Coding Perspectives for Collaborative Distributed Estimation Over Networks

John MacLaren Walsh and Sivagnanasundaram Ramanan
Dept. of Electrical and Computer Engineering
Drexel University
Philadelphia, PA 19104, USA
E-mail: jwalsh@coe.drexel.edu, sur23@drexel.edu

Abstract

A collaborative distributed estimation problem over a communication constrained network is considered from an information theory perspective. It is shown that the distributed estimation problem is related to multiterminal information theory and a suitable architecture for the codes for this multiterminal information theory problem is determined under source-channel separation. In particular, distributed source codes in which each node multicasts a different messages to each subset of other nodes are studied. This code construction hybridizes two important families of source codes: multiple description codes and codes for the CEO problem. The goal of this paper is to determine the fundamental relationship between the multicast communication rates and estimation performance obtainable, which is embodied in the rate distortion region for such distributed source codes. An achievable rate region for a given set of distortion constraints is proved to this problem and its structural properties are studied. Also, this achievable rate region is shown to simplify to the known bounds to some simpler problems.

I. Introduction

Consider a network of M nodes deployed to monitor a common phenomenon embodied by a sequence of random variables $T^{(n)}$. Each node j $(j \in \{1, \dots, M\})$ in the network makes indirect observations of this phenomenon, embodied as another sequence of random variables $Y_j^{(n)}$ statistically related to $T^{(n)}$. Let $\left\{T^{(n)}, Y_1^{(n)}, \dots, Y_M^{(n)}\right\}_{n=1}^{\infty}$ be temporally memoryless with instantaneous joint probability distribution $p_{T,Y_1,\dots,Y_M}\left(t^{(n)}, y_1^{(n)}, \dots, y_M^{(n)}\right)$ on $\mathcal{T} \times \mathcal{Y}_1 \times \dots \times \mathcal{Y}_M$.

Each node could use the local observations to obtain Bayesian estimates $\tilde{T}_j^{(n)}$ of $T^{(n)}$ that minimize some local cost function $\frac{1}{N}\sum_{n=1}^{N} E\left[d_j(\tilde{T}_j^{(n)}, T_j^{(n)}) \middle| \{Y_j^{(n)} | n \in [N]\}\right]$. Here, we allow the local cost functions to be different at different nodes taking into account the fact that different nodes may be interested in different parameters and different metrics may be appropriate for different parameters. Alternatively, the nodes in the network could communicate with each other in hopes of improving their estimates. We will study such collaborative distributed estimation [1] schemes which accomplish this with separated network/channel and source coding (despite the fact that such a separation is known to be suboptimal in some multiterminal problems). The network/channel codes see to it that messages sent over the network arrive at the intended receivers unaltered, while the distributed source code sees to it that the content of these messages provides the right information extracted from the observations at a node in order to lower the estimation error at the destination.

Our first insight, made in Section II, is that under this decomposition the proper source coding model reflecting the capabilities of the network code is one in which each node multicasts a different message to every possible subset of other nodes in the network. In particular, the source encoder at each node j encodes its observations $\{Y_j^{(n)} | n \in [N]\}$ into a common message $Q_{j\to\mathcal{A}} \in \{1, 2, \dots, 2^{NR_{j\to\mathcal{A}}}\}$ to each of the nodes with indices in some subset $\mathcal{A}$ of the other nodes using an average of $R_{j\to\mathcal{A}}$ bits per observation symbol. A different such message can be encoded at each node j for each such subset $\mathcal{A} \subseteq [M] \setminus j$, and then reliably multicasted via network coding to the nodes in $\mathcal{A}$. Of course, messages which are not to be sent are readily accommodated by setting the associated rate to zero.

One might think that a simpler natural model for such internode communication would be to suppose that each node j could reliably transmit a message $Q_{j\to i}$ exclusively to a single other node i with some rate $R_{j\to i}$ bits per observation symbol $Y_j^{(n)}$. However, owing to the difference in both wireless and wired networks between a multicast (multiple destination) transmission and a similar instantiation as a series of unicast (single destination) transmissions, the presented more versatile model is more directly coupled to network and/or wireless network channel coding capabilities. This idea is elaborated upon in Section II.

Supposing then, that these messages are reliably transmitted to their intended destinations (e.g. with the aid of some channel and network codes), we employ a classic technique from multiterminal information theory [2] [3] to study the relationship between the rates $\{R_{j\to\mathcal{A}} \mid j \in [M],\ \mathcal{A} \in 2^{[M]\setminus j}\}$ of the source code used, and the estimation errors D_j that each of the

J. M. Walsh and S. Ramanan were supported in part by the National Science Foundation under grant CCF-0728496 and part by the Air Force Office of Scientific Research under grant FA9550-09-C-0014. They wish to thank Jun Chen of McMaster University for helpful comments and suggestions in the early stages of this work.

nodes can obtain in estimating the sequence $T^{(n)}, n \in [N]$ from their own observations $\left\{Y_j^{(1)}, \ldots, Y_j^{(N)}\right\}$ and the messages $Q_{\mathcal{D}_j} := \left[Q_{i \to \mathcal{A}} \left| j \in \mathcal{A},\ \mathcal{A} \in 2^{[M]\setminus i}\right.\right]$ they have received. In this paper we give an inner bound (Section III) on the region of rates necessary to obtain a particular set of target average estimation errors $\{D_j | j \in [M]\}$. This notion and notation will be made precise in Section III. We also show that this inner bound is equal to the the known bounds for some simpler problems which can be obtained as simplifications of our model.

A. Related Work

Recent developments in sensor network signal processing have rekindled interest in the problems of remote and distributed estimation. Much of the work in these areas, both classical[4], [5], [2], [6] and modern[7], [8], [3], has focused on the optimized relationship between communication constraints and estimate performance. Such tradeoffs between communication rates and estimate performance, when optimized over all possible code constructions, lie squarely within the domain of information theory, and distributed source coding, in particular.

Although remote estimation and source coding may superficially appear to be completely seperate problems, Dobrushin and Tsybakov [4] first showed the relationship between Shannon's formulation of source coding and remote estimation within a point to point context. They considered the simplest case in which a single decoder estimates a source using the message received from a single encoder which indirectly observes the source via statistically related local observations. Such an indirect modification of Wyner and Ziv's [5] classic result, in which the decoder may also employ some local observations to help it estimate the source, is also possible [7].

With these results in point to point remote estimation long firmly established, attention has shifted to distributed estimation in a network context. Significant attention has focussed on the most simple distributed estimation model after point to point, in which a collection of nodes independently encode local observations into rate limited messages which are relayed to a single fusion center aiming to estimate an underlying sequence of parameters. The relationship between the rates and the estimate performance at the fusion center has been studied in the information theory community under the name of the CEO problem [2], [6], [9], [3], [10], and the closely related many-help-one problem [11], in which the parameter sequence to be estimated is directly observed by one of the nodes. In [12], Prabhakaran et al. explored the possibility of co-operation between the sensors in the CEO problem before communication with the fusion center, and showed that there is no sum-rate advantage of doing so for the Gaussian case. More generally, a distributed estimation problem may require coding schemes which allow multiple network nodes (which may also act as encoders) to act as fusion centers [2]. In the most extreme such case, which one might refer to as collaborative distributed estimation, each node in the network acts both as an encoder of messages to be sent to other nodes and as a fusion center decoder. The question then arises, what is the appropriate information theory model for such a problem? Is it a collection of indepedent CEO problems or a series of concatenated Wyner-Ziv problems[7], or does it have new characteristics? We suggest in the upcoming section that the appropriate generalization is a hybrid between a CEO problem and a multiple descriptions [13] problem, because in this context we argue it is important to allow an encoder to make multiple encodings of its observations bound for different subsets of other nodes in the network.

B. Notation

1) We denote the set $\{1, \ldots, M\}$ as $[M]$ for any natural number M. Also, $[M] \setminus i$ will denote the set $[M]$ with the element i removed.
2) For some set $\mathcal{A}$, $2^{\mathcal{A}}$ will be the power set of subsets from $\mathcal{A}$.
3) We use capital letters for random variables and small letters for realizations. We use superscripts for time indices and subscripts for node indices.
4) For a sequence of random variables $X^{(n)}$, $n \in [N]$, we define X^N as $X^N := \left[X^{(1)}, \ldots, X^{(N)}\right]$.
5) We denote the set of message indices leaving node i by $\mathcal{S}_i := \left\{(i \to \mathcal{A}) \,|\, \mathcal{A} \in 2^{[M]\setminus i}\right\}$ and define $\mathcal{S}$ to be the set of all message indices transmitted over the network, i.e. $\mathcal{S} := \bigcup_{i \in [M]} \mathcal{S}_i$.
6) We define the set of message indices arriving at node i to be $\mathcal{D}_i := \left\{(j \to \mathcal{A}) \,|\, i \in \mathcal{A},\, \mathcal{A} \in 2^{[M]\setminus j},\, j \in [M] \setminus i\right\}$. For any subset $\mathcal{C}_i \subseteq \mathcal{D}_i$, denote the set $\mathcal{E}(\mathcal{C}_i)$ to be the set of nodes which encoded these messages, i.e. $\mathcal{E}(\mathcal{C}_i) := \{j \,|\, \exists \mathcal{A},\, (j \to \mathcal{A}) \in \mathcal{C}_i\}$.
7) For any set of subscript indices $K = \{k_1, \ldots, k_L\}$, the vector $(F_{k_1}, \ldots, F_{k_L})$ is denoted with F_K.
8) For a set of length N sequences $X_1^N, \ldots, X_L^N$, the set of jointly strongly typical sequences [14] is denoted as $A_\epsilon^*(X_{[L]})$.
9) The notation $X \leftrightarrow Y \leftrightarrow Z$ means that X, Y, Z form a Markov chain.

II. Distributed Estimation and Multiterminal Source Coding

As outlined in the introduction, suppose we aim to separate the source coding part of the distributed estimation problem from the network/channel coding part, despite the fact that such a separation may be suboptimal. Here we argue that the best model for the distributed source code is one in which each encoder multicasts a message to each subset of other nodes in the network, rather than sending an individual message to each other node in the network.

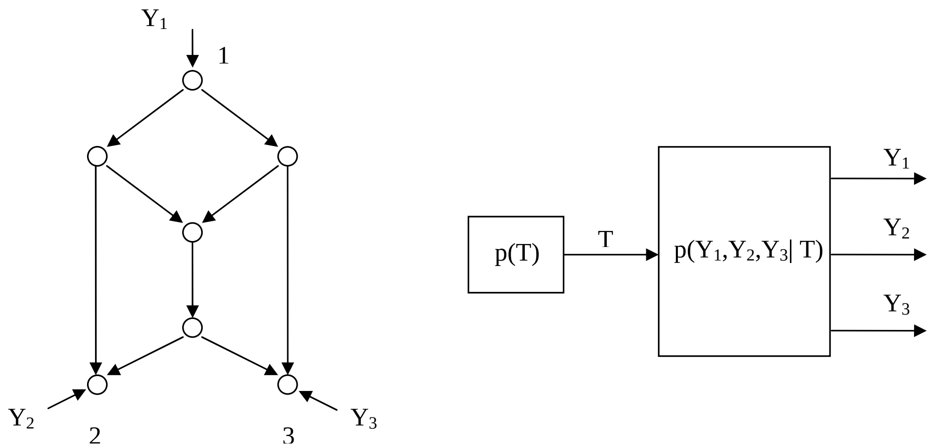

Fig. 1. This network demonstrates that considering a source code at node 1 which only encodes a dedicated message to node 2 and a dedicated message to node 3 is not general enough. Instead, the source encoder at node 1 should encode a separate message for each possible subset of other nodes in the network.

