Multi-Functional Compression with Side Information

Soheil Feizi, Muriel Médard
RLE at MIT
Emails: {sfeizi,medard}@mit.edu

***Abstract*—In this paper, 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.**

In [1], Wyner and Ziv considered this problem for the special case of $m = 1$ and $f_1(X, Y_1) = X$ and derived a rate-distortion function. Yamamoto extended this result in [2] to the case of having one general function $f_1(X, Y_1)$. Both of these results were in terms of an auxiliary random variable. For the case of zero distortion, in [3], Orlitsky and Roche gave an interpretation of this variable in terms of properties of the characteristic graph which led to a particular coding scheme. This result was extended in [4] by providing an achievable scheme based on colorings of the characteristic graph. In a recent work, reference [5] has considered this problem for a general tree network where intermediate nodes are allowed to perform some computations.

These previous works only considered the case where the receiver only wants to compute one function (m=1). Here, we want to consider the case in which the receiver wants to compute several functions with different side information random variables and zero distortion. Our results do not depend on the fact that all functions are desired in 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 Korner in [6]. We show that the minimum achievable rate for this problem is equal to the conditional multi-functional graph entropy of random variable X given side informations. We also propose a coding scheme based on graph colorings to achieve this rate.

***Index Terms*—Functional compression, graph Coloring, graph entropy.**

I. Introduction

In this paper, we consider the problem of functional compression for the case of having different desired functions at the receiver. While data compression problem considers the compression of sources at transmitters and their reconstruction at the receivers, functional compression does not consider the recovery of whole sources, but maybe one or some functions of sources are desired at the receiver(s). For instance, consider a network with one transmitter and one receiver. A sensor is measuring the temperature at the source node. The receiver only wants to compare this temperature with two or several thresholds. For instance, if the temperature is above a certain threshold or under another threshold, the receiver may announce an alarm. So, the receiver does not need to have the exact temperature of the source, and only some parts of source information is useful for the receiver. If the source node wants to send whole information without considering this fact, it should send its information at a rate at least equal to its entropy. But, considering this fact that the receiver only needs some part of source information with respect to its desired functions, we want to compute how much the source node can compress its information to have a lossless computation of functions at the receiver.

This work was supported by AFOSR award number FA9550-09-1-0196.

In the functional compression problem with side information, as shown in Figure 2, the receiver wants to compute a deterministic function of source random variable X. Another random variable Y, possibly correlated with X, is available at the receiver. First, Shannon considered this problem in [7] for the case $f(X, Y) = X$ and zero distortion. In [1], Wyner and Ziv derived a rate-distortion function for this problem in the case of $f(X, Y) = X$. Then, in [2], Yamamoto extended this result to the case of having a general function. Both of these results are in terms of an auxiliary random variable. In [3], Orlitsky and Roche considered this problem for the case of zero distortion and gave an interpretation of this variable in terms of properties of characteristic graph which led to a particular coding scheme. This result was extended in [4] by providing an achievable scheme based on colorings of the characteristic graph. In a recent work, [5] has considered this problem for a general tree network where intermediate nodes are allowed to perform some computations.

These previous works only considered the case that the receiver only wants to compute one function (m=1). Here, we want to consider the case that the receiver wants to compute several functions with different side information random variables and zero distortion (Figure 1). It is worthwhile to mention that our results do not depend on the fact that all functions are desired in one receiver. In other words, 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. We show that the minimum achievable rate for this problem is equal to the conditional multi-functional graph entropy of random variable X given the side informations. We also propose a coding scheme based on graph colorings to achieve this rate.

The rest of the paper is organized as follows. Section II

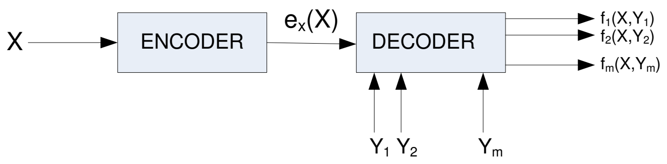

Fig. 1. General multi-functional compression problem with side information

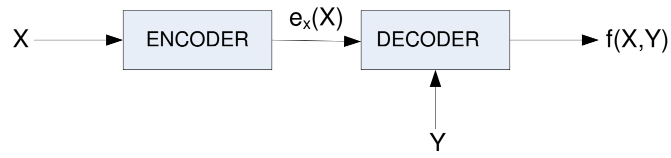

Fig. 2. Functional compression problem with side information, having one desired function at the receiver

sets up the problem statement, presents the necessary technical background and expresses previous results. Section III presents our main results and contributions. Conclusions are given in Section IV.

