On Network Functional Compression

Soheil Feizi, *Student Member, IEEE*, Muriel Médard, *Fellow, IEEE*

Abstract

In this paper, we consider different aspects of the problem of compressing for function computation across a network, which we call *network functional compression*. In network functional compression, computation of a function (or, some functions) of sources located at certain nodes in a network is desired at receiver(s), that are other nodes in the network. The rate region of this problem has been considered in the literature under certain restrictive assumptions, particularly in terms of the network topology, the functions and the characteristics of the sources. In this paper, we present results that significantly relax these assumptions. Firstly, we consider this problem for an arbitrary tree network and asymptotically lossless computation and derive rate lower bounds. We show that, for depth one trees with correlated sources, or for general trees with independent sources, a modularized coding scheme based on graph colorings and Slepian-Wolf compression performs arbitrarily closely to rate lower bounds. For a general tree network with independent sources, optimal computation to be performed at intermediate nodes is derived. We show that, for a family of functions and random variables called chain rule proper sets, computation at intermediate nodes is not necessary. We introduce a new condition on colorings of source random variables' characteristic graphs called the *coloring connectivity condition (C.C.C.)*. We show that, this condition is necessary and sufficient for any achievable coding scheme based on colorings, thus relaxing the previous sufficient zig-zag condition of Doshi *et al.* We also show that, unlike entropy, graph entropy does not satisfy the chain rule.

Secondly, we consider a multi-functional version of this problem with side information, where the receiver wants to compute several functions, with different side information random variables. We derive a rate region and propose a coding scheme based on graph colorings for this problem. Thirdly, we consider the functional compression problem with feedback. We show that, in this problem, unlike Slepian-Wolf compression, by having feedback, one may outperform rate bounds of the case without

feedback. These results extend those of Bakshi *et al.* Fourthly, we investigate the problem of distributed functional compression with distortion, where computation of a function within a distortion level is desired at the receiver. We compute a rate-distortion region for this problem. Then, we propose a simple suboptimal coding scheme with a non-trivial performance guarantee.

Our coding schemes are based on finding the minimum entropy coloring of the characteristic graph of the function we seek to compute. In general, it is shown by Cardinal *et al.* that finding this coloring is an NP-hard problem. However, we show that, depending on the characteristic graph's structure, there are some interesting cases where finding the minimum entropy coloring is not NP-hard, but tractable and practical. In one of these cases, we show that, by having a non-zero joint probability condition on random variables' distributions, for any desired function, finding the minimum entropy coloring can be solved in polynomial time. In another case, we show that, if the desired function is a quantization function with a certain structure, this problem is also tractable.

Index Terms

Functional Compression, Distributed Computation, Slepian-Wolf Compression, Graph Entropy, Graph Coloring, Feedback, Distortion.

I. INTRODUCTION

In this paper, we consider different aspects of the functional compression problem over networks. In the functional compression problem, we would like to compress source random variables for the purpose of computing a deterministic function (or some deterministic functions) at the receiver(s) when these sources and receivers are nodes in a network. Traditional data compression schemes are special cases of functional compression, where their desired function is the identity function. However, if the receiver is interested in computing a function (or some functions) of sources, further compressing is possible. In the rest of this section, we review some prior relevant research and illustrate some research challenges of this problem through some motivating examples which will be discussed in the following sections.

A. Prior Work in Functional Compression

We categorize prior work into the study of lossless functional compression and that of functional compression with distortion.
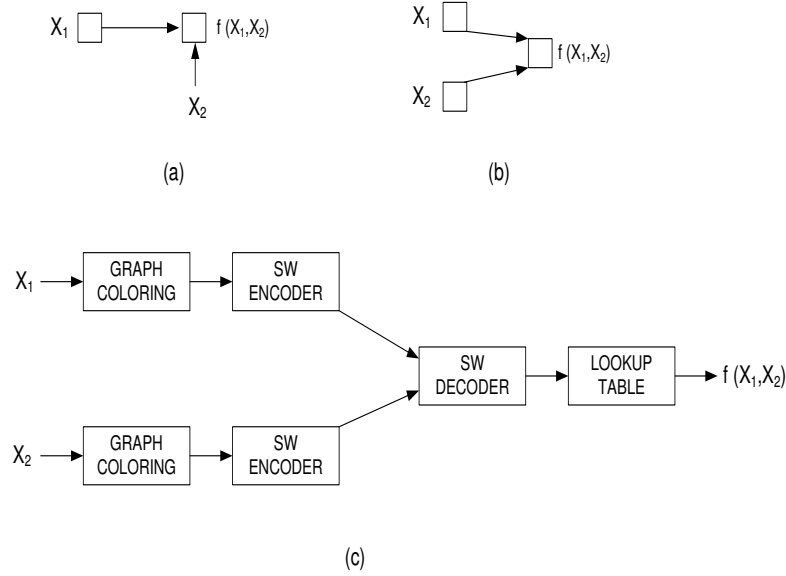
Fig. 1.   a) Functional compression with side information b) A distributed functional compression problem with two transmitters and a receiver c) An achievable encoding/decoding scheme for the functional compression.

1) Lossless Functional Compression: By lossless computation, we mean asymptotically lossless computation of a function: the error probability goes to zero as the block length goes to the infinity.

First, consider the network topology depicted in Figure 1-a which has two sources and a receiver. One of the sources is available at the receiver as the side information. Shannon was the first one who considered this problem in [1] for a special case when $f(X_1, X_2) = (X_1, X_2)$ (the identity function). For a general function, Orlitsky and Roche provided a single-letter characterization in [2]. In [3], Doshi et al. proposed an optimal coding scheme for this problem.

Now, consider the network topology depicted in Figure 1-b which has two sources and a receiver. This problem is a distributed compression problem. For the case that the desired function at the receiver is the identity function (i.e., $f(X_1, X_2) = (X_1, X_2)$), Slepian and Wolf provided a characterization of the rate region and an optimal achievable coding scheme in [4]. Some other practical but suboptimal coding schemes have been proposed by Pradhan and Ramchandran in [5]. Also, a rate-splitting technique for this problem is developed by Coleman *et al.* in [6]. Special cases when $f(X_1, X_2) = X_1$ and $f(X_1, X_2) = (X_1 + X_2) \mod 2$ have been investigated by Ahlswede and Körner in [7], and Körner and Marton in [8], respectively. Under some special
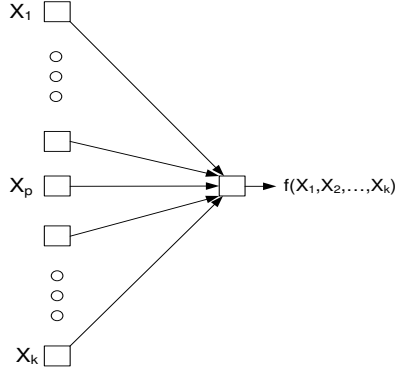
Fig. 2.   A general one-stage tree network with a desired function at the receiver.

conditions on source distributions, Doshi *et al.* in [3] investigated this problem for a general function and proposed some achievable coding schemes.

Sections II, III and IV of this paper consider different aspects of this problem (asymptotically lossless functional compression). In particular, we are going to answer to the following questions in these sections:

- For a depth one tree network with one desired function at the receiver (as shown in Figure 2), what is a necessary and sufficient condition for any coding scheme to guarantee that the network is solvable (i.e., the receiver is able to compute its desired function)?

- What is a rate region of the functional compression problem for a depth one tree network (a rate region is a set of rates for different links of the network under which the network is solvable)? How can a modularized coloring-based coding scheme perform arbitrarily closely to rate bounds?

- For a general tree network with one desired function at the receiver (as shown in Figure 3), when do intermediate nodes need to perform computation and what is an optimal computation to be performed? What is a rate-region for this network?

- How do results extend to the case of having several desired functions with the side information at the receiver?

- What happens if we have feedback in our system?

2) Functional Compression with Distortion: In this section, we review prior results in functional compression for the case of being allowed to compute the desired function at the receiver
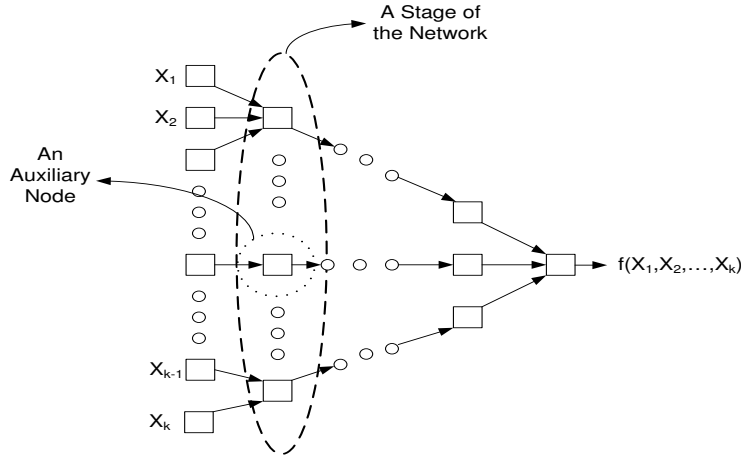
Fig. 3. An arbitrary tree network topology.

within a distortion level.

First, consider the network topology depicted in Figure 1-a, with the side information at the receiver. Wyner and Ziv [9] considered this problem for computing the identity function at the receiver with distortion D. Yamamoto solved this problem for a general function $f(X_1, X_2)$ in [10]. Doshi et al. gave another characterization of the rate distortion function given by Yamamoto in [3]. Feng et al. [11] considered the side information problem for a general function at the receiver in the case the encoder and decoder have some noisy information.

For the network topology depicted in 1-b and for a general function, the rate-distortion region has been unknown, but some bounds have been given by Berger and Yeung [12], Barros and Servetto [13], and Wagner et al. [14], where considered a specific quadratic distortion function.

In Section V of this paper, we answer to the following questions:

- What is a multi-letter characterization of a rate-distortion function for a distributed network depicted in Figure 1-b?

- For this problem, is there a simple suboptimal coding scheme based on graph colorings with a non-trivial performance guarantee?

Remarks:

- Our coding schemes in Sections [I-V] are based on finding a minimum entropy coloring of a characteristic graph of the function we seek to compute. In general, reference [15] showed

that, finding this coloring is an NP-hard problem. However, in Section VI, we consider whether there are some functions and/or source structures which lead to easy and practical coding schemes.

- Note that, our work is different in techniques and the problem setup from multi-round function computation (e.g., [16] and [17]). Also, some references consider the functional computation problem for specific functions. For example, [19] investigated computation of symmetric Boolean functions in tree networks and [20] and [18] studied the sum-network with three sources and three terminals. Note that, in our problem setup, the desired function at the receiver is an arbitrary function. Also, we are interested in asymptotically lossless or lossy computation of this function.

In the rest of this section, we explain some research challenges of the functional compression problem by some examples. In the next sections, we explain these issues with more detail.

B. Problem Outline

In this section, we address some problem outlines of functional compression. We use different simple examples to illustrate these issues which will be explained later in this paper.

Let us proceed by an example:

Example 1. *Consider the network shown in Figure 1-b, which has two source nodes and a receiver. Suppose source nodes have two independent source random variables (RVs) X_1 and X_2 such that X_1 takes values from the set $\mathcal{X}_1 = \{x_1^1, x_1^2, x_1^3, x_1^4\} = \{0, 1, 2, 3\}$, and X_2 takes values from the set $\mathcal{X}_2 = \{x_2^1, x_2^2\} = \{0, 1\}$, both with equal probability. Values of x_i^j for different i and j are shown in Figure 4. Suppose the receiver desires to compute a function $f(X_1, X_2) = (X_1 + X_2) \mod 2$.*

If $X_1 = 0$ or $X_1 = 2$, for all possible values of X_2, we have $f(X_1, X_2) = X_2$. Hence, we do not need to distinguish between $X_1 = 0$ and $X_1 = 2$. A similar argument holds for $X_1 = 1$ and $X_1 = 3$. However, cases $X_1 = 0$ and $X_1 = 1$ should be distinguished, because for $X_2 = 0$, the function value is different when $X_1 = 0$ than the one when $X_1 = 1$ (i.e., $f(0,0) = 0 \neq f(1,0) = 1$).

We notice that for each source random variable, depending on the function at the receiver and values of the other source random variable, we should *distinguish* some possible pair values.
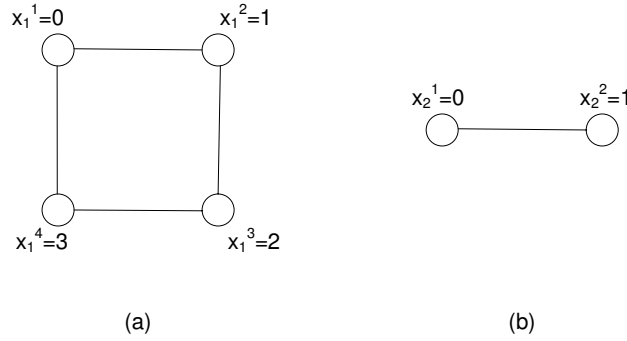
Fig. 4. Characteristic graphs described in Example 1: a) G_{X_1}, b) G_{X_2}.

In other words, values of source random variables which *potentially* can cause confusion at the receiver should be assigned to different *codes*. To determine which pair values of a random variable should be assigned to different codes, we make a graph for each random variable, called the *characteristic graph* or the *confusion graph* of that random variable ([1], [21]). Vertices of this graph are different possible values of that random variable. We connect two vertices if they should be distinguished. For the problem described in Example 1, the characteristic graph of X_1 (called G_{X_1}) is depicted in Figure 4-a. Note that we have not connected vertices which lead to the same function value for all values of X_2. The characteristic graph of X_2 (G_{X_2}) is shown in Figure 4-b.

Now, we seek to assign different codes to connected vertices, which corresponds a *graph coloring* where we assign different colors (codes) to connected vertices. Vertices that are not connected to each other can be assigned to the same or different colors (codes). Figure 5-(a,b) shows valid colorings for G_{X_1} and G_{X_2}.

Now, we propose a possible coding scheme for this example. First, we choose valid colorings for G_{X_1} and G_{X_2}. Instead of sending source random variables, we send these coloring random variables. At the receiver side, we use a look-up table to compute the desired function value by using the received colorings. Figure 5 demonstrates this coding scheme.

However, this coloring-based coding scheme is not necessarily an achievable scheme. In other words, if we send coloring random variables instead of source random variables, the receiver may not be able to compute its desired function. Hence, we need some conditions to guarantee the achievability of coloring-based coding schemes. We explain this required condition by an
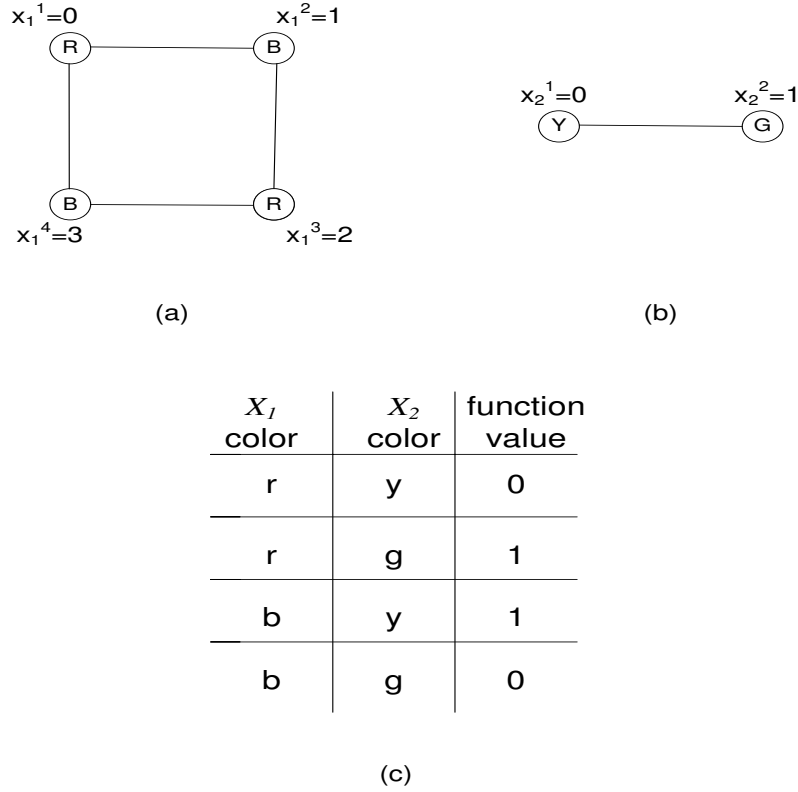
(a)   (b)

X_1 color	X_2 color	function value
r	y	0
r	g	1
b	y	1
b	g	0

(c)

Fig. 5.   a) G_{X_1} b) G_{X_2}, and c) a decoding look-up table for Example 1. (Different letters written over graph vertices indicate different colors.)

example.

Example 2. *Consider the same network topology as explained in Example 1 shown in Figure 1-b. Suppose $\mathcal{X}_1 = \{0, 1\}$ and $\mathcal{X}_2 = \{0, 1\}$. The function values are depicted in Figure 6-a. In particular, $f(0, 0) = 0$ and $f(1, 1) = 1$. Dark squares in this figure represent points with zero probability. Figure 6-b demonstrates characteristic graphs of these source random variables. Each has two vertices, not connected to each other. Hence, we can assign them to a same color. Figure 6-b shows these valid colorings for G_{X_1} and G_{X_2}. However, one may notice that if we send these coloring random variables instead of source random variables, the receiver would not be able to compute its desired function.*

Example 2 demonstrates a case where a coloring-based coding scheme fails to be an achievable scheme. Thus, we need a condition to avoid these situations. We investigate this necessary and
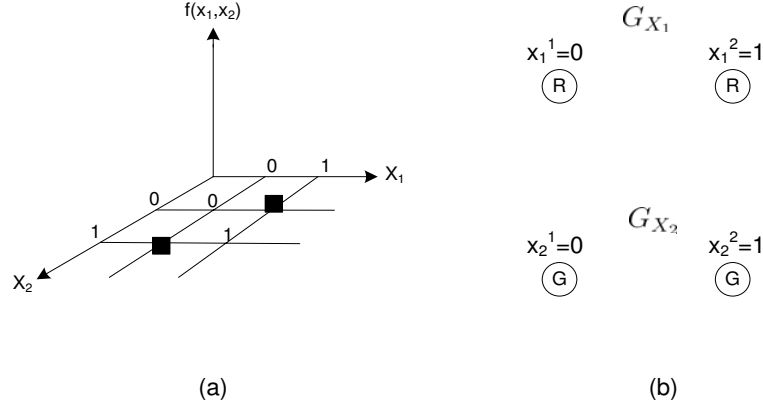
Fig. 6. An example for colorings not satisfying C.C.C. (Different letters written over graph vertices indicate different colors.)

sufficient condition in Section II. We call this condition the *coloring connectivity condition* or *C.C.C.* The situation of Example 2 happens when we have a disconnected coloring class (a coloring class is a set of source pairs with the same color for each coordinates). C.C.C. is a condition to avoid this situation.

Hence, an achievable coding scheme can be expressed as follows. Sources send, instead of source random variables, colorings of their random variables which satisfy C.C.C. Then, they perform source coding on these coloring random variables. Each receiver, by using these colors and a look-up table, can compute its desired function.

