Graph Coloring and Conditional Graph Entropy

Vishal Doshi, Devavrat Shah, Muriel Médard, Sidharth Jaggi

Laboratory for Information and Decision Systems

Massachusetts Institute of Technology

Cambridge, MA 02139

{vdoshi, devavrat, medard, jaggi} @ mit.edu

Abstract— We consider the remote computation of a function of two sources where one is receiver side information. Specifically, given side information Y, we wish to compute $f(X, Y)$ based on information transmitted by X over a noise-less channel. The goal is to characterize the minimal rate at which X must transmit information to enable the computation of the function f.

Recently, Orlitsky and Roche (2001) established that the conditional graph entropy of the characteristic graph of the function is a solution to this problem. Their achievability scheme does not separate "functional coding" from the well understood distributed source coding problem. In this paper, we seek a separation between the functional coding and the coding for correlation. In other words, we want to preprocess X (with respect to f), and then transmit the *preprocessed data* using a standard Slepian-Wolf coding scheme at the Orlitsky-Roche rate.

We establish that a minimum (conditional) entropy coloring of the product of characteristic graphs is asymptotically equal to the conditional graph entropy. This can be seen as a generalization of a result of Körner (1973) which shows that the minimum entropy coloring of product graphs is asymptotically equal to the graph entropy. Thus, graph coloring provides a modular technique to achieve coding gains when the receiver is interested in decoding some function of the source.

I. INTRODUCTION

Consider a network of sensors relaying correlated measurements to a central receiver with measurements of its own. Usually, the receiver is interested in computing a function of the received data. Hence, to the receiver the bulk of the data received from the various sensors is irrelevant. Contemporary data compression techniques (see [1], [2]) compress the data utilizing only the correlation of the measurements and do not take the receiver's function of interest into account. To increase the efficiency for such a system, we seek to implement better compression rates by exploiting the end-user's intentions (the computation of the function f of the data received).

Consider the following example: suppose there is a source (uniformly) producing an n-bit integer. The receiver also produces an n-bit integer and is only interested in the parity of the sum. Assuming independence of the integer data, contemporary techniques would have the source transmit all n bits, when it is clear that only the final bit is necessary. The use of such compression techniques enables possibly large gains over contemporary methods depending upon the complexity of the function.

As a first step towards the implementation of methods like this, we consider the case (as in the example) where there is a single source, denoted by the random variable X with support $\mathcal{X}$. The information available at the receiver is denoted by
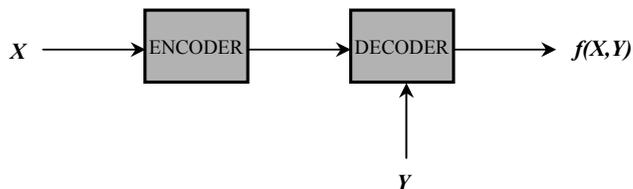

Fig. 1. X is encoded such that $f(X, Y)$ is recovered based on the coded information and side information Y.

the random variable Y with support $\mathcal{Y}$. Both X and Y are discrete memoryless sources. The receiver wishes to compute a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, where $\mathcal{Z}$ is some finite set. The goal here is to minimize the rate at which X transmits information to the receiver so as to enable computation of f at the receiver. Figure 1 illustrates the question of interest.

A. Setup and Related Work

This question has been very well studied in different contexts in the literature [3], [4], [5]. Among them, Orlitsky and Roche [5] recently characterized the minimum rate requirement as the conditional graph entropy of the characteristic graph. We introduce the necessary definitions and notation to explain these results as well as formally state our question of interest.

Assume that the random variables X and Y have some joint distribution $p(x, y)$. The n-sequence of random variables, denoted by $\mathbf{X}$ and $\mathbf{Y}$ are such that $(\mathbf{X}, \mathbf{Y}) = \{(X_i, Y_i)\}_{i=1}^n$ are i.i.d. with each (X_i, Y_i) drawn according to $p(x, y)$. The receiver is interested in computing a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ or $f : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}^n$, its obvious extension. The primary goal is to determine the minimal rate for X to transmit data to the receiver so that the receiver can compute the function f with vanishing probability of error.

