A Semi-Continuous Version of the Berger-Yeung Problem *

Toby Berger
School of Electrical Engineering, Cornell University, USA *berger@ee.cornell.edu*
and
Ram Zamir
Dept. of EE-Systems, Tel Aviv University, ISRAEL *zamir@eng.tau.ac.il*

February 19, 1999

Abstract

We present a continuous dual of multiterminal source encoding with one distortion criterion (the "Berger-Yeung problem"). A continuous source X is encoded with "high resolution" ($D_x \to 0$), with the aid of a "helper", i.e., a correlated discrete or continuous source Y, that is encoded *separately* subject to some arbitrary distortion criterion D_y. We find the asymptotic form of the set of achievable coding rates $\mathcal{R}(D_x, D_y)$ of the X- and Y-encoders as $D_x \to 0$. Two extreme cases of our result provide high resolution interpretations to the classical work of Wyner, Ahlswede, Korner and Ziv on source coding with side information.

Key Words: multiterminal source coding, Berger-Yeung problem, Wyner-Ziv problem, side information, Shannon lower bound, high resolution.

*This research was supported in part by the Wolfson Research Awards, administered by the Israel Academy of Science and Humanities, and by NSF Grants NCR-9216975, NCR-9632266 and IRI-9310670.

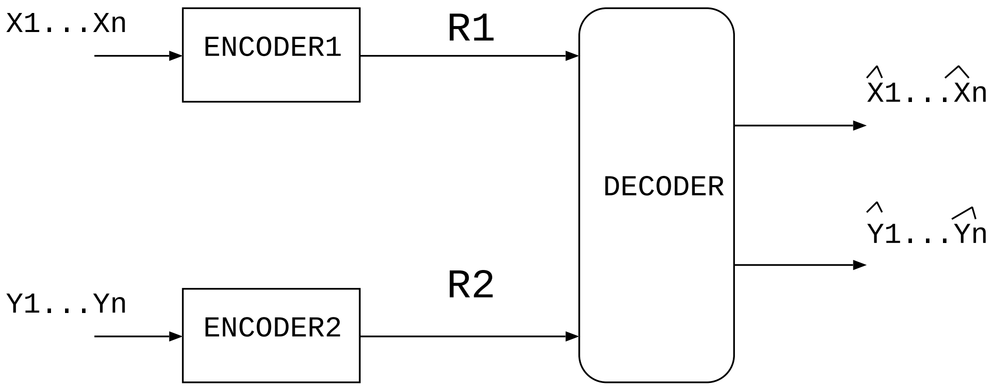

Figure 1: Separate encoding and joint decoding of correlated sources.

I. Introduction

A decoder receives encodings of memoryless correlated sources $\{X_k\}$ and $\{Y_k\}$ from separate encoders at rates R_1 and R_2, respectively, and needs to reconstruct the sources subject to fidelity criteria D_x and D_y. The classical multiterminal source coding problem, currently an open problem, is to find the set of admissible rate pairs $\mathcal{R}(D_x, D_y) = \{(R_1, R_2)\}$ [2, 4]. See Figure 1. Berger and Yeung [3] found the achievable rate region in the special case where $\{X_k\}$ is discrete and its reconstruction is required to be perfect in the usual Shannon sense. Their result consolidates earlier work of Slepian and Wolf [7], Wyner [8], Ahlswede and Korner [1], Wyner and Ziv [10], and Kaspi and Berger [5].

We give the Berger-Yeung problem a high resolution interpretation. We assume that X_k is real valued, has a "smooth" conditional distribution given Y_k, and is encoded with *small* mean squared error D_x. We determine the asymptotic form of the region $\mathcal{R}(D_x, D_y)$ in the limit as $D_x \to 0$ (or $R_1 \to \infty$). Thus, the assumption of lossless coding of $\{X_k\}$ in the Berger-Yeung problem is replaced here by high resolution coding of the continuous source $\{X_k\}$. This result extends our earlier work [12] in which both sources are encoded with high resolution.

The following scenarios may motivate this problem:

- *Y helps X:* We are mainly (or only) interested in an accurate reproduction of the source $\{X_k\}$, but the cost of sending bits through channel 1 is very high, whereas using channel 2 is cheap. Here the encoder of the correlated source $\{Y_k\}$ serves as a "helper" to reduce the rate in channel 1, and D_y is large (or $= \infty$).

- *X helps Y:* We are interested in encoding $\{Y_k\}$, but the cost of sending bits through channel 2 is very high, while the cost of using channel 1 is negligible, allowing a large rate R_1.

- *Joint Compression and Classification:* We have a mixed classification / quantization problem, where the "classifier" $\{Y_k\}$ and the "signal" $\{X_k\}$ have some joint distribution. Here $\{Y_k\}$ is discrete and D_y is small (or zero).

The Berger-Yeung problem reduces to the Wyner-Ziv (WZ) problem [10] (for lossy coding of $\{Y_k\}$ with perfect side information $\{X_k\}$) if the rate of the X-encoder is greater than $H(X)$, and reduces to the Wyner-Ahlswede-Korner (WAK) problem [8, 1] (for lossless coding of $\{X_k\}$ with compressed side information on $\{Y_k\}$) if no condition is imposed on the Y-distortion (i.e., if $D_y \geq D_{y,max}$). Likewise, our semi-continuous problem gives high resolution ($D_x \to 0$) interpretations to the WZ and the WAK problems when $R_1 \geq R_X(D_x)$ and when $D_y \geq D_{y,max}$, respectively.