However, we may need to consider coloring schemes of conflict graphs of vector extensions of the desired function. In the following, we explain this approach by an example:

Example 3. *Consider the network shown in Figure 1-b. Suppose X_1 is uniformly distributed over $\mathcal{X}_1 = \{0, 1, 2, 3, 4\}$. Consider X_2 and $f(X_1, X_2)$ such that we have a graph depicted in Figure 7 for G_{X_1}. Figure 7 also demonstrates a valid coloring for this graph. Let us call this coloring random variable $c_{G_{X_1}}$. Hence, we have $H(c_{G_{X_1}}) \approx 1.52$. Now, instead of X_1, suppose we encode $X_1 \times X_1$ (X_1^2), a random variable with 25 possibilities ($\{00, 01, ..., 44\}$). To make its characteristic graph, we connect two vertices when at least one of their coordinates are connected in G_{X_1}. Figure 8 illustrates the characteristic graph of X_1^2 (referred by $G_{X_1}^2$ and*
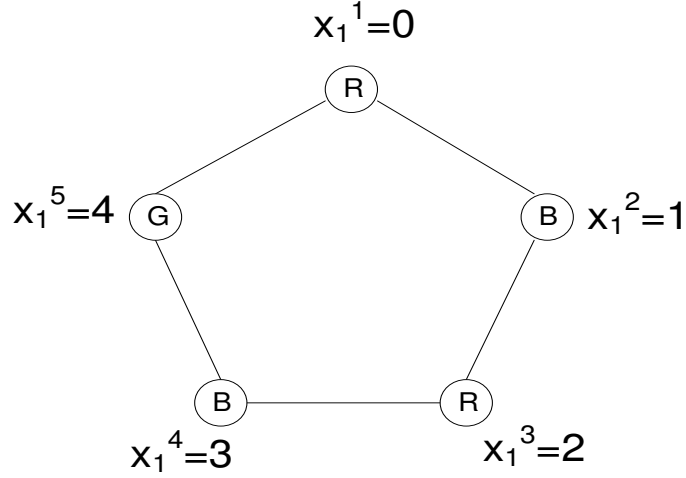
$$x_1{}^1=0$$

$$x_1{}^5=4 \quad G \qquad\qquad B \quad x_1{}^2=1$$

$$B \qquad R$$

$$x_1{}^4=3 \qquad x_1{}^3=2$$

Fig. 7. G_{X_1} described in Example 3. (Different letters written over graph vertices indicate different colors.)

called the second power of the graph G_{X_1}). A valid coloring of this graph, called $c_{G_{X_1}^2}$ is shown in this figure. One may notice that we use eight colors to color this graph. We have,

$$\frac{1}{2}H(c_{G_{X_1}^2}) \approx 1.48 < H(c_{G_{X_1}}) \approx 1.52. \tag{1}$$

Example 3 demonstrates this fact that if we assign colors to a sufficiently large power graph of G_{X_1}, we can compress source random variables more. In Section II, we show that sending colorings of sufficiently large power graphs of characteristic graphs which satisfy C.C.C. followed by a source coding (such as Slepian-Wolf compression) leads to an achievable coding scheme. On the other hand, any achievable coding scheme for this problem can be viewed as a coloring-based coding scheme satisfying C.C.C. In Section II, we shall explain these concepts in more detail.

Now, by another example, we explain some issues of the functional compression problem over tree networks.

Example 4. *Consider the network topology depicted in Figure 9. This is a tree network with four sources, two intermediate nodes and a receiver. Suppose source random variables are independent, with equal probability to be zero or one. In other words, $\mathcal{X}_i = \{0, 1\}$ for $i = 1, 2, 3, 4$. Suppose the receiver wants to compute a parity check function $f(X_1, X_2, X_3, X_4) =$*
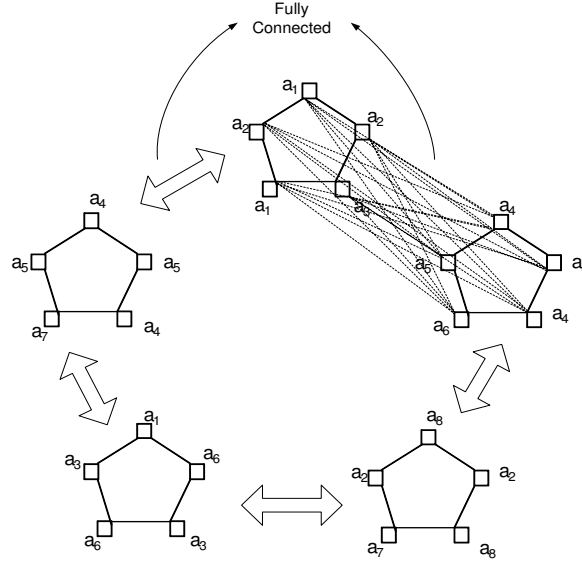
Fig. 8. $G_{X_1}^2$, the second power graph of G_{X_1}, described in Example 3. Letters a_1,...,a_8 written over graph vertices indicate different colors. Two subsets of vertices are fully connected if each vertex of one set is connected to every vertex in the other set.

$(X_1 + X_2 + X_3 + X_4) \mod 2$. *Intermediate nodes are allowed to perform computation.*

In Example 4, first notice that characteristic graphs of source random variables are complete graphs. Hence, coloring random variables of sources are equal to source random variables. If intermediate nodes act like relays (i.e., no computations are performed at intermediate nodes), the following set of rates is an achievable scheme:

$$R_{2j} \geq 1 \text{ for } 1 \leq j \leq 4$$
$$R_{1j} \geq 2 \text{ for } 1 \leq j \leq 2 \tag{2}$$

where R_{ij} are rates of different links depicted in Figure 9.

However, suppose intermediate nodes perform some computations. Assume source nodes send their coloring random variables satisfying C.C.C. (in this case, they are equal to source random variables because characteristic graphs are complete). Then, each intermediate node makes its own characteristic graph and by using the received colors, picks a corresponding color for its own characteristic graph and sends that color. The receiver, by using the received colors of
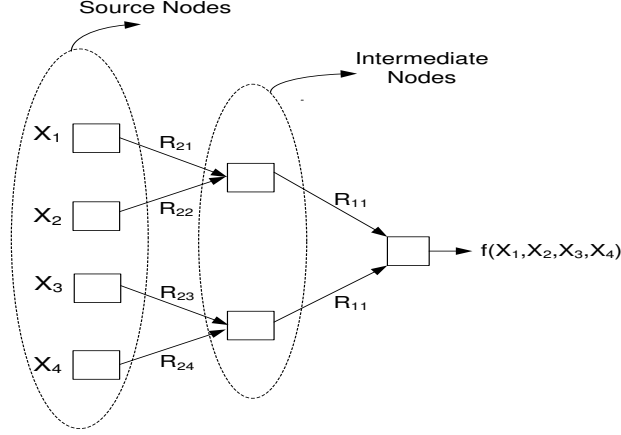
Fig. 9.   An example of a two stage tree network.

intermediate nodes' characteristic graphs and a look-up table, can compute its desired function. Figure 10 demonstrates this encoding/decoding scheme. For this example, intermediate nodes need to transmit one bit. Therefore, the following set of rates is achievable:

$$R_{ij} \geq 1. \tag{3}$$

for different possible i and j. Note that, in Example 4, by allowing intermediate nodes to compute, we can reduce transmission rates of some links. This problem is investigated in Section II for a tree network where optimal computation to be performed at intermediate nodes is derived. We also show that for a family of functions and source random variables, intermediate nodes *do not need to perform computation* and acting like relays is an optimal operation for them.

The problem of having different desired functions at the receiver with the side information is considered in Section III. For this problem, instead of a characteristic graph, we compute a new graph, called a *multi-functional characteristic graph*. This graph is basically an OR function of individual characteristic graphs with respect to different functions. In this section, we find a rate region and propose an achievable coding scheme for this problem.

The effect of feedback on the rate-region of functional compression problem is investigated in Section IV. If the function at the receiver is the identity function, this problem is the Slepian-Wolf compression with feedback. For this case, having feedback does not give us any gain in terms of the rate. For example, reference [22] considers both zero-error and asymptotically zero-error

G_{X_1,X_2}

$(X_1,X_2)=(0,0)$ $(X_1,X_2)=(0,1)$

R   B

B   R

$(X_1,X_2)=(1,0)$ $(X_1,X_2)=(1,1)$

(a)

G_{X_3,X_4}

$(X_3,X_4)=(0,0)$ $(X_3,X_4)=(0,1)$

Y   G

G   Y

$(X_3,X_4)=(1,0)$ $(X_3,X_4)=(1,1)$

(b)

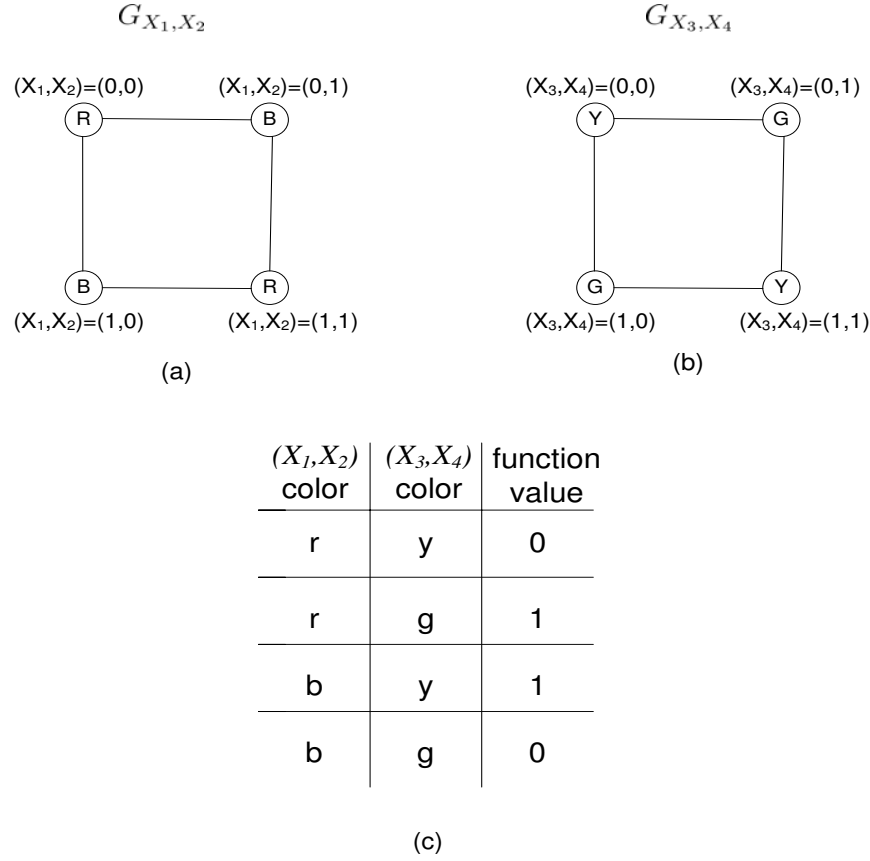| (X_1,X_2) color	(X_3,X_4) color	function value
r	y	0
r	g	1
b	y	1
b	g	0

(c)

Fig. 10. Characteristic graphs and a decoding look-up table for Example 4.

Slepian-Wolf compression with feedback. However, it is not the case when we have a general function f at the receiver. By having feedback, one may outperform rate bounds of the case without feedback.

We consider the problem of distributed functional compression with distortion in Section V. The objective is to compress correlated discrete sources so that an arbitrary deterministic function of those sources can be computed at the receiver within a distortion level. For this case, we compute a rate-distortion region and propose an achievable coding scheme.

In our proposed coding schemes for different functional compression problems, one needs to compute the minimum entropy coloring (a coloring random variable which minimizes entropy) of a characteristic graph. In general, finding this coloring is an NP-hard problem ([15]). However, in Section VI, we show that, depending on the characteristic graph's structure, there are some

interesting cases where finding a minimum entropy coloring is not NP-hard, but tractable and practical. Conclusions and future work are presented in Section VII.

II. FUNCTIONAL COMPRESSION OVER TREE NETWORKS

In this section, we consider the problem of functional compression for an arbitrary tree network. Suppose we have k possibly correlated source processes in a tree network, and a receiver at its root wishes to compute a deterministic function of these processes. Other nodes of this tree (called intermediate nodes) are allowed to perform computation to satisfy the node's demand. For this problem, we find a rate region (i.e., feasible rates for different links) of this network when sources are independent and a rate lower bound when sources are correlated.

The rate region of functional compression problem has been an open problem. However, it has been solved for some simple networks under some special conditions. For instance, [3] considered a rate region of a network with two transmitters and a receiver under a condition on source random variables. Here, we derive a rate lower bound for an arbitrary tree network based on the graph entropy. We introduce a new condition on colorings of source random variables' characteristic graphs called the coloring connectivity condition (C.C.C.). We show that unlike the condition used in [3], this condition is necessary and sufficient for any achievable coding scheme. We also show that, unlike entropy, graph entropy does not satisfy the chain rule. For one stage trees with correlated sources, and general trees with independent sources, we propose a modularized coding scheme based on graph colorings to perform arbitrarily closely to this rate lower bound. We show that in a general tree network case with independent sources, to achieve the rate lower bound, intermediate nodes should perform computation. However, for a family of functions and random variables, which we call chain-rule proper sets, it is sufficient to have intermediate nodes act like relays to perform arbitrarily closely to the rate lower bound.

In this section, after giving the problem statement and reviewing previous results, we explain our main contributions in this problem.

A. Problem Setup

Consider k discrete memoryless random processes, $\{X_1^i\}_{i=1}^\infty$, ..., $\{X_k^i\}_{i=1}^\infty$, as source processes. Memorylessness is not necessary, and one can approximate a source by a memoryless one with an arbitrary precision [23]. Suppose these sources are drawn from finite sets $\mathcal{X}_1 = \{x_1^1, x_1^2, ..., x_1^{|\mathcal{X}_1|}\}$,

..., $\mathcal{X}_k = \{x_k^1, x_k^2, ..., x_k^{|\mathcal{X}_k|}\}$. These sources have a joint probability distribution $p(x_1, ..., x_k)$. We express n-sequences of these random variables as $\mathbf{X_1} = \{X_1^i\}_{i=l}^{i=l+n-1}, ..., \mathbf{X_k} = \{X_k^i\}_{i=l}^{i=l+n-1}$ with the joint probability distribution $p(\mathbf{x}_1, ..., \mathbf{x}_k)$. Without loss of generality, we assume $l = 1$, and to simplify notation, n will be implied by the context if no confusion arises. We refer to the ith element of $\mathbf{x}_j$ as x_{ji}. We use $\mathbf{x}_j^1, \mathbf{x}_j^2,...$ as different n-sequences of $\mathbf{X}_j$. We shall omit the superscript when no confusion arises. Since the sequence $(\mathbf{x}_1, ..., \mathbf{x}_k)$ is drawn i.i.d. according to $p(x_1, ..., x_k)$, one can write $p(\mathbf{x}_1, ..., \mathbf{x}_k) = \prod_{i=1}^n p(x_{1i}, ..., x_{ki})$.

Consider a tree network shown in Figure 3. Suppose we have k source nodes in this network and a receiver in its root. We refer to other nodes of this tree as intermediate nodes. Source node j has an input random process $\{X_j^i\}_{i=1}^\infty$. The receiver wishes to compute a deterministic function $f : \mathcal{X}_1 \times ... \times \mathcal{X}_k \to \mathcal{Z}$, or $f : \mathcal{X}_1^n \times ... \times \mathcal{X}_k^n \to \mathcal{Z}^n$, its vector extension.

Note that sources can be at any nodes of the network. However, without loss of generality, we can modify the network by adding some fake leaves to source nodes which are not located in leaves of the network. So, in the achieved network, sources are located in leaves (as an example, look at Figure 11).

Also, by adding some auxiliary nodes, one can make sources to be in the same distance from the receiver. Hence, we consider source nodes to be in distance d_{max} from the receiver. Consider nodes of a tree with distance $i \geq 1$ from the receiver. We refer to them as the stage i of this tree. Let w_i be the number of such nodes. We refer to the jth node of the ith stage as n_{ij}. Its outgoing link is denoted by e_{ij}. Suppose this node sends M_{ij} over this edge with a rate R_{ij} (it maps length n blocks of M_{ij}, referred to as $\mathbf{M}_{ij}$, to $\{1, 2, ..., 2^{nR_{ij}}\}$.).

If this node is a source node (i.e., $n_{d_{max}j}$ for some j), then $\mathbf{M}_{ij} = en_{X_j}(\mathbf{X}_j)$, where en_{X_j} is the encoding function of the source j.

Now, suppose this node is an intermediate node (i.e., n_{ij}, $i \notin \{1, d_{max}\}$) with incoming edges $e_{(i+1)1}, ...,$ and $e_{(i+1)q}$. We allow this node to compute a function (say $g_{ij}(.)$). Hence, $\mathbf{M}_{ij} = g_{ij}(\mathbf{M}_{(i+1)1}, ..., \mathbf{M}_{(i+1)q})$.

The receiver has a decoder r which maps $r : \prod_{1 \leq j \leq w_1}\{1, ..., 2^{nR_{1j}}\} \to \mathcal{Z}^n$. Thus, the receiver computes $r(\mathbf{M}_{11}, ..., \mathbf{M}_{1w_1}) = r'(en_{X_1}(\mathbf{X}_1), ..., en_{X_k}(\mathbf{X}_k))$. We refer to this encoding/decoding scheme as an n-distributed functional code. Intermediate nodes are allowed to compute functions, but have no demand of their own. The desired function $f(\mathbf{X}_1, ..., \mathbf{X}_k)$ at the receiver is the only demand in the network. For any encoding/decoding scheme, the probability of error is defined
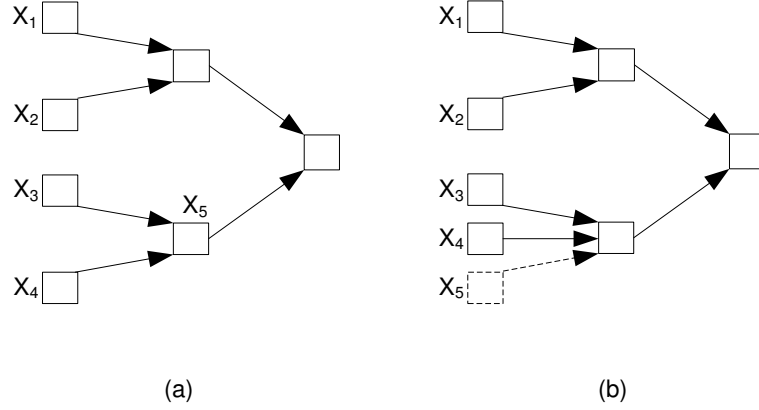
Fig. 11.   a) Sources are not necessarily located in leaves b) By adding some fake nodes, one can assume sources are in leaves of the modified tree.

as

$$P_e^n = Pr[(\mathbf{x}_1, ..., \mathbf{x}_k) : f(\mathbf{x}_1, ..., \mathbf{x}_k) \neq r'(en_{X_1}(\mathbf{x}_1), ..., en_{X_k}(\mathbf{x}_k))]. \tag{4}$$

A rate tuple of the network is the set of rates of its edges (i.e., $\{R_{ij}\}$ for valid i and j). We say a rate tuple is achievable iff there exist a coding scheme operating at these rates so that $P_e^n \to 0$ as $n \to \infty$. The achievable rate region is the set closure of the set of all achievable rates.

B. Definitions and Prior Results

In this part, first we present some definitions used in formulating our results. We also review some prior results.

Definition 5. *The characteristic graph* $G_{X_1} = (V_{X_1}, E_{X_1})$ *of* X_1 *with respect to* X_2, $p(x_1, x_2)$, *and function* $f(X_1, X_2)$ *is defined as follows:* $V_{X_1} = \mathcal{X}_1$ *and an edge* $(x_1^1, x_1^2) \in \mathcal{X}_1^2$ *is in* E_{X_1} *iff there exists a* $x_2^1 \in \mathcal{X}_2$ *such that* $p(x_1^1, x_2^1)p(x_1^2, x_2^1) > 0$ *and* $f(x_1^1, x_2^1) \neq f(x_1^2, x_2^1)$.

In other words, in order to avoid confusion about the function $f(X_1, X_2)$ at the receiver, if $(x_1^1, x_1^2) \in E_{X_1}$, descriptions of x_1^1 and x_1^2 must be different. Shannon first defined this when studying the zero error capacity of noisy channels [1]. Witsenhausen [24] used this concept to study a simplified version of our problem where one encodes X_1 to compute $f(X_1)$ with zero distortion. The characteristic graph of X_2 with respect to X_1, $p(x_1, x_2)$, and $f(X_1, X_2)$ is defined analogously and denoted by G_{X_2}. One can extend the definition of the characteristic graph to

the case of having more than two random variables. Suppose X_1, ..., X_k are k random variables defined in Section V-A.