To see that such a model is the appropriate one, consider a simple wired network depicted in Figure 1 in which three nodes $(1, 2, 3)$ making local observations $Y_1^{(n)}, Y_2^{(n)}, Y_3^{(n)}$ statistically related to a common underlying sequence $T^{(n)}$ would like to communicate over the butterfly network in order to form local estimates $\hat{T}_1^{(n)}, \hat{T}_2^{(n)}, \hat{T}_3^{(n)}$ of $T^{(n)}$. Because of the unidirectionality of the links, only node 1 may transmit information. Suppose further that the observations at node 2 and 3 are statistically identical and the distortion metrics are the same, and we wish to obtain the same target average estimation error $D_2 = D_3$ at the two nodes. If node 1 encodes a separate message for node 2 and node 3, then it would suffice to take these two messages to be the same in this symmetric case. However, the network code can not know this, because we have forced the source coding construction to have a separate message for each of nodes 2 and 3. Thus, the network code is forced to attempt to transmit two unicasts, one between 1 and 2 with rate $R_{1\to 2}$, and one in between 1 and 3 with rate $R_{1\to 3}$. If each link in the network is unit capacity, and the network code is forced to treat the information flowing in between nodes 1 and 2 as independently unicast from the unicast between 1 and 3, then the highest symmetric rate $R = R_{1\to 2} = R_{1\to 3}$ which can be obtained is $3/2$. However, had we chosen our source code as outputting three messages $S_{1\to 2}, S_{1\to 3}, S_{1\to 2,3}$, so that we included one which was *multicast* from 1 to both 2 and 3, then the network code could support a symmetric rate of $R_{1\to 2,3} = 2$ [15]. This would not send any unicast information at all $R_{1\to 2} = R_{1\to 3} = 0$. This way 33% more useful information flows from 1 to 2 and 3 as would have had we required only unicasts, and the distortion obtained at nodes 2 and 3 will thus be lower.

A similar conclusion concerning the insufficiency of a distributed source code creating unicasts in this context of networked estimation can be drawn in a wireless context as well. In particular, consider the network depicted in Figure 2 in which node 1 transmits a signal which is overheard by both 2 and 3, through independent Gaussian noise of differing powers, so that the received signal to noise power ratio at node 2 is lower than that at node 3. The capacity region of such a degraded broadcast channel is known, and a channel code can be constructed which reliably transmits the multicast messages $S_{1\to 2}, S_{1\to 3}, S_{1\to 2,3}$ of rates $R_{1\to 2}, R_{1\to 3}, R_{1\to 2,3}$ respectively if there is some $\alpha \in [0, 1]$ such that

$$R_{1\to 2} + R_{1\to 2,3} < \frac{1}{2}\log_2\left(1 + \frac{(1-\alpha)P_T}{\alpha P_T + N_2}\right) \tag{1}$$

$$R_{1\to 3} < \frac{1}{2}\log_2\left(1 + \alpha\frac{P_T}{N_3}\right) \tag{2}$$

If we again consider the application goal of the network for 1 to transmit information in order to help 2 and 3 with their estimation problems, and consider the special case that the local observations at 2 and 3 are statistically identical, then clearly the use of multicast messages (as opposed to unicast messages alone) is desirable. This is because any information sent to node 2 can also be heard at node 3 with no extra cost in information, and thus our source code ought to exploit this capability of the channel code.

From these two simple examples we can easily infer that a proper separated source and network/channel coding approach treats the source code within network node i as producing an array of 2^{M-1} multicast messages, with one message $X_{i\to\mathcal{A}}^N$ for each subset $\mathcal{A} \subseteq [M] \setminus i$. The capabilities of the possible network/channel codes are then summarized by a region $\mathcal{C}$ of vectors of such multicast rates

$$\mathbf{r} := [R_{j\to\mathcal{A}} \,|j \in [M],\ \mathcal{A} \subseteq [M] \setminus j] \tag{3}$$

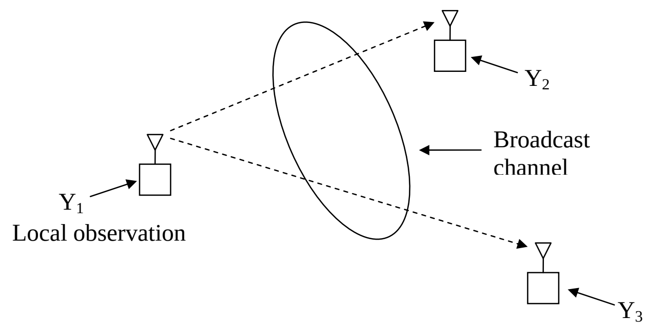

Fig. 2. Due to the broadcast nature of the wireless medium, an appropriate source coding model for collaborative inference over a wireless channel should involve communication with subsets of other users rather than only point to point communication.

which are simultaneously supportable by the network infrastructure. The capabilities of the possible source codes are summarized by a *rate distortion region* $\mathcal{RD}$ describing the set of simultaneously achievable multicast rates $\mathbf{r}$ and average estimation errors

$$\mathbf{d} := [D_i | i \in [M]], \quad D_i := \frac{1}{N} \sum_{n=1}^{N} E\left[d_i\left(T^{(n)}, \hat{T}_i^{(n)}\right)\right] \tag{4}$$

An overall source channel code achieving average estimation errors lower than $\mathbf{d}$ is selected by choosing a rate vector $\mathbf{r}$ that is in both $\mathcal{C}$ and also in $\mathcal{RD}$, i.e. with $(\mathbf{r}, \mathbf{d}) \in \mathcal{RD}$.

The capabilities of a given network infrastructure for reliable information transmission, and the design of the associated network/channel codes, are the topic of a large amount of research both prior and current. Thus, we now focus our efforts on describing the rate distortion region for the associated family of source codes we have selected.

III. Main Results

Having outlined our multiterminal source coding problem, we presently give an achievable rate distortion region to this problem. The rate distortion region explains the relationship between the length in bits of the different messages multicast between the nodes and the estimation errors (measured in terms of average costs for Bayesian estimation) that decoder/estimators at these nodes can obtain. In particular, the vector $(\mathbf{r}, \mathbf{d})$ of multicast rates $\mathbf{r} := \left[R_{j \to \mathcal{A}} \,\middle|\, j \in [M], \mathcal{A} \in 2^{[M]\setminus j}\right]$ and average estimation errors $\mathbf{d} := [D_j \,|\, j \in [M]]$ is said to be achievable if there exists a block length N, encoders and decoders

$$f_{j\to\mathcal{A}}^N : \mathcal{Y}_j^N \to [L_{j\to\mathcal{A}}^N], \quad g_i^N : \mathcal{Y}_i^N \times \prod_{(j\to\mathcal{A})\in\mathcal{D}_i} [L_{j\to\mathcal{A}}^N] \to \hat{\mathcal{T}}_i \tag{5}$$

with $\hat{T}_i^N = g_i^N(Y_i^N, Q_{\mathcal{D}_i})$ such that

$$R_{j\to\mathcal{A}} \geq \frac{1}{N} \log L_{j\to\mathcal{A}}^N, \quad E\left[\frac{1}{N}\sum_{n=1}^{N} d_i(T^{(n)}, \hat{T}_i^{(n)})\right] \leq D_i \tag{6}$$

The rate distortion region $\mathcal{RD}$ for this problem is defined as the closure of the region of achievable vectors $(\mathbf{r}, \mathbf{d})$.

The achievable rate region that we present is based on a well known bound in the distributed source coding literature, the Berger-Tung inner bound[2]. A slight modification of this bound to reflect differences from the CEO problem brought about by simultaneous encodings sent from the same encoder (i.e. multiple descriptions) to different overlapping subsets of decoders turns out to be necessary.

Theorem 1: Given a joint distribution $p_{T,Y_{[M]}}(t, y_{[M]})$, let $\Xi(\mathbf{d})$ be the collection of random vectors $\boldsymbol{\xi} = (U_{\mathcal{S}})$ which are jointly distributed with T and $Y_{[M]}$ such that the following conditions are satisfied

1) $T, Y_{[M]\setminus i}, U_{\mathcal{S}\setminus\mathcal{S}_i} \leftrightarrow Y_i \leftrightarrow U_{\mathcal{S}_i}$ for all $i \in [M]$
2) There exists a decoding function $g_i : \mathcal{U}_{\mathcal{D}_i} \times \mathcal{Y}_i \to \hat{\mathcal{T}}_i$ such that $E\left[d_i(T, g_i(U_{\mathcal{D}_i}, Y_i))\right] \leq D_i$ for all $i \in [M]$

For each $\boldsymbol{\xi} \in \Xi(\mathbf{d})$, let $\boldsymbol{\Phi}(\boldsymbol{\xi})$ be the collection of vectors $\boldsymbol{\phi} = (\tilde{R}_{\mathcal{S}})$

$$\boldsymbol{\Phi}(\boldsymbol{\xi}) := \left\{ \tilde{R}_{\mathcal{S}} \,\middle|\, \sum_{(j\to\mathcal{A})\in\mathcal{P}_j} \tilde{R}_{j\to\mathcal{A}} > \sum_{(j\to\mathcal{A})\in\mathcal{P}_j} H(U_{j\to\mathcal{A}}) - H(U_{\mathcal{P}_j}|Y_j) \text{ for each } \mathcal{P}_j \subseteq \mathcal{S}_j \text{ and } j \in [M] \right\}$$

For each $\boldsymbol{\xi} \in \Xi(\mathbf{d})$ and for each $\boldsymbol{\phi} \in \boldsymbol{\Phi}(\boldsymbol{\xi})$, let

$$\mathcal{RD}_{in}(\boldsymbol{\xi}, \boldsymbol{\phi}) \triangleq \left\{ R_{\mathcal{S}} \,\middle|\, \sum_{(j\to\mathcal{A})\in\mathcal{C}_i} R_{j\to\mathcal{A}} \geq \sum_{(j\to\mathcal{A})\in\mathcal{C}_i} \left(\tilde{R}_{j\to\mathcal{A}} - H(U_{j\to\mathcal{A}})\right) + H(U_{\mathcal{C}_i}|U_{\mathcal{D}_i\setminus\mathcal{C}_i}, Y_i), \forall \mathcal{C}_i \subseteq \mathcal{D}_i \text{ and } i \in [M] \right\}$$

Define

$$\mathcal{RD}_{in} := \bigcup_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \bigcup_{\boldsymbol{\phi}\in\boldsymbol{\Phi}(\boldsymbol{\xi})} \mathcal{RD}_{in}(\boldsymbol{\xi}, \boldsymbol{\phi})$$

Then, the convex hull $\mathbf{conv}\,(\mathcal{RD}_{in})$ of $\mathcal{RD}_{in}$ is an inner bound to the rate distortion region, i.e. $\mathbf{conv}\,(\mathcal{RD}_{in}) \subseteq \mathcal{RD}$.

