An Improved Outer Bound for the Multiterminal Source-Coding Problem

Aaron B. Wagner and Venkat Anantharam
Dept. of Electrical Engineering and Computer Sciences
University of California
Berkeley, CA 94720-1772 USA
{awagner,ananth}@eecs.berkeley.edu

***Abstract*— We prove a new outer bound on the rate-distortion region for the multiterminal source-coding problem. This bound subsumes the best known bound in the literature and improves upon it strictly in some cases. The improved bound enables us to obtain a new, conclusive result for the binary-erasure instance of the "CEO problem." The bound recovers many of the converse results that have been established for special cases of the problem, including the recent one for the Gaussian version of the CEO problem.**

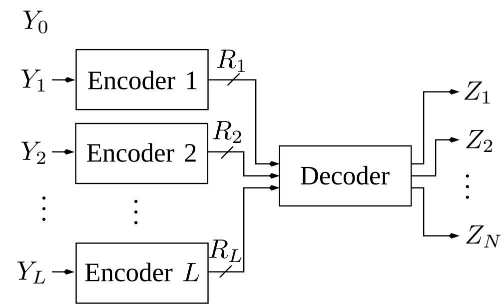

Fig. 1. The multiterminal source-coding problem.

I. INTRODUCTION

We consider a source-coding problem in which several encoders each observe their own stochastic process and then send messages to a decoder over rate-constrained, noiseless links. The decoder attempts to reproduce functions of the observed processes, possibly the processes themselves, subject to a fidelity constraint (see Fig. 1). The problem of computing the Shannon-theoretic rate-distortion region for this setup is traditionally called the *multiterminal source-coding problem*, even though this name suggests a more general network topology. The multiterminal source-coding problem has been open for some time.

Many special cases have been solved, however. For these, the reader is referred to the classical papers of Slepian and Wolf [1], Wyner [2], Ahlswede and Körner [3], Wyner and Ziv [4], Körner and Marton [5], and Gel'fand and Pinsker [6], and to the more recent papers of Berger and Yeung [7], Gastpar [8], Oohama [9], and Prabhakaran, Tse, and Ramchandran [10]. While these results are all conclusive, they are established using coding theorems that are tailored to the special cases under consideration.

For the general problem, there is a natural coding scheme described by Berger [11] and Tung [12]. Each encoder first quantizes its observed stochastic process as in classical rate-distortion theory. These quantized processes are then losslessly transmitted to the decoder using the binning scheme of Cover [13]. The decoder uses these quantized processes to form its reproductions. A description of the resulting achievable region is provided in Section III. This scheme is optimal in all of the special cases mentioned earlier except that of Körner and Marton [5].

The best outer bound in the literature can also be found in Berger [11] and Tung [12] and is also reproduced in Section III of this paper. Since the inner bound is not tight for the problem considered by Körner and Marton, it is clear that the outer bound must not coincide with the inner bound in all instances. This gap cannot be entirely attributed to the inner bound, however, as there are instances of the problem that can be solved from first principles for which the Berger-Tung outer bound is strictly bigger than the true rate region.

The aim of this paper is to establish an improved outer bound for this problem. Such a bound is proven in the next section, following a precise formulation of the problem. We show that this bound subsumes the Berger-Tung outer bound, while improving upon it strictly in some cases, in Section III. One case where we show a strict improvement is the binary-erasure version of the "CEO problem," introduced by Berger, Zheng, and Viswanathan [14]. For this instance our bound yields a conclusive result for the sum rate, in fact. Another case in which we show a strict improvement is the Gaussian version of the CEO problem, which was recently solved by Oohama [9] and independently by Prabhakaran, Tse and Ramchandran [10]. We show that the converse result of these authors can be recovered by applying our single-letter outer bound and then solving the optimization problem that it entails. The converse results used in all of the other solved special cases mentioned earlier are also recovered by our bound. This is discussed in Section IV.

