An Improved Outer Bound for Multiterminal Source Coding

Aaron B. Wagner, *Member, IEEE*, and Venkat Anantharam, *Fellow, IEEE*

***Abstract*—We prove a new outer bound on the rate–distortion region for the multiterminal source-coding problem. This bound subsumes the best outer 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 version 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 two-encoder problem.**

***Index Terms*—CEO problem, erasure distortion, multiterminal source coding, outer bound, rate region, rate–distortion.**

I. Introduction

In their lauded paper [1], Slepian and Wolf characterize the information rates needed to losslessly communicate two correlated, memoryless information sources when these sources are encoded separately. Their well-known result states that two discrete sources Y_1 and Y_2 can be losslessly reproduced if

$$R_1 > H(Y_1|Y_2)$$
$$R_2 > H(Y_2|Y_1)$$
$$R_1 + R_2 > H(Y_1, Y_2)$$

where R_1 is the rate of the encoder observing Y_1 and R_2 is the rate of the encoder observing Y_2. Conversely, lossless reproduction is not possible if (R_1, R_2) lies outside the closure of this region. See Cover and Thomas [2, Sec. 15.4] for a precise statement of the result and a modern proof. This result is naturally viewed as a multisource generalization of the classical result of Shannon [3], which says that, loosely speaking, a discrete memoryless source can be losslessly reproduced if and only if the data rate exceeds the entropy of the source. Shannon also studied a generalization of this result, albeit in a different direction. He studied the problem of reproducing a source imperfectly, subject to a fidelity constraint, and showed that the required rate is given by the well-known rate–distortion formula [3], [4]. One of the central problems of Shannon theory is to understand the limits of source coding for models that combine the two generalizations. That is, we seek to determine the rates required to reproduce two correlated sources, each subject to a fidelity constraint, when the sources are encoded separately (see Fig. 1). Determining the set of achievable rates and distortions for this setup is often called the multiterminal source-coding problem, even though this name suggests a more elaborate network topology. This problem has been unsolved for some time.

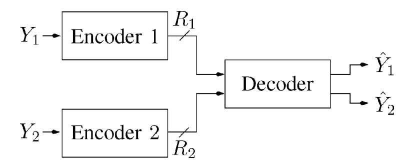

Fig. 1. Separate encoding of correlated sources.

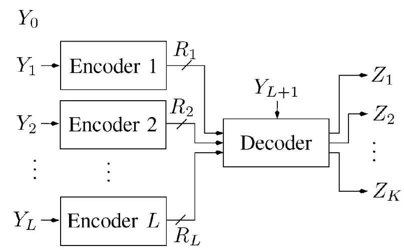

Fig. 2. A general model.

The model we consider in this paper is slightly more general and is depicted in Fig. 2. Beyond considering an arbitrary number of encoders L, we also allow for a hidden source Y_0, which is not directly observed by any encoder or the decoder, and a "side information" source Y_{L+1}, which is observed by the decoder but not by any encoder. We also permit arbitrary functions of the sources to be reproduced, in addition to, or in place of, the sources themselves. We will therefore use Z_1, Z_2, etc., to denote the instantaneous estimates instead of $\hat{Y}_1$, $\hat{Y}_2$, etc., as before. In this paper, we will refer to this more general problem as the multiterminal source-coding problem.

One might doubt the wisdom of embellishing the model when even the basic form shown in Fig. 1 is unsolved. But one of the contributions of this paper is to show that far from obscuring the problem, the added generality actually illuminates it. Of course, the more general problem is also unsolved.

Many special cases have been solved, however. For these, the reader is referred to the classical papers of Slepian and Wolf [1], mentioned earlier; Wyner [5]; Ahlswede and Körner [6];

Manuscript received November 29, 2005; revised February 10, 2008. This work was supported by DARPA under Grants F30602-00-2-0538 and N66001-00-C-8062, by the Office of Naval Research under Grant N00014-1-0637 and by the National Science Foundation under Grants ECS-0123512 and CCF-0500234. The work of A. B. Wagner was also supported by the National Science Foundation under Grant CCF-0642925 (CAREER). The work of V. Anantharam was also supported by the National Science Foundation under Grants CNS-0627161 and CCF-0635372, by Marvell Semiconductor, and by the University of California MICRO program. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Adelaide, Australia, September 2005. This work was primarily conducted when A. B. Wagner was with the University of California, Berkeley.

A. B. Wagner is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: wagner@ece.cornell.edu).

V. Anantharam is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: ananth@eecs.berkeley.edu).

Communicated by R. W. Yeung, Associate Editor for Shannon Theory.

Color version of Figure 4 in this paper is available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2008.920249

Wyner and Ziv [7]; Körner and Marton [8]; and Gel'fand and Pinsker [9]; and to the more recent papers of Berger and Yeung [10]; Zamir and Berger [11]; Gastpar [12]; Oohama [13]; Prabhakaran, Tse, and Ramchandran [14]; and Wagner, Tavildar, and Viswanath [15]. While all of these papers contain conclusive results, these results are established using coding theorems that are tailored to the special cases under consideration.

The solutions to these solved special cases suggest a coding technique for the general model [16], [17]. The idea is this. Each encoder first quantizes its observation as in single-user rate–distortion theory. The quantized processes are then losslessly communicated to the decoder using the binning scheme of Cover [18]. The decoder uses the quantized processes to produce the desired estimates. The set of rate–distortion vectors that can be achieved using this scheme is described in Section III. This inner bound to the rate–distortion region is tight in all of the special cases listed above except that of Körner and Marton [8]. Indeed, the Körner–Marton problem seems to require a custom coding technique that relies on the problem's unique structure. This suggests that the multiterminal source-coding problem may not have a classical single-letter solution.

We attack this problem, therefore, by proving single-letter inner and outer bounds on the rate–distortion region. The best inner bound in the literature has just been described. The best outer bound, which is due to Berger [16] and Tung [17], is described in Section III. In light of the result of Körner and Marton, it is clear that the two bounds must not coincide in all cases. 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–distortion region (see Section III-A of this paper).

Our aim is to provide an improved outer bound for the problem. We prove such a bound in the next section, following a precise formulation of the problem. We show that our bound is contained in (i.e., subsumes) the Berger–Tung outer bound in Section III. In that section, we also provide several examples for which the containment is strict.

One example is the binary erasure version of the "CEO problem." The CEO problem is a special case of the multiterminal source-coding problem in which the observed processes $Y_1, \ldots, Y_L$ are conditionally independent given the hidden process Y_0 and in which the decoder (the CEO) is only interested in estimating the hidden process.[1] Berger, Zhang, and Viswanathan [19] characterize the tradeoff between sum rate and Hamming distortion in the high-rate and many-encoder regime. Gel'fand and Pinsker [9] had earlier found the rate region in the lossless case. We consider the problem in which Y_0 is binary and uniform, and the encoders observe Y_0 through independent binary erasure channels. The decoder reproduces Y_0 subject to a constraint on the "erasure distortion" (see Section III-B or Cover and Thomas [2, p. 338]). For this problem, we show that our outer bound is tight in the sum rate for any number of users. In contrast, the Berger–Tung outer bound contains points whose sum rate is strictly smaller than the optimum.

In our view, this result is of interest in its own right. The binary erasure CEO problem arises naturally in sensor networks in which the sensors occasionally "sleep" to conserve energy. This application is described in Section III-B. The result also provides an example for which the binning-based coding scheme mentioned earlier is optimum. Finally, this is one of relatively few conclusive results for the multiterminal source-coding problem in general, and the CEO problem in particular.

Two of the other conclusive results available are for Gaussian versions of the problem. One is the Gaussian CEO problem, which was first studied by Viswanathan and Berger [20]. Here the encoders observe a hidden Gaussian source through independent Gaussian additive-noise channels. The distortion measure is squared error. The rate–distortion region for this problem was found by Oohama [21], [13] and independently by Prabhakaran, Tse, and Ramchandran [14]. We show that the converse result of these four authors can be recovered from our single-letter outer bound, while the Berger–Tung outer bound contains points that lie outside the true rate–distortion region. The other conclusive result is for the quadratic Gaussian version of the two-encoder problem depicted in Fig. 1. The rate region for this problem was recently determined by Wagner, Tavildar, and Viswanath [15]. For this problem too our outer bound is tight while the Berger–Tung outer bound is not.

The converse results used to solve all of the other special cases mentioned so far are also consequences of our bound. This is discussed in Section IV. Our outer bound therefore serves to unify most of what is known about the nonexistence of multiuser source codes. This unification is noteworthy in the case of the Gaussian problems because the connection between those converse results and the classical discrete results is not immediately apparent. As we will see, subject to some technical caveats, most of the key results in multiterminal source coding can be recovered by combining the general inner bound described earlier with the outer bound described next.

II. Formulation and Main Result

We work exclusively in discrete time. We use uppercase letters to denote random variables, lowercase letters to denote their realizations, and script letters to denote their ranges. Let $\{Y_0^n(t), Y_1^n(t), \ldots, Y_L^n(t), Y_{L+1}^n(t)\}_{t=1}^n$ be a vector-valued, finite-alphabet, memoryless source. The superscript n denotes the time horizon or *block length*. For $A \subseteq \{1, \ldots, L\}$, we denote $(Y_\ell^n(t))_{\{\ell \in A\}}$ by $\boldsymbol{Y}_A^n(t)$. If $A = \{1, \ldots, L\}$, we write this simply as $\boldsymbol{Y}^n(t)$. In this context, the set A^c should be interpreted as $\{1, \ldots, L\} \backslash A$ rather than $\{0, \ldots, L+1\} \backslash A$. When $A = \{\ell\}$, we shall write $Y_\ell^n(t)$ and $\boldsymbol{Y}_{\ell^c}^n(t)$ in place of $Y_{\{\ell\}}(t)$ and $\boldsymbol{Y}_{\{\ell\}^c}(t)$, respectively. Also, we use $Y_\ell^n(t_1 : t_2)$ to denote $\{Y_\ell^n(t)\}_{t=t_1}^{t_2}$, Y_ℓ^n to denote $Y_\ell^n(1 : n)$, and $Y_\ell^n(t^c)$ to denote

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

Similar notation will be used for other vectors that appear later.

[1]This definition is not as restrictive as it might seem. Indeed, any instance of the multiterminal source-coding problem with a single distortion constraint can be transformed into an instance of the CEO problem by lumping $Y_1, \ldots, Y_L$ into Y_0 and redefining the distortion measure as needed. Nonetheless, it defines a useful special case.

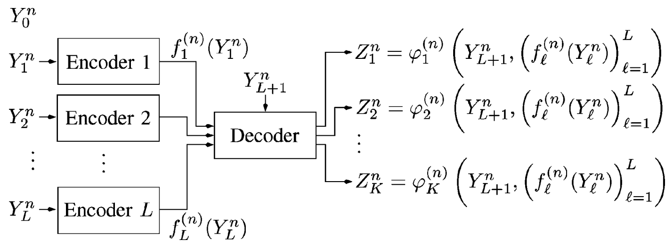

Fig. 3. Notation for the encoding and decoding rules.

The notation for the encoding and decoding rules is shown in Fig. 3. For each ℓ in $\{1, \ldots, L\}$, encoder ℓ observes a block of symbols of length n, Y_ℓ^n. It then employs a mapping

$$f_\ell^{(n)} : \mathcal{Y}_\ell^n \to \left\{1, \ldots, M_\ell^{(n)}\right\}$$

to convey information about the observed block to the decoder. The decoder observes Y_{L+1}^n and uses it and the received messages to estimate K functions of the vector-valued source using the mappings

$$\varphi_k^{(n)} : \mathcal{Y}_{L+1}^n \times \prod_{\ell=1}^{L} \left\{1, \ldots, M_\ell^{(n)}\right\} \mapsto \mathcal{Z}_k^n, \text{ for } k = 1, \ldots, K.$$

We assume that K distortion measures $d_k : \prod_{\ell=0}^{L+1} \mathcal{Y}_\ell \times \mathcal{Z}_k \mapsto \mathbb{R}_+$ are given.

We mention at this point that while the generality of this setup will be useful later when studying examples, it is not needed to appreciate the bounding technique itself. The reader is welcome to focus on the basic model shown in Fig. 1 for that purpose.

Definition 1: The rate–distortion vector

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

is achievable if there exists a block length n, encoders $f_\ell^{(n)}$, and a decoder

$$\left(\varphi_1^{(n)}, \ldots, \varphi_K^{(n)}\right)$$

such that[2]

$$R_\ell \geq \frac{1}{n} \log M_\ell^{(n)}, \quad \text{for all } \ell, \text{ and}$$
$$D_k \geq E\left[\frac{1}{n}\sum_{t=1}^{n} d_k(Y_0^n(t), \boldsymbol{Y}^n(t), Y_{L+1}^n(t), Z_k^n(t))\right], \quad \text{for all } k. \quad (1)$$

Let $\mathcal{RD}_\star$ be the set of achievable rate–distortion vectors. Its closure, $\overline{\mathcal{RD}_\star}$, is called the *rate–distortion region*.

We will sometimes be concerned with slices of the rate–distortion region. We denote these by, for example, $\overline{\mathcal{RD}_\star} \cap \{R_1 = 0\}$, meaning

$$\{(R_2, \ldots, R_L, D_1, \ldots, D_K) : (0, R_2, \ldots, R_L, D_1, \ldots, D_K) \in \overline{\mathcal{RD}_\star}\}.$$

In this paper, we view lossless compression as a limit of lossy compression with the distortion tending to zero. More precisely, if we wish to reproduce Y_1 losslessly, we will set, say, $\mathcal{Z}_1 = \mathcal{Y}_1$ with d_1 equal to Hamming distance, and then examine $\overline{\mathcal{RD}_\star} \cap \{D_1 = 0\}$. This notion of lossless compression is weaker than the one traditionally used. It is common instead to require that for all sufficiently large block lengths, there exists a code for which the probability of correctly reproducing the entire vector Y_1^n is arbitrarily close to 1. But a weaker notion is desirable here since we are proving an outer bound or converse result.

[2]All logarithms and exponentiations in this paper have base e.

Definition 2: Let $Y_0, Y_1, \ldots, Y_{L+1}$ be generic random variables with the distribution of the source at a single time. Let Γ_o denote the set of finite-alphabet random variables $\gamma = (U_1, \ldots, U_L, Z_1, \ldots, Z_K, W, T)$ satisfying the following conditions:

(i) (W, T) is independent of $(Y_0, \boldsymbol{Y}, Y_{L+1})$;
(ii) $U_\ell \leftrightarrow (Y_\ell, W, T) \leftrightarrow (Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1}, \boldsymbol{U}_{\ell^c})$, shorthand for "$U_\ell$, (Y_ℓ, W, T) and $(Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1}, \boldsymbol{U}_{\ell^c})$ form a Markov chain in this order," for all ℓ; and
(iii) $(Y_0, \boldsymbol{Y}, W) \leftrightarrow (\boldsymbol{U}, Y_{L+1}, T) \leftrightarrow \boldsymbol{Z}$.

It is straightforward to verify that Γ_o is precisely the set of finite-alphabet random variables

$$(U_1, \ldots, U_L, Z_1, \ldots, Z_K, W, T)$$

whose joint distribution with $(Y_0, \boldsymbol{Y}, Y_{L+1})$ factors as

$$p(y_0, \boldsymbol{y}, y_{L+1}, \boldsymbol{u}, \boldsymbol{z}, w, t) = p(y_0, \boldsymbol{y}, y_{L+1}) p(w, t) \prod_{\ell=1}^{L} p(u_\ell | y_\ell, w, t) p(\boldsymbol{z} | \boldsymbol{u}, y_{L+1}, t).$$

This description is useful in that it suggests a parametrization of the space Γ_o.

Definition 3: Let χ denote the set of finite-alphabet random variables X with the property that $Y_1, \ldots, Y_L$ are conditionally independent given (X, Y_{L+1}).

Note that χ is nonempty since it contains, e.g., $X = (Y_1, \ldots, Y_{L-1})$.

