1)}) - (V_c, V_r)$ and $(U_c^{(1)}, U_r^{(1)}) - U - (V_c, V_r)$ so that (45)-(47) are satisfied. Now define $U_c^{(2)} = U$ and let

$$\begin{aligned} C_{cc}^{(1)} &= I(U_c^{(1)}; V_c) \\ C_{cr}^{(1)} &= I(U_c^{(1)}; V_r) \\ C_{rr}^{(1)} &= I(U; V_r | U_c^{(1)}) \end{aligned}$$

and

$$\begin{aligned} C_{cc}^{(2)} &= I(U_c^{(2)}; V_c) &= I(U; V_c) \\ C_{cr}^{(2)} &= I(U_c^{(2)}; V_r) &= I(U; V_r) \\ C_{rr}^{(2)} &= I(U; V_r | U_c^{(2)}) &= 0 \; . \end{aligned}$$

By definition, both $(C_{cc}^{(1)}, C_{cr}^{(1)}, C_{rr}^{(1)})$ and $(C_{cc}^{(2)}, C_{cr}^{(2)}, C_{rr}^{(2)})$ belong to $\mathcal{C}_{\text{CRCR}}$. So does any convex combination of the two triplets. That is because if we define $Q \sim \text{Ber}(\lambda)$, so that

$$p(q, u_c, u, v_c, v_r) = p(u, v_c, v_r)p(q)p(u_c|u, q)$$

we can then write any convex combination as

$$\begin{aligned} C_{cc}^{(\lambda)} &= I(U_c^{(Q)}; V_c|Q) = I(U_c^{(Q)}, Q; V_c) \\ C_{cr}^{(\lambda)} &= I(U_c^{(Q)}; V_r|Q) = I(U_c^{(Q)}, Q; V_r) \\ C_{rr}^{(\lambda)} &= I(U; V_r|U_c^{(Q)}, Q) . \end{aligned}$$

Defining $U_c^{(\lambda)} = (U_c^{(Q)}, Q)$, one can see that $(C_{cc}^{(\lambda)}, C_{cr}^{(\lambda)}, C_{rr}^{(\lambda)}) \in \mathcal{C}_{\text{CRCR}}$.

It is clear that

$$C_{cr}^{(1)} \leq I(U_c^{(1)}; T, V_r) . \tag{54}$$

It also follows from the Markov chain $(U_c^{(1)}, T) - U - V_r$ that

$$I(U_c^{(1)}, T; V_r) \leq I(U; V_r) . \tag{55}$$

A fact which is not as obvious is

$$C_{cr}^{(1)} + C_{rr}^{(1)} \geq I(U_c^{(1)}; T, V_r) . \tag{56}$$

Towards proving (56), we observe using (55) that

$$\begin{aligned} I(U; V_r) &\geq I(U_c^{(1)}, T; V_r) \\ &= I(U_c^{(1)}; V_r|T) + I(T; V_r) \\ &= I(U_c^{(1)}; T, V_r) + I(T; V_r) - I(T; U_c^{(1)}) . \end{aligned} \tag{57}$$

But since $C_{cr}^{(1)} + C_{rr}^{(1)} = I(U; V_r)$, this yields (56) directly.

Next, we choose λ so that

$$C_{cr}^{(\lambda)} = I(U_c^{(1)}; T, V_r) .$$

That this can always be done follows from (54) and (56) together with the observation that $C_{cr}^{(1)} + C_{rr}^{(1)} = C_{cr}^{(2)}$. We then simultaneously have

$$C_{cc}^{(\lambda)} \geq C_{cc} \tag{58}$$

$$C_{cr}^{(\lambda)} \geq C_{cr} \tag{59}$$

$$C_{rr}^{(\lambda)} \geq C_{rr} . \tag{60}$$

Here, (58) follows from the fact that $C_{cc} \leq I(U_c^{(1)}; V_c) = C_{cc}^{(1)} \leq C_{cc}^{(2)}$. The fact that $C_{cr} \leq I(U_c^{(1)}; T, V_r) = C_{cr}^{(\lambda)}$ yields (59). Finally, (60) follows because

$$\begin{aligned} C_{rr}^{(\lambda)} &= I(U; V_r) - C_{cr}^{(\lambda)} \\ &= I(U; V_r) - I(U_c^{(1)}; T, V_r) \\ &\geq I(T; V_r) - I(T; U_c^{(1)}) \qquad (61) \\ &\geq C_{rr} \end{aligned}$$

where we used (57) in showing (61). ■

Lemma 3: If in Scheme CR-CR, $U = U_c + U_r$ with $U_c \perp U_r$ so that the addition operation has an inverse, i.e., $U_r = U - U_c$, then we obtain a special case of Scheme RC-CR.

Proof: Given U_c and U_r so that $U_c \perp U_r$, we can pick $T = U_c$ in Scheme RC-CR, which achieves the performance

$$\begin{aligned} C_{cc} &= I(T; V_c) - I(T; U_r) \\ &= I(U_c; V_c) \\ C_{cr} &= I(T; V_r) - I(T; U_r) \\ &= I(U_c; V_r) \\ C_{rr} &= I(U_r; T, V_r) \\ &= I(U_r; U_c) + I(U_r; V_r | U_c) \\ &= I(U_r + U_c; V_r | U_c) \\ &= I(U; V_r | U_c) \; . \end{aligned}$$

Thus, Scheme CR-CR is a special case of Scheme RC-CR. ■

V. Performance Analysis for the Quadratic Gaussian Problem

In this section, we analyze the distortion tradeoff of the proposed WZBC schemes for the quadratic Gaussian case. While Scheme 0 with DPC is developed only as a tool to be used in layered WZBC codes, Scheme 0 itself is a legitimate WZBC strategy. We thus analyze its performance in some detail first before proceeding with the layered schemes. It turns out that, somewhat surprisingly, Scheme 0 may in fact be the *optimal* strategy for an infinite family of source and channel parameters. Understanding the performance of Scheme 0 also gives insight into which receiver should be chosen as receiver c, and which one as receiver r.

A. Scheme 0

Using the test channel $X = Z + S$ with Gaussian S and Z where $S \perp Z$, and a Gaussian channel input U, (33) becomes (cf. (4))

$$\frac{1}{2}\log\left(1 - \mathbf{N}_k + \frac{\mathbf{N}_k}{\mathbf{S}}\right) \leq \frac{\kappa}{2}\log\left(1 + \frac{P}{\mathbf{W}_k}\right)$$

for $k = 1, \ldots, K$. In other words,

$$\frac{1}{\mathbf{S}} \leq 1 + \min_k \frac{\left(1 + \frac{P}{\mathbf{W}_k}\right)^\kappa - 1}{\mathbf{N}_k} .$$

Investigating (5), it is clear that $\mathbf{S}$ should be chosen so as to achieve the above inequality with equality. Substituting that choice in (5) yields

$$\frac{1}{D_k} = \frac{1}{\mathbf{N}_k} + \min_{k'} \frac{\left(1 + \frac{P}{\mathbf{W}_{k'}}\right)^\kappa - 1}{\mathbf{N}_{k'}} . \tag{62}$$

For all k^* that achieve the minimum in (62), we have

$$\frac{1}{D_{k^*}} = \frac{\left(1 + \frac{P}{\mathbf{W}_{k^*}}\right)^\kappa}{\mathbf{N}_{k^*}} .$$

Thus, as seen from (11), $D_{k^*} = D_{k^*}^{WZ}(\kappa C_{k^*})$. This, in particular, means that if

$$\frac{\left(1 + \frac{P}{\mathbf{W}_k}\right)^\kappa - 1}{\mathbf{N}_k}$$

is a constant, Scheme 0 achieves the trivial converse and there is no need for a layered WZBC scheme. Specialization of (62) to the case $\kappa = 1$ is also of interest:

$$\frac{1}{D_k} = \frac{1}{\mathbf{N}_k} + \frac{P}{\max_{k'}\left\{\mathbf{W}_{k'}\mathbf{N}_{k'}\right\}} . \tag{63}$$

In particular, all k^* maximizing $\mathbf{W}_{k^*}\mathbf{N}_{k^*}$ achieve $D_{k^*} = D_{k^*}^{WZ}(C_{k^*})$. Thus, the trivial converse is achieved if $\mathbf{W}_k\mathbf{N}_k$ is a constant.