II. Functional Compression Background

In this section, after giving the problem statement, we express necessary preliminaries to explain the used framework. We also state previous results.

A. Problem Setup

Consider two 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 RVs as $\mathbf{X} = \{X^i\}_{i=l}^{i=l+n-1}$ and $\mathbf{Y_k} = \{Y_k\}_{i=l}^{i=l+n-1}$ 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 the notations, n will be implied by the context. So, we have one encoder e_x and m decoders $r_1, ..., r_m$ (one for each function and its corresponding side information). Encoder e_x maps

$$e_x : \mathcal{X}^n \to \{1, ..., 2^{nR_x}\}, \tag{1}$$

and, each decoder r_k maps

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

The probability of error in decoding f_k is,

$$P_{e_k}^n = Pr[(\mathbf{x}, \mathbf{y}_k) : f_k(\mathbf{x}, \mathbf{y}_k) \neq r_k(e_x(\mathbf{x}), \mathbf{y}_k)] \tag{3}$$

and, total probability of error is,

$$P_e^n = 1 - \prod_k (1 - P_{e_k}^n) \tag{4}$$

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

B. Definitions and Prior Results

In this part, first we briefly explain some definitions used in formulating our results. Following definitions are for the case of considering having only one function at the receiver (i.e., $m = 1$). In the next section, we shall state some new concepts for the case of having several functions . To simplify notations, for the case of having one function, let us use Y, $f(X, Y)$ and $p(x, y)$ instead of Y_1, $f_1(X, Y_1)$ and $p_1(x, y_1)$ defined in the previous section.

Definition 1. *The characteristic graph $G_x = (V_x, E_x)$ of X with respect to Y, $p(x, y)$, and $f(x, y)$ is defined as follows: $V_x = \mathcal{X}$ and an edge $(x_1, x_2) \in \mathcal{X}^2$ is in E_x iff there exists a $y \in \mathcal{Y}$ such that $p(x_1, y)p(x_2, y) > 0$ and $f(x_1, y) \neq f(x_2, y)$.*

In other words, in order to avoid confusion about the function $f(x, y)$ at the receiver, if $(x_1, x_2) \in E_x$, then the descriptions of x_1 and x_2 must be different. Shannon first defined this when studying the zero error capacity of noisy channels [7]. Witsenhausen [8] used this concept to study a simplified version of our problem where one encodes X to compute $f(X)$ with zero distortion.

The n-th power of a graph G_x is a graph $G_x^n = (V_x^n, E_x^n)$ such that $V_x^n = \mathcal{X}^n$ and two sequences $\mathbf{x}_1$ and $\mathbf{x}_2$ are connected iff at least one of their coordinates is connected in G_x.

Definition 2. *Given a graph $G = (V, E)$ and a distribution on its vertices V, Körner [6] defines the graph entropy as follows:*

$$H_G(X) = \min_{X \in W \in \Gamma(G)} I(W; X), \tag{5}$$

where $\Gamma(G)$ is the set of all independent sets of G.

The notation $X \in W \in \Gamma(G)$ means that we are minimizing over all distributions $p(w, x)$ such that $p(w, x) > 0$ implies $x \in w$ where w is an independent set of the graph G. Witsenhausen [8] showed that the graph entropy is the minimum rate at which a single source can be encoded such that a function of that source can be computed with zero distortion.

Orlitsky and Roche [3] defined an extension of Körner's graph entropy, the *conditional graph entropy*.

Definition 3. *The conditional graph entropy is*

$$H_G(X|Y) = \min_{\substack{X \in W \in \Gamma(G) \\ W - X - Y}} I(W; X|Y). \tag{6}$$

Notation $W - X - Y$ indicates a Markov chain.

A vertex coloring of a graph is a function $c_G : V \to \mathbb{N}$ of a graph $G = (V, E)$ such that $(x_1, x_2) \in E$ implies $c_G(x_1) \neq c_G(x_2)$. The entropy of a coloring is the entropy of the induced distribution on colors. Here, $p(c_G(x)) = p(c_G^{-1}(c_G(x)))$ and

$c_G^{-1}(x) = \{\bar{x} : c_G(\bar{x}) = c_G(x)\}$ is called a *color class*. A valid coloring which has the minimum entropy among other valid colorings is called the minimum entropy coloring.