Wyner and Ziv [6] considered the side information rate distortion problem. Yamamoto [4] extended their rate distortion function to the case of functional compression. Their functions relied on auxilliary random variables. In a sense, this paper makes those variables explicit for the zero distortion case. Further, we describe them in such a way that our scheme builds upon a Slepian-Wolf coding scheme [7] layered with a preprocessing of the source data.

Define the *characteristic graph* of X with respect to Y, f, and $p(x, y)$ is $G = (V, E)$ with $V = \mathcal{X}$ and E defined as follows: an edge $(x_1, x_2) \in \mathcal{X}^2$ is in E if there exists a $y \in$

$\mathcal{Y}$ such that $p(x_1, y), p(x_2, y) > 0$ and $f(x_1, y) \neq f(x_2, y)$. Effectively, G is the "confusability graph" from the perspective of the receiver with side information Y. This was first defined by Witsenhausen [8].

The graph entropy of a graph G with some distribution on its vertices was introduced by Korner [3] and is defined as:

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

where $\Gamma(G)$ is the set of all independent sets[1] of G. Here, $X \in W$ means that the joint distribution $p(w, x)$ on $\Gamma(G) \times \mathcal{X}$ is such that $p(w, x) > 0$ implies $x \in w$. To motivate the importance of graph entropy, consider the following: given a function $g : \mathcal{X} \to \mathcal{Z}$, define deterministic random variable Y on $\mathcal{Y} = \{1\}$ and set $f(x, 1) = g(x)$ for all $x \in \mathcal{X}$. Then, the graph entropy $H_G(X)$ is the minimal rate required to distinguish functional values of g [8].

Orlitsky and Roche [5] established that the *conditional graph entropy*, an extension of graph entropy, is the minimum required transmission rate from X to receiver to compute function $f(X, Y)$ with small probability of error. The conditional graph entropy is defined as

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

A natural extension of this problem is the computation of rate-distortion region. Yamamoto characterized the rate-distortion function for the question posed here by extending the Wyner-Ziv side information result [6]. This was recently built upon by Feng, Effros and Savari [9].

Our work in this paper is closer (in spirit) to the result by Körner [3]. In essence, Körner showed that achieving the graph entropy is the same as coloring a large graph and sending the colors.

Consider the OR-product graph of G, denoted as $G^n = (V_n, E_n)$ where $V_n = V^n = \mathcal{X}^n$ and two vertices $(\mathbf{x}_1, \mathbf{x}_2)$ are in E_n if any component $(x_{1i}, x_{2i}) \in E$. By coloring, we mean standard graph vertex coloring, i.e. a function $c : V \to \mathbb{N}$ of a graph $G = (V, E)$ such that $(x_1, x_2) \in E$ implies $c(x_1) \neq c(x_2)$. The entropy of a coloring is given by the induced distribution on colors $p(c(x)) = p(c^{-1}(c(x)))$ where $c^{-1}(x) = \{\bar{x} : c(\bar{x}) = c(x)\}$.

Next, we define ε-colorings of any graph G (for any $\varepsilon > 0$). Let $\mathcal{A} \subset \mathcal{X} \times \mathcal{Y}$ be any subset such that $p(\mathcal{A}) \geq 1 - \varepsilon$. Define $\hat{p}(x, y) = p(x, y) \cdot \mathbf{1}_{(x,y) \in \mathcal{A}}$ (where $\mathbf{1}_Q$ is the indicator function for condition Q). While this is not a true probability, our definition of characteristic graph of X with respect to Y, f, and $\hat{p}$ is valid. Denote this characteristic graph as $\hat{G}$. Note that the set of edges of $\hat{G}$ is a subset of the set of edges of G. Finally, we say c is an ε-coloring of G if it is a valid colorings of $\hat{G}$ defined with respect to some high probability set $\mathcal{A}$.