Proof idea: This result is an adaptation of a well known inner bound in the multiterminal source coding community known as the Berger-Tung inner bound, as clarified by Han and Kobayashi [2], with the twist that the multiple (dependent) descriptions at each encoder require an additional set of encoder inequalities. The proof may be found in Appendix A. □

We next analyze the structure of the achievable rate region, because knowing the structure of the rate region may be helpful when we optimize some function of rates over the rate region. We indeed use some structural properties of the inner bound to simplify our bound to simpler problems in Section IV, and, thus present those structural properties below.

Proposition 1: For each $\boldsymbol{\xi} \in \Xi(\mathbf{d})$, $\Phi(\boldsymbol{\xi})$ is a contra-polymatroid.

Proof: The set $\mathcal{S}$ is implied to be the ground set, and the rank function $\rho : 2^{\mathcal{S}} \to \mathbb{R}$ is defined as

$$\rho(\mathcal{P}) \triangleq \sum_{j\in[M]} \sum_{(j\to\mathcal{A})\in\mathcal{P}\cap\mathcal{S}_j} H(U_{j\to\mathcal{A}}) - H(U_{\mathcal{P}\cap\mathcal{S}_j}|Y_j) \tag{7}$$

We must show that ρ is indeed a rank function. Consider two sets $\mathcal{Q}$ and $\mathcal{P}$ such that $\mathcal{Q} \subseteq \mathcal{P} \subseteq \mathcal{S}$, then

$$\begin{aligned} \rho(\mathcal{P}) - \rho(\mathcal{Q}) &= \sum_{j\in[M]} \left(\sum_{(j\to\mathcal{A})\in(\mathcal{P}\cap\mathcal{S}_j)\setminus(\mathcal{Q}\cap\mathcal{S}_j)} H(U_{j\to\mathcal{A}}) - H(U_{\mathcal{P}\cap\mathcal{S}_j}|Y_j) + H(U_{\mathcal{Q}\cap\mathcal{S}_j}|Y_j) \right) \\ &= \sum_{j\in[M]} \left(\sum_{(j\to\mathcal{A})\in(\mathcal{P}\cap\mathcal{S}_j)\setminus(\mathcal{Q}\cap\mathcal{S}_j)} H(U_{j\to\mathcal{A}}) - H(U_{(\mathcal{P}\cap\mathcal{S}_j)\setminus(\mathcal{Q}\cap\mathcal{S}_j)}|U_{\mathcal{Q}\cap\mathcal{S}_j}, Y_j) \right) \geq 0 \end{aligned}$$

This establishes that ρ is non-decreasing. Next consider any two sets $\mathcal{P} \subseteq \mathcal{S}$ and $\mathcal{Q} \subseteq \mathcal{S}$. We have

$$\begin{aligned} &\rho(\mathcal{P}) + \rho(\mathcal{Q}) - \rho(\mathcal{P}\cap\mathcal{Q}) - \rho(\mathcal{P}\cup\mathcal{Q}) = \\ &\sum_{j\in[M]} \left(H(U_{\mathcal{P}\cap\mathcal{Q}\cap\mathcal{S}_j}|Y_j) + H(U_{(\mathcal{P}\cup\mathcal{Q})\cap\mathcal{S}_j}|Y_j) - H(U_{\mathcal{P}\cap\mathcal{S}_j}|Y_j) - H(U_{\mathcal{Q}\cap\mathcal{S}_j}|Y_j) \right) \\ &= \sum_{j\in[M]} \left(H(U_{\mathcal{P}\cap\mathcal{Q}^c\cap\mathcal{S}_j}|U_{\mathcal{Q}\cap\mathcal{S}_j}, Y_j) - H(U_{\mathcal{P}\cap\mathcal{Q}^c\cap\mathcal{S}_j}|U_{\mathcal{P}\cap\mathcal{Q}\cap\mathcal{S}_j}, Y_j) \right) \leq 0 \end{aligned}$$

which implies that ρ is a rank function of a contra-polymatroid. To see that this contra-polymatroid is equal to $\Phi(\boldsymbol{\xi})$, simply note that evaluating the rank function ρ and writing the corresponding inequality for every subset of $\mathcal{S}_j$ gives the list of inequalities for node j. The collection of these inequalities over $j \in [M]$ then yields $\Phi(\boldsymbol{\xi})$. Finally, note that evaluating the rank function at any collection of indices corresponding to message sent from different encoders simply sums the corresponding individual inequalities for the different encoders. □

Corollary 1: For each $\boldsymbol{\xi} \in \Xi(\mathbf{d})$, the generating vertices of the polyhedron $\Phi(\boldsymbol{\xi})$ are exactly $\{\boldsymbol{\phi}(\boldsymbol{\pi})|\boldsymbol{\pi} \in \Pi(\mathcal{S})\}$ where $\Pi(\mathcal{S})$ is the set of permutations of the indices in $\mathcal{S}$, and $\boldsymbol{\phi}(\boldsymbol{\pi})$ is the vector given by

$$\phi_{\pi(1)}(\boldsymbol{\pi}) \triangleq \rho(\pi(1)) = I(U_{\pi(1)}; Y_{[M]}), \text{ and for every } i \in \{2, \ldots, |\mathcal{S}|\}, \tag{8}$$

$$\phi_{\pi(i)}(\boldsymbol{\pi}) \triangleq \rho(\{\pi(1), \ldots, \pi(i)\}) - \rho(\{\pi(1), \ldots, \pi(i-1)\}) = I(U_{\pi(i)}; U_{\{\pi(1),\ldots,\pi(i-1)\}}, Y_{[M]}) \tag{9}$$

and where ρ is the rank function defined in (7). Additionally, for any $\boldsymbol{\lambda} \in \mathbb{R}_+^{|\mathcal{S}|}$, then the solution to the linear program $\min_{\boldsymbol{\phi}\in\Phi(\boldsymbol{\xi})} \boldsymbol{\lambda}\cdot\boldsymbol{\phi}$ is attained by $\boldsymbol{\phi}(\boldsymbol{\pi})$ for $\boldsymbol{\pi}$ any permutation of the elements of $\mathcal{S}$ such that $\lambda_{\pi(1)} \geq \cdots \geq \lambda_{\pi(|\mathcal{S}|)}$.

Proof: These are standard properties of contra-polymatroids. See, for instance, Lemma 3.3 of [16]. □

IV. Simplification of Bounds to Simpler Problems

Because we have argued that the collaborative distributed estimation problem is essentially a hybrid between a collection of CEO problems and a multiple descriptions problem, it is important to show that the inner bound we have given specializes to known inner bounds for these problems in special cases.

A. Simplification to Multiple Descriptions Problem

The multiple descriptions problem for two descriptions can be obtained as a special case of our collaborative estimation problem for $M = 4$ nodes. Only one node, say node 1, gets to make observations which it would like to inform the other 3 network nodes about, so that $Y_1^{(n)} = T^{(n)}$ and $Y_i^{(n)} = 0$ for all $i \neq 1$. Additionally, node 1 structures its encodings so that nodes 2 and 3 receive different encodings, while node 4 receives everything that is available to node 2 and 3. This can be accomplished by having the only encodings sent from node 1 be $Q_{1\to\{2,4\}}, Q_{1\to\{3,4\}}, Q_{1\to\{4\}}$. The coding strategy introduced in [13] forms the two descriptions by dividing the bits of the message only to be used when both descriptions are available, i.e. $Q_{1\to\{4\}}$ up into two parts $Q_{1\to\{4\}} = (Q^1_{1\to\{4\}}, Q^2_{1\to\{4\}})$ containing $\Delta_1 \geq 0$ and $\Delta_2 \geq 0$ bits per symbol with $\Delta_1 + \Delta_2 = R_{1\to\{4\}}$. These two parts are included in the two descriptions $X_1 \triangleq (Q_{1\to\{2,4\}}, Q^1_{1\to\{4\}})$ and $X_2 \triangleq (Q_{1\to\{3,4\}}, Q^2_{1\to\{4\}})$. When only one of the two descriptions X_1 or X_2 is available, the achievability coding strategy introduced in [13] simply discards the part of the description associated with $Q_{1\to\{4\}}$ and utilizes only $U_{1\to\{3,4\}}$ or $U_{1\to\{2,4\}}$, respectively. When both descriptions are available, the achievability coding strategy introduced in [13] uses all of the encodings $(Q_{1\to\{2,4\}}, Q_{1\to\{3,4\}}, Q_{1\to\{4\}})$. We have thus set up our coding scheme so that user 4 is the user which gets both descriptions, while user 2 and user 3 get only the description X_1 and X_2, respectively. Additionally, since $R_1 = R_{2,4} + \Delta_1$ and $R_2 = R_{3,4} + \Delta_2$, we can remove the redundant variables Δ_1 and Δ_2, and rewrite the constraint for R_4 as $R_4 = R_1 - R_{2,4} + R_2 - R_{3,4}$. These identifications may be summarized with the following notation

$$Y_{1,n} \triangleq T_n,\ Y_{i,n} = \emptyset \quad \forall i \neq 1 \tag{10}$$
$$U_{1\to\{2,3\}} \triangleq U_{2,3},\ U_{1\to\{3,4\}} \triangleq U_{3,4},\ U_{1\to\{4,\}} \triangleq U_4, \quad \text{all other } U_{j\to\mathcal{A}} = \emptyset \tag{11}$$
$$R_{1\to\{2,3\}} \triangleq R_{2,3},\ R_{1\to\{3,4\}} \triangleq R_{3,4},\ R_{1\to\{4,\}} \triangleq R_4, \quad \text{all other } R_{j\to\mathcal{A}} = \emptyset \tag{12}$$
$$\tilde{R}_{1\to\{2,3\}} \triangleq \tilde{R}_{2,3},\ \tilde{R}_{1\to\{3,4\}} \triangleq \tilde{R}_{3,4},\ \tilde{R}_{1\to\{4\}} \triangleq \tilde{R}_4, \quad \text{all other } \tilde{R}_{j\to\mathcal{A}} = \emptyset \tag{13}$$

Where the auxiliary random variables $U_4, U_{2,4}, U_{3,4}, \hat{T}_2, \hat{T}_3, \hat{T}_4$ are selected such that

$$p(U_4, U_{2,4}, U_{3,4}, \hat{T}_2, \hat{T}_3, \hat{T}_4, T) = p(U_4, U_{2,4}, U_{3,4}|T)p(\hat{T}_2|U_{2,4})p(\hat{T}_3|U_{3,4})p(\hat{T}_4|U_4, U_{2,4}, U_{3,4})p(T)$$
$$D_1 \geq \mathbb{E}\left[d(T, \hat{T}_2)\right],\ D_2 \geq \mathbb{E}\left[d(T, \hat{T}_3)\right] D_0 \geq \mathbb{E}\left[d(T, \hat{T}_4)\right]$$