Definition 4. *Given $\mathcal{A} \subset \mathcal{X} \times \mathcal{Y}$, define $\hat{p}(x,y) = p(x,y)/p(\mathcal{A})$ when $(x,y) \in \mathcal{A}$ and $\hat{p}(x,y) = 0$ otherwise. In other words, $\hat{p}$ is the distribution over (x,y) conditioned on $(x,y) \in \mathcal{A}$. Denote the characteristic graph of X with respect to Y, $\hat{p}$, and f as $\hat{G}_x = (\hat{V}_x, \hat{E}_x)$ and the characteristic graph of Y with respect to X, $\hat{p}$, and f as $\hat{G}_y = (\hat{V}_y, \hat{E}_y)$. Note that $\hat{E}_x \subseteq E_x$ and $\hat{E}_y \subseteq E_y$. Finally, we say that c_{G_x} and c_{G_y} are ϵ-colorings of G_x and G_y if they are valid colorings of $\hat{G}_x$ and $\hat{G}_y$ defined with respect to some set $\mathcal{A}$ for which $p(\mathcal{A}) \geq 1-\epsilon$.*

In [9], the *Chromatic entropy* of a graph G is defined as

Definition 5.

$$H_G^{\chi}(X) = \min_{c_G \text{ is an } \epsilon\text{-coloring of } G} H(c_G(X)).$$

It means that the chromatic entropy is a representation of the chromatic number of high probability subgraphs of the characteristic graph. In [10], the conditional chromatic entropy is defined as

Definition 6.

$$H_G^{\chi}(X|Y) = \min_{c_G \text{ is an } \epsilon\text{-coloring of } G} H(c_G(X)|Y).$$

The above optimizations are minima, not infima. It is because that there are finitely many subgraphs for any fixed G and thus, there are only finitely many ϵ-colorings regardless of ϵ. There are some recent works considering the problem of finding the minimum entropy coloring of a graph which in general is NP-hard (e.g., [11]).

Körner proved in [6] that in the limit of long sequence length, the chromatic entropy approaches the graph entropy.

Theorem 7.

$$\lim_{n\to\infty} \frac{1}{n} H_{G^n}^{\chi}(\mathbf{X}) = H_G(X). \tag{7}$$

This result says that one can compute a function of a discrete memoryless source with vanishing probability of error by first coloring a sufficiently large power graph of the characteristic graph of the source with respect to the function, and then, encoding the colors using any code that achieves the entropy bound of the coloring random variable such as Slepian-Wolf code.

The conditional version of the above theorem was proved in [4].

Theorem 8.

$$\lim_{n\to\infty} \frac{1}{n} H_{G^n}^{\chi}(\mathbf{X}|\mathbf{Y}) = H_G(X|Y). \tag{8}$$

Now, consider the network shown in Figure 2 in which the receiver wants to compute a deterministic function of source random variable X; while another random variable Y, possibly correlated with X, is available at the receiver as the side information. First, Shannon considered this problem in [7] for the case $f(X,Y) = X$ and zero distortion. In [1], Wyner and Ziv derived a rate-distortion function for this problem for the case $f(X,Y) = X$. Then, in [2], Yamamoto extended this result to a general function. Both of these results are in terms of an auxiliary random variable. In [3], Orlitsky and Roche considered this problem for the case of zero distortion. They gave an interpretation of this variable in terms of properties of the characteristic graph which led to a particular coding scheme. In particular, they proved that the optimal achievable rate for this problem is $H_G(X|Y)$. This result was extended in [4] considering Theorem 8. An achievable encoding scheme was proposed in [4] which was based on colorings of the characteristic graph of X. It was shown that any colorings of high probability subgraphs of characteristics graph of X leads to an achievable encoding-decoding scheme. It is worthwhile to mention that although finding minimum entropy colorings required to achieve the optimal rate is NP-hard [11] but the simplest colorings will improve over the bound $H(X|Y)$ which arises when trying to completely recover X at the decoder. In a recent work, [5] has considered this problem for a general tree network where intermediate nodes are allowed to perform some computations.

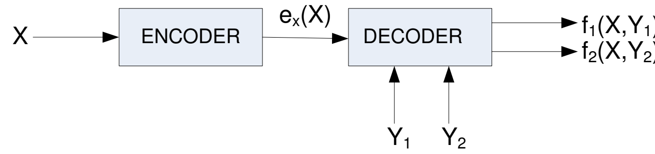

Fig. 3. Simple example of multi-functional compression problem

III. Main Results

The previous works only considered the case that the receiver only wants to compute one function (m=1). Consider the network shown in Figure 1. 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. In other words, one can apply them to the case of having several receivers with different desired functions (functions are separable). First, let us consider the case $m = 2$ (i.e., the receiver has two different functions to compute). Then, we extend results to the case of arbitrary m.