The *chromatic entropy* of a graph is

$$H_G^\chi(X) = \inf_{\varepsilon > 0} \{H(c(X)) : c \text{ is an } \varepsilon\text{-coloring of } G\}.$$

[1] A subset of vertices of a graph is an independent set if no two nodes in the subset are adjacent to each other in G.
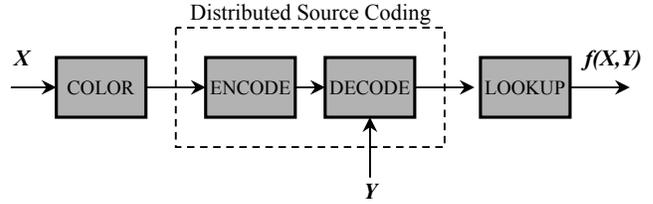

Fig. 2. Coloring based coding allows separation between functional coding and distributed source coding as shown above.

We define it's natural extension *conditional chromatic entropy* as follows:

$$H_G^\chi(X|Y) = \inf_{\varepsilon > 0} \{H(c(X)|Y) : c \text{ is an } \varepsilon\text{-coloring of } G\}.$$

Now, we state the Körner's result [3] as follows:

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

The result requires ε-colorings and not the simple colorings because, as seen in [10], the ratio of the above quantities can be arbitrarily large. The implications of this result are that for large enough n, one can color a high probability subgraph of the product graph G^n to achieve near optimum functional source coding when the function is only of one source. This effectively layers the solution to a problem of graph coloring followed by using standard coding techniques to transmit the color sequence.

The Orlitsky and Roche's solution is an end-to-end solution. That is, it includes functional and distributed source coding. Their achievability scheme involves coding over the space of independent sets. It is a larger class than the space of colors because each coloring gives rise to disjoint independent sets. Like the implication of Körner's result, the coloring will allow for separation between functional coding and distributed source coding. Hence we want a result similar to that of Körner's result so that the problem can be decoupled without losing any coding gains asymptotically.

B. Our Contribution

As the main result of this paper, we show that the minimum conditional entropy coloring of the OR-product graph is asymptotically the same as the conditional chromatic number.

Theorem 1:

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

This validates the idea of using graph coloring to achieve better compression ratios. The variable W and the distribution on it that achieves $H_G(X|Y)$ is the auxilliary random variable from the Yamamoto result. In the entropy sense, the induced variable, $c(X)$, converges to W. This method thus leads to a natural layering of the problem just as in Körner's case. Figure 2 illustrates the layering.

Nevertheless, we need large enough n to achieve rates close to the conditional graph entropy. Consider the following

example from [5]. We have $\mathcal{X} = \mathcal{Y} = \{1, 2, 3\}$, $p(x, y) = \frac{1}{6}$ for all $x \neq y$, and $f(x, y) = 1$ if $x > y$ and 0 otherwise. Then, our graph G has only one edge: $(1, 3)$. Because of uniformity, we know $H_G^\chi(X|Y) = \frac{2}{3}$ using the coloring $c(1) = c(2) = 1$ and $c(3) = 2$. But $H_G(X|Y) \approx 0.54$ bits as calculated in [5].

This is worrisome because minimum entropy graph coloring has been shown to be NP–complete [11]. Nevertheless, the problem is layered into two. The second part is that of distributed source coding which is well understood. There exist methods to achieve corner points of the Slepian-Wolf rate region (e.g. [1]), and there exist methods to achieve all other points on the rate region (e.g. [2]). Further, the first part, that of graph coloring, has very rich literature (e.g. [10], [12]) and heuristics exist (e.g. [13]) for this hard problem.

II. PROOF OF MAIN RESULT

We borrow some techniques that are present in Körner [3], Orlitsky and Roche [5]. First, we recall some useful known results. Then, we establish the claimed equality in Theorem 1.

A. Preliminaries

We begin by listing the results that we will use in our proof.