B. Layered WZBC Schemes

We begin by analyzing the channel coding performance for Schemes RC-CR and RC-RC separately, and then the source coding performance in terms of achievable capacities. Then closely examining the capacity regions, we determine whether $c = 1, r = 2$, or $c = 2, r = 1$ is more advantageous given κ, P, $\mathbf{N}_c$, $\mathbf{N}_r$, $\mathbf{W}_c$, and $\mathbf{W}_r$. The resultant expression when $\kappa = 1$ exhibits an interesting phenomenon which we will make use of in deriving closed form expressions for the (D_c, D_r) tradeoff in both schemes.

1) Effective Capacity Regions: For Scheme RC-CR, we choose channel variables U_c and U_r as independent zero-mean Gaussians with variances νP and $\bar{\nu} P$, respectively, with $0 \leq \nu \leq 1$, and use the superposition rule $U = U_c + U_r$. Motivated by Costa's construction for the auxiliary random variable T, we set $T = \gamma U_r + U_c$. Using (48)-(50), we obtain achievable (C_{cc}, C_{cr}, C_{rr}) as

$$
\begin{aligned}
C_{cc} &= I(\gamma U_r + U_c; U_c + U_r + W_c) - I(U_r; \gamma U_r + U_c) \\
&= h(U_c + U_r + W_c) + h(U_c) - h(\gamma U_r + U_c, U_c + U_r + W_c) \\
&= \frac{1}{2} \log \frac{[P + \mathbf{W}_c]\nu P}{\det \begin{bmatrix} \gamma^2 \bar{\nu} P + \nu P & \gamma \bar{\nu} P + \nu P \\ \gamma \bar{\nu} P + \nu P & P + \mathbf{W}_c \end{bmatrix}} \\
&= \frac{1}{2} \log \frac{1 + \frac{P}{\mathbf{W}_c}}{1 + \bar{\nu} P \left(\frac{\gamma^2}{\nu P} + \frac{(1-\gamma)^2}{\mathbf{W}_c} \right)} \qquad (64) \\
C_{cr} &= I(\gamma U_r + U_c; U_c + U_r + W_r) - I(U_r; \gamma U_r + U_c) \\
&= \frac{1}{2} \log \frac{1 + \frac{P}{\mathbf{W}_r}}{1 + \bar{\nu} P \left(\frac{\gamma^2}{\nu P} + \frac{(1-\gamma)^2}{\mathbf{W}_r} \right)} \qquad (65) \\
C_{rr} &= I(U_r; \gamma U_r + U_c, U_c + U_r + W_r) \\
&= h(\gamma U_r + U_c, U_c + U_r + W_r) - h(U_c, U_c + W_r) \\
&= h(\gamma U_r + U_c, U_c + U_r + W_r) - h(U_c) - h(W_r) \\
&= \frac{1}{2} \log \frac{\det \begin{bmatrix} \gamma^2 \bar{\nu} P + \nu P & \gamma \bar{\nu} P + \nu P \\ \gamma \bar{\nu} P + \nu P & P + \mathbf{W}_r \end{bmatrix}}{\nu P \mathbf{W}_r} \\
&= \frac{1}{2} \log \left(1 + \bar{\nu} P \left(\frac{\gamma^2}{\nu P} + \frac{(1-\gamma)^2}{\mathbf{W}_r} \right) \right) . \qquad (66)
\end{aligned}
$$

Here, (65) follows by replacing W_c with W_r in (64).

For Scheme RC-RC, using the same random variables as in Scheme RC-CR, (51)-(53) translate to the achievability of

$$
\begin{aligned}
C_{cc} &= I(\gamma U_r + U_c; U_c + U_r + W_c) - I(\gamma U_r + U_c; U_r) \\
&= \frac{1}{2} \log \frac{1 + \frac{P}{\mathbf{W}_c}}{1 + \bar{\nu} P \left(\frac{\gamma^2}{\nu P} + \frac{(1-\gamma)^2}{\mathbf{W}_c} \right)} \qquad (67) \\
C_{cr} &= I(\gamma U_r + U_c; U_c + U_r + W_r | U_r) \\
&= I(U_c; U_c + W_r) \\
&= \frac{1}{2} \log \left(1 + \frac{\nu P}{\mathbf{W}_r} \right) \qquad (68) \\
C_{rr} &= I(U_r; U_c + U_r + W_r) \\
&= \frac{1}{2} \log \left(1 + \frac{\bar{\nu} P}{\nu P + \mathbf{W}_r} \right) \qquad (69)
\end{aligned}
$$

where (67) follows from (64). Since the choice of γ affects only C_{cc}, it can be picked so as to maximize C_{cc}. In fact, this choice coincides with Costa's optimal γ for the point-to-point channel between U_c and V_c, where the CSI U_r is available at the encoder [3]. In other words, the optimal choice is given by (cf. [3, Equation (7)])

$$\gamma = \frac{\nu P}{\nu P + \mathbf{W}_c}$$

yielding

$$C_{cc} = \frac{1}{2} \log \left(1 + \frac{\nu P}{\mathbf{W}_c} \right) . \tag{70}$$

2) Source Coding Performance: We choose the auxiliary random variables so that $X = Z_r + S_r$ and $Z_r = Z_c + S_c'$ where S_r and S_c' are Gaussian random variables satisfying $S_r \perp Z_r$ and $S_c' \perp Z_c$. This choice imposes the Markov chain $X - Z_r - Z_c$, and implies $X = Z_c + S_c$ with $S_c \perp Z_c$ and $1 \geq \mathbf{S}_c \geq \mathbf{S}_r$. Using (4), one can then conclude

$$R_{cc} = \frac{1}{2} \log \left(1 - \mathbf{N}_c + \frac{\mathbf{N}_c}{\mathbf{S}_c} \right) \tag{71}$$

$$R_{cr} = \frac{1}{2} \log \left(1 - \mathbf{N}_r + \frac{\mathbf{N}_r}{\mathbf{S}_c} \right) \tag{72}$$

$$R_{rr} = \frac{1}{2} \log \left(\frac{1 - \mathbf{N}_r + \frac{\mathbf{N}_r}{\mathbf{S}_r}}{1 - \mathbf{N}_r + \frac{\mathbf{N}_r}{\mathbf{S}_c}} \right) . \tag{73}$$

For any achievable triplet (C_{cc}, C_{cr}, C_{rr}), (71)-(73) can be used to find the corresponding best (D_c, D_r). More specifically, (71)-(73) and (37)-(39) together imply

$$\frac{1}{\mathbf{S}_c} \leq \min \left\{ \frac{2^{2\kappa C_{cc}} - 1}{\mathbf{N}_c}, \frac{2^{2\kappa C_{cr}} - 1}{\mathbf{N}_r} \right\} + 1 \tag{74}$$

$$\frac{1}{\mathbf{S}_r} \leq \frac{2^{2\kappa C_{rr}} \left(1 - \mathbf{N}_r + \frac{\mathbf{N}_r}{\mathbf{S}_c} \right) - 1}{\mathbf{N}_r} + 1 . \tag{75}$$

Since we have from (5) that

$$D_k = \frac{\mathbf{N}_k}{1 - \mathbf{N}_k + \frac{\mathbf{N}_k}{\mathbf{S}_k}} \tag{76}$$

it is easy to conclude that both (74) and (75) should be satisfied with equality to obtain the best (D_c, D_r), which becomes

$$D_c = \frac{\mathbf{N}_c}{1 + \mathbf{N}_c \phi} \tag{77}$$

$$D_r = \frac{\mathbf{N}_r}{1 + \mathbf{N}_r \phi} 2^{-2\kappa C_{rr}} \tag{78}$$

where

$$\phi = \min \left\{ \frac{2^{2\kappa C_{cc}} - 1}{\mathbf{N}_c}, \frac{2^{2\kappa C_{cr}} - 1}{\mathbf{N}_r} \right\} . \tag{79}$$