There are many ways of coupling a given X in χ and γ in Γ_o to the source. In this paper, we shall only consider the unique coupling for which $X \leftrightarrow (Y_0, \boldsymbol{Y}, Y_{L+1}) \leftrightarrow \gamma$, which we call the *Markov coupling*. Whenever the joint distribution of X, $(Y_0, \boldsymbol{Y}, Y_{L+1})$, and γ arises, we assume that this coupling is in effect.

It is evident from the definition of χ that there is considerable latitude in choosing how X depends on Y_0. This is because the sole constraint on the choice of X only depends on the joint distribution of X and $(\boldsymbol{Y}, Y_{L+1})$. But as the following definition shows, this freedom is inconsequential since our outer bound only depends on the distributions of $(Y_0, \boldsymbol{Y}, Y_{L+1}, \gamma)$ and $(X, \boldsymbol{Y}, Y_{L+1}, \gamma)$ separately.

Definition 4: Let

$$\mathcal{RD}_o(X, \gamma) = \Bigg\{(\boldsymbol{R}, \boldsymbol{D}) : \sum_{\ell \in A} R_\ell \geq I(X; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, T) + \sum_{\ell \in A} I(Y_\ell; U_\ell | X, Y_{L+1}, W, T) \text{ for all } A \subseteq \{1, \ldots, L\},$$
$$\text{and } D_k \geq E[d_k(Y_0, \boldsymbol{Y}, Y_{L+1}, Z_k)] \text{ for all } k\Bigg\}.$$

Then define

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

Our main result is that $\mathcal{RD}_o$ is an outer bound on the rate region. In particular, for each $X \in \chi$

$$\bigcup_{\gamma \in \Gamma_o} \mathcal{RD}_o(X, \gamma)$$

is an outer bound. We can therefore view $\mathcal{RD}_o$ as an intersection of a family of outer bounds, one for each choice of X.

Theorem 1: The rate–distortion region is contained in $\overline{\mathcal{RD}_o}$. In fact, $\mathcal{RD}_\star \subseteq \mathcal{RD}_o$.

Proof: It suffices to show the second statement. Suppose $(\boldsymbol{R}, \boldsymbol{D})$ is achievable. Let $f_1^{(n)}, \ldots, f_L^{(n)}$ be encoders and $(\varphi_1^{(n)}, \ldots, \varphi_K^{(n)})$ a decoder satisfying (1). Take any X in χ and augment the sample space to include X^n so that

$$(X^n(t), Y_0^n(t), \boldsymbol{Y}^n(t), Y_{L+1}^n(t))$$

is independent over t. Next, let T be uniformly distributed over $\{1, \ldots, n\}$ and independent of X^n, Y_0^n, $\boldsymbol{Y}^n$, and Y_{L+1}^n. Then define

$$\begin{aligned}
X &= X^n(T) \\
Y_\ell &= Y_\ell^n(T) \text{ for each } \ell \text{ in } \{0, \ldots, L+1\} \\
U_\ell &= \left(f_\ell^{(n)}(Y_\ell^n), X^n(1:T-1), Y_{L+1}^n(T^c)\right) \text{ for each } \ell \\
Z_k &= Z_k^n(T) \text{ for each } k \\
W &= (X^n(T^c), Y_{L+1}^n(T^c)).
\end{aligned}$$

We will show that $\gamma = (\boldsymbol{U}, \boldsymbol{Z}, W, T)$ is in Γ_o and that, together with Y_0, $\boldsymbol{Y}$, Y_{L+1}, and X, it satisfies the Markov coupling. Condition (i) in the definition of Γ_o is satisfied because the source (including X) is independent and identically distributed (i.i.d.) over time. Condition (iii) holds because $\boldsymbol{Z}$ is a function of $(\boldsymbol{U}, Y_{L+1}, T)$. To show condition (ii), observe that since X is in χ

$$Y_\ell^n(t^c) \leftrightarrow (X^n(t^c), Y_{L+1}^n(t^c)) \leftrightarrow \boldsymbol{Y}_{\ell^c}^n(t^c), \quad \text{for all } t \text{ and } \ell.$$

Since the source is i.i.d., this implies

$$Y_\ell^n(t^c) \leftrightarrow (Y_\ell^n(t), X^n(t^c), Y_{L+1}^n(t^c)) \leftrightarrow (\boldsymbol{Y}_{\ell^c}^n, Y_0^n(t), Y_{L+1}^n(t)).$$

Since T is independent of the source, this implies

$$Y_\ell^n(T^c) \leftrightarrow (Y_\ell, W, T) \leftrightarrow (\boldsymbol{Y}_{\ell^c}^n, Y_0, Y_{L+1}).$$

Since U_ℓ is a function of (Y_ℓ^n, W, T), this implies

$$U_\ell \leftrightarrow (Y_\ell, W, T) \leftrightarrow (\boldsymbol{Y}_{\ell^c}^n, Y_0, Y_{L+1}).$$

Likewise, $\boldsymbol{U}_{\ell^c}$ is a function of $(\boldsymbol{Y}_{\ell^c}^n, W, T)$, so

$$U_\ell \leftrightarrow (Y_\ell, W, T) \leftrightarrow (\boldsymbol{Y}_{\ell^c}, Y_0, Y_{L+1}, \boldsymbol{U}_{\ell^c})$$

which is condition (ii) in the definition of Γ_o. This establishes that γ is in Γ_o. To show that the Markov coupling is satisfied, note that since the source is i.i.d.

$$I(X^n(t); X^n(t^c), \boldsymbol{Y}^n(t^c), Y_{L+1}^n(t^c) | Y_0^n(t), \boldsymbol{Y}^n(t), Y_{L+1}^n(t)) = 0.$$

By averaging this equation over t and using the fact that T is independent of the source, we obtain

$$I(X; X^n(T^c), \boldsymbol{Y}^n(T^c), Y_{L+1}^n(T^c), | Y_0, \boldsymbol{Y}, Y_{L+1}, T) = 0.$$

Also, since T is independent of the source

$$I(X; T | Y_0, \boldsymbol{Y}, Y_{L+1}) = 0.$$

By the chain rule for mutual information, these two equations imply

$$I(X; X^n(T^c), \boldsymbol{Y}^n(T^c), Y_{L+1}^n(T^c), T | Y_0, \boldsymbol{Y}, Y_{L+1}) = 0.$$

But γ is a function of

$$(X^n(T^c), \boldsymbol{Y}^n, Y_{L+1}^n, T)$$

so this implies

$$I(X; \gamma | Y_0, \boldsymbol{Y}, Y_{L+1}) = 0$$

which is the condition for the Markov coupling.

It remains to show that $(\boldsymbol{R}, \boldsymbol{D})$ is in $\mathcal{RD}_o(X, \gamma)$. First, note that (1) implies

$$D_k \geq E[d_k(Y_0^n(T), \boldsymbol{Y}^n(T), Y_{L+1}^n(T), Z_k^n(T))], \quad \text{for all } k,$$

i.e.,

$$D_k \geq E[d_k(Y_0, \boldsymbol{Y}, Y_{L+1}, Z_k)], \quad \text{for all } k.$$

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

$$n \sum_{\ell \in A} R_\ell \geq H\left(\left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A}\right).$$

Since conditioning reduces entropy, this implies

$$\begin{aligned}
n \sum_{\ell \in A} R_\ell &\geq H\left(\left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A} \middle| \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A^c}, Y_{L+1}^n\right) \\
&= I\left(X^n, \boldsymbol{Y}_A^n; \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A} \middle| \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A^c}, Y_{L+1}^n\right).
\end{aligned} \tag{2}$$

By the chain rule for mutual information

$$\begin{aligned}
I&\left(X^n, \boldsymbol{Y}_A^n; \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A} \middle| \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A^c}, Y_{L+1}^n\right) \\
&= I\left(X^n; \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A} \middle| \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A^c}, Y_{L+1}^n\right) \\
&\quad + I\left(\boldsymbol{Y}_A^n; \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A} \middle| \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A^c}, X^n, Y_{L+1}^n\right).
\end{aligned} \tag{3}$$

Applying the chain rule again gives the equation at the top of the following page. Consider next the second term on the right-hand side of (3). Since $X \in \chi$

$$\begin{aligned}
I&\left(\boldsymbol{Y}_A^n; \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A} \middle| \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell \in A^c}, X^n, Y_{L+1}^n\right) \\
&= \sum_{\ell \in A} I\left(Y_\ell^n; f_\ell^{(n)}(Y_\ell^n) \middle| X^n, Y_{L+1}^n\right).
\end{aligned}$$

$$I\left(X^n; \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell\in A} \,\middle|\, \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell\in A^c}, Y_{L+1}^n\right) = \sum_{t=1}^n I\left(X^n(t); \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell\in A} \,\middle|\, \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell\in A^c}, X^n(1:t-1), Y_{L+1}^n\right).$$

Applying the chain rule once more gives

$$I\left(Y_\ell^n; f_\ell^{(n)}(Y_\ell^n) \middle| X^n, Y_{L+1}^n\right) = \sum_{t=1}^n I\left(Y_\ell^n(t); f_\ell^{(n)}(Y_\ell^n) \middle| X^n, Y_\ell^n(1:t-1), Y_{L+1}^n\right).$$

But

$$\begin{aligned} I\Big(Y_\ell^n(t); f_\ell^{(n)}(Y_\ell^n) \Big| X^n, Y_\ell^n(1:t-1), Y_{L+1}^n\Big) &\\ + I(Y_\ell^n(t); Y_\ell^n(1:t-1) | X^n, Y_{L+1}^n) &\\ = I\left(Y_\ell^n(t); f_\ell^{(n)}(Y_\ell^n) \middle| X^n, Y_{L+1}^n\right) &\\ + I\left(Y_\ell^n(t); Y_\ell^n(1:t-1) \middle| f_\ell^{(n)}(Y_\ell^n), X^n, Y_{L+1}^n\right) & \end{aligned}$$

and the second term on the left-hand side is zero. Thus

$$I\left(Y_\ell^n(t); f_\ell^{(n)}(Y_\ell^n) \middle| X^n, Y_\ell^n(1:t-1), Y_{L+1}^n\right) \geq I\left(Y_\ell^n(t); f_\ell^{(n)}(Y_\ell^n) \middle| X^n, Y_{L+1}^n\right).$$

Substituting the results of these various calculations into (2) gives (4) at the bottom of the page. If A^c is nonempty, this can be rewritten as

$$\begin{aligned} \sum_{\ell\in A} R_\ell &\geq I(X^n(T); \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}^n(T), T) \\ &\quad + \sum_{\ell\in A} I(Y_\ell^n(T); U_\ell | X^n(T), X^n(T^c), Y_{L+1}^n(T), Y_{L+1}^n(T^c), T) \\ &= I(X; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, T) + \sum_{\ell\in A} I(Y_\ell; U_\ell | X, Y_{L+1}, W, T). \end{aligned}$$

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

$$\begin{aligned} &I\Big(X^n(t); \Big(f_\ell^{(n)}(Y_\ell^n)\Big)_{\ell\in A} \Big| \Big(f_\ell^{(n)}(Y_\ell^n)\Big)_{\ell\in A^c}, X^n(1:t-1), Y_{L+1}^n\Big) \\ &= I\left(X^n(t); \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell\in A} \middle| X^n(1:t-1), Y_{L+1}^n\right) \\ &= I\left(X^n(t); \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell\in A} \middle| X^n(1:t-1), Y_{L+1}^n\right) \\ &\quad + I(X^n(t); X^n(1:t-1), Y_{L+1}^n(t^c) | Y_{L+1}^n(t)) \\ &= I\Big(X^n(t); \Big(f_\ell^{(n)}(Y_\ell^n)\Big)_{\ell\in A}, X^n(1:t-1), Y_{L+1}^n(t^c) \Big| Y_{L+1}^n(t)\Big). \end{aligned}$$

Substituting this into (4) and proceeding as in the $A^c \neq \emptyset$ case completes the proof. □

It is worth noting that the proof uses classical techniques. Most of the manipulations in the latter part of the proof can be viewed as versions of the chain rule for mutual information. Since this chain rule holds in abstract spaces [22, eq. (3.6.6)], the proof can be readily extended to more general alphabets.

The key step in the proof is the introduction of X^n in (2). Unlike the other auxiliary random variables, X^n does not represent a component of the code. Rather, it is used to aid the analysis by inducing conditional independence among the messages sent by the encoders. This technique of augmenting the source to induce conditional independence was pioneered by Ozarow [23], who used it to solve the Gaussian two-descriptions problem. Wang and Viswanath [24] used it to determine the sum rate of the Gaussian vector multiple-descriptions problem with individual and central decoders. It was also used by Wagner, Tavildar, and Viswanath [15] to solve the Gaussian two-encoder source-coding problem. A step that is similar to (2) appeared in Gel'fand and Pinsker [9] and in later papers on the Gaussian CEO problem [13], [14], although in these works X^n is part of the source, so no augmentation is involved.

The significance of conditional independence has long been known in the related field of distributed detection (e.g., [25]). Given the similarity between distributed detection and the multiterminal source-coding problem, one expects conditional independence to play a significant role here as well. Indeed, most conclusive results for the multiterminal source-coding problem involve some kind of conditional independence assumption [9], [19], [12]–[14]. The motivation for introducing X^n is that it allows one to apply the approach used in these works to problems that lack conditional independence.

We do not consider the problem of computing $\mathcal{RD}_o$ in this paper. Note that we have not specified the alphabet sizes of the auxiliary random variables $\boldsymbol{U}$, W, and T. As such, the outer bound provided by Theorem 1 is not computable [26, p. 259] in the present form. One might question the utility of an outer bound that cannot be computed. The remainder of the paper, however, will show that the bound is still useful as a theoretical tool. In addition, cardinality bounds might be found later, although obtaining such bounds appears to be more difficult in this case than for related bounds.

$$\sum_{\ell\in A} R_\ell \geq \frac{1}{n}\sum_{t=1}^n \Bigg[I\left(X^n(t); \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell\in A} \,\middle|\, \left(f_\ell^{(n)}(Y_\ell^n)\right)_{\ell\in A^c}, X^n(1:t-1), Y_{L+1}^n\right) + \sum_{\ell\in A} I\left(Y_\ell^n(t); f_\ell^{(n)}(Y_\ell^n) \middle| X^n(t), X^n(t^c), Y_{L+1}^n(t), Y_{L+1}^n(t^c)\right)\Bigg]. \quad (4)$$

It should be mentioned that the time-sharing variable T is unnecessary; it can be absorbed into the other variables. We have included it to ease the comparison with existing inner and outer bounds, to which we turn next.

III. Relation to Existing Bounds

The coding scheme described in the Introduction gives rise to the following inner bound on the rate–distortion region.

Definition 5: Let Γ_i^{BT} denote the set of finite-alphabet random variables

$$\gamma = (U_1, \ldots, U_L, Z_1, \ldots, Z_K, T)$$

satisfying the following conditions:

(i) T is independent of $(Y_0, \boldsymbol{Y}, Y_{L+1})$;
(ii) $U_\ell \leftrightarrow (Y_\ell, T) \leftrightarrow (Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1}, \boldsymbol{U}_{\ell^c})$ for all ℓ; and
(iii) $(Y_0, \boldsymbol{Y}) \leftrightarrow (\boldsymbol{U}, Y_{L+1}, T) \leftrightarrow \boldsymbol{Z}$.

Then define

$$\begin{aligned}\mathcal{RD}_i^{BT}(\gamma) = \Bigg\{(\boldsymbol{R}, \boldsymbol{D}) : &\sum_{\ell \in A} R_\ell \geq I(\boldsymbol{Y}_A; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, T) \text{ for all } A, \\ &\text{and } D_k \geq E[d_k(Y_0, \boldsymbol{Y}, Y_{L+1}, Z_k)] \text{ for all } k\Bigg\}.\end{aligned}$$

Finally, let

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

Proposition 1 ([16], [17]): $\mathcal{RD}_i^{BT}$ is an inner bound, i.e., $\overline{\mathcal{RD}_i^{BT}} \subseteq \overline{\mathcal{RD}_\star}$.

