The Three-Terminal Interactive Lossy Source Coding Problem
Leonardo Rey Vega, Member, IEEE, Pablo Piantanida, Senior Member, IEEE, and Alfred O. Hero, III, Fellow, IEEE

Abstract—In this paper, we explore the three-node multi-terminal lossy source coding problem, which seems to offer a formidable mathematical complexity. We derive an inner bound to the general rate-distortion region of this problem, which is a natural extension of the seminal work by Kaspi on the interactive two-terminal source coding problem. It is shown that this (rather involved) inner bound contains several rate-distortion regions of some relevant source coding settings. In this way, besides the non-trivial extension of the interactive two terminal problem, our results can be seen as a generalization and hence unification of several previous works in the field. By specializing the inner bound to particular cases, we obtain some novel rate-distortion regions for several multi-terminal lossy source coding problems.

Index Terms—Multi-terminal source coding, Wyner-Ziv, rate-distortion region, Berger-Tung inner bound, interactive lossy source coding, distributed lossy source coding.

I. INTRODUCTION

A. Motivation and Related Works

Distributed source coding is an important branch of study in information theory with enormous relevance for the present and future technology. Efficient distributed data compression may be the only way to guarantee acceptable levels of performance when energy and bandwidth are severely limited as in many real world sensor networks. The distributed data collected by different nodes in a network can be highly correlated and this correlation can be exploited at the application layer, e.g., for target localization and tracking or anomaly detection. In such cases cooperative joint data-compression can achieve a better overall rate-distortion trade-off than independent compression at each node.

B. Main Contributions

Complete answers to the optimal trade-offs between rate and distortion for distributed source coding are scarce and the solution to many problems remains elusive. Two of the most important results in information theory, Slepian-Wolf solution to the distributed lossless source coding problem [2] and Wyner–Ziv [3] single letter solution for the rate-distortion region when side information is available at the decoder provided the kick-off for the study of these important problems. Berger [4] and Tung [5] generalized the Slepian-Wolf problem when lossy reconstructions are required at the decoder. It was shown that the obtained region, although not tight in general, is the optimal one in several special cases [6]–[9] and strictly suboptimal in others [10]. Heegard and Berger [11] considered the Wyner-Ziv problem when the side information at the decoder may be absent or when there are two decoders with degraded side information. Timo et al. [12] correctly extended the achievable region for many (> 2) decoders. In [13] and the references therein, the complementary delivery problem (closely related to the Heegard-Berger problem) is also studied. The use of interaction in a multi-terminal source coding setting has not been so extensively studied as the problems mentioned above. Through the use of multiple rounds of interactive exchanges of information explicit cooperation can take place using distributed/.successive refinement source coding. Transmitting “reduced pieces” of information, and constructing an explicit sequential cooperative exchange of information can be more efficient that transmitting the “total information” in one-shot.

The value of interaction for source coding problems was first recognized by Kaspi in his seminal work [1], where the interactive two-terminal lossy source coding problem was introduced and solved under the assumption of a finite number of communication rounds. In [14] it is shown that interaction strictly outperforms (in term of sum rate) the Wyner-Ziv rate function. There are also several extensions to the original Kaspi problem. In [15] the interactive source coding problem with a helper is solved when the sources satisfy a certain Markov chain property. In [16]–[18] other interesting cases where interactive cooperation can be beneficial are studied. To the best of our knowledge, a proper generalization of this setting to interactive multi-terminal (> 2) lossy source coding has not yet been reported.

In this paper, we consider the three-terminal interactive lossy source coding problem presented in Fig. 1. We have
It needs to recover a set of common and private messages. On the other hand, when each node is acting as a decoder and private messages destined to some restricted sets of nodes. Messages at each node: common messages destined to all nodes problem leads us to consider the generation of two sets of messages at each node: common messages destined to all nodes and private messages destined to some restricted sets of nodes. On the other hand, when each node is acting as a decoder it needs to recover a set of common and private messages generated at different nodes (i.e. during round \( t \) node 3, needs to recover the common descriptions generated at nodes 1 and 2 and the private ones generated also at nodes 1 and 2). This is reminiscent of the Berger-Tung problem, which is also an open problem. Again, the situation is more involved because of the cooperation induced by the multiple rounds of exchanged information. Particularly important is the fact that, in the case of the common descriptions, there is a potential cooperation naturally induced by the encoding-decoding ordering imposed by the network in addition to the cooperation based on the conditioning on the previous exchanged descriptions. This potential cooperation for the exchange of common messages is exploited through the use of a binning technique to be explained in Appendix B.

Despite the complexity of the problem, we give an inner bound to the rate-distortion region that allows us to recover the 2 node Kaspi’s region. We also recover several previous inner bounds and rate-distortion regions of some well-known cooperative and interactive –as well as non-interactive– lossy source coding problems.

### C. General Achievable Region

We derive a general achievable region by assuming a finite number of rounds. This region is not a trivial extension of Kaspi’s region [1] and the main ideas behind its derivation are the exchange of common and private descriptions between the nodes in the network in order to exploit the side information at the different nodes. As in the original Kaspi’s formulation, the key to obtaining the achievable region is the natural cooperation between the nodes induced by the generation of new descriptions based on the past exchanged descriptions. However, in comparison to Kaspi’s 2 node case, the 3 nodes interactions make significant differences in the optimal action of each node at the encoding and decoding procedure in a given round. At each encoding stage, each node needs to communicate to other two nodes with different side information. This is reminiscent of the Heegard-Berger problem mentioned above and whose complete solution is not known when the side information at the decoders is not degraded. Moreover, the situation is a bit more complex because of the presence of 3-way interaction. This similarity with the Heegard-Berger problem leads us to consider the generation of two sets of messages at each node: common messages destined to all nodes and private messages destined to some restricted sets of nodes. On the other hand, when each node is acting as a decoder it needs to recover a set of common and private messages.

### D. Special Cases

As the full problem seems to offer a formidable mathematical complexity, including several special cases which are known to be long-standing open problems, we cannot give a full converse proving the optimality of the general achievable region obtained. However, in Section V we provide a complete answer to the rate-distortion regions of several specific cooperative and interactive source coding problems:

1. Two encoders and one decoder subject to lossy/lossless reconstruction constraints without side information (see Fig. 2).
2. Two encoders and three decoders subject to lossless/lossy reconstruction constraints with side information (see Fig. 3).
3. Two encoders and three decoders subject to lossless/lossy reconstruction constraints, reversal delivery and side information (see Fig. 4).
4. Two encoders and three decoders subject to lossy reconstruction constraints with degraded side information (see Fig. 5).
5. Three encoders and three decoders subject to lossless/lossy reconstruction constraints with degraded side information (see Fig. 6).

*Fig. 1.* Three-Terminal Interactive Source Coding. There is a single noiseless rate-limited broadcast channel from each terminal to the other two terminals. \( D_{ij} \) denotes the average per-letter distortion between the source \( X^n_{ij} \) and \( X^n_{ij} \) measured at node \( i \) for each pair \( i \neq j \).

*Fig. 2.* Two encoders and one decoder subject to lossy/lossless reconstruction constraints without side information.
Interestingly enough, we show that for the two last problems, interaction through multiple rounds could be helpful. Whereas for the other three cases, it is shown that a single round of cooperatively exchanged descriptions suffices to achieve optimality. Table I summarizes the characteristics of each of the above mentioned cases.

Next we summarize the contents of the paper. In Section II we formulate the general problem. In Section III we present and discuss the inner bound of the general problem. In Section IV we show how our inner bound contains several results previously obtained in the past. In Section V we present the converse results and their tightness with respect to the inner bound for the special cases mentioned above, providing the optimal characterization for them. In Section VI we present a discussion of the obtained results, their limitations and some numerical results concerning the new optimal cases from the previous section. Finally in Section VII we provide some conclusions. The major mathematical details are relegated to the Appendixes.

**Notation:** With $x^n$ and upper-case letters $X^n$ we denote vectors and random vectors of $n$ components, respectively. The $i$-th component of vector $x^n$ is denoted interchangeably as $x_i$ or $x[i]$ and with $x_{[i:t]}$ we denote the components with indices ranging from $s$ to $t$ with $s \leq t$. We use $x_{[t]}$ to denote the vector $(x_{[1:t]}, \ldots, x_{[t+1:n]})$. Let $X$, $Y$ and $V$ be three random variables on some alphabets with probability distribution $p_{X,Y,V}$. We will denote as $(X,Y,V)$ the fact that $(X,Y,V)$ are distributed with $p_{X,Y,V}$. When clear from the context we will simply denote $p_X(x)$ with $p(x)$. If the probability distribution of random variables $X$, $Y$, $V$ satisfies $p(x|y,v) = p(x|y)$ for each $x$, $y$, $v$, then they form a Markov chain, which is denoted by $X \rightarrow Y \rightarrow V$. The probability of an event $\mathcal{A}$ is denoted by $\Pr[\mathcal{A}]$, where the measure used to compute it will be understood from the context. Conditional probability of a set $\mathcal{A}$ with respect to a set $\mathcal{B}$ is denoted as $\Pr[\mathcal{A}|\mathcal{B}]$. Entropy is denoted by $H(\cdot)$ and mutual information by $I(\cdot; \cdot)$. $H_2(p)$ denotes the entropy associated with a Bernoulli random variable with parameter $p$. With $h(\cdot)$ we denote differential entropy. Following [19] the set of strongly typical sequences associated with random variable $X$ is denoted by $\mathcal{T}_n^{X|\varepsilon}$, where $\varepsilon > 0$. Similarly, given $x^n \in X^n$ we also denote the set of conditional typical sequences by $\mathcal{T}_n^{y|x}(x^n)$. We simply denote these sets as $\mathcal{T}_n^{x}$ when the involved random variables are clear from the context. The cardinal of set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. The complement of a set is denoted by $\overline{\mathcal{A}}$. With $\mathbb{Z}_{\geq \alpha}$ and $\mathbb{R}_{\geq \beta}$ we denote the integers and reals numbers greater than $\alpha$ and $\beta$ respectively.

**II. PROBLEM FORMULATION**

Assume three discrete memoryless sources (DMS’s) on finite alphabets and probability mass function (pmf) given by $(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3, p_{X_1,X_2,X_3})$ and arbitrary bounded distortion...
measures: \( d_j : X_j \times \hat{X}_j \rightarrow \mathbb{R}_{\geq 0}, \ j \in \mathcal{M} \triangleq \{1, 2, 3\} \) where \( \{\hat{X}_j\}_{j \in \mathcal{M}} \) are finite reconstruction alphabets.\(^1\) We consider the problem of characterizing the rate-distortion region of the interactive source coding scenario described in Fig. 1. In this setting, through \( K \) rounds of information exchange between the nodes each one of them will attempt to recover a lossy description of the sources that the others nodes observe, e.g., node 1 must reconstruct—while satisfying distortion constraints—the realization of the sources \( (X^n_2, X^n_3) \) observed by nodes 2 and 3. Indeed, this setup can be seen as a generalization of the well-known Kaspi’s problem [1].

**Definition 1 (K-Step Interactive Source Code):** A K-step interactive \( n \)-length source code for the network model in Fig. 1, is defined by a sequence of encoder mappings:

\[
\begin{align*}
f_1 : X^n_1 & \times \left( J^n_2 \times J^n_3 \right) \rightarrow J^n_1, \\
f_2 : X^n_2 & \times \left( J^n_1 \times J^n_3 \right) \rightarrow J^n_2, \\
f_3 : X^n_3 & \times \left( J^n_1 \times J^n_2 \right) \rightarrow J^n_3,
\end{align*}
\]

with \( l \in [1:K] \) and message sets: \( J_i^n \triangleq \{1, 2, \ldots, T_i^n\}, \ T_i^n \in \mathbb{Z}_{\geq 0}, \ i \in \mathcal{M}, \) and reconstruction mappings:

\[
g_{ij} : X^n_i \times \bigotimes_{m \in \mathcal{M}, m \neq i} ( J_m^n \times \ldots \times J_m^K ) \rightarrow \hat{X}_j^n, \ i \neq j.
\]

The average per-letter distortion and the corresponding distortion levels achieved at the node \( i \) with respect to source \( j \) satisfy:

\[
E \left[ d_j \left( X^n_j, \hat{X}_j^n_i \right) \right] \leq D_{ij}, \quad j \in \mathcal{M}, \ i \neq j \quad (2)
\]

with \( d \left( x^n, y^n \right) \triangleq \frac{1}{n} \sum_{m=1}^n d(x_m, y_m) \). In compact form we denote a \( K \)-step interactive source coding by \((n, K, \mathcal{F}, \mathcal{G})\) where \( \mathcal{F} \) and \( \mathcal{G} \) denote the sets of encoders and decoders mappings.

**Definition 2 (Achievability and Rate-Distortion Region):** Consider tuples \( \mathbf{R} \triangleq (R_1, R_2, R_3) \) and \( \mathbf{D} \triangleq (D_{12}, D_{13}, D_{21}, D_{23}, D_{31}, D_{32}) \). The rate vector \( \mathbf{R} \) is \((\mathbf{D}, K)\)-achievable if \( \forall \epsilon > 0 \) there is \( n_0(\epsilon, K) \in \mathbb{N} \) such that \( \forall n > n_0(\epsilon, K) \) there exists a \( K \)-step interactive source code \((n, K, \mathcal{F}, \mathcal{G})\) with rates satisfying:

\[
\frac{1}{n} \sum_{i=1}^K \log \| J_i^n \| \leq R_i + \epsilon, \quad i \in \mathcal{M} \quad (3)
\]

and with average per-letter distortions at node \( i \) with respect to source \( j \) given by:

\[
E \left[ d_j \left( X^n_j, \hat{X}_j^n_i \right) \right] \leq D_{ij} + \epsilon, \quad i, j \in \mathcal{M}, \ i \neq j, \quad (4)
\]

where \( \hat{X}_j^n_i \triangleq g_{ij} \left( J^n_i, \bigotimes_{m \in \mathcal{M}, m \neq i} ( J_m^n \times \ldots \times J_m^K ) \right), \ i \neq j \in \mathcal{M}. \) The rate-distortion region \( \mathcal{R}_3(\mathbf{D}, K) \) is defined by:

\[
\mathcal{R}_3(\mathbf{D}, K) = \left\{ \mathbf{R} : \text{Ris} (\mathbf{D}, K)\text{-achievable} \right\}. \quad (5)
\]

---

\(^1\)The problem can be easily generalized to the case in which there are different reconstruction alphabets at the terminals. It can also be shown that all the results should be valid if we employ arbitrary bounded joint distortion functions, e.g. at node 1 we use \( d(X_2, X_3; \hat{X}_2, \hat{X}_3) \).
Similarly, the $\mathbf{D}$-achievable region $\mathcal{R}_3(\mathbf{D})$ is given by $\mathcal{R}_3(\mathbf{D}) = \bigcup_{K=1}^{\infty} \mathcal{R}_3(\mathbf{D}, K)$, that is:

$$\mathcal{R}_3(\mathbf{D}) = \left\{ \mathbf{R} : \mathbf{R} \text{ is } (\mathbf{D}, K)\text{-achievable for some } K \in \mathbb{Z}_{\geq 1} \right\}. $$

**Remark 1**: By definition $\mathcal{R}_3(\mathbf{D}, K)$ is closed and using a time-sharing argument it is easy to show that it is also convex $\forall K \in \mathbb{Z}_{\geq 1}$.

**Remark 2**: The code definition depends on the node ordering in the encoding procedure. Above we defined the encoding functions $\{f_{i,l}, f_{j,l}^2, f_{j,l}^3\}_{l=1}^{K}$ assuming that in each round node 1 acts first, followed by node 2, and finally by node 3, and the process beginning again at node 1. It is clear that $\mathcal{R}_3(\mathbf{D}, K)$ depends on the node ordering in the encoding procedure. In this paper we restrict the analysis to the canonical ordering $(1 \rightarrow 2 \rightarrow 3)$. However, there are $3! = 6$ different orderings that generally lead to different regions and the $(\mathbf{D}, K)$-achievable region defined above is more explicitly denoted as $\mathcal{R}_3(\mathbf{D}, K, \sigma_c)$, where $\sigma_c$ is the trivial permutation for $\mathcal{M}$. In this way the $(\mathbf{D}, K)$-achievable region is:

$$\mathcal{R}_3(\mathbf{D}, K) = \bigcup_{\sigma \in \Sigma(\mathcal{M})} \mathcal{R}_3(\mathbf{D}, K, \sigma) \quad (6)$$

where $\Sigma(\mathcal{M})$ contains the permutations of set $\mathcal{M}$. The theory presented in this paper for determining $\mathcal{R}_3(\mathbf{D}, K, \sigma_c)$ can be used for the other permutations $\sigma \neq \sigma_c$ to compute (6).\(^3\)

### III. Inner Bound on the Rate-Distortion Region

We first present a general achievable rate-region where each node at a given round $l$ will generate descriptions destined to the other nodes based on the realization of its own source, the past descriptions generated by this particular node and the descriptions generated at the other nodes and recovered by the given node up to the present round. In order to precisely describe the rate-distortion region, we need to introduce some definitions. For a set $\mathcal{A}$, let $C(\mathcal{A}) \triangleq 2^\mathcal{A} \setminus \{\emptyset, \mathcal{A}\}$ be the set of all subsets of $\mathcal{A}$ minus $\mathcal{A}$ and the empty set. Denote the auxiliary random variables:

$$U_{i \rightarrow S,l}, \quad S \in C(\mathcal{M}), \quad i \notin S, \quad l = 1, \ldots, K. \quad (7)$$

Auxiliary random variables $U_{i \rightarrow S,l}$ will be used to build the descriptions generated at node $i$ and at round $l$ and destined to a set of nodes $S \in C(\mathcal{M})$ with $i \notin S$. For example, $U_{1 \rightarrow 23,l}$ will be used for the description generated in node 1 at round $l$ and destined to nodes 2 and 3. Similarly, $U_{i \rightarrow 2,l}$ will be used to build the descriptions generated at node 1 at round $l$ and destined only to node 2. We define the following variables:

$$\mathcal{W}_{i[l]} \triangleq \text{Common information shared by the three nodes and available at node } i \text{ at round } l \text{ before encoding,}$$

$$\mathcal{V}_{[S,i,l]} \triangleq \text{Private information shared by nodes in } S \in C(\mathcal{M}) \text{ and available at node } i \in S, \text{ at round } l, \text{ before encoding.}$$

In precise terms, the quantities introduced above for our problem are defined by:

$$\mathcal{W}_{i[l]} = \{U_{1 \rightarrow 23,k}, U_{2 \rightarrow 13,k}, U_{3 \rightarrow 12,k}\}_{k=1}^{l-1},$$

$$\mathcal{W}_{[i,l]} = \mathcal{W}_{i[l]} \cup U_{1 \rightarrow 23,l}, \quad \mathcal{V}_{[3,l]} = \mathcal{W}_{[2,l]} \cup U_{2 \rightarrow 13,l},$$

$$\mathcal{V}_{[2,l]} = \mathcal{V}_{[2,l]} \cup U_{1 \rightarrow 23,l}, \quad \mathcal{V}_{[3,l]} = \mathcal{V}_{[2,l]} \cup U_{2 \rightarrow 13,l},$$

$$\mathcal{V}_{[23,l]} = \mathcal{V}_{[2,l]} \cup U_{2 \rightarrow 13,l}.$$  

Before presenting the general inner bound, we provide the basic idea of the random coding scheme that achieves the rate-region in Theorem 1 for the case of $K$ communication rounds. Assume that all codebooks are randomly generated and known to all the nodes before the information exchange begins and consider the encoding ordering given by $1 \rightarrow 2 \rightarrow 3$ so that we begin at round $l = 1$ in node 1. Also, and in order to maintain the explanation simple and to help the reader to grasp the essentials of the coding scheme employed, we will consider that all the terminals are able to recover the descriptions generated at other nodes (which will be the case under the conditions in our Theorem 1). From the observation of the source $X^n_1$, node 1 generates a set of descriptions for each of the other nodes connected to it. In particular it generates a common description to be recovered at nodes 2 and 3 in addition to two private descriptions for node 2 and 3, respectively, generated from a conditional codebook given the common description. Then, node 2 tries to recover the descriptions destined to it (the common description generated at 1 and its corresponding private description), using $X^n_2$ as side information, and generates its own descriptions, based on source $X^n_2$ and the recovered descriptions from node 1. Again, it generates a common description for nodes 1 and 3, and a private description for node 3 and another one for node 1. The same process goes on until node 3, which tries to recover jointly the common descriptions generated by node 1 and node 2, and then the private descriptions destined to him by node 1 and 2. Then generates its own descriptions (common and private ones) destined to nodes 1 and 2. Finally, node 1 tries to recover all the descriptions destined to it generated by nodes 2 and 3 in the same way as previously done by node 3. After this, round $l = 1$ is over, and round $l = 2$ begins with node 1 generating new descriptions using $X^n_1$, its encoding history (from the previous rounds) and the recovered descriptions from the other nodes. The process continues in a similar manner until we reach round $l = K$ where node 3 recovers the descriptions from the other nodes and generates its own ones. Node 1 recovers the last descriptions destined to it from nodes 2 and 3 but does not generate new ones. The same holds for node 2 who only recovers the last descriptions generated by node 3 and thus terminating the K-rounds information.

\(^2\)Notice that this limit exists because we have the union of a monotone increasing sequence of sets.

\(^3\)It should be mentioned that this is not the most general setting of the problem. The most general encoding procedure will follow from the definition of the transmission order by a sequence $t_1, t_2, t_3, \ldots, t_{|\mathcal{M}|} | \times K$ with $t_l \in \mathcal{M}$. This will cover even the situation in which the order can be changed in each round. To keep the mathematical presentation simpler we will not consider this more general setting.

\(^4\)Not to be confused with the Wyner’s definition of common information [20].
exchange procedure. Notice that at the end of round $K$ the decoding in node 1 and node 2 can be done simultaneously. This is due to the fact that node 1 is not generating new descriptions destined to node 2. However, in order to simplify the analysis and notation in the appendix we will consider that the last decoding of node 2 occurs in round $K + 1$. After all the exchanges are done, each node recovers an estimate of the other nodes source realizations by using all the available recovered descriptions from the $K$ previous rounds.

