Coding Perspectives for Collaborative Estimation Over Networks

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

***Abstract*—A collaborative distributed estimation problem over a communication constrained network is considered from an information theory perspective. 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 message to each subset of other nodes are studied. This code construction hybridizes 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. An achievable rate distortion region 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 $j \in [M]$ ($\{1, \dots, M\}$ is denoted as $[M]$) 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 the sequence $\left(T^{(n)}, Y_1^{(n)}, \dots, Y_M^{(n)}\right)$ be i.i.d. according to joint probability distribution $p_{T,Y_1,\dots,Y_M}$.

Each node could use the local observations to obtain Bayesian estimates $\tilde{T}_j^{(n)}$ of $T^{(n)}$ that minimize some local cost function $\frac{1}{N}\sum_{n=1}^{N} E\left[d_j(\tilde{T}_j^{(n)}, T_j^{(n)}) \,\middle|\, Y_j^N\right]$ (The vector $\left[Y_j^{(1)}, \dots, Y_j^{(N)}\right]$ is denoted as Y_j^N). Alternatively, the nodes in the network could communicate with each other in hopes of improving their estimates. We will study such collaborative distributed estimation 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 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 (e.g. with the aid of some channel and network codes) to the nodes in $\mathcal{A}$. For example, in the lowest dimensional $M = 3$ nontrivial case, each node will create multiple descriptions of its observations, one for each of the other two nodes in the network individually, and one for both of them as shown in Fig. 1.

S. Ramanan and J. M. Walsh 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.

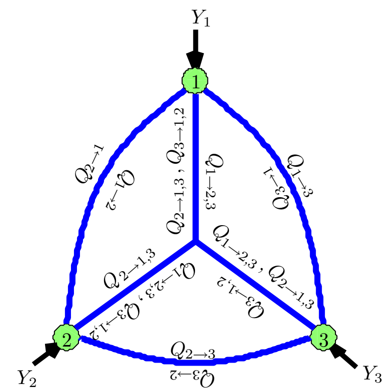

Fig. 1. The "peace" network, which depicts the lowest dimensional $M = 3$ nontrivial case of the problem of collaborative distributed estimation. To determine the direction of travel of a message, note that the messages flow in the direction in which they are read.

We employ a classic technique from multiterminal information theory [1] [2] to study the relationship between the rates $\{R_{j\to\mathcal{A}} \mid j \in [M],\ \mathcal{A} \in 2^{[M]\setminus j}\}$ ($2^{\mathcal{S}}$ is the power set of subsets of $\mathcal{S}$) of the source code used, and the estimation errors D_j that each of the nodes can obtain in estimating the sequence $T^{(n)}, n \in [N]$ from their own observations Y_j^N and the messages $Q_{\mathcal{D}_j} := \left[Q_{i\to\mathcal{A}} \,\middle|\, j \in \mathcal{A},\ \mathcal{A} \in 2^{[M]\setminus i}\right]$ they have received.

One might view this model as a generalization of two classes of multiterminal information theory problems: CEO problem [1], [3], [4], [2], [5] and multiple descriptions problem [6]. The CEO problem studies the rate - estimation error performance at the fusion center, which estimates an underlying sequence of parameters solely based on the rate constrained messages received from a collection of nodes which independently encode using noisy local observations. Two variations on the CEO problem are studied in [7] when decoder side information [8] is available at the CEO and in [9] under the name of many-help-one problem when one of the nodes directly observe the underlying sequence. In multiple descriptions problem rate estimation error performance is studied in the case in which a node encodes many descriptions of a source and sends different subset of descriptions to different decoders which use those descriptions to reproduce the source. A reader who is familiar with these two classes of problems may question the purpose of studying this model when both classes of problems are yet unsolved. Interestingly, we show that some known bounds for the CEO problem and the multiple

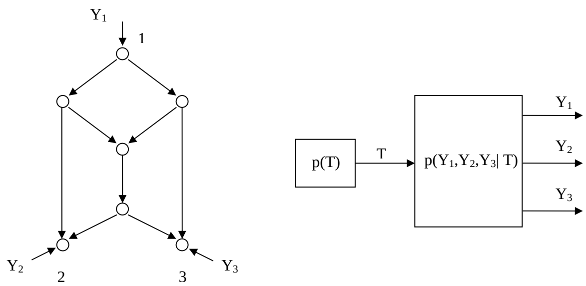

Fig. 2. This network demonstrates that considering a source code at node 1 which only encodes a dedicated message to node 2 and a dedicated message to node 3 is not general enough. Instead, the source encoder at node 1 should encode a separate message for each possible subset of other nodes in the network.

descriptions problem can be recovered from the results we derive in this paper by hybridizing the techniques from both classes of problems.