In the next section we state our main result. In Section III we consider the two special cases of our problem which correspond to the WZ and the WAK problems. In Section IV we give an explicit characterization of the asymptotic rate region when the sources are jointly Gaussian and fidelity is measured by the squared error distortion measure. The proof of the main result is given in the last section.

II. Statement of Result

Consider a sequence $\{(X_k, Y_k)\}$ of independent copies of a pair of correlated random variables $X \in \mathcal{R}^1$ and $Y \in \mathcal{Y}$, where $\mathcal{Y}$ is a discrete or a continuous alphabet. An encoder receives a vector $\underline{X} = X_1, \ldots, X_n$ and transmits an index $W_x = f_1(\underline{X})$. A separate encoder receives a vector $\underline{Y} = Y_1, \ldots, Y_n$ and transmits an index $W_y = f_2(\underline{Y})$. A joint decoder receives both indices and produces reconstructions of the sources $\underline{\hat{X}} = g_x(W_x, W_y)$ and $\underline{\hat{Y}} = g_y(W_x, W_y)$, where $\hat{X}_1, \ldots, \hat{X}_n$ belong to $\mathcal{R}^1$ (like the samples of the source $\underline{X}$), and $\hat{Y}_1, \ldots, \hat{Y}_n$ belong to some reconstruction alphabet $\hat{\mathcal{Y}}$. We are interested in the set $\mathcal{R}(D_x, D_y) = \{(R_1, R_2)\}$ of admissible coding rates

$$R_1 = \frac{1}{n} \log \|f_1\|, \qquad R_2 = \frac{1}{n} \log \|f_2\| \tag{1}$$

subject to a mean squared error constraint on $\underline{X}$

$$\frac{1}{n} \sum_{i=1}^{n} E(\hat{X}_i - X_i)^2 \leq D_x, \tag{2}$$

and some general additive distortion measure on $\underline{Y}$

$$\frac{1}{n} \sum_{i=1}^{n} E d_y(Y_i, \hat{Y}_i) \leq D_y \ , \tag{3}$$

where $d_y : \mathcal{Y} \times \hat{\mathcal{Y}} \to \mathcal{R}^+$. For $\hat{\mathcal{Y}}$ continuous, it is assumed that d_y satisfies Wyner's condition in [9].

We propose an asymptotic characterization for $\mathcal{R}(D_x, D_y)$ in the limit as $D_x \to 0$, where, since X is continuous, R_1 usually goes to infinity. Assume that the conditional differential entropy $h(X|Y)$ exists and is finite. For any random variable V which satisfies a Markov chain relation $X \leftrightarrow Y \leftrightarrow V$, define

$$\mathcal{R}^{**}(D_x, V) = \left\{ (R_1, R_2) \ : \ \begin{array}{rcl} R_1 & \geq & h(X|V) - \frac{1}{2} \log 2\pi e D_x \\ R_2 & \geq & I(Y;V|X) \\ R_1 + R_2 & \geq & I(Y;V|X) + h(X) - \frac{1}{2} \log 2\pi e D_x \end{array} \right\}, \tag{4}$$

and let

$$\mathcal{R}^*(D_x, D_y) = \bigcup_{V:\, X \leftrightarrow Y \leftrightarrow V,\, \mathcal{E}_d(Y|X,V) \leq D_y} \mathcal{R}^{**}(D_x, V) \tag{5}$$

where $\mathcal{E}_d(Y|X,V)$ is defined below. Note that due to the Markov relation $X \leftrightarrow Y \leftrightarrow V$, the expressions in the second and the third lines of (4) have the equivalent forms

$$I(Y;V|X) = I(Y;V) - I(X;V)$$

and

$$I(Y;V|X) + h(X) = I(Y;V) + h(X|V).$$

$\mathcal{E}_d(Y|X,V)$ in (5) denotes the minimum d_y-distortion in estimating Y from (X,V), i.e. (see [11]),

$$\mathcal{E}_d(Y|X,V) = \min_{g:\, \mathcal{X} \times \mathcal{V} \to \hat{\mathcal{Y}}} E d_y(Y, g(X,V)). \tag{6}$$

Thus, the union in (5) includes all random variables V which satisfy a Markov chain relation $X \leftrightarrow Y \leftrightarrow V$ and allow estimation of Y from (X, V) with distortion D_y. If $\mathcal{Y}$ is finite, then it is enough to take the union over V's with alphabet size $|\mathcal{Y}| + 2$; see [3]. Note that the Berger-Yeung rate region [3] has the form of (4), without the "$\frac{1}{2} \log 2\pi e D_x$" terms, and with regular entropies rather than differential entropies of X.