In Appendix F we show that $\mathcal{RD}_i^{BT}$ is in fact closed. We call $\mathcal{RD}_i^{BT}$ the Berger–Tung [16], [17] inner bound, since although these authors prove a bound that is less general than the one given here, their proof can be extended to prove Proposition 1. See Chen *et al.* [27] or Gastpar [12] for recent sketches of the proof that accommodate some of the generalizations included here.

To understand the difference between $\mathcal{RD}_i^{BT}$ and $\mathcal{RD}_o$, suppose that

$$(\boldsymbol{U}, \boldsymbol{Z}, W, T)$$

is in Γ_o and W is deterministic. Then $(\boldsymbol{U}, \boldsymbol{Z}, T)$ is in Γ_i^{BT}, and for all $A \subseteq \{1, \ldots, L\}$ and all $X \in \chi$

$$\begin{aligned}&I(X; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, T) + \sum_{\ell \in A} I(Y_\ell; U_\ell | X, Y_{L+1}, W, T) \\ &= I(X; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, T) + I(\boldsymbol{Y}_A; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, X, Y_{L+1}, W, T) \\ &= I(X; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, T) + I(\boldsymbol{Y}_A; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, X, Y_{L+1}, T) \\ &= I(X, \boldsymbol{Y}_A; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, T) \\ &= I(\boldsymbol{Y}_A; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, T).\end{aligned}$$

Thus

$$\mathcal{RD}_o(X, \boldsymbol{U}, \boldsymbol{Z}, W, T) = \mathcal{RD}_i^{BT}(\boldsymbol{U}, \boldsymbol{Z}, T). \tag{5}$$

Conversely, if $(\boldsymbol{U}, \boldsymbol{Z}, T)$ is in Γ_i^{BT}, then for any deterministic W, $(\boldsymbol{U}, \boldsymbol{Z}, W, T)$ is in Γ_o and (5) holds for any X. It follows that $\mathcal{RD}_i^{BT}$ is equal to $\mathcal{RD}_o$ with W restricted to be deterministic.

In particular, to obtain coincident inner and outer bounds, it suffices to show that restricting W to be deterministic does not reduce $\mathcal{RD}_o$. We will see later how this can be accomplished in several examples. Of course, it is not possible for the problem solved by Körner and Marton [8], since they show that the inner bound is not tight in that case.

The best general outer bound in the literature is the following.

Definition 6: Let Γ_o^{BT} denote the set of finite-alphabet random variables $\gamma = (\boldsymbol{U}, \boldsymbol{Z}, T)$ satisfying the following conditions:

(i) T is independent of $(Y_0, \boldsymbol{Y}, Y_{L+1})$;
(ii) $U_\ell \leftrightarrow (Y_\ell, T) \leftrightarrow (Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1})$ for all ℓ; and
(iii) $(Y_0, \boldsymbol{Y}) \leftrightarrow (\boldsymbol{U}, Y_{L+1}, T) \leftrightarrow \boldsymbol{Z}$.

Then define

$$\begin{aligned}\mathcal{RD}_o^{BT}(\gamma) = \Bigg\{(\boldsymbol{R}, \boldsymbol{D}) : &\sum_{\ell \in A} R_\ell \geq I(\boldsymbol{Y}; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, T) \text{ for all } A, \\ &\text{and } D_k \geq E[d_k(Y_0, \boldsymbol{Y}, Y_{L+1}, Z_k)] \text{ for all } k\Bigg\}.\end{aligned}$$

Finally, let

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

Proposition 2 ([16], [17]): $\mathcal{RD}_o^{BT}$ is an outer bound, i.e., $\mathcal{RD}_\star \subseteq \mathcal{RD}_o^{BT}$.

As with the inner bound, Berger [16] and Tung [17] prove the result for a model that is more restrictive than the one considered here, but their proof can be extended to this setup (cf. [27], [12]). The difference between $\mathcal{RD}_i^{BT}$ and $\mathcal{RD}_o^{BT}$ is that condition (ii) has been weakened in the latter. To understand this difference, it is helpful to consider the special case in which there are two encoders and no hidden source Y_0, side information Y_{L+1}, or time-sharing variable T. In this case, condition (ii) in the inner bound reduces to

$$U_1 \leftrightarrow Y_1 \leftrightarrow Y_2 \leftrightarrow U_2. \tag{6}$$

For the outer bound, on the other hand, condition (ii) reduces to

$$\begin{aligned}&U_1 \leftrightarrow Y_1 \leftrightarrow Y_2 \\ &U_1 \leftrightarrow Y_1 \leftrightarrow Y_2 \leftrightarrow U_2.\end{aligned} \tag{7}$$

The condition in (6) is sometimes called the "long Markov chain" [11] to contrast it with the "short Markov chains" in (7).

Proposition 3: The outer bound $\mathcal{RD}_o$ subsumes $\mathcal{RD}_o^{BT}$, i.e., $\mathcal{RD}_o \subseteq \mathcal{RD}_o^{BT}$.

Proof: First observe that for any $(\boldsymbol{U}, \boldsymbol{Z}, W, T)$ in Γ_o

$$U_\ell \leftrightarrow (Y_\ell, W, T) \leftrightarrow (Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1}), \quad \text{for each } \ell. \tag{8}$$

Now since (W, T) is independent of $(Y_0, \boldsymbol{Y}, Y_{L+1})$, we also have the following chain :

$$W \leftrightarrow (Y_\ell, T) \leftrightarrow (Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1}). \tag{9}$$

We will show that this pair of Markov chains implies that

$$U_\ell \leftrightarrow (Y_\ell, T) \leftrightarrow (Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1}), \quad \text{for each } \ell. \tag{10}$$

To see this, consider the quantity

$$I(U_\ell, W; Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1} | Y_\ell, T).$$

Using the chain rule, it can be decomposed in two ways

$$\begin{aligned} &I(U_\ell; Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1} | Y_\ell, T) + I(W; Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1} | Y_\ell, T, U_\ell) \\ &= I(W; Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1} | Y_\ell, T) + I(U_\ell; Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1} | Y_\ell, T, W). \end{aligned}$$

By (8) and (9), the right-hand side is zero. This implies that

$$I(U_\ell; Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1} | Y_\ell, T) = 0$$

which establishes (10). Thus $(\boldsymbol{U}, \boldsymbol{Z}, T)$ is in Γ_o^{BT}. From the definition of $\mathcal{RD}_o$, we see that if we choose $X = \boldsymbol{Y}$, then the mutual information expressions that define $\mathcal{RD}_o(X, \gamma)$ match those in $\mathcal{RD}_o^{BT}$. That is

$$\mathcal{RD}_o(\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{Z}, W, T) = \mathcal{RD}_o^{BT}(\boldsymbol{U}, \boldsymbol{Z}, T).$$

It follows that

$$\mathcal{RD}_o \subseteq \bigcup_{\gamma \in \Gamma_o} \mathcal{RD}_o(\boldsymbol{Y}, \gamma) \subseteq \bigcup_{\gamma \in \Gamma_o^{BT}} \mathcal{RD}_o^{BT}(\gamma) = \mathcal{RD}_o^{BT}. \quad \square$$

The proof reveals that $\mathcal{RD}_o$ improves upon $\mathcal{RD}_o^{BT}$ in two ways. The first is that $\mathcal{RD}_o$ allows for optimization over X while $\mathcal{RD}_o^{BT}$ effectively requires the choice $X = \boldsymbol{Y}$. The second is that Γ_o is "smaller" than Γ_o^{BT} in the sense that if $(\boldsymbol{U}, \boldsymbol{Z}, W, T)$ is in Γ_o then $(\boldsymbol{U}, \boldsymbol{Z}, T)$ is in Γ_o^{BT}. In particular, if there are two encoders and no hidden source Y_0, no side information Y_{L+1}, and no time-sharing variable T, then Γ_o^{BT} requires $\boldsymbol{U}$ to satisfy (7) while Γ_o requires the condition

$$U_1 \leftrightarrow (Y_1, W) \leftrightarrow (Y_2, W) \leftrightarrow U_2$$

for some W that is independent of $\boldsymbol{Y}$. Thus, Γ_o requires that $\boldsymbol{U}$ be a "mixture of long chains," whereas Γ_o^{BT} only requires that $\boldsymbol{U}$ satisfy the less-stringent short-chains condition.

The remainder of this section is devoted to showing that these improvements make the containment in Proposition 3 strict in some cases. As the reader will see, the first difference between the two outer bounds is entirely responsible for the gap that we expose between them in our examples. We hasten to add, however, that the second improvement is not empty in that Anantharam and Borkar [28] have shown that there can exist a $(\boldsymbol{U}, \boldsymbol{Z}, T)$ in Γ_o^{BT} with the property that there does not exist a W such that $(\boldsymbol{U}, \boldsymbol{Z}, W, T)$ is in Γ_o. It is interesting to note that the Anantharam–Borkar example arose independently of this work in the context of distributed stochastic control.

We will describe four examples for which $\mathcal{RD}_o^{BT}$ strictly contains $\mathcal{RD}_o$. In all four examples $\mathcal{RD}_o$ yields a conclusive result. The first is rather contrived and can be solved from first principles. It is included to illustrate the difference between the two bounds.

A. Toy Example

Let Y_{11}, Y_{12}, Y_{21}, and Y_{22} be i.i.d. random variables, uniformly distributed over $\{0, 1\}$. Consider two encoders ($L = 2$) with $Y_1 = (Y_{11}, Y_{12})$ and $Y_2 = (Y_{21}, Y_{22})$ (there is no hidden source or side information in this example). We have a single distortion constraint ($K = 1$) with $\mathcal{Z}_1 = \{0, 1\}^2$ and

$$d_1(\boldsymbol{Y}, Z_1) = \begin{cases} 0, & \text{if } Z_1 = (Y_{11}, Y_{21}) \text{ or } Z_1 = (Y_{12}, Y_{22}) \\ 1, & \text{otherwise.} \end{cases}$$

In words, the decoder attempts to guess either the first coordinate or the second coordinate of both encoders' observations. It incurs a distortion of zero if it guesses correctly the same coordinate of the two sources and one otherwise. Note that the decoder need not declare which coordinate it is attempting to guess.

One simple approach is to have both encoders always send, say, the first coordinate of their observations. This requires rate $\log 2$ for each encoder, and achieves a distortion of zero. It is not obvious that this scheme is optimal, however, because it does not make use of the flexibility afforded the decoder in choosing which coordinate to reconstruct. Using the outer bound, we can show that this scheme is indeed optimal.

Proposition 4: For this problem

$$\overline{\mathcal{RD}_o} \cap \{D_1 = 0\} \subseteq \{(R_1, R_2) : R_1 \geq \log 2, R_2 \geq \log 2\}.$$

Proof: Suppose (R_1, R_2, ϵ) is in $\mathcal{RD}_o$ and $\epsilon \leq 1/2$. Since Y_1 and Y_2 are independent, deterministic random variables are in χ. Thus, there exists γ in Γ_o such that

$$\begin{aligned} \epsilon &\geq E[d_1(\boldsymbol{Y}, Z_1)] \\ R_1 &\geq I(Y_1; U_1 | W, T) \\ R_2 &\geq I(Y_2; U_2 | W, T). \end{aligned}$$

By condition (ii) defining Γ_o

$$U_1 \leftrightarrow (Y_1, W, T) \leftrightarrow (Y_2, U_2). \tag{11}$$

Since Y_2 is independent of (Y_1, W, T) in this example, Y_2 must be independent of (Y_1, U_1, W, T). Thus

$$I(Y_1; U_1 | W, T) = I(Y_1; U_1 | W, T, Y_{21} \neq Y_{22}). \tag{12}$$

Likewise, Y_1 is independent of (Y_2, U_2, W, T) and hence given (W, T), Y_1 is independent of (Y_2, U_2). This observation combined with (11) implies $(Y_1, U_1) \leftrightarrow (W, T) \leftrightarrow (Y_2, U_2)$. In particular, $Y_1 \leftrightarrow (U_1, W, T) \leftrightarrow (Y_2, U_2)$. By condition (iii) defining Γ_o, $Y_1 \leftrightarrow (Y_2, U_1, U_2, W, T) \leftrightarrow (Y_2, Z_1)$. These last two chains imply that

$$Y_1 \leftrightarrow (U_1, W, T) \leftrightarrow (Y_2, Z_1).$$

Thus conditioned on (W,T) and the event $\{Y_{21} \neq Y_{22}\}$, we have $Y_1 \leftrightarrow U_1 \leftrightarrow (Y_2, Z_1)$. It follows that

$$\begin{aligned} & I(Y_1;U_1|W,T,Y_{21}\neq Y_{22}) \\ & \geq I(Y_1;Y_2,Z_1|W,T,Y_{21}\neq Y_{22}) \\ & = I(Y_1;Y_2,Z_1,d_1(\boldsymbol{Y},Z_1)|W,T,Y_{21}\neq Y_{22}) \\ & \quad - I(Y_1;d_1(\boldsymbol{Y},Z_1)|W,T,Y_2,Z_1,Y_{21}\neq Y_{22}) \\ & = H(Y_1|W,T,Y_{21}\neq Y_{22}) \\ & \quad - H(Y_1|Y_2,Z_1,d_1(\boldsymbol{Y},Z_1),W,T,Y_{21}\neq Y_{22}) \\ & \quad - H(d_1(\boldsymbol{Y},Z_1)|W,T,Y_2,Z_1,Y_{21}\neq Y_{22}), \end{aligned}$$

since $d_1(\boldsymbol{Y},Z_1)$ is a function of $\boldsymbol{Y}$ and Z_1. Next, observe that on the events $\{Y_{21}\neq Y_{22}\}$ and $\{d_1(\boldsymbol{Y},Z_1)=0\}$, Y_2 and Z_1 together must reveal one of the two bits of Y_1. Thus

$$H(Y_1|Y_2,Z_1,d_1(\boldsymbol{Y},Z_1)=0,W,T,Y_{21}\neq Y_{22}) \leq \log 2.$$

Continuing our chain of inequalities

$$\begin{aligned} & I(Y_1;U_1|W,T,Y_{21}\neq Y_{22}) \\ & \geq 2\log 2 - \log 2\cdot \Pr(d_1(\boldsymbol{Y},Z_1)=0|Y_{21}\neq Y_{22}) \\ & \quad - (2\log 2)\cdot \Pr(d_1(\boldsymbol{Y},Z_1)=1|Y_{21}\neq Y_{22}) \\ & \quad - H(d_1(\boldsymbol{Y},Z_1)|Y_{21}\neq Y_{22}) \\ & \geq \log 2 - (2\log 2)\cdot \Pr(d_1(\boldsymbol{Y},Z_1)=1|Y_{21}\neq Y_{22}) \\ & \quad - H(d_1(\boldsymbol{Y},Z_1)|Y_{21}\neq Y_{22}). \end{aligned} \tag{13}$$

Now

$$\begin{aligned} & \frac{1}{2}H(d_1(\boldsymbol{Y},Z_1)|Y_{21}\neq Y_{22}) + \frac{1}{2}H(d_1(\boldsymbol{Y},Z_1)|Y_{21}=Y_{22}) \\ & \quad = H(d_1(\boldsymbol{Y},Z_1)|1(Y_{21}=Y_{22})) \leq H(d_1(\boldsymbol{Y},Z_1)) \leq h(\epsilon) \end{aligned}$$

where, here and throughout, $h(\cdot)$ is the binary entropy function with natural logarithms. We conclude that

$$H(d_1(\boldsymbol{Y},Z_1)|Y_{21}\neq Y_{22}) \leq 2h(\epsilon).$$

Similarly

$$\Pr(d_1(\boldsymbol{Y},Z_1)=1|Y_{21}\neq Y_{22}) \leq 2\epsilon.$$

Substituting these two observations into (13) and recalling (12) yields

$$I(Y_1;U_1|W,T) \geq \log 2 - 4\epsilon\log 2 - 2h(\epsilon).$$

By symmetry, $I(Y_2;U_2|W,T)$ must satisfy the same inequality. This implies the desired conclusion. □

Since we know that $(\log 2, \log 2, 0)$ is achievable, it follows that

$$\begin{aligned} \overline{\mathcal{RD}_o} \cap \{D_1 = 0\} &= \overline{\mathcal{RD}_\star} \cap \{D_1 = 0\} \\ &= \{(R_1,R_2) : R_1 \geq \log 2, R_2 \geq \log 2\}. \end{aligned}$$

In particular, $\mathcal{RD}_o$ is tight in the zero-distortion limit. In contrast, we show next that the Berger–Tung outer bound is not.