Now, if

$$\frac{2^{2\kappa C_{cc}} - 1}{\mathbf{N}_c} \geq \frac{2^{2\kappa C_{cr}} - 1}{\mathbf{N}_r} \tag{80}$$

then $D_r = \mathbf{N}_r 2^{-2\kappa(C_{cr}+C_{rr})}$. But in both Schemes RC-CR and RC-RC, we have $C_{cr}+C_{rr} = C_r = \frac{1}{2}\log\left(1+\frac{P}{W_r}\right)$, implying $D_r = D_r^{WZ}(\kappa C_r)$, regardless of the chosen parameters. Thus, it suffices to consider only

$$\frac{2^{2\kappa C_{cc}}-1}{\mathbf{N}_c} \leq \frac{2^{2\kappa C_{cr}}-1}{\mathbf{N}_r} \tag{81}$$

because equality in (81) already gives $D_r = D_r^{WZ}(\kappa C_r)$. We thus have

$$D_c = \mathbf{N}_c 2^{-2\kappa C_{cc}} \tag{82}$$

$$\begin{aligned} D_r &= \frac{\mathbf{N}_r}{1+\frac{\mathbf{N}_r}{\mathbf{N}_c}[2^{2\kappa C_{cc}}-1]} 2^{-2\kappa C_{rr}} \\ &= \frac{\mathbf{N}_r}{1+\mathbf{N}_r\left[\frac{1}{D_c}-\frac{1}{\mathbf{N}_c}\right]} 2^{-2\kappa C_{rr}} . \end{aligned} \tag{83}$$

3) Choosing the Refinement Receiver: Note that setting $\nu = 1$ reduces both Scheme RC-CR and Scheme RC-RC to Scheme 0. This is regardless of which receiver is designated as c or r. This simple observation, along with the discussion in Section V-A, leads to the following lemma.

Lemma 4: In order to maximize the performance of any of the layered WZBC schemes, one must set c and r so that

$$\frac{\left(1+\frac{P}{\mathbf{W}_c}\right)^{\kappa}-1}{\mathbf{N}_c} \leq \frac{\left(1+\frac{P}{\mathbf{W}_r}\right)^{\kappa}-1}{\mathbf{N}_r} . \tag{84}$$

Remark 5: When $\kappa = 1$, (84) translates to

$$\mathbf{W}_c\mathbf{N}_c \geq \mathbf{W}_r\mathbf{N}_r . \tag{85}$$

Therefore, the product $\mathbf{W}_k\mathbf{N}_k$ determines the *combined* channel and side information quality, so that the "better" receiver is chosen to receive the RL information. Recall from the discussion in Section V-A that if $\mathbf{W}_k\mathbf{N}_k$ is constant, then in fact there is no need for refinement, as Scheme 0 already achieves the optimal performance.

Proof: When $\nu = 1$, i.e., when all the power is allocated to the CL, the layered schemes uniformly achieve the Scheme 0 performance. In particular, they achieve the capacity point

$$\begin{aligned} C_{cc} &= C_c = \frac{1}{2}\log\left(1+\frac{P}{\mathbf{W}_c}\right) \\ C_{cr} &= C_r = \frac{1}{2}\log\left(1+\frac{P}{\mathbf{W}_r}\right) \\ C_{rr} &= 0 . \end{aligned}$$

If (84) does not hold, then from (62), it follows that the schemes also achieve $D_r = D_r^{WZ}(\kappa C_r)$ and some $D_c > D_c^{WZ}(\kappa C_c)$. Now, if we set $\nu < 1$, it is obvious that D_r cannot be lowered any further. We claim that D_c cannot be lowered either. Therefore, none of the layered schemes would be able to achieve a better (D_c, D_r) than what Scheme 0 achieves. On the other hand, sending the refinement to receiver c could potentially result in a better performance than that.

Towards proving the above claim, observe from (79) that it suffices to show that neither C_{cc} nor C_{cr} can increase when $\nu < 1$ compared to the case $\nu = 1$. That, in turn, follows by closely examining the expressions for C_{cc} and C_{cr} in Section V-B1. In particular, expressions in (68) and (70) can obviously never increase, and for Scheme RC-CR, both (64) and (65) will be maximized by their corresponding optimal Costa parameters, i.e., by $\gamma = \frac{\nu P}{\nu P + \mathbf{W}_c}$ and by $\gamma = \frac{\nu P}{\nu P + \mathbf{W}_r}$, respectively. This results in $C_{cc} = \frac{1}{2}\log\left(1 + \frac{\nu P}{\mathbf{W}_c}\right)$ and $C_{cr} = \frac{1}{2}\log\left(1 + \frac{\nu P}{\mathbf{W}_r}\right)$ as the maximum possible values, which are strictly smaller than C_c and C_r, respectively. Therefore, the proof is complete. ■

C. Performance Comparisons for $\kappa = 1$

We first derive closed-form (D_c, D_r) tradeoffs for both Scheme RC-CR and RC-RC, and show that the former is superior to the latter. Thus, for the quadratic Gaussian problem, Scheme RC-CR prevails as the best among all the layered WZBC schemes we proposed.

Lemma 5: A distortion pair (D_c, D_r) is achievable using Scheme RC-CR if and only if $D_r \geq D_{\text{RCCR}}(D_c)$, where $D_{\text{RCCR}}(D_c)$ is the convex hull of

$$D^*_{\text{RCCR}}(D_c) = \frac{\mathbf{N}_r \mathbf{N}_c^2}{D_c \mathbf{N}_c + \mathbf{N}_r(\mathbf{N}_c - D_c)} \cdot \begin{cases} \frac{\mathbf{W}_r D_c}{(\mathbf{W}_r - \mathbf{W}_c)\mathbf{N}_c + (P + \mathbf{W}_c) D_c} & \mathbf{W}_c \geq \mathbf{W}_r \\ \frac{\mathbf{W}_c}{P + \mathbf{W}_c} & \mathbf{W}_c < \mathbf{W}_r \end{cases} \tag{86}$$

for

$$\frac{\mathbf{N}_c \mathbf{W}_c}{P + \mathbf{W}_c} \leq D_c \leq D_c^{\max}$$

with

$$D_c^{\max} = \mathbf{N}_c \cdot \begin{cases} \min\left\{1, \frac{\mathbf{N}_r(\mathbf{W}_c - \mathbf{W}_r)}{(P + \mathbf{W}_c)(\mathbf{N}_r - \mathbf{N}_c)}\right\} & \mathbf{N}_c < \mathbf{N}_r, \mathbf{W}_c > \mathbf{W}_r \\ 1 & \mathbf{N}_c \geq \mathbf{N}_r, \mathbf{W}_c \geq \mathbf{W}_r \\ \frac{\mathbf{W}_c}{P + \mathbf{W}_c} + \frac{P(\mathbf{W}_c \mathbf{N}_c - \mathbf{W}_r \mathbf{N}_r)}{(P + \mathbf{W}_c)(\mathbf{N}_c - \mathbf{N}_r)\mathbf{W}_r} & \mathbf{N}_c > \mathbf{N}_r, \mathbf{W}_c < \mathbf{W}_r \end{cases} . \tag{87}$$

Remark 6: The cases $\mathbf{N}_c \leq \mathbf{N}_r, \mathbf{W}_c < \mathbf{W}_r$ and $\mathbf{N}_c < \mathbf{N}_r, \mathbf{W}_c = \mathbf{W}_r$ are not considered in (87) because they are prohibited by the rule (85). The same rule also guarantees $\frac{\mathbf{N}_c \mathbf{W}_c}{P + \mathbf{W}_c} \leq D_c^{\max} \leq \mathbf{N}_c$.

As a byproduct of the proof, which is deferred to Appendix B, we observe that the Costa parameter γ is either 0 or 1, depending on whether $\mathbf{W}_c \geq \mathbf{W}_r$ or $\mathbf{W}_c < \mathbf{W}_r$, respectively. When it is 0, we have $T = U_c$, and as we argued in the proof of Lemma 3, this reduces Scheme RC-CR to Scheme CR-CR. On the other hand, when $\gamma = 1$, we have $T = U = U_c + U_r$. Thus, setting the auxiliary codeword T^m to be the same as the channel input U^m constitutes the optimal choice. To the best of our knowledge, this choice, which is typically encountered in DPC for binary symmetric channels, was never the optimal choice involving Gaussian channels.