For stating the forward part of Theorem 1 below, we need to impose a certain condition upon the joint distribution of (X, Y) and the distortion measure d_y. Let V be any random variable satisfying a Markov chain relation $X \leftrightarrow Y \leftrightarrow V$, and let N_ϵ be Gaussian with variance ϵ and independent of (X, Y, V). We will require that for any such V

$$\mathcal{E}_d(Y|X + N_\epsilon, V) \to \mathcal{E}_d(Y|X, V), \quad \text{as } \epsilon \to 0. \tag{7}$$

Property (7) is a smoothness condition in X for the estimation of Y, that is not easy to verify in general. Accordingly, we can give explicit sufficient conditions for (7) in two cases (see Appendix for the proof):

1. *(MSE distortion criterion:)* $\mathcal{Y} = \hat{\mathcal{Y}} = \mathcal{R}^1$, $d(y, \hat{y}) = (y - \hat{y})^2$, $EY^2 < \infty$. Condition (7) holds if for any real bounded function $g(x)$

$$E\{g(X)|X + N_\epsilon\} \to g(X) \quad \text{in quadratic mean as } \epsilon \to 0. \tag{8}$$

2. *(Finite reproduction alphabet:)* $|\hat{\mathcal{Y}}| < \infty$, $d(y, \hat{y}) \le d_{max} < \infty$. Condition (7) holds if for any function $g : \mathcal{R}^1 \to \hat{\mathcal{Y}}$

$$g(X + N_\epsilon) \to g(X) \quad \text{in probability as } \epsilon \to 0. \tag{9}$$

We believe that (7) actually holds under much weaker conditions.

Theorem 1 (Asymptotically Tight Outer Bound for the Berger-Yeung Problem:)

A. For any d_y-distortion D_y and mean squared error D_x,

$$\mathcal{R}(D_x, D_y) \subseteq \mathcal{R}^*(D_x, D_y)\ . \tag{10}$$

B. Furthermore, if $h(X|Y)$ and EX^2 are finite, and Condition (7) holds, then for any $\epsilon > 0$ and D_x small enough

$$\mathcal{R}(D_x, D_y) \supseteq \mathcal{R}^*(D_x, D_y)^{(-\epsilon)} \ , \tag{11}$$

where

$$\mathcal{R}^*(D_x, D_y)^{(-\epsilon)} \triangleq \{(R_1, R_2) : \ (R_1 - \epsilon, R_2 - \epsilon) \in \mathcal{R}^*(D_x, D_y)\} \ . \tag{12}$$

Theorem 1 extends the concept of "asymptotically tight Shannon outer bound" for multiterminal source coding, introduced in [12, thms. 1 and 2] on the basis of earlier results for the single terminal case [6]. The setting in [12] assumes that both terminals are continuous and encoded with high resolution. Here, only the rate of the X-terminal (the first row in (4)) is bounded by a "Shannon-type" lower bound, which becomes tight in the limit of high resolution. As in [12], this theorem can easily be extended to more general difference distortion measures for the source X, and to a many-terminal configuration (rather than two as above) in which all but one terminal operate with high resolution.

III. Analysis of the Asymptotic Rate Region

The union over the choice of the random variable V in the definition of $\mathcal{R}^*(D_x, D_y)$ in (5), makes it difficult to evaluate $\mathcal{R}^*(D_x, D_y)$ explicitly. Two important special cases will help us get insight into the structure of the asymptotic rate region $\mathcal{R}^*(D_x, D_y)$, and in some cases to calculate it. The first is when $D_y \geq D_{y,max}$, where

$$D_{y,max} \triangleq \mathcal{E}_d(Y|X) = \min_{g:\ \mathcal{X} \to \hat{\mathcal{Y}}} Ed_y(Y, g(X)). \tag{13}$$

Since for any V we have $\mathcal{E}_d(Y|X,V) \leq \mathcal{E}_d(Y|X) = D_{y,max}$, substituting $D_y \geq D_{y,max}$ in $\mathcal{R}^*(D_x, D_y)$ is equivalent to not imposing any constraint on the estimation of Y, meaning that the auxiliary random variable V in (5) must only satisfy the chain constraint $X \leftrightarrow Y \leftrightarrow V$. Hence, this setting parallels the Wyner-Ahlswede-Korner (WAK) problem of lossless discrete source coding with side information Y [8, 1, 4], but with lossless coding of a discrete source replaced by high resolution coding of the continuous source X. Indeed, for $D_y \geq D_{y,max}$ the region $\mathcal{R}^*(D_x, D_y)$ simplifies to the following form, reminiscent of the WAK region:

$$\mathcal{R}_{WAK}(D_x) = \bigcup_{V:\, X\leftrightarrow Y\leftrightarrow V} \left\{ (R_1, R_2) \;:\; \begin{array}{rcl} R_1 & \geq & h(X|V) - \frac{1}{2}\log 2\pi e D_x \\ R_2 & \geq & I(Y;V) \end{array} \right\}, \tag{14}$$

where the trivial point $(h(X) - \frac{1}{2}\log 2\pi e D_x, 0) \in \mathcal{R}^*(D_x, D_{y,max})$ and the convexity of the rate region $\mathcal{R}^*(D_x, D_y)$ allow in this case to compact the characterization of $\mathcal{R}_{WAK}(D_x)$ into two inequalities rather than three. We call the boundary of this region the "WAK curve".

The second case is when R_2 is equal to its minimal possible value

$$\begin{aligned} \min_{(R_1,R_2)\in\mathcal{R}^*(D_x,D_y)} R_2 &= \min_{\{V:\, X\leftrightarrow Y\leftrightarrow V, \mathcal{E}_d(Y|X,V)\leq D_y\}} I(Y;V|X) \\ &\triangleq R^{WZ}_{Y|X}(D_y)\,, \end{aligned} \tag{15}$$

where $R^{WZ}_{Y|X}(D_y)$ is the rate-distortion function of Y with side information X at the decoder, found by Wyner and Ziv [10, 9, 4, 11]. Hence, for R_1 large enough, the lower boundary of $\mathcal{R}^*(D_x, D_y)$ is given by the line

$$R_2 = R^{WZ}_{Y|X}(D_y), \tag{16}$$

which we call the "WZ line". The interpretation of this line is that in the high resolution limit $D_x \to 0$, the source X is effectively available at the decoder for the purpose of decoding Y.