II. FORMULATION AND MAIN RESULT

Let L denote the number of encoders. Let $\{Y_0^n(t), Y_1^n(t), \ldots, Y_L^n(t)\}_{t=1}^n$ be a vector-valued discrete memoryless source. For $A \subset \{1, \ldots, L\}$, we denote $\{Y_i^n(t), i \in A\}$ by $\mathbf{Y}_A^n(t)$. If $A = \{1, \ldots, L\}$, we write this simply as $\mathbf{Y}^n(t)$. In this context, the set A^c should be interpreted as $\{1,\ldots,L\}\backslash A$ rather than $\{0,1,\ldots,L\}\backslash A$. We use $Y_i^n(t_1 : t_2)$ to denote $(Y_i^n(t_1),\ldots,Y_i^n(t_2))$, Y_i^n to denote $Y_i^n(1:n)$, and $Y_i^n(t^c)$ to denote

$$(Y_i^n(1),\ldots,Y_i^n(t-1),Y_i^n(t+1),\ldots,Y_i^n(n)).$$

Similar notation will be used for other vectors to be introduced shortly.

For each i in $\{1,\ldots,L\}$, encoder i observes $\{Y_i^n(t)\}_{t=1}^n$ then employs a mapping $f_i^{(n)} : \mathcal{Y}_i^n \to \{1,\ldots,M_i^{(n)}\}$ to convey information about the observation to the decoder. The decoder uses the received messages to reproduce N functions of the observations according to the mappings

$$\varphi_j^{(n)} : \prod_{i=1}^{L} \left\{1,\ldots,M_i^{(n)}\right\} \mapsto \mathcal{Z}_j^n \text{ for } j=1,\ldots,N.$$

We assume that N distortion measures $d_j : \prod_{i=0}^{L} \mathcal{Y}_i \times \mathcal{Z}_j \mapsto \mathbb{R}^+$ are given. The process $\{Y_0^n(t)\}_{t=1}^n$ should be interpreted as a source of interest that is not directly observed by any encoder.

Definition 1: The rate-distortion vector

$$(\mathbf{R},\mathbf{D}) = (R_1,R_2,\ldots,R_L,D_1,D_2,\ldots,D_N)$$

is *achievable* if there exists a blocklength n, encoders $f_i^{(n)}$, and decoders $\varphi_j^{(n)}$ such that

$$\begin{aligned} \frac{1}{n}\log M_i^{(n)} &\le R_i \quad \text{for all } i, \text{ and} \\ E\left[\frac{1}{n}\sum_{t=1}^{n} d_j(Y_0^n(t),\mathbf{Y}^n(t),Z_j^n(t))\right] &\le D_j \quad \text{for all } j. \end{aligned} \tag{1}$$

Let $\mathcal{RD}^*$ be the set of achievable rate-distortion vectors.

Let $Y_0, Y_1,\ldots,Y_L$ be generic random variables with the distribution of the source at a single time. Let χ denote the set of discrete random variables X with the property that $Y_1,\ldots,Y_L$ are conditionally independent given X. Note that χ is nonempty since it contains, e.g., $X=\mathbf{Y}$. Let Γ_o denote the set of discrete random variables $\gamma = (U_1,\ldots,U_L,Z_1,\ldots,Z_N,W,T)$ satisfying

1) (T,W) is independent of $(Y_0,\mathbf{Y})$,
2) $U_i - (Y_i,T,W) - (Y_0,\mathbf{Y}_{i^c},\mathbf{U}_{i^c})$ for all i in $\{1,\ldots,L\}$ (meaning that U_i, (Y_i,T,W), and $(Y_0,\mathbf{Y}_{i^c},\mathbf{U}_{i^c})$ form a Markov chain in this order), and
3) $(Y_0,\mathbf{Y},W) - (\mathbf{U},T) - \mathbf{Z}$.

There are many ways of coupling a given X in χ and γ in Γ_o to $(Y_0,\mathbf{Y})$. In this paper, we shall focus on the unique coupling for which $X - (Y_0,\mathbf{Y}) - \gamma$, which we call the *Markov coupling*. Whenever the joint distribution of these variables appears, we assume that this coupling is in effect.