Lemma 6: A distortion pair (D_c, D_r) is achievable using Scheme RC-RC if and only if $D_r \geq D_{\text{RCRC}}(D_c)$, where $D_{\text{RCRC}}(D_c)$ is the convex hull of

$$D^*_{\text{RCRC}}(D_c) = \frac{\mathbf{N}_r \mathbf{W}_r}{P + \mathbf{W}_r} \cdot \frac{D_c \mathbf{N}_c + \frac{\mathbf{N}_c \mathbf{W}_c}{\mathbf{W}_r}(\mathbf{N}_c - D_c)}{D_c \mathbf{N}_c + \mathbf{N}_r(\mathbf{N}_c - D_c)} \tag{88}$$

for

$$\frac{\mathbf{N}_c \mathbf{W}_c}{P + \mathbf{W}_c} \leq D_c \leq \mathbf{N}_c \, .$$

Proof: It follows from (82), (83), and (109) that

$$D_c = \frac{\mathbf{N}_c \mathbf{W}_c}{\nu P + \mathbf{W}_c} \tag{89}$$

$$D_r = \frac{\mathbf{N}_r}{1 + \mathbf{N}_r \left[\frac{1}{D_c} - \frac{1}{\mathbf{N}_c}\right]} \cdot \frac{\nu P + \mathbf{W}_r}{P + \mathbf{W}_r} \tag{90}$$

provided

$$\frac{\nu P}{\mathbf{N}_c \mathbf{W}_c} \leq \frac{\nu P}{\mathbf{N}_r \mathbf{W}_r} .$$

Therefore, because of (85), (89) and (90) are always valid. Solving for ν in (89) and substituting it in (90) gives the desired result. ■

Lemma 7: For the quadratic Gaussian problem with $\kappa = 1$, the performance of Scheme RC-CR is superior to that of RC-RC.

Proof: See Appendix C. ■

We next compare Scheme RC-CR, our best scheme, to separate coding and uncoded transmission, case by case.

1) It is obvious by comparing (23) and (86) that when $\mathbf{W}_c \geq \mathbf{W}_r$ and $\mathbf{N}_c \geq \mathbf{N}_r$, Scheme RC-CR obtains the exact same performance as in separate source and channel coding (Note that $r = g, c = b$ in this case). This was expected because in this case, Scheme RC-CR is identical to Scheme CR-CR, and according to (42) and (43), the quality of the CL information is limited by the quality of the worse receiver, as in separate coding. This behavior is displayed in Figures 7(d) and (e).

   As for uncoded transmission, it can be better than the digital schemes. For example, consider the case $\mathbf{N}_c = \mathbf{N}_r = 1$ depicted in Figure 7(e), which corresponds to useless side information at both receivers. In this case, uncoded transmission actually achieves the trivial converse, and therefore, is the optimal strategy.

2) When $\mathbf{W}_c > \mathbf{W}_r$ and $\mathbf{N}_c < \mathbf{N}_r$, it follows from (24) and (86) that a sufficient condition for superiority of Scheme RC-CR over separate coding is given by

$$\frac{\mathbf{N}_r \mathbf{W}_r D_c}{(\mathbf{W}_r - \mathbf{W}_c)\mathbf{N}_c + (P + \mathbf{W}_c) D_c} \geq \frac{\mathbf{N}_r \mathbf{N}_c^2 \mathbf{W}_r D_c}{\Big(D_c \mathbf{N}_c + \mathbf{N}_r(\mathbf{N}_c - D_c)\Big)\Big((\mathbf{W}_r - \mathbf{W}_c)\mathbf{N}_c + (P + \mathbf{W}_c) D_c\Big)}$$

   which simplifies to

$$1 \geq \frac{\mathbf{N}_c^2}{D_c \mathbf{N}_c + \mathbf{N}_r(\mathbf{N}_c - D_c)}$$

   and is therefore granted since $\mathbf{N}_c < \mathbf{N}_r$. Moreover, equality is satisfied, i.e., the two schemes have equal performance, only when $D_c = D_c^{\max} = \mathbf{N}_c$. This behavior is exemplified in Figures 7(b) and (c). The difference between the two examples is that $D_c^{\max} = \mathbf{N}_c$ in (b), whereas $D_c^{\max} < \mathbf{N}_c$ in (c).

   Even though $\mathbf{N}_c = \mathbf{N}_r = 1$ is prohibited in this case, one can consider $\mathbf{N}_c = 1 - \epsilon$ and $\mathbf{N}_r = 1$ with arbitrarily small $\epsilon > 0$. Uncoded transmission is also superior to all the digital schemes in this limiting case also.

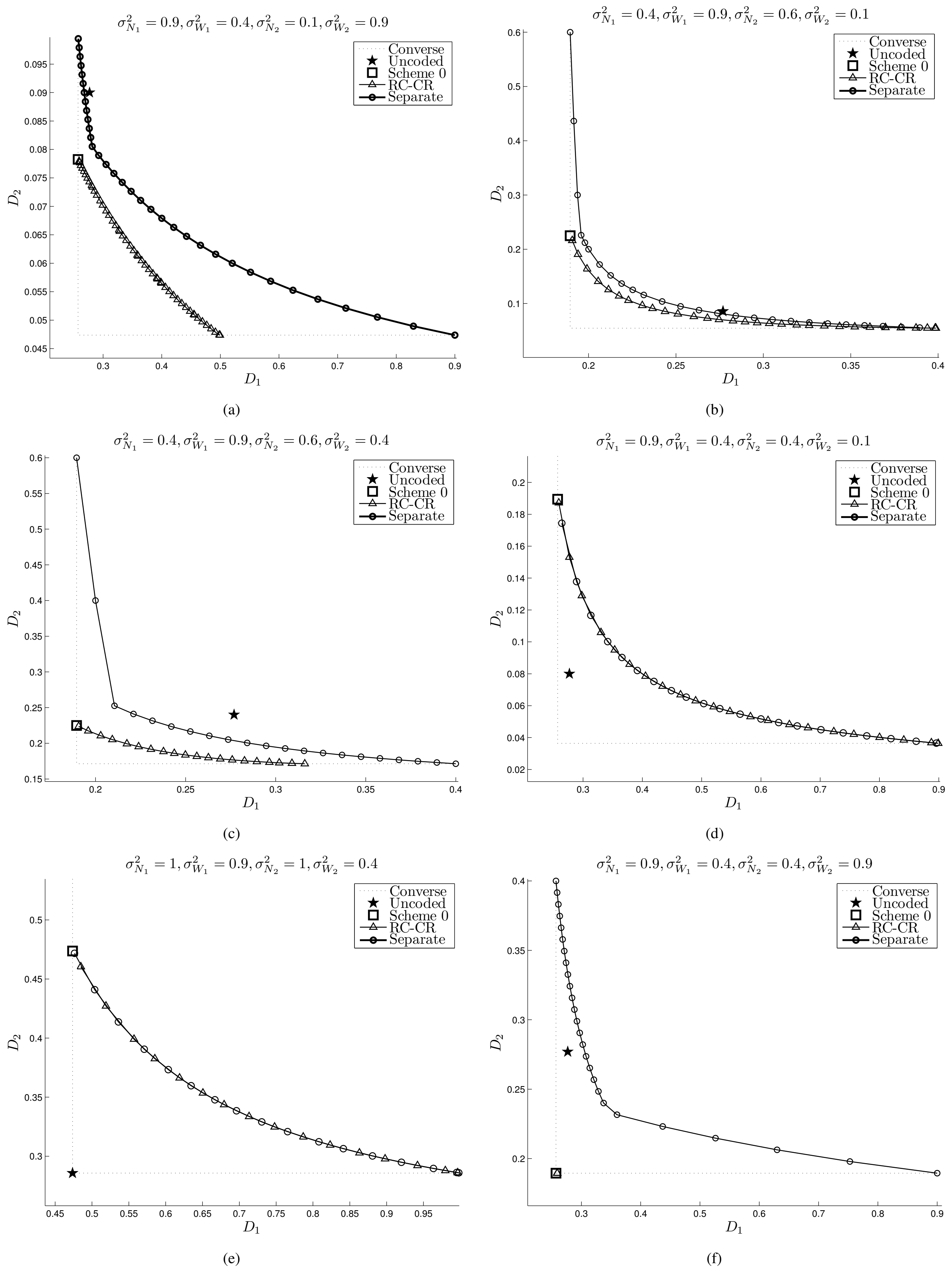

Fig. 7. Performance comparison for Gaussian sources and channels. In (a)-(e), $\text{var}(N_1)\text{var}(W_1) > \text{var}(N_2)\text{var}(W_2)$, and therefore the choice $c = 1$, $r = 2$ is made. In addition, in (e), $\text{var}(N_1) = \text{var}(N_2) = 1$, implying that there is no side information at either receiver and hence uncoded transmission is optimal. In (f), $\text{var}(N_1)\text{var}(W_1) = \text{var}(N_2)\text{var}(W_2)$ making Scheme 0 optimal.