For an arbitrary value of D_y, the WAK curve (14), the WZ line (16), and the line

$$R_1 + R_2 \geq R^{WZ}_{Y|X}(D_y) + h(X) - \frac{1}{2}\log 2\pi e D_x \tag{17}$$

provide an *outer* bound on $\mathcal{R}^*(D_x, D_y)$ as shown in Figure 2. Note that (17) is obtained by minimizing the third lower bound in (4) over V.

The outer bound above is easier to evaluate than $\mathcal{R}^*(D_x, D_y)$, but unfortunately slightly exceeds it. To see that, let $V^{WZ} = V^{WZ}(D_y)$ be the random variable achieving the minimization in (15). By definition we have $\mathcal{E}_d(Y|X, V^{WZ}) \leq D_y$. Substituting V^{WZ} in $\mathcal{R}^{**}(D_x, V)$, we see that any point (R_1, R_2) on the line (17) <u>up</u> to the point

$$(R_1, R_2) = (h(X|V^{WZ}) - \frac{1}{2}\log(2\pi e D_x), \quad I(Y;V^{WZ})) \tag{18}$$

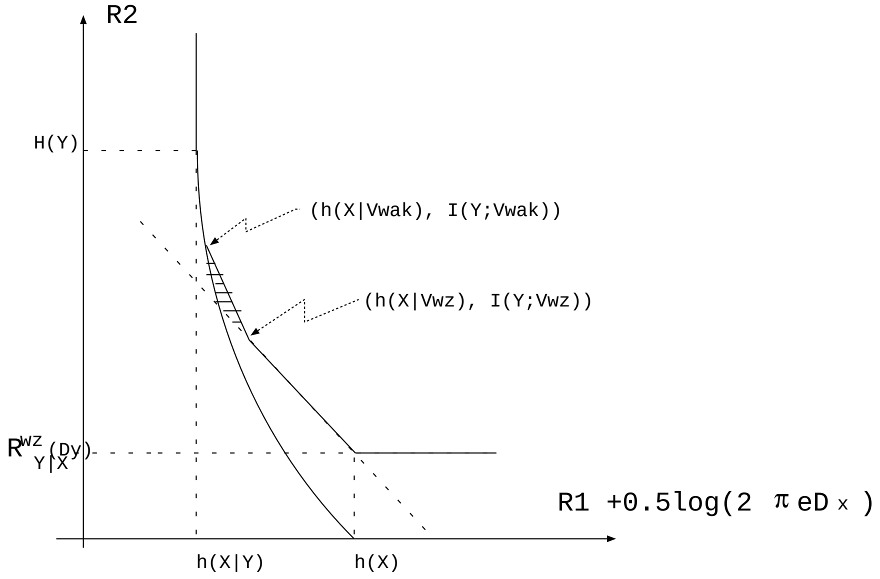

Figure 2: Typical inner and outer bounds on the rate region for a discrete source Y, by means of the WAK curve and the WZ rate-distortion function. The shaded area represents a gap between the bounds.

is achievable. Let $V^{WAK} = V^{WAK}(D_y)$ be a random variable such that the point

$$(R_1, R_2) = (h(X|V^{WAK}) - \frac{1}{2}\log(2\pi e D_x), \quad I(Y; V^{WAK})) \tag{19}$$

is on the WAK curve, and all random variables V associated with points above (19) on the WAK curve satisfy the distortion constraint $\mathcal{E}_d(Y|X,V) \leq D_y$.[1] Then we can achieve any point on the WAK curve down to the point (R_1, R_2) in (19).

Connecting the points (18) and (19) we obtain the inner bound on $\mathcal{R}^*(D_x, D_y)$ shown in Figure 2. This leaves a small portion between the outer and inner bounds that we cannot characterize in terms of the WAK curve and the WZ line. See the shaded region in Figure 2. Berger and Yeung observed similar phenomena in the case of a doubly symmetric binary source [3].

If we are lucky, the random variables V^{WZ} and V^{WAK} are the same, whereupon the outer and inner bounds coincide and fully characterize $\mathcal{R}^*(D_x, D_y)$. This is the case when $\{X_k, Y_k\}$ is Gaussian and squared error is the distortion measure, as described in the next section.

[1]Under natural assumptions on the distortion measure d_y, such a V^{WAK} exists for any $D_y > 0$, e.g., for Y discrete and $d_y =$ Hamming we can always pick $V = Y$ and get $\mathcal{E}_d(Y|X,V) = 0$.