Definition 2: Let

$$\begin{aligned} \mathcal{RD}_o(X,\gamma) = \Big\{ (\mathbf{R},\mathbf{D}) : \sum_{i\in A} R_i &\ge I(X;\mathbf{U}_A|\mathbf{U}_{A^c},T) \\ &+ \sum_{i\in A} I(Y_i;U_i|X,W,T) \text{ for all } A, \text{ and} \\ D_j &\ge E[d_j(Y_0,\mathbf{Y},Z_j)] \text{ for all } j \Big\}. \end{aligned}$$

Then define

$$\mathcal{RD}_o = \bigcap_{X\in\chi} \bigcup_{\gamma\in\Gamma_o} \mathcal{RD}_o(X,\gamma).$$

Theorem 1 is our main result.

Theorem 1: $\mathcal{RD}^* \subset \mathcal{RD}_o$.

Proof. Suppose $(\mathbf{R},\mathbf{D})$ is achievable. Let $f_1^{(n)},\ldots,f_L^{(n)}$ be encoders and $(\varphi_1^{(n)},\ldots,\varphi_N^{(n)})$ a decoder such that (1) holds. Take any X in χ and augment the sample space to include X^n in such a way that $(X^n(t),Y_0^n(t),\mathbf{Y}^n(t))$ is independent over t. Next let T be uniformly distributed over $\{1,\ldots,n\}$, independent of X^n, Y_0^n, and $\mathbf{Y}^n$. Then define

$$\begin{aligned} X &= X^n(T) \\ Y_i &= Y_i^n(T) \text{ for each } i \text{ in } \{0,\ldots,L\} \\ U_i &= \left(f_i^{(n)}(Y_i^n), X^n(1:T-1)\right) \text{ for each } i \text{ in } \{1,\ldots,L\} \\ Z_j &= Z_j^n(T) \text{ for each } j \\ W &= X^n(T^c). \end{aligned}$$

It can be verified that the variables $(\mathbf{U},\mathbf{Z},W,T)$ are in Γ_o and that, together with $Y_0,\mathbf{Y}$ and X, they satisfy the Markov coupling. Next note that (1) implies

$$E[d_j(Y_0^n(T),\mathbf{Y}^n(T),Z_j^n(T))] \le D_j \text{ for all } j,$$

or

$$E[d_j(Y_0,\mathbf{Y},Z_j)] \le D_j \text{ for all } j.$$

Now let $A \subset \{1,\ldots,L\}$. Then by the cardinality bound on entropy,

$$n\sum_{i\in A} R_i \ge H\left(f_i^{(n)}(Y_i^n), i\in A\right).$$

Since conditioning reduces entropy, this implies

$$\begin{aligned} n\sum_{i\in A} R_i &\ge H\left(f_i^{(n)}(Y_i^n), i\in A \middle| f_i^{(n)}(Y_i^n), i\in A^c\right) \\ &= I\left(X^n,\mathbf{Y}_A^n; f_i^{(n)}(Y_i^n), i\in A \middle| f_i^{(n)}(Y_i^n), i\in A^c\right). \end{aligned}$$

By the chain rule for mutual information,

$$\begin{aligned} &I\left(X^n,\mathbf{Y}_A^n; f_i^{(n)}(Y_i^n), i\in A \middle| f_i^{(n)}(Y_i^n), i\in A^c\right) \\ &\quad = I\left(X^n; f_i^{(n)}(Y_i^n), i\in A \middle| f_i^{(n)}(Y_i^n), i\in A^c\right) \\ &\quad + I\left(\mathbf{Y}_A^n; f_i^{(n)}(Y_i^n), i\in A \middle| X^n, f_i^{(n)}(Y_i^n), i\in A^c\right). \end{aligned} \tag{2}$$