**Theorem 1 (Inner Bound):** Let $R_3(D, K)$ be the closure of set of all rate tuples satisfying:

$$
R_1 = \sum_{i=1}^{K} (R^{(i)}_{1-23} + R^{(i)}_{1-2} + R^{(i)}_{1-3}),
$$

$$
R_2 = \sum_{i=1}^{K} (R^{(i)}_{2-13} + R^{(i)}_{2-1} + R^{(i)}_{2-3}),
$$

$$
R_3 = \sum_{i=1}^{K} (R^{(i)}_{3-12} + R^{(i)}_{3-1} + R^{(i)}_{3-2}),
$$

$$
R_1 + R_2 = \sum_{i=1}^{K} (R^{(i)}_{1-23} + R^{(i)}_{1-2} + R^{(i)}_{1-3} + R^{(i)}_{2-3} + R^{(i)}_{3-2} + R^{(i)}_{3-1}),
$$

$$
R_1 + R_3 = \sum_{i=1}^{K} (R^{(i)}_{1-23} + R^{(i)}_{1-3} + R^{(i)}_{3-2} + R^{(i)}_{3-1}),
$$

$$
R_2 + R_3 = \sum_{i=1}^{K} (R^{(i)}_{2-13} + R^{(i)}_{2-1} + R^{(i)}_{2-3} + R^{(i)}_{2-1} + R^{(i)}_{3-1}),
$$

where for each $i \in [1 : K]$:

$$
R^{(i)}_{1-23} > I \left( X_1; U_{1-23}, U_{2-13}, X_2 W_{1-2} W_{1-1} W_{13,1} W_{23,1}, X_3 W_{13,1}, V_{13,1} V_{23,1} \right)
$$

$$
R^{(i)}_{1-2} > I \left( X_1 X_2; U_{1-23}, U_{2-13}, X_2 W_{1-2} W_{1-1} W_{13,1} W_{23,1}, X_3 W_{13,1}, V_{13,1} V_{23,1} \right)
$$

$$
R^{(i)}_{1-3} > I \left( X_1 X_2; U_{1-23}, U_{2-13}, X_2 W_{1-2} W_{1-1} W_{13,1} W_{23,1}, X_3 W_{13,1}, V_{13,1} V_{23,1} \right)
$$

$$
R^{(i)}_{2-3} > I \left( X_1 X_2; U_{1-23}, U_{2-13}, X_2 W_{1-2} W_{1-1} W_{13,1} W_{23,1}, X_3 W_{13,1}, V_{13,1} V_{23,1} \right)
$$

$$
R^{(i)}_{1-3} + R^{(i)}_{2-3} > I \left( X_1 X_2; U_{1-23}, U_{2-13}, X_2 W_{1-2} W_{1-1} W_{13,1} W_{23,1}, X_3 W_{13,1}, V_{13,1} V_{23,1} \right)
$$

$$
R^{(i)}_{1-2} + R^{(i)}_{2-1} > I \left( X_1 X_2; U_{1-23}, U_{2-13}, X_2 W_{1-2} W_{1-1} W_{13,1} W_{23,1}, X_3 W_{13,1}, V_{13,1} V_{23,1} \right)
$$

$$
R^{(i)}_{1-2} + R^{(i)}_{2-3} > I \left( X_1 X_2; U_{1-23}, U_{2-13}, X_2 W_{1-2} W_{1-1} W_{13,1} W_{23,1}, X_3 W_{13,1}, V_{13,1} V_{23,1} \right)
$$

$$
R^{(i)}_{1-2} + R^{(i)}_{2-1} > I \left( X_1 X_2; U_{1-23}, U_{2-13}, X_2 W_{1-2} W_{1-1} W_{13,1} W_{23,1}, X_3 W_{13,1}, V_{13,1} V_{23,1} \right)
$$

$$
R^{(i)}_{1-2} + R^{(i)}_{2-1} > I \left( X_1 X_2; U_{1-23}, U_{2-13}, X_2 W_{1-2} W_{1-1} W_{13,1} W_{23,1}, X_3 W_{13,1}, V_{13,1} V_{23,1} \right)
$$

with $R^{(0)}_{1-2} = R^{(K+1)}_{1-2} = 0$ and $U_{i-S,0} = U_{i-S, K+1} = \emptyset$ for $S \in C(M)$ and $i \notin S$. With these definitions the rate-distortion region $R_3(D, K)$ satisfies:

$$
\bigcup_{p \in P(D, K)} \mathcal{R}_3(D, K) \subseteq \mathcal{R}_3(D, K),
$$

where $P(D, K)$ denotes the set of all joint probability measures of the auxiliary random variables that satisfy the following Markov chains for every $l \in [1 : K]$:

1. $U_{1-23,l} \rightarrow (X_1, W_{1-2}, W_{1-1}, W_{13,1}, W_{23,1}, X_2, X_3, V_{12,1}, V_{13,1}, V_{23,1})$
2. $U_{1-2,l} \rightarrow (X_1, W_{2-1}, W_{1-1}, W_{13,1}, W_{23,1}, X_2, X_3, V_{12,1}, V_{13,1}, V_{23,1})$
3. $U_{1-3,l} \rightarrow (X_1, W_{2-3}, W_{1-1}, W_{13,1}, W_{23,1}, X_2, X_3, V_{12,1}, V_{13,1}, V_{23,1})$
4. $U_{2-13,l} \rightarrow (X_2, W_{2-1}, W_{1-1}, W_{13,1}, W_{23,1}, X_2, X_3, V_{12,1}, V_{13,1}, V_{23,1})$
5. $U_{2-1,l} \rightarrow (X_2, W_{3-1}, W_{1-1}, W_{13,1}, W_{23,1}, X_2, X_3, V_{12,1}, V_{13,1}, V_{23,1})$
6. $U_{3-13,l} \rightarrow (X_1, W_{3-1}, W_{1-1}, W_{13,1}, W_{23,1}, X_2, X_3, V_{12,1}, V_{13,1}, V_{23,1})$
7. $U_{3-1,l} \rightarrow (X_2, W_{3-1}, W_{1-1}, W_{13,1}, W_{23,1}, X_2, X_3, V_{12,1}, V_{13,1}, V_{23,1})$
8. $U_{3-13,l} \rightarrow (X_3, W_{1-1}, W_{13,1}, W_{23,1}, X_1, X_2, V_{12,1}, V_{13,1}, V_{23,1})$
9. $U_{3-1,l} \rightarrow (X_3, W_{1-1}, W_{13,1}, W_{23,1}, X_1, X_2, V_{12,1}, V_{13,1}, V_{23,1})$

and such that there exist reconstruction mappings:

$$
\hat{g}_{ij} (X_i, Y_{ij, K+1}, W_{1, K+1}) = \hat{X}_{ij}
$$

with $E \left[ \sum_{j \in J} d_j (X_j, \hat{X}_{ij}) \right] \leq D_j$ for each $i, j \in M$ and $i \neq j$.

In order to save space, cardinalities of the auxiliary random variables are presented only for the special cases in Section V. The proof of this theorem is relegated to Appendix C and relies on the auxiliary results presented in Appendix A and the theorem on the cooperative Berger-Tung problem with side information presented in Appendix B.

**Remark 3:** It is worth mentioning here that our coding scheme is constrained to use successive decoding, i.e., by recovering first the coding layer of common descriptions and then the coding layer of private descriptions (for the recovering...
of each coding layer the nodes employ joint-decoding). Obviously, this is a sub-optimum procedure since the best scheme would be to use joint decoding where both common and private informations can be jointly recovered. However, the analysis of this scheme is much more involved. The associated achievable rate-distortion region involves a large number of equations that combine rates belonging to private and common messages from different nodes. Also, several mutual information terms in each of those rate equations cannot be combined, leading to a proliferation of many equations that offer little insight to the problem.

Remark 4: The idea behind our derivation of the achievable region can be extended to any number $M (> 3)$ of nodes in the network. This can be accomplished by generating a greater number of superimposed coding layers. The first layer of codes is composed by descriptions that should be decoded by all nodes. The next layer is composed by descriptions that should be recovered by subsets of $M - 1$ nodes. This continues until we reach the final layer, which is composed by codes with private descriptions for each of the nodes. Again, successive decoding is used at the nodes to recover the descriptions in each of these layers. Clearly, the number of required auxiliary random variables (and the number of equations in the rate-distortion region) will increase exponentially with the number of nodes.

Remark 5: It is interesting to compare the main ideas of our scheme with those of Kaspi [1]. The main idea in [1] is to have a single coding tree shared by the two nodes. Each leaf in the coding tree is a codeword generated either at node 1 or 2. At a given round each node knows (assuming no errors at the encoding and decoding procedures) the path followed in the tree. For example, at round $l$, node 1, using the knowledge of the path until round $l$ and its source realization generate a leaf (from a set of possible ones) using joint typicality encoding and binning. Node 2, using the same path known at node 1 and its source realization, uses joint typicality decoding to estimate the leaf generated at node 1. If there is no error at these encoding and decoding steps, the previous path is updated with the new leaf and both -node 1 and 2- know the updated path without error. Node 2 repeats the procedure. This is done until round $K$ where the final path is known at both nodes and used to reconstruct the desired sources.

In the case of three nodes the situation is more involved. At a given round, the encoder at an arbitrary node is seeing two decoders with different side information. In order to simplify the explanation consider that we are at round $l$ in the encoder 1, and that the listening nodes are 2 and 3. This situation forces node 1 to encode two sets of descriptions: one common for the other two nodes and a set of private ones associated with each of the listening nodes 2 and 3. Following the ideas of Kaspi, it is then natural to consider three different coding trees followed by node 1. One coding tree has leaves that are the common descriptions generated and shared by all the nodes in the network. The second tree is composed by leaves that are the private descriptions generated and shared with node 2. The third tree is composed by leaves that are the private descriptions generated and shared with node 3. As the private descriptions refine the common ones, depending on the quality of the side information of the node that is the intended recipient, it is clear that descriptions are correlated. For example, the private description destined to node 2, should depend not only on the past private descriptions generated and shared by nodes 1 and 2, but also on the common descriptions generated at all previous rounds in all the nodes and on the common description generated in the present round at node 1. Something similar happens for the private description destined to node 3. It is clear that as the common descriptions are to be recovered by all the nodes in the network, they can only be conditioned with respect to the past common descriptions generated at previous rounds and with respect to the common descriptions generated at the present round by a node who acted before (i.e. at round $l$ node 1 acts before than node 2). The private descriptions, as they are only required to be recovered at some set of nodes, can be generated conditional on the past exchanged common descriptions and the past private descriptions generated and recovered in the corresponding set of nodes (i.e., the private descriptions exchanged between nodes 1 and 2 at round $l$, can only be generated conditional on the past common descriptions generated at nodes 1, 2 and 3 and on the past private descriptions exchanged only between 1 and 2).

We can see clearly that there are basically four paths to be cooperatively followed in the network:

- One path of common descriptions shared by nodes 1, 2 and 3.
- One path of private descriptions shared by nodes 1 and 2.
- One path of private descriptions shared by nodes 1 and 3.
- One path of private descriptions shared by nodes 2 and 3.

It is also clear that each node only follows three of these paths simultaneously. The exchange of common descriptions deserves special mention. Consider round $l$ and node 3. This node needs to recover the common descriptions generated at nodes 1 and 2. But just before node 2 generates its own common description, it has recovered the common one generated at node 1. This recovered description could be used as some kind of coded side information by the encoder at node 2. This allows for a natural explicit cooperation between nodes 1 and 2 in order to help node 3 to recover both descriptions. Clearly, this is not the case for private descriptions from nodes 1 and 2 to be recovered at node 3. Node 2 does not recover the private description from node 1 to 3 and cannot generate an explicit collaboration to help node 3 to recover both private descriptions. Note, however, that as both private descriptions will be dependent on previous common descriptions an implicit collaboration (intrinsic to the code generation) is also in force. In appendix B we consider the problem (not in the interactive setting) of generating the explicit cooperation for the common descriptions through the use of what we call a super-binning procedure. The results of analysis will be used for our interactive three-node problem.
IV. KNOWN CASES AND RELATED WORK

Several inner bounds and rate-distortion regions on multi-terminal source coding problems can be derived by specializing the inner bound (8). Below we summarize only a few of them.

1) Distributed Source Coding With Side Information [5], [21]: Consider the distributed source coding problem where two nodes encode separately sources \(X^1_n\) and \(X^2_n\) with rates \((R_1, R_2)\) and a decoder by using side information \(X^K_n\) must reconstruct both sources with average distortion less than \(D_1\) and \(D_2\), respectively. By considering only one-round/one-way information exchange from nodes 1 and 2 (the encoders) to node 3 (the decoder), the results in [5] and [21] can be recovered as a special case of the inner bound (8). Specifically, we set:

\[
U_{1 \rightarrow 23, 3} = U_{2 \rightarrow 23, 3} = U_{3 \rightarrow 12, 3} = \emptyset, \\
U_{1 \rightarrow 3, 3} = U_{2 \rightarrow 3, 3} = U_{3 \rightarrow 3, 3} = \emptyset, \forall l > 1.
\]

In this case, the Markov chains of Theorem 1 reduce to:

\[
U_{1 \rightarrow 3, 1} \rightarrow X_1 \rightarrow (X_2, X_3, U_{2 \rightarrow 3, 1}), \\
U_{2 \rightarrow 3, 1} \rightarrow X_2 \rightarrow (X_1, X_3, U_{1 \rightarrow 3, 1}),
\]

and thus the inner bound from Theorem 1 specializes to the results in [21]:

\[
R_1 > I(X_1; U_{1 \rightarrow 3, 1} | X_3 U_{2 \rightarrow 3, 1}), \\
R_2 > I(X_2; U_{2 \rightarrow 3, 1} | X_1 U_{1 \rightarrow 3, 1}), \\
R_1 + R_2 > I(X_1, X_2; U_{1 \rightarrow 3, 1} U_{2 \rightarrow 3, 1} | X_3).
\]

2) Source Coding With Side Information at 2-Decoders [11], [12]: Consider the setting where one encoder observes \(X^1_n\) and transmits descriptions to two decoders with different side informations \((X^2_n, X^3_n)\) and distortion requirements \(D_2\) and \(D_3\). Again we consider only one way/round information exchange from node 1 (the encoder) to nodes 2 and 3 (the decoders). In this case, we set:

\[
U_{2 \rightarrow 13, 3} = U_{3 \rightarrow 12, 3} = \emptyset, \\
U_{1 \rightarrow 23, 3} = U_{2 \rightarrow 3, 3} = U_{3 \rightarrow 3, 3} = \emptyset, \forall l > 1.
\]

The above Markov chains imply:

\[
(U_{1 \rightarrow 23, 1}, U_{1 \rightarrow 2, 1}, U_{1 \rightarrow 3, 1}) \rightarrow X_1 \rightarrow (X_2, X_3) \tag{15}
\]

and thus the inner bound from Theorem 1 reduces to the results: in [11], [12]

\[
R_1 > \max \left\{ I(X_1; U_{1 \rightarrow 23, 1} | X_2), I(X_1; U_{1 \rightarrow 23, 1} | X_3) \right\} \\
+ I(X_1; U_{1 \rightarrow 2, 1} | X_2 U_{1 \rightarrow 23, 1}) \\
+ I(X_1; U_{1 \rightarrow 3, 1} | X_3 U_{1 \rightarrow 23, 1}). \tag{16}
\]

3) Two Terminal Interactive Source Coding [1]: Our inner bound (8) is basically the generalization of the interactive two terminal problem to the three-terminal setting. Assume only two encoders-decoders which observe \(X^1_n\) and \(X^2_n\) must reconstruct the other terminal source with distortion constraints \(D_1\) and \(D_2\), and after \(K\) rounds of information exchange. Let us set:

\[
U_{1 \rightarrow 23, 1} = U_{2 \rightarrow 13, 1} = U_{3 \rightarrow 12, 1} = \emptyset, \\
U_{1 \rightarrow 3, 1} = U_{3 \rightarrow 2, 1} = U_{2 \rightarrow 3, 1} = U_{3 \rightarrow 3, 1} = \emptyset, \forall l, X_3 = \emptyset.
\]

The Markov chains from Theorem 1 become

\[
U_{1 \rightarrow 2, 1} = (X_1, V_{12, 1}) \rightarrow X_2, \tag{17}
\]

\[
U_{2 \rightarrow 1, 1} = (X_2, V_{12, 1}) \rightarrow X_2, \tag{18}
\]

for \(l \in [1 : K]\) and thus the inner bound from Theorem 1 permit us to obtain the results in [1]:

\[
R_1 > I(X_1; V_{12, K+1, 1} | X_2), \tag{19}
\]

\[
R_2 > I(X_2; V_{12, K+1, 1} | X_1). \tag{20}
\]

4) Two Terminal Interactive Source Coding With a Helper [15]: Consider now two encoders/decoders, which observe \(X^2_n\) and \(X^3_n\), and must reconstruct the other terminal source with distortion constraints \(D_2\) and \(D_3\), respectively, using \(K\) communication rounds. Assume also that another encoder observes \(X^1_n\) and provides both nodes 2 and 3 with a common description before beginning the information exchange and then remains silent. Such common description can be exploited as coded side information. Let us set:

\[
U_{2 \rightarrow 13, 1} = U_{3 \rightarrow 12, 1} = \emptyset, \\
U_{1 \rightarrow 23, 1} = \emptyset, \forall l > 1, \\
U_{1 \rightarrow 3, 1} = U_{1 \rightarrow 3, 1} = U_{2 \rightarrow 3, 1} = U_{3 \rightarrow 1, 1} = \emptyset, \forall l.
\]

The Markov chains reduce to:

\[
U_{1 \rightarrow 23, 1} \rightarrow X_1 \rightarrow (X_2, X_3), \\
U_{2 \rightarrow 3, 1} \rightarrow (X_2, U_{1 \rightarrow 23, 1}, V_{23, 1, 2}) \rightarrow (X_1, X_3), \\
U_{3 \rightarrow 2, 1} \rightarrow (X_3, U_{1 \rightarrow 23, 1}, V_{23, 1, 3}) \rightarrow (X_1, X_2).
\]

An inner bound to the rate-distortion region for this problem reduces to (using the rate equations in our Theorem 1):

\[
R_1 > \max \left\{ I(X_1; U_{1 \rightarrow 23, 1} | X_2), I(X_1; U_{1 \rightarrow 23, 1} | X_3) \right\}, \\
R_2 > I(X_2; V_{23, K+1, 2} | X_3 U_{1 \rightarrow 23, 1}), \\
R_3 > I(X_3; V_{23, K+1, 2} | X_2 U_{1 \rightarrow 23, 1}).
\]

This inner bound contains as a special case the rate-distortion region in [15]. In that paper it is further assumed (in order to have a converse result) that \(X_1 \rightarrow X_3 \rightarrow X_2\). When this is the case, the lower bound of \(R_1\) is \(I(X_1; U_{1 \rightarrow 23, 1} | X_2)\).

V. NEW RESULTS ON INTERACTIVE AND COOPERATIVE SOURCE CODING

A. Two Encoders and One Decoder Subject to Lossy/Lossless Reconstruction Constraints Without Side Information

Consider now the problem described in Fig. 2 where encoder 1 wishes to communicate the source \(X^1_n\) to node 3 in a lossless manner while encoder 2 wishes to send a lossy description of the source \(X^2_n\) to node 3 with distortion constraint \(D_3\). The encoders use \(K\) communication rounds. This problem can be seen as the cooperating encoders version of the well-known Berger–Yeung [6] problem. A practical situation
where this setting could be of interest arises in wireless sensor networks (WSN) where the nodes collect measurements of spatially correlated sources and a fusion center (FC) requires the reconstruction of the different sources with distinct degrees of fidelity. As the information is correlated the agents can cooperate in order to efficiently exploit the sources dependence structure. The cooperation is done with exchanges of messages over the wireless medium. Because of the broadcast nature of the wireless channel the FC (node 3) have also access to those messages.

Theorem 2: The rate-distortion region of the setting described in Fig. 2 is given by the union, over all joint probability measures $p_{X_1 X_2 U_{2→13}}$, such that there exists a reconstruction mapping:

$$\hat{X}_{32} = g_{32}(X_1, U_{2→13})$$

with

$$E\left[ d(\hat{X}_{32}, X_{32}) \right] \leq D_{32},$$

of the sets of all tuples satisfying:

$$R_1 \geq H(X_1|X_2),$$

$$R_2 \geq I(X_2; U_{2→13}|X_1),$$

$$R_1 + R_2 \geq H(X_1) + I(X_2; U_{2→13}|X_1).$$

The auxiliary random variable $U_{2→13}$ has a cardinality bound of $\|U_{2→13}\| \leq \|X_1\| + 1$.

Remark 6: It is worth emphasizing that the rate-distortion region in Theorem 2 outperforms the non-cooperative rate-distortion region first derived in [6]. This is due to two facts: the conditional entropy given in the constraint for $R_1$ which is strictly smaller than the entropy $H(X_1|U_{2→13})$ present in the rate-region in [6], and the fact that the random description $U_{2→13}$ may be arbitrarily dependent on both sources $(X_1, X_2)$ which is not the case without cooperation [6]. Therefore, cooperation between encoders 1 and 2 reduces the rate needed to communicate the source $X_1$ while increasing the set of all admissible auxiliary random variables.

Remark 7: Notice that the rate-distortion region in Theorem 2 is achievable with a single round of interaction $K = 1$, which implies that multiple rounds do not improve the rate-distortion region in this case. This basically holds because node 3 reconstructs $X_1$ in a lossless fashion.

Remark 8: Although in the considered setting of Fig. 8 node 1 is not supposed to decode neither a lossy description nor the complete source $X_2$, if nodes 1 and 3 wish to recover exactly the same descriptions the optimal region remains the same as that in Theorem 2 with only a difference: it is necessary to impose the existence of a function $g_{12}(X_1, U_{2→13}) = \hat{X}_{12}$ which must satisfy an additional distortion constraint

$$E\left[ d(\hat{X}_{12}, X_{12}) \right] \leq D_{12}.$$ 

In order to show this, it is enough to check that in the converse proof given below the specific choice of the auxiliary random variable already allows node 1 to recover a general function $\hat{X}_{12}[t] = g_{12}[t](X_{1}[t], U_{2→13}[t])$ for each time $t \in [1 : n]$.

Proof: The direct part of the proof simply follows by choosing:

$$U_{1→23,1} = X_1, \quad U_{2→13,1} = U_{2→13},$$

and all of the rest auxiliary random variables being set to a constant for all $l$. It is straightforward to show that the rate-distortion region in Theorem 1 reduces to the desired region. We now proceed to the proof of the converse. If a pair of rates $(R_1, R_2)$ and distortion $D_{32}$ are admissible for the $K$-step interactive cooperative distributed source coding setting described in Fig. 2, then for all $\epsilon > 0$ there exists $n_0(\epsilon, K)$, such that $\forall n > n_0(\epsilon, K)$ there exists a $K$-step interactive source code $(n, K, F, G)$ with intermediate rates satisfying $\frac{1}{n} \sum_{i=1}^{K} \log \mathcal{J}_{i}^{K} \leq R_i + \epsilon$, $i \in \{1, 2\}$ and with average per-letter distortions with respect to the source $X_2$ and perfect reconstruction with respect to the source $X_1$ given by:

$$E\left[ d(X_2^n, \hat{X}_{32}^n) \right] \leq D_{32} + \epsilon,$$

$$\Pr(\hat{X}_{32}^n \neq X_{32}^n) \leq \epsilon,$$

where

$$\hat{X}_{32}^n \triangleq g_{32}(\mathcal{J}_{1}^{[1:K]}, \mathcal{J}_{2}^{[1:K]}),$$

$$\hat{X}_{31}^n \triangleq g_{31}(\mathcal{J}_{1}^{[1:K]}, \mathcal{J}_{2}^{[1:K]}).$$

For each $t \in [1 : n]$, we define random variables $U_{2→13[t]}$ as follows:

$$U_{2→13[t]} \triangleq (\mathcal{J}_{1}^{[1:K]}, \mathcal{J}_{2}^{[1:K]}, X_{1[t]}),$$

By the condition (22) which says that $\Pr(\hat{X}_{32}^n \neq \hat{X}_{31}^n) \leq \epsilon$ and Fano’s inequality [22], we have

$$H(X_1^n | \hat{X}_{31}^n) \leq \Pr(\hat{X}_{31}^n \neq X_{31}^n) \log_2(\|X_1^n\| - 1) + H_2(\Pr(\hat{X}_{31}^n \neq X_{31}^n)) \triangleq n\epsilon_n,$$

where $\epsilon_n(\epsilon) \to 0$ provided that $\epsilon \to 0$ and $n \to \infty$.

1) Rate at Node 1: For the first rate, we have

$$n(R_1 + \epsilon) \geq I(1^{[1:K]}, X_1^n | X_2^n)$$

$$(a) \geq n H(X_1^n | X_2^n) - H(X_1^n | X_2^n, \mathcal{J}_{1}^{[1:K]}, \mathcal{J}_{2}^{[1:K]}),$$

$$(b) \geq n H(X_1^n | X_2^n) - H(X_1^n | \hat{X}_{31}^n),$$

$$(c) \geq n H(X_1^n | X_2^n) - n\epsilon_n,$$

where

- step (a) follows from the fact that by definition of the code the sequence $\mathcal{J}_{1,2}^{[1:K]}$ is a function of the source $X_2^n$ and the vector of messages $\mathcal{J}_{1,2}^{[1:K]}$,
- step (b) follows from the converse assumptions that guarantees the existence of a reconstruction function $\hat{X}_{31}^n \triangleq g_{31}(\mathcal{J}_{1}^{[1:K]}, \mathcal{J}_{2}^{[1:K]}),$ and
- step (c) follows from Fano’s inequality in (25).
2) Rate at Node 2: For the second rate, we have
\[ n(R_2 + \epsilon) \geq I \left( \mathcal{J}_2^{[1:K]}; X_1^n \right) \]
\[ \geq \sum_{i=1}^{n} I \left( \mathcal{J}_2^{[1:K]}; X_2[i] \mid X_1[i] X_1[-i] X_2[1:2i-1] \right) \]
\[ \geq I \left( \mathcal{J}_2^{[1:K]}; X_1^n \right) + I \left( \tilde{U}_{2\rightarrow13}; X_2 \mid X_1 \right) \]
where
- step (a) follows from the chain rule for conditional mutual information and non-negativity of mutual information,
- step (b) follows from the memoryless property across time of the source \( X_1^n \) and standard properties of mutual information,
- step (c) follows from the non-negativity of mutual information \( I \left( \mathcal{J}_1^{[1:K]}; \mathcal{J}_2^{[1:K]} \right) \),
- step (d) follows from the use of a time sharing random variable \( Q \) uniformly distributed over the set \([1:n]\),
- step (e) follows from the definition of the conditional mutual information,
- step (f) follows by defining a new random variable \( \tilde{U}_{2\rightarrow13} \triangleq (U_{2\rightarrow13})_Q \),
- step (g) follows from the non-negativity of mutual information.

