Distributed Lossy Interactive Function Computation

Solmaz Torabi & John MacLaren Walsh
Dept. of Electrical and Computer Engineering
Drexel University
Philadelphia, PA 19104
Email: solmaz.t@drexel.edu & jwalsh@ece.drexel.edu

***Abstract*—Several users observing random sequences take turns sending error-free messages to a central estimation officer (CEO) and all other users one at a time. The CEO, which also observes a side information correlated with the users observations, aims to compute a function of the sources and the side information in a lossy manner. The users' observations are assumed to be conditionally independent given the side information. Inner and outer bounds to the rate distortion region for this lossy interactive distributed function computation problem are established and shown to coincide in some special cases. In addition to this newly closed case, two more examples are provided in which each user observes a subset of independent random variables, and the full rate distortion region is characterized. Additionally, the relationship between a zero-distortion limit and lossless distributed function computation is studied for a special class of extremum functions and related distortions.**

I. Introduction

In decentralized function computation, a series of two or more networked nodes, each storing or measuring local observation sequences, communicate with a sink, henceforth the central estimation officer (CEO), with the goal of enabling it to compute some function or statistic of their combined data across each element in the sequence. Over the years, substantial attention has been given to obtaining the minimum amount of information that must be exchanged in this problem in various senses and contexts.

In perhaps the simplest, *non-interactive lossless* context, each node sends exactly one message in parallel with all other nodes, and the CEO must compute the function losslessly, i.e. with diminishing block probability of error [1, 2]. An important early result [3] showed that for binary doubly symmetric sources, if the CEO aims to compute exclusively the binary sum of the sources, a lower rate is necessary than would be required for losslessly communicating the sources with Slepian and Wolf coding [4]. Additionally, [5] provided an inner bound which was shown to be optimal if the sources are conditionally independent given the function. Functions can be categorized based on whether their achievable rate region always coincides with the Slepian-Wolf region[6]. In some cases, the fundamental lossless non-interactive rate limit can be achieved by Slepian-Wolf coding of local graph colors [2]. Inner and outer bounds have also been proved a larger class of tree structured function computation networks, and these bounds have been shown to be equal if the source random variables obey certain Markov properties[1].

In the *non-interactive lossy* variant of this problem, the CEO may incur a small distortion with respect to some fidelity measure when computing the function [7]. The rate distortion region for the general CEO problem remains unknown, however, when the sources are jointly normally distributed and the distortion is squared error it has been determined in [8, 9]. The rate distortion region takes an especially simple form when the sources are independent, enabling a Blahut Arimoto-like algorithm to compute it numerically [10] for particular source distributions and distortion measures. Gastpar [11] studied the closely related multiterminal source coding problem, wherein the decoder, which has access to the side information, reproduces the users observations subject to separate individual distortion constraints. Inner and outer bounds for the rate region for the multiterminal source coding problem were derived [11], and were shown to be equal to one another under the special case where the sources are conditionally independent given the side information.

Interactive function computation has also been considered in another important variant of the CEO problem called *lossless function computation in collocated network*. In a collocated network, the users take turns sending messages to the CEO, with each user perfectly overhearing the message from every other user. In the Gaussian, squared error, collocated CEO problem, [12] the ability to interact does not result in reduction in the sum-rate compared with the non-interactive CEO problem. For decentralized function computation from discrete sources over collocated networks, [13] provided the worst case rate in the limit of large number of users for two classes of functions, the type threshold functions, and type sensitive functions. Building upon these results and information theoretic results from the point to point context [14–16], Ma et al, in [17] studied this problem in a distributed source coding framework, and proposed an iterative algorithm to compute the minimum sum rate of for any finite and infinite number of rounds.

In this paper, building upon this prior work, we study

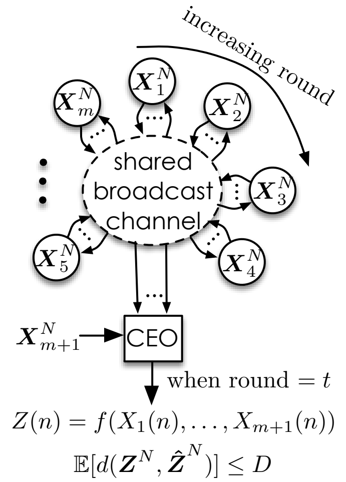

Fig. 1: System diagram

fundamental limits for lossy interactive function computation over a collocated network with side information. After formally defining the problem in §II, and reviewing a simple cutset outer bound and an inner bound based on a straightforward achievability construction in §III-A and III-B, we derive a tighter outer bound for the rate distortion region in §III-C. This outer bound is shown to be tight if the users observations and the side information are independent. In addition to this newly closed case, two more examples are provided in which each user observes a subset of independent random variables, and the full rate distortion region is characterized in §IV. Finally, §V considers example cases where the functions to compute are extrema, and reviews the relationship between lossless function computation limits and the zero distortion limit for distortion functions specifically tailored to this context.

II. Problem formulation

Consider a network with m source terminals and a single sink terminal as depicted in Fig. 1. Each source terminal $j \in \{1, ...m\}$ observes a random sequence $\boldsymbol{X}_j^N = (X_j(1), ..., X_j(N)) \in \mathcal{X}_j^N$. The sink node, the central estimation officer (CEO), observes a side information sequence $X_{m+1}(n), n \in [N]$ which is assumed to be correlated with the source variables. The random vectors $X_{1:m+1}(n) = (X_1(n), .., X_{m+1}(n))$ are iid in time $n \in [N]$, and, given the side information $X_{m+1}(n)$, the source variables $X_{1:m}(n)$ are independent. The sink terminal wishes to compute $f : \mathcal{X}_1 \times ... \times \mathcal{X}_m \times \mathcal{X}_{m+1} \to \mathcal{Z}$ in a lossy manner elementwise, estimating the sequence $\boldsymbol{Z}^N = (Z(1), ..., Z(N))$ with $Z(n) = f(X_1(n), X_2(n), ..., X_m(n), X_{m+1}(n))$. Let $d : \mathcal{X}_1 \times \mathcal{X}_2 \times ... \times \mathcal{X}_m \times \mathcal{X}_{m+1} \times \mathcal{Z} \to \mathbb{R}^+$ be a fidelity criterion for this function computation, yielding the block distortion metric

$$d^{(N)}(\boldsymbol{X}_{1:m+1}^N, \hat{\boldsymbol{Z}}^N) = \frac{1}{N} \sum_{n=1}^{N} d(x_1(n), ..., x_{m+1}(n), \hat{z}(n))$$

In order to enable the CEO to estimate the function computation, the nodes take turns broadcasting messages which are perfectly received at both the other nodes and the CEO. The communication is initiated by node 1, and at round i, node $j = ((i-1) \bmod m) + 1$ sends a message $M_i \in \mathcal{M}_i$, using the encoding function $\psi_i : \mathcal{X}_j^N \times \bigotimes_{k=1}^{i-1} \mathcal{M}_k \to \mathcal{M}_i$ to encode its observations based on the previously overheard messages. After t rounds of communication, the CEO estimates the function sequence based on the messages it has received and the side information using the decoding function $\phi : \mathcal{X}_{m+1}^N \times \bigotimes_{k=1}^{t} \mathcal{M}_k \to \mathcal{Z}^N$.