Proposition 5: The unachievable point $((3/4)\log 2, (3/4)\log 2, 0)$ is contained in $\mathcal{RD}_o^{BT}$.

Proof: Let the random variable W be uniformly distributed over $\{1,2\}$, and let $U_1 = Y_{1W}$ and $U_2 = Y_{2W}$. Let $Z_1 = (U_1, U_2)$. It is straightforward to verify that (U_1,U_2,Z_1) is in Γ_o^{BT} (the time-sharing random variable T is not needed and can be taken to be constant). Next note that $E[d_1(\boldsymbol{Y},Z_1)] = 0$. Finally, one can compute

$$I(\boldsymbol{Y};U_1,U_2) = \frac{5}{4}\log 2$$

and

$$I(\boldsymbol{Y};U_1) = \frac{1}{2}\log 2.$$

This implies that

$$\begin{aligned} I(\boldsymbol{Y};U_1|U_2) = I(\boldsymbol{Y};U_2|U_1) &= I(\boldsymbol{Y};\boldsymbol{U}) - I(\boldsymbol{Y};U_1) \\ &= \frac{3}{4}\log 2. \end{aligned}$$

The conclusion follows. □

B. Binary Erasure CEO Problem

Here Y_0 is uniformly distributed over $\{-1,1\}$, and

$$Y_\ell = N_\ell \cdot Y_0, \quad \text{for } \ell \text{ in } \{1,\ldots,L\}$$

where $N_1,\ldots,N_L$ are i.i.d., independent of Y_0, and satisfy

$$\begin{aligned} 0 < \Pr(N_1 = 0) &= p < 1 \\ \Pr(N_1 = 1) &= 1-p. \end{aligned}$$

Let $\mathcal{Z}_1 = \{-1,0,1\}$. We will assume that there is no side information and that the decoder is only interested in reproducing the hidden process Y_0. We measure the fidelity of its reproduction using a family of distortion measures $\{d_1^\lambda\}_{\lambda>0}$, where

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

We are particularly interested in the large-λ limit. In this regime, d_1^λ approximates the "erasure distortion measure" [2, p. 338]

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

We use a finite approximation because an infinite distortion measure is unforgiving of decoding errors that have negligible probability.

This example is motivated by the following problem arising in energy-limited sensor networks. We seek to monitor a remote source Y_0. To this end, we deploy an array of sensors, each of which is capable of observing the source with negligible probability of error. To lengthen the lifetime of the network, each sensor spends a fraction p of the time in a low-power "sleep" state. We assume that the sensors cycle between the awake and sleep states independently of each other and on a faster time scale than the sampling; at each discrete time, each sensor sleeps with probability p, independently of the other sensors and the past. Sensors do not make any observations while they are asleep, resulting in erasures. We permit the coding process to introduce additional erasures, but not errors, yielding the erasure distortion measure. What sum rate is required in order for the decoder to reproduce a fraction $1-D$ of the Y_0^n

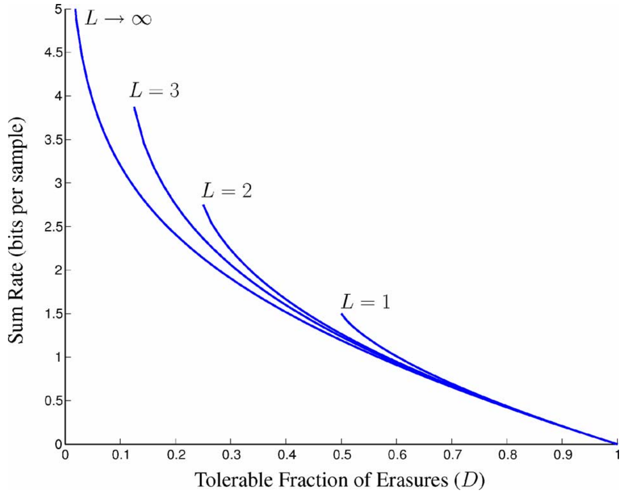

Fig. 4. Sum rate for the binary erasure CEO problem with $p = 1/2$ and $\lambda \to \infty$.

variables while almost never making an error? Of course, D must satisfy $D \geq p^L$.

Define

$$\mathcal{R}_\star(D,\lambda) = \inf\left\{\sum_{\ell=1}^{L} R_\ell : (R_1,\ldots,R_L,D) \in \overline{\mathcal{RD}_\star}(\lambda)\right\}$$

where $\overline{\mathcal{RD}_\star}(\lambda)$ is the rate–distortion region when the distortion measure is d_1^λ. We define $\mathcal{R}_o(D,\lambda)$ and $\mathcal{R}_i^{BT}(D,\lambda)$ analogously.

In Appendix A, we show that if $p^L \leq D \leq 1$, then

$$\lim_{\lambda\to\infty} \mathcal{R}_i^{BT}(D,\lambda) \leq (1-D)\log 2 + L\left[h\left(D^{1/L}\right) - (1-p)h\left(\frac{D^{1/L}-p}{1-p}\right)\right] \tag{14}$$

where $h(\cdot)$ again denotes the binary entropy function. In Appendix B, we show that the quantity on the right-hand side is also a lower bound to

$$\lim_{\lambda\to\infty} \mathcal{R}_o(D,\lambda).$$

This is accomplished by choosing $X = Y_0$. Hence, this expression must equal

$$\lim_{\lambda\to\infty} \mathcal{R}_\star(D,\lambda).$$

That is, the improved outer bound and the Berger–Tung inner bound together yield a conclusive result for the sum rate of the binary erasure CEO problem. In Appendix C, we show that $\overline{\mathcal{RD}_o^{BT}}$ contains points with a strictly smaller sum rate in general. Fig. 4 shows the correct sum rate for $p = 0.5$ and several values of L in the limit as $\lambda \to \infty$.

C. Quadratic Gaussian CEO Problem [20], [29], [13], [14]

We turn to a continuous example. Here $Y_0,\ldots,Y_L$ are jointly Gaussian and $Y_1,\ldots,Y_L$ are conditionally independent given Y_0. For $\ell \geq 1$, let us write $Y_\ell = Y_0 + N_\ell$, where $Y_0, N_1, \ldots, N_L$ are independent and

$$E\left[N_\ell^2\right] = \sigma_\ell^2 > 0, \quad \text{for all } \ell.$$

We will denote the variance of Y_0 by σ^2. Again there is no side information, and the decoder is only interested in reproducing the hidden process Y_0

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

The rate–distortion region for this problem was found by Oohama [21], [13] and Prabhakaran, Tse, and Ramchandran [14]. The two proofs are nearly the same, and build on earlier work of Oohama [29]. The primary contribution is the converse result, which uses the entropy power inequality [2, Theorem 17.7.3]. The Berger–Tung inner bound is used for achievability.

It is straightforward to extend Theorem 1 to this continuous setting. A statement of the continuous version is given in Appendix D, where we also use the techniques of Oohama [13] and Prabhakaran, Tse, and Ramchandran [14] to prove the following.

Proposition 6: For the Gaussian CEO problem,

$$\begin{aligned}\mathcal{RD}_o \subseteq \Bigg\{&(R_1,\ldots,R_L,D) \in \mathbb{R}_+^{L+1} :\\ &\text{there exists } (r_1,\ldots,r_L) \in \mathbb{R}_+^L : \text{ for all } A,\\ &\sum_{\ell\in A} R_\ell \geq \frac{1}{2}\log^+\left[\frac{1}{D}\left(\frac{1}{\sigma^2} + \sum_{\ell\in A^c}\frac{1-\exp(-2r_\ell)}{\sigma_\ell^2}\right)^{-1}\right]\\ &+ \sum_{\ell\in A} r_\ell\Bigg\}\end{aligned} \tag{15}$$

where $\log^+ x = \max(\log x, 0)$.

Since this expression equals $\overline{\mathcal{RD}_\star}$ [13], we conclude that $\mathcal{RD}_o$ is tight in this example. It also follows that the converse result of Oohama [13] and Prabhakaran, Tse, and Ramchandran [14] is a consequence of the outer bound provided in this paper. This does not imply, however, that the task of proving the converse result is made any easier by our bound. In fact, the proof that the outer bound is tight follows the same steps as the proof of the original converse, including the use of the entropy power inequality. But this is still an improvement over the Berger–Tung outer bound, which as we show in Appendix E contains points outside the rate–distortion region.

It should be mentioned that Oohama's [13] converse is actually more general than the result described here, in that Oohama permits one of the encoders to make noise-free observations (i.e., $\sigma_1^2 = 0$). Comparing Oohama's proof to Appendix D shows that the outer bound supplied in this paper also recovers this more general result.

D. Quadratic Gaussian Two-Encoder Problem With Separate Distortion Constraints [16], [17], [30], [11], [15]

In this problem, there are two sources ($L = 2$), Y_1 and Y_2, that are jointly Gaussian. The decoder attempts to reproduce each of the two sources individually

$$\begin{aligned} d_1(\boldsymbol{Y},Z_1) &= (Y_1 - Z_1)^2\\ d_2(\boldsymbol{Y},Z_2) &= (Y_2 - Z_2)^2.\end{aligned}$$

Recently, Wagner, Tavildar, and Viswanath [15] determined the rate region for this problem by showing that the Berger–Tung inner bound is tight. Although they do not explicitly use the outer bound described here, they do use the idea that underpins it, namely, augmenting the source to induce conditional independence. Partial characterizations of the rate region had been obtained earlier by Oohama [30] and Zamir and Berger [11].

Vamvatsikos [31] has shown that the outer bound is tight for this problem. The proof has two noteworthy aspects.

1) It utilizes an $X \in \chi$ that is not an existing component of the source (i.e., neither Y_1 nor Y_2). Thus, the proof requires that the source be augmented. This is different from the previous two examples, where the X that is used is already present in the source.
2) For the CEO problem, the intersection over $X \in \chi$ is not really needed in the sense that there exists a choice of X such that

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

For this problem, on the other hand, the proof requires one to intersect over several values of X to show that the outer bound is tight.

The reader is referred to Vamvatsikos [31] for the details of the proof. Tung [17, Appendix III] has shown that the Berger–Tung outer bound contains points that are now known to be outside the rate region.

IV. Recovery of Discrete Converse Results

Having seen that the new outer bound recovers the converse in two Gaussian examples, we show in this final section that it also recovers the converse for the discrete problems of Slepian and Wolf [1], Wyner [5], Ahlswede and Körner [6], Wyner and Ziv [7], Gel'fand and Pinsker [9], Berger and Yeung [10], and Gastpar [12]. The outer bound also recovers the converse result for the problem studied by Körner and Marton [8], although the proof of this fact is not as interesting. We shall therefore focus on the others. To recover these converse results, we shall use the following conclusive result for a special case of the problem.

Suppose that there exists a function $g : \mathcal{Y}_0 \mapsto \mathcal{Y}_0$ such that $Y_1, \ldots, Y_L$ are conditionally independent given $(g(Y_0), Y_{L+1})$. Also let

$$d_1(Y_0, \boldsymbol{Y}, Y_{L+1}, Z_1) = \begin{cases} 0, & \text{if } Z_1 = g(Y_0) \\ 1, & \text{otherwise.} \end{cases}$$

We make no other assumptions about the problem; in particular, the other distortion measures are arbitrary. We would like to characterize the set $\overline{\mathcal{RD}_\star} \cap \{D_1 = 0\}$. In words, conditioned on the side information and some function of the hidden variable, the observations are independent, and the function of the hidden variable must be reproduced losslessly (in addition to any other distortion constraints that are present). Note that $\overline{\mathcal{RD}_\star} \cap \{D_1 = 0\}$ will be empty unless $H(g(Y_0)|\boldsymbol{Y}, Y_{L+1}) = 0$. Gel'fand and Pinsker [9] refer to a similar condition as "completeness of observations."

Proposition 7: For this problem

$$\overline{\mathcal{RD}_o} \cap \{D_1 = 0\} = \overline{\mathcal{RD}_i^{BT}} \cap \{D_1 = 0\} = \overline{\mathcal{RD}_\star} \cap \{D_1 = 0\} \quad (16)$$

$$= \bigcup_{\gamma\in\Gamma_i^{BT}} \mathcal{RD}_i^{BT}(\gamma) \cap \{D_1 = 0\}. \quad (17)$$

Proof: To show (16), it suffices to show that $\overline{\mathcal{RD}_o} \cap \{D_1 = 0\}$ is contained in $\overline{\mathcal{RD}_i^{BT}} \cap \{D_1 = 0\}$. Suppose $(\boldsymbol{R}, \epsilon, D_2, \ldots, D_K)$ is a point in $\mathcal{RD}_o$ and $\epsilon \le 1/2$. By choosing $X = g(Y_0)$ in Definition 4, we see that there exists $(\boldsymbol{U}, \boldsymbol{Z}, W, \tilde{T})$ in Γ_o such that

$$\Pr(Z_1 \ne g(Y_0)) \le \epsilon$$
$$E[d_k(Y_0, \boldsymbol{Y}, Y_{L+1}, Z_k)] \le D_k, \quad \text{for all } k \ge 2$$

and for all A

$$\sum_{\ell\in A} R_\ell \ge I(g(Y_0); \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, \tilde{T}) + \sum_{\ell\in A} I(Y_\ell; U_\ell | g(Y_0), Y_{L+1}, W, \tilde{T}).$$

Now

$$\begin{aligned} I(g(Y_0); \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, \tilde{T}) &= H(g(Y_0)|\boldsymbol{U}_{A^c}, Y_{L+1}, \tilde{T}) - H(g(Y_0)|\boldsymbol{U}, Y_{L+1}, \tilde{T}) \\ &\ge H(g(Y_0)|\boldsymbol{U}_{A^c}, Y_{L+1}, W, \tilde{T}) - H(g(Y_0)|Z_1) \end{aligned}$$

where we have used the fact that

$$g(Y_0) \leftrightarrow Y_0 \leftrightarrow (\boldsymbol{U}, Y_{L+1}, \tilde{T}) \leftrightarrow Z_1.$$

By Fano's inequality [26, Lemma 1.3.8]

$$H(g(Y_0)|Z_1) \le h(\epsilon) + \epsilon \log(|\mathcal{Y}_0|).$$

Thus

$$\begin{aligned} I(g(Y_0); \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, \tilde{T}) &\ge H(g(Y_0)|\boldsymbol{U}_{A^c}, Y_{L+1}, W, \tilde{T}) - h(\epsilon) - \epsilon \log(|\mathcal{Y}_0|) \\ &\ge I(g(Y_0); \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, W, \tilde{T}) - h(\epsilon) - \epsilon \log(|\mathcal{Y}_0|). \end{aligned}$$

$$\begin{aligned} \sum_{\ell\in A} [R_\ell + h(\epsilon) + \epsilon \log(|\mathcal{Y}_0|)] &\ge I(g(Y_0); \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, W, \tilde{T}) \\ &\quad + \sum_{\ell\in A} I(Y_\ell; U_\ell | g(Y_0), Y_{L+1}, W, \tilde{T}) \\ &= I(g(Y_0); \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, W, \tilde{T}) \\ &\quad + I(\boldsymbol{Y}_A; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, g(Y_0), Y_{L+1}, W, \tilde{T}) \\ &= I(g(Y_0), \boldsymbol{Y}_A; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, W, \tilde{T}) \\ &\ge I(\boldsymbol{Y}_A; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, Y_{L+1}, W, \tilde{T}). \end{aligned}$$

If we now define $T = (W, \tilde{T})$, then $(\boldsymbol{U}, \boldsymbol{Z}, T)$ is in Γ_i^{BT} and the point

$$(R_1 + h(\epsilon) + \epsilon \log(|\mathcal{Y}_0|), \ldots, R_L + h(\epsilon) + \epsilon \log(|\mathcal{Y}_0|), \epsilon, D_2, \ldots, D_K)$$

is in $\mathcal{RD}_i^{BT}$. This implies that

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

which proves (16). To prove (17), it suffices to show that $\mathcal{RD}_i^{BT}$ is closed. This is shown in Appendix F. $\square$

The differences between this result and that of Gel'fand and Pinsker [9] are numerous but minor. Gel'fand and Pinsker's formulation does not include side information at the decoder or distortion constraints beyond the one on Y_0. They also require that Y_0, instead of $g(Y_0)$, be reproduced at the decoder. The region given here reduces to theirs when these extensions are ignored. Thus, this result seems to be a generalization of theirs, albeit a trivial one, since their proof can be modified to handle these extensions. A closer comparison, however, reveals that they define the rate region more stringently than we do here. Thus, our result does not recover theirs, strictly speaking, although it does recover the converse component of their result since our definitions are weaker.