IV. The Quadratic Doubly Gaussian Source Case

Suppose (X, Y) are zero mean jointly Gaussian with variances σ_x^2 and σ_y^2 and correlation coefficient ρ, and $d_y(y, \hat{y}) = (y - \hat{y})^2$ is the squared error distortion measure. In this case the WAK curve is generated by the collection of r.v.'s

$$\left\{ V = Y + Z : \; Z \perp\!\!\!\perp (X, Y), Z \sim \mathcal{N}(0, \sigma_z^2), 0 < \sigma_z^2 \leq \infty \right\} \tag{20}$$

where $\perp\!\!\!\perp$ denotes statistical independence. The WAK region becomes

$$\mathcal{R}_{WAK}(D_x) = \bigcup_{0 \leq \gamma < 1} \left\{ (R_1, R_2) \; : \; \begin{array}{lcl} R_1 & \geq & \frac{1}{2} \log \left(\frac{\sigma_x^2 (1 - \gamma \rho^2)}{D_x} \right) \\ R_2 & \geq & \frac{1}{2} \log \left(\frac{1}{1 - \gamma} \right) \end{array} \right\} . \tag{21}$$

The proof of (20) and (21) follows by a simple application of the entropy-power inequality to (14). Specifically, for (X, Y) jointly Gaussian we can write $X = aY + N$, where $a = \rho \sigma_x / \sigma_y$ and N is zero mean Gaussian with variance $(1 - \rho^2)\sigma_x^2$ that is independent of Y. Furthermore, $X \leftrightarrow Y \leftrightarrow V$, implies that N is independent of (Y, V). Using the conditional version of the entropy-power inequality we then have

$$\begin{aligned} h(X|V) &= h(aY + N|V) && (22) \\ &\geq \frac{1}{2} \log \left(2^{2h(aY|V)} + 2^{2h(N)} \right) && (23) \\ &= \frac{1}{2} \log 2\pi e \left(a^2 P + (1 - \rho^2)\sigma_x^2 \right) && (24) \end{aligned}$$

where $P \triangleq 2^{2h(Y|V)}/2\pi e$, and equality holds if and only if (Y, V) are jointly Gaussian. We also have

$$I(Y; V) = h(Y) - h(Y|V) = \frac{1}{2} \log \left(\frac{\sigma_y^2}{P} \right) .$$

Substituting in (14), and letting $\gamma = 1 - P/\sigma_y^2$, we arrive at (21). To relate the collection of V's in (20) with (21), note that $\gamma = \sigma_y^2/(\sigma_y^2 + \sigma_z^2)$.

For $0 < D_y \leq D_{y,max} = \sigma_y^2(1 - \rho^2)$, the WZ rate distortion function of Y with side information X is given by, [9],

$$R_{Y|X}^{WZ}(D_y) = \frac{1}{2} \log \left(\frac{\sigma_y^2 (1 - \rho^2)}{D_y} \right) . \tag{25}$$

Furthermore, $R_{Y|X}^{WZ}(D_y)$ is realized by $V^{WZ} = Y + Z^{WZ}$, where Z^{WZ} is Gaussian, independent of (X, Y) and has variance $\sigma_{WZ}^2 = [1/D_y - 1/\sigma_y^2(1 - \rho^2)]^{-1}$. Clearly for any $\sigma_z^2 < \sigma_{WZ}^2$, the

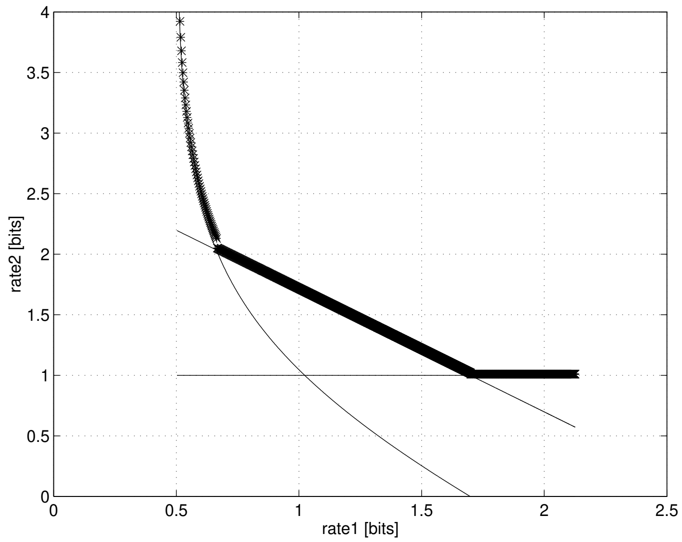

Figure 3: The WAK curve, the WZ line and the rate sum bound, forming a complete characterization of the rate region for correlated Gaussian sources with correlation coefficient $\rho = 0.9$, and signal-to-noise ratios $\sigma_x^2/D_x = 10$ dB and $\sigma_y^2/D_y = 13$ dB.

conditional variance of Y given $(X, V = Y + Z)$ is less than D_y. Thus we have $V^{WAK} = V^{WZ}$, and the outer and inner bounds on $\mathcal{R}^*(D_x, D_y)$ coincide. We conclude that the boundary of $\mathcal{R}^*(D_x, D_y)$ is completely characterized by the WAK curve given in (21), the WZ line $R_2 = R_{Y|X}^{WZ}(D_y)$ given in (25), and condition (17) which here becomes

$$R_1 + R_2 \geq \frac{1}{2} \log \left(\frac{\sigma_x^2 \sigma_y^2 (1 - \rho^2)}{D_x D_y} \right) . \tag{26}$$

Figure 3 shows this characterization of the asymptotic rate region $\mathcal{R}^*(D_x, D_y)$.