Definition 1. A rate-distortion tuple $(\mathbf{R}, D) = (R_1, R_2, ..., R_t, D)$ is *admissible* for t-message interactive function computation, if $\forall \epsilon > 0$, and $\forall N > n(\epsilon, t)$, there exist encoders $\psi_i, i \in \{1, \ldots, t\}$ and a decoder ϕ with parameters $(t, N, |\mathcal{M}_1|, ..., |\mathcal{M}_t|)$ satisfying $\frac{1}{N} \log_2 |\mathcal{M}_i| \leq R_i \ \forall i = 1, ..., t$ and $\mathbb{E}\big[d^{(N)}(\boldsymbol{X}_{1:m+1}^N, \hat{\boldsymbol{Z}}^N)\big] \leq D + \epsilon$ with $\hat{\boldsymbol{Z}} = \phi(M_1, \ldots, M_t, \boldsymbol{X}_{m+1}^N)$.

Finally, define the collection of admissible rate and distortion vectors $(\mathbf{R}, D)$ to be $\mathcal{RD}^t$.

III. Characterization of the Rate-Distortion region

A. Cut-Set outer bound

A simple outer bound for the rate distortion region $\mathcal{RD}^t$ can be created from a two terminal rate distortion function. In particular, let $R_{A \to B|s}^{TW}$ be the rate distortion function for terminal A, where it engages in a two terminal lossy interactive function computation, and a side information s is available at terminal B, [14], [16]:

$$R_{A \to B|s}^{TW}(D) = \min_{P_{U^t|X_A, X_B} \in \mathcal{C}(D)} I(X_A, U^t | X_B, X_s) \quad (1)$$

with $\mathcal{C}(D)$ the set of conditional distributions for $U^T = (U_1, \ldots, U_t)$ obeying $U_\ell \leftrightarrow (X_A, U_{1:\ell-1}) \leftrightarrow X_B, X_s$ for ℓ odd and $U_\ell \leftrightarrow (X_B, X_s, U_{1:\ell-1}) \leftrightarrow X_A$ for ℓ even, such that there exists some $\hat{g}(U^t, X_B, X_s)$ with $\mathbb{E}(d(X_A, X_B, X_s, \hat{g}(U^t, X_B, X_s))) \leq D$.

To lower bound the sum rate for node $j \in \{1, .., m\}$, we consider a cut between node X_j and super node $X_{[m+1] \setminus \{i\}}$. The supernode contains the side information and aims to compute the function $f(X_1, ..., X_{m+1})$ up to a distortion D. In the next lemma, the cutset outer bound is constructed by lower bounding the sum rate R_i for all the users $i = 1, ..., m$.

Lemma 1. *Cutset outer bound: Any achievable sum rate distortion tuples must lie in* $\mathcal{RD}_{cutset}^t$ *which can be characterized as follows*

$$\mathcal{RD}_{cutset}^t = \Big\{ R_{sum,j}(D) \geq R_{j \to [m] \setminus \{j\} | m+1}^{TW}(D), \ \forall j \in [m] \Big\}$$

where $R_{sum,j} = \sum_{k=0}^{\lfloor t/m \rfloor} R_{j+mk}$.

B. Inner bound: Achievability scheme

In this section we derive an achievable region, $\mathcal{RD}^t_{inn} \subseteq \mathcal{RD}^t$. The coding technique that leads to this region is a sequence of Wyner-Ziv coding.

Theorem 1. *An inner bound $\mathcal{RD}^t_{inn}$ for the t-message interactive lossy function computation with side information region $\mathcal{RD}^t$, consists of the set of $(\boldsymbol{R}, D)$ such that there exist a vector $U_{1:t} = (U_1, ..., U_t)$ of discrete random variables with $p(X_{1:m+1}, U_{1:t}) =$*

$$p(X_{m+1}) \prod_{i=1}^{t} p(U_i|U_{1:i-1}X_j) \prod_{k=1}^{m} p(X_k|X_{m+1})$$

which satisfy the following conditions.

$$\mathcal{RD}^t_{inn} = \left\{ (\mathbf{R}, D) \left| \begin{array}{c} R_i \geq I(U_i; X_j, U^{i-1}) - \min_{k \neq j \in [m+1]} \{I(U_i; X_k, U^{i-1})\} \\ U_i \leftrightarrow X_j U_{1:i-1} \leftrightarrow X_{[m+1]\setminus\{j\}}, \\ j = ((i-1) \bmod m) + 1, \quad \forall i \in [t] \\ \mathbb{E}\big[d(X_1, ., X_{m+1}, \hat{Z}(U_{1:t}, X_{m+1}))\big] \leq D \end{array} \right. \right\}$$

Proof. (Abbreviated) For each round $i \geq 1$, node $j = ((i-1) \bmod m) + 1$ uses the standard random binning code construction treating the previously received messages $U_{1:(i-1)}$, and the user with the worst observation as side information for the determination of U_i at each of the users and the CEO, notably ignoring the extra side information from local observations that each of these participants have. The messages are constructed in such a way that guarantee the worst user and all the other users as well as the decoder can reconstruct that. After reconstructing all of the messages, the CEO also utilizes the side information X_{m+1} in addition to all the messages it has received thus far $U_{1:t}$ to estimate $\hat{Z}$ with distortion D. □

C. General outer bound

In §III-A, we established a cut-set outer bound by allowing the super node to have access to the $X_{[m+1]\setminus i}$. In the following, we establish a tighter outer bound.

Theorem 2. *Any achievable $(\boldsymbol{R}, D)$ pair must lie in $\mathcal{RD}^t_{out}$ = the convex hull of the region*

$$\left\{ (\mathbf{R}, D) \left| \begin{array}{c} \forall i \in [t] \quad R_i \geq I(X_j; U_i|U_{1:i-1}, X_{m+1}), \\ U_i \leftrightarrow X_j U_{1:i-1} \leftrightarrow X_{[m+1]\setminus\{j\}}, \\ j = ((i-1) \bmod m) + 1, \\ \mathbb{E}\big[d(X_1, ., X_{m+1}, \hat{Z}(U_{1:t}, X_{m+1}))\big] \leq D \end{array} \right. \right\}$$

Proof. Let $(R_1, R_2, ..., R_t, D) \in \mathcal{RD}^t$ be a set of admissible rate and distortion, then $\forall \epsilon > 0$, and $\forall N > n(\epsilon, t)$, there exist an interactive distributed block source code with parameters $(t, N, |\mathcal{M}_1|, ..., |\mathcal{M}_t|)$ satisfying

$$\frac{1}{N} \log_2 |\mathcal{M}_i| \leq R_i \;\; \forall i = 1, ..., t \tag{2}$$

$$E\big[d^{(N)}(\boldsymbol{X}_1^N, ..., \boldsymbol{X}_{m+1}^N, \hat{\boldsymbol{Z}}^N)\big] \leq D + \epsilon \tag{3}$$