Definition 6. *The characteristic graph $G_{X_1} = (V_{X_1}, E_{X_1})$ of X_1 with respect to random variables $X_2,...,X_k$, $p(x_1, ..., x_k)$, and $f(X_1, ..., X_k)$ is defined as follows: $V_{X_1} = \mathcal{X}_1$ and an edge $(x_1^1, x_1^2) \in \mathcal{X}_1^2$ is in E_{x_1} if there exist $x_j^1 \in \mathcal{X}_j$ for $2 \leq j \leq k$ such that $p(x_1^1, x_2^1, ..., x_k^1)p(x_1^2, x_2^1, ..., x_k^1) > 0$ and $f(x_1^1, x_2^1, ..., x_k^1) \neq f(x_1^2, x_2^1, ..., x_k^1)$.*

Example 7. *To illustrate the idea of confusability and the characteristic graph, consider two random variables X_1 and X_2 such that $\mathcal{X}_1 = \{0, 1, 2, 3\}$ and $\mathcal{X}_2 = \{0, 1\}$ where they are uniformly and independently distributed on their own supports. Suppose $f(X_1, X_2) = (X_1 + X_2)$ mod 2 is to perfectly reconstructed at the receiver. Then, the characteristic graph of X_1 with respect to X_2, $p(x_1, x_2) = \frac{1}{8}$, and f is shown in Figure 4-a.*

The following definition can be found in [21].

Definition 8. *Given a graph $G_{X_1} = (V_{X_1}, E_{X_1})$ and a distribution on its vertices V_{X_1}, graph entropy is*

$$H_{G_{X_1}}(X_1) = \min_{X_1 \in W_1 \in \Gamma(G_{X_1})} I(X_1; W_1), \tag{5}$$

where $\Gamma(G_{X_1})$ is the set of all maximal independent sets of G_{X_1}.

The notation $X_1 \in W_1 \in \Gamma(G_{X_1})$ means that we are minimizing over all distributions $p(w_1, x_1)$ such that $p(w_1, x_1) > 0$ implies $x_1 \in w_1$, where w_1 is a maximal independent set of the graph G_{x_1}.

Example 9. *Consider the scenario described in Example 7. For the characteristic graph of X_1 shown in Figure 4-a, the set of maximal independent sets is $W_1 = \{\{0, 2\}, \{1, 3\}\}$. To minimize $I(X_1; W_1) = H(X_1) - H(X_1|W_1) = \log(4) - H(X_1|W_1)$, one should maximize $H(X_1|W_1)$. Because of the symmetry of the problem, to maximize $H(X_1|W_1)$, $p(w_1)$ must be uniform over two possible maximal independent sets of G_{X_1}. Since each maximal independent set $w_1 \in W_1$ has two X_1 values, thus, $H(X_1|w_1) = \log(2)$ bit, and since $p(w_1)$ is uniform, $H(X_1|W_1) = \log(2)$ bit. Therefore, $H_{G_{X_1}}(X_1) = \log(4) - \log(2) = 1$ bit. One can see if we want to encode X_1*

ignoring the effect of the function f, we need $H(X_1) = \log(4) = 2$ bits. We will show that, for this example, functional compression saves us 1 bit in every 2 bits compared to the traditional data compression.

Witsenhausen [24] showed that the graph entropy is the minimum rate at which a single source can be encoded so that a function of that source can be computed with zero distortion. Orlitsky and Roche [2] defined an extension of Körner's graph entropy, the *conditional graph entropy*.

Definition 10. *The conditional graph entropy is*

$$H_{G_{X_1}}(X_1|X_2) = \min_{\substack{X_1 \in W_1 \in \Gamma(G_{X_1}) \\ W_1 - X_1 - X_2}} I(W_1; X_1|X_2). \tag{6}$$

Notation $W_1 - X_1 - X_2$ indicates a Markov chain. If X_1 and X_2 are independent, $H_{G_{X_1}}(X_1|X_2) = H_{G_{X_1}}(X_1)$. To illustrate this concept, let us consider an example borrowed from [2].

Example 11. *When $f(X_1, X_2) = X_1$, $H_{G_{X_1}}(X_1|X_2) = H(X_1|X_2)$.*

To show this, consider the characteristic graph of X_1, denoted as G_{X_1}. Since $f(X_1, X_2) = X_1$, then for every $x_2^1 \in \mathcal{X}_2$, the set $\{x_1^i : p(x_1^i, x_2^1) > 0\}$ of possible x_1^i are connected to each other (i.e., this set is a clique of G_{X_1}). Since the intersection of a clique and a maximal independent set is a singleton, X_2 and the maximal independent set W_1 containing X_1 determine X_1. So,

$$\begin{aligned} H_{G_{X_1}}(X_1|X_2) &= \min_{\substack{X_1 \in W_1 \in \Gamma(G_{X_1}) \\ W_1 - X_1 - X_2}} I(W_1; X_1|X_2) \\ &= H(X_1|X_2) - \max_{X_1 \in W_1 \in \Gamma(G_{X_1})} H(X_1|W_1, X_2) \\ &= H(X_1|X_2). \end{aligned} \tag{7}$$

Definition 12. *A vertex coloring of a graph is a function $c_{G_{X_1}}(X_1) : V_{x_1} \to \mathbb{N}$ of a graph $G_{X_1} = (V_{X_1}, E_{X_1})$ such that $(x_1^1, x_1^2) \in E_{X_1}$ implies $c_{G_{X_1}}(x_1^1) \neq c_{G_{X_1}}(x_1^2)$. The entropy of a coloring is the entropy of the induced distribution on colors. Here, $p(c_{G_{X_1}}(x_1^i)) = p(c_{G_{X_1}}^{-1}(c_{G_{X_1}}(x_1^i)))$, where $c_{G_{X_1}}^{-1}(c_{G_{X_1}}(x_1^i)) = \{x_1^j : c_{G_{X_1}}(x_1^j) = c_{G_{X_1}}(x_1^i)\}$ for all valid j. This subset of vertices with the same color is called a color class. We refer to a coloring which minimizes the entropy as a minimum entropy coloring. We use $C_{G_{X_1}}$ as the set of all valid colorings of a graph G_{X_1}.*

Example 13. *Consider again the random variable X_1 described in Example 7, whose characteristic graph G_{X_1} is shown in Figure 4-a. A valid coloring for G_{X_1} is shown in Figure 5-a. One can see that, in this coloring, two connected vertices are assigned to different colors. Specifically, $c_{G_{X_1}}(X_1) = \{r, b\}$. So, $p(c_{G_{X_1}}(x_1^i) = r) = p(x_1^i = 0) + p(x_1^i = 2)$, and $p(c_{G_{X_1}}(x_1^i) = b) = p(x_1^i = 1) + p(x_1^i = 3)$.*

We define a power graph of a characteristic graph as follows:

Definition 14. *The n-th power of a graph G_{X_1} is a graph $G_{\mathbf{X}_1}^n = (V_{X_1}^n, E_{X_1}^n)$ such that $V_{X_1}^n = \mathcal{X}_1^n$ and $(\mathbf{x}_1^1, \mathbf{x}_1^2) \in E_{X_1}^n$ when there exists at least one i such that $(x_{1i}^1, x_{1i}^2) \in E_{X_1}$. We denote a valid coloring of $G_{\mathbf{X}_1}^n$ by $c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1)$.*

One may ignore atypical sequences in a sufficiently large power graph of a conflict graph and then, color that graph. This coloring is called an ϵ-coloring of a graph and is defined as follows:

Definition 15. *Given a non-empty set $\mathcal{A} \subset \mathcal{X}_1 \times \mathcal{X}_2$, define $\hat{p}(x_1, x_2) = p(x_1, x_2)/p(\mathcal{A})$ when $(x_1, x_2) \in \mathcal{A}$, and $\hat{p}(x, y) = 0$ otherwise. $\hat{p}$ is the distribution over (x_1, x_2) conditioned on $(x_1, x_2) \in \mathcal{A}$. Denote the characteristic graph of X_1 with respect to X_2, $\hat{p}(x_1, x_2)$, and $f(X_1, X_2)$ as $\hat{G}_{X_1} = (\hat{V}_{X_1}, \hat{E}_{X_1})$ and the characteristic graph of X_2 with respect to X_1, $\hat{p}(x_1, x_2)$, and $f(X_1, X_2)$ as $\hat{G}_{X_2} = (\hat{V}_{X_2}, \hat{E}_{X_2})$. Note that $\hat{E}_{X_1} \subseteq E_{X_1}$ and $\hat{E}_{X_2} \subseteq E_{X_2}$. Suppose $p(\mathcal{A}) \geq 1 - \epsilon$. We say that $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2)$ are ϵ-colorings of G_{X_1} and G_{x_2} if they are valid colorings of $\hat{G}_{X_1}$ and $\hat{G}_{X_2}$.*

In [25], the *Chromatic entropy* of a graph G_{X_1} is defined as

Definition 16.

$$H_{G_{X_1}}^{\chi}(X_1) = \min_{c_{G_{X_1}} \text{ is an } \epsilon\text{-coloring of } G_{X_1}} H(c_{G_{X_1}}(X_1)).$$

The chromatic entropy is a representation of the chromatic number of high probability subgraphs of the characteristic graph. In [3], the conditional chromatic entropy is defined as

Definition 17.

$$H_{G_{X_1}}^{\chi}(X_1|X_2) = \min_{c_{G_{X_1}} \text{ is an } \epsilon\text{-coloring of } G_{X_1}} H(c_{G_{X_1}}(X_1)|X_2).$$

Regardless of ϵ, the above optimizations are minima, rather than infima, because there are finitely many subgraphs of any fixed graph G_{X_1}, and therefore there are only finitely many ϵ-colorings, regardless of ϵ.

In general, these optimizations are NP-hard ([15]). But, depending on the desired function f, there are some interesting cases that they are not NP-hard. We discuss these cases in Section VI.

Körner showed in [21] that, in the limit of large n, there is a relation between the chromatic entropy and the graph entropy.

Theorem 18.

$$\lim_{n \to \infty} \frac{1}{n} H^{\chi}_{G^n_{\mathbf{X}_1}} (\mathbf{X}_1) = H_{G_{X_1}}(X_1). \tag{8}$$

This theorem implies that the receiver can asymptotically compute a deterministic function of a discrete memoryless source, by first coloring a sufficiently large power of the characteristic graph of the source random variable with respect to the function, and then, encoding achieved colors using any encoding scheme which achieves the entropy bound of the coloring RV. In the previous approach, to achieve the encoding rate close to graph entropy of X_1, one should find the optimal distribution over the set of maximal independent sets of G_{X_1}. But, this theorem allows us to find the optimal coloring of $G^n_{\mathbf{X}_1}$, instead of the optimal distribution on maximal independent sets. One can see that this approach modularizes the encoding scheme into two parts, a graph coloring module, followed by a Slepian-Wolf compression module.

The conditional version of the above theorem is proven in [3].

Theorem 19.

$$\lim_{n \to \infty} \frac{1}{n} H^{\chi}_{G^n_{\mathbf{X}_1}} (\mathbf{X}_1|\mathbf{X}_2) = H_{G_{X_1}}(X_1|X_2). \tag{9}$$

This theorem implies a practical encoding scheme for the problem of functional compression with side information where the receiver wishes to compute $f(X_1, X_2)$, when X_2 is available at the receiver as the side information. Orlitsky and Roche showed in [2] that $H_{G_{X_1}}(X_1|X_2)$ is the minimum achievable rate for this problem. Their proof uses random coding arguments and shows the existence of an optimal coding scheme. This theorem presents a modularized

encoding scheme where one first finds the minimum entropy coloring of $G_{\mathbf{X}_1}^n$ for large enough n, and then uses a compression scheme on the coloring random variable (such as Slepian-Wolf [4]) to achieve a rate arbitrarily close to $H(c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1)|\mathbf{X}_2)$. This encoding scheme guarantees computation of the function at the receiver with a vanishing probability of error.

All these results considered only functional compression with side information at the receiver (Figure 1-a). Consider the network shown in Figure 1-b. It shows a network with two source nodes and a receiver which wishes to compute a function of the sources' values. In general, the rate-region of this network has not been determined. However, [3] determined a rate-region of this network when source random variables satisfy a condition called the zigzag condition, defined below.

We refer to the ϵ-joint-typical set of sequences of random variables $\mathbf{X}_1$, ..., $\mathbf{X}_k$ as T_ϵ^n. k is implied in this notation for simplicity. T_ϵ^n can be considered as a strong or weak typical set ([23]).

Definition 20. *A discrete memoryless source* $\{(X_1^i, X_2^i)\}_{i\in\mathbb{N}}$ *with a distribution* $p(x_1, x_2)$ *satisfies the zigzag condition if for any* ϵ *and some* n, $(\mathbf{x}_1^1, \mathbf{x}_2^1)$, $(\mathbf{x}_1^2, \mathbf{x}_2^2) \in T_\epsilon^n$, *there exists some* $(\mathbf{x}_1^3, \mathbf{x}_2^3) \in T_\epsilon^n$ *such that* $(\mathbf{x}_1^3, \mathbf{x}_2^i), (\mathbf{x}_1^i, \mathbf{x}_2^3) \in T_{\frac{\epsilon}{2}}^n$ *for each* $i \in \{1, 2\}$, *and* $(x_{1j}^3, x_{2j}^3) = (x_{1j}^i, x_{2j}^{3-i})$ *for some* $i \in \{1, 2\}$ *for each* j.

In fact, the zigzag condition forces many source sequences to be typical. We first explain the results of [3]. Then, in Section II-C, we compute a rate-region without the need for the zigzag condition. Then, we extend our results to the case of having k source nodes.

Reference [3] shows that, if the source random variables satisfy the zigzag condition, an achievable rate region for this network is the set of all rates that can be achieved through graph colorings. The zigzag condition is a restrictive condition which does not depend on the desired function at the receiver. This condition is not necessary, but sufficient.

C. A Rate Region for One-Stage Tree Networks

In this section, we want to find a rate region for a general one stage tree network without having any restrictive conditions such as the zigzag condition. Consider the network shown in Figure 2 with k sources.

Definition 21. *A path with length m between two points $Z_1 = (x_1^1, x_2^1, ..., x_k^1)$, and $Z_m = (x_1^2, x_2^2, ..., x_k^2)$ is determined by $m-1$ points Z_i, $1 \leq i \leq m$ such that,*

i) $P(Z_i) > 0$, for all $1 \leq i \leq m$.

ii) Z_i and Z_{i+1} only differ in one of their coordinates.

Definition 21 can be generalized to two n-length vectors as follows.

Definition 22. *A path with length m between two points $\mathbf{Z}_1 = (\mathbf{x}_1^1, \mathbf{x}_2^1, ..., \mathbf{x}_k^1) \in T_\epsilon^n$, and $\mathbf{Z}_m = (\mathbf{x}_1^2, \mathbf{x}_2^2, ..., \mathbf{x}_k^2) \in T_\epsilon^n$ are determined by $m-1$ points $\mathbf{Z}_i$, $1 \leq i \leq m$ such that,*

i) $\mathbf{Z}_i \in T_\epsilon^n$, for all $1 \leq i \leq m$.

ii) $\mathbf{Z}_i$ and $\mathbf{Z}_{i+1}$ only differ in one of their coordinates.

Note that, each coordinate of $\mathbf{Z}_i$ is a vector with length n.

Definition 23. *A joint-coloring family J_C for random variables X_1, ..., X_k with characteristic graphs $G_{X_1},...,G_{X_k}$, and any valid colorings $c_{G_{X_1}},...,c_{G_{X_k}}$, respectively is defined as $J_C = \{j_c^1, ..., j_c^{n_{jc}}\}$ where j_c^i is the collection of points $(x_1^{i_1}, x_2^{i_2}, ..., x_k^{i_k})$ whose coordinates have the same color (i.e., $j_c^i = \left\{ (x_1^{i_1}, x_2^{i_2}, ..., x_k^{i_k}), (x_1^{l_1}, x_2^{l_2}, ..., x_k^{l_k}) : c_{G_{X_1}}(x_1^{i_1}) = c_{G_{X_1}}(x_1^{l_1}), ..., c_{G_{X_k}}(x_k^{i_k}) = c_{G_{X_k}}(x_k^{l_k}) \right\}$, for any valid $i_1,...i_k$, and $l_1,...l_k$). Each j_c^i is called a joint coloring class where n_{jc} is the number of joint coloring classes of a joint coloring family.*

We say a joint coloring class j_c^i is connected if between any two points in j_c^i, there exists a path that lies in j_c^i. Otherwise, it is disconnected. Definition 23 can be expressed for random vectors $\mathbf{X}_1,...,\mathbf{X}_k$ with characteristic graphs $G_{\mathbf{X}_1}^n,...,G_{\mathbf{X}_k}^n$, and any valid ϵ-colorings $c_{G_{\mathbf{X}_1}^n},...,c_{G_{\mathbf{X}_k}^n}$, respectively.

Definition 24. *Consider random variables X_1, ..., X_k with characteristic graphs G_{X_1}, ..., G_{X_k}, and any valid colorings $c_{G_{X_1}}$, ..., $c_{G_{X_k}}$. We say a joint coloring class $j_c^i \in J_C$ satisfies the Coloring Connectivity Condition (C.C.C.) when it is connected or disconnected parts of j_c^i have the same function value. We say colorings $c_{G_{X_1}}$, ..., $c_{G_{X_k}}$ satisfy C.C.C. when all joint coloring classes satisfy C.C.C.*

C.C.C. can be expressed for random vectors $\mathbf{X}_1$, ..., $\mathbf{X}_k$ with characteristic graphs $G_{\mathbf{X}_1}^n$, ..., $G_{\mathbf{X}_k}^n$, and any valid ϵ-colorings $c_{G_{\mathbf{X}_1}^n}$, ..., $c_{G_{\mathbf{X}_k}^n}$, respectively.
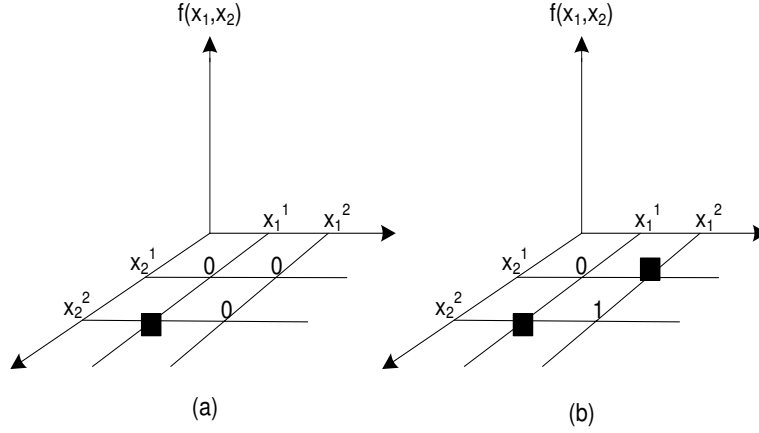
Fig. 12.   Two examples of a joint coloring class: a) satisfying C.C.C. b) not satisfying C.C.C. Dark squares indicate points with zero probability. Function values are depicted in the picture.