Under these identifications, the inner bound becomes

$$R_1 \geq R_{2,4} \tag{14}$$
$$R_2 \geq R_{3,4} \tag{15}$$
$$R_4 = R_1 - R_{2,4} + R_2 - R_{3,4} \tag{16}$$
$$R_4 \geq \tilde{R}_4 - H(U_4) + H(U_4|U_{3,4}, U_{2,4}) \tag{17}$$
$$R_{2,4} \geq \max\left\{\tilde{R}_{2,4} - H(U_{2,4}) + H(U_{2,4}), \tilde{R}_{2,4} - H(U_{2,4}) + H(U_{2,4}|U_{3,4}, U_4)\right\} \tag{18}$$
$$R_{3,4} \geq \max\left\{\tilde{R}_{3,4} - H(U_{3,4}) + H(U_{3,4}), \tilde{R}_{3,4} - H(U_{3,4}) + H(U_{3,4}|U_{2,4}, U_4)\right\} \tag{19}$$
$$R_4 + R_{3,4} \geq \tilde{R}_4 - H(U_4) + \tilde{R}_{3,4} - H(U_{3,4}) + H(U_4, U_{3,4}|U_{2,4}) \tag{20}$$
$$R_4 + R_{2,4} \geq \tilde{R}_4 - H(U_4) + \tilde{R}_{2,4} - H(U_{2,4}) + H(U_4, U_{2,4}|U_{3,4}) \tag{21}$$
$$R_{2,4} + R_{3,4} \geq \tilde{R}_{2,4} - H(U_{2,4}) + \tilde{R}_{3,4} - H(U_{3,4}) + H(U_{2,4}, U_{3,4}|U_4) \tag{22}$$
$$R_4 + R_{2,4} + R_{3,4} \geq \tilde{R}_4 - H(U_4) + \tilde{R}_{2,4} - H(U_{2,4}) + \tilde{R}_{3,4} - H(U_{3,4}) + H(U_4, U_{2,4}, U_{3,4}) \tag{23}$$

Here, the variable R_4 may be replaced everywhere with $R_1 - R_{2,4} + R_2 - R_{3,4}$. Thereby simplifying these expressions, we obtain

$$R_1 \geq R_{2,4} \tag{24}$$
$$R_2 \geq R_{3,4} \tag{25}$$
$$R_1 - R_{2,4} + R_2 - R_{3,4} \geq \tilde{R}_4 - H(U_4) + H(U_4|U_{3,4}, U_{2,4}) \tag{26}$$
$$R_{2,4} \geq \tilde{R}_{2,4} \tag{27}$$
$$R_{3,4} \geq \tilde{R}_{3,4} \tag{28}$$

Here (27) (resp. (19)) follows because the second term in the maximization (18) (resp. 28) is always less than the first term because it is the first term plus a negative semi-definite quantity. Also (21) (resp. (20)) follows from adding (26) and (27) (resp. (19)) and a negative semi-definite right hand quantity, so it is no longer needed. Similarly, (22) follows from adding (27) and (28) with a negative semi-definite right hand quantity, so it is no longer needed. Finally, (23) follows from adding (26), (27), and (28) with a negative semi-definite right hand quantity, so it is no longer needed as well.

Having made these simplifications, it becomes evident that the variables $R_{2,3}$ and $R_{2,4}$ are redundant, and thus, including now the source encoder node inequalities, we have that the rate region is given by

$$R_1 \geq \tilde{R}_{2,4} \tag{29}$$
$$R_2 \geq \tilde{R}_{3,4} \tag{30}$$
$$R_1 + R_2 \geq \tilde{R}_{2,4} + \tilde{R}_{3,4} + \tilde{R}_4 - H(U_4) + H(U_4|U_{3,4}, U_{2,4}) \tag{31}$$

We note that the minimum of $\tilde{R}_{2,4} + \tilde{R}_{3,4} + \tilde{R}_4$ from the encoder inequalities to be

$$H(U_4) + H(U_{2,4}) + H(U_{3,4}) - H(U_4, U_{2,4}, U_{3,4}|T)$$

Thus right hand side of (31) becomes

$$H(U_4) + H(U_{2,4}) + H(U_{3,4}) - H(U_4, U_{2,4}, U_{3,4}|T) - H(U_4) + H(U_4|U_{2,4}, U_{3,4}) = \tag{32}$$
$$H(U_{2,4}) + H(U_{3,4}) - H(U_4, U_{2,4}, U_{3,4}|T) + H(U_4|U_{2,4}, U_{3,4}) + H(U_{2,4}, U_{3,4}) - H(U_{2,4}, U_{3,4}) = \tag{33}$$
$$I(U_{2,4}; U_{3,4}) + I(T; U_4, U_{2,4}, U_{3,4}) \tag{34}$$

We next point out that by the contra-polymatroid property of the source encoder region describing the collection of variables $\tilde{R}_{2,4}, \tilde{R}_{3,4}, \tilde{R}_4$ by Corollary 1, this minimum is attained for 6 (permutations of $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 1$) possible solutions of $\tilde{R}_{2,4}, \tilde{R}_{3,4}, \tilde{R}_4$. However, we are interested in only two of the 6 solutions which are useful in finding the region of (R_1, R_2).

$$\tilde{R}_{2,4} = I(U_{2,4}; T), \quad \tilde{R}_{3,4} = I(U_{3,4}; U_{2,4}, T), \quad \tilde{R}_4 = I(U_4; U_{2,4}, U_{3,4}, T)$$

and/ or

$$\tilde{R}_{2,4} = I(U_{2,4}; U_{3,4}, T), \quad \tilde{R}_{3,4} = I(U_{3,4}; T), \quad \tilde{R}_4 = I(U_4; U_{2,4}, U_{3,4}, T)$$

Using time sharing argument of these two solutions we write the region of rates (R_1, R_2) as

$$R_1 \geq I(U_{2,4}; T) + \alpha \, I(U_{2,4}; U_{3,4}|T) \tag{35}$$
$$R_2 \geq I(U_{3,4}; T) + (1 - \alpha) \, I(U_{2,4}; U_{3,4}|T) \tag{36}$$
$$R_1 + R_2 \geq I(U_{2,4}; U_{3,4}) + I(T; U_4, U_{2,4}, U_{3,4}) \tag{37}$$

where $0 \leq \alpha \leq 1$. We next compare this region with the achievable rate regions for multiple descriptions available in the literature. El Gamal and Cover proved an achievable rate region (EGC region) for multiple descriptions problem in [13] which is given by

$$r_1 \geq I(U_{2,4}; T) \tag{38}$$
$$r_2 \geq I(U_{3,4}; T) \tag{39}$$
$$r_1 + r_2 \geq I(U_{2,4}; U_{3,4}) + I(T; U_4, U_{2,4}, U_{3,4}) \tag{40}$$

We now prove that any point (r_1, r_2) in the EGC region also lies in the region we proved. To prove this, we rewrite the EGC region in the following form

$$r_1 \geq I(U_{2,4}; T)$$
$$r_2 \geq \max\{I(U_{3,4}; T), I(U_{2,4}; U_{3,4}) + I(T; U_4, U_{2,4}, U_{3,4}) - r_1\}$$

and let

$$\alpha = \min\left\{\frac{r_1 - I(U_{2,4}; T)}{I(U_{2,4}; U_{3,4}|T)}, 1\right\} \tag{41}$$

Then

$$R_1 \geq \min\{r_1, I(U_{2,4}; T) + I(U_{2,4}; U_{3,4}|T)\} \leq r_1 \tag{42}$$

and

$$\begin{aligned}R_2 &\geq I(U_{3,4};T)+I(U_{2,4};U_{3,4}|T)-\min\{r_1-I(U_{2,4};T),\ I(U_{2,4};U_{3,4}|T)\}\\ &= \max\{I(U_{2,4};T)+I(U_{3,4};T)+I(U_{2,4};U_{3,4}|T)-r_1,\ I(U_{3,4};T)\}\\ &\leq \max\{I(U_{2,4};U_{3,4})+I(T;U_4,U_{2,4},U_{3,4})-r_1,\ I(U_{3,4};T)\}\\ &\leq r_2\end{aligned}$$

In the above proof we used the following inequality.

$$\begin{aligned}& I(U_{2,4};U_{3,4})+I(T;U_4,U_{2,4},U_{3,4})\\ =\ & I(U_{2,4};U_{3,4})+I(U_{2,4};T)+I(U_{3,4};T|U_{2,4})+I(U_4;T|U_{2,4},U_{3,4})\\ =\ & I(U_{2,4};T)+I(U_{3,4};U_{2,4},T)+I(U_4;T|U_{2,4},U_{3,4})\\ \geq\ & I(U_{2,4};T)+I(U_{3,4};T)+I(U_{2,4};U_{3,4}|T)\end{aligned}$$

This completes the proof that our inner bound contains every point in the EGC region.

B. Simplification to CEO problem

We next show that CEO problem can be obtained as a simplification of our model and that when our inner bound is simplified to this case gives the Berger-Tung inner bound.

Suppose now that the nodes $i \in [M] \setminus M$ observe the common phenomenon embodied by the sequence $T^{(n)}$ and send one message each to the CEO node M. Using the messages received from the nodes $i \in [M-1]$, the CEO node produces an estimate $\hat{T}$ ($\hat{T}_M = \hat{T}$) of T such that the expected distortion $E[d(T,\hat{T})] < D$. Denote the message sent from node i to the CEO with U_i for all $i \in [M-1]$. Let $\Xi(D)$ be the collection of random variables $\boldsymbol{\xi} = (U_{[M-1]})$ satisfying

- $T, Y_{[M-1]\setminus i}, U_{[M-1]\setminus i} \leftrightarrow Y_i \leftrightarrow U_i$ for all $i \in [M-1]$
- There exists a decoding function $g : \mathcal{U}_{[M-1]} \rightarrow \hat{\mathcal{T}}$ such that $D > E\left[d(T,\hat{T})\right]$

Let $\Psi(\boldsymbol{\xi})$ be the set of rates

$$\Psi(\boldsymbol{\xi}) = \left\{ R_{[M-1]} \mid \sum_{i\in\mathcal{C}} R_i > I(U_{\mathcal{C}};Y_{\mathcal{C}}|U_{\mathcal{D}\setminus\mathcal{C}}),\ \forall \mathcal{C} \subseteq [M-1] \right\}$$

Then the Berger-Tung inner bound for the CEO problem is given by

$$\mathcal{RD}_{in} = \left\{ (R_{[M-1]}, D) \middle| R_{[M-1]} \in \bigcup_{\boldsymbol{\xi}\in\boldsymbol{\Xi}(D)} \Psi(\boldsymbol{\xi}) \right\}$$

We prove that when our inner bound is simplified for the CEO problem we get the Berger-Tung bound.