For the first round, we have

$$\begin{aligned} &NR_1 \geq H(M_1) \geq H(M_1|\boldsymbol{X}^N_{2:m+1}) \\ &\geq I(\boldsymbol{X}^N_1; M_1|\boldsymbol{X}^N_{2:m+1}) \\ &= H(\boldsymbol{X}^N_1|\boldsymbol{X}^N_{m+1}) - H(\boldsymbol{X}^N_1|M_1, \boldsymbol{X}^N_{2:m+1}) \\ &= \sum_{n=1}^{N} H(X_1(n)|X_{m+1}(n)) \\ &\quad - \sum_{n=1}^{N} H(X_1(n)|M_1, \boldsymbol{X}^N_{2:m+1}, \boldsymbol{X}^{n-1}_1) \qquad (4) \\ &\geq \sum_{i=1}^{N} H(X_1(n)|X_{m+1}(n)) \\ &\quad - H(X_1(n)|M_1, \boldsymbol{X}^{n-1}_{2:m}, X_{m+1}(n), \boldsymbol{X}^{n+}_{m+1}\boldsymbol{X}^{n-1}_1) \\ &= \sum_{n=1}^{N} I\big(X_1(n); \underbrace{M_1, \boldsymbol{X}^{n-1}_{1:m}, \boldsymbol{X}^{n+}_{m+1}}_{U_1(n)}|X_{m+1}(n)\big) \qquad (5) \end{aligned}$$

with notation $\boldsymbol{X}^{n+}_{\mathcal{A}} = (X_{\mathcal{A}}(n+1), \ldots, X_{\mathcal{A}}(N))$ for any subset $\mathcal{A} \subset [m+1]$, and wherein we defined auxiliary random variables $U_1(n) = \{M_1, \boldsymbol{X}^{n-1}_{1:m}, \boldsymbol{X}^{n+}_{m+1}\}$, for $n \in [N]$. For the next rounds, $i \geq 2$ and $j = ((i-1) \bmod m)+1$ we have

$$\begin{aligned} &NR_i \geq H(M_i) \geq H(M_i|M_{1:i-1}, \boldsymbol{X}^N_{[m+1]\setminus j}) \\ &\overset{a}{=} I(\boldsymbol{X}^N_j; M_i|M_{1:i-1}, \boldsymbol{X}^N_{[m+1]\setminus j}) \\ &= \sum_{n=1}^{N} I(X_j(n); M_i|M_{1:i-1}, \boldsymbol{X}^N_{[m+1]\setminus j}, \boldsymbol{X}^{n-}_j) \\ &= \sum_{n=1}^{N} I\big(X_j(n); M_i|U_1(n), M_{2:i-1}, X_{[m+1]\setminus j}(n) \\ &\qquad\qquad , \boldsymbol{X}^{n+}_{m+1}, \boldsymbol{X}^{n+}_{[m]\setminus j}\big) \\ &\overset{b}{\geq} \sum_{n=1}^{N} I\big(X_j(n); M_i|U_1(n), M_{2:i-1}, X_{m+1}(n)\big) \\ &\overset{c}{=} \sum_{n=1}^{N} I\big(X_j(n); M_i|U_{1:i-1}(n), X_{m+1}(n)\big) \qquad (6) \end{aligned}$$

Here, in step a we have $H(M_i|M_{1:i-1}, \boldsymbol{X}^N_{1:m+1}) = 0$, since M_i is a deterministic function of $M_{1:i-1}, \boldsymbol{X}^N_j$, while in step c we defined $U_i(n) = M_i$ for $i \geq 2$. To prove step b, we need to prove that $X_j(n) \leftrightarrow M_{1:i-1}\boldsymbol{X}^{n-1}_{1:m}X_{m+1}(n)\boldsymbol{X}^{n+}_{m+1} \leftrightarrow X_{[m]\setminus j}(n)\boldsymbol{X}^{n+}_{[m]\setminus j}\boldsymbol{X}^{n-1}_{m+1}$. To prove this conditional independence, we use a technique from [18]. Note that the joint distribution of source variables and the messages can be factorized as follows:

$$\begin{aligned} &p(\boldsymbol{X}^N_1, \boldsymbol{X}^N_2, .., \boldsymbol{X}^N_{m+1}, M_1, .., M_t) \\ &= p(\boldsymbol{X}^N_{m+1}) \prod_{k=1}^{m} p(\boldsymbol{X}^{n-1}_k|\boldsymbol{X}^{n-1}_{m+1}) \prod_{\ell=1}^{m} p(X_\ell(n)|X_{m+1}(n)) \\ &\times \prod_{t=1}^{m} p(\boldsymbol{X}^{n+}_t|\boldsymbol{X}^{n+}_{m+1}) \prod_{\substack{i=1 \\ j=((i-1) \bmod m)+1}}^{t} p(M_i|\boldsymbol{X}^N_j, M_{1:i}) \end{aligned}$$

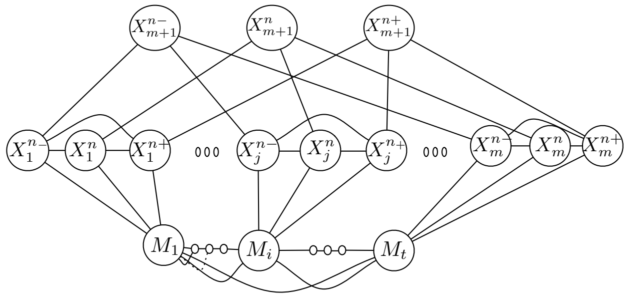

Fig. 2: Conditional independence structure

Using this factorized distribution, we can construct an undirected graph which exploit the conditional independence structure among the random variables. In this undirected graphs, the nodes are random variables appeared in the factorized distribution, and two nodes are connected if they appeared in the same factor. We have $X \perp\!\!\!\perp Y|\mathcal{V}$, if every path between X, and Y contains some node $V \in \mathcal{V}$.

Proposition 1. $X_j(n) \leftrightarrow M_{1:i-1}\boldsymbol{X}_{1:m}^{n-1}X_{m+1}(n)\boldsymbol{X}_{m+1}^{n+} \leftrightarrow X_{[m]\backslash j}(n)\boldsymbol{X}_{[m]\backslash j}^{n+}\boldsymbol{X}_{m+1}^{n-1}$.

Proof. As shown in Figure 2, any path from $X_j(n)$ to $X_{[m]\backslash j}(i)\boldsymbol{X}_{[m]\backslash j}^{n_+}\boldsymbol{X}_{m+1}^{n-1}$, contains some node from $M_{1:i-1}\boldsymbol{X}_{1:m}^{n-1}X_{m+1}(n)\boldsymbol{X}_{m+1}^{n+}$. ☐

Proposition 2. $U_i \leftrightarrow X_jU_{1:i-1} \leftrightarrow X_{[m+1]\backslash\{j\}}$.