The reason for including side information and additional distortion constraints in the model is that they enable us to also recover the converse results of the other problems mentioned earlier. For instance, Gastpar [12] considers the problem of reproducing the individual observations, subject to separate distortion constraints, under the assumption that the decoder is provided with side information that makes the observations conditionally independent. His converse result can be recovered by taking Y_0 to be constant. It is easily verified that, under this condition, our region coincides with his. The classical Wyner–Ziv problem [7] can be viewed as Gastpar's problem with a single encoder ($L = 1$). So that converse result is recovered as well.

Berger and Yeung [10] solve the two-encoder problem in which the individual observations are to be reproduced, with at least one of the two being reproduced losslessly. In our notation, this corresponds to setting $L = 2$, $Y_1 = Y_0$, and $g(Y_0) = Y_0$. Note that our conditional independence requirement trivially holds in this case.

To see that under these assumptions, our region reduces to theirs, suppose $(R_1, R_2, D_2) \in \mathcal{RD}_i^{BT}(\gamma) \cap \{D_1 = 0\}$ for some $\gamma \in \Gamma_i^{BT}$. Then

$$R_1 \geq I(Y_1; U_1 | U_2, T) = H(Y_1 | U_2, T). \tag{18}$$

Also

$$\begin{aligned} R_2 &\geq I(Y_2; U_2 | U_1, T) \geq I(Y_2; U_2 | U_1, Y_1, T) \\ &= I(Y_2; U_2 | Y_1, T) \end{aligned} \tag{19}$$

where we have used the fact that

$$U_2 \leftrightarrow (Y_2, U_1, T) \leftrightarrow Y_1$$

and

$$U_1 \leftrightarrow (Y_1, T) \leftrightarrow (Y_2, U_2)$$

(see Cover and Thomas [2, p. 35]). Finally

$$\begin{aligned} R_1 + R_2 &\geq I(Y_1, Y_2; U_1, U_2 | T) \\ &\geq I(Y_1; U_1, U_2 | T) + I(Y_2; U_1, U_2 | Y_1, T) \\ &\geq H(Y_1) + I(Y_2; U_2 | Y_1, T). \end{aligned} \tag{20}$$

It is now evident that the two regions are identical (cf. [10, p. 230]). Thus, the converse result of Berger and Yeung is a consequence of the outer bound provided here.

The classical problem of source coding with side information [5], [6] can be viewed as a special case of the Berger–Yeung problem in which D_2 exceeds the maximum value of d_2, the distortion measure for Y_2. Berger and Yeung demonstrate how, under this assumption, the region described above reduces to the one given by Wyner [5] and Ahlswede and Körner [6]. *Ipso facto*, the converse result for this problem is also recovered.

Finally, we return to the result of Slepian and Wolf [1]. Here the aim is to losslessly reproduce all of the observations. For two encoders ($L = 2$), this can be viewed as a special case of the problem of Berger and Yeung. These authors show how the region described in (18)–(20) reduces to the one given at the beginning of the paper. The result for more than two encoders can be viewed as a special case of Proposition 7 in which $Y_0 = \boldsymbol{Y}$ and $g(Y_0) = Y_0$. In this case, if $\boldsymbol{R} \in \overline{\mathcal{RD}_o} \cap \{D_1 = 0\}$, then for any A

$$\begin{aligned} \sum_{\ell \in A} R_\ell &\geq I(\boldsymbol{Y}_A; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, T) \\ &= H(\boldsymbol{Y}_A | \boldsymbol{U}_{A^c}, T) \\ &\geq H(\boldsymbol{Y}_A | \boldsymbol{Y}_{A^c}, T) \end{aligned}$$

since $\boldsymbol{Y}_A \leftrightarrow (\boldsymbol{Y}_{A^c}, T) \leftrightarrow (\boldsymbol{U}_{A^c}, T)$. Now $\boldsymbol{Y}$ is independent of T, so

$$\sum_{\ell \in A} R_\ell \geq H(\boldsymbol{Y}_A | \boldsymbol{Y}_{A^c})$$

which is the well-known rate region for this problem. Thus, the converse of Slepian and Wolf is also recovered. For this result, as with the others, our outer bound dispenses with the need to prove a custom converse coding theorem. In fact, Proposition 7 can be viewed as unifying all of the results in this discussion.

Appendix A
Sum-Rate Achievability for the Binary Erasure CEO Problem

Showing that a particular rate–distortion vector is achievable using the Berger–Tung inner bound is mostly a matter of finding the proper "test channels" $Y_\ell \to U_\ell$ for the encoders. To prove (14), we use binary erasure test channels that are identically distributed across the encoders. In this appendix and the next two, the notation is drawn from Section III-C.

Lemma 1: For any $p^L \leq D \leq 1$

$$\begin{aligned} &\lim_{\lambda \to \infty} \mathcal{R}_i^{BT}(D, \lambda) \\ &\leq (1 - D) \log 2 + L \left[h\left(D^{1/L}\right) - (1-p) h\left(\frac{D^{1/L} - p}{1 - p}\right) \right]. \end{aligned}$$

Proof: Fix D and let $\tilde{N}_1, \ldots, \tilde{N}_L$ be i.i.d., independent of $Y_0, \ldots, Y_L$, with

$$\begin{aligned} \Pr\left(\tilde{N}_1 = 0\right) &= \frac{D^{1/L} - p}{1 - p} \\ \Pr\left(\tilde{N}_1 = 1\right) &= 1 - \frac{D^{1/L} - p}{1 - p}. \end{aligned}$$

For ℓ in $\{1, \ldots, L\}$, let $U_\ell = Y_\ell \cdot \tilde{N}_i$. Then let

$$Z_1 = \operatorname{sgn}\left(\sum_{\ell=1}^{L} U_\ell\right) := \begin{cases} -1, & \text{if } \sum_{\ell=1}^{L} U_\ell < 0 \\ 0, & \text{if } \sum_{\ell=1}^{L} U_\ell = 0 \\ 1, & \text{otherwise.} \end{cases}$$

Then for all λ

$$E\left[d_1^\lambda(Y_0,\boldsymbol{Y},Z_1)\right] = \Pr(U_\ell = 0 \text{ for all } \ell) = \left[\Pr(U_1=0)\right]^L = D.$$

Thus, $(\boldsymbol{R},D)$ is contained in $\mathcal{RD}_i^{BT}$ for all λ if for all $A \subseteq \{1,\ldots,L\}$

$$\sum_{\ell\in A} R_\ell \geq I(\boldsymbol{Y}_A;\boldsymbol{U}_A|\boldsymbol{U}_{A^c}). \tag{21}$$

The rate vectors satisfying this collection of inequalities are known to form a contrapolymatroid [32], [27]. As such, there exist rate vectors $\boldsymbol{R}$ satisfying (21) such that

$$\sum_{\ell=1}^L R_\ell = I(\boldsymbol{Y};\boldsymbol{U}).$$

In particular, this holds for any vertex of (21) [32], [27]. Now

$$\begin{aligned} I(\boldsymbol{Y};\boldsymbol{U}) &= I(Y_0,\boldsymbol{Y};\boldsymbol{U}) \\ &= I(Y_0;\boldsymbol{U}) + I(\boldsymbol{Y};\boldsymbol{U}|Y_0) \\ &= I(Y_0;\boldsymbol{U}) + L\cdot I(Y_1;U_1|Y_0). \end{aligned}$$

But $I(Y_0;\boldsymbol{U}) = (1-D)\log 2$ and

$$\begin{aligned} I(Y_1;U_1|Y_0) &= H(U_1|Y_0) - H(U_1|Y_1) \\ &= h(D^{1/L}) - (1-p)h\left(\frac{D^{1/L}-p}{1-p}\right). \end{aligned}$$

Then for any λ, there exist vectors $(\boldsymbol{R},D)$ in $\mathcal{RD}_i^{BT}(\lambda)$ such that

$$\sum_{\ell=1}^L R_\ell = (1-D)\log 2 + L\left[h(D^{1/L}) - (1-p)h\left(\frac{D^{1/L}-p}{1-p}\right)\right].$$

The conclusion follows. □

Appendix B
Sum-Rate Converse for the Binary Erasure CEO Problem

We evaluate the outer bound's sum-rate constraint for the binary erasure CEO problem via a sequence of lemmas. Throughout this appendix, $g(\cdot)$ will denote the function on $[p,\infty)$ defined by

$$g(x) = \begin{cases} h(x) - (1-p)h(\frac{x-p}{1-p}), & p \leq x \leq 1 \\ 0, & x > 1. \end{cases}$$

We begin by proving several facts about $g(\cdot)$. For this, the following calculations are useful.

Lemma 2: For all x in $(\log p, 0]$

$$e^x\log(e^x-p) - xe^x \leq -p \tag{22}$$

and

$$e^x\log(e^x-p) - e^x(x+1) + \frac{e^{2x}}{e^x-p} \geq 0. \tag{23}$$

Proof: It is well known that

$$\log\left(\frac{1}{1-z}\right) \geq z, \quad \text{for all } z<1.$$

Replacing z with pe^{-x} and rearranging yields (22). To see (23), note that (22) implies that the first derivative of

$$(e^x-p)\log(e^x-p) - (e^x-p)(x+1) + e^x \tag{24}$$

is nonpositive on $(\log p, 0]$. Since the function in (24) is nonnegative at $x=0$, it follows that

$$(e^x-p)\log(e^x-p) - (e^x-p)(x+1) + e^x \geq 0 \tag{25}$$

for all x in $(\log p, 0]$. One can now obtain (23) by multiplying both sides by e^x and dividing both sides by (e^x-p). □

Lemma 3: The function $g(e^x)$ is nonincreasing and convex as a function of x on $[\log p,\infty)$.

Proof: The first derivative of $g(e^x)$ on $(\log p, 0)$ is

$$e^x\log(e^x-p) - xe^x.$$

This observation, the first conclusion of Lemma 2, and the continuity of $g(\cdot)$ together imply that $g(e^x)$ is nonincreasing on $[\log p, 0]$. Since $g(e^x)$ is constant on $[0,\infty)$, it follows that $g(e^x)$ is nonincreasing on $[\log p,\infty)$. The second derivative of $g(e^x)$ on $(\log p, 0)$ is

$$e^x\log(e^x-p) - e^x(x+1) + \frac{e^{2x}}{e^x-p}.$$

This observation, the second conclusion of Lemma 2, and the continuity of $g(\cdot)$ together imply that $g(e^x)$ is convex on $[\log p, 0]$. Since $g(e^x)$ is nonincreasing on $[\log p,\infty)$ and constant on $[0,\infty)$, it follows that $g(e^x)$ is convex on $[\log p,\infty)$. □

Corollary 1: The function $g(y^{1/L})$ is nonincreasing and convex in y on $[p^L,\infty)$.

Proof: $g(y^{1/L}) = g(e^x)$ with $x = (1/L)\log y$, and $g(e^x)$ is convex and nonincreasing while $(1/L)\log(\cdot)$ is concave and nondecreasing. □

The next lemma is central to our evaluation of the outer bound's sum rate. Note that condition (i) in the hypothesis implies that $\Pr(Y_0\cdot Z_1 < 0) = 0$. That is, the reproduction Z_1 is never in error (although it may be an erasure).

Lemma 4: Suppose $p^L \leq D$ and $(\boldsymbol{U},Z_1)$ is such that
(i) $E[d_1^\lambda(Y_0,Z_1)] \leq D$ for all λ;
(ii) $U_\ell \leftrightarrow Y_\ell \leftrightarrow (Y_0,\boldsymbol{Y}_{\ell^c},\boldsymbol{U}_{\ell^c})$ for all ℓ; and
(iii) $(Y_0,\boldsymbol{Y}) \leftrightarrow \boldsymbol{U} \leftrightarrow Z_1$.
Then

$$\frac{1}{L}\sum_{\ell=1}^L I(Y_\ell;U_\ell|Y_0) \geq g(D^{1/L}).$$

Proof: For each encoder ℓ, let

$$\begin{aligned} A_{\ell,+} &= \{u\in\mathcal{U}_\ell : \Pr(U_\ell=u|Y_0=1) > 0\} \\ A_{\ell,-} &= \{u\in\mathcal{U}_\ell : \Pr(U_\ell=u|Y_0=-1) > 0\}. \end{aligned}$$

Then define

$$\tilde{U}_\ell = \begin{cases} 1 & \text{if } U_\ell \in A_{\ell,+} \backslash A_{\ell,-} \\ -1 & \text{if } U_\ell \in A_{\ell,-} \backslash A_{\ell,+} \\ 0 & \text{otherwise.} \end{cases}$$

Finally, let

$$\delta_{\ell,+} = \Pr(\tilde{U}_\ell = 0 | Y_\ell = 1)$$
$$\delta_{\ell,-} = \Pr(\tilde{U}_\ell = 0 | Y_\ell = -1).$$

Then

$$\begin{aligned} \frac{1}{L}\sum_{\ell=1}^{L} I(Y_\ell; U_\ell | Y_0) &\geq \frac{1}{L}\sum_{\ell=1}^{L} I(Y_\ell; \tilde{U}_\ell | Y_0) \\ &= \frac{1}{L}\sum_{\ell=1}^{L}\left[H(\tilde{U}_\ell | Y_0) - H(\tilde{U}_\ell | Y_\ell)\right] \\ &= \frac{1}{L}\sum_{\ell=1}^{L}\left[\frac{1}{2}h\left(p + (1-p)\delta_{\ell,+}\right) + \frac{1}{2}h\left(p + (1-p)\delta_{\ell,-}\right)\right. \\ &\qquad \left. - \frac{1}{2}(1-p)h(\delta_{\ell,+}) - \frac{1}{2}(1-p)h(\delta_{\ell,-})\right]. \end{aligned}$$

Since $Y_0 \cdot Z_1 \geq 0$ a.s., on the event $Z_1 = 1$ we must have $Y_0 = 1$ and hence $U_\ell \in A_{\ell,+}$ for all ℓ. In addition, the condition $Y_0 \leftrightarrow \boldsymbol{U} \leftrightarrow Z_1$ dictates that when $Z_1 = 1$ we must have $U_\ell \in A_{\ell,+} \backslash A_{\ell,-}$ for some ℓ, for otherwise we would have $\Pr(Y_0 \cdot Z_1 = -1) > 0$. All of this implies that $\mathrm{sgn}(\sum_{\ell=1}^{L} \tilde{U}_\ell) = 1$ on the event that $Z_1 = 1$. Similarly, $\mathrm{sgn}(\sum_{\ell=1}^{L} \tilde{U}_\ell) = -1$ on the event $Z_1 = -1$. Thus, $\mathrm{sgn}(\sum_{\ell=1}^{L} \tilde{U}_\ell) = 0$ implies that $Z_1 = 0$, so

$$\Pr\left(\mathrm{sgn}\left(\sum_{\ell=1}^{L}\tilde{U}_\ell\right) = 0\right) \leq \Pr(Z_1 = 0) \leq D.$$