Example 25. *For example, suppose we have two random variables X_1 and X_2 with characteristic graphs G_{X_1} and G_{X_2}. Let us assume $c_{G_{X_1}}$ and $c_{G_{X_2}}$ are two valid colorings of G_{X_1} and G_{X_2}, respectively. Assume $c_{G_{X_1}}(x_1^1) = c_{G_{X_1}}(x_1^2)$ and $c_{G_{X_2}}(x_2^1) = c_{G_{X_2}}(x_2^2)$. Suppose j_c^1 represents this joint coloring class. In other words, $j_c^1 = \{(x_1^i, x_2^j)\}$, for all $1 \le i, j \le 2$ when $p(x_1^i, x_2^j) > 0$. Figure 12 considers two different cases. The first case is when $p(x_1^1, x_2^2) = 0$, and other points have a non-zero probability. It is illustrated in Figure 12-a. One can see that there exists a path between any two points in this joint coloring class. So, this joint coloring class satisfies C.C.C. If other joint coloring classes of $c_{G_{X_1}}$ and $c_{G_{X_2}}$ satisfy C.C.C., we say $c_{G_{X_1}}$ and $c_{G_{X_2}}$ satisfy C.C.C. Now, consider the second case depicted in Figure 12-b. In this case, we have $p(x_1^1, x_2^2) = 0$, $p(x_1^2, x_2^1) = 0$, and other points have a non-zero probability. One can see that there is no path between (x_1^1, x_2^2) and (x_1^2, x_2^2) in j_c^1. So, though these two points belong to a same joint coloring class, their corresponding function values can be different from each other. Thus, j_c^1 does not satisfy C.C.C. for this example. Therefore, $c_{G_{X_1}}$ and $c_{G_{X_2}}$ do not satisfy C.C.C.*

Lemma 26. *Consider two random variables X_1 and X_2 with characteristic graphs G_{X_1} and G_{X_2} and any valid colorings $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2)$ respectively, where $c_{G_{X_2}}(X_2)$ is a trivial coloring, assigning different colors to different vertices (to simplify the notation, we use $c_{G_{X_2}}(X_2) = X_2$ to refer to this coloring). These colorings satisfy C.C.C. Also, $c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1)$ and $c_{G_{\mathbf{X}_2}^n}(\mathbf{x}_2) = \mathbf{X}_2$ satisfy C.C.C., for any n.*

Proof: First, we know that any random variable X_2 by itself is a trivial coloring of G_{X_2} such that each vertex of G_{X_2} is assigned to a different color. So, J_C for $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2) = X_2$ can be written as $J_C = \{j_c^1, ..., j_c^{n_{j_c}}\}$ such that $j_c^1 = \{(x_1^i, x_2^i) : c_{G_{X_1}}(x_1^i) = \sigma_i\}$, where σ_i is a generic color. Any two points in j_c^1 are connected to each other with a path with length one. So, j_c^1 satisfies C.C.C. This arguments hold for any j_c^i for any valid i. Thus, all joint coloring classes and therefore, $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2) = X_2$ satisfy C.C.C. The argument for $c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1)$ and $c_{G_{\mathbf{X}_2}^n}(\mathbf{X}_2) = \mathbf{X}_2$ is similar. ∎

Lemma 27. *Consider random variables X_1, ..., X_k with characteristic graphs G_{X_1}, ..., G_{X_k}, and any valid colorings $c_{G_{X_1}}$, ..., $c_{G_{X_k}}$ with joint coloring class $J_C = \{j_c^i : i\}$. For any two points $(x_1^1, ..., x_k^1)$ and $(x_1^2, ..., x_k^2)$ in j_c^i, $f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2)$ if and only if j_c^i satisfies C.C.C.*

Proof: We first show that if j_c^i satisfies C.C.C., then, for any two points $(x_1^1, ..., x_k^1)$ and $(x_1^2, ..., x_k^2)$ in j_c^i, $f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2)$. Since j_c^i satisfies C.C.C., either $f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2)$, or there exists a path with length $m-1$ between these two points $Z_1 = (x_1^1, ..., x_k^1)$ and $Z_m = (x_1^2, ..., x_k^2)$, for some m. Two consecutive points Z_j and Z_{j+1} in this path, differ in just one of their coordinates. Without loss of generality, suppose they differ in their first coordinate. In other words, suppose $Z_j = (x_1^{j_1}, x_2^{j_2}..., x_k^{j_k})$ and $Z_{j+1} = (x_1^{j_0}, x_2^{j_2}..., x_k^{j_k})$. Since these two points belong to j_c^i, $c_{G_{X_1}}(x_1^{j_1}) = c_{G_{X_1}}(x_1^{j_0})$. If $f(Z_j) \neq f(Z_{j+1})$, there would exist an edge between $x_1^{j_1}$ and $x_1^{j_0}$ in G_{X_1} and they could not have the same color. So, $f(Z_j) = f(Z_{j+1})$. By applying the same argument inductively for all two consecutive points in the path between Z_1 and Z_m, one can get $f(Z_1) = f(Z_2) = ... = f(Z_m)$.

If j_c^i does not satisfy C.C.C., it means that there exists at least two points Z_1 and Z_2 in j_c^i such that no path exists between them with different function values. As an example, consider Figure 12-b. Hence, the function value is the same in a joint coloring class iff it satisfies C.C.C. ∎

Lemma 28. *Consider random variables $\mathbf{X}_1$, ..., $\mathbf{X}_k$ with characteristic graphs $G_{\mathbf{X}_1}^n$, ..., $G_{\mathbf{X}_k}^n$, and any valid ϵ-colorings $c_{G_{\mathbf{X}_1}^n}$, ..., $c_{G_{\mathbf{X}_k}^n}$ with the joint coloring class $J_C = \{j_c^i : i\}$. For any two points $(\mathbf{x}_1^1, ..., \mathbf{x}_k^1)$ and $(\mathbf{x}_1^2, ..., \mathbf{x}_k^2)$ in j_c^i, $f(\mathbf{x}_1^1, ..., \mathbf{x}_k^1) = f(\mathbf{x}_1^2, ..., \mathbf{x}_k^2)$ if and only if j_c^i satisfies C.C.C.*

Proof: The proof is similar to Lemma 27. The only difference is to use the definition of C.C.C. for $c_{G_{\mathbf{X}_1}^n}$, ..., $c_{G_{\mathbf{X}_k}^n}$. Since j_c^i satisfies C.C.C., either $f(\mathbf{x}_1^1, ..., \mathbf{x}_k^1) = f(\mathbf{x}_1^2, ..., \mathbf{x}_k^2)$, or there exists a path with length $m-1$ between any two points $\mathbf{Z}_1 = (\mathbf{x}_1^1, ..., \mathbf{x}_k^1) \in T_\epsilon^n$ and $\mathbf{Z}_m = (\mathbf{x}_1^2, ..., \mathbf{x}_k^2) \in T_\epsilon^n$ in j_c^i, for some m. Consider two consecutive points $\mathbf{Z}_j$ and $\mathbf{Z}_{j+1}$ in this path. They differ in one of their coordinates (suppose they differ in their first coordinate). In other words, suppose $\mathbf{Z}_j = (\mathbf{x}_1^{j_1}, \mathbf{x}_2^{j_2}..., \mathbf{x}_k^{j_k}) \in T_\epsilon^n$ and $\mathbf{Z}_{j+1} = (\mathbf{x}_1^{j_0}, \mathbf{x}_2^{j_2}..., \mathbf{x}_k^{j_k}) \in T_\epsilon^n$. Since these two points belong to j_c^i, $c_{G_{X_1}}(\mathbf{x}_1^{j_1}) = c_{G_{X_1}}(\mathbf{x}_1^{j_0})$. If $f(\mathbf{Z}_j) \neq f(\mathbf{Z}_{j+1})$, there would exist an edge between $\mathbf{x}_1^{j_1}$ and $\mathbf{x}_1^{j_0}$ in $G_{\mathbf{X}_1}^n$ and they could not get the same color. Thus, $f(\mathbf{Z}_j) = f(\mathbf{Z}_{j+1})$. By applying the same argument for all two consecutive points in the path between $\mathbf{Z}_1$ and $\mathbf{Z}_m$, one can get $f(\mathbf{Z}_1) = f(\mathbf{Z}_2) = ... = f(\mathbf{Z}_m)$. The converse part is similar to Lemma 27. ∎

Next, we want to show that, if X_1 and X_2 satisfy the zigzag condition given in Definition 20, any valid colorings of their characteristic graphs satisfy C.C.C., but not vice versa. In other words, we want to show that the zigzag condition used in [3] is sufficient but not necessary.

Lemma 29. *If two random variables X_1 and X_2 with characteristic graphs G_{X_1} and G_{X_2} satisfy the zigzag condition, any valid colorings $c_{G_{X_1}}$ and $c_{G_{X_2}}$ of G_{X_1} and G_{X_2} satisfy C.C.C., but not vice versa.*

Proof: Suppose X_1 and X_2 satisfy the zigzag condition, and $c_{G_{X_1}}$ and $c_{G_{X_2}}$ are two valid colorings of G_{X_1} and G_{X_2}, respectively. We want to show that these colorings satisfy C.C.C. To do this, consider two points (x_1^1, x_2^1) and (x_1^2, x_2^2) in a joint coloring class j_c^i. The definition of the zigzag condition guarantees the existence of a path with length two between these two point. Thus, $c_{G_{X_1}}$ and $c_{G_{X_2}}$ satisfy C.C.C.

The second part of this Lemma says that the converse part is not true. To have an example, one can see that in a special case considered in Lemma 26, those colorings always satisfy C.C.C. without having any condition such as the zigzag condition. ∎

Definition 30. *For random variables X_1, ..., X_k with characteristic graphs G_{X_1}, ..., G_{X_k}, the joint graph entropy is defined as follows:*

$$H_{G_{X_1}, ..., G_{X_k}}(X_1, ..., X_k) \triangleq \lim_{n \to \infty} \min_{c_{G_{\mathbf{X}_1}^n}, ..., c_{G_{\mathbf{X}_k}^n}} \frac{1}{n} H\left(c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1), ..., c_{G_{\mathbf{X}_k}^n}(\mathbf{X}_k)\right) \qquad (10)$$

in which $c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1)$, ..., $c_{G_{\mathbf{X}_k}^n}(\mathbf{X}_k)$ *are ϵ-colorings of* $G_{\mathbf{X}_1}^n$, ..., $G_{\mathbf{X}_k}^n$ *satisfying C.C.C. We refer*

to this joint graph entropy as $H_{[G_{X_i}]_{i \in S}}$ where $S = \{1, 2, ..., k\}$. Note that, this limit exists because we have a monotonically decreasing sequence bounded below. Similarly, we can define the conditional graph entropy.

Definition 31. *For random variables X_1, ..., X_k with characteristic graphs G_{X_1}, ..., G_{X_k}, the conditional graph entropy can be defined as follows:*

$$
\begin{aligned}
&H_{G_{X_1}, ..., G_{X_i}}(X_1, ..., X_i | X_{i+1}, ..., X_k) \\
&\triangleq \lim_{n \to \infty} \min \frac{1}{n} H(c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1), ..., c_{G_{\mathbf{X}_i}^n}(\mathbf{X}_i) | c_{G_{\mathbf{X}_{i+1}}^n}(\mathbf{X}_{i+1}), ..., c_{G_{\mathbf{X}_k}^n}(\mathbf{X}_k))
\end{aligned}
\tag{11}
$$

where the minimization is over $c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1)$, ..., $c_{G_{\mathbf{X}_k}^n}(\mathbf{X}_k)$, which are ϵ-colorings of $G_{\mathbf{X}_1}^n$, ..., $G_{\mathbf{X}_k}^n$ satisfying C.C.C.

Lemma 32. *For $k = 2$, Definitions 10 and 31 are the same.*

Proof: By using the data processing inequality, we have

$$
\begin{aligned}
H_{G_{X_1}}(X_1 | X_2) &= \lim_{n \to \infty} \min_{c_{G_{\mathbf{X}_1}^n}, c_{G_{\mathbf{X}_2}^n}} \frac{1}{n} H(c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1) | c_{G_{\mathbf{X}_2}^n}(\mathbf{X}_2)) \\
&= \lim_{n \to \infty} \min_{c_{G_{\mathbf{X}_1}^n}} \frac{1}{n} H(c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1) | \mathbf{X}_2).
\end{aligned}
$$

Then, Lemma 26 implies that $c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1)$ and $c_{G_{\mathbf{X}_2}^n}(\mathbf{x}_2) = \mathbf{X}_2$ satisfy C.C.C. A direct application of Theorem 19 completes the proof. ∎

Note that, by this definition, *the graph entropy does not satisfy the chain rule*.

Suppose $\mathcal{S}(k)$ denotes the power set of the set $\{1, 2, ..., k\}$ excluding the empty subset. Then, for any $S \in \mathcal{S}(k)$,

$$
X_S \triangleq \{X_i : i \in S\}.
$$

Let S^c denote the complement of S in $\mathcal{S}(k)$. For $S = \{1, 2, ..., k\}$, denote S^c as the empty set. To simplify notation, we refer to a subset of sources by X_S. For instance, $\mathcal{S}(2) = \{\{1\}, \{2\}, \{1, 2\}\}$, and for $S = \{1, 2\}$, we write $H_{[G_{X_i}]_{i \in S}}(X_s)$ instead of $H_{G_{X_1}, G_{X_2}}(X_1, X_2)$.

Theorem 33. *A rate region of the network shown in Figure 2 is determined by these conditions:*

$$\forall S \in \mathcal{S}(k) \implies \sum_{i \in S} R_{1i} \geq H_{[G_{X_i}]_{i \in S}}(X_S | X_{S^c}). \tag{12}$$

Proof: We first show the achievability of this rate region. We also propose a modularized en-coding/decoding scheme in this part. Then, for the converse, we show that no encoding/decoding scheme can outperform this rate region.

1)*Achievability*:

Lemma 34. *Consider random variables* $\mathbf{X}_1$, ..., $\mathbf{X}_k$ *with characteristic graphs* $G_{\mathbf{X}_1}^n$, ..., $G_{\mathbf{X}_k}^n$, *and any valid* ϵ-colorings $c_{G_{\mathbf{X}_1}^n}$, ..., $c_{G_{\mathbf{X}_k}^n}$ *satisfying C.C.C., for sufficiently large* n. *There exists*

$$\hat{f} : c_{G_{\mathbf{X}_1}^n}(\mathcal{X}_1) \times ... \times c_{G_{\mathbf{X}_k}^n}(\mathcal{X}_k) \rightarrow \mathcal{Z}^n \tag{13}$$

such that $\hat{f}(c_{G_{\mathbf{X}_1}^n}(\mathbf{x}_1), ..., c_{G_{\mathbf{X}_k}^n}(\mathbf{x}_k)) = f(\mathbf{x}_1, ..., \mathbf{x}_k)$, *for all* $(\mathbf{x}_1, ..., \mathbf{x}_k) \in T_\epsilon^n$.

Proof: Suppose the joint coloring family for these colorings is $J_C = \{j_c^i : i\}$. We proceed by constructing $\hat{f}$. Assume $(\mathbf{x}_1^1, ..., \mathbf{x}_k^1) \in j_c^i$ and $c_{G_{\mathbf{X}_1}^n}(\mathbf{x}_1^1) = \sigma_1$, ..., $c_{G_{\mathbf{X}_1}^n}(\mathbf{x}_1^1) = \sigma_k$. Define $\hat{f}(\sigma_1, ...\sigma_k) = f(\mathbf{x}_1^1, ..., \mathbf{x}_k^1)$.

To show this function is well-defined on elements in its support, we should show that for any two points $(\mathbf{x}_1^1, ..., \mathbf{x}_k^1)$ and $(\mathbf{x}_1^2, ..., \mathbf{x}_k^2)$ in T_ϵ^n, if $c_{G_{\mathbf{X}_1}^n}(\mathbf{x}_1^1) = c_{G_{\mathbf{X}_1}^n}(\mathbf{x}_1^2)$, ..., $c_{G_{\mathbf{X}_k}^n}(\mathbf{x}_k^1) = c_{G_{\mathbf{X}_k}^n}(\mathbf{x}_k^2)$, then $f(\mathbf{x}_1^1, ..., \mathbf{x}_k^1) = f(\mathbf{x}_1^2, ..., \mathbf{x}_k^2)$.

Since $c_{G_{\mathbf{X}_1}^n}(\mathbf{x}_1^1) = c_{G_{\mathbf{X}_1}^n}(\mathbf{x}_1^2)$, ..., $c_{G_{\mathbf{X}_k}^n}(\mathbf{x}_k^1) = c_{G_{\mathbf{X}_k}^n}(\mathbf{x}_k^2)$, these two points belong to a joint coloring class such as j_c^i. Since $c_{G_{\mathbf{X}_1}^n}$, ..., $c_{G_{\mathbf{X}_k}^n}$ satisfy C.C.C., by using Lemma 28, $f(\mathbf{x}_1^1, ..., \mathbf{x}_k^1) = f(\mathbf{x}_1^2, ..., \mathbf{x}_k^2)$. Therefore, our function $\hat{f}$ is well-defined and has the desired property. ∎

Lemma 34 implies that, given ϵ-colorings of characteristic graphs of random variables satisfy-ing C.C.C. at the receiver, we can successfully compute the desired function f with a vanishing probability of error as n goes to infinity. Thus, if the decoder at the receiver is given colors, it can look up f based on its table of $\hat{f}$. The question remains of at which rates encoders can transmit these colors to the receiver faithfully (with a probability of error less than ϵ).

Lemma 35. *(Slepian-Wolf Theorem)*

A rate-region of the network shown in Figure 2 where $f(X_1, ..., X_k) = (X_1, ..., X_k)$ *can be*

determined by these conditions:

$$\forall S \in \mathcal{S}(k) \Longrightarrow \sum_{i \in S} R_{1i} \geq H(X_S | X_{S^c}). \tag{14}$$

Proof: See [4]. ∎

We now use the Slepian-Wolf (SW) encoding/decoding scheme on achieved coloring random variables. Suppose the probability of error in each decoder of SW is less than $\frac{\epsilon}{k}$. Then, the total error in the decoding of colorings at the receiver is less than ϵ. Therefore, the total error in the coding scheme of first coloring $G_{\mathbf{X}_1}^n$, ..., $G_{\mathbf{X}_k}^n$, and then encoding those colors by using SW encoding/decoding scheme is upper bounded by the sum of errors in each stage. By using Lemmas 34 and 35, the total error is less than ϵ, and goes to zero as n goes to infinity. By applying Lemma 35 on achieved coloring random variables, we have,

$$\forall S \in \mathcal{S}(k) \Longrightarrow \sum_{i \in S} R_{1i} \geq \frac{1}{n} H(c_{G_{\mathbf{X}_S}^n} | c_{G_{\mathbf{X}_{S^c}}^n}), \tag{15}$$

where $c_{G_{\mathbf{X}_S}^n}$, and $c_{G_{\mathbf{X}_{S^c}}^n}$ are ϵ-colorings of characteristic graphs satisfying C.C.C. Thus, using Definition 31 completes the achievability part.

As an example, look at Figure 1-c. This network has two source nodes and a receiver. Source nodes compute ϵ-colorings of their characteristic graphs. These colorings should satisfy C.C.C. Then, an SW compression is performed on these colorings. The receiver, first, perform SW decoding to get the colors. Then, by using a look-up table, it can find the value of its desired function (As an example, look at Figure 5).

2) Converse: Here, we show that any distributed functional source coding scheme with a small probability of error induces ϵ-colorings on characteristic graphs of random variables satisfying C.C.C. Suppose $\epsilon > 0$. Define $\mathcal{F}_\epsilon^n$ for all (n, ϵ) as follows,

$$\mathcal{F}_\epsilon^n = \{\hat{f} : Pr[\hat{f}(\mathbf{X}_1, ..., \mathbf{X}_k) \neq f(\mathbf{X}_1, ..., \mathbf{X}_k)] < \epsilon\}. \tag{16}$$

In other words, $\mathcal{F}_\epsilon^n$ is the set of all functions equal to f with ϵ probability of error. For large enough n, all achievable functional source codes are in $\mathcal{F}_\epsilon^n$. We call these codes ϵ-achievable functional codes.

Lemma 36. *Consider some function $f : \mathcal{X}_1 \times ... \times \mathcal{X}_k \to \mathcal{Z}$. Any distributed functional code*

which reconstructs this function with zero error probability induces colorings on $G_{X_1}, ..., G_{X_k}$ with respect to this function, where these colorings satisfy C.C.C.

Proof: To show this lemma, let us assume we have a zero-error distributed functional code represented by encoders en_{X_1}, ..., en_{X_k} and a decoder r. Since it is error free, for any two points $(x_1^1, ..., x_k^1)$ and $(x_1^2, ..., x_k^2)$, if $p(x_1^1, ..., x_k^1) > 0$, $p(x_1^2, ..., x_k^2) > 0$, $en_{X_1}(x_1^1) = en_{X_1}(x_1^2)$, ..., $en_{X_k}(x_k^1) = en_{X_k}(x_k^2)$, then,

$$f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2) = r'(en_{X_1}(x_1^1), ..., en_{X_k}(x_k^1)). \tag{17}$$

We want to show that en_{X_1}, ..., en_{X_k} are some valid colorings of G_{X_1}, ..., G_{X_k} satisfying C.C.C. We demonstrate this argument for X_1. The argument for other random variables is analogous. First, we show that en_{X_1} induces a valid coloring on G_{X_1}, and then, we show that this coloring satisfies C.C.C. Let us proceed by contradiction. If en_{X_1} did not induce a coloring on G_{X_1}, there must be some edge in G_{X_1} with both vertices with the same color. Let us call these vertices x_1^1 and x_1^2. Since these vertices are connected in G_{X_1}, there must exist a $(x_2^1, ..., x_k^1)$ such that, $p(x_1^1, x_2^1, ..., x_k^1)p(x_1^2, x_2^1, ..., x_k^1) > 0$, $en_{X_1}(x_1^1) = en_{X_1}(x_1^2)$, and $f(x_1^1, x_2^1, ..., x_k^1) \neq f(x_1^2, x_2^1, ..., x_k^1)$. By taking $x_2^1 = x_2^2$, ..., $x_k^1 = x_k^2$ in (17), one can see that it is not possible. So, the contradiction assumption is wrong and en_{X_1} induces a valid coloring on G_{X_1}.