Proof. For $i = 1$, the Markov condition $M_1, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{n+} \leftrightarrow X_1(n) \leftrightarrow X_{2:m+1}(n)$.

$$
\begin{aligned}
0 &\leq I(M_1, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{n+}; X_{2:m+1}(n)|X_1(n)) && (7)\\
&\leq I(M_1, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{1:m}^{n+}, \boldsymbol{X}_{m+1}^{n+}; X_{2:m+1}(n)|X_1(n)) && (8)\\
&= I(\boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{1:m}^{n+}, \boldsymbol{X}_{m+1}^{n+}; X_{2:m+1}(n)|X_1(n)) = 0 && (9)
\end{aligned}
$$

And for $i \geq 2$ we need to show $M_i \leftrightarrow M_{1:i-1}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{n+}, X_j(n) \leftrightarrow X_{[m+1]\backslash\{j\}}(n)$. We have

$$
\begin{aligned}
0 &\leq I(M_i; X_{[m+1]\backslash\{j\}}(n)|M_{1:i-1}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{n+}, X_j(n))\\
&\leq I(M_i; X_{[m+1]\backslash\{j\}}(n), \boldsymbol{X}_{m+1}^{n+}|M_{1:i-1}, \boldsymbol{X}_{1:m}^{n-1}, X_j(n))\\
&\leq I(M_i, M_{1:i-1}; X_{[m+1]\backslash\{j\}}(n), \boldsymbol{X}_{m+1}^{n+}|\boldsymbol{X}_{1:m}^{n-1}, X_j(n))\\
&\leq I(M_i, M_{1:i-1}\boldsymbol{X}_{1:m}^{N}; X_{[m+1]\backslash\{j\}}(n), \boldsymbol{X}_{m+1}^{n+}|\boldsymbol{X}_{1:m}^{n-1}, X_j(n))\\
&= I(M_{1:i-1}\boldsymbol{X}_{1:m}^{N}; X_{[m+1]\backslash\{j\}}(n), \boldsymbol{X}_{m+1}^{n+}|\boldsymbol{X}_{1:m}^{n-1}, X_j(n)) = 0
\end{aligned}
$$

which results

$$
I(M_i; X_{[m+1]\backslash\{j\}}(n)|M_{1:i-1}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{n+}, X_j(n)) = 0.
$$

This proves the lemma. ☐

Next we prove that these set of auxiliary variables achieve the expected distortion less than D. By converse assumption, we have that there exists a decoding function $\phi(M_{1:t}, \boldsymbol{X}_{m+1}^{N})$, with nth element ϕ_n, obeying

$$
D \geq \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[d(X_{1:m+1}(n), \phi_n(M_{1:t}, \boldsymbol{X}_{m+1}^{N}))] \tag{10}
$$

Define the function $g : \mathcal{M}_{1:t} \times \mathcal{X}_{m+1}^{N} \to \mathcal{Z}^N$ with nth element g_n, to be the Bayes detector for Z_n from $M_{1:t}, \boldsymbol{X}_{m+1}^{N}$:

$$
g_n(M_{1:t}, \boldsymbol{X}_{m+1}^{N}) = \arg\min_{\hat{z}\in\mathcal{Z}}\mathbb{E}[d(X_{1:m}(n), \hat{z})|M_{1:t}, \boldsymbol{X}_{m+1}^{N}]. \tag{11}
$$

Defining g_n via (11) shows that

$$
\begin{aligned}
&\mathbb{E}[d(X_{1:m+1}(n), \phi_n(M_{1:t}, \boldsymbol{X}_{m+1}^{N}))] \geq\\
&\mathbb{E}[d(X_{1:m+1}(n), g_n(M_{1:t}, \boldsymbol{X}_{m+1}^{N}))]
\end{aligned} \tag{12}
$$

as $\phi_n(M_{1:t}, \boldsymbol{X}_{m+1}^{N})$ must select some value from $\mathcal{Z}$ for each $(M_{1:t}, \boldsymbol{X}_{m+1}^{N})$ which must necessarily have distortion lower bounded by the minimum selected in (11). Next, define $\tilde{g}_n$ to be the Bayes detector for Z_n from $M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{N}$, i.e. let $\tilde{g}_n(M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{N}) =$

$$
\arg\min_{\hat{z}\in\mathcal{Z}}\mathbb{E}[d(X_{1:m}(n), \hat{z})|M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{N}]. \tag{13}
$$

The optimality (13) then shows

$$
\begin{aligned}
&\mathbb{E}[d(X_{1:m+1}(n), g_n(M_{1:t}, \boldsymbol{X}_{m+1}^{N}))] \geq\\
&\mathbb{E}[d(X_{1:m+1}(n), \tilde{g}_n(M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{N}))]
\end{aligned} \tag{14}
$$

Next, observe that the Markov chain $X_{1:m+1}(n) \leftrightarrow M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{n+}, X_{m+1}(n) \leftrightarrow \boldsymbol{X}_{m+1}^{n-1}$ implies that the conditional expectation in (13) for any $\hat{z} \in \mathcal{Z}$ obeys

$$
\begin{aligned}
&\mathbb{E}[d(X_{1:m}(n), \hat{z})|M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{N}]\\
&\quad = \mathbb{E}[d(X_{1:m}(n), \hat{z})|M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, X_{m+1}(n), \boldsymbol{X}_{m+1}^{n+}]
\end{aligned} \tag{15}
$$

which in turn shows that the minimum $\tilde{g}_n(M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{N})$ for a given $M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, X_{m+1}(n), \boldsymbol{X}_{m+1}^{n+}$ can be constant in $\boldsymbol{X}_{m+1}^{n-1}$, so that there exists a function $\hat{g}_n(M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{n+}, X_{m+1}(n))$ such that

$$
\begin{aligned}
&\mathbb{E}[d(X_{1:m+1}(n), \tilde{g}_n(M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{N}))] =\\
&\mathbb{E}[d(X_{1:m}(n), \hat{g}_n(M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{n+}, X_{m+1}(n)))]
\end{aligned} \tag{16}
$$

Recognizing $(M_{1:t}, \boldsymbol{X}_{1:m}^{n-1}, \boldsymbol{X}_{m+1}^{n+}, X_{m+1}(n)) = (U_{1:t}(n), X_{m+1}(n))$, and putting together (10), (12), (14), and (16) we have

$$
D \geq \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[d(X_{1:m+1}(n), \hat{g}(U_{1:t}(n), X_{m+1}(n)))] \tag{17}
$$

Viewing (5) (25) and (17) as a convex combination with coefficients $\frac{1}{N}$ of N points in $\mathcal{RD}_{out}^{t}$ associated with random variables $U_{1:t}(m)$ obeying the conditions in its definition, we have proven that $(\boldsymbol{R}, D) \in \mathcal{RD}_{out}^{t}$. ☐

Proposition 3. *The cardinalities of U_i can be bounded by*

$$
|\mathcal{U}_i| \leq |\mathcal{X}_{j_i}|\prod_{r=1}^{i-1}|\mathcal{U}_r| + 1 + t - i \tag{18}
$$

Where we define $j_i := (i-1) \bmod m + 1$.

Proof. The proof may be found in [19]. □

Note that the inner bound is based on viewing the user with the noisiest (lowest quality source) as the side information. This guarantees that the user with a higher quality side information can also decode the message. The outer bound is based on the providing the side information to all the other sources. The interactive nature of this scheme requires that each message should be decoded not only by the CEO, but also by all the other sources. Therefore, one has to take care of an encoder which has access to the lower quality source information and views it as a side information. This is one of the complications that doesn't let the inner and outer bounds match in general. In the next section we study several cases where the outer bound can be proven to be tight.

IV. Some Special Cases

In this section, we study the conditions such that the two bounds coincide. Moreover, the rate distortion region is characterized for two examples in which the nodes observe particular subsets of random variables.