Consider Figure 3. In this problem, the receiver wants to compute two deterministic functions $f_1(X,Y_1)$ and $f_2(X,Y_2)$; while, Y_1 and Y_2 are available at the receiver as the side information. We want to find the minimum achievable rate in which the source node X may encode its information so that the decoder be able to compute its desired functions.

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

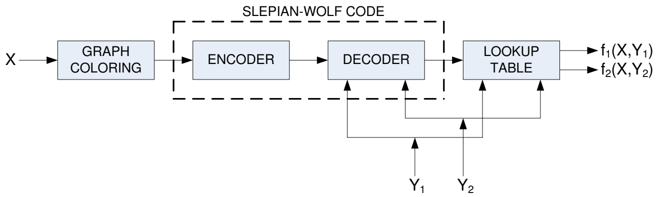

Fig. 4. Source coding scheme for multi-functional compression problem with side information

function of G_{f_1} and G_{f_2}. We call $G_{f_1f_2}$ the *multi-functional characteristic graph* of X.

Definition 9. *The multi-functional characteristic graph $G_{f_1f_2} = (V^x, E^x_{f_1f_2})$ 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:*

$V^x = \mathcal{X}$ and an edge $(x_1, x_2) \in \mathcal{X}^2$ is in $E^x_{f_1f_2}$ if 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 if 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 7, we define the *multi-functional graph entropy* as follows.

Theorem 10.

$$H_{G_{f_1f_2}}(X) = \lim_{n\to\infty} \frac{1}{n} H^{\mathcal{X}}_{G^n_{f_1f_2}}(\mathbf{X}). \tag{9}$$

Similarly to Definition 8, the *conditional multi-functional graph entropy* can be defined as follows.

Theorem 11.

$$H_{G_{f_1f_2}}(X|Y) = \lim_{n\to\infty} \frac{1}{n} H^{\mathcal{X}}_{G^n_{f_1f_2}}(\mathbf{X}|\mathbf{Y}). \tag{10}$$

where $G^n_{f_1f_2}$ is the n-th power of $G_{f_1f_2}$. Now, we can state the following theorem.

Theorem 12. *$H_{G_{f_1f_2}}(X|Y_1, Y_2)$ is the minimum achievable rate for the network shown in Figure 3.*

Proof: To show this, we first show that $R_x \geq H_{G_{f_1f_2}}(X|Y_1, Y_2)$ is an achievable rate (achievability), and no one can outperform this rate (converse). To do this, first, we show that any valid coloring of $G^n_{f_1f_2}$ 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 $G^n_{f_1f_2}$ (converse).

Achievablity: According to [4], any valid coloring of $G^n_{f_1}$ leads to successfully computation of $f_1(\mathbf{X}, \mathbf{Y}_1)$ at the receiver. If $c_{G^n_{f_1}}$ is a valid coloring of $G^n_{f_1}$, there exists a function r_1 such that $r_1(c_{G^n_{f_1}}(\mathbf{X}), \mathbf{Y}_1) = f_1(\mathbf{X}, \mathbf{Y}_1)$, with high probability. A similar argument holds for G_{f_2}. Now, assume that $c_{G^n_{f_1f_2}}$ is a valid coloring of $G^n_{f_1f_2}$. Since, $E^{\mathbf{x}}_{f_1} \subseteq E^{\mathbf{x}}_{f_1f_2}$ and $E^{\mathbf{x}}_{f_2} \subseteq E^{\mathbf{x}}_{f_1f_2}$, any valid coloring of $G^n_{f_1f_2}$ induces valid colorings for $G^n_{f_1}$ and $G^n_{f_2}$. Thus, any valid coloring of $G^n_{f_1f_2}$ leads to successful computation of $f_1(\mathbf{X}, \mathbf{Y}_1)$ and $f_2(\mathbf{X}, \mathbf{Y}_2)$ at the receiver. So, $c_{G^n_{f_1f_2}}$ leads to an achievable encoding scheme (i.e. there exist two functions r_1 and r_2 such that $r_1(c_{G^n_{f_1f_2}}(\mathbf{X}), \mathbf{Y}_1) = f_1(\mathbf{X}, \mathbf{Y}_1)$ and $r_2(c_{G^n_{f_1f_2}}(\mathbf{X}), \mathbf{Y}_2) = f_2(\mathbf{X}, \mathbf{Y}_2)$), with high probability.

