Distributed Functional Compression through Graph Coloring

Vishal Doshi, Devavrat Shah, Muriel Médard, and Sidharth Jaggi

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Email: {vdoshi, devavrat, medard, jaggi} @ mit.edu

Abstract

We consider the distributed computation of a function of random sources with minimal communication. Specifically, given two discrete memoryless sources, X and Y, a receiver wishes to compute $f(X, Y)$ based on (encoded) information sent from X and Y in a distributed manner. A special case, $f(X, Y) = (X, Y)$, is the classical question of distributed source coding considered by Slepian and Wolf (1973).

Orlitsky and Roche (2001) considered a somewhat restricted setup when Y is available as side information at the receiver. They characterized the minimal rate at which X needs to transmit data to the receiver as the conditional graph entropy of the characteristic graph of X based on f. In our recent work (2006), we further established that this minimal rate can be achieved by means of graph coloring and distributed source coding (e.g. Slepian-Wolf coding). This characterization allows for the separation between "function coding" and "correlation coding."

In this paper, we consider a more general setup where X and Y are both encoded (separately). This is a significantly harder setup for which to give a single-letter characterization for the complete rate region. We find that under a certain condition on the support set of X and Y (called the *zigzag* condition), it is possible to characterize the rate region based on graph colorings at X and Y separately. That is, any achievable pair of rates can be realized by means of first coloring graphs at X and Y separately (function coding) and then using Slepian-Wolf coding for these colors (correlation coding). We also obtain a single-letter characterization of the minimal joint rate. Finally, we provide simulation results based on graph coloring to establish the rate gains on real sequences.

I. INTRODUCTION

Consider a sensor network where the sensors make some correlated measurements and relay them to a central receiver. In practice, it is typical that the receiver wants to compute a specific function of the incoming data. Thus, most of the received information is irrelevant to the receiver. To increase system efficiency, we want to increase compression gains as much as possible. What can be done to obtain such coding gains?

To answer that question, we consider two sources of compression gains: correlation between the sources and the function to be computed in the end. Traditional distributed data compression would use the correlation between the sources (see, e.g., [1]), but would miss the gain associated with the end-users goal.

We begin with an example that illustrates the gains beyond simple correlation. Consider two independent sources producing n-bit integers, X and Y, uniformly. Suppose the receiver only wished to compute the parity of $X + Y$. If we ignored the receiver's goals,
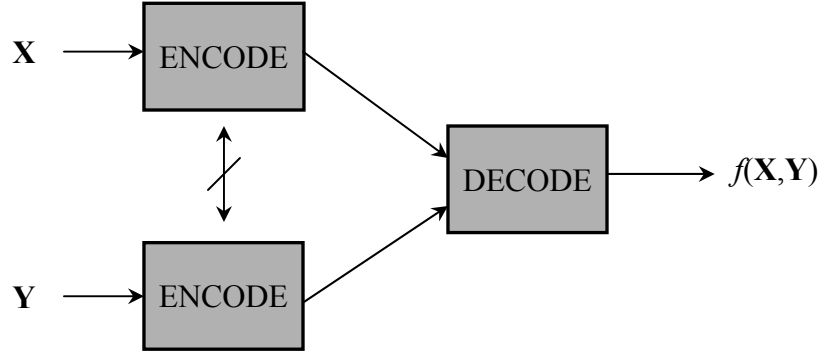
Fig. 1. The distributed encoding of X and Y such that $f(X, Y)$ is recovered at the decoder.

it would take at least n bits from the first source and n from the second. Nevertheless, in this example, it is clear that simply sending the final bit of each integer allows the receiver to compute the XOR to determine the parity of $X + Y$. Thus, using more clever compression, we have reduced $2n$ bits to 2 bits, a huge potential compression gain.

A. Setup

We consider two discrete memoryless sources, X and Y, drawn from discrete (finite) sets $\mathcal{X}$ and $\mathcal{Y}$, respectively, according to a joint distribution $p(x, y)$. 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 of this paper is to determine the set of all rates at which X and Y can be encoded in a distributed manner so that a decoder can reconstruct f with low probability of error.

Figure I-A illustrates the question of interest. Note that if $f(X, Y) = (X, Y)$, then this question is the classical distributed source coding problem considered in a seminal paper by Slepian and Wolf [2]. Next, we introduce necessary notation and definitions.

1) Formal problem statement: We denote n-sequences of random variables X and Y as $\mathbf{X}$ and $\mathbf{Y}$, respectively, where n is clear from context. Because the sequence $(\mathbf{X}, \mathbf{Y})$ is drawn i.i.d. according to $p(x, y)$, we can write $p(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{n} p(x_i, y_i)$. 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 vector extension.