Lemma 2. *If the source and the side information variables are mutually independent* $X_i \perp\!\!\!\perp X_j, \forall i \neq j \in [m+1]$*, the* $\mathcal{RD}^t_{inn} = \mathcal{RD}^t_{out} = \mathcal{RD}^t$.

$$\left\{ (\mathbf{R}, D) \;\middle|\; \begin{array}{c} \forall i \in [t] \quad R_i \geq I(X_j; U_i | U_{1:i-1}), \\ U_i \leftrightarrow X_j U_{1:i-1} \leftrightarrow X_{[m+1]\setminus\{j\}}, \\ j = ((i-1) \bmod m) + 1, \\ \mathbb{E}\big[d(X_1, ., X_{m+1}, \hat{Z}(U_{1:t}, X_{m+1}))\big] \leq D \end{array} \right\}$$

Proof. For simplicity, let's assume that there are only two nodes X_1, X_2 such that conditioning on the side information X_3 makes them independent. To achieve the first rate bound in Theorem 2, the following should hold. If we have

$$I(U_1; X_3) \leq I(U_1; X_2) \tag{19}$$

then, X_2 is less noisy than X_3 with respect to X_1, with U_1 obeying the Markov chain $U_1 \leftrightarrow X_1 \leftrightarrow X_2 X_3$. This condition makes X_2 a better side information to X_1 than X_3. Equation (19) combined with the Markov condition can be viewed as the following condition.

$$X_1 \leftrightarrow X_2 \leftrightarrow X_3 \tag{20}$$

Therefore, the first rate $R_1 \geq I(X_1; U_1 | X_3)$ can be achieved, since equations (19), and (20) ensure that the user that access to the lower quality side information can also decode the message, which guarantees that the user with a higher quality side information can also decode. In other words, to achieve the rate of the first round in Theorem 2, the source and the side information variables should be a set of particularly ordered degraded sources. The degradedness here means that the encoder which has access to the lower quality source information can be viewed as a side information.

Similarly, for the rate of the next user, if we have

$$I(U_2; X_3, U_1) \leq I(U_2; X_1 U_1) \tag{21}$$

then the rate bound $R_2 \geq I(X_2; U_2 | X_3, U_1)$ guarantees that both the CEO with the side information X_3, as well as the first node with observation X_1 can decode the message U_2. Proving equation (21) is equivalent to proving

$$U_2 \leftrightarrow X_1 \leftrightarrow U_1, X_3 \tag{22}$$

which holds only if $X_2 \leftrightarrow X_1 \leftrightarrow X_3$.

$$\begin{aligned} P(U_2 | X_1, X_3, U_1) &= \sum_{x_2} P(U_2, X_2 | X_1, X_3, U_1) \\ &= \sum_{x_2} P(X_2 | X_1, X_3, U_1) P(U_2 | X_{1:3}, U_1) \\ &\overset{a}{=} \sum_{x_2} P(X_2 | X_1) P(U_2 | X_1, X_2) \\ &= \sum_{x_2} P(X_2, U_2 | X_1) \\ &= P(U_2 | X_1) \end{aligned}$$

where in a we have $U_1 \leftrightarrow X_1 \leftrightarrow X_2$, and $X_2 \leftrightarrow X_1 \leftrightarrow X_3$ By data processing inequality in (22), and positivity of the conditional mutual information we have $I(U_1, X_3; U_2) \leq I(U_2; X_1) \leq I(U_2; X_1 U_1)$ which yields (21). Therefore, by assumption we have $X_1 \leftrightarrow X_3 \leftrightarrow X_2$, and to have a matching outer bound $X_1 \leftrightarrow X_2 \leftrightarrow X_3$, and $X_2 \leftrightarrow X_1 \leftrightarrow X_3$ should hold. These three Markov conditions results in mutual independence of the sources and the side information. □

Next we consider two examples where the nodes observe particular subsets of random variables, and we show the inner and outer bounds derived in sections III-C, and III-B match.

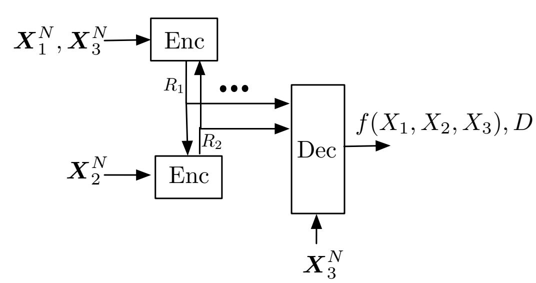

Fig. 3: Collocated network with degraded sources

1) Example one: Function computation where nodes observe particular subsets of independent variables: Suppose $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_2 \subseteq \{1, 2, 3\}$ such that user 1 observes $Z_{\mathcal{A}_1} = \{X_1, X_3\}$, and user 2 observes $Z_{\mathcal{A}_2} = \{X_2\}$, and $Z_{\mathcal{A}_3} = \{X_3\}$ is available as the side information to the CEO to compute $f(X_1, X_2, X_3)$ in an interactive manner over t rounds of interaction as depicted in Figure 3. Assume $\forall i, j \in \{1, 2, 3\}$, X_i is marginally independent of X_j, then the complete characterization of the rate distortion region $\mathcal{RD}^t$ consists of a set of $(\mathbf{R}, D)$ tuples such that for all

$i \in \{1, ..., t\}$,

$$\left\{ (\mathbf{R}, D) \;\middle|\; \begin{array}{c} R_i \geq I(X_1X_3; U_i|U_{1:i-1}), i \text{ odd} \\ R_i \geq I(X_2; U_i|U_{1:i-1}), i \text{ even} \\ U_i \leftrightarrow Z_{\mathcal{A}_j} U_{1:i-1} \leftrightarrow Z_{\mathcal{A}_j^c} \\ j = ((i-1) \bmod 2) + 1, \quad \forall i \in [t] \\ \mathbb{E}\big[d(X_{1:3}, \hat{Z}(U_{1:t}, X_3))\big] \leq D \end{array} \right\}$$

Proof. For simplicity let $t = 3$.

$$\begin{aligned}
NR_1 &\geq H(M_1) \\
&\geq H(M_1|\boldsymbol{Z}^N_{\mathcal{A}_2}) \\
&\geq I(\boldsymbol{Z}^N_{\mathcal{A}_1}, \boldsymbol{Z}^N_{\mathcal{A}_3}; M_1|\boldsymbol{Z}^N_{\mathcal{A}_2}) \\
&= H(\boldsymbol{Z}^N_{\mathcal{A}_1}, \boldsymbol{Z}^N_{\mathcal{A}_3}) - H(\boldsymbol{Z}^N_{\mathcal{A}_1}, \boldsymbol{Z}^N_{\mathcal{A}_3}|M_1, \boldsymbol{Z}^N_{\mathcal{A}_2}) \\
&= \sum_{n=1}^{N} H(Z_{\mathcal{A}_1}(n), Z_{\mathcal{A}_3}(n)|Z_{\mathcal{A}_2}(n)) \\
&\qquad - H(Z_{\mathcal{A}_1}(n), Z_{\mathcal{A}_3}(n)|M_1, \boldsymbol{Z}^N_{\mathcal{A}_2}, \boldsymbol{Z}^{n-}_{\mathcal{A}_1}, \boldsymbol{Z}^{n+}_{\mathcal{A}_3}) \\
&\stackrel{a_1}{\geq} \sum_{n=1}^{N} H(Z_{\mathcal{A}_1}(n)|Z_{\mathcal{A}_2}(n)) \\
&\qquad - H(Z_{\mathcal{A}_1}(n)|M_1, \boldsymbol{Z}^N_{\mathcal{A}_2}, \boldsymbol{Z}^{n-}_{\mathcal{A}_1}, \boldsymbol{Z}^{n+}_{\mathcal{A}_3}) \\
&\geq \sum_{n=1}^{N} H(X_1(n), X_3(n)) \\
&\qquad - H(X_1(n), X_3(n)|M_1, \boldsymbol{X}^{n-}_2, \boldsymbol{X}^{n-}_1, \boldsymbol{X}^{\setminus n}_3) \\
&\stackrel{a_2}{=} \sum_{n=1}^{N} I(X_1(n), X_3(n); U_1(n)) \qquad (23)
\end{aligned}$$