Applying the chain rule again gives

$$I\left(X^n; f_i^{(n)}(Y_i^n), i \in A \Big| f_i^{(n)}(Y_i^n), i \in A^c\right) = \sum_{t=1}^{n} I\Big(X^n(t); f_i^{(n)}(Y_i^n), i \in A \Big| f_i^{(n)}(Y_i^n), i \in A^c, X^n(1:t-1)\Big).$$

Consider next the second term on the right-hand side of (2). Since $X \in \chi$,

$$I\left(\mathbf{Y}_A^n; f_i^{(n)}(Y_i^n), i \in A \Big| X^n, f_i^{(n)}(Y_i^n), i \in A^c\right) = \sum_{i \in A} I\left(Y_i^n, f_i^{(n)}(Y_i^n) \Big| X^n\right).$$

Applying the chain rule for the third time gives

$$I\left(Y_i^n, f_i^{(n)}(Y_i^n) \Big| X^n\right) = \sum_{t=1}^{n} I\left(Y_i^n(t), f_i^{(n)}(Y_i^n) \Big| X^n, Y_i^n(1:t-1)\right).$$

But

$$I\left(Y_i^n(t), f_i^{(n)}(Y_i^n) \Big| X^n, Y_i^n(1:t-1)\right) = H(Y_i^n(t)|X^n, Y_i^n(1:t-1)) - H\left(Y_i^n(t) \Big| f_i^{(n)}(Y_i^n), X^n, Y_i^n(1:t-1)\right)$$

while

$$H(Y_i^n(t)|X^n, Y_i^n(1:t-1)) = H(Y_i^n(t)|X^n)$$

and

$$H\left(Y_i^n(t) \Big| f_i^{(n)}(Y_i^n), X^n, Y_i^n(1:t-1)\right) \le H\left(Y_i^n(t) \Big| f_i^{(n)}(Y_i^n), X^n\right).$$

It follows that

$$\sum_{i \in A} R_i \ge \frac{1}{n} \sum_{t=1}^{n} \Big[I\left(X^n(t); f_i^{(n)}(Y_i^n), i \in A \Big| f_i^{(n)}(Y_i^n), i \in A^c, X^n(1:t-1)\right) + \sum_{i \in A} I\left(Y_i^n(t); f_i^{(n)}(Y_i^n) \Big| X^n(t), X^n(t^c)\right) \Big]. \quad (3)$$

If A^c is nonempty, this can be rewritten as

$$\sum_{i \in A} R_i \ge I(X^n(T); \mathbf{U}_A | \mathbf{U}_{A^c}, T) + \sum_{i \in A} I(Y_i^n(T); U_i | X^n(T), X^n(T^c), T) = I(X; \mathbf{U}_A | \mathbf{U}_{A^c}, T) + \sum_{i \in A} I(Y_i; U_i | X, W, T).$$

The case $A = \{1, \ldots, L\}$ is handled separately. In this case, observe that

$$I\left(X^n(t); f_i^{(n)}(Y_i^n), i \in A \Big| f_i^{(n)}(Y_i^n), i \in A^c, X^n(1:t-1)\right) = I\left(X^n(t); f_i^{(n)}(Y_i^n), i \in A \Big| X^n(1:t-1)\right) = I\left(X^n(t); f_i^{(n)}(Y_i^n), i \in A, X^n(1:t-1)\right).$$

Substituting this into (3) and proceeding as in the $A^c \neq \emptyset$ case completes the proof. ■

III. Relation to Existing Bounds

It is instructive to compare this outer bound with the following inner bound described by Berger [11] and Tung [12].

Definition 3: Let Γ_i^{BT} denote the set of discrete random variables $\gamma = (U_1, \ldots, U_L, Z_1, \ldots, Z_N, T)$ satisfying

1) T is independent of $(Y_0, \mathbf{Y})$,
2) $U_i - (Y_i, T) - (Y_0, \mathbf{Y}_{i^c}, \mathbf{U}_{i^c})$ for all i in $\{1, \ldots, L\}$, and
3) $(Y_0, \mathbf{Y}) - (\mathbf{U}, T) - \mathbf{Z}$.