V. Proof of Theorem

Proof of Part A (Outer Bound):

Consider any multiterminal coding scheme for n-length source blocks with rates (R_1, R_2) which satisfies distortions D_x and D_y. We will show that $(R_1, R_2) \in \mathcal{R}^*(D_x, D_y)$. Define

$$V_i = (W_y, X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) , \quad i = 1 \ldots n$$

where $W_y = f_2(\underline{Y})$. Note that since (Y_i, X_i) are i.i.d. and W_y is a deterministic function of $\underline{Y}$, we have the following Markov chain relation

$$X_i \leftrightarrow Y_i \leftrightarrow V_i \ . \tag{27}$$

Clearly the rate of the Y-encoder can only decrease if $\underline{X}$ were available directly to the decoder; thus,

$$\begin{aligned} nR_2 &\geq H(W_y|\underline{X}) && (28)\\ &\geq \sum_{i=1}^{n} I(Y_i; V_i|X_i) \ , && (29) \end{aligned}$$

where (29) follows from standard manipulations in the proof of the converse rate-distortion theorem for coding $\underline{Y}$ with side information $\underline{X}$ at the decoder; see e.g. [10] or [4, sec. 14.9]. We next lower bound the rate of the X-encoder by

$$\begin{aligned} nR_1 &\geq H(W_x|W_y) && (30)\\ &= I(\underline{X}; W_x|W_y) && (31)\\ &= h(\underline{X}|W_y) - h(\underline{X}|W_y, W_x) && (32)\\ &= \sum_{i=1}^{n} h(X_i|W_y, X_1 \ldots X_{i-1}) - h(\underline{X} - \hat{\underline{X}}|W_y, W_x) && (33)\\ &\geq \sum_{i=1}^{n} h(X_i|V_i) - h(\underline{X} - \hat{\underline{X}}) && (34)\\ &\geq \sum_{i=1}^{n} h(X_i|V_i) - \frac{n}{2}\log 2\pi e D_x \ , && (35) \end{aligned}$$

where (30) follows by the same reasoning as (28); (31) follows since $W_x = f_1(\underline{X})$; the decomposition into a difference between differential entropies in (32) assumes that $h(\underline{X}|W_y) > -\infty$, otherwise $\sum_{i=1}^{n} h(X_i|V_i) = -\infty$, and (30)-(35) lead to the trivial lower bound $R_1 > -\infty$; (33) follows since $\hat{X}$ is a function of W_y and W_x; (34) follows since further conditioning reduces entropy; and in (35) we substituted the maximum entropy under the constraint $\|\underline{X} - \hat{\underline{X}}\|^2 \leq nD_x$ in (2). Finally, we lower bound the rate sum of both encoders by

$$\begin{aligned} n(R_1 + R_2) &\geq I(\underline{Y}, \underline{X}; W_y, W_x) && (36)\\ &= I(\underline{X}; W_y, W_x) + I(\underline{Y}; W_y, W_x|\underline{X}) && (37) \end{aligned}$$

$$= \quad h(\underline{X}) - h(\underline{X}|W_y, W_x) + H(W_y|\underline{X}) \tag{38}$$

$$\geq \quad \sum_{i=1}^{n} h(X_i) - \frac{n}{2}\log 2\pi e D_x + \sum_{i=1}^{n} I(Y_i; V_i|X_i) \ , \tag{39}$$

where (36) follows like (30) and (31); (37) follows from the chain rule; (38) follows like (32) and using the fact that W_y and W_x are functions of $\underline{Y}$ and $\underline{X}$; and (39) follows like (32)-(34) and (28)-(29), and since $X_1 \ldots X_n$ are independent. Combining (28)-(39) and comparing with the definition of $\mathcal{R}^{**}(D_x, V)$ in (4), we conclude that the pair (R_1, R_2) belongs to the uniform convex combination of the regions $\mathcal{R}^{**}(D_x, V_1), \ldots, \mathcal{R}^{**}(D_x, V_n)$. Letting

$$D_{yi} \triangleq \mathcal{E}_d(Y_i|X_i, V_i) \ , \quad i = 1 \ldots n \tag{40}$$

(see (6)), we can characterize the set containing (R_1, R_2) directly in terms of $\mathcal{R}^*(D_x, D_y)$, as defined in (5). Namely, since $\mathcal{R}^{**}(D_x, V_i) \subset \mathcal{R}^*(D_x, D_{yi})$, (R_1, R_2) belongs to the uniform convex combination of the regions $\mathcal{R}^*(D_x, D_{y1}), \ldots, \mathcal{R}^*(D_x, D_{yn})$, i.e.,

$$(R_1, R_2) \in \left\{ \left(\frac{\sum_{i=1}^{n} r_i'}{n}, \frac{\sum_{i=1}^{n} r_i''}{n} \right) \ : \ (r_i', r_i'') \in \mathcal{R}^*(D_x, D_{yi}) \quad \text{for } i = 1 \ldots n \right\} \ . \tag{41}$$

By a simple extension of the argument in [3, sec. IV] it follows that $\mathcal{R}^*(D_x, D_y)$ is convex in D_y; thus from (41),

$$(R_1, R_2) \in \mathcal{R}^* \left(D_x, \frac{\sum_{i=1}^{n} D_{yi}}{n} \right) \ . \tag{42}$$

Moreover, since $\underline{\hat{Y}}$ is a function of (W_y, W_x), we have for the $\underline{Y}$-distortion (see (6), (40))

$$\begin{aligned} nD_y &= \sum_{i=1}^{n} E d_y(Y_i, \hat{Y}_i) \\ &\geq \sum_{i=1}^{n} \mathcal{E}_d(Y_i|W_y, W_x) \\ &\geq \sum_{i=1}^{n} \mathcal{E}_d(Y_i|W_y, \underline{X}) \\ &= \sum_{i=1}^{n} \mathcal{E}_d(Y_i|V_i, X_i) \\ &= \sum_{i=1}^{n} D_{yi} \ , \end{aligned} \tag{43}$$