Step a_1 follows from positivity of the conditional mutual information. In step a_2 we defined auxiliary random variables $U_1(n) := \{M_1, \boldsymbol{X}^{n-}_{1:3}, \boldsymbol{X}^{n+}_3\}$, for $n \in [N]$. The rate for the second user in round 2:

$$\begin{aligned}
NR_2 &\geq H(M_2) \\
&\geq I(\boldsymbol{X}^N_2; M_2|M_1, \boldsymbol{X}^N_1, \boldsymbol{X}^N_3) \\
&= \sum_{n=1}^{n} H(X_2(n)|M_1, \boldsymbol{X}^N_1, \boldsymbol{X}^N_3, \boldsymbol{X}^{n-}_2) \\
&\qquad - \sum_{n=1}^{n} H(X_2(n)|M_1, M_2, \boldsymbol{X}^N_1, \boldsymbol{X}^N_3, \boldsymbol{X}^{n-}_2) \\
&\stackrel{b}{=} \sum_{n=1}^{n} H(X_2(n)|M_1, \boldsymbol{X}^{n-}_1, \boldsymbol{X}^{\setminus n}_3, \boldsymbol{X}^{n-}_2) \\
&\qquad - \sum_{n=1}^{n} H(X_2(n)|M_1, M_2, \boldsymbol{X}^N_1, \boldsymbol{X}^N_3, \boldsymbol{X}^{n-}_2) \\
&\stackrel{c}{\geq} \sum_{n=1}^{n} H(X_2(n)|M_1, \boldsymbol{X}^{n-}_1, \boldsymbol{X}^{\setminus n}_3, \boldsymbol{X}^{n-}_2) \\
&\qquad - \sum_{n=1}^{n} H(X_2(n)|M_1, M_2, \boldsymbol{X}^{n-}_1, \boldsymbol{X}^{\setminus n}_3, \boldsymbol{X}^{n-}_2) \\
&\stackrel{d}{=} \sum_{n=1}^{n} I(X_2(n); U_2(n)|U_1(n)) \qquad (24)
\end{aligned}$$

In step b, we used the following Markov condition: $X_2(n) \leftrightarrow M_1\boldsymbol{X}^{n-}_{1,2}, \boldsymbol{X}^{\setminus n}_3 \leftrightarrow X_1(n)X_3(n)\boldsymbol{X}^{n+}_1$ which follows the similar proof as in Proposition 1. Step c follows from the fact that conditioning reduces the entropy. In step d, the auxiliary random variable is chosen to be $U_2(n) := M_2$.

The lower bound for the rate of the first node in round 3 can be derived as follows:

$$\begin{aligned}
NR_3 &\geq H(M_3) \\
&\geq I(\boldsymbol{Z}^N_{\mathcal{A}_1}, \boldsymbol{Z}^N_{\mathcal{A}_3}; M_3|M_{1:2}, \boldsymbol{Z}^N_{\mathcal{A}_2}) \\
&= \sum_{n=1}^{N} I(Z_{\mathcal{A}_1}(n), Z_{\mathcal{A}_3}(n); M_3|M_{1:2}, \boldsymbol{Z}^N_{\mathcal{A}_2}, \boldsymbol{Z}^{n-}_{\mathcal{A}_1}, \boldsymbol{Z}^{n+}_{\mathcal{A}_3}) \\
&\geq \sum_{n=1}^{N} I(Z_{\mathcal{A}_1}(n); M_3|M_{1:2}, \boldsymbol{Z}^N_{\mathcal{A}_2}, \boldsymbol{Z}^{n-}_{\mathcal{A}_1}, \boldsymbol{Z}^{n+}_{\mathcal{A}_3}) \\
&= \sum_{n=1}^{N} I\big(X_1(n), X_3(n); M_3|U_1(n), U_2(n), \boldsymbol{X}^{n+}_2, X_2(n)\big) \\
&\stackrel{e}{=} \sum_{n=1}^{N} I\big(X_1(n), X_3(n); M_3|U_1(n), U_2(n)\big) \\
&\stackrel{f}{=} \sum_{n=1}^{N} I\big(X_1(n), X_3(n); U_3|U_{1:2}(n)\big) \qquad (25)
\end{aligned}$$

In step e we have $X_1(n)X_3(n) \leftrightarrow M_{1:2}\boldsymbol{X}^{n-}_{1:3}, \boldsymbol{X}^{n+}_3 \leftrightarrow \boldsymbol{X}^{n+}_2 X_2(n)$ using Figure 4, and the second term follows from conditioning reduces the entropy. In step f, the auxiliary $U_3(n) := M_3$. By the factorized distribution shown in Figure 4, we can check that these choice of auxiliary random variables obey the Markov chains, and the expected distortion constraints. The proof of achievability is available

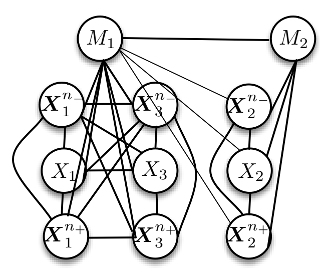

Fig. 4: Factorization of the joint distribution

in [19]. □

2) Example two: Nodes observe a subset of independent variables, and side information is available to the users: Suppose that $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_2 \subseteq \{1, 2, 3\}$, and user 1 observes $Z_{\mathcal{A}_1} = \{X_1, X_3\}$, and user 2 observes $Z_{\mathcal{A}_2} = \{X_2, X_3\}$, and $Z_{\mathcal{A}_3} = \{X_3\}$ is available as the side information to the CEO to compute $f(X_1, X_2, X_3)$ in an interactive manner described above. $X_i \perp\!\!\!\perp X_j, \forall i \neq j \in \{1, 2, 3\}$, then the complete characterization of the rate distortion region $\mathcal{RD}^t$ consists of a set of $(\mathbf{R}, D)$ tuples such that for all