The paper is organized as follows. We discuss the best model for the distributed source code in Section II. In Section III, we present our main results on achievable rate distortion region. In Section IV, we simplify the inner bound for some simpler problems.

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.

To see that such a model is the appropriate one, consider a simple wired network depicted in Fig. 2 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 $Q_{1\to 2}, Q_{1\to 3}, Q_{1\to 2,3}$, 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 $R_{1\to 2,3} = 2$ [10]. 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.

From this simple example 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 $Q_{i\to\mathcal{A}}$ 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

$$\mathbf{r} := [R_{j\to\mathcal{A}} \,|j \in [M],\ \mathcal{A} \subseteq [M] \setminus j] \tag{1}$$

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

$$\mathbf{d} := [D_i | i \in [M]]\,, \quad D_i := \frac{1}{N} \sum_{n=1}^{N} E\left[d_i\left(T^{(n)}, \hat{T}_i^{(n)}\right)\right] \tag{2}$$

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 $(\mathbf{r}, \mathbf{d}) \in \mathcal{RD}$. We now focus our efforts on describing the rate distortion region for the associated family of source codes we have selected.

III. Achievable Rate Distortion Region

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

$$f_{j\to\mathcal{A}}^N : \mathcal{Y}_j^N \to [L_{j\to\mathcal{A}}^N], \quad g_i^N : \mathcal{Y}_i^N \times \prod_{(j\to\mathcal{A})\in\mathcal{D}_i} [L_{j\to\mathcal{A}}^N] \to \hat{T}_i \tag{3}$$

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

$$R_{j\to\mathcal{A}} \geq \frac{1}{N} \log L_{j\to\mathcal{A}}^N, \quad E\left[\frac{1}{N}\sum_{n=1}^{N} d_i(T^{(n)}, \hat{T}_i^{(n)})\right] \leq D_i \tag{4}$$

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

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 the set $\{U_{i\to\mathcal{A}} \mid \mathcal{A} \in 2^{[M]\setminus i}\}$ as $U_{\mathcal{S}_i}$. If we define $\mathcal{S} := \bigcup_{i\in[M]} \mathcal{S}_i$, then we have the following theorem.

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 $\boldsymbol{\xi} \in \Xi(\mathbf{d})$, define $\boldsymbol{\Phi}(\boldsymbol{\xi})$ as in (5). Also, for each $\boldsymbol{\xi} \in \Xi(\mathbf{d})$ and for each $\boldsymbol{\phi} \in \boldsymbol{\Phi}(\boldsymbol{\xi})$, define $\mathcal{RD}_{in}(\boldsymbol{\xi}, \boldsymbol{\phi})$ as in (6).

Let

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

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

$$\mathbf{\Phi}(\boldsymbol{\xi}) = \left\{\tilde{R}_{\mathcal{S}} : \sum_{(j\to\mathcal{A})\in\mathcal{P}_j} \tilde{R}_{j\to\mathcal{A}} > \sum_{(j\to\mathcal{A})\in\mathcal{P}_j} H(U_{j\to\mathcal{A}}) - H(U_{\mathcal{P}_j}|Y_j),\ \forall\ \mathcal{P}_j \subseteq \mathcal{S}_j,\ j\in[M]\right\} \tag{5}$$

$$\mathcal{RD}_{in}(\boldsymbol{\xi},\phi) = \left\{R_{\mathcal{S}} : \sum_{(j\to\mathcal{A})\in\mathcal{C}_i} R_{j\to\mathcal{A}} \geq \sum_{(j\to\mathcal{A})\in\mathcal{C}_i} \Big(\tilde{R}_{j\to\mathcal{A}} - H(U_{j\to\mathcal{A}})\Big) + H(U_{\mathcal{C}_i}|U_{\mathcal{D}_i\setminus\mathcal{C}_i}, Y_i),\ \forall\ \mathcal{C}_i\subseteq\mathcal{D}_i,\ i\in[M]\right\} \tag{6}$$

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 [1], with the twist that the multiple (dependent) descriptions at each encoder require an additional set of encoder inequalities. A sketch of the proof is provided in Appendix A. A more detailed proof may be found in [11]. □

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

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

where $\mathcal{L}_j := (\mathcal{P}\cap\mathcal{S}_j)\setminus(\mathcal{Q}\cap\mathcal{S}_j)$. 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

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

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

$$\phi_{\pi(1)}(\boldsymbol{\pi}) \triangleq \rho(\pi(1)) = I(U_{\pi(1)}; Y_{[M]})$$

and for every $i\in\{2,\ldots,|\mathcal{S}|\}$

$$\begin{aligned} \phi_{\pi(i)}(\boldsymbol{\pi}) &\triangleq \rho(\{\pi(1),\ldots,\pi(i)\}) - \rho(\{\pi(1),\ldots,\pi(i-1)\}) \\ &= I(U_{\pi(i)}; U_{\{\pi(1),\ldots,\pi(i-1)\}}, Y_{[M]}) \end{aligned} \tag{8}$$

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 [12]. □

We next use these structural properties of the achievable rate distortion region to simplify this bound to two simpler problems: the multiple descriptions problem and the CEO problem.