3) Finally, when $\mathbf{W}_c < \mathbf{W}_r$ and $\mathbf{N}_c > \mathbf{N}_r$, since $r = b, c = g$ in this case, we need to explicitly write the best D_c for a given D_r for Scheme RC-CR. From (86), it follows that Scheme RC-CR can achieve

$$D_c = \frac{\mathbf{N}_c\mathbf{N}_r}{\mathbf{N}_c - \mathbf{N}_r}\left[\frac{\mathbf{N}_c\mathbf{W}_c}{(P+\mathbf{W}_c)D_r} - 1\right] \tag{91}$$

for $D_r^{WZ}(C_r) \leq D_r \leq \frac{\mathbf{N}_c\mathbf{N}_r\mathbf{W}_c}{\mathbf{N}_c\mathbf{W}_c + P\mathbf{N}_r}$. On the other hand, (24) implies that the minimum D_c that can be achieved by separate coding must necessarily satisfy

$$\begin{aligned} D_c &\geq \frac{\mathbf{N}_c\mathbf{N}_r\Big(\mathbf{N}_c\mathbf{W}_c - (P+\mathbf{W}_r)D_r - \mathbf{N}_r(\mathbf{W}_c - \mathbf{W}_r)\Big)}{\Big((\mathbf{W}_c - \mathbf{W}_r)\mathbf{N}_r + (P+\mathbf{W}_r)D_r\Big)(\mathbf{N}_c - \mathbf{N}_r)} \\ &= \frac{\mathbf{N}_c\mathbf{N}_r}{\mathbf{N}_c - \mathbf{N}_r}\left[\frac{\mathbf{N}_c\mathbf{W}_c}{(\mathbf{W}_c - \mathbf{W}_r)\mathbf{N}_r + (P+\mathbf{W}_c)D_r} - 1\right] . \end{aligned} \tag{92}$$

Superiority of Scheme RC-CR over separate coding then easily follows from (91) and (92). An example of this case is shown in Figure 7(a).

We next show that Scheme RC-CR always outperforms uncoded transmission in this case. In fact, uncoded transmission is even worse than Scheme 0. Since Scheme 0 achieves $D_r = D_r^{WZ}(C_r)$, it suffices to compare the D_r values. Comparing (31) and (63), this reduces to showing

$$\frac{\mathbf{N}_r\mathbf{W}_r}{\mathbf{W}_r + \mathbf{N}_r P} \geq \frac{\mathbf{N}_r\mathbf{N}_c\mathbf{W}_c}{\mathbf{N}_c\mathbf{W}_c + \mathbf{N}_r P}$$

or equivalently

$$\mathbf{W}_r \geq \mathbf{N}_c\mathbf{W}_c .$$

But since $\mathbf{W}_r > \mathbf{W}_c$, this is trivially true.

In Figure 7(f), we also include an example where $\mathbf{N}_c\mathbf{W}_c = \mathbf{N}_r\mathbf{W}_r$, i.e., where the combined channel and side information qualities are the same. Scheme 0 achieves the trivial converse as discussed in Section V-A. We also observed that uncoded transmission may achieve a distortion pair below the best known digital tradeoff, as shown in Figures 7(d) and (e). This was expected because it is well-known that the optimal scheme is uncoded transmission when there is no side information at either receiver, as is the case in Figure 7(e). For cases other than $\mathbf{N}_c\mathbf{W}_c = \mathbf{N}_r\mathbf{W}_r$, one could roughly say that the proposed digital schemes are better than uncoded transmission when the quality of the side information is sufficiently high, although we do not currently have the analytical means for comparison.

VI. Performance Analysis for the Binary Hamming Problem

In this section, we first analyze Scheme 0 for the binary Hamming problem and show that it can be optimal in this case as well. We then analyze the layered WZBC schemes and compare all the schemes numerically.

A. *Scheme 0*

It follows from Theorem 1 and Equations (7) and (8) that in the binary Hamming case, if there exists $0 \leq q \leq 1$ and $0 \leq \alpha \leq \frac{1}{2}$ such that

$$qr(\alpha, \beta_k) \leq \kappa[1 - H_2(p_k)] \tag{93}$$

for all k, then

$$D_k = (1 - q)\beta_k + q \min\{\alpha, \beta_k\} \tag{94}$$

can be achieved by Scheme 0. Unlike in the quadratic Gaussian case, the constraint (93) does not result in a single best value for q and α. Therefore, Scheme 0 produces a tradeoff of D_k's rather than one best point.

As discussed at the end of Section II-A, the distortion-rate function $D_k^{WZ}(R)$ is achieved either by $q = 1$ and $\alpha \leq \alpha_0(\beta_k)$, or by $0 \leq q < 1$ and $\alpha = \alpha_0(\beta_k)$. The implication of this fact regarding Scheme 0 is the following:

1) If β_k are not identical, neither are $\alpha_0(\beta_k)$, and thus we need $q = 1$ and some $\alpha \leq \min_k \alpha_0(\beta_k)$ to attain all $D_k^{WZ}(\kappa C_k)$ simultaneously, i.e.,

$$r(\alpha, \beta_k) = \kappa[1 - H_2(p_k)] \tag{95}$$

for all k. When this happens, we must necessarily have $D_k = \alpha$ i.e., D_k does not depend on k.

2) If $\beta_k = \beta$ for $k = 1, \ldots, K$, and thus $D_k^{WZ}(R)$ does not depend on k, we need $C_k = C$ (and hence $p_k = p$) so that the same test channel (q, α) achieves $D_k^{WZ}(C_k)$ simultaneously. But, this makes the problem trivial.

B. *Layered WZBC Schemes*

To evaluate R_{cc}, R_{cr} and R_{rr}, we first fix Z_c and Z_r as in Section II-C2 where subscripts g and b are to be replaced by r and c, respectively. Since we only analyze the Markov relation $X - Z_r - Z_c$, it suffices to analyze the case $q_c \leq q_r$ and $\alpha_c \geq \alpha_r$. This results in

$$\begin{aligned} R_{cc} &= q_c r(\alpha_c, \beta_c) \\ R_{cr} &= q_c r(\alpha_c, \beta_r) \\ R_{rr} &= q_r r(\alpha_r, \beta_r) - q_c r(\alpha_c, \beta_r) \; . \end{aligned}$$

We next make channel variable choices and derive the resulting channel coding rates for each scheme individually. Unlike in the quadratic Gaussian case, there is no power allocation parameter to vary. However, we have freedom in choosing the distributions of U_c and U_r as $\mathrm{Ber}(\gamma_c)$ and $\mathrm{Ber}(\gamma_r)$, respectively, as well as in choosing the auxiliary random variable as either $T = U_c$ (which reduces Scheme RC-CR to CR-CR) or $T = U_c \oplus U_r$.

1) Scheme RC-CR: In this case, with $T = U_c$, (45)-(47) become

$$\begin{aligned} C_{cc} &= I(U_c; U_c \oplus U_r \oplus W_c) \\ &= r(\gamma_r \star p_c, \gamma_c) \\ C_{cr} &= I(U_c; U_c \oplus U_r \oplus W_r) \\ &= r(\gamma_r \star p_r, \gamma_c) \\ C_{rr} &= I(U_c \oplus U_r; U_c \oplus U_r \oplus W_r | U_c) \\ &= I(U_r; U_r \oplus W_r) \\ &= r(p_r, \gamma_r) \ . \end{aligned} \tag{96}$$

But since $r(\cdot, \cdot)$ is increasing in its second argument, we have $\gamma_c = \frac{1}{2}$ as the optimal value achieving

$$C_{cc} = 1 - H_2(\gamma_r \star p_c) \tag{97}$$

$$C_{cr} = 1 - H_2(\gamma_r \star p_r) \ . \tag{98}$$

On the other hand, if $T = U_c \oplus U_r$, we obtain

$$\begin{aligned} C_{cc} &= I(U_c \oplus U_r; U_c \oplus U_r \oplus W_c) - I(U_r; U_c \oplus U_r) \\ &= r(p_c, \gamma_c \star \gamma_r) - r(\gamma_c, \gamma_r) \end{aligned} \tag{99}$$

$$\begin{aligned} C_{cr} &= I(U_c \oplus U_r; U_c \oplus U_r \oplus W_r) - I(U_r; U_c \oplus U_r) \\ &= r(p_r, \gamma_c \star \gamma_r) - r(\gamma_c, \gamma_r) \end{aligned} \tag{100}$$

$$\begin{aligned} C_{rr} &= I(U_r; U_c \oplus U_r, U_c \oplus U_r \oplus W_r) \\ &= I(U_r; U_c \oplus U_r) \\ &= r(\gamma_c, \gamma_r) \ . \end{aligned} \tag{101}$$