Definition 1 (Typicality): We use the notion of ε-strong typicality defined in [14, p. 358]. For any n, and any sequence $\mathbf{x}$ of length n, define the empirical frequency of the symbol x in $\mathbf{x}$ as

$$\nu_{\mathbf{x}}(x) = \frac{|\{i : x_i = x\}|}{n}.$$

Then a sequence $\mathbf{x}$ is ε-strongly typical for $\varepsilon > 0$ if for all $x \in \mathcal{X}$ with $p(x) > 0$,

$$|\nu_{\mathbf{x}}(x) - p(x)| \leq \frac{\varepsilon}{|\mathcal{X}|},$$

and for all $x \in \mathcal{X}$ with $p(x) = 0$, $\nu_{\mathbf{x}}(x) = 0$. The set of all such ε-strongly typical sequences, called the ε-typical set will be denoted as $T_\varepsilon^n(X)$, or T_ε^n when the variables are clear from context. A similar definition naturally extends for the case of joint variables.

Lemma 1: Suppose $(\mathbf{X}, \mathbf{Y})$ is a sequence of n random variables drawn independently and according to the joint distribution $p(x, y)$, which is the marginal of $p(w, x, y)$. Let an n-sequence $\mathbf{W}$ be drawn independently according to its marginal, $p(w)$. Suppose the joint distribution $p(w, x, y)$ forms a Markov Chain, $W - X - Y$ (i.e. $p(w, x, y) = p(w|x)p(x, y)$). Then, for all $\varepsilon > 0$, there is a $\varepsilon_1 = k \cdot \varepsilon$, where k depends only on the distribution $p(w, x, y)$, such that for sufficiently large n,

1) $P[\mathbf{X} \notin T_\varepsilon^n] < \varepsilon_1$, $P[\mathbf{Y} \notin T_\varepsilon^n] < \varepsilon_1$, and $P[(\mathbf{X}, \mathbf{Y}) \notin T_\varepsilon^n] < \varepsilon_1$,
2) For all $\mathbf{x} \in T_\varepsilon^n$, $P[(\mathbf{x}, \mathbf{W}) \in T_\varepsilon^n] \geq 2^{-n(I(W;X)+\varepsilon_1)}$.
3) For all $\mathbf{y} \in T_\varepsilon^n$, $P[(\mathbf{y}, \mathbf{W}) \in T_\varepsilon^n] \leq 2^{-n(I(W;X)-\varepsilon_1)}$.
4) For all $(\mathbf{w}, \mathbf{x}) \in T_\varepsilon^n$,

$$P[(\mathbf{w}, \mathbf{Y}) \in T_\varepsilon^n | (\mathbf{x}, \mathbf{Y}) \in T_\varepsilon^n] \geq 1 - \varepsilon_1.$$

Part 1 follows from Lemma 13.6.1, parts 2 and 3 follow from Lemma 13.6.2, and part 4 follows from Lemma 14.8.1, all from [14].
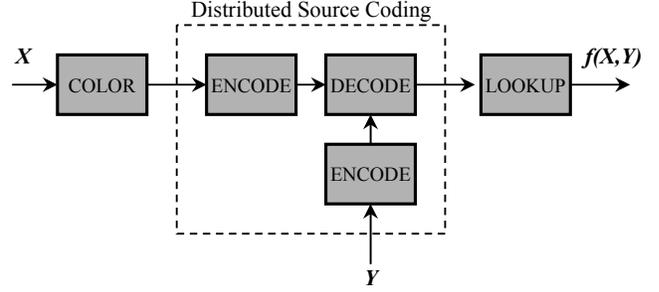

Fig. 3. We use Slepian-Wolf codes with Y sending at full rate to achieve the corner point $R_x = H(c(X)|Y)$.

B. Lower bound

To prove the lower bound, consider any n and the corresponding OR-product graph G^n. Let the coloring c of G^n be such that

$$H_{G^n}^\chi(\mathbf{X}|\mathbf{Y}) = H(c(\mathbf{X})).$$

Such a coloring exists because the number of possible colorings for a graph with n vertices is a finite set. Using this coloring, we will show that there is a scheme that requires rate $\frac{1}{n} H_{G^n}^\chi(\mathbf{X}|\mathbf{Y})$ and provides a way to compute the function at receiver with small probability of error. This will establish the achievability of rate $\frac{1}{n} H_{G^n}^\chi(\mathbf{X}|\mathbf{Y})$. By Orlitsky and Roche [5], no scheme can achieve rate smaller than $H_G(X|Y)$.