Then let

$$\mathcal{RD}_i^{BT}(\gamma) = \Big\{(\mathbf{R}, \mathbf{D}) : \sum_{i \in A} R_i \ge I(\mathbf{U}_A; \mathbf{Y}_A | \mathbf{U}_{A^c}, T) \text{ for all } A, \text{ and } E[d_j(Y_0, \mathbf{Y}, Z_j)] \le D_j \text{ for all } j\Big\}.$$

Finally, let

$$\mathcal{RD}_i^{BT} = \bigcup_{\gamma \in \Gamma_i^{BT}} \mathcal{RD}_i^{BT}(\gamma).$$

This bound and the next one were originally proven for a formulation of the problem that is less general than the one used here. It is evident from the proofs, however, that the bounds also hold for our formulation. One way to understand the difference between $\mathcal{RD}_o$ and $\mathcal{RD}_i^{BT}$ is to observe that if one restricts W to be deterministic in the definition of $\mathcal{RD}_o$, the resulting region is exactly $\mathcal{RD}_i^{BT}$. In practice, this means that in order to obtain coincident inner and outer bounds, it suffices to show that restricting W to be deterministic in the definition of $\mathcal{RD}_o$ does not reduce the region. The remainder of the paper contains several examples for which this is done.

The best outer bound on $\mathcal{RD}^*$ in the literature is the following one due to Berger [11] and Tung [12].

Definition 4: Let Γ_o^{BT} denote the set of discrete random variables $\gamma = (\mathbf{U}, \mathbf{Z}, T)$ such that

1) T is independent of $(Y_0, \mathbf{Y})$,
2) $U_i - (Y_i, T) - (Y_0, \mathbf{Y}_{i^c})$ for all i in $\{1, \ldots, L\}$, and
3) $(Y_0, \mathbf{Y}) - (\mathbf{U}, T) - \mathbf{Z}$.

Then let

$$\mathcal{RD}_o^{BT}(\gamma) = \Big\{(\mathbf{R}, \mathbf{D}) : \sum_{i \in A} R_i \ge I(\mathbf{U}_A; \mathbf{Y} | \mathbf{U}_{A^c}, T) \text{ for all } A, \text{ and } D_j \ge E[d_j(Y_0, \mathbf{Y}, Z_j)] \text{ for all } j\Big\}.$$

Finally, let

$$\mathcal{RD}_o^{BT} = \bigcup_{\gamma \in \Gamma_o^{BT}} \mathcal{RD}_o^{BT}(\gamma).$$

The difference between this bound and $\mathcal{RD}_i^{BT}$ is that condition 2) has been weakened. We next show that this outer bound is subsumed by the one in the previous section.

Proposition 1: $\mathcal{RD}_o \subset \mathcal{RD}_o^{BT}$.

Proof. If $(\mathbf{R}, \mathbf{D})$ is in $\mathcal{RD}_o$, then there exists γ in Γ_o such that

$$\sum_{i \in A} R_i \geq I(\mathbf{U}_A; \mathbf{Y} | \mathbf{U}_{A^c}, T) \text{ for all } A$$

and

$$D_j \geq E[d_j(Y_0, \mathbf{Y}, Z_j)] \text{ for all } j$$

(this can be seen by choosing $X = \mathbf{Y}$). Note that for each i,

$$U_i - (Y_i, T, W) - (Y_0, \mathbf{Y}_{i^c}),$$

from the definition of Γ_o. Since $(Y_0, \mathbf{Y}_{i^c}) - (Y_i, T) - W$, it holds

$$U_i - (Y_i, T) - (Y_0, \mathbf{Y}_{i^c}).$$

It follows that $(\mathbf{R}, \mathbf{D})$ is in $\mathcal{RD}_o^{BT}$. ■

The balance of this section consists of two examples for which the containment in Proposition 1 is strict. The first example has been analyzed before. The second is novel.