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. The coding strategy introduced in [6] to this problem can be accomplished by dividing $Q_{1\to\{4\}}$ 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 $\Delta_1+\Delta_2 = R_{1\to\{4\}}$ and forming 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 [6] 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 [6] uses all of the encodings $(Q_{1\to\{2,4\}}, Q_{1\to\{3,4\}}, Q_{1\to\{4\}})$. 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

$$U_{1\to\{2,3\}} \triangleq U_{2,3},\ U_{1\to\{3,4\}} \triangleq U_{3,4},\ U_{1\to\{4,\}} \triangleq U_4 \quad (9)$$
$$R_{1\to\{2,3\}} \triangleq R_{2,3},\ R_{1\to\{3,4\}} \triangleq R_{3,4},\ R_{1\to\{4,\}} \triangleq R_4 \quad (10)$$
$$\tilde{R}_{1\to\{2,3\}} \triangleq \tilde{R}_{2,3},\ \tilde{R}_{1\to\{3,4\}} \triangleq \tilde{R}_{3,4},\ \tilde{R}_{1\to\{4\}} \triangleq \tilde{R}_4 \quad (11)$$
$$U_{j\to\mathcal{A}} = \emptyset,\ R_{j\to\mathcal{A}} = \emptyset,\ \tilde{R}_{j\to\mathcal{A}} = \emptyset \text{ all other } \mathcal{A} \quad (12)$$

Where the auxiliary random variables $U_4, U_{2,4}, U_{3,4}$ are selected such that

$$p(U_4, U_{2,4}, U_{3,4}, T) = p(T)p(U_4, U_{2,4}, U_{3,4}|T) \quad (13)$$

$$D_1 \geq \mathbb{E}\left[d(T,\hat{T}_2)\right],\ D_2 \geq \mathbb{E}\left[d(T,\hat{T}_3)\right],\ D_0 \geq \mathbb{E}\left[d(T,\hat{T}_4)\right]$$

Under these identifications, the inner bound becomes

$$\begin{aligned} R_4 &\geq \tilde{R}_4 - H(U_4) + H(U_4|U_{3,4}, U_{2,4}) & (14)\\ R_{2,4} &\geq \tilde{R}_{2,4} & (15)\\ R_{3,4} &\geq \tilde{R}_{3,4} & (16)\end{aligned}$$

Having the inequalities $R_1 \geq R_{2,4},\ R_2 \geq R_{3,4}$ (because $\Delta_1, \Delta_2 \geq 0$) in hand, we replace R_4 with $R_1 - R_{2,4} + R_2 - R_{3,4}$ in (14) and use the inequalities (14)-(16) to obtain a bound on the rate region (R_1, R_2) which is given by

$$\begin{aligned} R_1 &\geq \tilde{R}_{2,4} & (17)\\ R_2 &\geq \tilde{R}_{3,4} & (18)\\ R_1 + R_2 &\geq \tilde{R}_{2,4} + \tilde{R}_{3,4} + \tilde{R}_4 & \\ &\quad -H(U_4) + H(U_4|U_{3,4}, U_{2,4}) & (19)\end{aligned}$$

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 (19) becomes

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

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) and present the values of $\tilde{R}_{2,4}, \tilde{R}_{3,4}$ below.

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

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

$$\begin{aligned} R_1 &\geq I(U_{2,4}; T) + \alpha\ I(U_{2,4}; U_{3,4}|T) & (20)\\ R_2 &\geq I(U_{3,4}; T) + (1-\alpha)\ I(U_{2,4}; U_{3,4}|T) & (21)\\ R_1 + R_2 &\geq I(U_{2,4}; U_{3,4}) + I(T; U_4, U_{2,4}, U_{3,4}) & (22)\end{aligned}$$

where $0 \leq \alpha \leq 1$. We next show that any point in the achievable rate region (ECG region) proved in [6] also lies in the region we proved above. To prove this, we rewrite the EGC region in the following form

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

and let

$$\alpha = \min\left\{\frac{r_1 - I(U_{2,4}; T)}{I(U_{2,4}; U_{3,4}|T)}, 1\right\} \quad (23)$$

Then

$$R_1 \geq \min\left\{r_1, I(U_{2,4}; T) + I(U_{2,4}; U_{3,4}|T)\right\} \leq r_1 \quad (24)$$

and

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

In the above proof we used the following inequality which can be easily proved.