3) Sum-Rate of Nodes 1 and 2: For the sum-rate, we have
\[ n(R_1 + R_2 + 2\epsilon) \geq H \left( \mathcal{J}_1^{[1:K]} \right) + n(R_2 + \epsilon) \]
\[ \geq H \left( \mathcal{J}_1^{[1:K]} \right) + I \left( \mathcal{J}_1^{[1:K]}; X_1^n \right) \]
\[ \geq H \left( \mathcal{J}_1^{[1:K]} \right) + I \left( \mathcal{J}_2^{[1:K]}; X_1^n \right) \]
\[ \geq H \left( \mathcal{J}_1^{[1:K]} \right) + I \left( \mathcal{J}_2^{[1:K]}; X_1^n \right) + I \left( \tilde{U}_{2\rightarrow13}; X_2 \mid X_1 \right) \]
\[ \geq n \left[ H(X_1) + I \left( \tilde{U}_{2\rightarrow13}; X_2 \mid X_1 \right) \right] \]
\[ - H \left( X_1^n \mid \mathcal{J}_1^{[1:K]} \right) \]
\[ + I \left( X_1^n \mid \mathcal{J}_2^{[1:K]} \right) \]
\[ \geq n \left[ H(X_1) + I \left( \tilde{U}_{2\rightarrow13}; X_2 \mid X_1 \right) \right] \]
\[ - H \left( X_1^n \mid \tilde{X}_{32} \right) \]
\[ \geq n \left[ H(X_1) + I \left( \tilde{U}_{2\rightarrow13}; X_2 \mid X_1 \right) \right] - \epsilon_n, \]
where
- step (a) follows from inequality (26),
- step (b) follows from non-negativity of mutual information \( I \left( \mathcal{J}_1^{[1:K]}; \mathcal{J}_2^{[1:K]} \right) \),
- step (c) follows from the code assumption that guarantees the existence of reconstruction function \( \hat{X}_{31} = g_31 \left( \mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]} \right) \) and from the fact that conditioning reduces entropy,
- step (e) follows from Fano’s inequality in (25).

4) Distortion at Node 3: Node 3 computes \( \tilde{X}_{32} = g_{32} \left( \mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]} \right) \) and \( \hat{X}_{32} = g_{32} \left( \mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]} \right) \). For each \( t \in [1:n] \), let us define function \( \hat{X}_{32[t]} \) as the \( t \)-th coordinate of \( \hat{X}_{32} \):
\[ \hat{X}_{32[t]} \left( X_{1[t]}, U_{2\rightarrow13[t]} \right) \triangleq g_{32[t]} \left( \mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]} \right) \]
The component-wise mean distortion verifies
\[ D_{32} + \epsilon \]
\[ \geq \mathbb{E} \left[ d \left( X_2, g_{31} \left( \mathcal{J}_1^{[1:K]}, \mathcal{J}_2^{[1:K]} \right) \right) \right] \]
\[ = \sum_{t=1}^{n} \mathbb{E} \left[ d \left( X_{2[t]}, \hat{X}_{32[t]} \left( X_{1[t]}, U_{2\rightarrow13[t]} \right) \right) \mid Q = t \right] \]
\[ = \mathbb{E} \left[ d \left( X_2, \hat{X}_{32} \left( X_{1}, \tilde{U}_{2\rightarrow13} \right) \right) \right] \]
where we defined function \( \hat{X}_{32} \) by
\[ \hat{X}_{32} \left( X_{1}, \tilde{U}_{2\rightarrow13} \right) \triangleq \hat{X}_{32} \left( X_{1}, U_{2\rightarrow13} \right) \]
\[ \text{With Side Information} \]

B. Two Encoders and Three Decoders Subject to Lossless/Lossy Reconstruction Constraints

Consider now the problem described in Fig. 3 where encoder 1 wishes to communicate losslessly the source \( X_1^n \) to nodes 2 and 3 while encoder 2 wishes to send a lossy
description of its source $X^n$ to nodes 1 and 3 with distortion constraints $D_{12}$ and $D_{32}$, respectively. In addition to this, the encoders exchange messages using $K$ communication rounds. This problem can be seen as a generalization of the settings previously investigated in [4] and [6]. A practical motivation for this model could be a heterogeneous, decentralized and multi-task WSN where the nodes collect spatially correlated information and require to estimate, with varying degrees of quality, the source measurements of other nodes or some predefined functions of them (see [23] and the references therein for more specific examples).

**Theorem 3:** The rate-distortion region of the setting described in Fig. 3 is given by the union, over all joint probability measures $p_{X_1,X_2,U_{2\rightarrow 13}U_{2\rightarrow 3}}$ that satisfy the Markov chain $$(U_{2\rightarrow 13}, U_{2\rightarrow 3}) \rightarrow (X_1, X_2) \rightarrow X_3$$ (28) and such that there exists reconstruction mappings: $$\hat{X}_{32} = g_{32}(X_1, X_3, U_{2\rightarrow 13}, U_{2\rightarrow 3}), \quad E[d(X_2, \hat{X}_{32})] \leq D_{32},$$ $$\hat{X}_{12} = g_{12}(X_1, U_{2\rightarrow 13}), \quad E[d(X_2, \hat{X}_{12})] \leq D_{12},$$ of the sets of all tuples satisfying: $$R_1 \geq H(X_1|X_2),$$ $$R_2 \geq I(U_{2\rightarrow 13}; X_2|X_1) + I(U_{2\rightarrow 3}; U_{2\rightarrow 13}X_1X_3),$$ $$R_1 + R_2 \geq H(X_1) + I(U_{2\rightarrow 13}U_{2\rightarrow 3}; X_1X_3).$$ The auxiliary random variables have cardinality bounds: $$\|U_{2\rightarrow 3}\| \leq \|X_1\|\|X_2\| + 2 \quad \text{and} \quad \|U_{2\rightarrow 13}\| \leq \|X_1\|\|X_2\| \times \|U_{2\rightarrow 13}\| + 1.$$ **Remark 9:** Notice that the rate-distortion region in Theorem 3 is achievable using a single round of interactions $K = 1$, which implies that multiple rounds do not improve the rate-distortion region in this case.

**Remark 10:** It is worth mentioning that cooperation between encoders reduces the rate needed to communicate the source $X_2$ while increasing the optimization set of all admissible auxiliary random variables.

**Proof:** The direct part of the proof follows by choosing $U_{2\rightarrow 13,1} \triangleq U_{2\rightarrow 13}$ and $U_{2\rightarrow 3,1} \triangleq U_{2\rightarrow 3}$ and setting equal to a constant the rest of the auxiliary random variables. Notice that according to Theorem 1 the auxiliary random variables should satisfy: $$U_{2\rightarrow 13} \rightarrow (X_1, X_2) \rightarrow X_3,$$ $$U_{2\rightarrow 3} \rightarrow (U_{2\rightarrow 13}, U_{12}, X_2) \rightarrow X_3.$$ (30) It is easy to see that these Markov chains are equivalent to (28). From the rate equations in Theorem 1, and the above choices for the auxiliary random variables we obtain: $$R_{1\rightarrow 23} > H(X_1|X_2),$$ $$R_{2\rightarrow 13} > I(X_2; U_{2\rightarrow 13}|X_1),$$ $$R_{1\rightarrow 23} + R_{2\rightarrow 13} > H(X_1|X_3) + I(X_2; U_{2\rightarrow 13}|X_1X_3),$$ $$R_{2\rightarrow 3} > I(X_2; U_{2\rightarrow 3}|U_{2\rightarrow 13}X_1X_3),$$ where for obtaining the bound for $R_{2\rightarrow 23}$ we used that $I(X_2; U_{2\rightarrow 13}|X_1) \geq I(X_2; U_{2\rightarrow 13}|X_1X_3)$. Noticing that $R_1 \triangleq R_{1\rightarrow 23}$ and $R_2 \triangleq R_{2\rightarrow 13} + R_{2\rightarrow 3}$ the rate-distortion region (8) reduces to the desired region. We now proceed to the proof of the converse. If a pair of rates $(R_1, R_2)$ and distortions $(D_{12}, D_{32})$ are admissible for the $K$-step interactive cooperative distributed source coding setting described in Fig. 3, then for all $\epsilon > 0$ there exists $n_0(\epsilon, K)$, such that $\forall n > n_0(\epsilon, K)$ there is a $K$-step interactive source code $(n, K, \mathcal{F}, \mathcal{G})$ with intermediate rates satisfying: $$\frac{1}{n} \sum_{i=1}^{K} \log \|J_i^n\| \leq R_i + \epsilon, \quad i \in \{1, 2\}$$ (31) and with average per-letter distortions with respect to the source 2 and perfect reconstruction with respect to the source 1 at all nodes: $$\Pr(X^n \neq \hat{X}_{21}^n) \leq \epsilon, \quad \Pr(X^n \neq \hat{X}_{31}^n) \leq \epsilon,$$ $$E[d(X^n_2, \hat{X}_{12}^n)] \leq D_{12} + \epsilon,$$ $$E[d(X^n_2, \hat{X}_{32}^n)] \leq D_{32} + \epsilon,$$ (32) (33) (34) where $$\hat{X}_{32}^n \triangleq g_{32}\left(J_1^{[1:K]}, J_2^{[1:K]}, X^n_3\right),$$ $$\hat{X}_{12}^n \triangleq g_{12}\left(J_2^{[1:K]}, X^n_2\right),$$ $$\hat{X}_{31}^n \triangleq g_{31}\left(J_1^{[1:K]}, J_2^{[1:K]}, X^n_3\right),$$ $$\hat{X}_{21}^n \triangleq g_{21}\left(J_1^{[1:K]}, X^n_2\right).$$ (35) For each $t \in [1: n]$, we define random variables $U_{2\rightarrow 13[t]}$ and $U_{2\rightarrow 3|t}$ as follows: $$U_{2\rightarrow 13[t]} \triangleq \left(J_1^{[1:K]}, J_2^{[1:K]}, X_{1[t-1]}, X_{3[1:t-1]}\right),$$ $$U_{2\rightarrow 3|t} \triangleq \left(U_{2\rightarrow 13[t]}, X_{3[t+1:n]}, X_{2[1:t-1]}\right).$$ (36) The fact that these choices of the auxiliary random variables satisfy the Markov chain (28) can be obtained from point 6) in Lemma 5. By the conditions (32) and Fano’s inequality, we have $$H(X_1^n|\hat{X}_{31}^n) \leq \Pr(X_1^n \neq \hat{X}_{31}^n) \log_2(\|X_1^n\| - 1) + H_2(\Pr(X_1^n \neq \hat{X}_{31}^n)) \leq n \epsilon_n,$$ $$H(X^n_2|\hat{X}_{21}^n) \leq \Pr(X_2^n \neq \hat{X}_{21}^n) \log_2(\|X_2^n\| - 1) + H_2(\Pr(X_2^n \neq \hat{X}_{21}^n)) \leq n \epsilon_n,$$ (37) (38) where $\epsilon_n(\epsilon) \rightarrow 0$ provided that $\epsilon \rightarrow 0$ and $n \rightarrow \infty$.

1) **Rate at Node 1:** From cut-set arguments similar to the ones used in Theorem 2 and Fano inequality, we can easily obtain: $$n(R_1 + \epsilon) \geq n\left[H(X_1|X_2) - \epsilon_n\right].$$ (39)
2) Rate at Node 2: For the second rate, we have

\[
n(R_2 + \epsilon) \overset{(a)}{\geq} I \left( \mathcal{J}_2^{[1:K]} ; X_1^n, X_2^n, X_3^n \right) \\
\overset{(b)}{\geq} I \left( \mathcal{J}_2^{[1:K]} ; X_2^n, X_3^n \mid X_1^n \right) \\
\overset{(c)}{=} \sum_{i=1}^{n} \left[ I \left( \mathcal{J}_2^{[1:K]} ; X_{1\mid t} \mid X_{3\mid t} ; X_{3\mid t} \mid X_{1\mid t} \right) \\
+ I \left( \mathcal{J}_2^{[1:K]} ; X_{1\mid t} \mid X_{3\mid t} ; X_{2\mid t-1} \mid X_{1\mid t}, X_{3\mid t} \right) \\
+ I \left( \mathcal{J}_2^{[1:K]} ; X_{1\mid t} \mid X_{3\mid t} \mid U_{2\rightarrow 13} \vert t \right) \right] \\
\overset{(d)}{=} \sum_{i=1}^{n} \left[ I \left( U_{2\rightarrow 13} \mid t \mid X_{1\mid t} ; X_{3\mid t} \right) \\
+ I \left( U_{2\rightarrow 13} \mid t \mid X_{1\mid t} ; X_{2\mid t} \mid X_{3\mid t} \right) \right] \\
\overset{(e)}{=} \sum_{i=1}^{n} \left[ I \left( U_{2\rightarrow 13} \mid Q \mid X_{1\mid Q}, Q = t \right) \\
+ I \left( U_{2\rightarrow 13} \mid Q \mid X_{1\mid Q} ; X_{2\mid Q} \right) U_{2\rightarrow 13} \mid Q, Q = t \right] \\
\overset{(f)}{=} \sum_{i=1}^{n} \left[ I \left( \tilde{U}_{2\rightarrow 13} ; X_{2\mid 1} \right) \\
+ I \left( \tilde{U}_{2\rightarrow 3} ; X_{2\mid 1} ; X_{3\mid 1} \tilde{U}_{2\rightarrow 13} \right) \right],
\]

where

- step (a) follows from the fact that \( \mathcal{J}_2^{[1:K]} \) is a function of the sources \( X_1^n, X_2^n, X_3^n \),
- step (b) follows from the fact that \( \mathcal{J}_2^{[1:K]} \) is a function of \( \mathcal{J}_1^{[1:K]} \) and \( X_1^n \) and from the non-negativity of mutual information,
- step (c) follows from the chain rule for conditional mutual information and from the memoryless property across time of the sources \( X_1^n, X_2^n, X_3^n \),
- step (d) follows from the chain rule for conditional mutual information and the definitions in (36),
- step (e) follows from the Markov chain \( U_{2\rightarrow 13} \mid t \rightarrow (X_{1\mid t}, X_{3\mid t}) \rightarrow X_{3\mid t} \), for all \( t \in [1 : n] \),
- step (f) follows from the use of a time-sharing random variable \( Q \) uniformly distributed over the set \( [1 : n] \),
- step (g) follows by letting new random variables \( \tilde{U}_{2\rightarrow 13} \triangleq (U_{2\rightarrow 13} \mid Q), \tilde{U}_{2\rightarrow 3} \triangleq (U_{2\rightarrow 3} \mid Q) \).

3) Sum-Rate of Nodes 1 and 2: For the sum-rate, we have

\[
n(R_1 + R_2 + 2\epsilon) \overset{(a)}{\geq} H \left( \mathcal{J}_1^{[1:K]} \right) + H \left( \mathcal{J}_2^{[1:K]} \right) \\
\overset{(b)}{=} I \left( \mathcal{J}_1^{[1:K]} ; X_2^nX_3^nX_2^n \mid X_1^n \right) + I \left( \mathcal{J}_1^{[1:K]} ; X_2^n \mid X_1^nX_2^n \right) \\
\overset{(c)}{=} H \left( X_1^n \mid X_3^n \right) - H \left( X_1^n \mid \mathcal{J}_1^{[1:K]} \mathcal{J}_2^{[1:K]} \right) \\
+ I \left( \mathcal{J}_1^{[1:K]} ; \mathcal{J}_2^{[1:K]} ; X_2^nX_3^nX_2^n \right) \\
\overset{(d)}{\geq} H \left( X_1^n \mid X_3^n \right) - H \left( X_1^n \mid \tilde{X}_3^n \right) \\
+ I \left( \mathcal{J}_1^{[1:K]} ; \mathcal{J}_2^{[1:K]} ; X_2^n \right) \\
\overset{(e)}{=} n \left[ H \left( X_1 \mid X_3 \right) - \epsilon_n \right] + I \left( \mathcal{J}_1^{[1:K]} ; \mathcal{J}_2^{[1:K]} ; X_2^n \right) \\
\overset{(f)}{=} n \left[ H \left( X_1 \mid X_3 \right) - \epsilon_n \right] \\
+ \sum_{i=1}^{n} I \left( \mathcal{J}_1^{[1:K]} ; \mathcal{J}_2^{[1:K]} ; X_{1\mid t} \mid X_{3\mid t} ; X_{2\mid t-1} \right) ; \\
X_{2\mid t} \mid X_{1\mid t} X_{3\mid t} \right],
\]

where

- step (a) follows from the fact that \( \mathcal{J}_1^{[1:K]} \) and \( \mathcal{J}_2^{[1:K]} \) are functions of the sources \( X_1^n, X_2^n, X_3^n \),
- step (b) follows non-negativity of mutual information,
- step (c) follows from the converse assumptions in (35) that guarantees the existence of reconstruction function \( \tilde{X}_3^n \triangleq g_{31} \left( \mathcal{J}_1^{[1:K]} , \mathcal{J}_2^{[1:K]} , X_3^n \right) \),
- step (d) follows from Fano’s inequality in (37),
- step (e) follows from the chain rule of conditional mutual information and the memoryless property across time of the source \( X_1^n, X_2^n, X_3^n \),
- step (f) follows from steps (36),
- step (g) follows from the use of a time sharing variable \( Q \) uniformly distributed over \( [1 : n] \),
- step (h) follows by letting new random variables \( \tilde{U}_{2\rightarrow 13} \triangleq (U_{2\rightarrow 13} \mid Q), \tilde{U}_{2\rightarrow 3} \triangleq (U_{2\rightarrow 3} \mid Q) \).

4) Distortion at Node 1: Node 1 calculate \( \hat{X}_{12}^n \). It is clear that we can write without loss of generality \( \tilde{X}_{12}^n \triangleq g_{12} \left( \mathcal{J}_1^{[1:K]} , \mathcal{J}_2^{[1:K]} , X_3^n \right) \). For each \( t \in [1 : n] \), we define a function \( \tilde{X}_{12}^t \) as being the \( t \)-th coordinate of this estimate \( \tilde{X}_{12} \left( \tilde{U}_{2\rightarrow 13} \mid t \right) , X_{1\mid t} \right) \). It is straightforward to show, as in Theorem 2, that the component-wise mean distortion verifies

\[
D_{12} + \epsilon \overset{(a)}{\geq} E \left[ d \left( X_2^n, \tilde{X}_{12} \left( \tilde{U}_{2\rightarrow 13}, X_1^n \right) \right) \right],
\]

where we defined function \( \tilde{X}_{12} \) by

\[
\tilde{X}_{12} \left( \tilde{U}_{2\rightarrow 13}, X_1^n \right) = \tilde{X}_{12} \left( Q, U_{2\rightarrow 13} \mid Q, X_{1\mid t} \right) \triangleq \tilde{X}_{12} \mid Q \left( U_{2\rightarrow 13} \mid Q, X_{1\mid t} \right) .
\]
5) Distortion at Node 3: Node 3 reconstructs a lossy description \( \hat{X} \) is given by the union of all joint probability measures. For each \( t \in [1 : n] \), we define a function \( \hat{X}_{32} \) as being the \( t \)-th coordinate of this estimate:

\[
\hat{X}_{32[t]} (U_{2 \rightarrow 13}, U_{2 \rightarrow 3}, X_3) \triangleq g_{32[t]} \left( J_{1}^{[1 : K]}, J_{2}^{[1 : K]}, X \right).
\]

Similarly to the distortion at node 1, the component-wise mean distortion verifies

\[
D_{32} + \epsilon \geq \mathbb{E} \left[ d (X_2, \hat{X}_{32} (U_{2 \rightarrow 13}, U_{2 \rightarrow 3}, X_3)) \right]
\]

where we defined function \( \hat{X}_{32} \) by

\[
\hat{X}_{32} (U_{2 \rightarrow 13}, U_{2 \rightarrow 3}, X_3) \triangleq \hat{X}_{32} (U_{2 \rightarrow 13}, U_{2 \rightarrow 3}, X_3) (Q)
\]

the proof is concluded.

C. Two Encoders and Three Decoders Subject to Lossless/Lossy Reconstruction Constraints, Reversal Delivery and Side Information

Consider now the problem described in Fig. 4 where encoder 1 attempts to communicate losslessly the source \( X_1 \) to node 2 and a lossy description to node 3. Encoder 2 wishes to send a lossy description of its source \( X_2 \) to node 1 and a lossless one to node 3. The corresponding distortion levels at node 1 and 3 are \( D_{12} \) and \( D_{31} \), respectively. In addition to this, the encoders use \( K \) communication rounds. This problem is very similar to the problem described in Fig. 3, with the difference that the decoding goals at node 3 are inverted. Again, a motivating example for this problem could be in the field of multitasking on decentralized WSN. The resulting optimal region for this problem can be seen to be a special case of Theorem 3.

**Theorem 1:** The rate-distortion region of the setting described in Fig. 4 is given by the union over all joint probability measures \( p_{X_1, X_2, X_3, U_{2 \rightarrow 13}} \) satisfying the Markov chain

\[
U_{2 \rightarrow 13} \rightarrow (X_1, X_2) \rightarrow X_3 \tag{40}
\]

and such that there exist reconstruction mappings:

\[
\hat{X}_{31} = g_{31} (X_2, X_3, U_{2 \rightarrow 13}) \quad \text{with} \quad \mathbb{E} \left[ d (X_1, \hat{X}_{31}) \right] \leq D_{31},
\]

\[
\hat{X}_{12} = g_{12} (X_1, U_{2 \rightarrow 13}) \quad \text{with} \quad \mathbb{E} \left[ d (X_2, \hat{X}_{12}) \right] \leq D_{12},
\]

of the set of all tuples satisfying:

\[
R_1 \geq H (X_1 | X_2),
\]

\[
R_2 \geq I (U_{2 \rightarrow 13}; X_2 | X_1) + H (X_2 | U_{2 \rightarrow 13} X_1 X_3),
\]

\[
R_1 + R_2 \geq H (X_1 X_2 | X_3).
\]

The auxiliary random variable has cardinality bounds:

\[
||U_{2 \rightarrow 13}|| \leq ||X_1|| ||X_2|| + 3.
\]

**Remark 11:** Notice that the rate-distortion region in Corollary 1 is achievable with a single round of interactions \( K = 1 \), which implies that multiple rounds do not improve the rate-distortion region.

**Remark 12:** Notice that, although node 3 requires only the lossy recovery of \( X_1 \), it can in fact recover \( X_1 \) perfectly. This is due to the fact that node 3 requires also the lossless reconstruction of \( X_2 \). In this way, node 3 has the same information than node 2, which also has to recover \( X_1 \) losslessly. This explains the sum-rate term, which can be recognized to be the rate that guarantees the perfect recovery of \( X_1 \) and \( X_2 \) at node 3. We also see, that the cooperation helps in the Wyner-Ziv problem that exists between node 2 and 1, with an increasing of the set of the allowable auxiliary random variables thanks to the Markov chain (40).

**Proof:** The direct part of the proof follows by making exactly the same choices for the auxiliary variables than in Theorem 3, with the exception of \( U_{2 \rightarrow 3,1} \) which is chosen as \( X_2 \). The converse proof follows along the same lines of the corresponding one for Theorem 3.

D. Two Encoders and Three Decoders Subject to Lossy Reconstruction Constraints With Degraded Side Information

Consider now the problem described in Fig. 5 where encoder 1 has access to \( X_1 \) and \( X_2 \) and wishes to communicate a lossy description of \( X_1 \) to nodes 2 and 3 with distortion constraints \( D_{12} \) and \( D_{31} \), while encoder 2 wishes to send a lossy description of its source \( X_2 \) to nodes 1 and 3 with distortion constraints \( D_{12} \) and \( D_{32} \). To achieve this, encoders 1 and 2 communicate their descriptions using \( K \) communication rounds. This problem can be seen as a generalization of the settings previously investigated in [21]. This setup is motivated by the following application. Consider that node 1 transmits a probing signal \( X_3 \) which is used to explore a spatial region (i.e., a radar transmitter). After transmission of this probing signal, node 1 measures the response \( X_3 \) at its own location. Similarly, in a different location node 2 measures the response \( X_2 \). A Markov chain \( X_1 \rightarrow X_3 \rightarrow X_2 \) can be assumed for modeling this situation. Descriptions of signals \( X_1 \) and \( X_2 \) have to be sent to node 3 (e.g., the fusion center) which has knowledge of the probing signal \( X_3 \) and wants to reconstruct a lossy estimate of \( X_1 \) and \( X_2 \). Although it is not strictly necessary for this application we also requires that nodes 1 and 2 recover in a lossy fashion \( X_2 \) and \( X_1 \) respectively. Nodes 1 and 2 cooperate through multiple rounds to accomplish the task.