Lemma 2:

$$\liminf_{n \to \infty} \frac{1}{n} H_{G^n}^\chi(\mathbf{X}|\mathbf{Y}) \geq H_G(X|Y).$$

Proof: Given $n > 0$, let the coloring c, as above, denote one the coloring that achieves $H_{G^n}^\chi(\mathbf{X}|\mathbf{Y})$ for the OR-product graph G^n. For every color, σ and $\mathbf{y}$, define $g(\sigma, \mathbf{y}) = f(\mathbf{x}, \mathbf{y})$ where $\mathbf{x}$ is any element of $c^{-1}(\sigma) = \{\mathbf{x} : c(\mathbf{x}) = \sigma\}$ such that $p(\mathbf{x}, \mathbf{y}) > 0$. If no such $\mathbf{x}$ exists, g is undefined.

Suppose $\mathbf{x}$ is the information at the source, and $\mathbf{y}$ is the information at the receiver. Without loss of generality, presume $p(\mathbf{x}, \mathbf{y}) > 0$. Suppose $c(\mathbf{x})$ and $\mathbf{y}$ are available at the decoder. Then, we have an error if $g(\sigma, \mathbf{y}) \neq f(\mathbf{x}, \mathbf{y})$. Suppose $\mathbf{x}' \in \mathcal{X}$, $c(\mathbf{x}') = c(\mathbf{x})$, and $p(\mathbf{x}', \mathbf{y}) > 0$. Because $\mathbf{x}$ and $\mathbf{x}'$ have the same color, we know $(\mathbf{x}, \mathbf{x}') \notin E_n$, and therefore for all $\mathbf{y}$ where $p(\mathbf{x}, \mathbf{y}), p(\mathbf{x}', \mathbf{y}) > 0$, including ours, we must have $f(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}', \mathbf{y})$. Because f has the same value for all $\mathbf{x}'$ with $p(\mathbf{x}', \mathbf{y}) > 0$, $f(\mathbf{x}, \mathbf{y}) = g(\sigma, \mathbf{y})$. Thus, we have shown that the color of $\mathbf{X}$ and $\mathbf{Y}$ are sufficient to determine the function $f(\mathbf{X}, \mathbf{Y})$.

The Slepian-Wolf Theorem [7] states that for two sources, X and Y, there exists a sequence of codes with vanishing probability of error as n grows without bound for the rates: $R_1 \geq H(X|Y)$, $R_2 \geq H(Y|X)$, and $R_1 + R_2 \geq H(X, Y)$.

We apply that theorem to the source $(c(\mathbf{X}), \mathbf{Y})$. In our case, Y is side information. In other words, $R_2 \geq H(Y)$. See Figure 3 for an illustration.

Thus, we get the following *per symbol* rates are achievable for an encoding of X where c is any ε-coloring of G^n (with

$\varepsilon > 0$).

$$R_1 \geq \frac{1}{n} H(c(\mathbf{X})|\mathbf{Y}).$$

Thus, the minimal value of R_1 is by definition $H_{G^n}^\chi(\mathbf{X}|\mathbf{Y})$.

Putting the above discussion together, we obtain the desired result:

$$\liminf_{n\to\infty} \frac{1}{n} H_{G^n}^\chi(\mathbf{X}|\mathbf{Y}) \geq H_G(X|Y)$$

■

C. Upper Bound

Next we show that, indeed, the information needed to recover $c_n(\mathbf{X})$ given $\mathbf{Y}$ is at most $H_G(X|Y)$.

Lemma 3:

$$\liminf_{n\to\infty} \frac{1}{n} H_{G^n}^\chi(\mathbf{X}|\mathbf{Y}) \leq H_G(X|Y).$$

Proof: Suppose $\varepsilon, \delta > 0$. Suppose n is large enough so that the following holds: (1) Lemma 1 applies with some $\varepsilon_1 < 1$, (2) $2^{-n\delta} < \varepsilon_1$, and (3) $n > 2 + \frac{3}{2\varepsilon_1}$.