Now, we should show that these induced colorings satisfy C.C.C. If it was not true, it would mean that there must exist two point $(x_1^1, ..., x_k^1)$ and $(x_1^2, ..., x_k^2)$ in a joint coloring class j_c^i such that there is no path between them in j_c^i. So, Lemma 27 says that the function f can get different values in these two points. In other words, it is possible to have $f(x_1^1, ..., x_k^1) \neq f(x_1^2, ..., x_k^2)$, where $c_{G_{X_1}}(x_1^1) = c_{G_{X_1}}(x_1^2)$, ..., $c_{G_{X_k}}(x_k^1) = c_{G_{X_k}}(x_k^2)$, which is in contradiction with (17). Thus, these colorings satisfy C.C.C. ∎

In the last step, we should show that any achievable functional code represented by $\mathcal{F}_\epsilon^n$ induces ϵ-colorings on characteristic graphs satisfying C.C.C.

Lemma 37. *Consider random variables $\mathbf{X}_1$, ..., $\mathbf{X}_k$. All ϵ-achievable functional codes of these random variables induce ϵ-colorings on characteristic graphs satisfying C.C.C.*

Proof: Suppose $g(\mathbf{x}_1, ..., \mathbf{x}_k) = r'(en_{X_1}(\mathbf{x}_1), ..., en_{X_k}(\mathbf{x}_k)) \in \mathcal{F}_\epsilon^n$ is such a code. Lemma 36 says that a zero-error reconstruction of g induces some colorings on characteristic graphs satisfy-
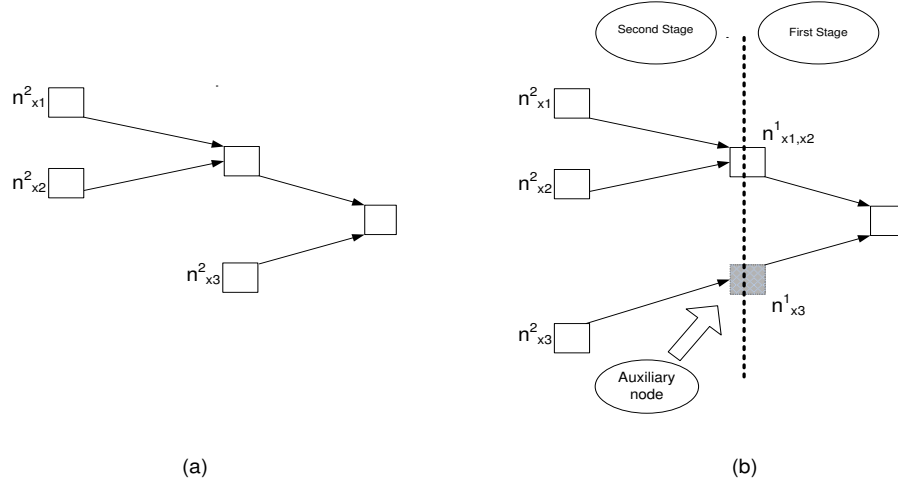
Fig. 13.   (a) A simple tree topology. (b) A completed tree.

ing C.C.C., with respect to g. Suppose the set of all points $(\mathbf{x}_1, ..., \mathbf{x}_k)$ such that $g(\mathbf{x}_1, ..., \mathbf{x}_k) \neq f(\mathbf{x}_1, ..., \mathbf{x}_k)$ be denoted by $\mathcal{C}$. Since $g \in \mathcal{F}_\epsilon^n$, $Pr[\mathcal{C}] < \epsilon$. Therefore, functions en_{X_1}, ..., en_{X_k} restricted to $\mathcal{C}$ are ϵ-colorings of characteristic graphs satisfying C.C.C. (by definition).   ∎

Lemmas 36 and 37 establish the converse part and complete the proof.   ∎

If we have two transmitters ($k = 2$), Theorem 33 can be simplified as follows.

Corollary 38. *A rate region of the network shown in Figure 1-b is determined by these three conditions:*

$$
\begin{aligned}
R_{11} &\geq H_{G_{X_1}}(X_1|X_2) \\
R_{12} &\geq H_{G_{X_2}}(X_2|X_1) \\
R_{11} &+ R_{12} \geq H_{G_{X_1}, G_{X_2}}(X_1, X_2).
\end{aligned}
\tag{18}
$$

D. A Rate Lower Bound for a General Tree Network

In this section, we compute a rate lower bound of an arbitrary tree network with k correlated sources at its leaves and a receiver in its root (see Figure 3). We refer to other nodes of this tree as intermediate nodes. The receiver wishes to compute a deterministic function of source random variables. Intermediate nodes have no demand of their own, but they are allowed to

perform computation. Computing the desired function f at the receiver is the only demand of the network. For independent sources, we propose a modularized coding scheme to perform arbitrarily closely to derived rate lower bounds.

First, we propose a framework to categorize any tree networks and their nodes.

Definition 39. *For an arbitrary tree network,*

- *The distance of each node is the number of hops in the path between that node and the receiver.*

- d_{max} *is the distance of the farthest node from the receiver.*

- *A complete tree is a tree such that all source nodes are in a distance d_{max} from the receiver.*

- *An auxiliary node is a new node connected to a leaf of a tree and increases the leaf's distance by one. The added link is called an auxiliary link. The leaf in the original tree to which is added an auxiliary node is called the actual node corresponding to that auxiliary node. The link in the original tree connected to the actual node is called the actual link corresponding to that auxiliary link.*

- *For any given tree, one can complete it by adding some consecutive auxiliary nodes to its leaves whose distances are less than d_{max}. The achieved tree is called the completed tree and this process is called tree completion.*

These concepts are depicted in Figure 13. Auxiliary nodes in the completed tree network act as intermediate nodes. Note that, all functions that may be computed in auxiliary nodes can be gathered in their corresponding actual node in the original tree. So, the rate of the actual link in the original tree network is the minimum of rates of corresponding auxiliary links in the completed tree. Thus, if we compute the rate-region for the completed tree of any given arbitrary tree, we can compute the rate-region of the original tree. Therefore, in the rest of this section, we consider the rate-region of completed tree networks.

Definition 40. *Consider a completed tree network with k source nodes in distance d_{max} from the receiver. Nodes in distance i from the receiver are called the ith stage of the tree. w_i is the number of nodes in the ith stage. Ξ_{ij} is a subset of source random variables whose paths to the receiver have the last i common hops. Ξ_{ij} is called source variables of node n_{ij}. A connection set of a completed tree is defined as $S_T = \{s_t^i : 1 \leq i \leq d_{max}\}$, where $s_t^i = \{\Xi_{ij} : 1 \leq j \leq w_i\}$.*

Note that, any completed tree can be expressed by a connection set.

For example, consider the network shown in Figure 13-b. Its connection set is $S_T = \{s_t^1, s_t^2\}$ such that $s_t^1 = \{(X_1, X_2), X_3\}$ and $s_t^2 = \{X_1, X_2, X_3\}$. In other words, $\Xi_{11} = (X_1, X_2)$, $\Xi_{12} = X_3$, $\Xi_{21} = X_1$, $\Xi_{22} = X_2$ and $\Xi_{23} = X_3$. One can see that S_T completely describes the structure of the tree.

We have three types of nodes: source nodes, intermediate nodes and a receiver. Source nodes process their messages and transmit them. Intermediate nodes can compute some functions of their received information. The receiver processes the received information to compute its desired function. For example, consider the network shown in Figure 13-b. Random variables sent through links e_{21}, e_{22}, e_{23}, e_{11} and e_{12} are M_{21}, M_{22}, M_{23}, M_{11} and M_{12} such that $M_{11} = g_{11}(M_{21}, M_{22})$, and $M_{12} = g_{12}(M_{23})$.

1) A Rate Lower Bound: Consider node n_{ij} of a tree. Let $\mathcal{S}(w_i)$ be the power set of the set $\{1, 2, ..., w_i\}$ and $s_i \in \mathcal{S}(w_i)$ be a non-empty subset of $\{1, 2, ..., w_i\}$.

Theorem 41. *A rate lower bound of a tree network with the connection set $S_T = \{s_t^i : i\}$ can be determined by these conditions,*

$$\forall s_i \in \mathcal{S}(w_i) \implies \sum_{j \in s_i} R_{ij} \geq H_{[G_{\Xi_{iz}}]_{z \in s_i}}(\Xi_{is_i} | \Xi_{is_i^c}) \tag{19}$$

for all $i = 1, ..., |S_T|$ where $\Xi_{is_i} = [\Xi_{ij}]_{j \in s_i}$ and $\Xi_{is_i^c} = \{X_1, ..., X_k\} \backslash \{\Xi_{is_i}\}$.

Proof: In this part, we want to show that no coding scheme can outperform this rate region. Consider nodes in the i-th stage of this network, n_{ij} for $1 \leq j \leq w_i$. Suppose they are directly connected to the receiver. So, the information sent in links of this stage should be sufficient to compute the desired function. Suppose their parent nodes sent all their information without doing any compression. So, by direct application of Theorem 33, (19) can be derived. This argument can be repeated for all stages. Thus, no coding scheme can outperform these bounds. ∎

In the following, we express some cases under which we can achieve the derived rate lower bound of Theorem 41.

2) Tightness of the Rate Lower Bound for Independent Sources: In this part, we propose a functional coding scheme to achieve the rate lower bound. Suppose random variables $\mathbf{X}_1$, ..., $\mathbf{X}_k$ with characteristic graphs $G_{\mathbf{X}_1}^n$, ..., $G_{\mathbf{X}_k}^n$ are independent. Assume $c_{G_{\mathbf{X}_1}^n}$, ..., $c_{G_{\mathbf{X}_k}^n}$ are valid

ϵ-colorings of these characteristic graphs satisfying C.C.C. The proposed coding scheme can be described as follows: source nodes first compute colorings of high probability subgraphs of their characteristic graphs satisfying C.C.C., and then, perform source coding on these coloring random variables. Intermediate nodes first compute their parents' coloring random variables, and then, by using a look-up table, find corresponding source values of their received colorings. Then, they compute ϵ-colorings of their own characteristic graphs. The corresponding source values of their received colorings form an independent set in the graph. If all are assigned to a single color in the minimum entropy coloring, intermediate nodes send this coloring random variable followed by a source coding. But, if vertices of this independent set are assigned to different colors, intermediate nodes send the coloring with the lowest entropy followed by source coding (Slepian-Wolf). The receiver first performs a minimum entropy decoding ([23]) on its received information and achieves coloring random variables. Then, it uses a look-up table to compute its desired function by using achieved colorings.

To show the achievability, we show that, if nodes of each stage were directly connected to the receiver, the receiver could compute its desired function. Consider the node n_{ij} in the i-th stage of the network. Since the corresponding source values of its received colorings form an independent set on its characteristic graph ($G_{\Xi_{ij}}$) and this node computes the minimum entropy of this graph, it is equivalent to the case where it would receive the exact source information, because both of them lead to the same coloring random variable. So, if all nodes of stage i were directly connected to the receiver, the receiver could compute its desired function and link rates would satisfy the following conditions.

$$\forall s_i \in \mathcal{S}(w_i) \Longrightarrow \sum_{j \in s_i} R_{ij} \geq H_{[G_{\Xi_{iz}}]_{z \in s_i}}(\Xi_{is_i}). \tag{20}$$

Thus, by using a simple induction argument, one can see that the proposed scheme is achievable and it can perform arbitrarily closely to the derived rate lower bound, while sources are independent.

E. A Case When Intermediate Nodes Do not Need to Compute

Though the proposed coding scheme in Section II-D2 can perform arbitrarily closely to the rate lower bound, it may require computation at intermediate nodes.

Definition 42. *Suppose $f(X_1, ..., X_k)$ is a deterministic function of random variables $X_1,...,X_k$. $(f, X_1, ..., X_k)$ is called a chain-rule proper set when for any $s \in \mathcal{S}(k)$, $H_{[G_{X_i}]_{i \in S}} = H_{G_{X_s}}(X_s)$.*

Theorem 43. *In a general tree network, if sources $X_1,...,X_k$ are independent random variables and $(f, X_1, ..., X_k)$ is a chain-rule proper set, it is sufficient to have intermediate nodes as relays to perform arbitrarily closely to the rate lower bound mentioned in Theorem 41.*

Proof: Consider an intermediate node n_{ij} in the i-th stage of the network whose corresponding source random variables are X_s where $s \in \mathcal{S}(k)$ (i.e., $X_s = \Xi_{ij}$). Since random variables are independent, one can write rate bounds of Theorem 41 as,

$$\forall s_i \in \mathcal{S}(w_i) \implies \sum_{j \in s_i} R_{ij} \geq H_{[G_{\Xi_{iz}}]_{z \in s_i}}(\Xi_{is_i}). \tag{21}$$

Now, consider the outgoing link rate of the node n_{ij}. If this intermediate node acts like a relay, we have $R_{ij} = H_{[G_{X_i}]_{i \in S}(X_s)}$ (since $X_s = \Xi_{ij}$). If $(f, X_1, ..., X_k)$ is a chain-rule proper set, we can write,

$$
\begin{aligned}
R_{ij} &= H_{[G_{X_i}]_{i \in S}}(X_s) \\
&= H_{G_{X_s}}(X_s) \\
&= H_{G_{\Xi_{ij}}}(\Xi_{ij}).
\end{aligned}
\tag{22}
$$

For any intermediate node n_{ij} where $j \in s_i$ and $s_i \in \mathcal{S}(w_i)$, we can write a similar argument which lead to conditions (21). As mentioned in Theorem 41, to perform arbitrarily closely to the rate lower bound, this node needs to compress its received information up to the rate $H_{G_{X_s}}(X_s)$. If this node acted as a relay and forwarded the received information from the previous stage, it would lead to an achievable rate $H_{\bigcup_{i \in s} G_{X_i}}(X_s)$ in the next stage, which in general is not equal to $H_{G_{X_s}}(X_s)$. So, this scheme cannot achieve the rate lower bound. However, if for any $s \in \mathcal{S}(k)$, $H_{\bigcup_{i \in s} G_{X_i}}(X_s) = H_{G_{X_s}}(X_s)$, this scheme can perform arbitrarily closely to the rate lower bound by having intermediate nodes as relays. ∎

In the following Lemma, we provide a sufficient condition to guarantee that a set is a chain-rule proper set.

Lemma 44. *Suppose X_1 and X_2 are independent and $f(X_1, X_2)$ is a deterministic function. If*
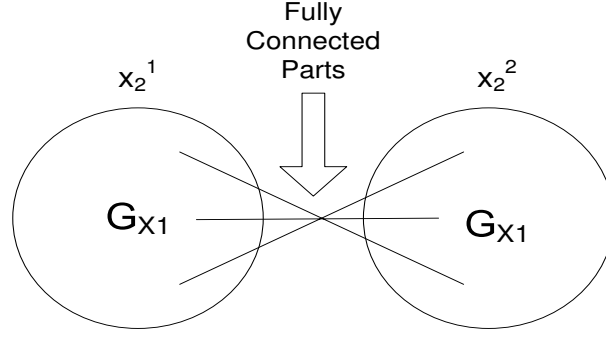
Fig. 14. An example of G_{X_1, X_2} satisfying conditions of Lemma 44, when $\mathcal{X}_2$ has two members.

for any x_2^1 and x_2^2 in $\mathcal{X}_2$ we have $f(x_1^i, x_2^1) \neq f(x_1^j, x_2^2)$ for any possible i and j, then, (f, X_1, X_2) is a chain-rule proper set.

Proof: We show that under this condition any colorings of the graph G_{X_1, X_2} can be expressed as colorings of G_{X_1} and G_{X_2}, and vice versa. The converse part is straightforward because any colorings of G_{X_1} and G_{X_2} can be viewed as a coloring of G_{X_1, X_2}.

Consider Figure 14 which illustrates conditions of this lemma. Under these conditions, since all x_2 in $\mathcal{X}_2$ have different function values, graph G_{X_1, X_2} can be decomposed to some subgraphs which have the same topology as G_{X_1}, corresponding to each x_2 in $\mathcal{X}_2$. These subgraphs are fully connected to each other under conditions of Corollary 44. Thus, any coloring of this graph can be represented as two colorings of G_{X_1} and G_{X_2}, which is a complete graph. Hence, the minimum entropy coloring of G_{X_1, X_2} is equal to the minimum entropy coloring of (G_{X_1}, G_{X_2}). Therefore, $H_{G_{X_1}, G_{X_2}}(X_1, X_2) = H_{G_{X_1, X_2}}(X_1, X_2)$. ∎

III. MULTI-FUNCTIONAL COMPRESSION WITH SIDE INFORMATION

In this section, we consider the problem of multi-functional compression with side information. The problem is how we can compress a source X so that the receiver is able to compute some deterministic functions $f_1(X, Y_1), ..., f_m(X, Y_m)$, where Y_i, $1 \leq i \leq m$, are available at the receiver as side information.

Section II only considers the case where the receiver desires to compute one function (m=1). Here, we consider a case where computation of several functions with different side information is desired at the receiver. Our results do not depend on the fact that all desired functions are in
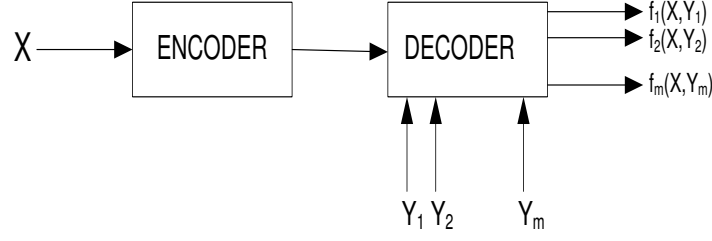
Fig. 15.   A multi-functional compression problem with side information.

one receiver and one can apply them to the case of having several receivers with different desired functions (i.e., functions are separable). We define a new concept named the *multi-functional graph entropy* which is an extension of the graph entropy defined by Körner in [21]. We show that the minimum achievable rate for this problem is equal to the conditional multi-functional graph entropy of the source random variable given side informations. We also propose a coding scheme based on graph colorings to achieve this rate.

A. Problem Setup

Consider discrete memoryless sources (i.e., $\{X^i\}_{i=1}^{\infty}$ and $\{Y_k^i\}_{i=1}^{\infty}$) and assume that these sources are drawn from finite sets $\mathcal{X}$ and $\mathcal{Y}_k$ with a joint distribution $p_k(x, y_k)$. We express n-sequence of these random variables as $\mathbf{X} = \{X^i\}_{i=l}^{i=l+n-1}$ and $\mathbf{Y_k} = \{Y_k^i\}_{i=l}^{i=l+n-1}$ with joint probability distribution $p_k(\mathbf{x}, \mathbf{y}_k)$. Assume $k = 1, ..., m$ (we have m random variables as side information at the receiver).

The receiver wants to compute m deterministic functions $f_k : \mathcal{X} \times \mathcal{Y}_k \to \mathcal{Z}_k$ or $f_k : \mathcal{X}^n \times \mathcal{Y}_k^n \to \mathcal{Z}_k^n$, its vector extension. Without loss of generality, we assume $l = 1$ and to simplify notations, n will be implied by the context. We have one encoder en_X and m decoders $r_1, ..., r_m$ (one for each function and its corresponding side information). Encoder en_X maps

$$en_X : \mathcal{X}^n \to \{1, ..., 2^{nR}\}, \tag{23}$$

and, each decoder r_k maps

$$r_k : \{1, ..., 2^{nR}\} \times \{1, ..., 2^n\} \to \mathcal{Z}_k^n. \tag{24}$$

The probability of error in decoding f_k is

$$P_{e_k}^n = Pr[(\mathbf{x}, \mathbf{y}_k) : f_k(\mathbf{x}, \mathbf{y}_k) \neq r_k(en_X(\mathbf{x}), \mathbf{y}_k)], \tag{25}$$

and, the total probability of error is

$$P_e^n = 1 - \prod_k (1 - P_{e_k}^n). \tag{26}$$

We declare an error when we have an error in computation of at least one function. A rate R is achievable if $P_e^n \to 0$ when $n \to \infty$. Our aim here is to find the minimum achievable rate.