**Theorem 4:** The rate-distortion region for the setting described in Fig. 5 where \( X_1 \rightarrow X_3 \rightarrow X_2 \) is given by the union, over all joint probability measures \( p_{X_1, X_2, X_3, U_{1 \rightarrow 3, K}} \) satisfying the following Markov chains:

\[
U_{1 \rightarrow 23,l} \rightarrow (X_1, X_3, W_{l[1, K]}) \rightarrow X_2, \tag{41}
\]

\[
U_{2 \rightarrow 13,l} \rightarrow (X_2, W_{l[2, K]}) \rightarrow (X_1, X_3), \tag{42}
\]

\[
U_{1 \rightarrow 3,K} \rightarrow (X_1, X_3, W_{l[1, K + 1]}) \rightarrow X_2, \tag{43}
\]

for all \( l = [1 : K] \), and such that there exist reconstruction mappings:

\[
\hat{X}_{12} = g_{12} (X_1, X_3, U_{1 \rightarrow 3, K}, W_{l[1, K + 1]}), \tag{44}
\]

\[
\hat{X}_{21} = g_{21} (X_2, W_{l[1, K + 1]}) \tag{45}
\]
with \( \mathbb{E} \left[ d(X_2, \hat{X}_{12}) \right] \leq D_{12} \) and \( \mathbb{E} \left[ d(X_1, \hat{X}_{21}) \right] \leq D_{21} \) and
\[
\hat{X}_{31} = g_{31} \left( X, \mathcal{W}_{[1,K+1]}, U_{1 \rightarrow 3,K} \right),
\]
\[
\hat{X}_{32} = g_{32} \left( X, \mathcal{W}_{[1,K+1]}, U_{1 \rightarrow 3,K} \right),
\]
with \( \mathbb{E} \left[ d(X_1, \hat{X}_{31}) \right] \leq D_{31}, \mathbb{E} \left[ d(X_2, \hat{X}_{32}) \right] \leq D_{32} \) and
\[
\text{where } \mathcal{W}_{[1,i]} = \left\{ U_{1 \rightarrow 23,i}, U_{2 \rightarrow 13,i} \right\}_{i=1}^{K} \text{ for all } i = [1 : K],
\]
of the set of all tuples satisfying:
\[
R_1 \geq I (\mathcal{W}_{[1,K+1]}; X, X_3 | X_2) + I (U_{1 \rightarrow 3,K}; X_3 | \mathcal{W}_{[1,K+1]}),
\]
\[
R_2 \geq I (\mathcal{W}_{[1,K+1]}; X_2 | X_3).
\]

The auxiliary random variables have cardinality bounds:
\[
\| U_{i \rightarrow 23,i} \| \leq \| X_i \| \| X_3 \| \prod_{i=1}^{l-1} \| U_{i \rightarrow 23,i} \| \| U_{2 \rightarrow 13,i} \| + 2,
\]
\[
\| U_{2 \rightarrow 13,i} \| \leq \| X_2 \| \| U_{i \rightarrow 23,i} \| \prod_{i=1}^{l-1} \| U_{i \rightarrow 23,i} \| \| U_{2 \rightarrow 13,i} \| + 2,
\]
\[
\| U_{1 \rightarrow 3,K} \| \leq \| X_1 \| \| X_3 \| \prod_{i=1}^{K} \| U_{i \rightarrow 23,i} \| \| U_{2 \rightarrow 13,i} \| + 3
\]
for each \( i \in [1 : K] \).

**Remark 13:** Notice that multiple rounds are needed to achieve the rate-distortion region in Theorem 4. It is worth to mention that encoders 1 and 2 cooperate over the \( K \) rounds while only during the last round node 1 send a private description to node 3. Because of the Markov chain assumed for the sources we observe the following:

- Only node 1 sends a private description to node 3 which is due to the fact that node 3 disposes of a better side information than that intended to 2.
- For the transmission of node 2, both nodes 1 and 3 can be thought as an unique node and there is no reason for node 2 to send a private description to node 1 or node 3.
- Notice that the there is sum-rate. Node 3 could recover the descriptions generated at nodes 1 and 2 without resorting to joint-decoding. That is, node 3 can recover the descriptions generated at nodes 1 and 2 separately and independently.

**Proof:** The direct part of the proof follows by setting all auxiliary random variables equal to constants with the exception of \( U_{1 \rightarrow 23,i}, U_{2 \rightarrow 13,i} \) and \( U_{1 \rightarrow 3,K} \) which are random variables which, according to Theorem 1, should satisfy the Markov chains (41)-(43) and the required distortion constraints. Cumberbough but straightforward calculations allows to obtain the desired rate-distortion region. We now proceed to the proof of the converse. If a pair of rates \((R_1, R_2)\) and distortions \((D_{12}, D_{21}, D_{13}, D_{23})\) are admissible for the \( K \)-step interactive cooperative distributed source coding setting described in Fig. 5, then for all \( \epsilon > 0 \) there exists \( n_0(\epsilon,K) \) such that \( \forall n > n_0(\epsilon,K) \) there exists a \( K \)-step interactive source code \((n, K, \mathcal{F}, G)\) with intermediate rates satisfying
\[
\frac{1}{n} \sum_{i=1}^{K} \log \| J_i \| \leq R_i + \epsilon, \quad i \in \{1, 2\}
\]
and average per-letter distortions:
\[
\mathbb{E} \left[ d(X_1^n, \hat{X}_{12}^n) \right] \leq D_{21} + \epsilon, \quad \mathbb{E} \left[ d(X_2^n, \hat{X}_{31}^n) \right] \leq D_{31} + \epsilon,
\]
\[
\mathbb{E} \left[ d(X_3^n, \hat{X}_{21}^n) \right] \leq D_{32} + \epsilon,
\]
where
\[
\hat{X}_{21}^n = g_{21} \left( J_1^{[1,K]}, J_2^{[1,K]} \right), X_2^n,
\]
\[
\hat{X}_{31}^n = g_{31} \left( J_1^{[1,K]}, J_2^{[1,K]} \right), X_3^n,
\]
\[
\hat{X}_{32}^n = g_{32} \left( J_1^{[1,K]}, J_2^{[1,K]} \right), X_2^n.
\]
(44)

For each \( t \in [1 : n] \), define the following random variables:
\[
U_{1 \rightarrow 23,i,[t]} = J_{1,i}^{[1,K]}, X_{3,[t]}^{[1,K]} | X_{2,[t]}^{[1,K]},
\]
\[
U_{1 \rightarrow 23,K,[t]} = J_{1,K}^{[1,K]}, \forall k = [2 : K],
\]
\[
U_{2 \rightarrow 13,[t]} = J_{2,[t]}^{[1,K]}, \forall k = [1 : K],
\]
\[
U_{1 \rightarrow 3,K,[t]} = X_{3,[t]}^{[1,K]}.
\]
(45)

From Corollary 4 in the Appendices we see that these choices satisfy equations (41), (42) and (43).

1) **Rate at Node 1:** For the first rate, we have
\[
n(R_1 + \epsilon) \geq I \left( J_1^{[1,K]}, X_1^n | X_2^n \right)
\]
\[
\geq I \left( J_1^{[1,K]}, J_2^{[1,K]} | X_1^n, X_2^n \right)
\]
\[
\geq I \left( J_1^{[1,K]} X_1^n | J_2^{[1,K]} X_2^n \right)
\]
\[
= \sum_{i=1}^{n} I \left( J_1^{[1,K]} J_2^{[1,K]} X_{3,[t]}^{[1,K]} | X_{2,[t]}^{[1,K]} \right)
\]
\[
+ I \left( J_1^{[1,K]} J_2^{[1,K]} X_{2,[t]}^{[1,K]} | X_{3,[t]}^{[1,K]} \right)
\]
\[
= \sum_{i=1}^{n} I \left( J_1^{[1,K]} J_2^{[1,K]} X_{2,[t]}^{[1,K]} | X_{3,[t]}^{[1,K]} \right)
\]
\[
= \sum_{i=1}^{n} I \left( J_1^{[1,K]} J_2^{[1,K]} X_{3,[t]}^{[1,K]} | X_{2,[t]}^{[1,K]} \right)
\]
\[
+ I \left( J_1^{[1,K]} J_2^{[1,K]} X_{3,[t]}^{[1,K]} | X_{2,[t]}^{[1,K]} \right)
\]
\[
\geq \sum_{i=1}^{n} I \left( U_{1 \rightarrow 23,i,[t]} U_{2 \rightarrow 13,[t]} | X_{3,[t]}^{[1,K]} \right)
\]
\[
+ I \left( U_{1 \rightarrow 3,K,[t]} | X_{3,[t]}^{[1,K]} \right)
\]
\[
t = \sup \left[ I \left( \tilde{U}_{1 \rightarrow 23, [K]} \tilde{U}_{2 \rightarrow 13, [K]} | X_{3,[t]}^{[1,K]} \right) \right)
\]
\[
= I \left( \tilde{U}_{1 \rightarrow 3, [K]} | X_{3,[t]}^{[1,K]} \right)
\]

\[
\hat{X}_{21}^n = g_{21} \left( J_1^{[1,K]}, J_2^{[1,K]} \right), X_2^n,
\]
\[
\hat{X}_{31}^n = g_{31} \left( J_1^{[1,K]}, J_2^{[1,K]} \right), X_3^n,
\]
\[
\hat{X}_{32}^n = g_{32} \left( J_1^{[1,K]}, J_2^{[1,K]} \right), X_2^n.
\]
(44)

\[
\hat{X}_{31}^n = g_{31} \left( J_1^{[1,K]}, J_2^{[1,K]} \right), X_3^n,
\]
\[
\hat{X}_{32}^n = g_{32} \left( J_1^{[1,K]}, J_2^{[1,K]} \right), X_2^n.
\]
(44)

Notice that \( U_{3 \rightarrow 12,i} = \emptyset \) for all \( i \) because \( R_3 = 0 \).
• step (a) follows from the fact that $J^{[1:K]}_1$ is a function of the sources $(X^n_1, X^n_2, X^n_3)$,
• step (b) follows from the non-negativity of mutual information and from the fact that $J^{[1:K]}_2$ is a function of $J^{[1:K]}_1$ and the source $X^n_2$,
• step (c) follows from the chain rule for conditional mutual information and the memoryless property across time of the sources $(X^n_1, X^n_2, X^n_3)$,
• step (d) follows from the non-negativity of mutual information,
• step (e) follows from the Markov chain $X_{2[t]} \implies (J^{[1:K]}_1, J^{[1:K]}_2, X_{3[t+1:n]}) \implies (X_{3[t+1:n]} X_{1[t-1:j]}$ (Corollary 4 in the appendix A), for all $t \in [1 : n]$ which follows from $X_1 \iff X_3 \not{\iff} X_2$,
• step (f) follows from definitions in (45) and from the non-negativity of mutual information,
• step (g) follows from standard time-sharing arguments and the definition of new random variables, (i.e. $\bar{U}_{1 \rightarrow 3}[1 : K] \triangleq (U_{1 \rightarrow 3}, [1 : K], \bar{Q})$ and $X_1 \triangleq (X_{1[1 : K]} \iff \bar{Q})$).

The last step follows from the definition of the past shared common descriptions $V_{\{1,t\}} \forall t$. It is also immediate to show that $(\bar{U}_{1 \rightarrow 3}, \bar{U}_{2 \rightarrow 3})$ satisfies the Markov chains in (41)-(43) for all $t \in [1 : K]$.

2) Rate at Node 2: For the second rate, by following similar steps as before we can obtain:

$$n(R_2 + \epsilon) \geq n I (\tilde{V}_{\{1,K+1\}}; X_2 X_3).$$

(46)

3) Distortion at Nodes 1 and 2: Node 1 reconstructs an estimate $\hat{X}_{12}^{n} \triangleq g_{12}(J^{[1:K]}_1, J^{[1:K]}_2, X_1^n, X_2^n)$ while node 2 reconstructs $\hat{X}_{21}^{n} \triangleq g_{21}(J^{[1:K]}_1, J^{[1:K]}_2, X_2^n)$. For each $t \in [1 : n]$, define functions $\hat{X}_{12[t]}$ and $\hat{X}_{21[t]}$ as being the $t$-th coordinate of the corresponding estimates of $\hat{X}_{12}^{n}$ and $\hat{X}_{21}^{n}$, respectively:

$$\tilde{X}_{12[t]} (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_1^n, X_3^n) \triangleq g_{12[t]},$$

$$\tilde{X}_{21[t]} (V_{\{1,K+1\},[t]}, U_{2 \rightarrow 3,K,[t]}, X_2^n) \triangleq g_{21[t]}.$$  

(47)

We can easily show the following:

$$D_{12} + \epsilon \geq \mathbb{E} \left[ d (X_{2,t}, g_{12}(J^{[1:K]}_1, J^{[1:K]}_2, X_1^n)) \right]$$

(48)

$$\geq \frac{1}{n} \sum_{t=1}^{n} \mathbb{E} \left[ d (X_{2,t}, \tilde{X}_{12[t]}^{n}) \right]$$

(49)

$$\geq \frac{1}{n} \sum_{t=1}^{n} \mathbb{E} \left[ d (X_{2,t}, \tilde{X}_{21[t]}^{n}) \right]$$

(50)

$$\geq \mathbb{E} \left[ d (X_{2,t}, \tilde{X}_{12} (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_1^n, X_3^n)) \right]$$

(51)

where

• step (a) follows from (47),
• step (b) can be obtained from the Markov chain $X_{2[t]} \implies (X_{1[t]}, X_{3[t]}, J^{[1:K]}_1, J^{[1:K]}_2, X_{3[t-1]}, X_{2[t+1:n]}) \implies X_{1[t]}$ for each $t \in [1 : n]$ (which follows from Corollary 4 in the appendices) and the definition of a function $\tilde{X}_{12[t]}$ which depends only on $(V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_1^n, X_3^n)$.

By following the same steps, we can also show that:

$$D_{21} + \epsilon \geq \mathbb{E} \left[ d (X_{1,t}, \tilde{X}_{21} (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_1^n, X_3^n)) \right],$$

(52)

where

$$\tilde{X}_{31} \triangleq g_{31}(J^{[1:K]}_1, J^{[1:K]}_2, X_3^n) \quad \text{and} \quad \tilde{X}_{32} \triangleq g_{32}(J^{[1:K]}_1, J^{[1:K]}_2, X_3^n).$$

(53)

We can easily show that:

$$\tilde{X}_{31} \triangleq g_{31}(J^{[1:K]}_1, J^{[1:K]}_2, X_3^n) \quad \text{and} \quad \tilde{X}_{32} \triangleq g_{32}(J^{[1:K]}_1, J^{[1:K]}_2, X_3^n).$$

(54)

We can easily show that:

$$\tilde{X}_{31} \triangleq g_{31}(J^{[1:K]}_1, J^{[1:K]}_2, X_3^n) \quad \text{and} \quad \tilde{X}_{32} \triangleq g_{32}(J^{[1:K]}_1, J^{[1:K]}_2, X_3^n).$$

(55)

4) Distortions at Node 3: Node 3 compute lossy reconstructions $X_{31}^{n} \triangleq g_{31}(J^{[1:K]}_1, J^{[1:K]}_2, X_3^n)$ and $X_{32}^{n} \triangleq g_{32}(J^{[1:K]}_1, J^{[1:K]}_2, X_3^n)$. For each $t \in [1 : n]$, define functions $\tilde{X}_{31[t]}$ and $\tilde{X}_{32[t]}$ as being the $t$-th coordinate of the corresponding estimates of $X_{31}^{n}$ and $X_{32}^{n}$, respectively:

$$\tilde{X}_{31[t]} (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_3^n) \triangleq g_{31[t]},$$

$$\tilde{X}_{32[t]} (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_3^n) \triangleq g_{32[t]}.$$  

(56)

The component-wise mean distortions can be easily analyzed following similar arguments as before to obtain

$$D_{31} + \epsilon \geq \mathbb{E} \left[ d (X_{2,t}, \tilde{X}_{31} (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_3^n)) \right],$$

(57)

with

$$\tilde{X}_{31} (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_3^n) \triangleq g_{31[t]},$$

(58)

$$\tilde{X}_{32} (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_3^n) \triangleq g_{32[t]}.$$  

(59)

We can easily show that:

$$\tilde{X}_{31} (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_3^n) \triangleq g_{31[t]}(Q) (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_3^n) \triangleq g_{31[t]},$$

(60)

and

$$\tilde{X}_{32} (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_3^n) \triangleq g_{32[t]}(Q) (V_{\{1,K+1\},[t]}, U_{1 \rightarrow 3,K,[t]}, X_3^n) \triangleq g_{32[t]}.$$  

(61)

E. Three Encoders and Three Decoders Subject to Lossless/Lossy Reconstruction Constraints With Degraded Side Information

Consider now the problem described in Fig. 6 where encoder 1 wishes to communicate losslessly the source $X^n_1$ to nodes 2 and 3 while encoder 2 wishes to send a lossy description of its source $X^n_2$ to node 3 with distortion constraints $D_{23}$. Encoder 3 wishes to send a lossy description of its source $X^n_3$ to node 2 with distortion constraints $D_{32}$. To achieve this, the encoders perform the exchanges using $K$ communication rounds. This problem can be seen as a generalization of the settings previously investigated in [11]. This setting can model a situation in which node 1 generate a process $X_1$ (e.g., a radar probing signal). This process physically propagates to the locations of nodes 2 and 3. These nodes measure $X_2$ and $X_3$ respectively. If node 2 is closer to node 1 than node 3 we can assume that $X_1 \iff X_2 \not{\iff} X_3$. In this manner, we can think...
that signals $X_2$ and $X_3$ incorporate some kind of information about the characteristics of the surrounding area (which clearly influence the way signal $X_1$ propagates) where node 1, 2 and 3 are located. Nodes 2 and 3 then interact between them and with node 1, in order to reconstruct $X_1$ in lossless fashion and $X_2$ and $X_3$ with some distortion level. In this way, nodes 2 and 3 could use the knowledge of $X_1$ and the received signals $X_2$ and $X_3$ (with the predefined levels of distortion in order to limit the rate of their exchanges) to get some knowledge (in a cooperative manner) about the physical characteristics of the area where they are.

Theorem 5: The rate-distortion region of the setting described in Fig. 6 where $X_1 \rightarrow X_2 \rightarrow X_3$ form a Markov chain is given by the union over all joint probability measures $p_{X_1 X_2 X_3 U_3,2 (j), K (U_2, x_2, j)}$, satisfying the Markov chains

\begin{equation}
U_{2 \rightarrow 3, l} \rightarrow (X_1, X_2, \mathcal{V}_{[23, l, 2]}), \quad X_3,
\end{equation}
\begin{equation}
U_{3 \rightarrow 2, l} \rightarrow (X_1, X_3, \mathcal{V}_{[23, l, 3]}), \quad X_2, \quad \forall l \in [1 : K],
\end{equation}

and such that there exist reconstruction mappings:

\begin{equation}
\hat{X}_{23} = g_{23} (X_1, X_2, \mathcal{V}_{[23, K + 1, 2]}), \quad E \left[ d (X_3, \hat{X}_{23}) \right] \leq D_{23},
\end{equation}
\begin{equation}
\hat{X}_{32} = g_{32} (X_1, X_3, \mathcal{V}_{[23, K + 1, 2]}), \quad E \left[ d (X_2, \hat{X}_{32}) \right] \leq D_{32},
\end{equation}

of the set of all tuples satisfying:

\begin{align*}
R_1 & \geq H (X_1 | X_2), \\
R_2 & \geq I (V_{[23, K + 1, 2]}; X_2 | X_1 X_3), \\
R_3 & \geq I (V_{[23, K + 1, 2]}; X_3 | X_1 X_2), \\
R_1 + R_2 & \geq H (X_1 | X_3) + I (V_{[23, K + 1, 2]}; X_2 | X_1 X_3).
\end{align*}

The auxiliary random variables have cardinality bounds:

\begin{align*}
\| U_{2 \rightarrow 3, l} \| & \leq \| X_1 \| \| X_2 \| \prod_{i=1}^{l-1} \| U_{2 \rightarrow 3, i} \| \| U_{3 \rightarrow 2, l} \| + 3, \\
\| U_{3 \rightarrow 2, l} \| & \leq \| X_1 \| \| X_3 \| \prod_{i=1}^{l-1} \| U_{2 \rightarrow 3, i} \| \| U_{2 \rightarrow 2, l} \| + 3,
\end{align*}

for all $l \in [1 : K]$.

Remark 14: Theorem 5 shows that several exchanges between nodes 2 and 3 can be helpful. Node 1 transmit only once its full source at the beginning.

Proof: The direct part of the proof follows according to Theorem 1 by choosing for $l \in [1 : K]$:

\begin{equation}
U_{1 \rightarrow 3, l} = U_{1 \rightarrow 2, l} = U_{3 \rightarrow 1, l} = U_{2 \rightarrow 1, l} = U_{3 \rightarrow 2, l} = \emptyset, \\
U_{1 \rightarrow 23, l} = U_{2 \rightarrow 13, l} = \emptyset,
\end{equation}

and $U_{1 \rightarrow 23, 1} = U_{2 \rightarrow 13, 1} = X_1$. The remaining auxiliary random variables satisfy $\forall l \in [1 : K]$ the Markov chains in (50). With these choices and after some manipulations we obtain the desired region. For the converse, assume that rates $(R_1, R_2, R_3)$ and distortions $(D_{23}, D_{32})$ are admissible for the $K$-step interactive cooperative distributed source coding setting described in Fig. 6. Then for all $\epsilon > 0$ there exists $n_0 (\epsilon, K)$, such that $\forall n > n_0 (\epsilon, K)$ there exists a $K$-step interactive source code $(n, K, F, G)$ with intermediate rates satisfying $\sum_{i=1}^{K} \log \| J_i \| \leq R_i + \epsilon$, $i \in \{1, 2, 3\}$ and with average per-letter distortions with respect to the source 2 and 3 and perfect reconstruction with respect to the source 1 at all nodes given by:

\begin{align}
E \left[ d (X_2^n, \hat{X}_{23}^n) \right] & \leq D_{23} + \epsilon,
\end{align}
\begin{align}
E \left[ d (X_3^n, \hat{X}_{32}^n) \right] & \leq D_{32} + \epsilon,
\end{align}
\begin{align}
\Pr \left( X_1^n \neq \hat{X}_{23}^n \right) & \leq \epsilon,
\end{align}
\begin{align}
\Pr \left( X_1^n \neq \hat{X}_{32}^n \right) & \leq \epsilon,
\end{align}

where

\begin{align}
\hat{X}_{23} & \triangleq g_{23} (J_1^{[1:K]}, J_2^{[1:K]}, J_3^{[1:K]}, X_3^n), \\
\hat{X}_{32} & \triangleq g_{32} (J_1^{[1:K]}, J_2^{[1:K]}, J_3^{[1:K]}, X_2^n), \\
\hat{X}_{31} & \triangleq g_{31} (J_1^{[1:K]}, J_2^{[1:K]}, J_3^{[1:K]}, X_3^n), \\
\hat{X}_{21} & \triangleq g_{21} (J_1^{[1:K]}, J_2^{[1:K]}, J_3^{[1:K]}, X_2^n). \end{align}

For each $t \in [1 : n]$ and $l \in [1 : K]$, we define random variables $U_{2 \rightarrow 3, l [t]}$ and $U_{3 \rightarrow 2, l [t]}$ as follows:

\begin{align}
U_{2 \rightarrow 3, l [t]} & \triangleq (J_{11}^{[1]}, J_{12}^{[1]}, X_{[t]}^{[1]}, X_{2[t]}^{[1]} X_{3[t]}, X_{3[t]}^{[1]}), \\
U_{3 \rightarrow 2, l [t]} & \triangleq (J_{11}^{[1]}, J_{21}^{[1]}, l \in [2 : K], \\
U_{3 \rightarrow 2, l [t]} & \triangleq (J_{11}^{[1]}, l \in [1 : K]).
\end{align}

These auxiliary random variables satisfy the Markov conditions (50). This can be verified from Lemma 6 in the appendix. By the conditions in (53) and Fano’s inequality, we have

\begin{equation}
H (X_1^n | \hat{X}_{31}^n) \leq \Pr \left( X_1^n \neq \hat{X}_{31}^n \right) \log \| l X_1^n \| - 1
\end{equation}

\begin{equation}
+ H_2 \left( \Pr \left( X_1^n \neq \hat{X}_{31}^n \right) \right) \leq n \epsilon_n,
\end{equation}

\begin{equation}
H (X_1^n | \hat{X}_{21}^n) \leq \Pr \left( X_1^n \neq \hat{X}_{21}^n \right) \log \| l X_1^n \| - 1
\end{equation}

\begin{equation}
+ H_2 \left( \Pr \left( X_1^n \neq \hat{X}_{21}^n \right) \right) \leq n \epsilon_n,
\end{equation}

where $\epsilon_n (\epsilon) \rightarrow 0$ provided that $\epsilon \rightarrow 0$ and $n \rightarrow \infty$.