Since $\mathcal{R}^*(D_x, D_y)$ is monotonically decreasing with D_y, (43) implies that $\mathcal{R}^*(D_x, \sum_{i=1}^{n} D_{yi}/n) \subseteq \mathcal{R}^*(D_x, D_y)$. So, by (42), (R_1, R_2) belongs to $\mathcal{R}^*(D_x, D_y)$ as claimed. $\square$

Proof of Part B (Asymptotic Achievability):

As in the proof of the asymptotic tightness of the Shannon outer bound in the double high resolution case [12, thm. 2], the main idea here is to use additive noise in the "test channel" associated with the high resolution link, namely $X \to X + N$, where N is Gaussian with variance D_x and is independent of (V, Y, X). If V satisfies $X \leftrightarrow Y \leftrightarrow V$, then this test channel induces a "long" Markov chain $V \leftrightarrow Y \leftrightarrow X \leftrightarrow (X+N)$ that is associated with an achievable rate region, [2],

$$\mathcal{R}^{**}_{add}(D_x, V) = \left\{ (R_1, R_2) \ : \ \begin{array}{rcl} R_1 & \geq & I(X; X+N|V) \\ R_2 & \geq & I(Y; V|X+N) \\ R_1 + R_2 & \geq & I(Y, X; V, X+N) \end{array} \right\}, \quad N \sim \mathcal{N}(0, D_x) \ . \tag{44}$$

Furthermore, the rate pairs in (44) allow us to decode $\underline{X}$ with mean squared error D_x (by setting $\hat{X} = X + N$ in the scheme of [2]), and to decode $\underline{Y}$ with average distortion $\mathcal{E}_d(Y|X+N, V)$. This implies the following inner bound on the rate region

$$\mathcal{R}(D_x, D_y) \supseteq \mathcal{R}^*_{add}(D_x, D_y) \triangleq \bigcup_{\{V:\, X \leftrightarrow Y \leftrightarrow V,\, \mathcal{E}_d(Y|X+N,V) \leq D_y\}} \mathcal{R}^{add}(D_x, V) \ . \tag{45}$$

We now show that asymptotically the reverse inclusion, $\subseteq$, also holds. For that, let us examine the difference between the three lower bounds in (44) and the corresponding lower bounds characterizing $\mathcal{R}^{**}(D_x, V)$ in (4). For the first term we have

$$I(X; X+N|V) - [h(X|V) - \frac{1}{2}\log 2\pi e D_x] = h(X+N|V) - h(X|V) \ . \tag{46}$$

For the second we have, due to the Markov chain relation $V \leftrightarrow Y \leftrightarrow X \leftrightarrow (X+N)$,

$$I(Y; V|X+N) - I(Y; V|X) = I(V; X) - I(V; X+N) = [h(X+N|V) - h(X|V)] - [h(X+N) - h(X)] \ . \tag{47}$$

Finally, we have for the last term

$$I(Y, X; V, X+N) - [I(Y; V|X) + h(X) - \frac{1}{2}\log 2\pi e D_x] = h(X+N|V) - h(X|V) \ . \tag{48}$$

Following the same arguments in the proof of [12, thm. 2] we have, as $D_x \to 0$, $h(X+N) \to h(X)$ and $h(X+N|V) \to h(X|V)$. Thus the quantities in (46)-(48) vanish. Furthermore, by Condition (7) we have $\mathcal{E}_d(Y|X+N,V) \to \mathcal{E}_d(Y|X,V)$. Comparing with the definition of $\mathcal{R}^*_{add}(D_x,D_y)$, we conclude that for any $\delta > 0$ there exists D_x small enough such that $\mathcal{R}^*(D_x,D_y) \subseteq \mathcal{R}^*_{add}(D_x,D_y+\delta)^{(+\delta)}$. Since $\mathcal{R}^*(D_x,D_y)$ is continuous in D_y, we also have $\mathcal{R}^*(D_x,D_y) \subseteq \mathcal{R}^*_{add}(D_x,D_y)^{(+\epsilon)}$ for any $\epsilon > 0$ and D_x small enough. In view of (45) and Part A of the theorem, this completes the proof of Part B. $\square$

Appendix: Sufficient Conditions for (7)

A. MSE criterion

For $d_y(y,\hat{y}) = (y-\hat{y})^2$, the minimum estimation error becomes the conditional variance, i.e.,

$$\mathcal{E}_d(U|W) = VAR(U|W) \triangleq E[U - E(U|W)]^2 \ .$$

We need to show that Condition (8) implies, as $\epsilon \to 0$,

$$VAR(Y|X+N_\epsilon,V) \to VAR(Y|X,V) \ . \tag{49}$$

Let $\hat{Y}_\epsilon = E(Y|X+N_\epsilon,V)$, and $\hat{Y}_0 = E(Y|X,V)$, be the estimators which achieve the conditional variances in (49). The proof of (49) follows from two simple lemmas.