This implies that

$$\frac{1}{2}\prod_{\ell=1}^{L}(p + (1-p)\delta_{\ell,+}) + \frac{1}{2}\prod_{\ell=1}^{L}(p + (1-p)\delta_{\ell,-}) \leq D.$$

Thus

$$\begin{aligned} &\frac{1}{L}\sum_{\ell=1}^{L} I(Y_\ell; U_\ell | Y_0) \\ &\geq \inf\left\{\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{2}\Big[h(p + (1-p)\delta_{\ell,+}) - (1-p)h(\delta_{\ell,+})\right. \\ &\quad + h(p + (1-p)\delta_{\ell,-}) - (1-p)h(\delta_{\ell,-})\Big] : \delta_{\ell,+}, \delta_{\ell,-} \in [0,1] \\ &\qquad\qquad \text{for all } \ell \text{ and} \\ &\quad \left.\frac{1}{2}\prod_{\ell=1}^{L}(p + (1-p)\delta_{\ell,+}) + \frac{1}{2}\prod_{\ell=1}^{L}(p + (1-p)\delta_{\ell,-}) \leq D\right\}. \end{aligned}$$

This optimization problem is not convex, but if we change variables to

$$\Delta_{\ell,+} = \log(p + (1-p)\delta_{\ell,+})$$
$$\Delta_{\ell,-} = \log(p + (1-p)\delta_{\ell,-})$$

then it can be rewritten as

$$\begin{aligned} &\inf\left\{\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{2}\left[h\left(e^{\Delta_{\ell,+}}\right) - (1-p)h\left(\frac{e^{\Delta_{\ell,+}} - p}{1-p}\right)\right.\right. \\ &\qquad\qquad \left. + h\left(e^{\Delta_{\ell,-}}\right) - (1-p)h\left(\frac{e^{\Delta_{\ell,-}} - p}{1-p}\right)\right] : \\ &\quad \Delta_{\ell,+}, \Delta_{\ell,-} \in [\log p, 0] \text{ for all } \ell \text{ and} \\ &\quad \left.\frac{1}{2}\exp\left(\sum_{\ell=1}^{L}\Delta_{\ell,+}\right) + \frac{1}{2}\exp\left(\sum_{\ell=1}^{L}\Delta_{\ell,-}\right) \leq D\right\} \\ &= \inf\left\{\frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{2}\left[g\left(e^{\Delta_{\ell,+}}\right) + g\left(e^{\Delta_{\ell,-}}\right)\right] :\right. \\ &\quad \Delta_{\ell,+}, \Delta_{\ell,-} \in [\log p, 0] \text{ for all } \ell \text{ and} \\ &\quad \left.\frac{1}{2}\exp\left(\sum_{\ell=1}^{L}\Delta_{\ell,+}\right) + \frac{1}{2}\exp\left(\sum_{\ell=1}^{L}\Delta_{\ell,-}\right) \leq D\right\} \end{aligned}$$

which is convex by Lemma 3. Thus, we may assume without loss of optimality that

$$\Delta_{1,+} = \Delta_{2,+} = \cdots = \Delta_{L,+} =: \Delta_+$$

and

$$\Delta_{1,-} = \Delta_{2,-} = \cdots = \Delta_{L,-} =: \Delta_-.$$

This gives

$$\begin{aligned} &\frac{1}{L}\sum_{\ell=1}^{L} I(Y_\ell; U_\ell | Y_0) \\ &\geq \inf\left\{\frac{1}{2}g\left(e^{\Delta_+}\right) + \frac{1}{2}g\left(e^{\Delta_-}\right) : \Delta_+, \Delta_- \in [\log p, 0] :\right. \\ &\qquad\qquad \left.\frac{1}{2}e^{L\Delta_+} + \frac{1}{2}e^{L\Delta_-} \leq D\right\} \\ &= \inf\left\{g\left(e^{\Delta}\right) : \Delta \in [\log p, 0] : e^{L\Delta} \leq D\right\} \\ &\geq g(D^{1/L}) \end{aligned}$$

where we have used Lemma 3 again. □

The quantity $I(Y_\ell; U_\ell | Y_0)$ can be interpreted as the amount of information per symbol that the ℓth encoder sends about the erasure pattern of its observation. Lemma 4 then says that if a fraction D of the output symbols is allowed to be erased and no errors are allowed, then the amount of information that the average encoder must send about its erasure pattern is at least $g(D^{1/L})$ nats per symbol. We would like to extend this last assertion to allow "few" decoding errors instead of none. To this end, we will employ the following cardinality bound on the alphabet sizes of the auxiliary random variables $U_1, \ldots, U_L$.

Lemma 5: Let $(\boldsymbol{U}, Z_1)$ be such that
(i) $U_\ell \leftrightarrow Y_\ell \leftrightarrow (Y_0, \boldsymbol{Y}_{\ell^c}, \boldsymbol{U}_{\ell^c})$ for all ℓ, and
(ii) $(Y_0, \boldsymbol{Y}) \leftrightarrow \boldsymbol{U} \leftrightarrow Z_1$.

Then for any λ, there exist alternate random variables $\tilde{\boldsymbol{U}}$ and $\tilde{Z}_1$ also satisfying (i) and (ii) such that

$$E[d_1^\lambda(Y_0, \tilde{Z}_1)] \leq E[d_1^\lambda(Y_0, Z_1)]$$
$$I(Y_\ell; \tilde{U}_\ell | Y_0) = I(Y_\ell; U_\ell | Y_0) \text{ for all } \ell,$$

and

$$|\mathcal{U}_\ell| \le 4 \text{ for all } \ell.$$

See Csiszár and Körner [26, Theorem 3.4.6] or Lemma 11 to follow for proofs of similar results. The next lemma is the desired extension of Lemma 4.

Lemma 6: Suppose $p^L \le D$ and $(\boldsymbol{U}, Z_1)$ is such that
(i) $E[d_1^\lambda(Y_0, Z_1)] \le D$;
(ii) $U_\ell \leftrightarrow Y_\ell \leftrightarrow (Y_0, \boldsymbol{Y}_{\ell^c}, \boldsymbol{U}_{\ell^c})$ for all ℓ; and
(iii) $(Y_0, \boldsymbol{Y}) \leftrightarrow \boldsymbol{U} \leftrightarrow Z_1$.
If

$$\frac{32L}{p(1-p)}\left(\frac{2D}{\lambda}\right)^{1/L} \le \delta \le \frac{1}{2}$$

then

$$\frac{1}{L}\sum_{\ell=1}^{L} I(Y_\ell; U_\ell|Y_0) \ge g\left((D+\delta)^{1/L}\right) + 2\delta \log\frac{\delta}{5}.$$

Proof: By Lemma 5, we may assume that $\mathcal{U}_\ell = \{1,\dots,4\}$ for each ℓ. We may also assume that Z_1 is a deterministic function of $\boldsymbol{U}$: $Z_1 = \phi(\boldsymbol{U})$. Define

$$A_{\ell,+} = \left\{u_\ell \in \mathcal{U}_\ell : \exists\, \boldsymbol{u}_{\ell^c} : \phi(\boldsymbol{u}) = 1 \text{ and } \Pr(U_\ell = u_\ell|Y_0 = -1) = \min_{j\in\{1,\dots,L\}} \Pr(U_j = u_j|Y_0 = -1)\right\}$$

$$A_{\ell,-} = \left\{u_\ell \in \mathcal{U}_\ell : \exists\, \boldsymbol{u}_{\ell^c} : \phi(\boldsymbol{u}) = -1 \text{ and } \Pr(U_\ell = u_\ell|Y_0 = 1) = \min_{j\in\{1,\dots,L\}} \Pr(U_j = u_j|Y_0 = 1)\right\}.$$

We now define random variables $(\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{Z}})$ to replace $(\boldsymbol{U}, \boldsymbol{Z})$. The replacements will be close to the originals in distribution but will have the property that $\Pr(Y_1 \cdot \tilde{Z}_1 < 0) = 0$. That is, $\tilde{Z}_1$ will never be in error. Set $\tilde{\mathcal{U}}_\ell = \{1,\dots,5\}$ for each ℓ, and let

$$\Pr(\tilde{U}_\ell = i|Y_\ell = 1) = \begin{cases} 0, & \text{if } i \in A_{\ell,-} \\ \Pr(U_\ell \in A_{\ell,-}|Y_\ell = 1), & \text{if } i = 5 \\ \Pr(U_\ell = i|Y_\ell = 1), & \text{otherwise} \end{cases}$$

$$\Pr(\tilde{U}_\ell = i|Y_\ell = -1) = \begin{cases} 0, & \text{if } i \in A_{\ell,+} \\ \Pr(U_\ell \in A_{\ell,+}|Y_\ell = -1), & \text{if } i = 5 \\ \Pr(U_\ell = i|Y_\ell = -1), & \text{otherwise} \end{cases}$$

$$\Pr(\tilde{U}_\ell = i|Y_\ell = 0) = \begin{cases} 0, & \text{if } i \in A_{\ell,+} \cup A_{\ell,-} \\ \Pr(U_\ell \in A_{\ell,+} \cup A_{\ell,-}|Y_\ell = 0), & \text{if } i = 5 \\ \Pr(U_\ell = i|Y_\ell = 0), & \text{otherwise.} \end{cases}$$

Then define

$$\tilde{Z}_1 = \tilde{\phi}(\tilde{\boldsymbol{U}}) := \begin{cases} \phi(\tilde{u}), & \text{if } \tilde{u}_\ell \le 4 \text{ for all } \ell \\ 0, & \text{otherwise.} \end{cases}$$

There is a natural way of coupling $\tilde{\boldsymbol{U}}$ to $\boldsymbol{U}$ such that if $\tilde{U}_\ell$ is in $\{1,\dots,4\}$ then $\tilde{U}_\ell = U_\ell$. With this coupling in mind, it is evident that

$$\begin{aligned} E[d_1^\lambda(Y_0, \tilde{Z}_1)] &= E[d_1^\lambda(Y_0, \tilde{Z}_1)1(\max(\tilde{U}_1,\dots,\tilde{U}_L) \le 4)] \\ &\quad + E[d_1^\lambda(Y_0, \tilde{Z}_1)1(\max(\tilde{U}_1,\dots,\tilde{U}_L) = 5)] \\ &\le E[d_1^\lambda(Y_0, Z_1)1(\max(\tilde{U}_1,\dots,\tilde{U}_L) \le 4)] \\ &\quad + \Pr(\max(\tilde{U}_1,\dots,\tilde{U}_L) = 5) \\ &\le D + \Pr(\max(\tilde{U}_1,\dots,\tilde{U}_L) = 5). \end{aligned}$$

Now for any ℓ in $\{1,\dots,L\}$

$$\begin{aligned} \Pr(\tilde{U}_\ell = 5) &= \frac{1-p}{2}\Pr(U_\ell \in A_{\ell,-}|Y_\ell = 1) \\ &\quad + \frac{1-p}{2}\Pr(U_\ell \in A_{\ell,+}|Y_\ell = -1) \\ &\quad + p\Pr(U_\ell \in A_{\ell,+} \cup A_{\ell,-}|Y_\ell = 0). \end{aligned}$$

By the union bound, this is upper-bounded by

$$\left[\frac{1-p}{2}\Pr(U_\ell \in A_{\ell,-}|Y_\ell = 1) + p\Pr(U_\ell \in A_{\ell,-}|Y_\ell = 0)\right] + \left[\frac{1-p}{2}\Pr(U_\ell \in A_{\ell,+}|Y_\ell = -1) + p\Pr(U_\ell \in A_{\ell,+}|Y_\ell = 0)\right].$$

Since $U_\ell \leftrightarrow Y_\ell \leftrightarrow Y_0$

$$\begin{aligned} \Pr(\tilde{U}_\ell = 5) &\le \left[\frac{1-p}{2}\Pr(U_\ell \in A_{\ell,-}|Y_\ell = 1, Y_0 = 1) \right. \\ &\quad \left. + p\Pr(U_\ell \in A_{\ell,-}|Y_\ell = 0, Y_0 = 1)\right] \\ &\quad + \left[\frac{1-p}{2}\Pr(U_\ell \in A_{\ell,+}|Y_\ell = -1, Y_0 = -1) \right. \\ &\quad \left. + p\Pr(U_\ell \in A_{\ell,+}|Y_\ell = 0, Y_0 = -1)\right] \\ &\le \Pr(U_\ell \in A_{\ell,-}|Y_0 = 1) + \Pr(U_\ell \in A_{\ell,+}|Y_0 = -1). \end{aligned}$$

However

$$\sum_{\boldsymbol{u}:\phi(\boldsymbol{u})=1} \frac{1}{2}\prod_{j=1}^{L}\Pr(U_j = u_j|Y_0 = -1)\lambda \le D$$

which implies that for any ℓ

$$\sum_{\boldsymbol{u}:\phi(\boldsymbol{u})=1, u_\ell \in A_{\ell,+}} \prod_{j=1}^{L}\Pr(U_j = u_j|Y_0 = -1) \le \frac{2D}{\lambda}. \tag{26}$$

By the definition of $A_{\ell,+}$, for each $u_\ell \in A_{\ell,+}$, there exists at least one $\boldsymbol{u}_{\ell^c}$ such that $\phi(\boldsymbol{u}) = 1$ and

$$\prod_{j=1}^{L}\Pr(U_j = u_j|Y_0 = -1) \ge \Pr(U_\ell = u_\ell|Y_0 = -1)^L.$$

Together with (26), this implies

$$\sum_{u_\ell \in A_{\ell,+}} \Pr(U_\ell = u_\ell|Y_0 = -1)^L \le \frac{2D}{\lambda}.$$

Applying Hölder's inequality [33, p. 121] then gives

$$\begin{aligned}\Pr(U_\ell \in A_{\ell,+}|Y_0=-1) &= \sum_{u_\ell \in A_{\ell,+}} \Pr(U_\ell = u_\ell|Y_0=-1)\\ &\le \left[\sum_{u_\ell \in A_{\ell,+}} \Pr(U_\ell = u_\ell|Y_0=-1)^L\right]^{1/L} \cdot |A_{\ell,+}|^{(L-1)/L}\\ &\le 4\left(\frac{2D}{\lambda}\right)^{1/L}.\end{aligned}$$

Likewise

$$\Pr(U_\ell \in A_{\ell,-}|Y_0=1) \le 4\left(\frac{2D}{\lambda}\right)^{1/L}.$$

Thus

$$\Pr(\tilde{U}_\ell = 5) \le 8\left(\frac{2D}{\lambda}\right)^{1/L}.$$

By the union bound, it follows that

$$\Pr(\max(\tilde{U}_1,\dots,\tilde{U}_L)=5) \le 8L\left(\frac{2D}{\lambda}\right)^{1/L}$$

and therefore

$$E[d_1^\lambda(Y_0,\tilde{Z}_1)] \le D + 8L\left(\frac{2D}{\lambda}\right)^{1/L} \le D+\delta.$$

Note that $\tilde{Z}_1 = 1$ only if $\tilde{U}_\ell$ is in $A_{\ell,+}$ for some ℓ, and

$$\Pr(\tilde{U}_\ell \in A_{\ell,+}|Y_0=-1) = 0, \quad \text{for all } \ell.$$

Thus, $\Pr(Y_0 = -1, \tilde{Z}_1 = 1) = 0$ and similarly, $\Pr(Y_0 = 1, \tilde{Z}_1 = -1) = 0$. It follows from Lemma 4 that

$$\frac{1}{L}\sum_{\ell=1}^{L} I(Y_\ell;\tilde{U}_\ell|Y_0) \ge g\left((D+\delta)^{1/L}\right). \tag{27}$$

The remainder of the proof is devoted to showing that $I(Y_\ell;\tilde{U}_\ell|Y_0)$ is close to $I(Y_\ell;U_\ell|Y_0)$. For this we use the decomposition

$$I(Y_\ell;\tilde{U}_\ell|Y_0) = H(\tilde{U}_\ell|Y_0) - H(\tilde{U}_\ell|Y_\ell).$$

Observe that

$$\Pr(\tilde{U}_\ell = 5|Y_\ell=0)p \le \Pr(\tilde{U}_\ell=5) \le 8\left(\frac{2D}{\lambda}\right)^{1/L}.$$

Thus

$$\Pr(\tilde{U}_\ell = 5|Y_\ell = 0) \le \frac{\delta}{2}.$$

Similarly

$$\Pr(\tilde{U}_\ell = 5|Y_\ell = 1) \le \frac{\delta}{2}$$

and

$$\Pr(\tilde{U}_\ell = 5|Y_\ell = -1) \le \frac{\delta}{2}.$$

Therefore, if we view U_ℓ as a random variable on $\{1,\dots,5\}$, for any i in $\{-1,0,1\}$

$$\sum_{j=1}^{5} |\Pr(\tilde{U}_\ell = j|Y_\ell=i) - \Pr(U_\ell=j|Y_\ell=i)| = 2\Pr(\tilde{U}_\ell=5|Y_\ell=i) \le \delta.$$

A standard result on the continuity of entropy [26, Lemma 1.2.7] now implies that (recall $\delta \le 1/2$)

$$|H(\tilde{U}_\ell|Y_\ell=i) - H(U_\ell|Y_\ell=i)| \le -\delta\log\frac{\delta}{5}$$

so

$$|H(\tilde{U}_\ell|Y_\ell) - H(U_\ell|Y_\ell)| \le -\delta\log\frac{\delta}{5}.$$

Likewise, for any i in $\{-1,1\}$

$$\frac{1}{2}\Pr(\tilde{U}_\ell = 5|Y_0=i) \le 8\left(\frac{2D}{\lambda}\right)^{1/L}.$$

Thus

$$\Pr(\tilde{U}_\ell = 5|Y_0=i) \le \frac{\delta}{2}$$

so

$$|H(\tilde{U}_\ell|Y_0) - H(U_\ell|Y_0)| \le -\delta\log\frac{\delta}{5}$$

as before. It follows that

$$|I(Y_\ell;\tilde{U}_\ell|Y_0) - I(Y_\ell;U_\ell|Y_0)| \le -2\delta\log\frac{\delta}{5}.$$

Combining this with (27) yields

$$\frac{1}{L}\sum_{\ell=1}^{L} I(Y_\ell;U_\ell|Y_0) \ge g\left((D+\delta)^{1/L}\right) + 2\delta\log\frac{\delta}{5}. \qquad \square$$

We are now in a position to prove the main result of this Appendix.