1) Rate at Node 1: For the first rate, it is straightforward to obtain (following similar arguments used for the previous theorems)

\begin{equation}
n (R_1 + \epsilon) \geq n [H (X_1 | X_2) - \epsilon_n],
\end{equation}

2) Rate at Nodes 2 and 3: For the second rate, we have

\begin{align}
n (R_2 + \epsilon) & \geq I (J_2^{[1:K]}, X_1^n X_2^n X_3^n), \\
& \geq I \left( J_2^{[1:K]}, X_2^n \right), \\
& \geq I \left( J_1^{[1:K]} J_2^{[1:K]} J_3^{[1:K]} X_3^n X_2^n \right), \\
& \geq \sum_{i=1}^{n} I \left( V_{[23, K + 1, 2]}^{[i]}, X_{2[t]}^{[i]} X_{3[t]}^{[i]} \right), \\
& \geq \sum_{i=1}^{n} I \left( V_{[23, K + 1, 2]}^{[i]}, X_{2[t]}^{[i]} X_{3[t]}^{[i]} \right), \\
& \geq n I \left( V_{[23, K + 1, 2]}^{[i]}, X_{2[t]}^{[i]} X_{3[t]}^{[i]} \right), \\
& \geq n I \left( V_{[23, K + 1, 2]}^{[i]}, X_{2[t]}^{[i]} X_{3[t]}^{[i]} \right),
\end{align}
By following similar steps, it is not difficult to check that

\[ n(R_3 + \varepsilon) \geq \sum_{t=1}^{n} I(\tilde{V}_{23,K+1,2}[t]; X_3[t], X_1[t], X_2[t]) \]

\[ = \sum_{t=1}^{n} I(\tilde{V}_{23,K+1,2} Q[t]; X_3[t], X_1[t], X_2[t], Q = t) \]

\[ \geq n I(\tilde{V}_{23,K+1,2}; X_3, X_1, X_2). \]

3) Sum-Rate of Nodes 1 and 2: For the sum-rate, we have

\[ n(R_1 + R_2 + 2\varepsilon) \geq H(\tilde{V}_{1}^{1:K}) + H(\tilde{V}_{2}^{1:K}) \]

(a) \[ = I(\tilde{V}_{1}^{1:K}, \tilde{V}_{2}^{1:K}, X_1^n, X_2^n, X_3^n ) \]

\[ + I(\tilde{V}_{1}^{1:K}, \tilde{V}_{2}^{1:K}, X_3^n, X_1^n, X_2^n) \]

(b) \[ = H(X_1^n, X_2^n, X_3^n) - H(X_1^n, X_2^n | \tilde{V}_{31}) \]

\[ + I(\tilde{V}_{1}^{1:K}, \tilde{V}_{2}^{1:K}, X_3^n, X_1^n, X_2^n). \]

(c) \[ = n I(\tilde{V}_{1}^{1:K}; X_3) - \epsilon_n \]

\[ + \sum_{t=1}^{n} I(\tilde{V}_{1}^{1:K}, \tilde{V}_{2}^{1:K}, X_3^n; X_{1}[t] | X_{2}[t], X_{3}[t]), \]

(d) \[ = n I(\tilde{V}_{1}^{1:K}; X_3) - \epsilon_n \]

\[ + \sum_{t=1}^{n} I(\tilde{V}_{23,K+1,2}; X_{2}[t] | X_{1}[t], X_{3}[t]). \]

(e) \[ = n I(\tilde{V}_{23,K+1,2}; X_2 | X_1, X_3). \]

where

- step (a) follows from the fact that \( \tilde{V}_{1}^{1:K} \) and \( \tilde{V}_{2}^{1:K} \) are functions of the sources \( (X_1^n, X_2^n, X_3^n) \),
- step (b) follows from the code assumption in (54) that guarantees the existence of reconstruction function \( \tilde{V}_{31} \triangleq g_{31}(\tilde{V}_{1}^{1:K}, \tilde{V}_{2}^{1:K}, \tilde{V}_{3}^{1:K}, X_3^n) \),
- step (c) follows from Fano’s inequality in (56), the chain rule of conditional mutual information and the memoryless property across time of the sources \( (X_1^n, X_2^n, X_3^n) \) and non-negativity of mutual information,
- step (d) from follows from the definitions in (55),
- step (e) follows from the use of a time sharing random variable \( Q \) uniformly distributed over \( [1 : n] \) and from definition \( \tilde{V}_{23,K+1,2} \triangleq (\tilde{V}_{23,K+1,2}[Q], Q) \).

4) Distortion at Node 2: Node 2 reconstructs a lossy \( \tilde{X}_{23} \triangleq g_{23}(\tilde{V}_{1}^{1:K}, \tilde{V}_{2}^{1:K}, \tilde{V}_{3}^{1:K}, X_3^n) \). For each \( t \in [1 : n] \), we define a function \( \tilde{X}_{32} \) as the \( t \)-th coordinate of this estimate:

\[ \tilde{X}_{32} \triangleq (\tilde{V}_{23,K+1,2}[Q], X_{2}[t]) \triangleq g_{32}[Q,t]. \]

The component-wise mean distortion can be easily analyzed mimicking previous developments to obtain

\[ D_{23} + \varepsilon \geq \mathbb{E} [d(X_3, \tilde{X}_{23}(\tilde{V}_{23,K+1,2}, X_2))], \]

where we defined function \( \tilde{X}_{23} \) by

\[ \tilde{X}_{23}(\tilde{V}_{23,K+1,2}, X_2) \triangleq \tilde{X}_{32} \triangleq g_{32}(\tilde{V}_{1}^{1:K}, \tilde{V}_{2}^{1:K}, \tilde{V}_{3}^{1:K}, X_3^n). \]

5) Distortion at Node 3: Node 3 reconstructs a lossy description \( \tilde{X}_{32} \triangleq g_{32}(\tilde{V}_{1}^{1:K}, \tilde{V}_{2}^{1:K}, \tilde{V}_{3}^{1:K}, X_3^n) \). For each \( t \in [1 : n] \), we define a function \( \tilde{X}_{32} \) as the \( t \)-th coordinate of this estimate:

\[ \tilde{X}_{32} \triangleq (\tilde{V}_{23,K+1,2}[Q], X_{2}[t]) \triangleq g_{32}[Q,t]. \]

The component-wise mean distortion analysis is identical to the one corresponding to node 2. We can easily obtain:

\[ D_{32} + \varepsilon \geq \mathbb{E} [d(X_3, \tilde{X}_{23}(\tilde{V}_{23,K+1,2}, X_2))], \]

where we defined function \( \tilde{X}_{32} \) by

\[ \tilde{X}_{32}(\tilde{V}_{23,K+1,2}, X_3) \triangleq g_{32}(\tilde{V}_{23,K+1,3}[Q], X_{3}[Q]). \]

VI. DISCUSSION

A. Numerical Example

In order to obtain further insight into the gains obtained from cooperation, we consider the case of two encoders and one decoder subject to lossy/lossless reconstruction constraints without side information (Theorem 2) in which the sources are distributed according to:

\[ p_{X_1 X_2}(x_1, x_2) = a I \{ x_1 = 1 \} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp \left( -\frac{x_1^2}{2\sigma_1^2} \right) \]

\[ + (1 - a) I \{ x_1 = 0 \} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left( -\frac{x_1^2}{2\sigma_0^2} \right). \]

(59)

We observe that \( X_1 \) follows a Bernoulli distribution with parameter \( a \in [0, 1] \) while \( X_2 \) given \( X_1 \) follows a Gaussian distribution with different variance according to the value of \( X_1 \in [0, 1] \). In this sense, \( X_2 \) follows a Gaussian mixture distribution.\(^9\) The optimal rate-distortion region for

\(^9\)Although the inner bound region in Theorem 1 is strictly valid for discrete sources with finite alphabets, the Gaussian distribution is sufficiently well-behaved to apply a uniform quantization procedure prior to the application of the results of Theorem 1. By limiting argument using a sequence of decreasing quantization step-sizes will deliver the desired result in [24, Ch. 3].
this case is given by Theorem 2 and can be alternatively written as:

\[
R_{\text{coop}}(D) = \bigcup_{p \in \mathcal{L}} \left\{ (R_1, R_2) : R_1 \geq H(X_1|X_2), \right. \\
R_2 \geq I(X_2; U|X_1), \left. \quad R_1 + R_2 \geq H(X_1) + I(X_2; U|X_1) \right\},
\]

where

\[
\mathcal{L} = \left\{ p_{U|X_2} : \text{there exists } (x_1, u) \mapsto g(x_1, u) \right. \\
\text{such that } \mathbb{E}[d(X_2, g(X_1, U))] \leq D \}. \tag{60}
\]

The corresponding non-cooperative region for the same problem was characterised in [6]:

\[
R_{\text{no-coop}}(D) = \bigcup_{p \in \mathcal{L}^*} \left\{ (R_1, R_2) : R_1 \geq H(X_1|U), \right. \\
R_2 \geq I(X_2; U|X_1), \left. \quad R_1 + R_2 \geq H(X_1) + I(X_2; U|X_1) \right\},
\]

where

\[
\mathcal{L}^* = \left\{ p_{U|X_2} : \text{there exists } (x_1, u) \mapsto g(x_1, u) \right. \\
\text{such that } \mathbb{E}[d(X_2, g(X_1, U))] \leq D \}. \tag{63}
\]

From the previous expressions, it is evident that the cooperative case offers some gains with respect to the non-cooperative setup. This is clearly evidenced from the lower limit in \(R_1\) and the fact that \(\mathcal{L}^* \subseteq \mathcal{L}\). We have the following result.

**Theorem 6 (Cooperative Region for a Mixed Gaussian Source):** Assume the source distribution is given by (59) and that, without loss of generality, \(\sigma_0^2 \leq \sigma_1^2\). The rate-distortion region from Theorem 2 can be written as:

\[
R_1 > \frac{1 - \alpha}{\sqrt{2\pi \sigma_0^2}} \int_{-\infty}^{\infty} \exp \left( -\frac{x^2}{2\sigma_0^2} \right) H_2(g(x))dx \\
+ \frac{1}{\sqrt{2\pi \sigma_1^2}} \int_{-\infty}^{\infty} \exp \left( -\frac{x^2}{2\sigma_1^2} \right) H_2(g(x))dx,
\]

where \(\mathcal{L}^* = \left\{ p_{U|X_2} : \text{there exists } (x_1, u) \mapsto g(x_1, u) \right. \\
\text{such that } \mathbb{E}[d(X_2, g(X_1, U))] \leq D \}. \tag{61}
\]

where \(|x|^+ = \max\{0, x\}\) and

\[
g(x) \triangleq \frac{\alpha}{\sqrt{2\pi \sigma_1^2}} \exp \left( -\frac{x^2}{2\sigma_1^2} \right) + \frac{1 - \alpha}{\sqrt{2\pi \sigma_0^2}} \exp \left( -\frac{x^2}{2\sigma_0^2} \right).
\]

**Proof:** The converse proof is straightforward by observing that when \(D \leq \sigma_0^2\):

\[
I(X_2; U|X_1) = h(X_2|X_1) - h(X_2|U, X_1) \\
\geq h(X_2|X_1) - \frac{1}{2} \log(2\pi eD) \tag{65}
\]

and \(h(X_2|X_1) = \frac{a}{2} \log(2\pi e\sigma_1^2) + \frac{1 - a}{2} \log(2\pi e\sigma_0^2)\). For the case when \(\sigma_0^2 > a\sigma_1^2 + (1 - a)\sigma_0^2\) we can write:

\[
I(X_1; U|X_1) \geq h(X_2|X_1) - ah(X_2|U, X_1 = 1) \\
- (1 - a)h(X_2|X_1 = 0) \tag{66}
\]

\[
= \frac{a}{2} \log(2\pi e\sigma_1^2) - a \log \left( \frac{2\pi e(D - (1 - a)\sigma_0^2)}{a} \right) \\
= \frac{a}{2} \log \left( \frac{a\sigma_1^2}{D - (1 - a)\sigma_0^2} \right).
\]

When \(D > a\sigma_1^2 + (1 - a)\sigma_0^2\) we can lower bound the mutual information by zero.

The achievability follows from the choice:

\[
g(U, X_1) = \begin{cases} 
\frac{\sigma_0^2}{\sigma_0^2 + \sigma_2^0} U & \text{if } X_1 = 0 \\
\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^1} U & \text{if } X_1 = 1,
\end{cases} \tag{67}
\]

and by setting the auxiliary random variable:

\[
U = \begin{cases} 
X_2 + Z_0 & \text{if } X_1 = 0 \\
X_2 + Z_1 & \text{if } X_1 = 1,
\end{cases} \tag{68}
\]

where \(Z_0, Z_1\) are zero-mean Gaussian random variables, independent from \(X_2\) and \(X_1\) and with variances given by:

\[
\sigma_{Z_0}^2 = \frac{D\sigma_0^2}{\sigma_0^2 - D}, \quad \sigma_{Z_1}^2 = \frac{D\sigma_1^2}{\sigma_1^2 - D}, \tag{69}
\]

when \(D \leq \sigma_0^2\), while for \(\sigma_0^2 < D \leq a\sigma_1^2 + (1 - a)\sigma_0^2\), we choose:

\[
\sigma_{Z_0}^2 \to \infty, \quad \sigma_{Z_1}^2 = \frac{D - (1 - a)\sigma_0^2}{a\sigma_1^2 - D} \tag{70}
\]

Finally, for \(D > a\sigma_1^2 + (1 - a)\sigma_0^2\), we let \(\sigma_{Z_0}^2 \to \infty\) and \(\sigma_{Z_1}^2 \to \infty\).

Unfortunately, the non-cooperative region is hard to evaluate for the assumed source model.10 In order to present some comparison between the cooperative and non-cooperative case let us fix the same value for the rate \(R_1\) in both cases and

\[10\] However, there are cases where an exact characterization is possible. This is the case, for example, when \(X_1\) and \(X_2\) are the input and output of a binary channel with crossover probability \(a\) and the distortion function is the Hamming distance [6].
In order to guarantee that (72) and (73) are achievable, under

\[ R \]

\[ H \]

= \left[ X_2 - \mathbb{E}_{X_2|X_1} \left[ X_2 | U, X_1 = 0 \right] \right]^2 | X_1 = 0 \right],

\[ \beta_1 = \mathbb{E}_{X_2|X_1} \left[ (X_2 - \mathbb{E}_{X_2|X_1} \left[ X_2 | U, X_1 = 1 \right])^2 | X_1 = 1 \right]. \]

The distortion constraint imposes the condition:

\[ (1 - \alpha) \beta_0 + \alpha \beta_1 \leq D. \]  \hspace{1cm} (73)

In order to guarantee that (72) and (73) are achievable, under the Markov constraint \( U \leftrightarrow X_2 \leftrightarrow X_1 \), the following conditions on \( p_{U|X_2}(u|x_2) \) should be satisfied:

\[ p_{U|X_2}(u|x_2) \frac{1}{\sqrt{2\pi \sigma_0}} \exp \left( -\frac{x_2^2}{2\sigma_0^2} \right) \]

\[ \int_{-\infty}^{\infty} p_{U|X_2}(u|x_2) \frac{1}{\sqrt{2\pi \sigma_0}} \exp \left( -\frac{x_2^2}{2\sigma_0^2} \right) dx_2 = N(f_0(u), \beta_0), \]

where we denote with \( N(\mu, \sigma^2) \) the density of Gaussian random variable with mean \( \mu \) and variance \( \sigma^2 \). The functions \( f_0(U) \) and \( f_1(U) \) are given by:

\[ f_0(U) \triangleq \mathbb{E}_{X_2|X_1} \left[ X_2 | U, X_1 = 0 \right], \]

\[ f_1(U) \triangleq \mathbb{E}_{X_2|X_1} \left[ X_2 | U, X_1 = 1 \right]. \]

The characterization of all distributions \( p_{U|X_2}(u|x_2) \) that satisfy (74) and (75) appears to be a difficult problem. In order to show a numerical example, we shall simply assume that:

\[ p_{U|X_2}(u|x_2) = N(u, \sigma_w^2). \]  \hspace{1cm} (76)

Indeed, this choice satisfies simultaneously expressions (74) and (75). In this way, we can calculate the corresponding values of \( \beta_0 \) and \( \beta_1 \) obtaining a parametrization of \( I(X_2; U|X_1) \) as a function of \( \sigma_w^2 \):

\[ I(X_2; U|X_1) = \frac{1 - \alpha}{2} \log \left( \frac{\sigma_0^2 + \sigma_w^2}{\sigma_0^2} \right) + \frac{\alpha}{2} \log \left( \frac{\sigma_1^2 + \sigma_w^2}{\sigma_1^2} \right) \]

\[ \int_{-\infty}^{\infty} p_{U|X_2}(u|x_2) \frac{1}{\sqrt{2\pi \sigma_1}} \exp \left( -\frac{x_2^2}{2\sigma_1^2} \right) dx_2 = N(f_1(u), \beta_1), \]

\[ \int_{-\infty}^{\infty} p_{U|X_2}(u|x_2) \frac{1}{\sqrt{2\pi \sigma_1}} \exp \left( -\frac{x_2^2}{2\sigma_1^2} \right) dx_2 = N(f_1(u), \beta_1), \]

\[ (1 - \alpha) \frac{\sigma_0^2 \sigma_w^2}{\sigma_0^2 + \sigma_w^2} + \alpha \frac{\sigma_1^2 \sigma_w^2}{\sigma_1^2 + \sigma_w^2} = D. \]  \hspace{1cm} (78)

We can replace (77) in (71) to obtain an indication of the performance of the non-cooperative case when \( R_1 \) is fixed.

We present now some numerical evaluations. As equation (71) is valid for both the cooperative and the non-cooperative setups, it is sufficient to compare the mutual information term \( I(X_2; U|X_1) \) for each of them. Let us consider the next scenarios:

1. \( \alpha = 0.1, \sigma_0^2 = 0.01, \sigma_1^2 = 2, \)
2. \( \alpha = 0.1, \sigma_0^2 = 0.5, \sigma_1^2 = 2. \)

From Fig. 7 we see that in the case \( \sigma_0^2 \ll \sigma_1^2 \) the gain of the cooperative scheme is pretty noticeable. However, as \( \sigma_0^2 \) becomes comparable to \( \sigma_1^2 \) the gains are reduced. This was expected from the fact that as \( \sigma_0^2 \to \sigma_1^2 \), the distribution of the random variable \( X_2 \) converges to a Gaussian distribution. In that case, the reconstruction of \( X_2 \) at Node 3 is equivalent, for the cooperative scenario, to a lossy source coding problem with side information \( X_1 \) at both the encoder and the decoder, while for the non-cooperative setting to the standard Wyner-Ziv problem.

### B. Interactive Lossless Source Coding

Consider now the problem represented in Fig. 8 where encoder 1 wishes to communicate losslessly the source \( X_1^n \) to two decoders which observe sources \( X_2^n \) and \( X_3^n \), respectively. At the same time node 1 wishes to recover \( X_2^n \) and \( X_3^n \) in lossless fashion. Similarly the other encoders want to communicate losslessly their sources and recover the sources.
from the rest in a similar manner. It is desired to do this through $K$ rounds of exchanges.

Theorem 7 (Interactive Lossless Source Coding): The rate region of the setting described in Fig. 8 is given by the set of all tuples satisfying:

\[
R_1 > H(X_1|X_2X_3), \\
R_2 > H(X_2|X_1X_3), \\
R_3 > H(X_3|X_1X_2), \\
R_1 + R_2 > H(X_1X_2X_3), \\
R_1 + R_3 > H(X_1X_3|X_2), \\
R_2 + R_3 > H(X_2X_3|X_1).
\]

Remark 15: It is worth observing that the multiple exchanges of descriptions between all nodes can not improve the rate region that could be obtained using Slepian-Wolf coding [2].

Proof: The achievability part is a standard exercise. The converse proof is straightforward from cut-set arguments. Both proofs are omitted.

We should note that for this important case, Theorem 1 does not provide the optimal rate region. That is, the coding scheme used is not optimal for this case. In fact, from Theorem 1 we can obtain the following achievable region\(^{11}\):

\[
R_1 > H(X_1|X_2), \\
R_2 > H(X_2|X_1X_3), \\
R_3 > H(X_3|X_1X_2), \\
R_1 + R_2 > H(X_1X_2|X_3), \\
R_1 + R_3 > H(X_1X_3|X_2), \\
R_2 + R_3 > H(X_2X_3|X_1).
\]

It is easily seen that in this region, node 2 is not performing joint decoding of the descriptions generated at node 1 and 3. Because of the encoding ordering assumed (1 $\rightarrow$ 2 $\rightarrow$ 3) and the fact that the common description generated in node 2 should be conditionally generated on the common description generated at node 1, node 2 has to recover this common description first. At the end, it recovers the common description generated at node 3. On the other hand, nodes 1 and 3 perform joint decoding of the common information generated at nodes 2 and 3, and at nodes 1 and 2, respectively. Clearly, this is a consequence of the sequential encoding and decoding structure imposed to the nodes in the network and which is the basis of the interaction. If all the nodes would be allowed to perform a joint decoding procedure, in order to recover all the exchanged descriptions only at the end of each round, this problem would not appear. However, this would destroy the sequential encoding-decoding structure assumed by our coding scheme which seems to be optimal in other situations.

\[\]
uniformly for every \((x^n, w^n, u^n_1, u^n_2) \in T^n_{[XWV_1V_2]}\) provided that:
\[
\log \mathbb{E}[A_1, A_2]_n = I(U_1; XV_2U_2|W_1) + I(U_2; XV_1U_1|W_2) - I(U_1; U_2|XV_1W_2) - \delta,
\]
where \(\delta = \delta(\epsilon, \epsilon', \epsilon_1, \epsilon_2, n) \to 0\) when \(\epsilon, \epsilon', \epsilon_1, \epsilon_2 \to 0\) and \(n \to \infty\).

**Corollary 2:** Assume the conditions in Lemma 1 are satisfied, and \(\Pr\left\{ (X^n, W^n, V^n_1, V^n_2) \in T^n_{[XWV_1V_2]} \right\} \to 1\) as \(n \to \infty\).

Then,
\[
\Pr\left\{ (U^n_1(m_1), U^n_2(m_2), X^n, W^n, V^n_1, V^n_2) \in T^n_{[U_{1,2}XWV_1V_2]} \text{ for some } (m_1, m_2) \right\} \to 0
\]
provided that (79) is satisfied.

**Lemma 3 (Generalized Markov Lemma [25]):** Consider a pmf \(p_{UXY}\) that satisfies \(Y \xrightarrow{\epsilon} X \xrightarrow{\epsilon} U\). Consider \((x^n, y^n) \in T^n_{[XY\epsilon]}\) and random vectors \(U^n\) generated according to:
\[
\Pr \left\{ U^n = u^n | x^n, y^n, U^n \in T^n_{[U|X,Y]}(x^n, y^n) \right\} = \frac{1}{\|T^n_{[U|X,Y]}(x^n, y^n)\|}.
\]
For sufficiently small \(\epsilon, \epsilon', \epsilon''\) the following holds uniformly for every \((x^n, y^n) \in T^n_{[XY\epsilon]}\):
\[
\Pr \left\{ U^n \notin T^n_{[U|X,Y]}(x^n, y^n), x^n, y^n, U^n \in T^n_{[U|X,Y]}(x^n, y^n) \right\} = O(\epsilon^{-n})
\]
where \(c > 1\).

**Corollary 3:** Assume the conditions in Lemma 2 are satisfied, and \(\Pr\left\{ (X^n, Y^n) \in T^n_{[XY\epsilon]} \right\} \to 1\), and for every \((x^n, y^n) \in T^n_{[XY\epsilon]}\), \(\Pr\left\{ (U^n, X^n, Y^n) \in T^n_{[U|XY\epsilon]}(x^n, y^n) \right\} \to 1\).