For the case of having the identity function at the receiver (i.e., the receiver wants the whole information of the source node), Slepian and Wolf proposed a technique in [12] 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 mentioned bound.

Converse: Now, we show that any achievable encoding-decoding scheme performing on blocks with length n, induces a valid coloring of $G^n_{f_1f_2}$. In other words, we want to show that if there exist functions e, r_1 and r_2 such that $r_1(e(\mathbf{X}), \mathbf{Y}_1) = f_1(\mathbf{X}, \mathbf{Y}_1)$ and $r_2(e(\mathbf{X}), \mathbf{Y}_2) = f_2(\mathbf{X}, \mathbf{Y}_2)$, $e(\mathbf{X})$ is a valid coloring of $G^n_{f_1f_2}$.

Let us proceed by contradiction. If $e(\mathbf{X})$ were not a valid coloring of $G^n_{f_1f_2}$, there must be some edge in $E^{\mathbf{x}}_{f_1f_2}$ 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., $e(\mathbf{x}_1) = e(\mathbf{x}_2)$), but also are connected. Since they are connected to each other, by definition of $G^n_{f_1f_2}$, 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, $r_1(e(\mathbf{x}_1), \mathbf{y}_1) \neq r_1(e(\mathbf{x}_2), \mathbf{y}_1)$. Since $e(\mathbf{x}_1) = e(\mathbf{x}_2)$, then, $r_1(e(\mathbf{x}_1), \mathbf{y}_1) \neq r_1(e(\mathbf{x}_1), \mathbf{y}_1)$. 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 $G^n_{f_1f_2}$ and it completes the proof. ■

Now, let us consider the network shown in Figure 1 where the receiver wishes to compute m deterministic functions of source information having some side informations.

Theorem 13. *$H_{G_{f_1,...,f_m}}(X|Y_1, ..., Y_m)$ is the minimum achievable rate for the network shown in Figure 1.*

The argument here is pretty similar to the case $m = 2$ mentioned in Theorem 12. So, for brevity, we only sketch the proof. To show this, one may first show that any colorings of multi-functional characteristic graph of X with respect to desired functions ($G_{f_1,...,f_m}$) leads to an achievable scheme. Then, showing that any achievable encoding-decoding scheme induces a coloring on $G_{f_1,...,f_m}$ completes the proof.

IV. Conclusion

In this paper, we considered 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 is available at the receiver as the side information.

There are some works considered this problem for the case of having one desired function at the receiver. In [1], Wyner and Ziv considered this problem for the special case of $m = 1$ and $f_1(X, Y_1) = X$ and derived a rate-distortion function. In [2], Yamamoto extended this result to the case of having one general function $f_1(X, Y_1)$. Both of these results were in terms of an auxiliary random variable. For the case of zero distortion, Orlitsky and Roche gave in [3] an interpretation of this variable in terms of properties of characteristic graph which led to a particular coding scheme. This results was extended in [4] by providing an achievable scheme based on colorings of the characteristic graph. Also, in a recent work, reference [5] has considered this problem for a general tree network where intermediate nodes allowed to perform some computations.

Here, we considered the case that the receiver wants to compute several deterministic functions with different side information random variables and zero distortion. Our results do not depend on the fact that all functions are desired in one receiver and one can apply them to the case of having several receivers with different desired functions (i.e., functions are separable). We defined a new concept named the *multi-functional graph entropy* which is an extension of the graph entropy defined by korner in [6]. We showed that minimum achievable rate for this problem is equal to the conditional multi-functional graph entropy of random variable X given the side informations. We also proposed a coding scheme based on graph colorings to achieve this rate.

REFERENCES

[1] 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.

[2] 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.

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

[4] V. Doshi, D. Shah, M. Médard, and S. Jaggi, "Graph coloring and conditional graph entropy," in *2006 Asilomar Conference on Signals, Systems, and Computers*, Asilomar, CA, Oct.-Nov. 2006, pp. 2137–2141.

[5] S. Feizi and M. Médard, "Only sources need to compute, on functional compression in tree networks," invited paper in Allerton Conference, Sep. 2009.

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

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

[8] 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.

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

[10] V. Doshi, D. Shah, M. Médard, and S. Jaggi, "Distributed functional compression through graph coloring," in *2007 Data Compression Conference*, Snowbird, UT, Mar. 2007, pp. 93–102.

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

[12] 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.