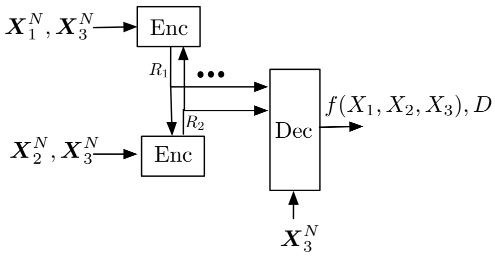

Fig. 5: Collocated network: nodes observing subsets of independent sources

$i \in \{1, ..., t\}$

$$\left\{ (\mathbf{R}, D) \;\middle|\; \begin{array}{c} R_i \geq I(X_1; U_i | X_3, U_{1:i-1}), i \text{ odd} \\ R_i \geq I(X_2; U_i | X_3, U_{1:i-1}), i \text{ even} \\ U_i \leftrightarrow Z_{\mathcal{A}_j} U_{1:i-1} \leftrightarrow Z_{\mathcal{A}_j^c} \\ j = ((i-1) \bmod 2) + 1, \quad \forall i \in [t] \\ \mathbb{E}\big[d(X_{1:3}, \hat{Z}(U_{1:t}, X_3))\big] \leq D \end{array} \right\}$$

Proof. For simplicity let $t = 3$. Achievability uses the random binning argument and follows from the proof in *Example one*, and it's omitted here. We provide the converse analysis.

$$\begin{aligned} NR_1 &\geq H(M_1) \\ &\geq I(\boldsymbol{X}_1^N; M_1 | \boldsymbol{X}_2^N \boldsymbol{X}_3^N) \\ &= \sum_{n=1}^{N} H(X_1(n) | X_2(n), X_3(n)) \\ &\quad - H(X_1(n) | M_1, \boldsymbol{X}_2^N, \boldsymbol{X}_3^N, \boldsymbol{X}_1^{n-}) \\ &\geq \sum_{n=1}^{N} H(X_1(n) | X_3(n)) \\ &\quad - H(X_1(n) | M_1, \boldsymbol{X}_{1:3}^{n-}, \boldsymbol{X}_3^{n+}, X_3(n)) \\ &\stackrel{a}{=} \sum_{n=1}^{N} I(X_1(n); U_1(n) |, X_3(n)) \end{aligned}$$

In step a we defined auxiliary random variables $U_1(n) := \{M_1, \boldsymbol{X}_{1:3}^{n-}, \boldsymbol{X}_3^{n+}\}$, for $n \in [N]$.

$$\begin{aligned} NR_2 &\geq H(M_2) \\ &\geq I(\boldsymbol{X}_2^N; M_2 | M_1, \boldsymbol{X}_1^N, \boldsymbol{X}_3^N) \\ &= \sum_{n=1}^{n} H(X_2(n) | M_1, \boldsymbol{X}_1^N, \boldsymbol{X}_3^N, \boldsymbol{X}_2^{n-}) \\ &\quad - \sum_{n=1}^{n} H(X_2(n) | M_1, M_2, \boldsymbol{X}_1^N, \boldsymbol{X}_3^N, \boldsymbol{X}_2^{n-}) \\ &\stackrel{b}{=} \sum_{n=1}^{n} H(X_2(n) | M_1, \boldsymbol{X}_{1:3}^{n-}, \boldsymbol{X}_3^{n+}, X_3(n)) \\ &\quad - \sum_{n=1}^{n} H(X_2(n) | M_1, M_2, \boldsymbol{X}_{1:3}^{n-}, \boldsymbol{X}_3^{n+}, X_3(n)) \\ &\stackrel{c}{=} \sum_{n=1}^{n} I(X_2(n); U_2(n) | U_1(n) X_3(n)) \end{aligned}$$

In step b, we used the following Markov condition: $X_2(n) \leftrightarrow M_1 \boldsymbol{X}_{1:3}^{n-} \boldsymbol{X}_3^{n+} X_3(n) \leftrightarrow X_1(n) \boldsymbol{X}_1^{n+}$ which follows the similar proof as in Proposition 1, and the second part follows from the fact that conditioning reduces the entropy. In step c, the auxiliary random variable is chosen to be $U_2(n) := \{M_2\}$.

$$\begin{aligned} NR_3 &\geq H(M_3) \\ &\geq I(\boldsymbol{X}_1^N; M_3 | M_1, M_2, \boldsymbol{X}_2^N, \boldsymbol{X}_3^N) \\ &= \sum_{n=1}^{N} H(X_1(n) | M_{1:2}, \boldsymbol{X}_2^N, \boldsymbol{X}_3^N, \boldsymbol{X}_1^{n-}) \\ &\quad - H(X_1(n) | M_{1:3}, \boldsymbol{X}_2^N, \boldsymbol{X}_3^N, \boldsymbol{X}_1^{n-}) \\ &\stackrel{d}{\geq} \sum_{n=1}^{N} H(X_1(n) | M_{1:2}, \boldsymbol{X}_{1:3}^{n-}, \boldsymbol{X}_3^{n+}, X_3(n)) \\ &\quad - H(X_1(n) | M_{1:3}, \boldsymbol{X}_{1:3}^{n-}, \boldsymbol{X}_3^{n+}, X_3(n)) \\ &\stackrel{f}{=} \sum_{n=1}^{N} I(X_1(n); U_3(n) | U_{1:2}(n), X_3(n)) \end{aligned}$$

In step d, we used the following Markov condition: $X_1(n) \leftrightarrow M_{1:2} \boldsymbol{X}_{1:3}^{n-} \boldsymbol{X}_3^{n+} X_3(n) \leftrightarrow X_2(n) \boldsymbol{X}_2^{n+}$. In step f we defined auxiliary random variables $U_3(n) := \{M_3\}$, for $n \in [N]$. Using the graphical technique, one can verify that the Markov conditions $M_1 \boldsymbol{X}_{1:3}^{n-} \boldsymbol{X}_3^{n+} \leftrightarrow X_1(n) X_3(n) \leftrightarrow X_2(n)$, and $M_2 \leftrightarrow M_1 \boldsymbol{X}_{1:3}^{n-} \boldsymbol{X}_3^{n+} X_2(n) X_3(n) \leftrightarrow X_1(n)$ hold. $\square$

V. Extremum function computation

Assume that CEO computes either the maximum value across users (the $\max$), or a user attaining this maximum (the $\arg\max$). Let the function f to be $f = \max_i X_i$, and $f = \arg\max_i X_i$. The CEO needs to compute these functions up to a distortion D. We define the distortion measure for computing the $\max$ function as follows,

$$d_M\left(x_1(i), ..., x_{m+1}(i), \hat{z}_M(i)\right) = \begin{cases} z_M(i) & \text{if } \hat{z}_M(i) > z_M(i) \\ z_M(i) - \hat{z}_M(i) & \text{o.w.} \end{cases}$$

Unlike the Hamming distortion, we choose the costs of underestimating and overestimating to be different. This distortion measure penalizes over-estimation more heavily than under-estimation. Next we define the distortion measure to compute the $\arg\max$ as follows:

$$d_A\left(x_1(i), ..., x_{m+1}(i), \hat{z}_A(i)\right) = \begin{cases} 0 & \text{if } \hat{z}_A(i) \in \mathcal{Z}_A(i) \\ x_{z_A(i)}(i) - x_{\hat{z}_A(i)}(i) & \text{o.w.} \end{cases}$$

The above distortion measures the loss between source value of the user with the actual max and the source value for the user estimated to have the max. $\mathcal{Z}_A$ is a set of admissible functions in zero distortion regime, and is to be defined in definition 2. Note that in computing the argmax a tie happens when two or more users attain the maximum value, and in this case, the CEO can choose any user that achieves the maximum. Bearing this in mind, we show that the rate savings is possible when CEO aims to compute argmax function interactively with predefined distortion measure.