Proof: Since the nodes $i \in [M-1]$ send messsages only to node M, for all $i \in [M-1]$, set

$$R_{i\to\mathcal{A}} = 0 \text{ and } \tilde{R}_{i\to\mathcal{A}} = 0\ \forall\ \mathcal{A} \in 2^{[M]\setminus i} \setminus \{M\}$$

Since node M does not send any message, set

$$R_{M\to\mathcal{A}} = 0 \text{ and } \tilde{R}_{M\to\mathcal{A}} = 0\ \forall\ \mathcal{A} \in 2^{[M]\setminus M}$$

If we define $\mathcal{D}$ as

$$\mathcal{D} = \{(j \to M) | j \in [M-1]\}$$

then $\Phi(\boldsymbol{\xi})$ becomes

$$\Phi(\boldsymbol{\xi}) = \{\tilde{R}_{\mathcal{D}} | \tilde{R}_{j\to M} > H(U_{j\to M}) - H(U_{j\to M}|Y_j), \forall j \in [M-1]\}$$

Here, $\tilde{R}_{j\to M}$ can be selected such that $\tilde{R}_{j\to M} = I(U_{j\to M};Y_{j\to M}) + \epsilon_j$ for all $j \in [M-1]$ where ϵ_j can be made arbitrarily small. Note that selecting the rates so will not change the rate region. If we select $\tilde{R}_{j\to M} = I(U_{j\to M};Y_{j\to M}) + \epsilon_j$, there will be only 1 rate vector $\tilde{R}_{\mathcal{D}}$ in the set $\Phi(\boldsymbol{\xi})$. Thus, Ψ is only a function of $\boldsymbol{\xi}$, i.e. $\Psi(\boldsymbol{\xi},\phi)=\Psi(\boldsymbol{\xi})$. Hence, $\Psi(\boldsymbol{\xi})$ is the collection of rate vectors $R_{\mathcal{D}}\geq 0$ obeying

$$\begin{aligned}\sum_{(j\to M)\in\mathcal{C}} R_{j\to M} &> \sum_{(j\to M)\in\mathcal{C}} (\tilde{R}_{j\to M}-H(U_{j\to M}))+H(U_{\mathcal{C}}|U_{\mathcal{D}\backslash\mathcal{C}})\\ &= H(U_{\mathcal{C}}|U_{\mathcal{D}\backslash\mathcal{C}})-\sum_{(j\to M)\in\mathcal{C}} H(U_{j\to M}|Y_j)\\ &= H(U_{\mathcal{C}}|U_{\mathcal{D}\backslash\mathcal{C}})-H(U_{\mathcal{C}}|Y_{\mathcal{C}})\\ &= H(U_{\mathcal{C}}|U_{\mathcal{D}\backslash\mathcal{C}})-H(U_{\mathcal{C}}|Y_{\mathcal{C}},U_{\mathcal{D}\backslash\mathcal{C}})\\ &= I(U_{\mathcal{C}};Y_{\mathcal{C}}|U_{\mathcal{D}\backslash\mathcal{C}})\end{aligned}$$

for all $\mathcal{C}\subseteq\mathcal{D}$. Here, we have used the facts that node M (CEO) does not have any side information ($Y_M=0$) and $U_{\mathcal{C}}\leftrightarrow Y_{\mathcal{C}}\leftrightarrow U_{\mathcal{D}\backslash\mathcal{C}}$. If we redefine the rates, variables and distortion constraints as

$$\begin{aligned}R_{j\to M} &\triangleq R_j\ \forall\ j\in[M-1]\\ U_{j\to M} &\triangleq U_j\ \forall\ j\in[M-1]\\ D_M &\triangleq D\end{aligned}$$

then an inner bound for the rate-distortion region for the CEO problem becomes

$$\mathcal{RD}_{in}=\left\{(R_{[M-1]},D)\,\middle|\,R_{[M-1]}\in\bigcup_{\boldsymbol{\xi}\in\boldsymbol{\Xi}(D)}\Psi(\boldsymbol{\xi})\right\}$$

where $\boldsymbol{\Xi}(D)$ is the collection random variables $\boldsymbol{\xi}=(U_{[M-1]})$ satisfying

- $T,Y_{[M-1]\backslash i},U_{[M-1]\backslash i}\leftrightarrow Y_i\leftrightarrow U_i$ for all $i\in[M-1]$
- There exists a decoding function $g:\mathcal{U}_{[M-1]}\to\hat{\mathcal{T}}$ such that $D>E\left[d(T,\hat{T})\right]$

■

C. Side Information May Be Absent at the Decoder

The "side information may be absent at the decoder" problem studied by Heegard and Berger in [17] can also be obtained as a simplification of our model. To see this, let the number of nodes $M=3$ and, suppose that node 3 directly observes the source, i.e. $Y_3=T$, and node 1 has side information about the source $Y_1=Y$ while node 2 has no side information. Also, suppose that node 3 sends a common description to both $1,2$ and an individual description to only node 1 as it is implicitly done in [17]. We show that sum of the rates of these two descriptions derived from our inner bound is equal to the rate-distortion function proved for the sum-rate in [17].

To prove this, set the rates and variables which are not involved in the problem zero and redefine the necessary rates and variables as follows.

$$Y_1\triangleq Y,\ Y_3\triangleq T,\quad \text{all other } Y_i=\emptyset \tag{43}$$
$$U_{3\to1}\triangleq U,\ U_{3\to\{1,2\}}\triangleq W,\quad \text{all other } U_{j\to\mathcal{A}}=\emptyset \tag{44}$$
$$\tilde{R}_{3\to1}\triangleq\tilde{R}_1,\ \tilde{R}_{3\to\{1,2\}}\triangleq\tilde{R}_{1,2},\quad \text{all other } \tilde{R}_{j\to\mathcal{A}}=\emptyset \tag{45}$$
$$R_{3\to1}\triangleq R_1,\ R_{3\to\{1,2\}}\triangleq R_{1,2},\quad \text{all other } R_{j\to\mathcal{A}}=\emptyset \tag{46}$$

Note that the variables T,Y,U,W satisfy the following conditions.

1) $Y\leftrightarrow T\leftrightarrow U,W$.
2) There exist functions $\hat{T}_1(U,W,Y)$ and $\hat{T}_2(W)$ such that $E[d_1(T,\hat{T}_1)]\leq D_1$ and $E[d_2(T,\hat{T}_2)]\leq D_2$.

With the redefined variables, the constraints on $\tilde{R}_1,\tilde{R}_{1,2}$ become

$$\tilde{R}_1 \geq H(U)-H(U|T) \tag{47}$$
$$\tilde{R}_{1,2} \geq H(W)-H(W|T) \tag{48}$$
$$\tilde{R}_1+\tilde{R}_{1,2} \geq H(U)+H(W)-H(U,W|T) \tag{49}$$

and the constraints on $R_1, R_{1,2}$ become

$$R_1 \geq \tilde{R}_1 - H(U) + H(U|W,Y) \tag{50}$$
$$R_{1,2} \geq \tilde{R}_{1,2} - H(W) + H(W|U,Y) \tag{51}$$
$$R_{1,2} \geq \tilde{R}_{1,2} \tag{52}$$
$$R_1 + R_{1,2} \geq \tilde{R}_1 - H(U) + \tilde{R}_{1,2} - H(W) + H(U,W|Y) \tag{53}$$

Observing that the tight bound on $R_1 + R_{1,2}$ comes from adding (50) and (52), we write

$$R_1 + R_{1,2} \geq \tilde{R}_1 + \tilde{R}_{1,2} - H(U) + H(U|W,Y) \tag{54}$$

Using the bound on $\tilde{R}_1 + \tilde{R}_{1,2}$ in (49), we obtain

$$R_1 + R_{1,2} \geq \tilde{R}_1 + \tilde{R}_{1,2} - H(U) + H(U|W,Y) \tag{55}$$
$$\geq H(W) - H(U,W|T) + H(U|W,Y) \tag{56}$$
$$= I(T;W) + (T;U|W,Y) \tag{57}$$

This is exactly the rate-distortion function for the sum-rate which Heegard and Berger proved in [17].

V. Conclusion

We analyzed optimized code constructions for collaborative distributed estimation via multiterminal information theory. We argued that the proper model for a distributed source code for collaborative distributed estimation involves multiple multicast messages from each encoder rather than unicast messages, yielding a hybrid coding problem between multiple descriptions and the CEO problem. An achievable rate region which hybridized the Berger Tung inner bound and multiple descriptions proof techniques were presented. The inner bound was shown to be equal to the known bounds for some simpler problems by exploiting the structural properties of the rate region.

Appendix A
Proof of Theorem

We present the detailed proof of the inner bound given in Section III here.

Proof: Select a joint conditional distribution $p\left(u_{\mathcal{S}} \mid t, y_{[M]}\right)$, a set of encoding functions $\{f^N_{j\to\mathcal{A}} \mid (j \to \mathcal{A}) \in \mathcal{S}\}$ and a set of decoding functions $\{g^N_i \mid i \in [M]\}$ such that the rates $R_{\mathcal{S}}$ are in $\mathcal{RD}_{in}$. Calculate the marginal distributions $p(u_{j\to\mathcal{A}})$.

Codebook Generation:
At each node $j \in [M]$, for each subset of nodes $\mathcal{A} \subseteq 2^{[M]\setminus j}$, generate a codebook with $2^{n\tilde{R}_{j\to\mathcal{A}}}$ length-N codewords by randomly drawing the elements such that they are i.i.d. according to the distribution $p(u_{j\to\mathcal{A}})$, where $\sum_{(j\to\mathcal{A})\in\mathcal{P}_j} \tilde{R}_{j\to\mathcal{A}} > \sum_{(j\to\mathcal{A})\in\mathcal{P}_j} H(U_{j\to\mathcal{A}}) - H(U_{\mathcal{P}_j}|Y_j)$ for each $\mathcal{P}_j \subseteq \mathcal{S}_j$. Index the codewords by $m_{j\to\mathcal{A}} \in \{1,\ldots,2^{n\tilde{R}_{j\to\mathcal{A}}}\}$. Partition the codewords into $2^{nR_{j\to\mathcal{A}}}$ bins by randomly and uniformly assigning the indices to the bins. Index the bins by $b_{j\to\mathcal{A}} \in \{1,\ldots,2^{nR_{j\to\mathcal{A}}}\}$ and denote the set of codewords in bin $b_{j\to\mathcal{A}}$ by $\mathcal{B}_{j\to\mathcal{A}}(b_{j\to\mathcal{A}})$.

Encoding:
At each node $j \in [M]$, encode the observation sequence Y^N_j by selecting one codeword $U^N_{j\to\mathcal{A}}(m_{j\to\mathcal{A}})$ from each codebook $C_{j\to\mathcal{A}}$, for each $(j \to \mathcal{A}) \in \mathcal{S}_j\}$, such that $\left(U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j}),\, Y^N_j\right) \in A^*_\epsilon(U_{\mathcal{S}_j}, Y_j)$. If there are more than one such $U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j})$, select the codewords with the smallest indices under lexicographic ordering. If there is no such $U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j})$, select an arbitrary set of codewords. For each subset of nodes $\mathcal{A} \subseteq 2^{[M]\setminus j}$, send the index $b_{j\to\mathcal{A}}$ of the bin that contains $U^N_{j\to\mathcal{A}}(m_{j\to\mathcal{A}})$ to the nodes in $\mathcal{A}$, i.e. $U^N_{j\to\mathcal{A}}(m_{j\to\mathcal{A}}) \in \mathcal{B}_{j\to\mathcal{A}}(b_{j\to\mathcal{A}})$. This requires $R_{j\to\mathcal{A}}$ bits to multicast a message to a subset of nodes $\mathcal{A} \subseteq 2^{[M]\setminus j}$.