Then,
\[
\Pr\left\{ (U^n, X^n, Y^n) \in T^n_{[U|XY\epsilon]} \right\} \to 1
\]
Lemma 2 and Corollary 3 will be central for us. They will guarantee the joint typicality of the descriptions generated in different encoders considering the pmf of the chosen descriptions induced by the coding scheme used. The original proof of this result is given in [5] and involves a combination of rather sophisticated algebraic and combinatorial arguments over finite alphabets. An alternative proof was also provided in [24], which strongly relies on a rather obscure result by Uhlig [26] on combinatorics. In [25] a short and more general proof of this result is given.

Next, we present a result which will be useful for proving Theorem 1. In order to use the Markov lemma we need to show that the descriptions induced by the encoding procedure in each node satisfy the hypotheses in Lemma 2.

**Lemma 3 (Encoding induced distribution):** Consider a pmf \(p_{UXW}\) and \(\epsilon' \geq \epsilon\). Be \((U^n(m))_{m=1}^S\) random vectors independently generated according to
\[
\Pr \left\{ U^n \in T^n_{[U|W\epsilon]}(w^n) \right\} = \frac{1}{\|T^n_{[U|W\epsilon]}(w^n)\|}
\]
and where \((W^n, X^n)\) are generated with an arbitrary distribution. Once these vectors are generated, and given \(x^n\) and \(w^n\), we choose one of them if \((u^n(m), w^n, x^n) \in T^n_{[U|WX\epsilon]}\) for some \(m \in [1 : S]\). If there are various vectors \(u^n\) that satisfies this we choose the one with smallest index. If there are none we choose an arbitrary one. Let \(M\) denote the index chosen.

Then we have that:
\[
\Pr \left\{ U^n(M) = u^n | x^n, w^n, U^n(M) \in T^n_{[U|WX\epsilon]}(x^n, w^n) \right\} = \frac{1}{\|T^n_{[U|WX\epsilon]}(x^n, w^n)\|}.
\]

**Lemma 4 (Reconstruction functions for degraded RVs [11]):** Consider random variables \((X, Y, Z)\) such that \(X \xrightarrow{\epsilon} Y \xrightarrow{\epsilon} Z\), an arbitrary function \(\tilde{X} = f(Y, Z)\), and an arbitrary positive distortion function \(d(\cdot, \cdot)\). Then, \(E[d(X, g^*(Y))] \leq E[d(X, f(Y, Z))]\).

Finally, we present two lemmas, about Markov chains induced by the interactive encoding scheme used in this paper, which will be relevant for our converse results.

**Lemma 5 (Interactive encoding Markov chains I):** Consider a set of three sources \((X^n, Y^n, Z^n) \sim \prod_{t=1}^n p_{X,Y,Z}(x_t, y_t, z_t)\) and integer \(K \in \mathbb{N}\). For each \(l \in [1 : K]\) consider arbitrary message sets \(T_l^x, T_l^y\) and arbitrary functions
\[
\mathcal{F}_l^x = f_l^x \left( X^n, \mathcal{J}_l^{[1:l-1]}, \mathcal{J}_l^{[1:l]} \right), \quad \mathcal{F}_l^y = f_l^y \left( Y^n, \mathcal{J}_l^{[1:l]}, \mathcal{J}_l^{[1:l-1]} \right),
\]
with \(\mathcal{J}_l^x \in T_l^x\) and \(\mathcal{J}_l^y \in T_l^y\). The following Markov chain relations are valid for each \(i \in [1 : n]\) and \(l \in [1 : K]\):
1. \((\mathcal{J}_l^x, X_{[1:i-1]}, Y_{[t+1:n]}) \xrightarrow{\epsilon} X_t \xrightarrow{\epsilon} Y_t \xrightarrow{\epsilon} Z_t\).
2. \((\mathcal{J}_l^y, X_{[t+1:n]}) \xrightarrow{\epsilon} \left( \mathcal{J}_l^{[1:i-1]}, \mathcal{J}_l^{[1:l]}, X_{[1:i]}, Y_{[t+1:n]} \right) \xrightarrow{\epsilon} (Y_t, Z_t)\).
3. \((\mathcal{J}_l^x, Y_{[1:i-1]}) \xrightarrow{\epsilon} \left( \mathcal{J}_l^{[1:i]}, \mathcal{J}_l^{[1:l-1]}, X_{[1:i-1]}, Y_{[t+1:n]} \right) \xrightarrow{\epsilon} (X_t, Z_t)\).
4. \((\mathcal{J}_l^{[1:i]}, \mathcal{J}_l^{[1:l]}, X_{[1:i]}, Z_{[1:i-1]}, Y_{[1:i-1]}) \xrightarrow{\epsilon} (X_t, Y_t) \xrightarrow{\epsilon} Z_t\).

**Corollary 4:** Consider the setting in Lemma 5 with the following modifications:
- \(X \xrightarrow{\epsilon} Z \xrightarrow{\epsilon} Y\).
- \(f_l^x \left( X^n, Z^n, \mathcal{J}_l^{[1:l-1]}, \mathcal{J}_l^{[1:l]} \right) = \mathcal{J}_l^x\).

The following are true:
1. \((\mathcal{J}_l^x, Z_{[1:i-1]}, Y_{[t+1:n]}) \xrightarrow{\epsilon} Z_t \xrightarrow{\epsilon} Y_t\).
2. \((\mathcal{J}_l^y, Y_{[t+1:n]}) \xrightarrow{\epsilon} \left( \mathcal{J}_l^{[1:i]}, \mathcal{J}_l^{[1:l-1]}, Z_{[1:i]}, Y_{[t+1:n]} \right) \xrightarrow{\epsilon} (Y_t)\).
3. \((\mathcal{J}_l^x, Y_{[1:i-1]}) \xrightarrow{\epsilon} \left( \mathcal{J}_l^{[1:i]}, \mathcal{J}_l^{[1:l]}, Z_{[1:i-1]}, Y_{[t+1:n]} \right) \xrightarrow{\epsilon} (X_t, Z_t)\).
4. \((Z_{[t+1:n]}, X^n) \xrightarrow{\epsilon} \left( \mathcal{J}_l^{[1:i]}, \mathcal{J}_l^{[1:l]}, Z_{[1:i]}, Y_{[t+1:n]} \right) \xrightarrow{\epsilon} Y_{[1:]}, Z_t)\).
Fig. 9. Cooperative Berger-Tung problem.

Lemma 6 (Interactive encoding Markov chains II): Consider a set of three sources \( (X^n, Y^n, Z^n) \sim \prod_{t=1}^n P_{X^n Y^n Z^n} (x_t, y_t, z_t) \) and integer \( K \in \mathbb{N} \). For each \( l \in [1 : K] \) consider arbitrary message sets \( T_x^l, T_y^l, T_z^l \) and arbitrary functions

\[
\begin{align*}
J_x^l &= f_x^l \left( x^n, J_y^{[1:l-1]}, J_z^{[1:l-1]} \right), \\
J_y^l &= f_y^l \left( y^n, J_x^{[1:l]}, J_z^{[1:l]} \right), \\
J_z^l &= f_z^l \left( z^n, J_x^{[1:l]}, J_y^{[1:l]} \right)
\end{align*}
\]

with \( J_x^l \in T_x^l, J_y^l \in T_y^l \) and \( J_z^l \in T_z^l \). The following Markov chain relations are valid for each \( t \in [1 : n] \) and \( l \in [1 : K] \):

1. \( (J_x^l, J_y^l, X_{[t-1]}, Y_{[t+1:n]}, Y_{[t+1:n]}, Z_{[t-1]}) \sim (X_{[t]}, Y_{[t]}) \)
2. \( Z_{[t]} \sim (J_y^{[1:t-1]}, J_z^{[1:t-1]}, X^n, Y_{[t:n]}, Z_{[t-1]}) \)
3. \( Y_{[t]} \sim (J_x^{[1:t]}, J_y^{[1:t]}, J_z^{[1:t]}, X^n, Y_{[t+1:n]}, Z_{[1:t]}) \)
4. \( Z_{[t+1:n]} \sim (J_x^{[1:t+1:n]}, J_y^{[1:t+1:n]}, J_z^{[1:t+1:n]}, X^n, Y_{[t+1:n]}, Z_{[t+1:n]}) \)
5. \( Y_{[1:t-1]} \sim (J_x^{[1:k]}, J_y^{[1:k]}, J_z^{[1:k]}, X^n, Y_{[t+1:n]}, Z_{[1:t-1]}) \)

Definition 3 (Cooperative Code): A code \( (n, f_1^n, f_2^n, g^n, M_1, M_2) \) for the setup in Fig. 9 is given by:

- Two sets of indices \( M_1, M_2 \).
- An encoding function \( f_1^n : X^n \times Y^n \rightarrow M_1 \), such that \( f_1^n(x^n, v^n_1) = m_1 \).
- An encoding function \( f_2^n : X^n \times Y^n \rightarrow M_2 \), such that \( f_2^n(x^n, v^n_2, m_1) = m_2 \).
- A decoding function \( g^n : X^n \times Y^n \rightarrow M_1 \times M_2 \rightarrow \mathcal{U}_1 \times \mathcal{U}_2 \), such that \( g^n(x^n, v^n_1, v^n_2, m_1, m_2) = (u^n_1, u^n_2) \).

Definition 4 (Achievable Rates): We say that \( (R_1, R_2) \) are \( \epsilon \)-achievable if there exists a code \( (n, f_1^n, f_2^n, g^n, M_1, M_2) \) such that:

\[
\frac{1}{n} \log |M_1| \leq R_1 + \epsilon, \quad \frac{1}{n} \log |M_2| \leq R_2 + \epsilon
\]

and

\[
\Pr \left( (\hat{U}_1^n, \hat{U}_2^n, V_1^n, V_2^n, X_1^n, X_2^n, X_3^n) \in \mathcal{T}_{[X_1^n, X_2^n, X_3^n]} \right) \xrightarrow{n \to \infty} 1.
\]

The closure of the set of all achievable rates \( (R_1, R_2) \) is denoted by \( \mathcal{R}_{CBT} \).

The following theorem presents an inner bound for \( \mathcal{R}_{CBT} \).

Theorem 8 (Inner Bound for the Cooperative BT Problem): Consider \( \mathcal{R}_{CBT}^{inner} \) to be the closure of the set of rates satisfying:

\[
R_1 > I(X_1; U_1|X_2 V_1),
R_2 > I(X_2; U_2|X_1 V_2 U_1),
R_1 + R_2 > I(X_1; U_1 U_2|X_3 V_2),
\]

for a pmf \( p_{X_1 X_2 X_3 U_1 U_2 V_1} \) that verifies the Markov chains in (81). Then \( \mathcal{R}_{CBT}^{inner} \subset \mathcal{R}_{CBT} \).

Remark 16: Notice that we are not asking for \((X_1^n, X_2^n, X_3^n, V_1^n, V_2^n)\) to be independently and identically distributed. This is in fact not needed for the result that follows. For us, when trying to use this result, the case of most interest will be when \((X_1^n, X_2^n, X_3^n)\) is generated using the product measure \( \prod_{t=1}^n p_{X_1 X_2 X_3}(x_{1:t}, x_{2:t}, x_{3:t}) \), (that is,
when \((X_1X_2X_3)\) is a DMS). However, \((V^n_1, V^n_2)\) will not be identically and independently distributed. Still, (82) will be satisfied.

**Remark 17:** Notice that unlike the classical rate-distortion problem we are not interested in an average per-symbol distortion constraints at the decoder. We only require that the obtained sequences be jointly typical with the sources. Clearly the problem can be slightly modified to consider the case in which reconstruction distortion constraints are of interest. In fact, case (C) reported in [27], considers a similar setting. Here, given the importance of this result for our interactive scheme, we present a slightly different and more direct proof of the inner bound. We will discuss the key points in the encoding and decoding procedures which will be relevant for our extension to the interactive problem.

**Proof:** Our proof uses standard ideas from multi-terminal source coding. As \(V^n_1\) is common to both encoders and decoder we can set without loss of generality \(V^n_1 = \emptyset\). We can take into account the situation in which \(V^n_1 \neq \emptyset\) conditioning with respect to \(V_1\) the mutual information terms obtained in the \(V^n_1 = \emptyset\) case.

**A. Coding Generation**

We randomly generate \(2^{n \hat{R}_1}\) codewords \(U^n_1(k), k \in [1 : 2^{\hat{R}_1}]\) according to

\[
U^n_1(k) \sim \frac{1}{\left| T^n_{U_1| k, \epsilon_d} \right|}, \quad \epsilon_{cd} > 0. \tag{84}
\]

These \(2^{n \hat{R}_1}\) codewords are distributed uniformly over \(2^{n \hat{R}_1}\) bins denoted by \(B_1(m_1)\), where \(m_1 \in [1 : 2^{n \hat{R}_1}]\). For each codeword \(u^n_1(k)\) with \(k \in [1 : 2^{n \hat{R}_1}]\), we randomly generate \(2^{n \hat{R}_2}\) codewords according to:

\[
U^n_2(l, k) \sim \frac{1}{\left| T^n_{U_2| U_1, \epsilon_d} (u^n_1(k)) \right|}, \quad \epsilon_{cd} > 0 \quad \tag{85}
\]

with \(l \in [1 : 2^{n \hat{R}_2}]\). The \(2^{(n \hat{R}_1+\hat{R}_2)}\) resulting codewords are distributed uniformly in \(2^{n \hat{R}_2}\) bins, denoted by \(B_2(m_2)\), \(m_2 \in [1 : 2^{n \hat{R}_2}]\). It is worth to mention that the codewords \(U^n_2(l, k)\) are not distributed in a different structure of bins for each \(k\), but on only one super-bin structure where the size of each bin is approximately given by \(2^{n(\hat{R}_1+\hat{R}_2)}/2^{n \hat{R}_2}\) and where \(B_2(m_2)\) is not indexed with \(k\). As will be clear, this will not constrain the decoder to use **successive decoding**, using instead **joint decoding** in order to recover the desired codewords \((U^n_1, U^n_2)\).

After the generation of codewords and bin indexing is finished, the codebooks are revealed to all parties.

**B. Encoding at Node 1**

Given \(x^n_1\), the encoder search for \(k \in [1 : 2^{n \hat{R}_1}]\) in such a way that:

\[
(x^n_1, u^n_1(k)) \in T^n_{x_1| U_1, \epsilon_d}, \quad \epsilon_2 > 0. \tag{86}
\]

If more than one index satisfies this condition, the one with the smallest index is chosen. Otherwise, if no such index exists, an arbitrary one is selected and an error is declared. Finally, the encoder selects \(m_1\) as the index of the bin which contains the codeword \(u^n_1(k)\) and transmit it to nodes 2 and 3.

**C. Decoding at Node 2**

Given \(x^n_2\) and \(m_1\), the decoder looks for an index \(k \in [1 : 2^{n \hat{R}_1}]\) in the bin \(B_1(m_1)\) such that:

\[
(x^n_2, u^n_1(k)) \in T^n_{X_2| U_1, \epsilon_3}, \quad \epsilon_3 > 0. \quad \tag{87}
\]

If there only one index that satisfies this we declare it as the index generated at node 1. If there several or none we choose a predefined one and declare an error. The chosen index is denoted as \(\hat{k}(2)\).

**D. Encoding at Node 2**

Given \(x^n_2\) and \(\hat{k}(2)\) we search for \(l \in [1 : 2^{n \hat{R}_2}]\) such that:

\[
(x^n_2, u^n_1(\hat{k}(2)), u^n_2(l, \hat{k}(2))) \in T^n_{x_2| U_2, \epsilon_4}, \quad \epsilon_4 > 0. \quad \tag{88}
\]

If more than one index satisfies this condition, then we choose the one with the smallest index. Otherwise, if no such index exists, we choose an arbitrary one and declare an error. Finally we select \(m_2\) as the index of the bin which contains the codeword \(u^n_2(l, \hat{k}(2))\) selected and transmit it to node 3.

**E. Decoding at Node 3**

Given \(x^n_3, v^n_2\) and \(m_1, m_2\), the decoder search in the bins \(B_1(m_1)\) and \(B_2(m_2)\) for a pair of indices \((k, l) \in [1 : 2^{n \hat{R}_1}] \times [1 : 2^{n \hat{R}_2}]\) such that

\[
(x^n_3, u^n_1(k), u^n_2(l, k)) \in T^n_{X_3, U_2, U_1, \epsilon}, \quad \epsilon > 0. \quad \tag{89}
\]

If there are only one pair of indices that satisfy this we declare them as the indices generated at node 1 and 2. If there several or none pairs we choose a predefined one and declare an error. The chosen pair is denoted by \((\hat{k}(3), \hat{l}(3))\). Finally, the decoder declares \((\hat{u}^n_1, \hat{u}^n_2) = (u^n_1(\hat{k}(3)), u^n_2(\hat{l}(3), \hat{k}(3)))\).

**F. Error Probability Analysis**

Consider \((K, L)\) the description indices generated at node 1 and 2, and \((M_1, M_2)\) the corresponding bin indices. With \(\hat{K}(2)\) and \((\hat{K}(3), \hat{L}(3))\) we denote the indices recovered at nodes 2 and 3. We want to prove that \(\Pr[\mathcal{E}] \leq \epsilon'\) when \(n\) is sufficiently large, where

\[
\mathcal{E} = \left\{ \left( X^n_1, X^n_2, X^n_3, V^n_2, U^n_1(\hat{K}(3)), U^n_2(\hat{L}(3), \hat{K}(3)) \right) \mid \notin T^n_{X_1, X_2, X_3, V_2, U_1, U_2, \epsilon} \right\}.
\]
We consider the following events of error:

\[ \mathcal{E}_1 = \left\{ (X_1^n, X_2^n, X_3^n, V_2^n) \notin T_{X_1^nX_2^nX_3^nV_2^n}(1) \right\}, \]

\[ \mathcal{E}_2 = \left\{ (X_1^n, U_1^n(k)) \notin T_{X_1^nU_1^n}(1 : k) \quad \forall k \in [1 : 2n^R] \right\}, \]

\[ \mathcal{E}_3 = \left\{ (X_1^n, X_2^n, X_3^n, V_2^n, U_1^n(K)) \notin T_{X_1^nX_2^nX_3^nU_1^n}(1) \right\}, \]

\[ \mathcal{E}_4 = \left\{ \exists \tilde{k} \neq K, \tilde{k} \in B_1(M_1), (X_2^n, U_1^n(\tilde{k})) \in T_{X_1^nU_1^n}(1) \right\}, \]

\[ \mathcal{E}_5 = \left\{ (X_2^n, U_1^n(\tilde{k})), (X_2^n, U_2^n(l, \tilde{k})) \right\} \notin T_{X_1^nX_2^nU_1^nU_2^n}(1) \right\} \forall l \in [1 : 2n^R]. \]

\[ \mathcal{E}_6 = \left\{ \left( X_2^n, U_2^n(K), U_2^n(L) \right) \right\} \notin T_{X_1^nX_2^nX_3^nU_2^nU_1^n}(1) \right\} \right\} \forall l \in [1 : 2n^R]. \]

\[ \mathcal{E}_7 = \left\{ \exists \tilde{k} \neq K, \tilde{k} \in B_1(M_1), \hat{k} \in B_2(M_2), \right\} \]

\[ \left( X_3^n, \hat{V}_2^n, U_3^n(\tilde{k}), U_3^n(l, \hat{k}) \right) \in T_{X_3^nX_2^nU_3^nU_2^n}(1) \right\} \forall l \in [1 : 2n^R]. \]

Clearly \( \mathcal{E} \subseteq \bigcup_{i=1}^{7} \mathcal{E}_i \). In fact, it is easy to show that \( \left\{ (\hat{k}(3), \hat{l}(3)) \right\} \notin (K, L), \hat{k}(2) \notin K \right\} \subseteq \bigcup_{i=1}^{7} \mathcal{E}_i \). From the assumed hypotheses, we obtain that \( \lim_{n \to \infty} \Pr[\mathcal{E}] = 0 \). Choosing \( \epsilon_1 < \frac{\epsilon_{cd} \alpha_1}{n \log n} \) and \( \epsilon_2 < \epsilon_{cd} \) we can use the covering lemma [24] to obtain \( \lim_{n \to \infty} \Pr[\mathcal{E}_2] = 0 \) if

\[ \hat{R}_1 > I(U_1; X_1) + \delta(\epsilon_1, \epsilon_2, \epsilon_{cd}, n). \quad (90) \]

For the analysis of \( \Pr[\mathcal{E}_3] \) we can use Lemma 2, its corollary and Lemma 3 defining \( Y \triangleq X_2X_3, X \triangleq X_1 \) and \( U \triangleq U_1 \) and using \( \epsilon_2, \epsilon_3 \) and \( \epsilon_{cd} \) sufficiently small to obtain \( \lim_{n \to \infty} \Pr[\mathcal{E}_3] = 0 \). For the analysis of \( \Pr[\mathcal{E}_4] \) we can write:

\[ \Pr[\mathcal{E}_4] = \mathbb{E}_n \left[ \Pr[\mathcal{E}_4 | K = k, M_1 = m_1] \right] = \mathbb{E}_n \left[ \left| \sum \sum \left( X_2^n, U_1^n(\tilde{k}) \right) \in T_{X_1^nU_1^n}(1) | M_1 = m_1 \right| \right] \]

where the union is over \( \left\{ \tilde{k} \neq k : \tilde{k} \in B_1(m_1) \right\} \). Using Lemma 1 (with the appropriate equivalences on the involved random variables) and the statistical properties of the codebooks, binning and encoding, we have that for each \( k \) and \( m_1 \):

\[ \lim_{n \to \infty} \Pr \left\{ \left( X_2^n, U_1^n(\tilde{k}) \right) \in T_{X_1^nU_1^n}(1) \right\} \right\} \right\} \left\{ M_1 = m_1 \right\} = 0 \]

provided that:

\[ \frac{1}{n} \log \mathbb{E}[B_1(m_1)] < I(X_2 U_1) - \delta. \quad (91) \]

As \( \mathbb{E}[B_1(m_1)] = 2^{n(R_1 - R_1)} \) \( \forall m_1 \) we have that \( \lim_{n \to \infty} \Pr[\mathcal{E}_4] = 0 \) when:

\[ \hat{R}_1 - R_1 < I(X_2 U_1) - \delta. \quad (92) \]

The analysis of \( \mathcal{E}_5 \) follows the same lines of \( \mathcal{E}_2 \). The above analysis implies that:

\[ \lim_{n \to \infty} \Pr \left\{ \left( X_2^n, U_1^n(\tilde{k})(2) \right) \right\} \in T_{X_2^nU_1^n}(1) = 1. \quad (93) \]

\[ \text{lim}_{n \to \infty} \Pr \left\{ \left( X_2^n, U_1^n(\tilde{k})(2) \right) \right\} \in T_{X_2^nU_1^n}(1) = 1. \quad (93) \]

\[ \text{lim}_{n \to \infty} \Pr \left\{ \left( X_2^n, U_1^n(\tilde{k})(2) \right) \right\} \in T_{X_2^nU_1^n}(1) = 1. \quad (93) \]
Notice that equation (95) remains inactive because of (97). Equations (90), (92), (94), (96) and (97) can be combined with \( \hat{R}_1 > R_1 \) and \( R_1 + R_2 > R_2 \) which follow from the binning structure assumed for the generated codebooks. A Fourier-Motzkin elimination procedure allows to eliminate \( \hat{R}_1 \) and \( \hat{R}_2 \) obtaining the desired region (conditioning also the mutual information terms on \( V_1 \)).

The following corollary considers the case in which a genie gives node 2 the value of \( M_1 \). Indeed, this case will be important for our main result.

**Corollary 5:** If a genie gives \( M_1 \) to node 2, the achievable region \( \mathcal{R}^*_{CBT} \) reduces to:
\[
R_2 > I(X_2; U_2|X_1V_1V_2U_1), \\
R_1 + R_2 > I(X_1X_2; U_1U_2|X_1V_1V_2).
\]
The proof of this result is straightforward and it will not be presented.

**APPENDIX C**

**PROOF OF THEOREM 1**

Let us describe the coding generation, encoding and decoding procedures. We will consider the following notation. With \( M_i \rightarrow S_f \) we will denote the index corresponding to the true description \( U^n_{i \rightarrow S_f} \) generated at node \( i \) at round \( l \) and destined to the group of nodes \( S \in \mathcal{C}(M) \) with \( i \not\in S \). With \( M_i \rightarrow S_f(j) \) where \( S \in \mathcal{C}(M)_i \), \( i \not\in S \), \( j \in S \) we denote the corresponding estimated index at node \( j \).