B. Main Results

Prior work in the functional compression problem consider a case when computation of a function is desired at the receiver (m=1). In this section, we consider a case when computation of several functions is desired at the receiver. As an example, consider the network shown in Figure 15. The receiver wants to compute m functions with different side information random variables. We want to compute the minimum achievable rate for this case. Note that our results do not depend on the fact that all functions are desired in one receiver and one can extend them to the case of having several receivers with different desired functions (i.e., functions are separable.).

First, let us consider the case of $m = 2$. Then, we extend our results to the case of arbitrary m. In this problem, the receiver wants to compute two deterministic functions $f_1(X, Y_1)$ and $f_2(X, Y_2)$, where Y_1 and Y_2 are available at the receiver as side information. We wish to find the minimum achievable rate for this problem.

Let us call $G_{X,f_1} = (V, E_{f_1})$ the characteristic graph of X with respect to Y_1, $p_1(x, y_1)$ and $f_1(X, Y_1)$, and $G_{X,f_2} = (V, E_{f_2})$ the characteristic graph of X with respect to Y_2, $p_2(x, y_2)$ and $f_2(X, Y_2)$. Now, define $G_{X,f_1,f_2} = (V, E_{f_1,f_2})$ such that $E_{f_1,f_2} = E_{f_1} \bigcup E_{f_2}$. In other words, G_{X,f_1,f_2} is the *or* function of G_{X,f_1} and G_{X,f_2}. We call G_{X,f_1,f_2} the *multi-functional characteristic graph* of X.

When we deal with one function, we drop f from notations (as in Section II).

Definition 45. *The multi-functional characteristic graph $G_{X,f_1,f_2} = (V, E_{f_1,f_2})$ of X with respect to Y_1, Y_2, $p_1(x, y_1)$, $p_2(x, y_2)$,and $f_1(x, y_1), f_2(x, y_2)$ is defined as follows:*
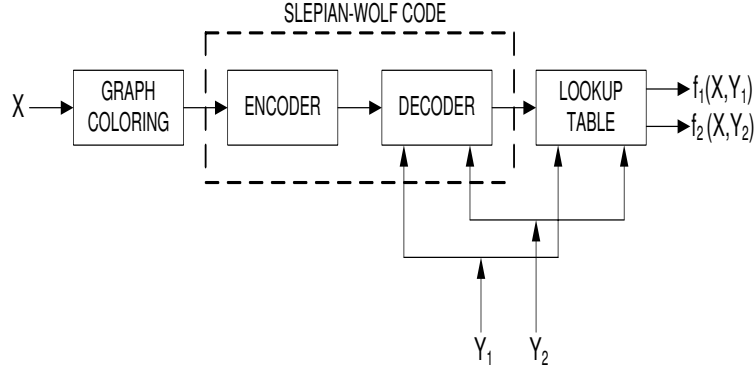
Fig. 16. A source coding scheme for multi-functional compression problem with side information.

$V = \mathcal{X}$ and an edge $(x_1, x_2) \in \mathcal{X}^2$ is in E_{f_1, f_2} iff there exists a $y_1 \in \mathcal{Y}_1$ such that $p_1(x_1, y_1)p_1(x_2, y_1) > 0$ and $f_1(x_1, y_1) \neq f_1(x_2, y_1)$ or there exists a $y_2 \in \mathcal{Y}_2$ such that $p_2(x_1, y_2)p_2(x_2, y_2) > 0$ and $f_2(x_1, y_2) \neq f_2(x_2, y_2)$.

Similarly to Definition 18, we define the *multi-functional graph entropy* as follows:

Theorem 46.

$$H_{G_{X, f_1, f_2}}(X) = \lim_{n \to \infty} \frac{1}{n} H^\chi_{G^n_{X, f_1, f_2}}(\mathbf{X}). \tag{27}$$

The *conditional multi-functional graph entropy* can be defined similarly to Definition 19 as follows:

Theorem 47.

$$H_{G_{X, f_1, f_2}}(X|Y) = \lim_{n \to \infty} \frac{1}{n} H^\chi_{G^n_{X, f_1, f_2}}(\mathbf{X}|\mathbf{Y}), \tag{28}$$

where G^n_{X, f_1, f_2} is the n-th power of G_{X, f_1, f_2}. Now, we can state the following theorem.

Theorem 48. $H_{G_{X, f_1, f_2}}(X|Y_1, Y_2)$ *is the minimum achievable rate for the network shown in Figure 15 when* $m = 2$.

Proof: To show this, we first show that $R_X \geq H_{G_{X, f_1, f_2}}(X|Y_1, Y_2)$ is an achievable rate (achievability), and no one can outperform this rate (converse). To do this, first, we show that

any valid coloring of G_{X,f_1,f_2}^n for any n leads to an achievable encoding-decoding scheme for this problem (achievability). Then, we show that every achievable encoding-decoding scheme performing on blocks with length n, induces a valid coloring of G_{X,f_1,f_2}^n (converse).

Achievablity: According to [3], any valid coloring of G_{X,f_1}^n leads to successfully computing of $f_1(\mathbf{X}, \mathbf{Y}_1)$ at the receiver. If c_{G_{X,f_1}^n} is a valid coloring of G_{X,f_1}^n, there exists a function r_1 such that $r_1(c_{G_{X,f_1}^n}(\mathbf{X}), \mathbf{Y}_1) = f_1(\mathbf{X}, \mathbf{Y}_1)$, with high probability. A similar argument holds for G_{X,f_2}. Now, assume that c_{G_{X,f_1,f_2}^n} is a valid coloring of G_{X,f_1,f_2}^n. Since, $E_{f_1}^n \subseteq E_{f_1,f_2}^n$ and $E_{f_2}^n \subseteq E_{f_1,f_2}^n$, any valid coloring of G_{X,f_1,f_2}^n induces valid colorings for G_{X,f_1}^n and G_{X,f_2}^n. Thus, any valid coloring of G_{X,f_1,f_2}^n leads to successful computation of $f_1(\mathbf{X}, \mathbf{Y}_1)$ and $f_2(\mathbf{X}, \mathbf{Y}_2)$ at the receiver. So, c_{G_{X,f_1,f_2}^n} leads to an achievable encoding scheme (i.e. there exist two functions r_1 and r_2 such that $r_1(c_{G_{X,f_1,f_2}^n}(\mathbf{X}), \mathbf{Y}_1) = f_1(\mathbf{X}, \mathbf{Y}_1)$ and $r_2(c_{G_{X,f_1,f_2}^n}(\mathbf{X}), \mathbf{Y}_2) = f_2(\mathbf{X}, \mathbf{Y}_2)$, with high probability.).

When the receiver wants the whole information of the source node, Slepian and Wolf proposed a technique in [4] to compress source random variable X up to the rate $H(X|Y)$ when Y is available at the receiver. Here, one can perform Slepian-Wolf compression technique on the minimum entropy coloring of large enough power graph and get the given bound.

Converse: Now, we show that any achievable encoding-decoding scheme performing on blocks with length n, induces a valid coloring of G_{X,f_1,f_2}^n. In other words, we want to show that if there exist functions en_X, r_1 and r_2 such that $r_1(en_X(\mathbf{X}), \mathbf{Y}_1) = f_1(\mathbf{X}, \mathbf{Y}_1)$ and $r_2(en_X(\mathbf{X}), \mathbf{Y}_2) = f_2(\mathbf{X}, \mathbf{Y}_2)$, $en_X(\mathbf{X})$ is a valid coloring of G_{X,f_1,f_2}^n.

Let us proceed by contradiction. If $en_X(\mathbf{X})$ were not a valid coloring of G_{X,f_1,f_2}^n, there must be some edge in E_{f_1,f_2}^n with both vertices with the same color. Let us call these two vertices $\mathbf{x}_1$ and $\mathbf{x}_2$ which take the same values (i.e., $en_X(\mathbf{x}_1) = en_X(\mathbf{x}_2)$), but also are connected. Since they are connected to each other, by definition of G_{X,f_1,f_2}^n, there exists a $\mathbf{y}_1 \in \mathcal{Y}_1$ such that $p_1(\mathbf{x}_1, \mathbf{y}_1)p_1(\mathbf{x}_2, \mathbf{y}_1) > 0$ and $f_1(\mathbf{x}_1, \mathbf{y}_1) \neq f_1(\mathbf{x}_2, \mathbf{y}_1)$, or there exists a $\mathbf{y}_2 \in \mathcal{Y}_2$ such that $p_2(\mathbf{x}_1, \mathbf{y}_2)p_2(\mathbf{x}_2, \mathbf{y}_2) > 0$ and $f_2(\mathbf{x}_1, \mathbf{y}_2) \neq f_1(\mathbf{x}_2, \mathbf{y}_2)$. Without loss of generality, assume that the first case occurs. Thus, we have a $\mathbf{y}_1 \in \mathcal{Y}_1$ such that $p_1(\mathbf{x}_1, \mathbf{y}_1)p_1(\mathbf{x}_2, \mathbf{y}_1) > 0$ and $f_1(\mathbf{x}_1, \mathbf{y}_1) \neq f_1(\mathbf{x}_2, \mathbf{y}_1)$. So, $r_1(en_X(\mathbf{x}_1), \mathbf{y}_1) \neq r_1(en_X(\mathbf{x}_2), \mathbf{y}_1)$. Since $en_X(\mathbf{x}_1) = en_X(\mathbf{x}_2)$, then, $r_1(en_X(\mathbf{x}_1), \mathbf{y}_1) \neq r_1(en_X(\mathbf{x}_1), \mathbf{y}_1)$. But, it is not possible. Thus, our contradiction assumption was not true. In other words, any achievable encoding-decoding scheme for this problem induces a valid coloring of G_{X,f_1,f_2}^n and thus completes the proof. ∎

Now, let us consider the network shown in Figure 15, where the receiver wishes to compute m deterministic functions of source information having some side information.

Theorem 49. $H_{G_{X,f_1,...,f_m}}(X|Y_1,...,Y_m)$ *is the minimum achievable rate for the network shown in Figure 15.*

The argument here is similar to the case of $m = 2$ given in Theorem 48. We only sketch the proof. One may first show that any colorings of multi-functional characteristic graph of X with respect to desired functions (i.e., $G_{X,f_1,...,f_m}$) leads to an achievable scheme. Then, showing that any achievable encoding-decoding scheme induces a coloring on $G_{X,f_1,...,f_m}$ completes the proof.

IV. FEEDBACK IN FUNCTIONAL COMPRESSION

In this section, we investigate the effect of having feedback on the rate-region of the functional compression problem. If the function at the receiver is the identity function, this problem is Slepian-Wolf compression with feedback. For this case, having feedback does not improve rate bounds. For example, reference [22] considers both zero-error and asymptotically zero-error Slepian-Wolf compression with feedback. However, for a general desired function at the receiver, having feedback may improve rate bounds of the case without feedback.

A. Main Results

Consider a distributed functional compression problem with two sources and a receiver depicted in Figure 17-a. This network does not have feedback. In Section II, we derived a rate-region for this network. In this section, we consider the effect of having feedback on the rate-region of the network. For simplicity, we consider a simple distributed network topology with two sources. However, one can extend all discussions to more general networks of the type considered in Sections II and III.

Consider the network shown in Figure 17-b. If the desired function at the receiver is the identity function, this problem is Slepian-Wolf compression with feedback. For this case, having feedback does not change the rate region ([4] and [22]). However, when we have a general function at the receiver, by having feedback, one may improve the rate bounds of Theorem 38.

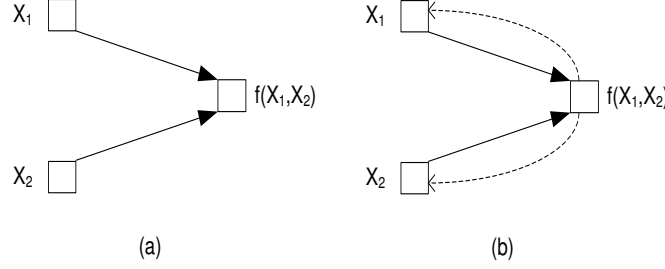**Theorem 50.** *Having feedback may improve rate bounds of Theorem 38.*

Fig. 17. A distributed functional compression network a) without feedback b) with feedback.

Proof: Consider a network without feedback depicted in Figure 17-a. In Section II, we showed an achievable scheme where sources send their minimum entropy colorings of high probability subgraphs of their characteristic graphs satisfying C.C.C., followed by Slepian-Wolf compression. This scheme performs arbitrarily closely to rate bounds derived in Theorem 38. Now, we seek to show that, in some cases, by having feedback, one can outperform these bounds. Consider source random variables X_1 and X_2 with characteristic graphs G_{X_1} and G_{X_2}, respectively. Suppose $S_{c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min}}$ and $S_{c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}}$ are two sets of joint colorings of source random variables defined as follows,

$$
S_{c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min}} = \underset{(c_{G_{\mathbf{X}_1}^n}, c_{G_{\mathbf{X}_2}^n}) \in C_{G_{\mathbf{X}_1}^n} \times C_{G_{\mathbf{X}_2}^n}}{\arg \min} \frac{1}{n} H(c_{G_{\mathbf{X}_1}^n}, c_{G_{\mathbf{X}_2}^n})
$$

$$
S_{c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}} = \underset{\substack{(c_{G_{\mathbf{X}_1}^n}, c_{G_{\mathbf{X}_2}^n}) \in C_{G_{\mathbf{X}_1}^n} \times C_{G_{\mathbf{X}_2}^n} \\ \text{satisfying C.C.C.}}}{\arg \min} \frac{1}{n} H(c_{G_{\mathbf{X}_1}^n}, c_{G_{\mathbf{X}_2}^n}). \tag{29}
$$

Now, consider the case when $S_{c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min}} \cap S_{c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}} = \emptyset$, i.e., suppose any $c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min} \in S_{c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min}}$ does not satisfy C.C.C. Thus, C.C.C. restricts the link sum rates of any achievable scheme, because $H(c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min}) < H(c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n})$ for any $c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min} \in S_{c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min}}$ and $c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n} \in S_{c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}}$.

Choose any two joint colorings $c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min} \in S_{c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min}}$ and $c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n} \in S_{c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}}$. Suppose set A contains all points $(\mathbf{x}_1, \mathbf{x}_2)$ such that their corresponding colors in the joint-coloring class of $c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min}$ do not satisfy C.C.C. Now, we propose a coding scheme with feedback which can outperform rate bounds of the case without having feedback. If sources know whether or not
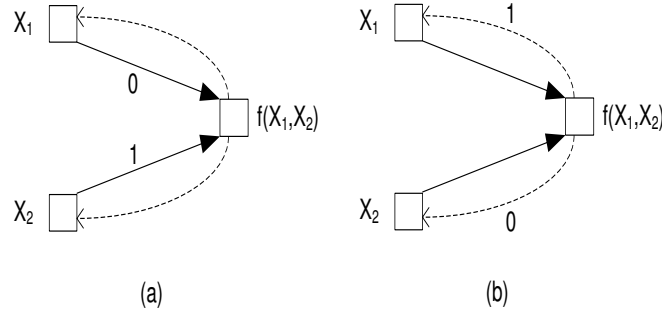
Fig. 18. An example of the proposed feedback scheme. a) Since $\mathbf{x}_1 \notin A_{X_1}$, source X_1 sends 0. Since $\mathbf{x}_2 \in A_{X_2}$, source X_2 sends 1. b) The receiver forward signaling bits to the sources. Then, sources can use the coloring scheme $c^{min}_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}$.

they have some sequences in A, they can switch between $c^{min}_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}$ and $c'_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}$ in their coding scheme with feedback. Since $H(c^{min}_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}) < H(c'_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}})$, this approach outperforms the one without feedback in terms of rates. In the following, we present a possible feedback scheme.

Before sending each sequence, sources first check if their sequences belong to A or not. Say A_{X_1} is the set of all $\mathbf{x}_1$ such that there exists a $\mathbf{x}_2$ such that $(\mathbf{x}_1, \mathbf{x}_2) \in A$. A_{X_2} is defined similarly. One can see that $A \subseteq A_{X_1} \times A_{X_2}$. So, instead of checking if a sequence is in A or not, by exchanging some information, sources check if the sequence belongs to $A_{X_1} \times A_{X_2}$ or not. In order to do this, source X_1 sends a one to the receiver when $\mathbf{x}_1 \in A_{X_1}$. Otherwise, it sends a zero. Source X_2 uses a similar scheme. The receiver exchanges these bits using feedback channels. When a source sends a one, and receives a one from its feedback channel, it uses $c'_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}$ as its joint coloring. Otherwise, it uses $c^{min}_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}$ in its coding scheme. Depending on which joint coloring scheme has been used by sources, the receiver uses a corresponding look-up table to compute the desired function. Hence, this scheme is achievable. An example of this scheme is depicted in Figure 18.

Since the length of sequences is arbitrarily large, one can ignore these four extra signaling bits in rate computation. If we did not have feedback, according to Theorem 38,

$$R_{11} + R_{12} \geq \frac{1}{n} H(c'_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}). \tag{30}$$

Say $P_a = Pr[(\mathbf{x}_1, \mathbf{x}_2) \in A_{X_1} \times A_{X_2}]$. Thus, for the proposed coding scheme with feedback,

we have,

$$R_{11}^f + R_{12}^f \geq \frac{1}{n}[P_a H(c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}) + (1 - P_a)H(c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min})] \tag{31}$$

where R_{1i}^f is the transmission rate of source i with feedback. Thus,

$$[R_{11}^f + R_{12}^f] - [R_{11} + R_{12}] \geq \frac{1}{n}(1 - P_a)[H(c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min}) - H(c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n})]. \tag{32}$$

The right-hand side of (32) represents a gain in link sum rates by having feedback. When, $P_a \neq 1$ and $c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min} \neq c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}$, this is strictly positive, which means the proposed coding scheme with feedback outperforms the one without having feedback in terms of rate bounds. For the identity function at the receiver, $c_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}^{min} = c'_{G_{\mathbf{X}_1}^n, G_{\mathbf{X}_2}^n}$, and the proposed coding scheme with feedback does not improve rate bounds. Note that, for the identity function at the receiver, Slepian-Wolf compression can perform arbitrarily closely to min-cut max-flow bounds. ∎

Note that, in a general network, for cases where the minimum entropy colorings of sources satisfy C.C.C., it is not known whether or not feedback can improve rate bounds.

V. A RATE-DISTORTION REGION FOR DISTRIBUTED FUNCTIONAL COMPRESSION

In this section, we consider the problem of distributed functional compression with distortion. The objective is to compress correlated discrete sources so that an arbitrary deterministic function of those sources can be computed up to a distortion level at the receiver. In this section, we derive a rate-distortion region for a network with two transmitters and a receiver. All discussions can be extended to more general networks considered in Sections II and III.

A recent result is presented in [3] which computes a rate-distortion region for the side information problem. The result in [3] gives a characterization of Yamamoto's rate distortion function [10] in terms of a reconstruction function. Here, we extend these results to the distributed functional compression problem. In this case, we compute a rate-distortion region and then, propose a practical coding scheme with a non-trivial performance guarantee. Note that this proposed characterization is not a single letter characterization.

A. Problem Setup

Consider two sources as described in Section V-A. The receiver wants to compute a deterministic function $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Z}$ or $f : \mathcal{X}_1^n \times \mathcal{X}_2^n \to \mathcal{Z}^n$, its vector extension up to distortion

D with respect to a given distortion function $d : \mathcal{Z} \times \mathcal{Z} \to [0, \infty)$. A vector extension of the distortion function is defined as follows:

$$d(\mathbf{z}_1, \mathbf{z}_2) = \frac{1}{n} \sum_{i=1}^{n} d(z_{1i}, z_{2i}), \tag{33}$$

where $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^n$. As in [9], we assume that $d(z_1, z_2) = 0$ if and only if $z_1 = z_2$. This assumption causes vector extension to satisfy the same property (i.e., $d(\mathbf{z}_1, \mathbf{z}_2) = 0$ if and only if $\mathbf{z}_1 = \mathbf{z}_2$).