Decoding:
At each node $i \in [M]$, decode the messages received at the node by selecting the codeword $U^N_{j\to\mathcal{A}}(\ell_{j\to\mathcal{A}})$ in bin $\mathcal{B}_{j\to\mathcal{A}}(b_{j\to\mathcal{A}})$ for each $(j \to \mathcal{A}) \in \mathcal{D}_i$ such that $\left(U^N_{\mathcal{D}_i}(\boldsymbol{\ell}_{\mathcal{D}_i}), Y^N_i\right) \in A^*_\epsilon\left(U_{\mathcal{D}_i}, Y_i\right)$, where $U_{\mathcal{D}_i} \triangleq (U_{j\to\mathcal{A}})_{(j\to\mathcal{A})\in\mathcal{D}_i}$. If there is no such a set of codewords, select an arbitrary set of codewords. Reproduce the underlying sequence T^N by $\hat{T}^N_i = g^N_i\left(Y^N_i, U^N_{\mathcal{D}_i}(\boldsymbol{\ell}_{\mathcal{D}_i})\right)$.

We presently prove that the expected distortion achieved by this coding technique is approximately D_i at each node $i \in [M]$. We say that the coding is successful when

1) At each encoder, the codewords are selected such that they are collectively jointly strongly typical with the observation sequence and
2) At each decoder, the same codewords selected at the encoders are selected from the corresponding bins, and they are collectively jointly strongly typical with the underlying sequence (T^N) and the side information at the decoder.

We show that when the coding is successful, the expected distortion is approximately D_i. We also show that the probabilities of occurrence of any errors during coding are small, and, contribution to the overall expected distortion by these events is insignificant since the distortion measure is bounded. Finally, we prove that since the probability of successful coding $\to 1$ as $N \to \infty$, overall expected distortion is approximately D_i. We begin the proof by listing the possible errors.

1) E_0: $\left(T^N, Y_{[M]}^N\right) \notin A_\epsilon^*(T, Y_{[M]})$
2) E_1: At least at one (encoder) node $j \in [M]$, $\left(U_{\mathcal{S}_j}^N(\boldsymbol{m}_{\mathcal{S}_j}), Y_j^N\right) \notin A_\epsilon^*\left(U_{\mathcal{S}_j}, Y_j\right)$ for all $\boldsymbol{m}_{\mathcal{S}_j} \in \prod_{(j\to\mathcal{A})\in\mathcal{S}_j}\left[2^{N\tilde{R}_{j\to\mathcal{A}}}\right]$.
3) E_2: At least at one (decoder) node $i \in [M]$, $\left(U_{\mathcal{D}_i}^N(\boldsymbol{m}_{\mathcal{D}_i}), Y_{\mathcal{E}(\mathcal{D}_i)}^N, Y_i^N, T^n\right) \notin A_\epsilon^*\left(U_{\mathcal{D}_i}, Y_{\mathcal{E}(\mathcal{D}_i)}, Y_i, T\right)$ where $\boldsymbol{m}_{\mathcal{D}_i}$ is defined as follows
$$\boldsymbol{m}_{\mathcal{D}_i} := (m_{j\to\mathcal{A}})_{(j\to\mathcal{A})\in\mathcal{D}_i}$$
where $m_{j\to\mathcal{A}}$ is the index of the codeword selected by the encoding function $f_{j\to\mathcal{A}}^N$.
4) E_3: At least at one (decoder) node $i \in [M]$, $\exists\, U_{\mathcal{D}_i}^N(\boldsymbol{m}_{\mathcal{D}_i})$ such that $\left(U_{\mathcal{D}_i}^N(\boldsymbol{m}_{\mathcal{D}_i}), Y_{\mathcal{E}(\mathcal{D}_i)}^N, Y_i^N, T^n\right) \in A_\epsilon^*(U_{\mathcal{D}_i}, Y_{\mathcal{E}(\mathcal{D}_i)}, Y_i, T)$, but $U_{\mathcal{D}_i}^N(\boldsymbol{\ell}_{j\to\mathcal{A}}) \neq U_{\mathcal{D}_i}^N(\boldsymbol{m}_{j\to\mathcal{A}})$ where $\boldsymbol{\ell}_{\mathcal{D}_i}$ is defined as follows
$$\boldsymbol{\ell}_{\mathcal{D}_i} := (\ell_{j\to\mathcal{A}})_{(j\to\mathcal{A})\in\mathcal{D}_i}$$
where $\ell_{j\to\mathcal{A}}$ is the index of the codeword selected at the decoder i such that $U_{j\to\mathcal{A}}^N(\ell_{j\to\mathcal{A}}) \in B_{j\to\mathcal{A}}(b_{j\to\mathcal{A}})$.

Define the coding error E as $E := \cup_{i=0}^3 E_i$. Then the probability of error $Pr(E)$ is bounded above by
$$Pr(E) \leq Pr(E_0) + \sum_{i=1}^{3} Pr(E_i \cap E_0^c)$$
We now show that the probabilities of these errors are small.

1) Clearly $Pr(E_0) \to 0$ as $N \to \infty$.
2) To prove $Pr(E_1 \cap E_0^c)$ is small, for each $y_j^N \in A_\epsilon^*(Y_j)$, $j \in [M]$, define the random set $G_{\mathcal{S}_j}(y_j^N)$ as
$$G_{\mathcal{S}_j}(y_j^N) := \left\{\boldsymbol{m}_{\mathcal{S}_j} \,\middle|\, U_{\mathcal{S}_j}^N(\boldsymbol{m}_{\mathcal{S}_j}) \in A_\epsilon^*(U_{\mathcal{S}_j}|y_j^n)\right\}$$
For each set $\mathcal{S}_j$, define the event $E_{1,\mathcal{S}_j}$ as
$$E_{1,\mathcal{S}_j} := \{(U_{\mathcal{S}_j}^N(\boldsymbol{m}_{\mathcal{S}_j}), Y_j^N) \notin A_\epsilon^*(U_{\mathcal{S}_j}, Y_j) \quad \text{for all} \quad \boldsymbol{m}_{\mathcal{S}_j} \in \prod_{(j\to\mathcal{A})\in\mathcal{S}_j}[2^{N\tilde{R}_{j\to\mathcal{A}}}]\}$$
Then, we have
$$\begin{aligned} Pr\left[E_{1,\mathcal{S}_j} \text{ and } Y_j^N \in A_\epsilon^*(Y_j)\right] &= Pr\left[|G_{\mathcal{S}_j}(Y_j^N)| = 0 \text{ and } Y_j^N \in A_\epsilon^*(Y_j)\right] \\ &\leq \max_{y_j^N \in A_\epsilon^*(Y_j)} Pr\left[|G_{\mathcal{S}_j}(y_j^N)| = 0\right] \end{aligned}$$
Using Chebyshev's inequality [18] [13], for all $y_j^N \in A_\epsilon^*(Y_j)$ and $0 < \alpha < 1$, we write
$$\begin{aligned} Pr\left[|G_{\mathcal{S}_j}(y_j^N)| = 0\right] &\leq Pr\left[\left||G_{\mathcal{S}_j}(y_j^N)| - E\left[|G_{\mathcal{S}_j}(y_j^N)|\right]\right| \geq \alpha E\left[|G_{\mathcal{S}_j}(y_j^N)|\right]\right] \\ &\leq \frac{\text{Var}\left[|G_{\mathcal{S}_j}(y_j^N)|\right]}{\alpha^2\left(E\left[|G_{\mathcal{S}_j}(y_j^N)|\right]\right)^2} \end{aligned}$$
We now bound $E\left[|G_{\mathcal{S}_j}(y_j^N)|\right]$. Define the indicator function
$$1\left(U_{\mathcal{S}_j}^N(\boldsymbol{m}_{\mathcal{S}_j}) \in G_{\mathcal{S}_j}(y_j^N)\right) = \begin{cases} 1 & \text{if } U_{\mathcal{S}_j}^N(\boldsymbol{m}_{\mathcal{S}_j}) \in G_{\mathcal{S}_j}(y_j^N) \\ 0 & \text{otherwise} \end{cases}$$
Then the cardinality of the set $G_{\mathcal{S}_j}(y_j^N)$ is given by
$$\left|G_{\mathcal{S}_j}(y_j^N)\right| = \sum_{\boldsymbol{m}_{\mathcal{S}_j} \in [2^{N\tilde{R}_{\mathcal{S}_j}}]} 1\left(U_{\mathcal{S}_j}^N(\boldsymbol{m}_{\mathcal{S}_j}) \in G_{\mathcal{S}_j}(y_j^N)\right)$$

where $\left[2^{N\tilde{R}_{\mathcal{S}_j}}\right]$ denotes the Cartesian product of the sets $\left[2^{N\tilde{R}_{j\to\mathcal{A}}}\right]$, $(j\to\mathcal{A})\in\mathcal{S}_j$. Since $E\left[1\left(U^N_{\mathcal{S}_j}\in G_{\mathcal{S}_j}(y_j^N)\right)\right]\geq 2^{-N\left[\sum_{(j\to\mathcal{A})\in\mathcal{S}_j}H(U_{j\to\mathcal{A}})-H(U_{\mathcal{S}_j}|Y_j)+\epsilon_1\right]}$ where $\epsilon_1\to 0$ as $\epsilon\to 0$, we have

$$\begin{aligned} E\left[|G_{\mathcal{S}_j}(y_j^N)|\right] &\geq \left|\left[2^{N\tilde{R}_{\mathcal{S}_j}}\right]\right| \; E\left[1\left(U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j})\in G_{\mathcal{S}_j}(y_j^N)\right)\right] \\ &\geq 2^{N\left[\sum_{(j\to\mathcal{A})\in\mathcal{S}_j}\left(\tilde{R}_{j\to\mathcal{A}}-H(U_{j\to\mathcal{A}})\right)+H(U_{\mathcal{S}_j}|Y_j)-\epsilon_1\right]} \end{aligned}$$