Lemma 1

$$\hat{Y}_\epsilon = E(\hat{Y}_0|X+N_\epsilon,V).$$

Proof: By the low of iterated expectation we have

$$\begin{aligned}\hat{Y}_\epsilon &= E(Y|X+N_\epsilon,V) = E[E(Y|X+N_\epsilon,X,V)|X+N_\epsilon,V] = E[E(Y|N_\epsilon,X,V)|X+N_\epsilon,V] \\ &= E[E(Y|X,V)|X+N_\epsilon,V] = E(\hat{Y}_0|X+N_\epsilon,V) \end{aligned} \tag{50}$$

where we used the fact that N_ϵ is independent of (X,Y,V). $\square$

Lemma 2

$$VAR(Y|X+N_\epsilon,V) = VAR(Y|X,V) + E(\hat{Y}_\epsilon - \hat{Y}_0)^2 \ .$$

Proof: Let $\Delta_0 = Y - \hat{Y}_0$, $\Delta_\epsilon = Y - \hat{Y}_\epsilon$ and $\Delta_{0\epsilon} = \hat{Y}_0 - \hat{Y}_\epsilon$. Lemma 2 amounts to $E\Delta_\epsilon^2 = E\Delta_0^2 + E\Delta_{0\epsilon}^2$. Since $\Delta_\epsilon = \Delta_0 + \Delta_{0\epsilon}$, this is equivalent to $E(\Delta_0\Delta_{0\epsilon}) = 0$. We show that as follows. Since N_ϵ is independent of (X, Y, V), the triple $(X + N_\epsilon) - (X, V) - Y$ form a Markov chain, implying that $\hat{Y}_\epsilon \leftrightarrow (X, V) \leftrightarrow Y$ form a Markov chain, implying that $\Delta_{0\epsilon} - (X, V) - \Delta_0$ form a Markov chain. Hence we can write

$$E(\Delta_0\Delta_{0\epsilon}) = E\{E(\Delta_0\Delta_{0\epsilon}|X, V)\} = E\{E(\Delta_0|X, V) \cdot E(\Delta_{0\epsilon}|X, V)\} = 0,$$

since $E(\Delta_0|X, V) = 0$ by the definition of $\hat{Y}_0$. $\square$

Now, let $g_v(x) \triangleq E(Y|X = x, V = v)$. From Lemma 1 we have $\hat{Y}_\epsilon = E(g_V(X)|X + N_\epsilon)$. Condition (8) then implies that $\hat{Y}_\epsilon \to g_V(X) = \hat{Y}_0$ in quadratic mean as $\epsilon \to 0$. Thus $E(\hat{Y}_\epsilon - \hat{Y}_0)^2 \to 0$, and Lemma 2 implies the desired result.

B. Finite Reproduction Alphabet

Let

$$g_v(x) = \arg\min_{\hat{y}} E\{d_y(Y, \hat{y})|X = x, V = v\}$$

and $\hat{Y}_0 = g_V(X)$ be the estimation function and the estimator achieving $\mathcal{E}_d(Y|X, V)$. Let $\widetilde{Y}_\epsilon = g_V(X + N_\epsilon)$ be a possible, sub-optimum estimator of Y from $X + N_\epsilon$ and V. It is easy to verify that

$$\mathcal{E}_d(Y|X + N_\epsilon, V) \leq Ed_y(Y, \widetilde{Y}_\epsilon) \leq \mathcal{E}_d(Y|X, V) + \Pr(\widetilde{Y}_\epsilon \neq \hat{Y}_0)d_{max}$$

where d_{max} is an upper bound on $d(y, \hat{y})$. By Condition (9), the probability $\Pr(\widetilde{Y}_\epsilon \neq \hat{Y}_0)$ vanishes as $\epsilon \to 0$, and the result follows.

Acknowledgments

We gratefully acknowledge helpful discussions with Raymond Yeung and Robert Liptser.

References

[1] R. Ahlswede and J. Korner. Source coding with side information and a converse for the degraded broadcast channel. *IEEE Trans. Information Theory*, 21:629–637, 1975.

[2] T. Berger. *Multiterminal Source Coding.* In G.Longo, editor, the Information Theory Approach to Communications, Springer-Verlag, New York, 1977.

[3] T. Berger and R. Yeung. Multiterminal source encoding with one distortion criterion. *IEEE Trans. Information Theory*, IT-35:228–236, March 1989.

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

[5] A. Kaspi and T. Berger. Rate-distortion for correlated sources with partially separated encoders. *IEEE Trans. Information Theory*, IT-28:828–840, Nov. 1982.

[6] T. Linder and R. Zamir. On the asymptotic tightness of the Shannon lower bound. *IEEE Trans. Information Theory*, IT-40:2026–2031, Nov. 1994.

[7] D. Slepian and J.K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Information Theory*, IT-19:471–480, July 1973.

[8] A.D. Wyner. On source coding with side information at the decoder. *IEEE Trans. Information Theory*, IT-21:294–300, 1975.

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

[10] A.D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Information Theory*, IT-22:1–10, Jan., 1976.

[11] R. Zamir. The rate loss in the Wyner-Ziv problem. *IEEE Trans. Information Theory*, pages 2073–2084, Nov. 1996.

[12] R. Zamir and T. Berger. Multiterminal source coding with high resolution. *IEEE Trans. Information Theory*, to appear Jan. 1999.

List of Figures

List of Footnotes

1. Under natural assumptions on the distortion measure d_y, such a V^{WAK} exists for any $D_y > 0$, e.g., for Y discrete and $d_y =$ Hamming we can always pick $V = Y$ and get $\mathcal{E}_d(Y|X, V) = 0$.