Let $p(w,x,y)$ be the distribution that achieves $H_G(X|Y)$ with the Markov chain, $W - X - Y$. Let marginals be denoted simply as $p(w)$, $p(x)$, $p(y)$. For any integer M, we define an M-*system* $(\mathbf{W}_1, \ldots, \mathbf{W}_M)$ where each $\mathbf{W}_i$ is drawn independently according to $p(\mathbf{w}) = \prod_{i=1}^n p(w_i)$.

Declare an error if $(\mathbf{x}, \mathbf{y}) \notin T_\varepsilon^n$. (Thus, we are in fact looking at colorings over T_ε^n, or ε_1-colorings.) This, by construction, happens with probability less than ε_1. Henceforth assume that $(\mathbf{x}, \mathbf{y}) \in T_\varepsilon^n$.

Declare an error if there is no i such that $(\mathbf{W}_i, \mathbf{x}) \in T_\varepsilon^n$. The probability for this event is

$$
\begin{aligned}
P[(\mathbf{W}_i, \mathbf{x}) \notin T_\varepsilon^n \forall i] &\overset{(a)}{\leq} \prod_{i=1}^M P[(\mathbf{W}_i, \mathbf{x}) \notin T_\varepsilon^n] \\
&\overset{(b)}{=} (1 - P[(\mathbf{W}, \mathbf{x}) \in T_\varepsilon^n])^M \\
&\overset{(c)}{\leq} \left(1 - 2^{-n(I(W;X)+\varepsilon_1)}\right)^M \\
&\overset{(d)}{\leq} 2^{-M \cdot 2^{-n(I(W;X)+\varepsilon_1)}}
\end{aligned}
$$

where (a) and (b) follow because the $\mathbf{W}_i$ are i.i.d., (c) follows from Lemma 1, part (2), and (d) follows because for $\alpha \in [0,1]$, $(1-\alpha)^n \leq 2^{-n\alpha}$. Assuming $M > 2^{n(I(W;X)+\varepsilon_1+\delta)}$,

$$P[(\mathbf{W}_i, \mathbf{x}) \notin T_\varepsilon^n \forall i] \leq 2^{-\delta n} < \varepsilon_1,$$

because n is large enough.

Henceforth, fix an M-system $(\mathbf{W}_1, \ldots, \mathbf{W}_M)$ for some $M > 2^{n(I(W;X)+\varepsilon_1+\delta)}$. Further, assume there is some i such that $(\mathbf{W}_i, \mathbf{x}) \in T_\varepsilon^n$.

For every $\mathbf{x}$, let the smallest (or any) such i be denoted as $\hat{c}(\mathbf{x})$. Note that $\hat{c}$ is a valid coloring of the graph G^n.

Next, for each $\mathbf{y}$, define the following:

$$
\begin{aligned}
S(\mathbf{y}) &= \{\hat{c}(\mathbf{x}) : (\mathbf{x}, \mathbf{y}) \in T_\varepsilon^n\}, \\
Z(\mathbf{y}) &= \{\mathbf{W}_i : (\mathbf{W}_i, \mathbf{y}) \in T_\varepsilon^n\}, \\
s(\mathbf{y}) &= |S(\mathbf{y})|, \\
z(\mathbf{y}) &= |Z(\mathbf{y})|.
\end{aligned}
$$

Using these definitions, $s(\mathbf{y}) = \sum_{i=1}^M \mathbf{1}_{i \in S(\mathbf{y})}$, where we have used the fact that our coloring scheme $\hat{c}$ is simply an assignment of the indices of the M-system. Thus, we know

$$E[s(\mathbf{Y})] = \sum_{i=1}^M P[i \in S(\mathbf{Y})].$$

Similarly, we get $z(\mathbf{y}) = \sum_{i=1}^M \mathbf{1}_{\mathbf{W}_i \in Z(\mathbf{y})}$. Thus,

$$
\begin{aligned}
E[z(\mathbf{Y})] &= \sum_{i=1}^M P[\mathbf{W}_i \in Z(\mathbf{Y})] \\
&\geq \sum_{i=1}^M P[\mathbf{W}_i \in Z(\mathbf{Y}) \text{ and } i \in S(\mathbf{Y})] \\
&= \sum_{i=1}^M P[i \in S(\mathbf{Y})] P[\mathbf{W}_i \in Z(\mathbf{Y})|i \in S(\mathbf{Y})]
\end{aligned}
$$