Lemma 7: For any $p^L \le D$

$$\lim_{\lambda\to\infty} \mathcal{R}_o(D,\lambda) \ge (1-D)\log 2 + Lg\left(D^{1/L}\right).$$

Proof: Fix $p^L \le D \le 1$ and $\delta \in (0,1/2]$, and suppose λ satisfies

$$\lambda \ge \max\left[4\left(\frac{32L}{\delta p(1-p)}\right)^{2L}, \left(\frac{D}{\delta}\right)^2\right]. \tag{28}$$

By taking $X = Y_0$ in the definition of $\mathcal{RD}_o(\lambda)$, it follows that there exist $\boldsymbol{R}$ in $\mathbb{R}_+^L$ and γ in Γ_o such that

$$D+\delta \ge E[d_1^\lambda(Y_0,Z_1)], \text{ and}$$
$$\mathcal{R}_o(D,\lambda)+\delta \ge \sum_{\ell=1}^{L} R_\ell \ge I(Y_0;\boldsymbol{U}|T) + \sum_{\ell=1}^{L} I(Y_\ell;U_\ell|Y_0,W,T). \tag{29}$$

For each realization (w,t) of (W,T), let

$$D_{w,t} = E[d_1^\lambda(Y_0,Z_1)|W=w,T=t].$$

Let $S = \{(w,t) : D_{w,t} \leq \sqrt{\lambda}\}$. Then by Markov's inequality

$$\Pr((W,T) \notin S) \leq \frac{D}{\sqrt{\lambda}} \leq \delta. \tag{30}$$

In particular, $\Pr((W,T) \in S) > 0$. Also, for any $(w,t) \in S$

$$\frac{32L}{p(1-p)} \left(\frac{2D_{w,t}}{\lambda}\right)^{1/L} \leq \delta$$

by (28). Thus, by Lemma 6, if $(w,t) \in S$

$$\frac{1}{L}\sum_{\ell=1}^{L} I(Y_\ell; U_\ell | Y_0, W=w, T=t) \geq g\left((D_{w,t}+\delta)^{1/L}\right) + 2\delta \log\frac{\delta}{5}.$$

By averaging over $(w,t) \in S$ and invoking Corollary 1, we obtain

$$\sum_{(w,t)\in S} \frac{1}{L}\sum_{\ell=1}^{L} I(Y_\ell; U_\ell | Y_0, W=w, T=t) \cdot \frac{\Pr(W=w, T=t)}{\Pr((W,T)\in S)} \geq g((D+\delta)^{1/L}) + 2\delta\log\frac{\delta}{5}.$$

From (30), it follows that

$$\sum_{\ell=1}^{L} I(Y_\ell; U_\ell|Y_0, W, T) \geq L(1-\delta)\left[g\left((D+\delta)^{1/L}\right) + 2\delta\log\frac{\delta}{5}\right]. \tag{31}$$

Now by the data processing inequality

$$\begin{aligned} I(Y_0; \boldsymbol{U}|T) &= I(Y_0; \boldsymbol{U}, T) \\ &\geq I(Y_0; Z_1). \end{aligned}$$

Let $\varepsilon = 1(Y_0 \cdot Z_1 = -1)$. Continuing

$$\begin{aligned} I(Y_0; \boldsymbol{U}|T) &\geq H(Y_0) - H(Y_0|Z_1) \\ &= \log 2 - H(Y_0, \varepsilon|Z_1) \\ &= \log 2 - H(\varepsilon|Z_1) - H(Y_0|\varepsilon, Z_1) \\ &\geq \log 2 - h(D/\lambda) - \Pr(Z_1 = 0)\log 2 \\ &\geq (1-D)\log 2 - h(\delta). \end{aligned}$$

Substituting this and (31) into (29) yields

$$\begin{aligned} R_o(D,\lambda) \geq\ & (1-D)\log 2 - h(\delta) \\ & + L(1-\delta)\left[g\left((D+\delta)^{1/L}\right) + 2\delta\log\frac{\delta}{5}\right] - \delta. \end{aligned}$$

The proof is terminated by letting $\lambda \to \infty$ and then $\delta \to 0$. $\square$

Appendix C
The Berger–Tung Outer Bound Is Loose for the Binary Erasure CEO Problem

We will show numerically that for one instance of the binary erasure CEO problem, $\mathcal{RD}_o^{BT}$ contains points with a strictly superoptimal sum rate. Let $L = 2$ and $p = 1/2$. Let W_1 and W_2 be $\{0,1\}$-valued random variables with the joint distribution

$$\begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 0 \end{bmatrix}$$

i.e.,

$$\begin{aligned} &\Pr(W_1 = 0, W_2 = 0) = \frac{1}{5} \\ &\Pr(W_1 = 1, W_2 = 0) = \Pr(W_1 = 0, W_2 = 1) = \frac{2}{5}. \end{aligned}$$

We assume that (W_1, W_2) is independent of (Y_0, Y_1, Y_2). Let $U_\ell = Y_\ell \cdot W_\ell$ for ℓ in $\{1,2\}$, and let $Z_1 = \mathrm{sgn}(U_1 + U_2)$. Since Y_ℓ can be written as $Y_\ell = Y_0 \cdot N_\ell$ where N_1 and N_2 are i.i.d. with $\Pr(N_1 = 0) = \Pr(N_1 = 1) = 1/2$ (recall the notation of Section III-C), we have $U_\ell = Y_0 \cdot N_\ell \cdot W_\ell$. Note that $N_1 \cdot W_1$ and $N_2 \cdot W_2$ have the joint distribution

$$\begin{bmatrix} 3/5 & 1/5 \\ 1/5 & 0 \end{bmatrix}.$$

Thus, for any λ

$$E[d_1^\lambda(Y_0, Z_1)] = \Pr(N_1 \cdot W_1 = N_2 \cdot W_2 = 0) = 3/5.$$

Now we can compute

$$I(Y_1, Y_2; U_1, U_2) \leq 0.6273 \text{ nats}$$

and

$$\begin{aligned} I(Y_1, Y_2; U_1|U_2) &= I(Y_1, Y_2; U_1, U_2) - I(Y_1, Y_2; U_2) \\ &= I(Y_1, Y_2; U_1, U_2) - I(Y_2; U_2) \\ &\leq 0.3248 \text{ nats}. \end{aligned}$$

It follows that $(0.3248, 0.3248, 3/5)$ is in $\mathcal{RD}_o^{BT}(\lambda)$ for any λ. Thus

$$\lim_{\lambda\to\infty} \inf\left\{R_1 + R_2 : \left(R_1, R_2, \frac{3}{5}\right) \in \overline{\mathcal{RD}_o^{BT}}(\lambda)\right\} \leq 0.6496 \text{ nats}.$$

From the previous two appendices, the correct sum rate is

$$\left(1 - \frac{3}{5}\right)\log 2 + 2g\left(\sqrt{\frac{3}{5}}\right) \geq 0.6562 \text{ nats}.$$

Appendix D
Evaluation of the Outer Bound for the Gaussian CEO Problem

Two lemmas are needed for our proof of Proposition 6. The first is a simple extension of Theorem 1 to the Gaussian CEO setting of Section III-D. For this appendix, let us redefine χ to be the set of real-valued random variables X such that $Y_1, \ldots, Y_L$ are conditionally independent given X (the side information Y_{L+1} is unneeded and shall be ignored). Let us also redefine Γ_o to be the set of random variables $(U_1, \ldots, U_L, Z_1, W, T)$ such that each takes values in a finite-dimensional Euclidean space, and collectively they satisfy the Markov conditions defining the original Γ_o:

(i) (W,T) is independent of $(Y_0, \boldsymbol{Y})$;
(ii) $U_\ell \leftrightarrow (Y_\ell, W, T) \leftrightarrow (Y_0, \boldsymbol{Y}_{\ell^c}, \boldsymbol{U}_{\ell^c})$ for all ℓ; and
(iii) $(Y_0, \boldsymbol{Y}, W) \leftrightarrow (\boldsymbol{U}, T) \leftrightarrow Z_1$;

and one new technical condition:

(iv) the conditional distribution of U_ℓ given W and T is discrete for each ℓ.

Note that any conditional distribution involving these random variables is well defined [34, Theorem 6.3]. As such, so is any conditional mutual information [22, Ch. 3, especially the translator's notes at the end].

Lemma 8: For the Gaussian CEO problem, $\mathcal{RD}_\star \subseteq \mathcal{RD}_o$ if $\mathcal{RD}_o$ is defined using the χ and Γ_o just described.

The proof follows the original and is omitted. The second ingredient is a consequence of an intriguing result of Oohama [13] and Prabhakaran, Tse, and Ramchandran that relates information the encoders send about the hidden source to information they send about their observation noise.

Lemma 9 (cf. [13, Lemma 3]): If γ is in Γ_o, then for all $A \subseteq \{1, \ldots, L\}$

$$\exp(2I(Y_0; \boldsymbol{U}_A|W,T)) \le 1 + \sum_{\ell \in A} \frac{1 - \exp(-2I(Y_\ell; U_\ell|Y_0, W, T))}{\sigma_\ell^2/\sigma^2}.$$

Proof: For any realization of (W,T), it follows from Lemma 3 in Oohama [13] that[3]

$$\exp(2I(Y_0; \boldsymbol{U}_A|W=w,T=t)) \le 1 + \sum_{\ell \in A} \frac{1 - \exp(-2I(Y_\ell; U_\ell|Y_0, W=w, T=t))}{\sigma_\ell^2/\sigma^2}.$$

We now average over (w,t) and invoke the convexity of $\exp(\cdot)$ twice, once on each side. □

Proof of Proposition 6: If $(\boldsymbol{R}, D)$ is in $\mathcal{RD}_o$, then there exists γ in Γ_o such that $E[(Y_0 - Z_1)^2] \le D$ and for all $A \subseteq \{1, \ldots, L\}$

$$\sum_{\ell \in A} R_\ell \ge I(Y_0; \boldsymbol{U}_A|\boldsymbol{U}_{A^c}, T) + \sum_{\ell \in A} I(Y_\ell; U_\ell|Y_0, W, T). \quad (32)$$

Now

$$I(Y_0; \boldsymbol{U}_A|\boldsymbol{U}_{A^c}, T) + I(Y_0; \boldsymbol{U}_{A^c}|T) = I(Y_0; \boldsymbol{U}|T). \quad (33)$$

Since $Y_0 \leftrightarrow (\boldsymbol{U}, T) \leftrightarrow Z_1$, the right-hand side can be lower-bounded as follows:

$$\begin{aligned} I(Y_0; \boldsymbol{U}|T) &= I(Y_0; \boldsymbol{U}, T) \\ &\ge I(Y_0; Z_1) \\ &\ge \frac{1}{2} \log \frac{\sigma^2}{E[(Y_0 - Z_1)^2]} \end{aligned}$$

where we have used the rate–distortion theorem for Gaussian sources [2, Theorem 10.3.2]. In particular

$$I(Y_0; \boldsymbol{U}|T) \ge \frac{1}{2} \log \frac{\sigma^2}{D}. \quad (34)$$

Let us address the second term on the left-hand side of (33). Observe that

$$\begin{aligned} &I(Y_0; \boldsymbol{U}_{A^c}|T) + I(Y_0; W|\boldsymbol{U}_{A^c}, T) \\ &\quad = I(Y_0; W|T) + I(Y_0; \boldsymbol{U}_{A^c}|W, T) \\ &\quad = I(Y_0; \boldsymbol{U}_{A^c}|W, T). \end{aligned}$$

[3] Oohama's result assumes that U_ℓ is a deterministic function of Y_ℓ for each ℓ, but the proof shows that condition (ii) above is actually sufficient.

Defining $r_\ell = I(Y_\ell; U_\ell|Y_0, W, T)$ and applying Lemma 9 to the right-hand side gives

$$I(Y_0; \boldsymbol{U}_{A^c}|T) \le \frac{1}{2} \log \left[1 + \sum_{\ell \in A^c} \frac{1 - \exp(-2r_\ell)}{\sigma_\ell^2/\sigma^2}\right]. \quad (35)$$

Substituting (34) and (35) into (33) gives

$$I(Y_0; \boldsymbol{U}_A|\boldsymbol{U}_{A^c}, T) \ge \frac{1}{2} \log^+ \left\{ \frac{1}{D} \left[\frac{1}{\sigma^2} + \sum_{\ell \in A^c} \frac{1 - \exp(-2r_\ell)}{\sigma_\ell^2} \right]^{-1} \right\}.$$

The conclusion follows upon substitution of this inequality and the definition of r_ℓ into (32). □

Appendix E
The Berger–Tung Outer Bound is Loose for the Gaussian CEO Problem

We have just seen that the improved outer bound is capable of recovering the converse result of Oohama [13] and Prabhakaran, Tse, and Ramchandran [14] for the Gaussian CEO problem. Here we will show that the Berger–Tung outer bound does not recover this result. As with the binary erasure CEO problem, we will show that, in general, the Berger–Tung outer bound contains points with a strictly superoptimal sum rate.

Consider the case in which, in the notation of Section III-D, $L = 2$ and $\sigma^2 = \sigma_1^2 = \sigma_2^2 = 1$. In words, two encoders each observe a unit variance, i.i.d. Gaussian process in additive Gaussian noise with a signal-to-noise ratio of unity. It follows from Proposition 6 that the minimum sum rate needed to achieve the distortion $1/2$ is at least $(3/2)\log 2$ nats.

Let W, V_1, and V_2 be Gaussian random variables, independent of each other and of Y_0, Y_1, and Y_2. Let V_1 and V_2 have unit variance; we denote the variance of W by σ_W^2. Let

$$\begin{aligned} U_1 &= Y_1 + V_1 + W \\ U_2 &= Y_2 + V_2 - W. \end{aligned}$$

Note that the sum of U_1 and U_2 is a sufficient statistic for Y_0 given U_1 and U_2. This observation makes it easy to verify that if $Z_1 = E[Y_0|U_1, U_2]$, then $E[(Y_0 - Z_1)^2] = 1/2$. Note that this distortion is independent of σ_W^2.