A. Gaussian CEO Problem [15], [16], [9], [10]

In this example, $Y_0, \ldots, Y_L$ are jointly Gaussian and $Y_1, \ldots, Y_L$ are conditionally independent given Y_0. For $i \geq 1$, let us write $Y_i = Y_0 + N_i$, where $Y_0, N_1, \ldots, N_L$ are independent and

$$E\left[N_i^2\right] = \sigma_i^2 > 0.$$

There is a single distortion measure

$$d_1(Y_0, \mathbf{Y}, Z_1) = (Y_0 - Z_1)^2.$$

This example models the situation in which several sensors observe a single process of interest in the presence of thermal noise. This problem has been studied by a number of authors. Its rate-distortion region was recently found by Oohama [9] and independently by Prabhakaran, Tse, and Ramchandran [10]. It should be mentioned that the formulation studied by Oohama [9] is slightly more general than the one described here.

It is straightforward to extend Theorem 1 to this continuous setting. We omit the details for brevity and instead focus on the results. Using ideas from Oohama [16] and Prabhakaran, Tse, and Ramchandran [10], one can show that

$$\begin{aligned} \mathcal{RD}_o = \{(R_1, \ldots, R_L, D) :\ & \text{there exists} \\ & (r_1, \ldots, r_L) : \text{for all } A, \sum_{i \in A} R_i \geq \frac{1}{2} \log \frac{1}{D} \\ & - \frac{1}{2} \log \left(\frac{1}{\sigma_{Y_0}^2} + \sum_{i \in A^c} \frac{1 - \exp(-2 r_i)}{\sigma_i^2} \right) + \sum_{i \in A} r_i \right\}. \end{aligned}$$

It follows from Theorem 1 in Oohama [9] that this region equals $\overline{\mathcal{RD}^*}$, the closure of the rate-distortion region. That is, $\mathcal{RD}_o$ is tight for this example. The Berger-Tung outer bound can also be extended to the continuous case without difficulty [11]. It can be shown, however, that $\mathcal{RD}_o^{BT}$ contains points outside the rate-distortion region.

B. Binary-Erasure CEO Problem

We turn to a discrete example. Here Y_0 is uniformly distributed over $\{-1, 1\}$, and $Y_i = N_i \cdot Y_0$ for $i \geq 1$, where $N_1, \ldots, N_L$ are i.i.d. with $0 < \Pr(N_1 = 0) = \epsilon < 1$ and $\Pr(N_1 = 1) = 1 - \epsilon$. Let $\mathcal{Z}_1 = \{-1, 0, 1\}$ and

$$d_1(Y_0, \mathbf{Y}, Z_1) = \begin{cases} 0 & \text{if } Y_0 = Z_1, \\ 1 & \text{if } Z_1 = 0, \\ K & \text{otherwise.} \end{cases} \tag{4}$$

We are particularly interested in the large-K limit.

This example is motivated by the following problem arising in power-limited sensor networks. We seek to detect a sequence of i.i.d. binary uniform random variables. To this end, we deploy an array of sensors, each of which is capable of detecting the random variables with negligible probability of error. To lengthen the lifetime of the network, each sensor spends a fraction ϵ of the time in a low-power "sleep" state. We assume that the sensors sleep asynchronously; at each discrete time, each sensor sleeps with probability ϵ, independent of the other sensors and the past. Sensors do not receive any information while they are asleep, resulting in erasures. We permit the coding process to introduce additional erasures, but not errors, yielding the distortion measure (4) and our interest in large-K asymptotics. What sum rate is required in order for the decoder to reproduce a fraction $1-D$ of the Y_0^n variables? Of course, D must satisfy $D \geq \epsilon^L$.