Consider the network depicted in Figure 1-b. The sources encode their data at rates R_{11} and R_{12} by using encoders en_{X_1} and en_{X_2}, respectively . The receiver decodes the received data by using decoder r. Hence, we have:

$$en_{X_1} : \mathcal{X}_1^n \to \{1, ..., 2^{nR_{11}}\}$$

$$en_{X_2} : \mathcal{X}_2^n \to \{1, ..., 2^{nR_{12}}\}$$

and a decoder maps,

$$r : \{1, ..., 2^{nR_{11}}\} \times \{1, ..., 2^{nR_{12}}\} \to \mathcal{Z}^n.$$

The probability of error is

$$P_e^n = \Pr[\{(\mathbf{x}_1, \mathbf{x}_2) : d(f(\mathbf{x}_1, \mathbf{x}_2), r(en_{X_1}(\mathbf{x}_1), en_{X_2}(\mathbf{x}_2))) > D\}].$$

We say a rate pair (R_{11}, R_{12}) is achievable up to distortion D if there exist en_{X_1}, en_{X_2} and r such that $P_e^n \to 0$ when $n \to \infty$.

Our aim is to find feasible rates for different links of the network shown in Figure 1-b when the receiver wants to compute $f(X_1, X_2)$ up to distortion D.

B. Prior Results

In this part, we overview prior relevant work. Consider the network shown in Figure 1-a. For this network, in [10], Yamamoto gives a characterization of a rate-distortion function for the side information functional compression problem (i.e., X_2 is available at the receiver). The

TABLE I
RESEARCH PROGRESS ON NONZERO-DISTORTION SOURCE CODING PROBLEMS

Problem types	$f(X_1, X_2) = (X_1, X_2)$	General $f(X_1, X_2)$
Side information	Wyner and Ziv [9]	Feng et al. [11] Yamamoto [10] Doshi et al. [3]
Distributed	Coleman et al. [6] Berger and Yeung [12] Barros and Servetto [13] Wagner et al. [14]	*

rate-distortion function proposed in [10] is a generalization of the Wyner-Ziv side-information rate-distortion function [9]. Specifically, Yamamoto gives the rate distortion function as follows:

Theorem 51. *The rate distortion function for the functional compression problem with side information is*

$$R(D) = \min_{p \in \mathcal{P}(D)} I(W_1; X_1 | X_2)$$

where $\mathcal{P}(D)$ is the collection of all distributions on W_1 given X_1 such that there exists a $g : \mathcal{W}_1 \times \mathcal{X}_2 \to \mathcal{Z}$ satisfying $E[d(f(\mathbf{X}_1, \mathbf{X}_2), g(\mathbf{W}_1, \mathbf{X}_2))] \leq D$.

This is an extension of the Wyner-Ziv rate-distortion result [9]. Further, the variable $W_1 \in \Gamma(G_{X_1})$ in the definition of the Orlitsky-Roche rate, Definition 10, (a variable over the independent sets of G_{X_1}) can be seen as an interpretation of Yamamoto's auxiliary variable, W_1, for the zero-distortion case.

A new characterization of the rate distortion function given by Yamamoto was discussed in [3]. It was shown in [3] that finding a suitable reconstruction function, $\hat{f}$, is equivalent to find g on $\mathcal{W}_1 \times \mathcal{X}_2$ from Theorem 51. Let $\mathcal{F}_m(D)$ denote the set of all functions $\hat{f}_m : \mathcal{X}_1^m \times \mathcal{X}_2^m \to \mathcal{Z}^m$ such that

$$\lim_{n \to \infty} E[d(f(\mathbf{X}_1, \mathbf{X}_2), \hat{f}_m(\mathbf{X}_1, \mathbf{X}_2))] \leq D,$$

and let $\mathcal{F}(D) = \bigcup_{m \in N} \mathcal{F}_m(D)$. Also, let $G_{X_1, \hat{f}}$ denote the characteristic graph of $\mathbf{X}_1$ with respect to $\mathbf{X}_2$, $p(\mathbf{x}_1, \mathbf{x}_2)$, and $\hat{f}$ for any $\hat{f} \in \mathcal{F}(D)$. For each m and all functions $\hat{f} \in \mathcal{F}(D)$, denote for brevity the normalized graph entropy $\frac{1}{m} H_{G_{X_1, \hat{f}}}(\mathbf{X}_1 | \mathbf{X}_2)$ as $H_{G_{X_1, \hat{f}}}(X_1 | X_2)$. The following

theorem was given in [3]:

Theorem 52. *A rate distortion function for the network shown in Figure 1-a can be expressed as follows:*

$$R(D) = \inf_{\hat{f} \in \mathcal{F}(D)} H_{G_{X_1, \hat{f}}}(X_1 | X_2).$$

The problem of finding an appropriate function $\hat{f}$ is equivalent to finding a new graph whose edges are a subset of the edges of the characteristic graph. A graph parameterization by D was proposed in [3] to look at a subset of $\mathcal{F}(D)$. The resulting bound is not tight, but it provides a practical technique to tackle a very difficult problem.

Define the D-characteristic graph of X_1 with respect to X_2, $p(x_1, x_2)$, and $f(X_1, X_2)$, as having vertices $V = \mathcal{X}_1$ and the pair (x_1^1, x_1^2) is an edge if there exists some $x_2^1 \in \mathcal{X}_2$ such that $p(x_1^1, x_2^1)p(x_1^2, x_2^1) > 0$ and $d(f(x_1^1, x_2^1), f(x_1^2, x_2^1)) > D$. We call this graph as $G_{X_1}(D)$. Because $d(z_1, z_2) = 0$ if and only if $z_1 = z_2$, the 0-characteristic graph is the characteristic graph (i.e., $G_{X_1}(0) = G_{X_1}$). The following corollary was given in [3]:

Corollary 53. *The rate $H_{G_{X_1}(D)}(X_1 | X_2)$ is achievable.*

C. Main Results

This section contains our contributions in this problem. Our aim is to find a rate-distortion region for the network shown in Figure 1-b. Recall the Yamamoto rate distortion function (Theorem 51) and Theorem 52. These theorems explain a rate distortion function for the side information problem. Now, we are considering the case when we have distributed functional compression.

Again, for any m, let $\mathcal{F}_m(D)$ denote the set of all functions $\hat{f}_m : \mathcal{X}_1^m \times \mathcal{X}_2^m \to \mathcal{Z}^m$ such that

$$\lim_{n \to \infty} E[d(f(\mathbf{X}_1, \mathbf{X}_2), \hat{f}_m(\mathbf{X}_1, \mathbf{X}_2))] \le D.$$

In other words, we consider n blocks of m-vectors; thus, the functions in the expectation above will be on $\mathcal{X}_1^{mn} \times \mathcal{X}_2^{mn}$. Let $\mathcal{F}(D) = \bigcup_{m \in N} \mathcal{F}_m(D)$. Let $G_{X_1, \hat{f}}$ denote the characteristic graph of $\mathbf{X}_1$ with respect to $\mathbf{X}_2$, $p(\mathbf{x}_1, \mathbf{x}_2)$, and $\hat{f}$ for any $\hat{f} \in \mathcal{F}(D)$ and $G_{X_2, \hat{f}}$ denote the characteristic graph of $\mathbf{X}_2$ with respect to $\mathbf{X}_1$, $p(\mathbf{x}_1, \mathbf{x}_2)$, and $\hat{f}$ for any $\hat{f} \in \mathcal{F}(D)$. For each m and all functions

$\hat{f} \in \mathcal{F}(D)$, denote for brevity normalized graph entropies $\frac{1}{m} H_{G_{X_1,\hat{f}}}(\mathbf{X}_1|\mathbf{X}_2)$ as $H_{G_{X_1,\hat{f}}}(X_1|X_2)$, $\frac{1}{m} H_{G_{X_2,\hat{f}}}(\mathbf{X}_2|\mathbf{X}_1)$ as $H_{G_{X_2,\hat{f}}}(X_2|X_1)$ and $\frac{1}{m} H_{G_{X_1,\hat{f}},G_{X_2,\hat{f}}}(\mathbf{X}_1,\mathbf{X}_2)$ as $H_{G_{X_1,\hat{f}},G_{X_2,\hat{f}}}(X_1,X_2)$.

Now, for a specific function $\hat{f} \in \mathcal{F}(D)$, define $R_{\hat{f}}(D) = (R_{11}^{\hat{f}}(D), R_{12}^{\hat{f}}(D))$ such that,

$$
\begin{aligned}
R_{11}^{\hat{f}} &\geq H_{G_{X_1,\hat{f}}}(X_1|X_2) \\
R_{12}^{\hat{f}} &\geq H_{G_{X_2,\hat{f}}}(X_2|X_1) \\
R_{11}^{\hat{f}} + R_{12}^{\hat{f}} &\geq H_{G_{X_1,\hat{f}},G_{X_2,\hat{f}}}(X_1,X_2).
\end{aligned}
\tag{34}
$$

Theorem 54. *A rate-distortion region for the network shown in Figure 1-b is determined by* $\bigcup_{\hat{f}\in\mathcal{F}(D)} R_{\hat{f}}(D)$.

Proof: We want to show that $\bigcup_{\hat{f}\in\mathcal{F}(D)} R_{\hat{f}}(D)$ determines a rate-distortion region for the considered network. We first show this rate-distortion region is achievable for any $\hat{f} \in \mathcal{F}(D)$, and then we prove every achievable rate region is a subregion of it (converse).

According to Theorem 38, $R_{\hat{f}}(D)$ is sufficient to determine the function $\hat{f}(\mathbf{X}_1, \mathbf{X}_2)$ at the receiver. Also, by definition,

$$
\lim_{n\to\infty} E[d(f(\mathbf{X}_1, \mathbf{X}_2), \hat{f}(\mathbf{X}_1, \mathbf{X}_2))] \leq D.
$$

Thus, for a specific $\hat{f} \in \mathcal{F}(D)$, $R_{\hat{f}}(D)$ is achievable. Therefore, the union of these achievable regions for different $\hat{f} \in \mathcal{F}(D)$ (i.e., $\bigcup_{\hat{f}\in\mathcal{F}(D)} R_{\hat{f}}(D)$) is also achievable.

Next, we show that any achievable rate region is a subregion of $\bigcup_{\hat{f}\in\mathcal{F}(D)} R_{\hat{f}}(D)$. Assume that we have an achievable scheme in which source 1 encodes its data to $en_{X_1}(\mathbf{X}_1)$ and source 2 encodes its data to $en_{X_2}(\mathbf{X}_2)$. At the receiver, we compute $r(en_{X_1}(\mathbf{X}_1), en_{X_2}(\mathbf{X}_2))$. Since it is an achievable scheme up to distortion D, there exists $\hat{f} \in \mathcal{F}(D)$ such that $r(en_{X_1}(\mathbf{X}_1), en_{X_2}(\mathbf{X}_2)) = \hat{f}(\mathbf{X}_1, \mathbf{X}_2)$. Thus, considering Theorem 38, this achievable rate-distortion region is a subregion of $\bigcup_{\hat{f}\in\mathcal{F}(D)} R_{\hat{f}}(D)$. This completes the proof. ∎

Next, we present a simple scheme which satisfies Theorem 54. Again, the problem of finding an appropriate function $\hat{f}$ is equivalent to finding a new graph whose edges are a subset of the edges of the characteristic graph of random variables. This motivates Corollary 55 where we use a similar graph parameterization by D. Our scheme is as follows:

Define the D-characteristic graph of X_1 with respect to X_2, $p(x_1, x_2)$, and $f(X_1, X_2)$, as

having vertices $V = \mathcal{X}_1$ and the pair (x_1^1, x_1^2) is an edge if there exists some $x_2^1 \in \mathcal{X}_2$ such that $p(x_1^1, x_2^1)p(x_1^2, x_2^1) > 0$ and $d(f(x_1^1, x_2^1), f(x_1^2, x_2^1)) > D$. Denote this graph as $G_{X_1}(D)$. Similarly, we define $G_{X_2}(D)$. Following Corollary 53 and Theorem 54, we have the following Corollary:

Corollary 55. *For independent sources, if the distortion function is a metric and (R_{11}, R_{12}) satisfies the following conditions, then, (R_{11}, R_{12}) is achievable.*

$$
\begin{aligned}
R_{11} &\geq H_{G_{X_1}(D/2)}(X_1) & (35) \\
R_{12} &\geq H_{G_{X_2}(D/2)}(X_2) \\
R_{11} + R_{12} &\geq H_{G_{X_1}(D/2), G_{X_2}(D/2)}(X_1, X_2).
\end{aligned}
$$

Proof: From Theorem 38, by sending colorings of high probability subgraphs of sources's $D/2$-characteristic graphs satisfying C.C.C., one can achieve the rate region described in (35). For simplicity, we assume the power of the graphs is one. Extensions to an arbitrary power are analogous. Suppose the receiver gets two colors from sources (say c_1 from source 1, and c_2 from source 2). To show that the receiver is able to compute its desired function up to distortion level D, we need to show that for every (x_1^1, x_2^1) and (x_1^2, x_2^2) such that $C_{G_{X_1}(D/2)}(x_1^1) = C_{G_{X_1}(D/2)}(x_1^2)$ and $C_{G_{X_2}(D/2)}(x_2^1) = C_{G_{X_2}(D/2)}(x_2^2)$, we have $d(f(x_1^1, x_2^1), f(x_1^2, x_2^2)) \leq D$. Since the distortion function d is a metric, we have,

$$
\begin{aligned}
d(f(x_1^1, x_2^1), f(x_1^2, x_2^2)) &\leq d(f(x_1^1, x_2^1), f(x_1^2, x_2^1)) + d(f(x_1^2, x_2^1), f(x_1^2, x_2^2)) \\
&\leq D/2 + D/2 = D. & (36)
\end{aligned}
$$

This completes the proof. ∎

VI. POLYNOMIAL TIME CASES FOR FINDING THE MINIMUM ENTROPY COLORING OF A CHARACTERISTIC GRAPH

In this section, we consider the problem of finding the minimum entropy coloring of a characteristic graph. This problem arises in the functional compression problem where computation of a function of sources is desired at the receiver. We considered some aspects of this problem in Sections [II- V] and proposed some coding schemes. In those proposed coding schemes, one

needs to compute the minimum entropy coloring (a coloring random variable which minimizes the entropy) of a characteristic graph. In general, finding this coloring is an NP-hard problem (as shown by Cardinal et al. [15]) . However, in this section, we show that depending on the characteristic graph's structure, there are some interesting cases where finding the minimum entropy coloring is not NP-hard, but tractable and practical. In one of these cases, we show that, having a non-zero joint probability condition on random variables' distributions, for any desired function f, makes characteristic graphs to be formed of some non-overlapping fully-connected maximal independent sets. Therefore, the minimum entropy coloring can be solved in polynomial time. In another case, we show that, if the function we seek to compute is a type of quantization functions, this problem is also tractable.

We also consider this problem in a general case. By using Huffman or Lempel-Ziv coding notions, we heuristically relate finding the minimum entropy coloring to finding the maximum independent set of a graph. While the minimum-entropy coloring problem is a recently studied problem, there are some heuristic algorithms to solve approximately the maximum independent set problem.

We proceed this section by stating the problem setup. Then, we explain our contributions to this problem.

A. Problem Setup

In some problems such as the functional compression problem, we need to find a coloring random variable of a characteristic graph which minimizes the entropy. The problem is how to compute such a coloring for a given characteristic graph.

Given a characteristic graph G_{X_1} (or, its n-th power, $G^n_{\mathbf{X}_1}$), one can assign different colors to its vertices. Suppose $C_{G_{X_1}}$ is the collection of all valid colorings of this graph. Among these colorings, one which minimizes the entropy of the coloring random variable is called the minimum-entropy coloring, and we refer to it by $c^{min}_{G_{X_1}}$:

$$c^{min}_{G_{X_1}} = \operatorname*{arg\,min}_{c_{G_{X_1}} \in C_{G_{X_1}}} H(c_{G_{X_1}}). \tag{37}$$

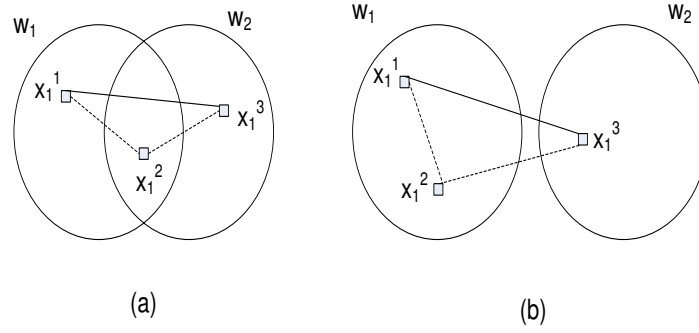The problem is how to compute $c^{min}_{G_{X_1}}$ given G_{X_1}.

Fig. 19. Having non-zero joint probability distribution, a) maximal independent sets cannot overlap with each other (this figure is to depict the contradiction) b) maximal independent sets should be fully connected to each other. In this figure, a solid line represents a connection, and a dashed line means no connection exists.

B. Main Results

In general, finding $c_{G_{X_1}}^{min}$ is an NP-hard problem ([15]). However, in this section, we investigate cases where this coloring can be computed in polynomial time, depending on the characteristic graph's structure. In one of these cases, we show that, by having a non-zero joint probability condition on random variables' distributions, for any desired function, finding $c_{G_{X_1}}^{min}$ can be solved in polynomial time. In another case, we show that, if the function we seek to compute is a quantization function, this problem is also tractable. We also consider this problem in a general case. By using Huffman or Lempel-Ziv coding notions, we heuristically relate finding the minimum entropy coloring to finding the maximum independent set of a graph.

For simplicity, we consider functions with two input random variables, but one can extend all discussions to functions with more input random variables than two.

1) Non-Zero Joint Probability Distribution Condition: Consider the network shown in Figure 1-b. Source random variables have a joint probability distribution $p(x_1, x_2)$, and the receiver wishes to compute a deterministic function of sources (i.e., $f(X_1, X_2)$). In Section II, we showed that, in an achievable coding scheme, one needs to compute minimum entropy colorings of characteristic graphs. The question is how source nodes can compute minimum entropy colorings of their characteristic graphs G_{X_1} and G_{X_2} (or, similarly the minimum entropy colorings of $G_{\mathbf{X}_1}^n$ and $G_{\mathbf{X}_2}^n$, for some n). For an arbitrary graph, this problem is NP-hard ([15]). However, in certain cases, depending on the probability distribution or the desired function, the characteristic graph has some special structure which leads to a tractable scheme to find the minimum entropy

coloring. In this section, we consider the effect of the probability distribution.

Theorem 56. *Suppose for all $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, $p(x_1, x_2) > 0$. Then, maximal independent sets of the characteristic graph G_{X_1} (and, its n-th power $G_{\mathbf{X}_1}^n$, for any n) are some non-overlapping fully-connected sets. Under this condition, the minimum entropy coloring can be achieved by assigning different colors to its different maximal independent sets.*

Proof: Suppose $\Gamma(G_{X_1})$ is the set of all maximal independent sets of G_{X_1}. Let us proceed by contradiction. Consider Figure 19-a. Suppose w_1 and w_2 are two different non-empty maximal independent sets. Without loss of generality, assume x_1^1 and x_1^2 are in w_1, and x_1^2 and x_1^3 are in w_2. These sets have a common element x_1^2. Since w_1 and w_2 are two different maximal independent sets, $x_1^1 \notin w_2$ and $x_1^3 \notin w_1$. Since x_1^1 and x_1^2 are in w_1, there is no edge between them in G_{X_1}. The same argument holds for x_1^2 and x_1^3. But, we have an edge between x_1^1 and x_1^3, because w_1 and w_2 are two different maximal independent sets, and at least there should exist such an edge between them. Now, we want to show that it is not possible.

Since there is no edge between x_1^1 and x_1^2, for any $x_2^1 \in \mathcal{X}_2$, $p(x_1^1, x_2^1)p(x_1^2, x_2^1) > 0$, and $f(x_1^1, x_2^1) = f(x_1^2, x_2^1)$. A similar argument can be expressed for x_1^2 and x_1^3. In other words, for any $x_2^1 \in \mathcal{X}_2$, $p(x_1^2, x_2^1)p(x_1^3, x_2^1) > 0$, and $f(x_1^2, x_2^1) = f(x_1^3, x_2^1)$. Thus, for all $x_2^1 \in \mathcal{X}_2$, $p(x_1^1, x_2^1)p(x_1^3, x_2^1) > 0$, and $f(x_1^1, x_2^1) = f(x_1^3, x_2^1)$. However, since x_1^1 and x_1^3 are connected to each other, there should exist a $x_2^1 \in \mathcal{X}_2$ such that $f(x_1^1, x_2^1) \neq f(x_1^3, x_2^1)$ which is not possible. So, the contradiction assumption is not correct and these two maximal independent sets do not overlap with each other.