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

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 our inner bound simplifies the Berger-Tung inner bound for this case. To see this, suppose 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 these 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 $E[d(T, \hat{T})] < D$.

Since the nodes $i \in [M-1]$ send messsages only to node M, we set the rates corresponding to the other messages to 0 and redefine the rates and variables relevent to this problem as follows.

$$R_{j\to M} \triangleq R_j,\ U_{j\to M} \triangleq U_j\ \forall\ j \in [M-1]$$
$$R_{j\to\mathcal{A}} = 0,\ \tilde{R}_{j\to\mathcal{A}} = 0, U_{j\to\mathcal{A}} = \emptyset \text{ all other } \mathcal{A}, j \in [M-1]$$
$$D_M \triangleq D,\ R_{M\to\mathcal{A}} = 0,\ \tilde{R}_{M\to\mathcal{A}} = 0, U_{M\to\mathcal{A}} = \emptyset$$

Note that the random vectors $\boldsymbol{\xi} = (U_{[M-1]})$ satisfy the following constraints.

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

If we denote the set $[M-1]$ as $\mathcal{D} := [M-1]$ as, then $\Phi(\boldsymbol{\xi})$ becomes

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

Here, $\tilde{R}_j$ can be selected such that $\tilde{R}_j = I(U_j; Y_j) + \epsilon_j$ 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 $\tilde{R}_j = I(U_j; Y_j) + \epsilon_j$, 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

$$\begin{aligned}\sum_{j\in\mathcal{C}} R_j &> \sum_{j\in\mathcal{C}}(\tilde{R}_j - H(U_j)) + H(U_{\mathcal{C}}|U_{\mathcal{D}\setminus\mathcal{C}}) \\ &= H(U_{\mathcal{C}}|U_{\mathcal{D}\setminus\mathcal{C}}) - \sum_{j\in\mathcal{C}} H(U_j|Y_j) \\ &= H(U_{\mathcal{C}}|U_{\mathcal{D}\setminus\mathcal{C}}) - H(U_{\mathcal{C}}|Y_{\mathcal{C}}) \\ &= H(U_{\mathcal{C}}|U_{\mathcal{D}\setminus\mathcal{C}}) - H(U_{\mathcal{C}}|Y_{\mathcal{C}},U_{\mathcal{D}\setminus\mathcal{C}}) \\ &= I(U_{\mathcal{C}};Y_{\mathcal{C}}|U_{\mathcal{D}\setminus\mathcal{C}})\end{aligned}$$

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 $U_{\mathcal{C}} \leftrightarrow Y_{\mathcal{C}} \leftrightarrow U_{\mathcal{D}\setminus\mathcal{C}}$. Thus the 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 of random vectors $\boldsymbol{\xi}$. This is exactly the Berger Tung inner bound for the CEO problem given in [4].

C. *Simplification to* Side Information May be Absent *case*

The problem studied in [13] 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 [13]. We can 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 [13]. We skip the proof to conserve the space and refer any interested reader to [11]. We can also show that the sum rate result proved for the general problem with degraded side information can be retrieved from our inner bound.

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 a sketch of the proof of the inner bound given in Section III here.

Proof Select a joint conditional distribution $p\left(u_{\mathcal{S}} \,\middle|\, t, y_{[M]}\right)$, 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 $\left(U^N_{\mathcal{S}_j}(\boldsymbol{m}_{\mathcal{S}_j}), Y^N_j\right)\in A^*_\epsilon(U_{\mathcal{S}_j}, Y_j)$, where A^*_ϵ is the set of strongly typical sequences. 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 $\left(U^N_{\mathcal{D}_i}(\boldsymbol{\ell}_{\mathcal{D}_i}), Y^N_i\right)\in A^*_\epsilon(U_{\mathcal{D}_i}, Y_i)$, 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 $\hat{T}^N_i = g^N_i\left(Y^N_i, U^N_{\mathcal{D}_i}(\boldsymbol{\ell}_{\mathcal{D}_i})\right)$. $\square$

REFERENCES

[1] 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.

[2] 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.

[3] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem," *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[4] 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.

[5] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Transactions on Information Theory*, vol. IT-43, no. 6, pp. 1912–1923, November 1997.

[6] 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.

[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] 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.

[9] Y. Oohama, "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.

[10] 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.

[11] J. M. Walsh and S. Ramanan, "Coding perspectives for collaborative distributed estimation over networks," proof of inner bound. [Online]. Available: http://www.ece.drexel.edu/walsh/web/listOfPubs.html

[12] 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.

[13] 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.