One can show that for $\epsilon^L \leq D \leq 1$, in the limit as $K \to \infty$,

$$\begin{aligned} \inf \left\{ \sum_{i=1}^{L} R_i : (R_1, \ldots, R_L, D) \in \mathcal{RD}_o \right\} &= (1 - D) \log 2 \\ &+ L \left[h\left(D^{1/L}\right) - (1 - \epsilon) h \left(\frac{D^{1/L} - \epsilon}{1 - \epsilon} \right) \right], \end{aligned}$$

where $h(\cdot)$ is the binary entropy function. Using the Berger-Tung inner bound, one can show that this sum rate is attainable. That is, the outer bound yields a conclusive result for the sum rate of the binary-erasure CEO problem. Evidently this problem was previously unsolved. One can numerically verify that, in general, $\mathcal{RD}_o^{BT}$ contains points with a strictly smaller sum rate.

IV. Recovery of Existing Results

We have seen that the new outer bound recovers the converse of Oohama [9] and Prabhakaran, Tse, and Ramchandran [10] for the Gaussian CEO problem. It turns out that it also recovers the converses used by Slepian and Wolf [1], Wyner [2], Ahlswede and Körner [3], Wyner and Ziv [4], Körner and Marton [5], Gel'fand and Pinsker [6], Berger and Yeung [7], and Gastpar [8]. In the case of Körner and Marton, the outer bound can be evaluated and shown to coincide with the true rate-distortion region by proceeding along the lines of the original converse proof [5]. We omit the details for brevity and focus on the other results. For these, we show something more, namely that our outer bound can be used to solve the following problem that subsumes all of these results as special cases.

Assume that $Y_2, \ldots, Y_L$ are conditionally independent given Y_1, and $\mathcal{Z}_1 = \mathcal{Y}_1$ with

$$d_1(Y_0, \mathbf{Y}, Z_1) = \begin{cases} 0 & \text{if } Z_1 = Y_1, \\ 1 & \text{otherwise.} \end{cases}$$

We do not make any assumptions about Y_0 or the other distortion metrics. We would like to characterize the set $\overline{\mathcal{RD}^*} \cap \{D_1 = 0\}$. Thus, the first encoder's observation is particular. It must be reproduced with negligible probability of error, and conditioned on it, the other encoders' observations are independent.

Proposition 2: For this problem,

$$\overline{\mathcal{RD}_o} \cap \{D_1 = 0\} = \overline{\mathcal{RD}_i^{BT}} \cap \{D_1 = 0\} = \overline{\mathcal{RD}^*} \cap \{D_1 = 0\}.$$

Proof (sketch). It suffices to show that $\overline{\mathcal{RD}_o} \cap \{D_1 = 0\}$ is contained in $\overline{\mathcal{RD}_i^{BT}} \cap \{D_1 = 0\}$. For brevity, we will show instead that $\mathcal{RD}_o \cap \{D_1 = 0\}$ is contained in $\mathcal{RD}_i^{BT} \cap \{D_1 = 0\}$, the proof of which conveys the main idea. Suppose $(\mathbf{R}, 0, D_2, \ldots, D_N)$ is contained in $\mathcal{RD}_o$. By choosing $X = Y_1$ in Definition 2, we see that there exists $(\mathbf{U}, \mathbf{Z}, W, \tilde{T})$ in Γ_o such that $Z_1 = Y_1$,

$$E[d_j(Y_0, \mathbf{Y}, Z_j)] \leq D_j \text{ for all } j \geq 2,$$

and for all A,

$$\sum_{i \in A} R_i \geq I(Y_1; \mathbf{U}_A | \mathbf{U}_{A^c}, \tilde{T}) + \sum_{i \in A} I(Y_i, U_i | Y_1, W, \tilde{T}).$$

Now

$$\begin{aligned} I(Y_1; \mathbf{U}_A | \mathbf{U}_{A^c}, \tilde{T}) &= H(Y_1 | \mathbf{U}_{A^c}, \tilde{T}) - H(Y_1 | \mathbf{U}_A, \mathbf{U}_{A^c}, \tilde{T}) \\ &\geq H(Y_1 | \mathbf{U}_{A^c}, W, \tilde{T}) \\ &\quad - H(Y_1 | \mathbf{U}_A, \mathbf{U}_{A^c}, W, \tilde{T}), \end{aligned}$$

since conditioning reduces entropy and the second term is zero. This implies

$$I(Y_1; \mathbf{U}_A | \mathbf{U}_{A^c}, \tilde{T}) \geq I(Y_1; \mathbf{U}_A | \mathbf{U}_{A^c}, W, \tilde{T}).$$