2) Scheme RC-RC: Making the same choices as in Scheme RC-CR, it follows from (45)-(47) that when $T = U_c$,

$$\begin{aligned} C_{cc} &= I(U_c; U_c \oplus U_r \oplus W_c) \\ &= r(p_c \star \gamma_r, \gamma_c) \end{aligned} \tag{102}$$

$$\begin{aligned} C_{cr} &= I(U_c; U_c \oplus U_r \oplus W_r | U_r) \\ &= I(U_c; U_c \oplus W_r) \\ &= r(p_r, \gamma_c) \end{aligned} \tag{103}$$

$$\begin{aligned} C_{rr} &= I(U_r; U_c \oplus U_r \oplus W_r) \\ &= r(p_r \star \gamma_c, \gamma_r) \end{aligned} \tag{104}$$

and when $T = U_c \oplus U_r$, (102) is replaced by

$$\begin{aligned} C_{cc} &= I(U_c \oplus U_r; U_c \oplus U_r \oplus W_c) - I(U_r; U_c \oplus U_r) \\ &= r(p_c, \gamma_c \star \gamma_r) - r(\gamma_c, \gamma_r) \; . \end{aligned} \tag{105}$$

After some algebra, the difference between the right-hand sides of (102) and (105) can be written as

$$r(p_c \star \gamma_r, \gamma_c) - [r(p_c, \gamma_c \star \gamma_r) - r(\gamma_c, \gamma_r)] = r(\gamma_c, \gamma_r) - r(p_c, \gamma_r) \; .$$

Recalling the fact that $r(\cdot, \cdot)$ is decreasing in its first argument, we see that if $\gamma_c \geq p_c$, one should use $T = U_c \oplus U_r$, and otherwise, use $T = U_c$.

C. Performance Comparisons for $\kappa = 1$.

Analytical performance comparisons prove more difficult for the binary Hamming problem. Even which receiver should be designated as c and which as r is not straightforward to decide. That is because (i) there is no power allocation parameter we can control, and (ii) even Scheme 0 can produce a curve which could achieve both $D_c = D_c^{WZ}(\kappa C_c)$ and $D_r = D_r^{WZ}(\kappa C_r)$, rather than a single best point.

It is also not clear that our choice of source random variables are the best. As mentioned earlier, our main motivation in adopting the same test channel as in point-to-point coding for all digital schemes is its simplicity. The alphabet size bounds in [14], [16], however, are much higher and therefore it might be possible to further improve the performance of all the tested digital schemes.

The performance of the various schemes for certain source-channel pairs at rate $\kappa = 1$ is presented in Figure 8. For each of the new schemes, convex hull of two curves is shown, where in one $c = 2, r = 1$ and in the other $c = 1, r = 2$. In all our examples, Scheme RC-CR is at least as good as Scheme RC-RC and separate source and channel coding. In Figures 8(a)-(d), the parameters β_1, β_2, and p_1 are fixed so that (95) is satisfied for $k = 1$, and p_2 is varying. As p_2 increases, the collective behavior of the schemes dramatically changes. In Figure 8(a), $c = 1, r = 2$ is consistently the best choice among all schemes. As the quality of the second channel decreases, and reaches the point where (95) is also satisfied for $k = 2$, Scheme 0 becomes optimal, as shown in Figure 8(b). When p_2 is increased even further, as in Figure 8(c), $c = 2, r = 1$ becomes the better choice. When p_2 reaches the point where the first receiver has access to both the better channel and the better side information, as in Figures 8(d) and (e), separate coding and Scheme RC-CR become identical as in the quadratic Gaussian case. However, uncoded transmission can still outperform all the digital schemes as shown in Figure 8(e) for the case of trivial side information. Finally, Figure 8(f) exemplifies the interesting phenomenon mentioned above, where Scheme 0 produces a curve, rather than a point, which happens to be the best along with the layered WZBC schemes.

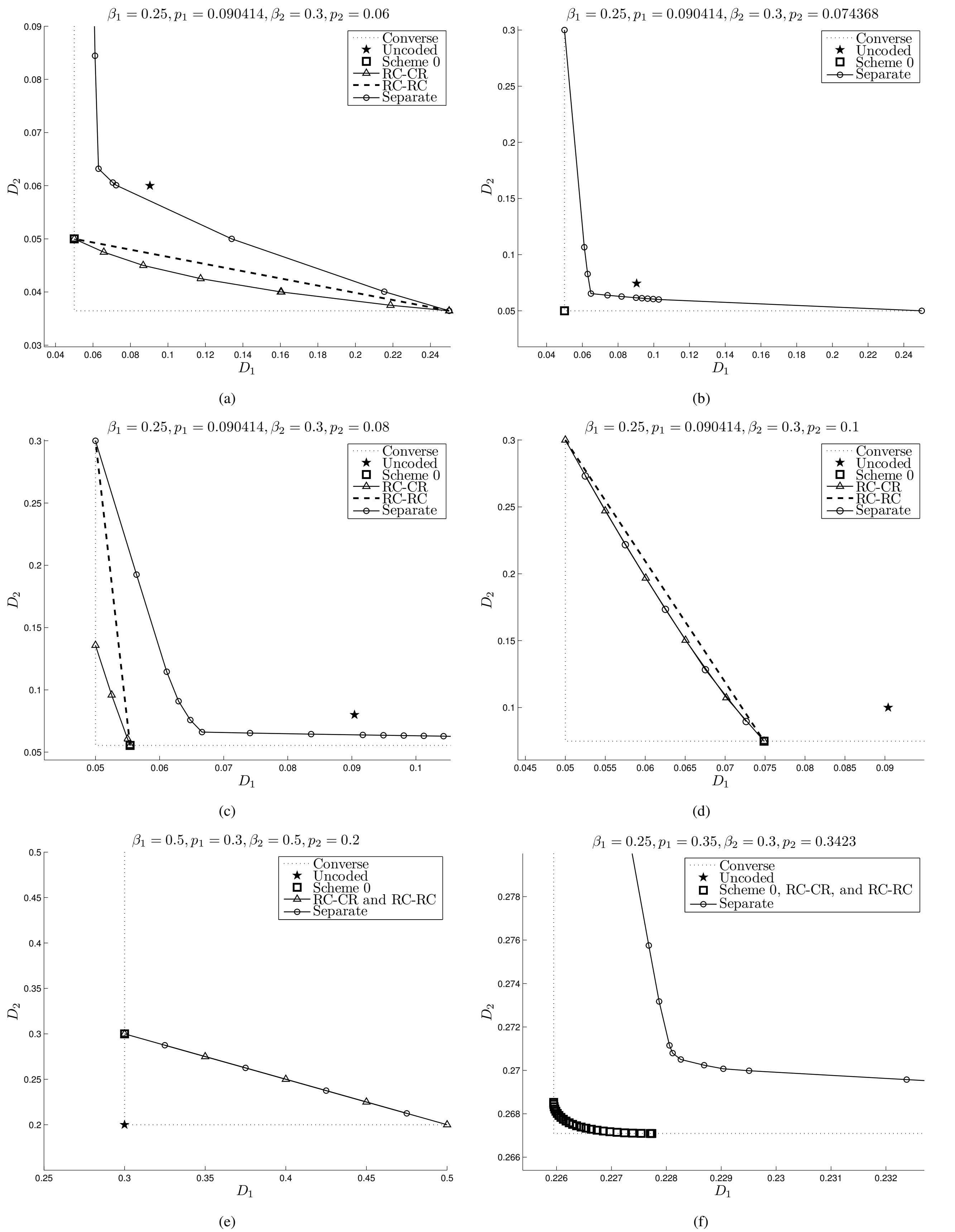

Fig. 8. Performance comparison for binary sources and channels. In (a)-(d), β_1, β_2, and p_1 are fixed, and as p_2 increases, how all the schemes compare changes. In (e), uncoded transmission is optimal. In (f), Scheme 0, along with all the other layered schemes, is the best. It is also noteworthy to observe that it touches both trivial converse bounds simultaneously.