Definition 2. A function $Z_A : \mathcal{X}^{m+1} \to [m]$, and $Z_M : \mathcal{X}^{m+1} \to [|\mathcal{X}|]$, is a candidate argmax and max function in a zero distortion regime respectively, if

$$E\big[d_A(x_1(i), ..., x_m(i), x_{m+1}(i), \hat{z}_A(i)\big] = 0 \tag{26}$$
$$E\big[d_M(x_1(i), ..., x_m(i), x_{m+1}(i), \hat{z}_M(i)\big] = 0 \tag{27}$$

Therefore, $\hat{z}_A \in \mathcal{Z}_A$, and $\hat{z}_M \in \mathcal{Z}_M$. In another words, $\mathcal{Z}_A(\mathcal{Z}_M)$ is a set of admissible $\arg\max(\max)$ function which can be computed with zero expected distortion.

If the per sample distortion is chosen to be hamming distortion, in the case of two-terminal function computation, with correlated sources, the characterization sum rate with zero distortion is the same as sum rate in the lossless regime [16]. In this section, we defined another type of distortion measure for the purpose of computing the extrema function. We illustrate that in a network that contains multiple mutually independent sources, with each source broadcasting its message, the sum rate in zero distortion, and lossless regime have a similar characterization.

Lemma 3. *The optimal rate to compute $f(X_1, ..., X_m) = \arg\max_i X_i$ with zero distortion is equal to the optimal rate for losslessly computing the function.*

Proof. Let $\mathcal{R}_t$ be the rate region for t-round lossless $\arg\max$ computation. First we show that $\mathcal{R}_t \subseteq \{\mathbf{R}|(\mathbf{R}, 0) \in \mathcal{RD}_t\}$. Let $\mathbf{R} \in \mathcal{R}_t$, so $\mathbf{R}$ is losslessly achievable, meaning that $p(\boldsymbol{X}_{\boldsymbol{Z}_A} \neq \boldsymbol{X}_{\hat{\boldsymbol{Z}}_A}) < \epsilon$. The expected distortion is then

$$\begin{aligned}
&\mathbb{E}\Big[d^{(N)}(\boldsymbol{X}_1^N, .., \boldsymbol{X}_{m+1}^N, \hat{\boldsymbol{Z}}^N)\Big] \\
&= \mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^{N} d\left(x_1(i), ..., x_{m+1}(i), \hat{z}(i)\right)\Big] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(x_{z_A(i)}(i) - x_{\hat{z}_A(i)}(i)\right) p\left(x_{z_A(i)}(i) \neq x_{\hat{z}_A(i)}(i)\right) \\
&\leq p\left(\boldsymbol{X}_{\boldsymbol{Z}_A} \neq \boldsymbol{X}_{\hat{\boldsymbol{Z}}_A}\right) < \epsilon
\end{aligned}$$

We prove that $\{\mathbf{R}|(\mathbf{R}, 0) \in \mathcal{RD}_t\} \subseteq \mathcal{R}_t$. Let $\mathbf{R}$ be a set of rates that can be achieved with zero distortion. $\exists$ a function g s.t. $d_A\left(X_1, ..., X_{m+1}, g_A(U^t, X_{m+1})\right) = 0$. Hence, $\left(X_{Z_A} - X_{g_A(U^t, X_{m+1})}\right) \mathbf{1}\left(g_A(U_{1:t}, X_{m+1}) \notin \mathcal{Z}_A\right) = 0$. Therefore, $g_A(U^t, X_{m+1}) \in \mathcal{Z}_A$. This satisfies the conditional entropy condition in $H(\arg\max(X_1, ..., X_m)|U_{1:t}) = H(g_A(U^t, X_{m+1})|U^t) = 0$. Hence, $\mathbf{R}$ is achievable with vanishing block error probability. $\square$

Lemma 4. *The optimal rate to compute $f(X_1, ..., X_m) = \max_i X_i$ with zero distortion is equal to the optimal rate for losslessly computing the function.*

Proof. The proof of this may be found in [19]. $\square$

VI. Conclusion

We established inner and outer bound for lossy interactive function computation in a collocated network. Users observe conditionally independent random variables and broadcast their messages to the CEO one at a time. The CEO has a dependent side information and aims to compute a function of the sources and side information up to a distortion measure. The inner and outer bound are shown to coincide in some special cases. When the function of interest is the extremum function, the distortion measure is chosen to be different than the hamming distortion. We showed that in a special case, the characterization of the zero distortion rate region is equivalent to the lossless characterization for computing either the max or the argmax function.

References

[1] M. Sefidgaran and A. Tchamkerten, "Distributed function computation over a rooted directed tree," *arXiv preprint arXiv:1312.3631*, 2013.

[2] V. Doshi, D. Shah, M. Medard, and S. Jaggi, "Functional compression through graph coloring," *IEEE Trans. Inform. Theory*, vol. 56, no. 8, pp. 3901–3917, August 2010.

[3] J. Korner and K. Marton, "How to encode the modulo-two sum of binary sources (corresp.)," *IEEE Trans. Inform. Theory*, vol. 25, no. 2, pp. 219–221, 1979.

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

[5] S. I. Gel'fand and M. S. Pinsker, "Coding of sources on the basis of observations with incomplete information," *Prob. Peredachi Informatsii*, vol. 15, no. 2, pp. 45–57, 1979.

[6] T. S. Han and K. Kobayashi, "A dichotomy of functions f(x,y) of correlated sources (x,y) from the viewpoint of the achievable rate region," *IEEE Trans. Inform. Theory*, vol. 33, no. 1, pp. 69–76, 1987.

[7] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem [multiterminal source coding]," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[8] Y. Oohama, "Rate-distortion theory for gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, 2005.

[9] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic gaussian ceo problem," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2004.

[10] G. Ku, J. Ren, and J. M. Walsh, "Computing the rate distortion region for the ceo problem with independent sources," *IEEE Trans. Sig. Proc.*, vol. 63, no. 3, pp. 567–575, Feb. 2015.

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

[12] V. Prabhakaran, K. Ramchandran, and D. Tse, "On the role of interaction between sensors in the ceo problem," in *Proc. Allerton Conference*. Citeseer, 2004.

[13] A. Giridhar and P. R. Kumar, "Computing and communicating functions over sensor networks," *IEEE J. Sel. Areas Comm.*, vol. 23, no. 4, pp. 755–764, 2005.

[14] A. H. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. 31, no. 6, p. 735, November 1985.

[15] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 903–917, March 2001.

[16] N. Ma and P. Ishwar, "Some results on distributed source coding for interactive function computation," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 6180–6195, September 2011.

[17] N. Ma, P. Ishwar, and P. Gupta, "Interactive source coding for function computation in collocated networks," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4289–4305, 2012.

[18] H. H. Permuter, Y. Steinberg, and T. Weissman, "Two-way source coding with a helper," *Information Theory, IEEE Transactions on*, vol. 56, no. 6, pp. 2905–2919, 2010.

[19] S. Torabi and J. M. Walsh, "Distributed lossy interactive function computation." [Online]. Available: http://www.ece.drexel.edu/walsh/torInteractfunctcomp.pdf