So

$$\begin{aligned} \sum_{i \in A} R_i &\geq I(Y_1; \mathbf{U}_A | \mathbf{U}_{A^c}, W, \tilde{T}) + \sum_{i \in A} I(Y_i, U_i | Y_1, W, \tilde{T}) \\ &= I(Y_1; \mathbf{U}_A | \mathbf{U}_{A^c}, W, \tilde{T}) \\ &\quad + I(\mathbf{Y}_A, \mathbf{U}_A | \mathbf{U}_{A^c}, Y_1, W, \tilde{T}) \\ &= I(Y_1, \mathbf{Y}_A; \mathbf{U}_A | \mathbf{U}_{A^c}, W, \tilde{T}) \\ &\geq I(\mathbf{Y}_A; \mathbf{U}_A | \mathbf{U}_{A^c}, W, \tilde{T}). \end{aligned}$$

If we now define $T = (W, \tilde{T})$, it is evident from Definition 3 that $(\mathbf{R}, 0, D_2, \ldots, D_N)$ is in $\mathcal{RD}_i^{BT}$. ■

The result of Berger and Yeung [7] can be recovered by particularizing to the case $L = 2$, in which case the conditional independence assumption always holds. Berger and Yeung describe how further reductions recover the results of Slepian and Wolf [1], Wyner [2], Ahlswede and Körner [3], and Wyner and Ziv [4]. It should be mentioned that some of these earlier workers define the problem more stringently than is done in this paper or in Berger and Yeung, by requiring that the chance of guessing any of the n symbols $Y_1^n(1), \ldots, Y_1^n(n)$ incorrectly be made arbitrarily small. But this is not a concern since it is known that the two definitions give rise to the same answer. Of course, our outer bound continues to hold under the more stringent definition.

Gel'fand and Pinsker [6] solve the special case in which $R_1 = 0$ and only the distortion constraint on Y_1 is present ($N = 1$). They also define the problem more stringently than we do here, but again this is not a concern. Gastpar [8] solves the problem with Y_1 presented to the decoder as side information. His result can be recovered by considering points in $\overline{\mathcal{RD}^*} \cap \{D_1 = 0\} \cap \{R_1 = H(Y_1)\}$. In all cases, the rate region in Proposition 2 can be shown to coincide with those in the original works. As such, the outer bound supplied here is sufficient to recover the earlier results.

ACKNOWLEDGMENT

It is a pleasure to acknowledge informative discussions with Vinod Prabhakaran of U.C. Berkeley. In particular, the results in Section III-A are due to him.

REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[2] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, May 1975.

[3] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, Nov. 1975.

[4] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[5] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 219–221, Mar. 1979.

[6] S. I. Gel'fand and M. S. Pinsker, "Coding of sources on the basis of observations with incomplete information," *Probl. Pered. Inform.*, vol. 15, no. 2, pp. 45–57, Apr.–Jul. 1979, in Russian.

[7] T. Berger and R. W. Yeung, "Multiterminal source encoding with one distortion criterion," *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 228–236, Mar. 1989.

[8] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2762–2768, Nov. 2004.

[9] Y. Oohama, "Multiterminal source coding for correlated memoryless Gaussian sources with several side informations at the decoder," in *Proc. IEEE Inform. Theory and Comm. Workshop*, 1999, p. 100.

[10] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. Int. Symp. Inform. Theory*, 2004, p. 117.

[11] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, ser. CISM Courses and Lectures, G. Longo, Ed. Springer-Verlag, 1978, vol. 229, pp. 171–231.

[12] S.-Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of Electrical Engineering, Cornell University, Ithaca, NY, May 1978.

[13] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 226–228, Mar. 1975.

[14] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[15] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, Sept. 1997.

[16] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.