We know that if $i \in S(\mathbf{Y})$, there is some $\mathbf{x}$ such that $\hat{c}(\mathbf{x}) = i$ and $(\mathbf{x}, \mathbf{Y}) \in T_\varepsilon^n$. For each such $\mathbf{x}$, we must have (by definition of our coloring), $(\mathbf{W}_i, \mathbf{x}) \in T_\varepsilon^n$. For each such $\mathbf{x}$ where $(\mathbf{W}_i, \mathbf{x}) \in T_\varepsilon^n$,

$$
\begin{aligned}
P[\mathbf{W}_i \in Z(\mathbf{Y})|i \in S(\mathbf{Y})] &= P[(\mathbf{W}_i, \mathbf{Y}) \in T_\varepsilon^n|(\mathbf{x}, \mathbf{Y}) \in T_\varepsilon^n] \\
&\geq 1 - \varepsilon_1
\end{aligned}
$$

by Lemma 1, part (4). Thus, we have $E[z(\mathbf{Y})] \geq (1 - \varepsilon_1) E[s(\mathbf{Y})]$. This and Jensen's inequality imply that

$$
\begin{aligned}
E[\log s(\mathbf{Y})] &\leq \log E[s(\mathbf{Y})] \\
&\leq \log E\left[\frac{z(\mathbf{Y})}{1-\varepsilon_1}\right].
\end{aligned}
$$

Finally, using a Taylor series expansion, we know $\log \frac{1}{1-\varepsilon_1} \leq \varepsilon_2 = 2\varepsilon_1 + \frac{1}{2}$ when $0 < \varepsilon_1 < 1$. Thus,

$$E[\log s(\mathbf{Y})] \leq \log E[z(\mathbf{Y})] + \varepsilon_2, \tag{1}$$

Next, we compute

$$
\begin{aligned}
E[z(\mathbf{Y})] &= \sum_{i=1}^M P[(\mathbf{W}_i, \mathbf{Y}) \in T_\varepsilon^n] \\
&= M \cdot P[(\mathbf{W}_i, \mathbf{Y}) \in T_\varepsilon^n]
\end{aligned}
$$

because the $\mathbf{W}_i$ are i.i.d. Therefore,

$$E[z(\mathbf{Y})] \leq M \cdot 2^{-n(I(W;Y)-\varepsilon_1)} \tag{2}$$

by Lemma 1, part (3).

By the definition of $S(\mathbf{y})$, we know that determining $\hat{c}$ given $\mathbf{Y} = \mathbf{y}$ requires at most $\log s(\mathbf{y})$ bits. Therefore, we have

$$H(\hat{c}(\mathbf{X})|\mathbf{Y}) \leq E[\log s(\mathbf{Y})]. \tag{3}$$

Putting it all together, we have

$$
\begin{aligned}
H_{G^n}^{\chi}(\mathbf{X}|\mathbf{Y}) &\overset{(a)}{\le} H(\hat{c}(\mathbf{X})|\mathbf{Y}) \\
&\overset{(b)}{\le} E[\log s(\mathbf{Y})] \\
&\overset{(c)}{\le} \log E[z(\mathbf{Y})] + \varepsilon_2 \\
&\overset{(d)}{\le} \log\left(M \cdot 2^{-n(I(W;Y)-\varepsilon_1)}\right) + \varepsilon_2 \\
&\overset{(e)}{=} \log\left(2^{n(I(W;X)-I(W;Y)+2\varepsilon_1+\delta)} + 1\right) + \varepsilon_2 \\
&\overset{(f)}{\le} n(I(W;X) - I(W;Y) + 2\varepsilon_1 + \delta) + 1 + \varepsilon_2
\end{aligned}
$$

where (a) follows by definition of the conditional chromatic entropy, (b) follows from inequality (3), (c) follows from inequality (1), (d) follows from inequality (2), (e) follows by setting $M = \lceil 2^{n(I(W;X)+\varepsilon_1+\delta)} \rceil$, and (f) follows because $\log(\alpha + 1) \le \log(\alpha) + 1$ for $\alpha \ge 1$.