VII. Conclusions and Future Work

We proposed coding schemes for the WZBC problem, and analyzed their distortion performance for the quadratic Gaussian and binary Hamming cases. Even though our schemes are for a general rate κ channel uses per source symbol, the achievability regions are easiest to compute for $\kappa = 1$. In fact, for the quadratic Gaussian case, we were able to derive closed form expressions for the entire distortion tradeoff and show that our best scheme, namely Scheme RC-CR, is always at least as good as (in fact, except for one certain case, always better than) separate coding. By numerical comparisons, we observed the same phenomenon for the binary Hamming case under the regime where all the test channels are constrained to be of the form which achieves the Wyner-Ziv rate-distortion function. On the other hand, even Scheme RC-CR may not achieve the same performance as uncoded transmission. This is not surprising, since when there is no (or trivial) side information, it is known that uncoded transmission is optimal.

In an upcoming paper, we combine the digital schemes we proposed with uncoded transmission to extract the benefits of both methods. In fact, as we show in a preliminary version [8], the hybrid scheme is more than the sum of its parts and distortions outside the convexification of the digital and analog regions are achievable.

Appendix

A. *Proof of Lemma 1*

The key to the proof is the observation that for optimal performance, (20) needs to be satisfied with equality for any κ. To see this, assume that (D_b, D_g) with $D_b < \mathbf{N}_b$ satisfies (20) with strict inequality for some $0 < \nu \leq 1$. Then one can decrease ν until equality is obtained in (20), and still satisfy (21) or (22), depending on whether $X - Y_g - Y_b$ or $X - Y_b - Y_g$, respectively. That, in turn, follows because the right-hand side of either of (21) or (22) are decreasing in ν.

When $\kappa = 1$, equality in (20) translates to

$$\bar{\nu} P = \frac{D_b(P + \mathbf{W}_b)}{\mathbf{N}_b} - \mathbf{W}_b \ .$$

For the case $X - Y_g - Y_b$, (21) then becomes

$$D_g \geq \frac{\mathbf{N}_g \mathbf{N}_b^2 \mathbf{W}_g D_b}{\Big(D_b \mathbf{N}_b + \mathbf{N}_g(\mathbf{N}_b - D_b)\Big)\Big((\mathbf{W}_g - \mathbf{W}_b)\mathbf{N}_b + (P + \mathbf{W}_b)D_b\Big)} \ .$$

If $X - Y_b - Y_g$, on the other hand, (22) implies

$$D_g \geq \frac{\mathbf{N}_g \mathbf{W}_g D_b}{\Big((\mathbf{W}_g - \mathbf{W}_b)\mathbf{N}_b + (P + \mathbf{W}_b)D_b\Big)}$$

and

$$D_g \geq \frac{\mathbf{N}_b \mathbf{N}_g \Big(\mathbf{N}_g \mathbf{W}_g - (\mathbf{W}_g - \mathbf{W}_b)\mathbf{N}_b - (P + \mathbf{W}_b)D_b\Big)}{\Big((\mathbf{W}_g - \mathbf{W}_b)\mathbf{N}_b + (P + \mathbf{W}_b)D_b\Big)\,(\mathbf{N}_g - \mathbf{N}_b)}$$

simultaneously, which is the desired result.

B. Proof of Lemma 5

It follows from (82) and (83) that by varying ν and γ, we obtain the tradeoff

$$D_c = \mathbf{N}_c \frac{Pa(\nu,\gamma) + \mathbf{W}_c}{P + \mathbf{W}_c} \tag{106}$$

$$D_r = \frac{\mathbf{N}_r}{1 + \mathbf{N}_r\left[\frac{1}{D_c} - \frac{1}{\mathbf{N}_c}\right]} \cdot \frac{1}{1 + \frac{Pb(\nu,\gamma)}{\mathbf{W}_r}} \tag{107}$$

where

$$a(\nu,\gamma) = \bar{\nu}\left(\frac{\mathbf{W}_c}{\nu P}\gamma^2 + (1-\gamma)^2\right)$$
$$b(\nu,\gamma) = \bar{\nu}\left(\frac{\mathbf{W}_r}{\nu P}\gamma^2 + (1-\gamma)^2\right) .$$

We next fix D_c, which, in turn, fixes $a(\nu,\gamma)$ as

$$a(\nu,\gamma) = \frac{D_c[P + \mathbf{W}_c] - \mathbf{W}_c\mathbf{N}_c}{\mathbf{N}_c P} \tag{108}$$

and minimize D_r, which reduces to maximizing $b(\nu,\gamma)$. Since neither C_{cc} nor C_{cr} can be negative, we need both $a(\nu,\gamma) \leq 1$ and $b(\nu,\gamma) \leq 1$ to be satisfied. The former requirement is guaranteed because we naturally limit ourselves to $D_c \leq \mathbf{N}_c$. The latter, on the other hand, becomes vacuous since rewriting (81) gives

$$b(\nu,\gamma) \leq \frac{\mathbf{N}_c[Pa(\nu,\gamma) + \mathbf{W}_c] - \mathbf{N}_r\mathbf{W}_r[1 - a(\nu,\gamma)]}{\mathbf{N}_c[Pa(\nu,\gamma) + \mathbf{W}_c] + P\mathbf{N}_r[1 - a(\nu,\gamma)]} \tag{109}$$

whose right-hand side is always less than or equal to 1.

Now if $\mathbf{W}_c \geq \mathbf{W}_r$, we always have $a(\nu,\gamma) \geq b(\nu,\gamma)$ since

$$b(\nu,\gamma) = a(\nu,\gamma) - \frac{\bar{\nu}\gamma^2}{\nu P}[\mathbf{W}_c - \mathbf{W}_r] .$$

Thus, among all choices of γ and ν which satisfy (108), the one that potentially minimizes D_r is $\gamma = 0$ and

$$\nu = \left(1 - \frac{D_c}{\mathbf{N}_c}\right)\left(1 + \frac{\mathbf{W}_c}{P}\right) .$$

That is because with this choice we have $b(\nu,\gamma) = a(\nu,\gamma)$. It then remains to check (109), which can be written after some algebra as

$$\mathbf{N}_c\mathbf{N}_r[\mathbf{W}_c - \mathbf{W}_r] \geq D_c[P + \mathbf{W}_c][\mathbf{N}_r - \mathbf{N}_c] .$$

This is granted if $\mathbf{N}_r \leq \mathbf{N}_c$ and is equivalent to

$$D_c \leq \frac{\mathbf{N}_c\mathbf{N}_r[\mathbf{W}_c - \mathbf{W}_r]}{[P + \mathbf{W}_c][\mathbf{N}_r - \mathbf{N}_c]} \tag{110}$$

if $\mathbf{N}_r > \mathbf{N}_c$. The constraint (110), on the other hand, is in effect only if

$$\mathbf{N}_c(P + \mathbf{W}_c) < \mathbf{N}_r(P + \mathbf{W}_r)$$

for otherwise, it is trivially satisfied because $D_c \leq \mathbf{N}_c$. Substituting $b(\nu,\gamma) = a(\nu,\gamma)$ in (107) yields

$$D_r = \frac{\mathbf{N}_r \mathbf{W}_r \mathbf{N}_c^2 D_c}{\Big(D_c \mathbf{N}_c + \mathbf{N}_r(\mathbf{N}_c - D_c)\Big)\Big((\mathbf{W}_r - \mathbf{W}_c)\mathbf{N}_c + (P + \mathbf{W}_c) D_c\Big)} \, .$$

On the other hand, if $\mathbf{W}_c < \mathbf{W}_r$, it is more helpful to write

$$b(\nu,\gamma) = \frac{\mathbf{W}_r}{\mathbf{W}_c} a(\nu,\gamma) - \bar{\nu}(1-\gamma)^2 \left[\frac{\mathbf{W}_r}{\mathbf{W}_c} - 1\right]$$

as this reveals $b(\nu,\gamma) \leq \frac{\mathbf{W}_r}{\mathbf{W}_c} a(\nu,\gamma)$. Thus, the optimal choice of parameters is potentially $\gamma = 1$ and

$$\nu = \frac{\mathbf{W}_c \mathbf{N}_c}{D_c[P + \mathbf{W}_c]}$$

provided this choice satisfies (109). Once again, after some algebra, that translates to

$$D_c \leq \frac{\mathbf{N}_c P[\mathbf{W}_c \mathbf{N}_c - \mathbf{W}_r \mathbf{N}_r] + \mathbf{W}_c \mathbf{W}_r \mathbf{N}_c [\mathbf{N}_c - \mathbf{N}_r]}{[P + \mathbf{W}_c][\mathbf{N}_c - \mathbf{N}_r]\mathbf{W}_r}$$

Substituting $b(\nu,\gamma) = \frac{\mathbf{W}_r}{\mathbf{W}_c} a(\nu,\gamma)$ in (107) yields

$$D_r = \frac{\mathbf{N}_r \mathbf{N}_c^2 \mathbf{W}_c}{\Big(D_c \mathbf{N}_c + \mathbf{N}_r(\mathbf{N}_c - D_c)\Big)\Big(P + \mathbf{W}_c\Big)} \, .$$

Combining all the above results yields (86) and (87).