We next bound Var $\left[|G_{\mathcal{S}_j}(y_j^N)|\right]$. Consider

$$\begin{aligned} &\text{Var}\left[|G_{\mathcal{S}_j}(y_j^N)|\right]\leq E\left[\left|G_{\mathcal{S}_j}(y_j^N)\right|^2\right] \\ &= E\left[\left(\sum_{\boldsymbol{m}_{\mathcal{S}_j}\in[2^{N\tilde{R}_{\mathcal{S}_j}}]}\;\sum_{\boldsymbol{m}'_{\mathcal{S}_j}\in[2^{N\tilde{R}_{\mathcal{S}_j}}]} 1\left(U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j})\in G_{\mathcal{S}_j}(y_j^N), U^N_{\mathcal{S}_j}(\boldsymbol{m}'_{\mathcal{S}_j})\in G_{\mathcal{S}_j}(y_j^N)\right)\right)\right] \\ &= \sum_{\boldsymbol{m}_{\mathcal{S}_j}}\sum_{\boldsymbol{m}'_{\mathcal{S}_j}} Pr\left[U^N_{\mathcal{S}_j}(\boldsymbol{m}'_{\mathcal{S}_j})\in A^*_\epsilon(U_{\mathcal{S}_j}|y_j^n)\,\middle|\, U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j})\in A^*_\epsilon(U_{\mathcal{S}_j}|y_j^n)\right]\; Pr\left[U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j})\in A^*_\epsilon(U_{\mathcal{S}_j}|y_j^n)\right] \\ &= \sum_{\boldsymbol{m}_{\mathcal{S}_j}}\left(1+\sum_{\boldsymbol{m}_{\mathcal{S}_j}\neq\boldsymbol{m}'_{\mathcal{S}_j}} Pr\left[U^N_{\mathcal{S}_j}(\boldsymbol{m}'_{\mathcal{S}_j})\in A^*_\epsilon(U_{\mathcal{S}_j}|y_j^n)\,\middle|\, U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j})\in A^*_\epsilon(U_{\mathcal{S}_j}|y_j^n)\right]\right)\; Pr\left[U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j})\in A^*_\epsilon(U_{\mathcal{S}_j}|y_j^n)\right] \\ &= \sum_{\boldsymbol{m}_{\mathcal{S}_j}}\left(1+\sum_{\mathcal{P}_j\subset\mathcal{S}_j}\;\sum_{\boldsymbol{m}'_{\mathcal{P}_j}|m_{j\to\mathcal{A}}\neq m'_{j\to\mathcal{A}}\;\;\forall(j\to\mathcal{A})\in\bar{\mathcal{P}}_j} Pr\left[\left(U^N_{\mathcal{P}_j}(\boldsymbol{m}'_{\mathcal{P}_j}),U^N_{\bar{\mathcal{P}}_j}(\boldsymbol{m}'_{\bar{\mathcal{P}}_j})\right)\in A^*_\epsilon(U_{\mathcal{S}_j}|y_j^n)\;|\right.\right. \\ &\qquad\left.\left.\left(U^N_{\mathcal{P}_j}(\boldsymbol{m}_{\mathcal{P}_j}),U^N_{\bar{\mathcal{P}}_j}(\boldsymbol{m}_{\bar{\mathcal{P}}_j})\right)\in A^*_\epsilon(U_{\mathcal{S}_j}|y_j^n)\right]\right)\; Pr\left[U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j})\in A^*_\epsilon(U_{\mathcal{S}_j}|y_j^n)\right] \\ &\leq \sum_{\boldsymbol{m}_{\mathcal{S}_j}}\left(1+\sum_{\mathcal{P}_j\subset\mathcal{S}_j}2^{N\left[\sum_{(j\to\mathcal{A})\in\bar{\mathcal{P}}_j}(\tilde{R}_{j\to\mathcal{A}}-H(U_{j\to\mathcal{A}}))+H(U_{\bar{\mathcal{P}}_j}|U_{\mathcal{P}_j},Y_j)+\epsilon_2\right]}\right)\; Pr\left[U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j})\in A^*_\epsilon(U_{\mathcal{S}_j}|y_j^n)\right] \\ &= \left(1+\sum_{\mathcal{P}_j\subset\mathcal{S}_j}2^{N\left[\sum_{(j\to\mathcal{A})\in\bar{\mathcal{P}}_j}(\tilde{R}_{j\to\mathcal{A}}-H(U_{j\to\mathcal{A}}))+H(U_{\bar{\mathcal{P}}_j}|U_{\mathcal{P}_j},Y_j)+\epsilon_2\right]}\right) E\left[|G_{\mathcal{S}_j}(y_j^N)|\right] \end{aligned}$$

Here $\bar{\mathcal{P}}_j=\mathcal{S}_j\setminus\mathcal{P}_j$ and $\epsilon_2\to 0$ as $\epsilon\to 0$. Hence, we have

$$\begin{aligned} Pr\left[\left|G_{\mathcal{S}_j}(y_j^N)\right|=0\right] &\leq 2^{-N\left[\sum_{(j\to\mathcal{A})\in\mathcal{S}_j}(\tilde{R}_{j\to\mathcal{A}}-H(U_{j\to\mathcal{A}}))+H(U_{\mathcal{S}_j}|Y_j)-\epsilon_1\right]} \\ &\quad+\sum_{\mathcal{P}_j\subset\mathcal{S}_j}2^{-N\left[\sum_{(j\to\mathcal{A})\in\mathcal{P}_j}(\tilde{R}_{j\to\mathcal{A}}-H(U_{j\to\mathcal{A}}))+H(U_{\mathcal{P}_j}|Y_j)-\epsilon_1-\epsilon_2\right]} \end{aligned}$$

Note that $Pr(E_1\cap E_0^c)\leq\sum_{j\in[M]}Pr(E_{1,\mathcal{S}_j}\cap E_0^c)$. Thus, $Pr(E_1\cap E_0^c)$ can be made arbitrary small by selecting the rates $\tilde{R}_{j\to\mathcal{A}}$, $(j\to\mathcal{A})\in\mathcal{S}_j$ such that for each subset $\mathcal{P}_j\subseteq\mathcal{S}_j$ the following rate condition is satisfied

$$\sum_{(j\to\mathcal{A})\in\mathcal{P}_j}\tilde{R}_{j\to\mathcal{A}}>\sum_{(j\to\mathcal{A})\in\mathcal{P}_j}H(U_{j\to\mathcal{A}})-H(U_{\mathcal{P}_j}|Y_j)+\epsilon_1+\epsilon_2 \tag{58}$$

3) When the rate conditions in (58) are satisfied at each encoder $j\in[M]$, by Lemma 1 (see Appendix) $Pr(E_2\cap E_0^c)\to 0$ as $N\to\infty$.
4) To prove $Pr(E_3\cap E_0^c)$ is small, for each $i\in[M]$ define $E_{3,\mathcal{D}_i}$ as

$$\begin{aligned} E_{3,\mathcal{D}_i} = &\left\{\left(U^N_{\mathcal{D}_i}(\boldsymbol{m}_{\mathcal{D}_i}),Y^N_{\mathcal{E}(\mathcal{D}_i)},Y_i^N,T^N\right)\in A^*_\epsilon\left(U_{\mathcal{D}_i},Y_{\mathcal{E}(\mathcal{D}_i)},Y_i,T\right)\;\text{ and }\;\left(U^N_{\mathcal{D}_i}(\boldsymbol{\ell}_{\mathcal{D}_i}),Y_i^N\right)\in A^*_\epsilon\left(U_{\mathcal{D}_i},Y_i\right)\right. \\ &\left.\text{for some }\; U^N_{\mathcal{D}_i}(\boldsymbol{\ell}_{\mathcal{D}_i})\in\mathcal{B}_{\mathcal{D}_i}(b_{\mathcal{D}_i})\;\text{ and }\; U^N_{\mathcal{D}_i}(\boldsymbol{m}_{\mathcal{D}_i})\neq U^N_{\mathcal{D}_i}(\boldsymbol{\ell}_{\mathcal{D}_i})\right\} \end{aligned}$$

Note that $\boldsymbol{\ell}_{\mathcal{D}_i} \neq \boldsymbol{m}_{\mathcal{D}_i}$ if $\ell_{j\to\mathcal{A}} \neq m_{j\to\mathcal{A}}$ for at least one $(j \to \mathcal{A}) \in \mathcal{D}_i$. Define the event $E'_{3,\mathcal{D}_i}$ as

$$E'_{3,\mathcal{D}_i} := \Big\{ \left(U^N_{\mathcal{D}_i}(\boldsymbol{\ell}_{\mathcal{D}_i}), Y_i^N\right) \in A^*_\epsilon(U_{\mathcal{D}_i}, Y_i) \quad \text{for some} \quad U^N_{\mathcal{D}_i}(\boldsymbol{\ell}_{\mathcal{D}_i}) \in \mathcal{B}_{\mathcal{D}_i}(b_{\mathcal{D}_i}) \quad \text{and} \quad U^N_{\mathcal{D}_i}(\boldsymbol{m}_{\mathcal{D}_i}) \neq U^N_{\mathcal{D}_i}(\boldsymbol{\ell}_{\mathcal{D}_i}) \Big\}$$

Then

$$Pr(E_{3,\mathcal{D}_i}) \leq Pr\big[E'_{3,\mathcal{D}_i} \,|\, \left(U^N_{\mathcal{D}_i}(\boldsymbol{m}_{\mathcal{D}_i}), Y_i^N\right) \in A^*_\epsilon(U_{\mathcal{D}_i}, Y_i)\big] = P_e$$

Let the set $\mathcal{C}_i$ be

$$\mathcal{C}_i := \{(j \to \mathcal{A}) \in \mathcal{D}_i | \ell_{j\to\mathcal{A}} \neq m_{j\to\mathcal{A}}\}$$

For a particular $\boldsymbol{\ell}_{\mathcal{D}_i} = \boldsymbol{\ell}'_{\mathcal{D}_i} (\neq \boldsymbol{m}_{\mathcal{D}_i})$, define $E_{3,\mathcal{C}_i}(\boldsymbol{\ell}'_{\mathcal{C}_i})$ as

$$\begin{aligned} E_{3,\mathcal{C}_i}(\boldsymbol{\ell}'_{\mathcal{C}_i}) \quad := \quad & \Big\{ \Big(U^N_{\mathcal{D}_i}(\boldsymbol{\ell}'_{\mathcal{D}_i}), Y_i^N\Big) \in A^*_\epsilon(U_{\mathcal{D}_i}, Y_i) \quad \text{such that} \quad \ell'_{j\to\mathcal{A}} \neq m_{j\to\mathcal{A}} \,\forall (j \to \mathcal{A}) \in \mathcal{C}_i \\ & \text{and} \quad \ell'_{j\to\mathcal{A}} = m_{j\to\mathcal{A}} \,\forall (j \to \mathcal{A}) \in \mathcal{D}_i \setminus \mathcal{C}_i \Big\} \end{aligned}$$

Consider the code corresponding to a particular $(j \to \mathcal{A}) \in \mathcal{C}_i$. There are $(|\mathcal{B}_{j\to\mathcal{A}}| - 1)$ sequences in the bin $\mathcal{B}_{j\to\mathcal{A}}(b_{j\to\mathcal{A}})$ such that $\ell_{j\to\mathcal{A}} \neq m_{j\to\mathcal{A}}$, where $|\mathcal{B}_{j\to\mathcal{A}}|$ is the size of the bin. Thus for a particular $\mathcal{C}_i$, the number of possible events in which $\boldsymbol{\ell}_{j\to\mathcal{A}} \neq \boldsymbol{m}_{j\to\mathcal{A}}$ (the number of events $E_{3,\mathcal{C}_i}(\boldsymbol{\ell}'_{\mathcal{C}_i})$) is $\prod_{(j\to\mathcal{A})\in\mathcal{C}_i} (|\mathcal{B}_{j\to\mathcal{A}}| - 1)$. Note that the probabilities of these events are equal. Thus the probability of error P_e is bounded by

$$P_e \leq \sum_{\mathcal{C}_i \subseteq \mathcal{D}_i} \left(\prod_{(j\to\mathcal{A})\in\mathcal{C}_i} (|\mathcal{B}_{j\to\mathcal{A}}| - 1) \right) Pr\big[E_{3,\mathcal{C}_i}(\boldsymbol{\ell}'_{\mathcal{C}_i}) \,|\, \left(U^N_{\mathcal{D}_i}(\boldsymbol{m}_{\mathcal{D}_i}), Y_i^N\right) \in A^*_\epsilon(U_{\mathcal{D}_i}, Y_i)\big]$$