**A. Codebook Generation**

Consider the round \( l \) in \([1 : K]\). For simplicity let us consider the descriptions at node 1. We generate \( 2^nR_1^{(l)} \) i.i.d. \( n \)-length codewords \( U^n_{1 \rightarrow 23,l}(m_{1 \rightarrow 23,l}, m_{\hat{V}_1[l],1}) \) according to:
\[
U^n_{1 \rightarrow 23,l}(m_{1 \rightarrow 23,l}, m_{\hat{V}_1[l],1}) \sim \mathbb{I}\left\{ u^n_{1 \rightarrow 23,l} \in T^n_{U_{1 \rightarrow 23,l}, V_{l}[1:2]}|E_1(1,23) \right\} u^n_{1[l],1},
\]
where \( E_1(1,23,l) > 0 \), \( m_{1 \rightarrow 23,l} \in [1 : 2^nR_1^{(l)} - 23] \) and with \( m_{\hat{V}_1[l],1} \) denoting the indices of the common descriptions generated in rounds \( t \in [1 : l - 1] \). For example, \( m_{\hat{V}_1[l],1} = \{ m_{1 \rightarrow 23,1}, m_{2 \rightarrow 13,2}, m_{3 \rightarrow 12,3} \}_{j=1}^{l-1} \). With \( w^n_{1[l],1} \) we denote the set of \( n \)-length common information code words from previous rounds corresponding to the indices \( m_{\hat{V}_1[l],1} \). For each \( m_{\hat{V}_1[l],1} \) consider the set of \( 2^nR_1^{(l)} \) codewords \( U^n_{1 \rightarrow 23,l}(m_{1 \rightarrow 23,l}, m_{3 \rightarrow 12,1 - 1}, m_{\hat{V}_1[l],1}) \). These \( n \)-length codewords are distributed independently and uniformly over \( 2^nR_1^{(l)} \) bins denoted by \( B_{1 \rightarrow 23,l}(p_{1 \rightarrow 23,l}, m_{\hat{V}_1[l],1}) \) with \( p_{1 \rightarrow 23,l} \in [1 : 2^nR_1^{(l)} - 23] \). Notice that this binning structure is exactly the same we have used for the cooperative Berger-Tung problem in Appendix B. Node 1 distributes codewords \( U^n_{1 \rightarrow 23,l}(m_{1 \rightarrow 23,l}, m_{3 \rightarrow 12,l - 1}, m_{\hat{V}_1[l],1}) \) in a super-binning structure. This will allow node 2 to recover both, \( m_{1 \rightarrow 23,l} \) and \( m_{3 \rightarrow 12,l - 1} \), using the same procedure as in the Berger-Tung problem described above. Notice that a different super-binning structure is generated for every \( m_{\hat{V}_1[l],1} \). This is without loss of generality, because at round \( l \) nodes 1, 2 and 3, will have a very good estimate of it (see below).

We also generate \( 2^nR_1^{(l)} - 2 \) and \( 2^nR_1^{(l)} - 3 \) independent and identically distributed \( n \)-length codewords \( U^n_{1 \rightarrow 2,j}(m_{1 \rightarrow 2,l}, m_{\hat{V}_2[l],j}, m_{\hat{V}_1[l],1}) \), and \( U^n_{1 \rightarrow 3,j}(m_{1 \rightarrow 3,l}, m_{\hat{V}_2[l],j}, m_{\hat{V}_1[l],1}) \) according to:
\[
U^n_{1 \rightarrow 2,j}(m_{1 \rightarrow 2,l}, m_{\hat{V}_2[l],j}, m_{\hat{V}_1[l],1}) \sim \mathbb{I}\left\{ u^n_{1 \rightarrow 2,l} \in T^n_{U_{1 \rightarrow 2,l}, V_{l}[1:2]}|E_1(1,2) \right\} u^n_{1[l],1},
\]
where \( \epsilon(l, 2, l) > 0 \), \( \epsilon(l, 3, l) > 0 \), and \( m_{1 \rightarrow 2,l} \in [1 : 2^nR_1^{(l)} - 2] \) and \( m_{1 \rightarrow 3,l} \in [1 : 2^nR_1^{(l)} - 3] \). These codewords are distributed uniformly on \( 2^nR_1^{(l)} \) bins denoted by \( B_{1 \rightarrow 2,j}(p_{1 \rightarrow 2,l}, m_{\hat{V}_2[l],j}, m_{\hat{V}_1[l],1}) \) and indexed with \( p_{1 \rightarrow 2,l} \in [1 : 2^nR_1^{(l)} - 2] \) and on \( 2^nR_1^{(l)} \) bins denoted by \( B_{1 \rightarrow 3,j}(p_{1 \rightarrow 3,l}, m_{\hat{V}_2[l],j}, m_{\hat{V}_1[l],1}) \) and indexed with \( p_{1 \rightarrow 3,l} \in [1 : 2^nR_1^{(l)} - 1] \), respectively. Notice that these codewords (which will be used to generate private descriptions to node 2 and 3) are not distributed in a super-binning structure. This is because there is not explicit cooperation between the nodes at this level. That is, node 2 is not compelled to recover the private description that node 1 generate for node 3, and for that reason the private description that node 2 generate for node 3 is not superimposed over the former. Notice that the binning structure used for the codewords to be utilized by node 1 impose the following relationships:
\[
R_1^{(l)} < R_2^{(l)} + R_3^{(l)} + R_3^{(l-1)} - 2,
\]
\[
R_1^{(l)} < R_2^{(l)} + R_3^{(l)} - 3,
\]
The common and private codewords to be utilized by nodes 2 and 3, for every round, are generated by following a similar procedure and theirs corresponding rates have analogous relationships. Finally, the generated codebooks are revealed to all the nodes in the network.

**B. Encoding Technique**

Consider node 1 at round \( l \) in \([1 : K]\). Upon observing \( x^n_1 \) and given all of its encoding and decoding history up to round \( l \), encoder 1 first looks for a codeword \( u^n_{1 \rightarrow 2,j}(m_{1 \rightarrow 2,l}, m_{\hat{V}_2[l],j}(1)) \) such that \( \epsilon(l, 2, l) > 0 \),
\[
(x^n_1, u^n_{1[l],j}(m_{\hat{V}_2[l],j}(1))), u^n_{1 \rightarrow 2,j}(m_{1 \rightarrow 2,l}, m_{\hat{V}_2[l],j}(1)) \in T^n_{U_{1 \rightarrow 2,j}, X_1}(1,23,l).
\]
Notice that some components of \( m_{\hat{V}_2[l],j}(1) \) are generated at node 1 and are perfectly known. If more than one codeword satisfies this condition, then we choose the one with the smallest index. Otherwise, if no such codeword exists, an
arbitrary index is chosen and an error is declared. With the chosen index $m_{1 \rightarrow 23,l}$, and with $\tilde{m}_{V[3,1,j]}(1)$, encoder 1 finds the index $p_{1 \rightarrow 23,l}$ of the bin $B_{1 \rightarrow 23,l}(p_{1 \rightarrow 23,l}, \tilde{m}_{V[3,1,j]}(1))$ to which $u_{1 \rightarrow 23,l}^{n}(m_{1 \rightarrow 23, l}, m_{3 \rightarrow 12, l-1}, \tilde{m}_{V[3,1,j]}(1))$ belongs. After this, Encoder 1 generates the private descriptions looking for codewords $u_{1 \rightarrow 23,l}^{n}(m_{1 \rightarrow 23, l}, m_{3 \rightarrow 12, l-1}, \tilde{m}_{V[3,1,j]}(1))$, $u_{1 \rightarrow 3, l}^{n}(m_{1 \rightarrow 3, l}, \tilde{m}_{V[3,1,j]}(1), \tilde{m}_{V[13,1,j]}(1))$ such that

$$
\left( x_{1}^{n}, w_{0,1,j}^{n}, \tilde{m}_{V[3,1,j]}(1) \right), \left( m_{1 \rightarrow 23,l}, m_{3 \rightarrow 12, l-1}, \tilde{m}_{V[3,1,j]}(1) \right) \in T_{[1]}^{n}(x_{1}, w_{0,1,j}, \tilde{m}_{V[3,1,j]}(1))
$$

respectively, where $\epsilon_{1}(1,2,l) > 0$ and $\epsilon_{1}(1,3,l) > 0$. Given $(\tilde{m}_{V[3,1,j]}(1), \tilde{m}_{V[13,1,j]}(1))$, the encoding procedure continues by determining the bin indices $p_{1 \rightarrow 23,l}$ and $p_{1 \rightarrow 3, l}$ to which the generated private descriptions belong to. Node 1 then transmits to node 2 and 3 the indices $(p_{1 \rightarrow 23,l}, p_{1 \rightarrow 2,l}, p_{1 \rightarrow 3, l})$. The encoding in nodes 2 and 3 follows along the same lines and for that reason are not described.

C. Decoding Technique

Consider round $l \in [1 : K + 1]$ and node 2. During round $l$, node 2 can recover the private information indices by looking for codewords $u_{1 \rightarrow 23,l}^{n}(m_{1 \rightarrow 23, l}, m_{3 \rightarrow 12, l-1}, \tilde{m}_{V[3,1,j]}(2))$, $u_{1 \rightarrow 3, l}^{n}(m_{1 \rightarrow 3, l}, \tilde{m}_{V[3,1,j]}(2), \tilde{m}_{V[13,1,j]}(2))$ that satisfies:

$$
\left( x_{2}^{n}, w_{3,1-l}^{n}, \tilde{m}_{V[3,1,j-1]}(2) \right), \left( m_{1 \rightarrow 23, l}, m_{3 \rightarrow 12, l-1}, \tilde{m}_{V[3,1,j]}(2) \right), \left( m_{1 \rightarrow 23, l}, m_{3 \rightarrow 12, l-1}, \tilde{m}_{V[3,1,j]}(2) \right)
$$

with $\epsilon_{1}(2,l) > 0$ and that also belongs to the bins indicated by $p_{1 \rightarrow 23,l}$ and $p_{1 \rightarrow 23,l}$-1. If there are more than one pair of codewords, or none that satisfies this, we choose a predefined one and declare an error. After this is done, node 2 can recover the private information indices by looking for codewords $u_{1 \rightarrow 23,l}^{n}(m_{1 \rightarrow 23, l}, m_{3 \rightarrow 12, l-1}, \tilde{m}_{V[3,1,j]}(2))$, $u_{3 \rightarrow 2,l-1}^{n}(m_{3 \rightarrow 2,l-1}, \tilde{m}_{V[3,1,j]}(2))$ which satisfy

$$
\left( x_{2}^{n}, w_{12,1-l}^{n}, \tilde{m}_{V[12,1,j]}(2), \tilde{m}_{V[13,1,j]}(2) \right)
$$

with $\epsilon_{1}(2,l) > 0$ and are in the bins given by $p_{1 \rightarrow 2,l}$ and $p_{3 \rightarrow 2,l-1}$. If there are more than one pair of codewords, or none that satisfies this, we choose a predefined one and declare an error. The decoding in nodes 1 and 3 is exactly the same and for that reason are not described.

D. Lossy Reconstructions

When the exchange of information is completed, each node needs to estimate the other nodes sources. For instance, node 1 reconstruct the source of node 2 by computing:

$$
\hat{x}_{12, l} = g_{12} \left( x_{1, l}, w_{12, l+1,3}, w_{1, l+1,3} \right)
$$

and similarly, for the source of node 3:

$$
\hat{x}_{13, l} = g_{13} \left( x_{1, l}, w_{13, l+1,3}, w_{1, l+1,3} \right)
$$

Reconstruction at nodes 2 and 3 is done in a similar way using the adequate reconstruction functions.

E. Error and Distortion Analysis

In order to maintain expressions simple, in the following when we denote a description without the corresponding index, i.e. $U_{i \rightarrow S,l}^{n}$ or $W_{i,j}^{n}$, we will assume that the corresponding index is the true one generated in the corresponding nodes through the detailed encoding procedure. Consider round $l$ and the event $D_{l} = G_{l} \cap F_{l}$, where for $\epsilon_{1} > 0$

$$
G_{l} = \left\{ \left( x_{1}^{n}, x_{2}^{n}, x_{3}^{n}, W_{1,j}^{n}, W_{12,1,j}^{n}, W_{13,1,j}^{n}, W_{23,1,j}^{n} \right) \in T_{[1]}^{n}(x_{1}, x_{2}, x_{3}, W_{1,j}, W_{12,1,j}, W_{13,1,j}, W_{23,1,j}) \cap \epsilon_{1} \left( 2, l \right) \right\}
$$

The set $G_{l}$ indicates that all the descriptions generated in the network, up to round $l$, are jointly typical with the sources. The occurrence of this depends mainly on the encoding procedure in the nodes. Set $F_{l}$ indicates, that up to round $l$, all nodes were able to recover the true indices of the descriptions. This clearly implies that there were not errors at the decoding procedures in all the nodes in the network. The condition in $F_{l}$ on $M_{3 \rightarrow 12, l-1,2}$, $M_{3 \rightarrow 2,l-1}$ is due to the fact, that the decoding of those descriptions in node 2 occurs during round $l$. The occurrence of $D_{l}$ guarantees that at the beginning of round 1:

- Node 1 and 2 share a common path of descriptions $W_{i,j}^{n} \cup W_{12,1,j}^{n}$ which are typical with $(X_{1}^{n}, X_{2}^{n}, X_{3}^{n})$.
- Node 1 and 3 share a common path of descriptions $W_{i,j}^{n} \cup W_{13,1,j}^{n}$ which are typical with $(X_{1}^{n}, X_{2}^{n}, X_{3}^{n})$. 
• Node 2 and 3 share a common path of descriptions $\mathcal{W}^n_{[3,j-1]} \cup \mathcal{W}^n_{[23,j-1,3]}$ which are typical with $(X^n_1, X^n_2, X^n_3)$.

Let us also define the event $\mathcal{E}_i$:

$$
\mathcal{E}_i = \left\{ \text{there exists at least an error at the encoding or decoding in a node during round } i \right\}
$$

$$
= \bigcup_{i \in \mathcal{M}} \mathcal{E}_{\text{enc}}(i, l) \cup \mathcal{E}_{\text{dec}}(i, l)
$$

where $\mathcal{E}_{\text{enc}}(i, l)$ contains the errors at the encoding in node $i$ during round $l$ and $\mathcal{E}_{\text{dec}}(i, l)$ denotes the event that at node $i$ during round $l$ there is a failure to recovering an index generated previously in other node. For example, at node 1 and during round 1:

$$
\mathcal{E}_{\text{enc}}(1, l, 23) = \mathcal{E}_{\text{enc}}(1, l, 23) \cup \mathcal{E}_{\text{enc}}(1, l, 2) \cup \mathcal{E}_{\text{enc}}(1, l, 3)
$$

$$
= \left\{ (X^n_1, \mathcal{W}^n_{[1,j]}(\hat{M}W_{[1,j]}(1))), U^n_{1 \rightarrow 23, f}(m_{1 \rightarrow 23,j}, \hat{M}V_{[1,j]}(1))) \notin T^n_{U_{1 \rightarrow 23,j}X_{1}[1,j]} \right\}
$$

$$
\forall m_{1 \rightarrow 23, j} \in [1 : 2^nR_{1,23}^{(1)}],
$$

$$
\mathcal{E}_{\text{dec}}(1, l, 23) = \left\{ (X^n_1, \mathcal{W}^n_{[2,j]}(\hat{M}W_{[2,j]}(1))), \mathcal{V}^n_{[12,j,1]}(\hat{M}V_{[12,j,1]}(1)), U^n_{1 \rightarrow 2,j}(m_{1 \rightarrow 2,j}, \hat{M}V_{[12,j,1]}(1))) \notin T^n_{U_{1 \rightarrow 2,j}X_{1}[12,j,1]} \right\}
$$

$$
\forall m_{1 \rightarrow 2, j} \in [1 : 2^nR_{1,2}^{(2)}],
$$

$$
\mathcal{E}_{\text{dec}}(1, l, 2) = \left\{ (X^n_1, \mathcal{W}^n_{[3,j]}(\hat{M}W_{[3,j]}(1))), \mathcal{V}^n_{[13,j,1]}(\hat{M}V_{[13,j,1]}(1)), U^n_{1 \rightarrow 3,j}(m_{1 \rightarrow 3,j}, \hat{M}V_{[13,j,1]}(1))) \notin T^n_{U_{1 \rightarrow 3,j}X_{1}[13,j,1]} \right\}
$$

$$
\forall m_{1 \rightarrow 3, j} \in [1 : 2^nR_{1,3}^{(3)}],
$$

Event $\mathcal{E}_{\text{dec}}(i, l)$ can be decomposed as:

$$
\mathcal{E}_{\text{dec}}(i, l) = \bigcup_{S \in C(M)} \bigcup_{i \in S} \left\{ \hat{M}_{j \rightarrow S,l}(i) \neq M_{j \rightarrow S,l} \right\}
$$

At the end of the information exchange phase we would expect the occurrence of $\mathcal{D}_{K+1} \cap \hat{\mathcal{E}}_{K+1}$, where $\hat{\mathcal{E}}_{K+1}$ is the event of an error during round $K + 1$. As during round $K + 1$ only node 2 tries to recover the descriptions generated during round $K$ in node 3, we have:

$$
\mathcal{E}_{\text{dec}}(2, K + 1) = \left\{ \hat{M}_{3 \rightarrow 12, K}(2) \neq M_{3 \rightarrow 12, K} \text{ or } \hat{M}_{3 \rightarrow 2, K}(2) \neq M_{3 \rightarrow 2, K} \right\}
$$

For simplicity this last event is relabeled as $\mathcal{E}_{K+1}$. The occurrence of $\mathcal{D}_{K+1} \cap \hat{\mathcal{E}}_{K+1}$ guarantees that all the descriptions generated during the $K$ rounds of information exchange in the network are jointly typical with the sources realizations and that those descriptions can be perfectly recovered in all the nodes. If we can guarantee that $\Pr \{ \mathcal{D}_{K+1} \cap \hat{\mathcal{E}}_{K+1} \} \rightarrow \infty$, then with probability converging to one we obtain:

• Node 1 and 2 share a common path of descriptions $\mathcal{W}^n_{[1,K+1]} \cup \mathcal{W}^n_{[12,K+1]}$ which are typical with $(X^n_1, X^n_2, X^n_3)$.

• Node 1 and 3 share a common path of descriptions $\mathcal{W}^n_{[1,K+1]} \cup \mathcal{W}^n_{[13,K+1]}$ which are typical with $(X^n_1, X^n_2, X^n_3)$.

• Node 2 and 3 share a common path of descriptions $\mathcal{W}^n_{[2,K+1]} \cup \mathcal{W}^n_{[23,K+1]}$ which are typical with $(X^n_1, X^n_2, X^n_3)$.

Using standard analysis ideas, the average distortions (over the codebooks) at the reconstruction stages in all the nodes satisfy the required fidelity constraints. From there is straightforward to prove the existence of good codebooks for the network.

In order to prove that $\Pr \{ \mathcal{D}_{K+1} \cap \hat{\mathcal{E}}_{K+1} \} \rightarrow \infty$ let us write:

$$
\Pr \left\{ \mathcal{D}_{K+1} \cap \hat{\mathcal{E}}_{K+1} \right\} = \Pr \left\{ \mathcal{D}_{K+1} \cup \hat{\mathcal{E}}_{K+1} \right\}
$$

$$
= \Pr \{ \mathcal{D}_{K+1} \} + \Pr \{ \mathcal{D}_{K+1} \cap \mathcal{E}_{K+1} \}
$$

$$
\leq \Pr \{ \mathcal{D}_{K+1} \cap \mathcal{D}_{K} \} + \Pr \{ \hat{\mathcal{E}}_{K+1} \}
$$

$$
+ \Pr \{ \mathcal{D}_{K+1} \cap \mathcal{D}_{K} \cap \hat{\mathcal{E}}_{K+1} \}
$$

$$
\leq \Pr \{ \hat{\mathcal{D}}_{K+1} \} + \sum_{l=1}^{K} \Pr \{ \mathcal{D}_{l} \cap \mathcal{E}_{l} \}
$$

$$
+ \sum_{l=1}^{K} \Pr \{ \mathcal{D}_{l+1} \cap (\mathcal{D}_{l} \cap \hat{\mathcal{E}}_{l}) \}.
$$

Notice that

$$
\mathcal{D}_{l} = \left\{ (X^n_1, X^n_2, X^n_3) \in T^n_{X_1 X_2 X_3 e} \right\}, \quad \epsilon_1 > 0.
$$

From the conditional typicality lemma [19], [24], we see that for every $\epsilon_1 > 0$, $\Pr \{ \mathcal{D}_{l} \} \rightarrow \infty$. Then, it is easy to see that

$$
\Pr \{ \mathcal{D}_{K+1} \cap \hat{\mathcal{E}}_{K+1} \} \rightarrow \infty
$$

will hold if the coding generation, the encoding and decoding procedures described above allow us to have the following:

1) If $\Pr \{ \mathcal{D}_{l} \} \rightarrow \infty$ then $\Pr \{ \mathcal{D}_{l+1} \} \rightarrow \infty$ for every $l \in [1 : K + 1]$.

2) $\Pr \{ \mathcal{D}_{l} \cap \mathcal{E}_{l} \} \rightarrow \infty$ for all $l \in [1 : K + 1]$.

In the following we will prove these facts. Observe that, at round $l$ the nodes act sequentially:

Encoding at node 1 → Decoding at node 2 → ⋅⋅⋅ → Encoding at node 3 → Decoding at node 1.

Then, using (98) we can write condition 2) as:

$$
\Pr \{ \mathcal{E}_{l} \cap \mathcal{E}_{l} \}
$$

$$
= \Pr \{ \mathcal{D}_{l} \cap \mathcal{E}_{l} \}
$$

$$
+ \Pr \{ \mathcal{D}_{l} \cap \mathcal{D}_{l+1} \cap \mathcal{E}_{l+1} \}
$$

$$
+ \Pr \{ \mathcal{D}_{l} \cap \mathcal{E}_{l} \cap \mathcal{E}_{l} \cap \mathcal{E}_{l} \}
$$

$$
+ \Pr \{ \mathcal{D}_{l} \cap \mathcal{D}_{l+1} \cap \mathcal{E}_{l+1} \cap \mathcal{E}_{l+1} \}
$$

$$
+ \cdots + \Pr \{ \mathcal{D}_{l} \cap \mathcal{D}_{l+1} \cap \mathcal{E}_{l+1} \cap \mathcal{E}_{l+1} \cap \cdots \cap \mathcal{E}_{l+1} \}.
$$
Assume then that at the beginning of round $l$ we have $\Pr \{ D_l \} \underset{n \to \infty}{\longrightarrow} 1$. Let us analyze the encoding procedure at node 1. Let us consider $\Pr \{ D_l \cap \mathcal{E}_{enc}(1, l) \}$. We can write:

$$\Pr \{ D_l \cap \mathcal{E}_{enc}(1, l) \} = \Pr \{ \mathcal{E}_{enc}(1, l, 23) \cap D_l \} + \Pr \{ \mathcal{E}_{enc}(1, l, 2) \cap \mathcal{D}_l \cap \tilde{\mathcal{E}}_{enc}(1, l, 23) \} + \Pr \{ \mathcal{E}_{enc}(1, l, 3) \cap \mathcal{D}_l \cap \tilde{\mathcal{E}}_{enc}(1, l, 23) \}.$$  

From the fact that $\lim_{n \to \infty} \Pr \{ G_l \} = 1$ we have that $\lim_{n \to \infty} \Pr \{ A_l(1, 23) \} = 1$ where:

$$A_l(1, 23) = \left\{ \left( X_1^n, W_{1,l}^n \right) \in T_{X_1}^n W_{1,l}[1,|l|] \right\}.$$  

Then, we can use the covering lemma [24] to obtain:

$$\lim_{n \to \infty} \Pr \{ \mathcal{E}_{enc}(1, l, 23) \cap D_l \} = 0,$$

provided that:

$$\hat{r}_{1 \to 23}^{(l)} = I \left( X_1; U_{1 \to 23,l} \right) W_{1,l}^n + \delta_{l}(1, l, 23) \quad (100)$$

where $\delta_{l}(1, l, 23)$ can be made arbitrarily small. On the other hand we can write:

$$\Pr \{ \mathcal{E}_{enc}(1, l, 2) \cap \mathcal{D}_l \cap \tilde{\mathcal{E}}_{enc}(1, l, 23) \} \leq \Pr \{ \mathcal{E}_{enc}(1, l, 2) \cap G_l(1, 2) \cap \mathcal{F}_l \} + \Pr \{ \tilde{G}_l(1, 2) \} \quad (101)$$

where:

$$G_l(1, 2) = \left\{ \left( X_1^n, X_2^n, X_3^n, W_{1,l}^m, W_{12,l}^m, W_{13,l}^m, W_{23,l}^m, \left[ V_{13,l}^m \right] \left[ V_{23,l}^m \right] \right) \in T_{X_1 X_2 X_3}^n W_{12,l}[1,|l|], W_{13,l}[1,|l|], W_{23,l}[1,|l|] \right\}.$$  

where $\epsilon_{l}(1, 2) > 0$. As explained before, $\Pr \{ G_l \} \underset{n \to \infty}{\longrightarrow} 1$. Then, from condition (100) we have:

$$\Pr \{ \tilde{\mathcal{E}}_{enc}(1, l, 23) \cap \mathcal{F}_l \} \underset{n \to \infty}{\longrightarrow} 1.$$  

Moreover, from the coding generation and the encoding procedure proposed is immediate to use Lemma 3 to show that:

$$\Pr \{ \mathcal{U}_{1 \to 23,l} = u_{1 \to 23,l} \} \mathcal{E}_{enc}(1, l, 23) \cap \mathcal{F}_l \} \leq \Pr \{ U_{1 \to 23,l} \}$$

Then, from Markov chain

$$U_{1 \to 23,l} \leftrightarrow (X_1, W_{1,l}) \leftrightarrow (X_2, X_3, V_{12,l}, V_{13,l}, V_{23,l})$$

and from Lemma 2, for sufficiently small $(\epsilon_{c}(1, l, 23), \epsilon_{l}, \epsilon_{l}(1, 2))$ and after some minor manipulations, we can obtain $\Pr \{ G_l(1, 2) \} \underset{n \to \infty}{\longrightarrow} 1$. Looking at equation (101) it is clear that we need to analyze term $\Pr \{ \mathcal{E}_{enc}(1, l, 2) \cap G_l(1, 2) \cap \mathcal{F}_l \}$. Similarly as before $\lim_{n \to \infty} \Pr \{ A_l(1, 2) \} = 1$.