C. Proof of Lemma 7

Let us first compare (88) to (86) for the $\mathbf{W}_c \geq \mathbf{W}_r$ case. We shall show for all $\frac{\mathbf{N}_c \mathbf{W}_c}{P + \mathbf{W}_c} \leq D_c \leq \mathbf{N}_c$ that

$$\frac{\mathbf{N}_r \mathbf{N}_c^2}{D_c \mathbf{N}_c + \mathbf{N}_r(\mathbf{N}_c - D_c)} \cdot \frac{\mathbf{W}_r D_c}{(\mathbf{W}_r - \mathbf{W}_c)\mathbf{N}_c + (P + \mathbf{W}_c) D_c} \leq \frac{\mathbf{N}_r \mathbf{W}_r}{P + \mathbf{W}_r} \cdot \frac{D_c \mathbf{N}_c + \frac{\mathbf{N}_c \mathbf{W}_c}{\mathbf{W}_r}(\mathbf{N}_c - D_c)}{D_c \mathbf{N}_c + \mathbf{N}_r(\mathbf{N}_c - D_c)}$$

or equivalently that

$$D_c \mathbf{N}_c (P + \mathbf{W}_r) \leq \left(D_c + \frac{\mathbf{W}_c}{\mathbf{W}_r}(\mathbf{N}_c - D_c)\right)\Big((\mathbf{W}_r - \mathbf{W}_c)\mathbf{N}_c + (P + \mathbf{W}_c) D_c\Big) \, . \tag{111}$$

Adding $D_c \mathbf{N}_c (\mathbf{W}_c - \mathbf{W}_r)$ to both sides of (111) yields

$$D_c \mathbf{N}_c (P + \mathbf{W}_c) \leq \left(D_c + \frac{\mathbf{W}_c}{\mathbf{W}_r}(\mathbf{N}_c - D_c)\right)(P + \mathbf{W}_c) D_c + \frac{\mathbf{W}_c}{\mathbf{W}_r}(\mathbf{N}_c - D_c)(\mathbf{W}_r - \mathbf{W}_c)\mathbf{N}_c \, . \tag{112}$$

Taking the first term on the right-hand side of (112) to the left-hand side, we obtain

$$D_c (P + \mathbf{W}_c)(\mathbf{N}_c - D_c)\left(1 - \frac{\mathbf{W}_c}{\mathbf{W}_r}\right) \leq \frac{\mathbf{W}_c}{\mathbf{W}_r}(\mathbf{N}_c - D_c)(\mathbf{W}_r - \mathbf{W}_c)\mathbf{N}_c$$

or equivalently

$$D_c (P + \mathbf{W}_c) \geq \mathbf{W}_c \mathbf{N}_c$$

which is guaranteed. Equality is satisfied in only three trivial cases: (i) When $D_c = D_c^{WZ}(C_c)$, which coincides with Scheme 0, (ii) when $\mathbf{W}_c = \mathbf{W}_r$, and (iii) when $D_c = \mathbf{N}_c$, which should be excluded if $D_c^{\max} < \mathbf{N}_c$.

As for the $\mathbf{W}_c < \mathbf{W}_r$ case, to prove that Scheme RC-CR is superior, we need to show

$$\frac{\mathbf{N}_r\mathbf{N}_c^2}{D_c\mathbf{N}_c+\mathbf{N}_r(\mathbf{N}_c-D_c)}\cdot\frac{\mathbf{W}_c}{P+\mathbf{W}_c}\leq\frac{\mathbf{N}_r\mathbf{W}_r}{P+\mathbf{W}_r}\cdot\frac{D_c\mathbf{N}_c+\frac{\mathbf{N}_c\mathbf{W}_c}{\mathbf{W}_r}(\mathbf{N}_c-D_c)}{D_c\mathbf{N}_c+\mathbf{N}_r(\mathbf{N}_c-D_c)}$$

or equivalently that

$$\frac{\mathbf{N}_c\mathbf{W}_c}{P+\mathbf{W}_c}\leq\frac{D_c\mathbf{W}_r+\mathbf{W}_c(\mathbf{N}_c-D_c)}{P+\mathbf{W}_r}\ . \tag{113}$$

Rearranging (113), we have

$$\mathbf{N}_c\mathbf{W}_c(P+\mathbf{W}_r)\leq(P+\mathbf{W}_c)\Big(D_c(\mathbf{W}_r-\mathbf{W}_c)+\mathbf{W}_c\mathbf{N}_c\Big)$$

which is once again equivalent to

$$D_c(P+\mathbf{W}_c)\geq\mathbf{W}_c\mathbf{N}_c\ .$$

Equality in this case is satisfied if and only if $D_c = D_c^{WZ}(C_c)$.

References

[1] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Transactions on Information Theory*, 19(2):197–207, March 1973.

[2] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel,", *IEEE Transactions on Information Theory*, 49(7):1691–1706, July 2003.

[3] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, 29(3):439-441, May 1983.

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

[5] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Peredach. Inform.*, 10(3):3–14, 1974.

[6] S. I. Gel'fand and M. S. Pinsker, "Coding for channels with random parameters," *Problems of Control and Information Theory*, 9(1):19–31, 1980.

[7] D. Gündüz and E. Erkip, "Reliable cooperative source transmission with side information," *IEEE Information Theory Workshop*, Bergen, Norway, July 2007.

[8] D. Gündüz, J. Nayak and E. Tuncel, "Wyner-Ziv coding over broadcast channels using hybrid digital/analog transmission," *IEEE International Symposium on Information Theory*, Toronto, ON, July 2008.

[9] G. Kramer and S. Shamai, "Capacity for classes of broadcast channels with receiver side information," *IEEE Information Theory Workshop*, Lake Tahoe, CA, September 2007.

[10] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Transactions on Information Theory*, 47(3):903–917, March 2001.

[11] R. Puri, K. Ramchandran and S. Pradhan, "On seamless digital upgrade of analog transmission systems using coding with side information," *Proceedings of 40th Allerton Conference on Communications, Control and Computing*, Allerton, IL, October 2002.

[12] 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.

[13] S. Shamai, S. Verdú and R. Zamir, "Systematic lossy source/channel coding," *IEEE Transactions on Information Theory*, 44(2):564–579, March 1998.

[14] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem", *IEEE Transactions on Information Theory*, 50(8):1636–1654, August 2004.

[15] A. Sutivong, M. Chiang, T. M. Cover, Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Transactions on Information Theory*, 51(4):1486–1495, April 2005.

[16] C. Tian and S. Diggavi, "Side information scalable source coding," submitted to *IEEE Transactions on Information Theory*, arXiv:0707.4597v1 [cs.IT].

[17] C. Tian and S. Diggavi, "On multistage successive refinement for Wyner–Ziv source coding with degraded side informations," *IEEE Transactions on Information Theory*, 53(8):2946–2960, August 2007.

[18] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Transactions on Information Theory*, 52(4):1469-1482, April 2006.

[19] M. P. Wilson, K. Narayanan and G. Caire, "Joint source channel coding with side information using hybrid digital analog codes," submitted to *IEEE Transactions on Information Theory*, arXiv:0802.3851v1 [cs.IT].

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

[21] 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.