We showed that maximal independent sets cannot have overlaps with each other. Now, we want to show that they are also fully connected to each other. Again, let us proceed by contradiction. Consider Figure 19-b. Suppose w_1 and w_2 are two different non-overlapping maximal independent sets. Suppose there exists an element in w_2 (call it x_1^3) which is connected to one of elements in w_1 (call it x_1^1) and is not connected to another element of w_1 (call it x_1^2). By using a similar discussion to the one in the previous paragraph, we may show that it is not possible. Thus, x_1^3 should be connected to x_1^1. Therefore, if for all $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, $p(x_1, x_2) > 0$, then maximal independent sets of G_{X_1} are some separate fully connected sets. In other words, the complement of G_{X_1} is formed by some non-overlapping cliques. Finding the minimum entropy coloring of this graph is trivial and can be achieved by assigning different colors to these non-overlapping
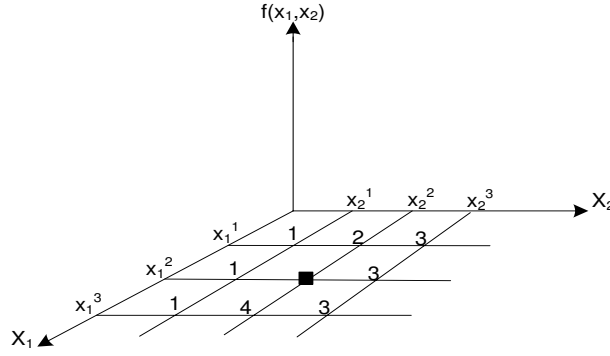
Fig. 20. Having non-zero joint probability condition is necessary for Theorem 56. A dark square represents a zero probability point.

fully-connected maximal independent sets.

This argument also holds for any power of G_{X_1}. Suppose $\mathbf{x}_1^1$, $\mathbf{x}_1^2$ and $\mathbf{x}_1^3$ are some typical sequences in $\mathcal{X}_1^n$. If $\mathbf{x}_1^1$ is not connected to $\mathbf{x}_1^2$ and $\mathbf{x}_1^3$, it is not possible to have $\mathbf{x}_1^2$ and $\mathbf{x}_1^3$ connected. Therefore, one can apply a similar argument to prove the theorem for $G_{\mathbf{X}_1}^n$, for some n. This completes the proof. ∎

Here are some remarks about Theorem 56:

- If the characteristic graph satisfying conditions of Theorem 56 is sparse, its power graph would also remain sparse (a sparse graph with m vertices is a graph whose number of edges is much smaller than $\frac{m(m-1)}{2}$).

- The condition $p(x_1, x_2) > 0$, for all $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, is a necessary condition for Theorem 56. In order to illustrate this, consider Figure 20. In this example, x_1^1, x_1^2 and x_1^3 are in $\mathcal{X}_1$, and x_2^1, x_2^2 and x_2^3 are in $\mathcal{X}_2$. Suppose $p(x_1^2, x_2^2) = 0$. By considering the value of function f at these points depicted in the figure, one can see that, in G_{X_1}, x_1^2 is not connected to x_1^1 and x_1^3. However, x_1^1 and x_1^3 are connected to each other. Thus, Theorem 56 does not hold here.

- The condition used in Theorem 56 only restricts the probability distribution and it does not depend on the function f. Thus, for any function f at the receiver, if we have a non-zero joint probability distribution of source random variables (for example, when source random variables are independent), finding the minimum-entropy coloring is easy and tractable.

2) Quantization Functions: In Section VI-B1, we introduced a condition on the joint probability distribution of random variables which leads to a specific structure of the characteristic graph so that finding the minimum entropy coloring is not NP-hard. In this section, we consider some special functions which to lead to some special graph structures.

An interesting function is a quantization function. A natural quantization function is a function which separates the $X_1 - X_2$ plane into some rectangles such that each rectangle corresponds to a different value of that function. Sides of these rectangles are parallel to the plane axes. Figure 21-a depicts such a quantization function.

Given a quantization function, one can extend different sides of each rectangle in the $X_1 - X_2$ plane. This may make some new rectangles. We call each of them *a function region*. Each function region can be determined by two subsets of $\mathcal{X}_1$ and $\mathcal{X}_2$. For example, in Figure 21-b, one of the function regions is distinguished by the shaded area.

Definition 57. *Consider two function regions $\mathcal{X}_1^1 \times \mathcal{X}_2^1$ and $\mathcal{X}_1^2 \times \mathcal{X}_2^2$. If for any $x_1^1 \in \mathcal{X}_1^1$ and $x_1^2 \in \mathcal{X}_1^2$, there exist x_2^1 such that $p(x_1^1, x_2^1)p(x_1^2, x_2^1) > 0$ and $f(x_1^1, x_2^1) \neq f(x_1^2, x_2^1)$, we say these two function regions are pairwise X_1-proper.*

Theorem 58. *Consider a quantization function f such that its function regions are pairwise X_1-proper. Then, G_{X_1} (and $G_{\mathbf{X}_1}^n$, for any n) is formed of some non-overlapping fully-connected maximal independent sets, and its minimum entropy coloring can be achieved by assigning different colors to different maximal independent sets.*

Proof: We first prove it for G_{X_1}. Suppose $\mathcal{X}_1^1 \times \mathcal{X}_2^1$, and $\mathcal{X}_1^2 \times \mathcal{X}_2^2$ are two X_1-proper function regions of a quantization function f, where $\mathcal{X}_1^1 \neq \mathcal{X}_1^2$. We show that $\mathcal{X}_1^1$ and $\mathcal{X}_1^2$ are two non-overlapping fully-connected maximal independent sets. By definition, $\mathcal{X}_1^1$ and $\mathcal{X}_1^2$ are two non-equal partition sets of $\mathcal{X}_1$. Thus, they do not have any element in common.

Now, we want to show that vertices of each of these partition sets are not connected to each other. Without loss of generality, we show it for $\mathcal{X}_1^1$. If this partition set of $\mathcal{X}_1$ has only one element, this is a trivial case. So, suppose x_1^1 and x_1^2 are two elements in $\mathcal{X}_1^1$. By definition of function regions, one can see that, for any $x_2^1 \in \mathcal{X}_2$ such that $p(x_1^1, x_2^1)p(x_1^2, x_2^1) > 0$, then $f(x_1^1, x_2^1) = f(x_1^2, x_2^1)$. Thus, these two vertices are not connected to each other. Now, suppose x_1^3 is an element in $\mathcal{X}_1^2$. Since these function regions are X_1-proper, there should exist at least
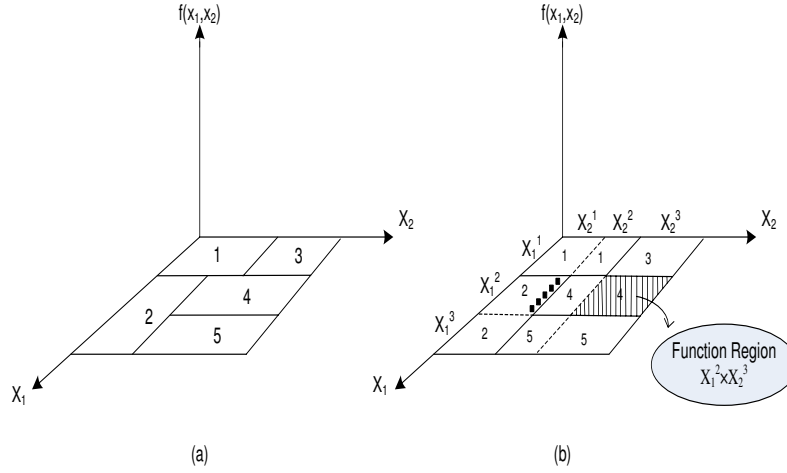
Fig. 21.   a) A quantization function. Function values are depicted in the figure on each rectangle. b) By extending sides of rectangles, the plane is covered by some function regions.

one $x_2^1 \in \mathcal{X}_2$, such that $p(x_1^1, x_2^1)p(x_1^3, x_2^1) > 0$, and $f(x_1^1, x_2^1) \neq f(x_1^3, x_2^1)$. Thus, x_1^1 and x_1^3 are connected to each other. Therefore, $\mathcal{X}_1^1$ and $\mathcal{X}_1^2$ are two non-overlapping fully-connected maximal independent sets. One can easily apply this argument to other partition sets. Thus, the minimum entropy coloring can be achieved by assigning different colors to different maximal independent sets (partition sets). The proof for $G_{\mathbf{X}_1}^n$, for any n, is similar to the one mentioned in Theorem 56. This completes the proof.   ∎

Note that without X_1-proper condition of Theorem 58, assigning different colors to different partitions still leads to an achievable coloring scheme. However, it is not necessarily the minimum entropy coloring. In other words, without this condition, maximal independent sets may overlap.

Corollary 59. *If a function f is strictly monotonic with respect to X_1, and $p(x_1, x_2) \neq 0$, for all $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, then, G_{X_1} (and, $G_{\mathbf{X}_1}^n$ for any n) is a complete graph.*

Under conditions of Corollary 59, functional compression does not give us any gain, because, in a complete graph, one should assign different colors to different vertices. Traditional compression where f is the identity function is a special case of Corollary 59.

3) Minimum Entropy Coloring for an Arbitrary Graph: Finding the minimum entropy coloring of an arbitrary graph (called the chromatic entropy) is NP-hard ([15]). Reference [15] showed that, even finding a coloring whose entropy is within $(\frac{1}{7} - \epsilon) \log m$ of its chromatic entropy

is NP-hard, for any $\epsilon > 0$, where m is the number of vertices of the graph. That is a reason we introduced some special structures on the characteristic graph to have some tractable and practical schemes to find the minimum entropy coloring. While cases investigated in Sections VI-B1 and VI-B2 cover certain practical cases, in this part, we want to consider this problem without assuming any special structure of the graph. In particular, we show that, by using a notion of an empirical Huffman coding scheme or a Lempel-Ziv coding scheme, one can heuristically relate finding the minimum-entropy coloring problem and finding the maximum independent set problem. While the minimum-entropy coloring problem is a recently studied problem, there are some heuristic algorithms to solve the maximum independent set problem [29].

Suppose G_{X_1} is the characteristic graph of X_1. Without loss of generality, in this section, we consider $n = 1$. All discussions can be extended to $G_{\mathbf{X}_1}^n$, for any n. Suppose $p(x_1)$ is the probability distribution of X_1. Let us define the adjacency matrix $A = [a_{ij}]$ for this graph as follows: $a_{ij} = 1$ when x_1^i and x_1^j are connected to each other in G_{X_1}, otherwise, $a_{ij} = 0$. One can see that the adjacency matrix is symmetric, with all zeros in its diagonal. A *one* in this matrix means that its corresponding vertices should be assigned to different colors.

Let us define a permutation matrix P with the same size of A. This matrix has only a *one* in each of its rows and columns. The matrix PAP^t would be a matrix such that rows and columns of A are reordered simultaneously, with respect to this permutation matrix P. For any valid coloring, there exists a permutation matrix P, such that PAP^t has zero square matrices on its diagonal. This reordering is such that, vertices with the same color are adjacent to each other in PAP^t. Each of these zero square matrices on the diagonal of PAP^t represents a maximal independent set, or equivalently a color class. One can see that there exists a bijective mapping between any valid coloring and any permutation matrix P which leads to have some zero square matrices on the diagonal of PAP^t.

Example 60. *For an example, consider the coloring of the graph depicted in Figure 7. This coloring leads to the following PAP^t matrix:*

$$PAP^t = \begin{pmatrix} 0 & 0 & & & D_1 & D_2 \\ 0 & 0 & & & & \\ & & 0 & 0 & & D_3 \\ D_1 & & 0 & 0 & & \\ D_2 & & D_3 & & 0 & \end{pmatrix} \tag{38}$$

where D_i, $i = 1, 2, 3$ are non-zero matrices. Each of zero square matrices on the diagonal represents a color class, or a maximal independent set of this graph. The permutation matrix P in this case is,

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{39}$$

Now, we want to take the probability distribution into account. To do this, we repeat each vertex x_1^i in the adjacency matrix, n_i times, such that $\frac{p(x_1^i)}{p(x_1^j)} = \frac{n_i}{n_j}$, for any valid i and j. We call the achieved matrix, the weighted adjacency matrix and denote it by A_w. The above argument about the permutation remains the same. Any valid coloring can be represented by a permutation matrix P such that PA_wP^t has some zero square matrices on its diagonal. Since we represent the probability distribution of each vertex as its number of repetitions in A_w, the proportional sizes of zero square matrices on the diagonal of PA_wP^t represent the corresponding probability of that color class. In other words, a color class of a larger zero square matrix has more probability than a color class with a smaller zero square matrix.

Now, one can heuristically use Huffman coding technique to find a coloring (or its corresponding permutation matrix) to minimize the entropy. To do this, we first find a permutation matrix which leads to the largest zero square matrix on the diagonal of PA_wP^t. Then, we assign a color to that class, and eliminate its corresponding rows and columns. We repeat this algorithm till all vertices are assigned to some colors. One can see that, finding the largest zero square matrix on the diagonal of PA_wP^t is equivalent to finding the maximum independent set of a graph. Note that, it is a heuristic algorithm, and does not necessarily reach to the minimum entropy coloring. The other point is that, here, we have assumed that the probability distribution of X_1 is known.

If we do not know this probability distribution, one can use an empirical distribution, instead of the actual distribution. In that case, using a Lempel-Ziv coding notion instead of Huffman coding leads to a similar algorithm.

VII. Conclusions and Future Work

In this paper, we considered different aspects of the functional compression problem where computing a function (or, some functions) of sources is desired at the receiver(s). The rate region of this problem has been considered in the literature under certain restrictive assumptions, particularly in terms of the network topology, the functions and the characteristics of the sources. In this paper, we significantly relaxed these assumptions. In Section II of this paper, we considered this problem for an arbitrary tree network and asymptotically lossless computation and derived rate lower bounds. We showed that, for one stage tree networks with correlated sources, or for general trees with independent sources, these lower bounds are tight. For these cases, we proposed a modularized coding scheme based on graph colorings and Slepian-Wolf compression which performs arbitrarily closely to rate lower bounds. Optimal computations that should be performed at intermediate nodes are derived, for a general tree network with independent sources. We showed that, for a family of functions and random variables called chain rule proper sets, computation at intermediate nodes is not necessary. We also introduced a new condition on colorings of source random variables' characteristic graphs called the coloring connectivity condition (C.C.C.) and showed that, unlike the condition mentioned in Doshi *et al.*, this condition is necessary and sufficient for any achievable coding scheme based on colorings. We also showed that, unlike entropy, graph entropy does not satisfy the chain rule.

The problem of having different desired functions with side information at the receiver was considered in Section III. For this problem, we defined a new concept named multi-functional graph entropy, an extension of graph entropy defined by Körner to the multi-functional case. We showed that, the minimum achievable rate for this problem with side information is equal to conditional multi-functional graph entropy of the source random variable given the side information. We also proposed a coding scheme based on graph colorings to achieve this rate.

In Section IV, we investigated the effect of having feedback on the rate region of the functional compression problem. If the function at the receiver is the identity function, this problem reduces to the Slepian-Wolf compression with feedback, for which having feedback does not increase

the rate. However, we showed that, in general, feedback can improve rate bounds.

The problem of distributed functional compression with distortion was investigated in Section V. The objective is to compress correlated discrete sources so that an arbitrary deterministic function of those sources can be computed within a distortion level at the receiver. In this case, we computed a rate-distortion region for this problem which is not a single letter characterization. Then, we proposed a simple suboptimal coding scheme with a non-trivial performance guarantee.

In these proposed coding schemes, one needs to compute the minimum entropy coloring of a characteristic graph. In general, finding this coloring is an NP-hard problem ([15]). However, in Section VI, we showed that depending on the characteristic graph's structure, there are certain cases where finding the minimum entropy coloring is not NP-hard, but tractable and practical. In one of these cases, we showed that, by having a non-zero joint probability condition on random variables' distributions, for any desired function, finding the minimum entropy coloring can be solved in polynomial time. In another case, we showed that, if the desired function is a type of quantization functions, this problem is also tractable.

For possible future work, one may consider a general network topology rather than tree networks. For instance, one can consider a general multi-source multicast network in which receivers desire to have a deterministic function of source random variables. For the case of having the identity function at the receivers, this problem is well-studied in [26], [27] and [28] under the name of network coding for multi-source multicast networks. Reference [28] shows that, random linear network coding can perform arbitrarily closely to min-cut max-flow bounds. To have an achievable scheme for the functional version of this problem, one may perform random network coding on coloring random variables satisfying C.C.C. If receivers desire different functions, one can use colorings of multi-functional characteristic graphs satisfying C.C.C., and then use random network coding for these coloring random variables. This achievable scheme can be extended to disjoint multicast and disjoint multicast plus multicast cases described in [27]. This scheme is an achievable scheme; however it is not optimal in general. If sources are independent, one may use encoding/decoding functions derived for tree networks at intermediate nodes, along with network coding.

Throughout this paper, we considered asymptotically lossless or lossy computation of a function. For possible future work, one may consider this problem for the zero-error computation of a function which leads to a communication complexity problem. One can use tools and schemes

we have developed in this paper to attain some achievable schemes in the zero error computation case.

REFERENCES

[1] C. E. Shannon, "The zero error capacity of a noisy channel," *IEEE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.

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

[3] V. Doshi, D. Shah, M. Médard, and M. Effros, "Functional compression through graph coloring," *Information Theory, IEEE Transactions on*, vol. 56, no. 8, pp. 3901–3917, 2010.

[4] 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.

[5] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 626–643, Mar. 2003.

[6] T. P. Coleman, A. H. Lee, M. Médard, and M. Effros, "Low-complexity approaches to Slepian-Wolf near-lossless distributed data compression," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3546–3561, Aug. 2006.

[7] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, Nov. 1975.

[8] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 219–221, 1979.

[9] A. 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.

[10] H. Yamamoto, "Wyner-Ziv theory for a general function of the correlated sources," *IEEE Trans. Inf. Theory*, vol. 28, no. 5, pp. 803–807, Sep. 1982.

[11] H. Feng, M. Effros, and S. Savari, "Functional source coding for networks with receiver side information," in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Sep. 2004, pp. 1419–1427.

[12] 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.

[13] J. Barros and S. Servetto, "On the rate-distortion region for separate encoding of correlated sources," IEEE Symposium on Information Theory (ISIT), Yokohama, Japan, 2003, p. 171.

[14] A. Wagner, S. Tavildar, and P. Viswanath, "The rate region of the quadratic Gaussian two-terminal source-coding problem," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 1938–1961, 2008.

[15] J. Cardinal, S. Fiorini, and G. Joret, "Tight results on minimum entropy set cover," *Algorithmica*, vol. 51, no. 1, pp. 49–60, May 2008.

[16] N. Ma and P. Ishwar, "Distributed source coding for interactive function computation," *Submitted to IEEE Transactions on Information Theory*, 2008.

[17] N. Ma, P. Ishwar, and P. Gupta, "Information-theoretic bounds for multiround function computation in collocated networks," in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*. IEEE, 2009, pp. 2306–2310.

[18] A. Ramamoorthy, "Communicating the sum of sources over a network," in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*. IEEE, 2008, pp. 1646–1650.

[19] H. Kowshik and P. R. Kumar, "Optimal computation of symmetric boolean functions in tree networks," in *2010 IEEE International Symposium on Information Theory, ISIT 2010*, TX, US, 2010, pp. 1873–1877.

[20] S. Shenvi and B. K. Dey, "A necessary and sufficient condition for solvability of a 3s/3t sum-network," in *2010 IEEE International Symposium on Information Theory, ISIT 2010*, TX, US, 2010, pp. 1858–1862.

[21] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," 6-th Prague Conference on Information Theory, 1973, pp. 411–425.

[22] M. Bakshi and M. Effros, "On zero-error source coding with feedback," in *2010 IEEE International Symposium on Information Theory, ISIT 2010*, TX, US, 2010.

[23] I. Csiszar and J. Körner, in *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, New York, 1981.

[24] H. S. Witsenhausen, "The zero-error side information problem and chromatic numbers," *IEEE Trans. Inf. Theory*, vol. 22, no. 5, pp. 592–593, Sep. 1976.

[25] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1329–1339, Sep. 1996.

[26] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1204–1216, 2000.

[27] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. on networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[28] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[29] T. Cormen, in *Introduction to algorithms*. MIT Press.