Using again the covering lemma [24] we obtain that:

$$\hat{r}_{1 \to 23}^{(l)} > I \left( X_1; U_{1 \to 23,l} \right) W_{1,l}^n V_{13,l}^m + \delta_{l}(1, l, 23) \quad (102)$$

where $\delta_{l}(1, l, 2)$ can be made arbitrarily small. For the analysis of $\Pr \{ \mathcal{E}_{enc}(1, l, 3) \cap \mathcal{D}_l \cap \tilde{\mathcal{E}}_{enc}(1, l, 23) \}$ we follow the same procedure. We can write:

$$\Pr \{ \mathcal{E}_{enc}(1, l, 3) \cap \mathcal{D}_l \cap \tilde{\mathcal{E}}_{enc}(1, l, 23) \} \leq \Pr \{ \mathcal{E}_{enc}(1, l, 3) \cap G_l(1, 3) \cap \mathcal{F}_l \} + \Pr \{ \tilde{G}_l(1, 3) \}$$

with:

$$G_l(1, 3) = \left\{ \left( X_1^n, X_2^n, X_3^n, W_{1,l}^m, W_{12,l}^m, W_{13,l}^m, W_{23,l}^m, \left[ V_{13,l}^m \right] \left[ V_{23,l}^m \right] \right) \in T_{X_1 X_2 X_3}^n W_{12,l}[1,|l|], V_{13,l}[1,|l|], V_{23,l}[1,|l|] \right\}.$$  

Using the Markov chain

$$U_{1 \to 23,l} \leftrightarrow (X_1, W_{1,l}, V_{13,l}) \leftrightarrow (X_2, X_3, V_{12,l}^m, V_{23,l}^m),$$

the fact that $\Pr \{ G_l(1, 2) \} \underset{n \to \infty}{\longrightarrow} 1$ and the Markov Lemma 2, and Lemma 3 for appropriately chosen values of $(\epsilon_{c}(1, l, 2), \epsilon_{l}(1, 2), \epsilon_{l}(1, 3))$ we have:

$$\Pr \{ G_l(1, 3) \} \underset{n \to \infty}{\longrightarrow} 1.$$  

Following exactly the same reasoning as above, we have that in order to have:

$$\Pr \{ \mathcal{E}_{enc}(1, l, 3) \cap \mathcal{D}_l \cap \tilde{\mathcal{E}}_{enc}(1, l, 23) \} \underset{n \to \infty}{\longrightarrow} 0,$$

besides conditions (100) and (102) we need:

$$\hat{r}_{1 \to 3}^{(l)} > I \left( X_1; U_{1 \to 3,l} \right) W_{1,l}^n V_{13,l}^m + \delta_{l}(1, l, 3) \quad (104)$$

for sufficiently small $\delta_l(1, l, 3)$. With these conditions we have proved that the encoding procedure in node 1 during round $l$ permit us to have:

$$\Pr \{ \mathcal{D}_l \cap \mathcal{E}_{enc}(1, l) \} \underset{n \to \infty}{\longrightarrow} 0.$$  

Another instance of Lemma 2, jointly with Markov chain

$$U_{1 \to 3,l} \leftrightarrow (X_1, W_{1,l}, V_{13,l}) \leftrightarrow (X_2, X_3, V_{12,l}^m, V_{23,l}^m)$$

and Lemma 3 allow us to have:

$$\Pr \{ G_l(2, 13) \} \underset{n \to \infty}{\longrightarrow} 1,$$

where:

$$G_l(2, 13) = \left\{ \left( X_1^n, X_2^n, X_3^n, W_{1,l}^m, W_{12,l}^m, W_{13,l}^m, W_{23,l}^m, \left[ V_{13,l}^m \right] \left[ V_{23,l}^m \right] \right) \in T_{X_1 X_2 X_3}^n W_{12,l}[1,|l|], V_{13,l}[1,|l|], V_{23,l}[1,|l|] \right\}.$$  

At this point we have to analyze the decoding in node 2. If that decoding is successful, and using (105), the analysis of the encoding at node 2 follows the same lines as above.\textsuperscript{13} The same can be said of the encoding at node 3 (after successful decoding). In this way, we terminate round $l$ with:

$$\Pr \{ D_{l+1} \} = \Pr \{ G_{l+1} \cap \mathcal{F}_{l+1} \} \underset{n \to \infty}{\longrightarrow} 1 \quad (106)$$

\textsuperscript{13}See that $G_l(2, 13)$ has, for the encoding at node 2, the same role that $G_l$ has for the encoding at node 1 during round $l$.\textsuperscript{13}
which is one the results we wanted. Clearly, analyzing now the decoding at node 2 (from which we can easily extrapolate the analysis to the decoding at node 1 and 3) we will be able to obtain $\Pr\{D_l \cap E_{\text{dec}}(2, l) \cap \tilde{E}_{\text{enc}}(1, l)\} \xrightarrow{n \to \infty} 0$ which is the other required result.

The decoding in each of nodes follows the approach of successive decoding. Decoder 2 will try to find first the common descriptions $M_{1\rightarrow 2, l}$ and $M_{3\rightarrow 2, l-1}$ Then, it will try to find the private descriptions $M_{1\rightarrow 2, l}$ and $M_{3\rightarrow 2, l-1}$ (using of course the previously obtained common descriptions as side information). Clearly, the use of joint-decoding could improve the rate-distortion region. However, the analysis of this strategy, besides of being more difficult to analyze, it will give rise to a more complex rate-distortion region. It can be easily seen, that the joint-decoding region will contain several sum-rate equations that will contains common and private rates. Sum-rate decoding allows for a rate-distortion region where the sum-rate equations contains only common or private rates, being more easy to analyze and understand. In order to analyze the decoding at node 2, we can write:

$$\Pr\{D_l \cap E_{\text{dec}}(2, l) \cap \tilde{E}_{\text{enc}}(1, l)\} \leq \Pr\{E_{\text{dec}}(2, l) \cap F_l \cap G_l(2, 13)\} + \Pr\{\tilde{G}_l(2, 13)\}.$$ 

As $\Pr\{G_l(2, 13)\} \xrightarrow{n \to \infty} 1$ we can concentrate on the first term. Event $E_{\text{dec}}(2, l)$ can be written as:

$$E_{\text{dec}}(2, l) = \mathcal{H}_{\text{common}}(2, l) \cup \mathcal{H}_{\text{private}}(2, l),$$

where

$$\mathcal{H}_{\text{common}}(2, l) = \left\{ \tilde{M}_{3\rightarrow 2, l-1} = 2, \tilde{M}_{1\rightarrow 23, l}(2) \right\} \neq \left\{ M_{3\rightarrow 2, l-1}, M_{1\rightarrow 23, l} \right\},$$

$$\mathcal{H}_{\text{private}}(2, l) = \left\{ \tilde{M}_{3\rightarrow 2, l-1} = 2, \tilde{M}_{1\rightarrow 23, l}(2) \right\} \neq \left\{ M_{3\rightarrow 2, l-1}, M_{1\rightarrow 23, l} \right\}.$$ 

From these definitions, we can easily deduce that:

$$\Pr\{E_{\text{dec}}(2, l) \cap F_l \cap G_l(2, 13)\} = \Pr\{\mathcal{H}_{\text{private}}(2, l) \cap F_l \cap G_l(2, 13)\} \leq \Pr\{\mathcal{H}_{\text{common}}(2, l)\} + \Pr\{\mathcal{H}_{\text{private}}(2, l)\},$$

$$\Pr\{E_{\text{dec}}(2, l) \cap F_l \cap G_l(2, 13)\},$$

whence

$$\mathcal{K}_{\text{common}}(2, l) = \left\{ \tilde{M}_{3\rightarrow 2, l-1} = 2, \tilde{M}_{1\rightarrow 23, l}(2) \right\} \neq \left\{ M_{3\rightarrow 2, l-1}, M_{1\rightarrow 23, l} \right\},$$

$$\mathcal{K}_{\text{private}}(2, l) = \left\{ \tilde{M}_{3\rightarrow 2, l-1} = 2, \tilde{M}_{1\rightarrow 23, l}(2) \right\} \neq \left\{ M_{3\rightarrow 2, l-1}, M_{1\rightarrow 23, l} \right\}.$$ 

In Fig. 10, we have a representation of the problem seen at decoder 2. Node 3 generates a common description at rate $R^{(l-1)}_{3\rightarrow 2, l}$ using $W^{(l)}_{3\rightarrow 1, l}$ as side information. Similarly node 1, after decoding the common description from node 3, generates its own description using the recovered one and also $W^{(l)}_{3\rightarrow 1, l}$ as side information. All these operations are done using the super-binning structure as in the cooperative Berger-Tung problem in Appendix B. Then, node 2, using $(X^{2}_2, W^{(l)}_{1\rightarrow 3, l}, W^{(l)}_{3\rightarrow 1, l})$ as side information tries to recover the descriptions generated at node 1 and 3. Remember the fact that the encoding procedure at nodes 1 and 3 requires:

$$\hat{R}^{(l-1)}_{3\rightarrow 1, l} = I(X_3; U^{(l)}_{3\rightarrow 2, l} | W^{(l)}_{1\rightarrow 3, l}, W^{(l)}_{3\rightarrow 1, l}) + \delta_3(1, l - 1, 12),$$

$$\hat{R}^{(l)}_{3\rightarrow 1, l} = I(X_1; U^{(l)}_{1\rightarrow 2, l} | W^{(l)}_{1\rightarrow 1, l}) + \delta_3(1, l, 1, 23),$$

$$\hat{R}^{(l)}_{3\rightarrow 1, l} = I(X_3; U^{(l)}_{3\rightarrow 2, l} | W^{(l)}_{1\rightarrow 3, l}, W^{(l)}_{3\rightarrow 1, l}),$$

and that the following Markov chains:

$$U^{(l)}_{3\rightarrow 1, l} \perp (X_3, W^{(l)}_{1\rightarrow 3, l}) \sim (X_1, X_2, W^{(l)}_{1\rightarrow 1, l}, W^{(l)}_{2\rightarrow 1, l}),$$

$$U^{(l)}_{1\rightarrow 2, l} \perp (X_1, W^{(l)}_{1\rightarrow 1, l}) \sim (X_1, X_2, W^{(l)}_{1\rightarrow 1, l}, W^{(l)}_{2\rightarrow 1, l}),$$

are implied by the Markov chains in the conditions of Theorem 1. In this way, we can use the results in Appendix B to show that the following rates imply

$$\Pr\{\mathcal{K}_{\text{common}}(2, l)\} \xrightarrow{n \to \infty} 0:$$

$$\hat{R}^{(l)}_{3\rightarrow 2, l} > I(X_3; U^{(l)}_{3\rightarrow 2, l} | X_2, W^{(l)}_{1\rightarrow 1, l}, W^{(l)}_{2\rightarrow 1, l}, W^{(l)}_{3\rightarrow 1, l}) + \delta_3(2, l) R^{(l)}_{3\rightarrow 2, l} + \delta_3(1, l, 12) + \delta_3(1, l, 23),$$

$$\hat{R}^{(l)}_{3\rightarrow 2, l} < I(X_1; U^{(l)}_{1\rightarrow 2, l} | W^{(l)}_{1\rightarrow 1, l}) + \delta_3(2, l) R^{(l)}_{3\rightarrow 2, l} + \delta_3(1, l, 12) + \delta_3(1, l, 23),$$

where $\delta_3(2, l), \delta_3(1, l, 23), \delta_3(1, l, 12)$ can be made arbitrarily small.\footnote{Here we considered the corollary to Theorem 8. That is we assumed, that node 1 knows perfectly the value of $M_{3\rightarrow 12, l-1}$. This follows from the assumed fact, that at the beginning of round $l$, the probability of decoding errors at previous rounds in all nodes is goes to zero when $n \to \infty$. In this way, the constraint on rate $R^{(l)}_{3\rightarrow 2, l-1}$ that should be considered, according to Theorem 8 is not needed. In fact, constraints on rate $R^{(l)}_{3\rightarrow 2, l-1}$ will arise when at node 1 we consider the recovering of $M_{2\rightarrow 12, l-1}$ and $M_{2\rightarrow 13, l-1}$. For that reason, the analysis carried on is valid. Through this analysis we avoid carrying a lengthy and difficult Fourier-Motzkin procedure to eliminate $\hat{R}^{(l)}_{3\rightarrow 1, l}, \hat{R}^{(l)}_{3\rightarrow 2, l}, \hat{R}^{(l)}_{3\rightarrow 3, l}$ for $l = 1 : K$.}
The decoding of the private descriptions can be seen as a standard Berger-Tung decoding problem (see Fig. 11) where the binning used to transmit the descriptions generated in node 3 is not the common descriptions. Lemma 1 can be easily used to analyze $\mathbb{P}(K_{\text{private}}(2, I))$. The following conditions guarantee that $\mathbb{P}(K_{\text{private}}(2, I)) < \infty$:

$$
\hat{R}_{1 \rightarrow 3}^{(i)} = I \left( X_2; U_{2 \rightarrow 13, I} \mid W_{2, I} | V_{23, I, 1} \right) + \delta_p(2, I),
$$

$$
\hat{R}_{3 \rightarrow 1}^{(i)} < \hat{R}_{1 \rightarrow 2}^{(i)} - \delta_{dp}(2, I),
$$

where $\delta_{dp}(2, I), \delta'_{dp}(2, I), \delta''_{dp}(2, I)$ can be made arbitrarily small.

Then, combining all the obtained results, we have that:

$$
\mathbb{P} \left\{ D_I \cap E_{\text{dec}}(2, I) \cap E_{\text{enc}}(1, I) \right\} \xrightarrow{n \to \infty} 0.
$$

At this point, the story is as it was at the encoding stage in node 1 and all the steps can be repeated with minor modifications, proving the desired results at the end of round $I$:

$$
\mathbb{P} \left\{ D_{I+1} \right\} \xrightarrow{n \to \infty} 1, \quad \mathbb{P} \left\{ D_I \cap E_I \right\} \xrightarrow{n \to \infty} 0.
$$

The rates equations for the encoding and decoding at the other nodes are as follows.

**Encoding at node 2:**

$$
\hat{R}_{2 \rightarrow 13}^{(i)} = I \left( X_2; U_{2 \rightarrow 13, I} \mid W_{2, I} | V_{23, I, 1} \right) + \delta_r(2, I, 13),
$$

$$
\hat{R}_{2 \rightarrow 1}^{(i)} = I \left( X_2; U_{2 \rightarrow 1, I} \mid W_{3, I} | V_{23, I, 2} \right) + \delta_r(2, I, 1),
$$

$$
\hat{R}_{2 \rightarrow 3}^{(i)} = I \left( X_2; U_{2 \rightarrow 3, I} \mid W_{3, I} | V_{23, I, 2} \right) + \delta_r(2, I, 3).
$$

**Decoding at node 2:**

$$
\hat{R}_{2 \rightarrow 13}^{(i)} > I \left( X_2; U_{2 \rightarrow 13, I} \mid W_{2, I} | V_{23, I, 1} \right) + \delta_r(2, I, 13),
$$

$$
\hat{R}_{2 \rightarrow 1}^{(i)} > I \left( X_2; U_{2 \rightarrow 1, I} \mid W_{3, I} | V_{23, I, 2} \right) + \delta_r(2, I, 1),
$$

$$
\hat{R}_{2 \rightarrow 3}^{(i)} > I \left( X_2; U_{2 \rightarrow 3, I} \mid W_{3, I} | V_{23, I, 2} \right) + \delta_r(2, I, 3).
$$

**Encoding at node 3:**

$$
\hat{R}_{3 \rightarrow 12}^{(i)} > I \left( X_3; U_{3 \rightarrow 12, I} \mid W_{3, I} \right) + \delta_r(3, I, 12),
$$

$$
\hat{R}_{3 \rightarrow 1}^{(i)} > I \left( X_3; U_{3 \rightarrow 1, I} \mid W_{1, I+1} | V_{13, I, 3} \right) + \delta_r(3, I, 1),
$$

**Decoding at node 1:**

$$
\hat{R}_{3 \rightarrow 12}^{(i)} > I \left( X_3; U_{3 \rightarrow 12, I} \mid W_{3, I} \right) + \delta_r(1, I),
$$

$$
\hat{R}_{3 \rightarrow 1}^{(i)} > I \left( X_3; U_{3 \rightarrow 1, I} \mid W_{1, I+1} | V_{13, I, 3} \right) + \delta_r(1, I).
$$

The final private rate equations in Theorem 1 follow from a rather simple Fourier-Motzkin elimination procedure.
ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and the Associate Editor Prof. Jun Chen by their insightful comments and advice, which significantly helped improve the presentation of the paper.

REFERENCES

[1] A. H. Kaspi, “Two-way source coding with a fidelity criterion,” IEEE Trans. Inf. Theory, vol. 31, no. 6, pp. 735–740, Nov. 1985.

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

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

[4] T. Berger, “Multiterminal source coding,” Inf. Theory Approach Commun., vol. 229, pp. 171–231, Jul. 1977.

[5] S. Y. Tung, “Multiterminal source coding,” Ph.D. Dissertation, Dept. Elect. Eng., Cornell Univ., Ithaca, NY, USA, May 1978.

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

[7] T. Berger, Z. Zhang, and H. Viswanathan, “The CEO problem [multiterminal source coding],” IEEE Trans. Inf. Theory, vol. 42, no. 3, pp. 857–870, May 1996.

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

[9] A. B. Wagner, S. Tavildar, and P. Viswanathan, “Rate region of the quadratic Gaussian two encoder-source coding problem,” IEEE Trans. Inf. Theory, vol. 54, no. 5, pp. 1938–1961, May 2008.

[10] A. Wagner, B. Kelly, and Y. Altug, “Distributed rate-distortion with common components,” IEEE Trans. Inf. Theory, vol. 57, no. 7, pp. 4035–4057, Jul. 2011.

[11] C. Heegard and T. Berger, “Rate distortion when side information may be absent,” IEEE Trans. Inf. Theory, vol. 31, no. 6, pp. 727–734, Nov. 1985.

[12] R. Timo, T. Chan, and A. Grant, “Rate distortion with side-information at many decoders,” IEEE Trans. Inf. Theory, vol. 57, no. 8, pp. 5240–5257, Aug. 2011.

[13] R. Timo, A. Grant, and G. Kramer, “Lossy broadcasting with complementary side information,” IEEE Trans. Inf. Theory, vol. 59, no. 1, pp. 104–131, Jan. 2013.

[14] N. Ma and P. Ishwar, “Interaction strictly improves the Wyner-Ziv rate-distortion function,” in Proc. Int. Symp. Inf. Theory, Jun. 2010, pp. 61–65.

[15] H. H. Permutter, Y. Steinberg, and T. Weissman, “Two-way source coding with a helper,” IEEE Trans. Inf. Theory, vol. 56, no. 6, pp. 2905–2919, Jun. 2010.

[16] N. Ma and P. Ishwar, “Some results on a distributed source coding for interactive function computation,” IEEE Trans. Inf. Theory, vol. 57, no. 9, pp. 6180–6195, Sep. 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, Jul. 2012.

[18] L. Satikar and H. V. Poor, “Distributed estimation in multi-agent networks,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jul. 2012, pp. 329–333.

[19] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, New York, NY, USA: Academic, 1981.

[20] A. D. Wyner, “The common information of two dependent random variables,” IEEE Trans. Inf. Theory, vol. 21, no. 2, pp. 163–179, Mar. 1975.

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

[22] T. Cover and J. Thomas, Elements of Information Theory, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[23] J. Chen, C. Richard, and A. H. Sayed, “Diffusion LMS over multitask networks,” IEEE Trans. Signal Process., vol. 63, no. 11, pp. 2733–2748, Jun. 2015.

[24] A. El Gamal and Y.-H. Kim, Network Information Theory, Cambridge, U.K.: Cambridge Univ. Press, 2011.

[25] P. Piantanida, L. R. Vega, and A. Hero, “A proof of the generalized Markov lemma with countable infinite sources,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jul. 2014, pp. 591–595.

[26] W. Uhrmann, “Vergleich der hypergeometrischen mit der binomialverteilung,” Metrika, vol. 10, no. 1, pp. 145–158, 1966.

[27] A. Kaspi and T. Berger, “Rate-distortion for correlated sources with partially separated encoders,” IEEE Trans. Inf. Theory, vol. 28, no. 6, pp. 828–840, Nov. 1982.

Leonardo Rey Vega (S’07–M’11) received both B.Sc (with honors) and PhD (summa cum laude) degrees in Electrical Engineering from the University of Buenos Aires (Argentina) in 2004 and 2010, respectively. In 2007 and 2008 he was invited at the INRS-EMT in Montreal, Canada and in the first semester 2012 he was a visitor at the Department of Telecommunications at SUPELEC, France. He is currently an Assistant Professor at the University of Buenos Aires. Dr. Rey Vega’s research interests include statistical signal processing for smart-grid applications, information theory, machine learning and wireless sensor networks.

Alfred O. Hero, III (F’–) received the B.S. (summa cum laude) from Boston University (1980) and the Ph.D from Princeton University (1984), both in Electrical Engineering. Since 1984 he has been with the University of Michigan, Ann Arbor, where he is the John H. Holland Distinguished University Professor of Electrical Engineering and Computer Science and the R. Jamison and Betty Williams Professor of Engineering. He is also the Co-Director of the University’s Michigan Institute for Data Science (MIDAS). His primary appointment is in the Department of Electrical Engineering and Computer Science and he also has appointments, by courtesy, in the Department of Biomedical Engineering and the Department of Statistics. From 2008-2013 he held the Digeito Chaire d’Excellence at the Ecole Superieure d’Electricite, Gif-sur-Yvette, France. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) and several of his research articles have received best paper awards. Alfred Hero was awarded the University of Michigan Distinguished Faculty Achievement Award (2011). He received the IEEE Signal Processing Society Meritorious Service Award (1998), the IEEE Third Millennium Medal (2000), and the IEEE Signal Processing Society Technical Achievement Award (2014). In 2015 he received the Society Award, which is the highest career award bestowed by the IEEE Signal Processing Society. Alfred Hero was President of the IEEE Signal Processing Society (2006-2008) and was on the Board of Directors of the IEEE (2009-2011) where he served as Director of Division IX (Signals and Applications). He served on the IEEE TAB Nominations and Appointments Committee (2012-2014). Alfred Hero is currently a member of the Big Data Special Interest Group (SIG) of the IEEE Signal Processing Society. Since 2011 he has been a member of the Committee on Applied and Theoretical Statistics (CATS) of the US National Academies of Science.

Alfred Hero’s recent research interests are in information theory, high dimensional spatio-temporal data, multi-modal data integration, statistical signal processing, and machine learning. Of particular interest are applications to social networks, network security and forensics, computer vision, and personalized health.