It follows that for any value of σ_W^2, $\mathcal{RD}_o^{BT}$ contains points of the form $(R_1, R_2, 1/2)$ with

$$R_1 + R_2 = \max(I(Y_1, Y_2; U_1, U_2), 2\, I(Y_1, Y_2; U_1|U_2)).$$

However

$$I(Y_1, Y_2; U_1, U_2) = \frac{3}{2} \log 2 + \frac{1}{2} \log \frac{1 + \sigma_W^2}{1 + 2\sigma_W^2}$$

and

$$\begin{aligned} I(Y_1, Y_2; U_1|U_2) &= I(Y_1, Y_2; U_1, U_2) - I(Y_1, Y_2; U_2) \\ &= I(Y_1, Y_2; U_1, U_2) - I(Y_2; U_2) \\ &= \frac{3}{2} \log 2 + \frac{1}{2} \log \frac{1 + \sigma_W^2}{1 + 2\sigma_W^2} \\ &\quad - \frac{1}{2} \log \left(1 + \frac{2}{\sigma_W^2 + 1}\right). \end{aligned}$$

Observe that, when viewed as functions of σ_W^2, $I(Y_1,Y_2;U_1,U_2)$ is strictly decreasing and $I(Y_1,Y_2;U_1|U_2)$ is continuous. Since $\sigma_W^2=0$ yields

$$I(Y_1,Y_2;U_1,U_2)=\frac{3}{2}\log 2$$
$$2I(Y_1,Y_2;U_1|U_2)<\frac{3}{2}\log 2$$

it follows that there exists $\sigma_W^2>0$ such that

$$\max(I(Y_1,Y_2;U_1,U_2),2I(Y_1,Y_2;U_1|U_2))<\frac{3}{2}\log 2.$$

Appendix F
$\mathcal{RD}_i^{BT}$ Is Closed

The main step in proving that $\mathcal{RD}_i^{BT}$ is closed is to show that one can limit the ranges of the auxiliary random variables without reducing the region.

Definition 7: Let $\hat{\Gamma}_i^{BT}$ denote the set of finite-alphabet random variables

$$\gamma=(U_1,\dots,U_L,Z_1,\dots,Z_K,T)$$

in Γ_i^{BT} such that

$$|\mathcal{U}_\ell|=|\mathcal{Y}_\ell|+2^L+K-2 \text{ for all } \ell$$

and

$$|\mathcal{T}|=2^L+K-1.$$

Then let

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

We shall show that $\mathcal{RD}_i^{BT}=\hat{\mathcal{RD}}_i^{BT}$ in two steps, first handling the case in which T is deterministic, and then bootstrapping to the general case. Both steps involve now-standard uses of Carathéodory's theorem. We give proofs of both steps, albeit condensed ones, due to the complexity of the setup.

Lemma 10: Suppose that $\gamma=(\boldsymbol{U},\boldsymbol{Z},T)$ is in Γ_i^{BT} and T is deterministic. Then there exists $\hat{\gamma}=(\hat{\boldsymbol{U}},\hat{\boldsymbol{Z}},\hat{T})$ in $\hat{\Gamma}_i^{BT}$ such that $\hat{T}$ is deterministic and $\mathcal{RD}_i^{BT}(\gamma)=\mathcal{RD}_i^{BT}(\hat{\gamma})$.

Proof: For any $A\subseteq\{1,\dots,L\}$ containing 1, we have

$$\begin{aligned}&I(\boldsymbol{Y}_A;\boldsymbol{U}_A|\boldsymbol{U}_{A^c})\\&=H(\boldsymbol{Y}_A|\boldsymbol{U}_{A^c})-H(\boldsymbol{Y}_A|\boldsymbol{U})\\&=H(\boldsymbol{Y}_A|\boldsymbol{U}_{A^c})-\sum_{u_1}H(\boldsymbol{Y}_A|\boldsymbol{U}_{1^c},U_1=u_1)\Pr(U_1=u_1)\end{aligned}$$

while for any nonempty A not containing 1, we have

$$\begin{aligned}&I(\boldsymbol{Y}_A;\boldsymbol{U}_A|\boldsymbol{U}_{A^c})\\&\quad=\sum_{u_1}I(\boldsymbol{Y}_A;\boldsymbol{U}_A|\boldsymbol{U}_{A^c\setminus\{1\}},U_1=u_1)\Pr(U_1=u_1).\end{aligned}$$

Carathéodory's theorem [26, Lemma 4.3.4] guarantees that we can find a $\hat{U}_1$ with $\hat{\mathcal{U}}_1\subseteq\mathcal{U}_1$ such that $|\hat{\mathcal{U}}_1|=|\mathcal{Y}_1|+2^L+K-2$ and the equations at the bottom of the page hold, and similarly for $I(\boldsymbol{Y}_A;\boldsymbol{U}_A|\boldsymbol{U}_{A^c\setminus\{1\}},U_1=u_1)$ and $E[d_k(Y_0,\boldsymbol{Y},Y_{L+1},Z_k)|U_1=u_1]$. Since $U_1\leftrightarrow Y_1\leftrightarrow(Y_0,\boldsymbol{Y}_{1^c},\boldsymbol{U}_{1^c},Y_{L+1})$, if we substitute $\hat{U}_1$ for U_1, the resulting γ is in Γ_i^{BT} and $\mathcal{RD}_i^{BT}(\gamma)$ is unchanged. Repeating this procedure for $U_2,\dots,U_L$ completes the proof. $\square$

Lemma 11: $\mathcal{RD}_i^{BT}=\hat{\mathcal{RD}}_i^{BT}$.

Proof: Let $(\boldsymbol{U},\boldsymbol{Z},T)$ be in Γ_i^{BT}. For each t in $\mathcal{T}$, let $(\boldsymbol{U},\boldsymbol{Z},t)$ denote the joint distribution of $(\boldsymbol{U},\boldsymbol{Z},T)$ conditioned on the event $\{T=t\}$. By Lemma 10, for each t, there exists $(\hat{\boldsymbol{U}},\hat{\boldsymbol{Z}})$ such that $(\hat{\boldsymbol{U}},\hat{\boldsymbol{Z}},t)$ is in $\hat{\Gamma}_i^{BT}$ and $\mathcal{RD}_i^{BT}(\boldsymbol{U},\boldsymbol{Z},t)=\mathcal{RD}_i^{BT}(\hat{\boldsymbol{U}},\hat{\boldsymbol{Z}},t)$. By replacing $(\boldsymbol{U},\boldsymbol{Z})$ with $(\hat{\boldsymbol{U}},\hat{\boldsymbol{Z}})$ for each value of T, we obtain $(\hat{\boldsymbol{U}},\hat{\boldsymbol{Z}},T)$ in Γ_i^{BT} such that $|\hat{\mathcal{U}}_\ell|=|\mathcal{Y}_\ell|+2^L+K-2$ for all ℓ and $\mathcal{RD}_i^{BT}(\boldsymbol{U},\boldsymbol{Z},T)=\mathcal{RD}_i^{BT}(\hat{\boldsymbol{U}},\hat{\boldsymbol{Z}},T)$. Now

$$\begin{aligned}&\mathcal{RD}_i^{BT}(\hat{\boldsymbol{U}},\hat{\boldsymbol{Z}},T)\\&=\Bigg\{(\boldsymbol{R},\boldsymbol{D}):\sum_{\ell\in A}R_\ell\geq\\&\qquad\sum_{t\in\mathcal{T}}I(\boldsymbol{Y}_A;\boldsymbol{U}_A|\boldsymbol{U}_{A^c},Y_{L+1},T=t)\Pr(T=t)\text{ for all }A,\text{and}\\&\quad D_k\geq\sum_{t\in\mathcal{T}}E[d_k(Y_0,\boldsymbol{Y},Y_{L+1},Z_k)|T=t]\Pr(T=t)\text{ for all }k\Bigg\}.\end{aligned}$$

Carathéodory's theorem implies that we can find a $\hat{T}$ with $\hat{\mathcal{T}}\subseteq\mathcal{T}$ and $|\hat{\mathcal{T}}|=2^L+K-1$ such that

$$\begin{aligned}&\sum_{t\in\mathcal{T}}I(\boldsymbol{Y}_A;\boldsymbol{U}_A|\boldsymbol{U}_{A^c},Y_{L+1},T=t)\Pr(T=t)\\&\qquad=\sum_{t\in\hat{\mathcal{T}}}I(\boldsymbol{Y}_A;\boldsymbol{U}_A|\boldsymbol{U}_{A^c},Y_{L+1},T=t)\Pr(\hat{T}=t)\end{aligned}$$

for all nonempty A and similarly for

$$E[d_k(Y_0,\boldsymbol{Y},Y_{L+1},Z_k)|T=t].$$

Then $(\hat{\boldsymbol{U}},\hat{\boldsymbol{Z}},\hat{T})$ is in $\hat{\Gamma}_i^{BT}$ and

$$\mathcal{RD}_i^{BT}(\boldsymbol{U},\boldsymbol{Z},T)=\mathcal{RD}_i^{BT}(\hat{\boldsymbol{U}},\hat{\boldsymbol{Z}},\hat{T}).$$

Since $(\boldsymbol{U},\boldsymbol{Z},T)$ in Γ_i^{BT} was arbitrary, this implies $\mathcal{RD}_i^{BT}\subseteq\hat{\mathcal{RD}}_i^{BT}$. This completes the proof because the reverse containment is obvious. $\square$

$$\sum_{u_1\in\hat{\mathcal{U}}_1}\Pr(Y_1=y_1|U_1=u_1)\Pr(\hat{U}_1=u_1)=\Pr(Y_1=y_1)\text{ for all }y_1\text{ in }\mathcal{Y}_1\text{ but one}$$

$$\sum_{u_1\in\mathcal{U}_1}H(\boldsymbol{Y}_A|\boldsymbol{U}_{1^c},U_1=u_1)\Pr(U_1=u_1)=\sum_{u_1\in\hat{\mathcal{U}}_1}H(\boldsymbol{Y}_A|\boldsymbol{U}_{1^c},U_1=u_1)\Pr(\hat{U}_1=u_1)\text{ for all }A\text{ containing }1$$

The cardinality bounds provided by the last two lemmas, while finite, are exponential in L and hence impractical for moderate numbers of encoders. One can improve upon these bounds by exploiting the contrapolymatroid structure [32], [27] of $\mathcal{RD}_i^{BT}$. While this would be useful if one wished to numerically evaluate the bound, our aim here is merely to show that it is closed.

Lemma 12: $\hat{\mathcal{RD}}_i^{BT}$ is closed.

Proof: The Markov conditions defining $\hat{\Gamma}_i^{BT}$ can be expressed as

$$I(T; Y_0, \boldsymbol{Y}, Y_{L+1}) = 0$$
$$I(U_\ell; Y_0, \boldsymbol{Y}_{\ell^c}, Y_{L+1}, \boldsymbol{U}_{\ell^c} | Y_\ell, T) = 0, \quad \text{for all } \ell$$
$$I(Y_0, \boldsymbol{Y}; \boldsymbol{Z} | \boldsymbol{U}, Y_{L+1}, T) = 0.$$

Since the conditional mutual information function is continuous, $\hat{\Gamma}_i^{BT}$ is compact when viewed as a subset of Euclidean space. Thus, if $(\boldsymbol{R}^{(n)}, \boldsymbol{D}^{(n)})$ is a sequence in $\hat{\mathcal{RD}}_i^{BT}$ that converges to $(\boldsymbol{R}, \boldsymbol{D})$, by considering subsequences we may assume that $(\boldsymbol{R}^{(n)}, \boldsymbol{D}^{(n)})$ is in $\hat{\mathcal{RD}}_i^{BT}(\gamma^{(n)})$ for each n and $\gamma^{(n)} \to \gamma = (\boldsymbol{U}, \boldsymbol{Z}, T) \in \hat{\Gamma}_i^{BT}$. By invoking the continuity of mutual information once again, we obtain

$$\sum_{\ell \in A} R_\ell \geq I(\boldsymbol{Y}_A; \boldsymbol{U}_A | \boldsymbol{U}_{A^c}, T)$$

for each A. Likewise

$$D_k \geq E[d_k(Y_0, \boldsymbol{Y}, Y_{L+1}, Z_k)], \quad \text{for all } k.$$

It follows that $(\boldsymbol{R}, \boldsymbol{D})$ is in $\mathcal{RD}_i^{BT}(\gamma)$ and therefore also in $\hat{\mathcal{RD}}_i^{BT}$. $\square$

Corollary 2: $\mathcal{RD}_i^{BT}$ is closed.

ACKNOWLEDGMENT

A. B. Wagner would like to thank Vinod Prabhakaran for many helpful discussions. The results in Section III-D are due to him. This work has also benefited from the helpful comments of Stark C. Draper, Pramod Viswanath, Anant Sahai, and the anonymous reviewers.

REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2006.
[3] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.
[4] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, vol. 7, no. 4, pp. 142–163, Mar. 1959.
[5] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 3, pp. 294–300, May 1975.
[6] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 6, pp. 629–637, Nov. 1975.
[7] A. D. Wyner and J. Ziv, "The rate–distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
[8] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 2, pp. 219–221, Mar. 1979.
[9] S. I. Gel'fand and M. S. Pinsker, "Coding of sources on the basis of observations with incomplete information," (in Russian) *Probl. Pered. Inform.*, vol. 15, no. 2, pp. 45–57, Apr.–Jul. 1979.
[10] 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.
[11] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 106–117, Jan. 1999.
[12] M. Gastpar, "The Wyner–Ziv problem with multiple sources," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2762–2768, Nov. 2004.
[13] Y. Oohama, "Rate–distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, Jul. 2005.
[14] V. Prabhakaran, D. N. C. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, 2004, p. 117.
[15] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, 54, no. 5, pp. 1938–1961, May 2008.
[16] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1978, vol. 229, CISM Courses and Lectures, pp. 171–231.
[17] S.-Y. Tung, "Multiterminal Source Coding," Ph.D. dissertation, Cornell Univ., School of Electrical Engineering, Ithaca, NY, 1978.
[18] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 226–228, Mar. 1975.
[19] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.
[20] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, Sep. 1997.
[21] Y. Oohama, "Multiterminal source coding for correlated memoryless Gaussian sources with several side informations at the decoder," in *Proc. IEEE Information Theory and Communications Workshop*, Kruger National Park, South Africa, Jun. 1999, p. 100.
[22] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes.* Transl.: from Russian by A. Feinstein. San Francisco, CA: Holden-Day, 1964.
[23] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.
[24] H. Wang and P. Viswanath, "Vector Gaussian multiple description with individual and central receivers," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2133–2153, Jun. 2007.
[25] J. N. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Trans. Autom. Control*, vol. 30, no. 5, pp. 440–446, May 1985.
[26] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest, Hungary: Akadémiai Kiadó, 1981.
[27] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 977–987, Aug. 2004.
[28] V. Anantharam and V. S. Borkar, "Common randomness and distributed control: A counterexample," *Syst. Control Lett.*, vol. 56, pp. 568–72, 2007.
[29] 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.
[30] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1912–1923, Nov. 1997.
[31] A. Vamvatsikos, "On the Wagner-Anantharam Outer Bound and Achievable Gaussian Source Coding Exponents," Master's thesis, Cornell Univ., School of Electrical and Computer Engineering, Ithaca, NY, 2007.
[32] P. Viswanath, "Sum rate of a class of Gaussian multiterminal source coding problems," in *Advances in Network Information Theory*, ser. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, P. Gupta, G. Kramer, and A. J. van Wijngaarden, Eds. Providence, RI: AMS, 2004, vol. 66, pp. 43–60.
[33] H. L. Royden, *Real Analysis*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1988.
[34] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed. New York: Springer-Verlag, 2002.