We have the following

$$Pr[E_{3,\mathcal{C}_i}(\boldsymbol{\ell}'_{\mathcal{C}_i}) \,|\, \left(U^N_{\mathcal{D}_i}(\boldsymbol{m}_{\mathcal{D}_i}), Y_i^N\right) \in A^*_\epsilon(U_{\mathcal{D}_i}, Y_i)] \quad \leq \quad 2^{N\left[H(U_{\mathcal{C}_i}|U_{\mathcal{D}_i\setminus\mathcal{C}_i}, Y_i) - \sum_{(j\to\mathcal{A})\in\mathcal{C}_i} H(U_{j\to\mathcal{A}}) + \epsilon_3 + \epsilon_4\right]}$$

where $\epsilon_3, \epsilon_4 \to 0$ as $\epsilon \to 0$. We bound

$$\prod_{(j\to\mathcal{A})\in\mathcal{C}_i} (|\mathcal{B}_{j\to\mathcal{A}}| - 1) < \prod_{(j\to\mathcal{A})\in\mathcal{C}_i} |\mathcal{B}_{j\to\mathcal{A}}| = 2^{N\sum_{(j\to\mathcal{A})\in\mathcal{C}_i}(\tilde{R}_{j\to\mathcal{A}} - R_{j\to\mathcal{A}})}$$

Hence

$$P_e \leq \sum_{\mathcal{C}_i \subseteq \mathcal{D}_i} 2^{-N\left[\sum_{(j\to\mathcal{A})\in\mathcal{C}_i} R_{j\to\mathcal{A}} - \sum_{(j\to\mathcal{A})\in\mathcal{C}_i} \tilde{R}_{j\to\mathcal{A}} + \sum_{(j\to\mathcal{A})\in\mathcal{C}_i} H(U_{j\to\mathcal{A}}) - H(U_{\mathcal{C}_i}|U_{\mathcal{D}_i\setminus\mathcal{C}_i}, Y_i) - \epsilon_3 - \epsilon_4\right]}$$

Thus, by selecting sufficiently large N and the rates such that

$$\sum_{(j\to\mathcal{A})\in\mathcal{C}_i} R_{j\to\mathcal{A}} > \sum_{(j\to\mathcal{A})\in\mathcal{C}_i} \tilde{R}_{j\to\mathcal{A}} - \sum_{(j\to\mathcal{A})\in\mathcal{C}_i} H(U_{j\to\mathcal{A}}) + H(U_{\mathcal{C}_i}|U_{\mathcal{D}_i\setminus\mathcal{C}_i}, Y_i) + \epsilon_3 + \epsilon_4$$

P_e can be made arbitrarily small. Note that $Pr(E_3 \cap E_0^c) \leq \sum_{i\in[M]} Pr(E_{3,\mathcal{D}_i} \cap E_0^c)$. Thus, $Pr(E_3 \cap E_0^c)$ is small.

At node $i \in [M]$ when $\left(U^N_{\mathcal{D}_i}(\boldsymbol{\ell}_{\mathcal{D}_i}), Y^N_{\mathcal{E}(\mathcal{D}_i)}, Y_i^N, T^N\right) \in A^*_\epsilon(U_{\mathcal{D}_i}, Y_{\mathcal{E}(\mathcal{D}_i)}, Y_i, T)$, the empirical distribution on $\mathcal{T} \times \mathcal{Y}_i \times \mathcal{U}_{\mathcal{D}_i}$ is approximately equal to the true distribution, and, thus the expected distortion is approximately D_i. From Lemma 1, the probability of succesful coding $\to 1$ as $N \to \infty$. Thus, the overall expected distortion is approximately D_i at node $i \in [M]$. This completes the proof, i.e. **conv** $(\mathcal{RD}_{in}) \subseteq \mathcal{RD}$. ■

APPENDIX B
EXTENDED MARKOV LEMMA AND PROOF

We state Lemma 1 below which we use to show that $P(E_2)$ is small. We derive this Lemma from the *Generalized Markov Lemma* (Lemmas 3.3 and 3.4) in [2].

Lemma 1: Let $\{U_{j\to\mathcal{A}} \,|\, (j \to \mathcal{A}) \in \mathcal{S}\}$ be a set of auxiliary random variables which are jointly distributed with T and $Y_{[M]}$ such that $T, Y_{[M]\setminus j}, U_{\mathcal{S}\setminus\mathcal{S}_j} \leftrightarrow Y_j \leftrightarrow U_{\mathcal{S}_j}$ for all $j \in [M]$. For each $(j \to \mathcal{A}) \in \mathcal{S}$, let $\left\{U^N_{j\to\mathcal{A}}(k_{j\to\mathcal{A}}) \,|\, k_{j\to\mathcal{A}} \in \{1, \ldots, 2^{n\tilde{R}_{j\to\mathcal{A}}}\}, \tilde{R}_{j\to\mathcal{A}} > I(Y_j; U_{j\to\mathcal{A}})\right\}$ be a set of mutually independent random vectors (code-words) elements of which are drawn i.i.d. according to $\prod_{n=1}^N p_{U_{j\to\mathcal{A}}}(u^{(n)}_{j\to\mathcal{A}})$. The random vectors are independent from $T, Y_{[M]}$

and $U_{\mathcal{S}}$. If we select the codeword $U^N_{j\to\mathcal{A}}(m_{j\to\mathcal{A}})$ from the set $\left\{U^N_{j\to\mathcal{A}}(1),\ldots,U^N_{j\to\mathcal{A}}(2^{n\tilde{R}_{j\to\mathcal{A}}})\right\}$ for all $(j\to\mathcal{A})\in\mathcal{S}$ such that $\left(U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j}),\, Y^N_j\right)\in A^*_\epsilon(U_{\mathcal{S}_j},Y_j)$, $j\in[M]$, then for all $i\in[M]$

$$Pr\left\{\left(U^N_{\mathcal{D}_i}(\boldsymbol{m}_{\mathcal{D}_i}), Y^N_{\mathcal{E}(\mathcal{D}_i)}, Y^N_i, T^N\right)\in A^*_\epsilon\Big(U_{\mathcal{D}_i}, Y_{\mathcal{E}(\mathcal{D}_i)}, Y_i, T\Big)\right\}\geq 1-\delta \tag{59}$$

where $\delta\to 0$ as $\epsilon\to 0$ for sufficiently large N.

Proof: This Lemma can be proved by imitating the proof of Lemma 3.4 in [2] which is proved by induction. Suppose that $U^N_{\mathcal{S}_1}(\boldsymbol{m}_{\mathcal{S}_1}),\ldots,U^N_{\mathcal{S}_{j-1}}(\boldsymbol{m}_{\mathcal{S}_{j-1}})$ have been defined such that they satisfy the following properties.

1) $U^N_{\mathcal{S}_a}(\boldsymbol{m}_{\mathcal{S}_a})$ take values in $\left\{U^N_{\mathcal{S}_a}(1),\ldots,U^N_{\mathcal{S}_a}(2^{n\tilde{R}_{\mathcal{S}_a}})\right\}$ for all $a\in\{1,\ldots,j-1\}$
2) For all $i\in[M]$

$$Pr\left\{\left(U^N_{\mathcal{D}_i\cap\mathcal{P}_{j-1}}(\boldsymbol{m}_{\mathcal{D}_i\cap\mathcal{P}_{j-1}}), U^N_{\mathcal{D}_i\cap\bar{\mathcal{P}}_{j-1}}, Y^N_{\mathcal{E}(\mathcal{D}_i)}, Y^N_i, T^N\right)\in A^*_\epsilon(U_{\mathcal{D}_i}, Y_{\mathcal{E}(\mathcal{D}_i)}, Y_i, T)\right\}\geq 1-\delta_{j-1}$$

where $\delta_{j-1}\to 0$ as $\epsilon\to 0$ and $\mathcal{P}_{j-1}:=\{\mathcal{S}_a\mid a\in\{1,\ldots,j-1\}\}$. Here $\bar{\mathcal{P}}_{j-1}=\mathcal{S}\setminus\mathcal{P}_{j-1}$.

Note that when $j=1$, $\{1,\ldots,j-1\}=\{\}$. Thus the properties (a) and (b) are satisfied for $j=1$ by the condition (1) of Theorem 1.

Define $V=U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j})$ and $Y=Y_j$ and for all $\{i\mid i\in\mathcal{A},\ (j\to\mathcal{A})\in\mathcal{S}_j\}$

$$W^*_i=U^N_{\mathcal{D}_i\cap\mathcal{P}_{j-1}}(\boldsymbol{m}_{\mathcal{D}_i\cap\mathcal{P}_{j-1}}), U^N_{\mathcal{D}_i\cap\bar{\mathcal{P}}_j}, Y^N_{\mathcal{E}(\mathcal{D}_i)\setminus j}, Y^N_i, T^N$$

Then from condition (1) of Theorem 1 and properties (a) and (b), all the conditions of Lemma 3.3 in [2] are satisfied. Thus

$$Pr\left\{\left(U^N_{\mathcal{D}_i\cap\mathcal{P}_j}(\boldsymbol{m}_{\mathcal{D}_i\cap\mathcal{P}_{j-1}}), U^N_{\mathcal{D}_i\cap\bar{\mathcal{P}}_j}, Y^N_{\mathcal{E}(\mathcal{D}_i)}, Y^N_i, T^N\right)\in A^*_\epsilon\left(U_{\mathcal{D}_i}, Y_{\mathcal{E}(\mathcal{D}_i)}, Y_i, T\right)\right\}\geq 1-\delta_j$$

where $\delta_j=\delta_j(\delta_{j-1},\epsilon)\to 0$ as δ_{j-1} and $\epsilon\to 0$. Thus properties (a) and (b) hold for j instead of $j-1$. Repeating the induction until $j=M$ proves the lemma. ■

REFERENCES

[1] V. Delouille, R. Neelamani, and R. Baraniuk, "Robust distributed estimation in sensor networks using the embedded polygons algorithm," in *Proceedings of the 3rd international symposium on Information processing in sensor networks*, 2004.
[2] T. S. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Transactions on Information Theory*, vol. IT-26, no. 3, pp. 277–288, May 1980.
[3] A. B. Wagner and V. Anantharam, "An improved outer bound for multiterminal source coding," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1919–1937, May 2008.
[4] R. L. Dobrushin, "Survey of soviet research in information theory," *IEEE Transactions on Information Theory*, vol. IT-18, no. 6, pp. 703–724, November 1972.
[5] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. IT-22, no. 1, pp. 1–10, January 1976.
[6] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem," *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 887–902, May 1996.
[7] S. C. Draper and G. W. Wornell, "Side information aware coding strategies for sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 966–976, August 2004.
[8] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2210–2219, June 2005.
[9] 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 Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 977–987, August 2004.
[10] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Transactions on Information Theory*, vol. IT-43, no. 6, pp. 1912–1923, November 1997.
[11] ——, "Rate-distortion theory for gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2577–2593, July 2005.
[12] V. Prabhakaran, K. Ramchandran, and D. Tse, "On the role of interaction between sensors in the ceo problem," in *Forty-Second Annual Allerton Conference on Communication, Control, and Computing*, 2004.
[13] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. IT-28, no. 6, pp. 851–857, November 1982.
[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc, 2006.
[15] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371–381, February 2003.
[16] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels - part i: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, November 1998.
[17] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 727–734, November 1985.
[18] A. A. El Gamal and E. C. Van Der Meulen, "A proof of marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Transactions on Information Theory*, vol. IT-27, no. 1, pp. 120–122, January 1981.