We know for Markov chains $W - X - Y$,

$$
I(W;X) - I(W;Y) = I(W;X|Y).
$$

Thus, for our optimal distribution $p(w, x, y)$, we have

$$
H_{G^n}^{\chi}(\mathbf{X}|\mathbf{Y}) \le n(H_G(X|Y) + 2\varepsilon_1 + \delta) + 1 + \varepsilon_2
$$

Because $n > 2 + \frac{3}{2\varepsilon_1}$, $\frac{1+\varepsilon_2}{n} < \varepsilon_1$. Thus, $\frac{1}{n} H_{G^n}^{\chi}(\mathbf{X}|\mathbf{Y}) \le H_G(X|Y) + 3\varepsilon_1 + \delta$. Therefore,

$$
\limsup_{n\to\infty} \frac{1}{n} H_{G^n}^{\chi}(\mathbf{X}|\mathbf{Y}) \le H_G(X|Y).
$$

∎

Combining the lower bound, Lemma 2, and the upper bound, Lemma 3, we have our result:

$$
\lim_{n\to\infty} \frac{1}{n} H_{G^n}^{\chi}(\mathbf{X}|\mathbf{Y}) = H_G(X|Y). \tag{4}
$$

III. CONCLUSION

We have validated the idea of graph coloring to achieve all possible rates. As a consequence, it is possible to decouple the side information functional source coding problem into a problem of graph coloring followed by a coding scheme that removed the redundancy between the colors (like Slepian-Wolf). While the graph coloring problem is shown to be NP–complete, this decoupling allows for the decomposition of a hard problem into one for which many heuristics and copious literature exist (graph coloring) and one that is solved (distributed source coding).

We plan to show that this decoupling is possible in the more general case when Y is a separate source (and not side information). In this case, we will wish to code both X and Y. This will generalize our current results where we assume the rate for Y is $H(Y)$. Further, we plan to consider a relaxation of the fidelity criterion. We will wish to see if Yamamoto's rate distortion function can be achieved with graph coloring, as well as considering the case where Y is not side information. Thus, we hope to have a unified solution that will allow for the distributed compression of functions of correlated sources.

REFERENCES

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

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

[3] 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.

[4] H. Yamamoto, "Wyner-ziv theory for a general function of the correlated sources," *IEEE Trans. Inform. Theory*, vol. 28, no. 5, pp. 803–807, Sept. 1982.

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

[6] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

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

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

[9] H. Feng, M. Effros, and S. Savari, "Functional source coding for networks with receiver side information," Allerton Conference on Communication, Control, and Computing, Sept. 2004.

[10] J. Cardinal, S. Fiorini, and G. Joret, "Minimum entropy coloring," Lecture Notes on Computer Science, ser. International Symposium on Algorithms and Computation, X. Deng and D. Du, Eds., vol. 3827. Springer-Verlag, 2005, pp. 819–828.

[11] J. Cardinal, S. Fiorini, and G. V. Assche, "On minimum entropy graph colorings," ISIT 2004, June–July 2004, p. 43.

[12] C. McDiarmid, "Colourings of random graphs," in *Graph Colourings*, ser. Pitman Research Notes in Mathematics Series, R. Nelson and R. J. Wilson, Eds. Longman Scientific & Technical, 1990, pp. 79–86.

[13] G. Campers, O. Henkes, and J. P. Leclerq, "Graph coloring heuristics: A survey, some new propositions and computational experiences on random and "Leighton's" graphs," in *Operations research '87*, G. K. Rand, Ed., 1988, pp. 917–932.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.