For any n, R_x, and R_y, using the notation for the distributed source coding problem from [3], we define a $\left(\left(2^{nR_x}, 2^{nR_y} \right), n \right)$-*distributed functional code* for the joint source (X, Y) and function f as two encoder maps,

$$
\begin{aligned}
e_x &: \mathcal{X}^n \to \left\{ 1, \ldots, 2^{nR_x} \right\}, \\
e_y &: \mathcal{Y}^n \to \left\{ 1, \ldots, 2^{nR_y} \right\},
\end{aligned}
$$

and a decoder map,

$$
d : \left\{ 1, \ldots, 2^{nR_x} \right\} \times \left\{ 1, \ldots, 2^{nR_y} \right\} \to \mathcal{Z}^n.
$$

The *probability of error* is

$$
P_e^n = \Pr[\{(\mathbf{x}, \mathbf{y}) : f(\mathbf{x}, \mathbf{y}) \neq d(e_x(\mathbf{x}), e_y(\mathbf{y}))\}].
$$

A rate pair, (R_x, R_y), is *achievable* if there exists a sequence of $\left(\left(2^{nR_x}, 2^{nR_y}\right), n\right)$-distributed functional codes such that $P_e^n \to 0$ as $n \to \infty$. The *achievable rate region* is the closure of the set of all achievable rates.

2) The goal: Our goal is to find the achievable rate region for the above problem. Furthermore, we would like a modular architecture that effectively decouples the "correlation coding" from the "function coding," where the correlation coding could be implemented using Slepian-Wolf codes.

B. Previous results

To state some important previous results, as well as our own, we need more notation. The *characteristic graph* $G_x = (V_x, E_x)$ of X with respect to Y, f, and $p(x, y)$ is defined as follows: $V_x = \mathcal{X}$ and an edge $(x_1, x_2) \in \mathcal{X}^2$ is in E_x 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_x is the "confusability graph" from the perspective of the receiver. This was first defined by Witsenhausen [4]. The characteristic graph of Y with respect to X, f, and $p(x, y)$ is defined analogously and denoted G_y.

The graph entropy of any graph G with a distribution on its vertices was defined by Körner [5] 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. We explain the notation: $X \in W$ is equivalent to saying that $p(w, x) > 0$ implies $x \in w$ for the joint distribution $p(w, x)$. Witsenhausen [4] showed that the graph entropy is the minimum rate at which a variable can be encoded such that a function of that source is recovered (when the graph is defined so x and x' have an edge when $f(x) \neq f(x')$).

Orlitsky and Roche [6] defined an extension of Körner's graph entropy, the *conditional graph entropy*, as:

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

Further, when G is the characteristic graph of X with respect to Y, f, and p, they established that $H_G(X|Y)$ is the minimum required transmission rate from X to receiver to compute function $f(X, Y)$ with small probability of error when Y is available as side information.

A natural extension of this problem is the computation of rate distortion region. Yamamoto gives a full characterization of the rate distortion function for the side information functional compression problem [7] as a generalization of the Wyner-Ziv side-information rate-distortion region [8]. Feng, Effros, and Savari [9] build upon the Yamamoto function. The rate distortion when Y is a separate source is an open problem and future work.

To describe some of the more interesting results for graph entropies, we require more notation. Alon and Orlitsky denoted [10] the OR-power graph of G as $G^n = (V_n, E_n)$ where $V_n = V^n$ and two vertices $(\mathbf{x}_1, \mathbf{x}_2) \in E_n \subseteq V_n^2$ if any component $(x_{1i}, x_{2i}) \in E$.

We introduce standard graph vertex coloring where a coloring is any 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

---

[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.

a coloring is the entropy of the induced distribution on colors $p(c(x)) = p(c^{-1}(c(x)))$ where $c^{-1}(x) = \{\bar{x} : c(\bar{x}) = c(x)\}$ and is called a *color class*.

Next, we define ε-*colorings* (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 $\hat{p}(x, y) = p(x, y) \cdot \mathbf{1}_{(x,y) \in \mathcal{A}}$. While this is not a true probability, our definition of characteristic graph of X (or Y) with respect to Y (or X), f, and $\hat{p}$ is valid. Denote these characteristic graphs $\hat{G}_x$ and $\hat{G}_y$. Note that the edges of $\hat{G}_x$ and $\hat{G}_y$ are subsets of the edges of G_x and G_y. Finally, we say c_x and c_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 high probability set $\mathcal{A}$.

Alon and Orlitsky [10] defined the *chromatic entropy* of a graph G as:

$$H_G^\chi(X) = \min_{G\text{-colorings } c} H(c(X)).$$

We defined [11] a natural extension, the *conditional chromatic entropy*, as follows:

$$H_G^\chi(X|Y) = \min_{G\text{-colorings } c} H(c(X)|Y).$$

It should be noted that finding minimum entropy colorings is an NP-hard problem [12].

Now, we restate Körner's result [5] as follows:

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

The implications of this result are that for large enough n, one can color a high probability subgraph of the power graph G^n and send the colors to achieve near optimal functional source coding (assuming G is the characteristic graph).

We extended Körner's result to the side information case in [11] to obtain:

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

This can also decouple the function (graph coloring) from the correlation. We want to extend such graph coloring techniques to our current problem of interest.

C. Our Contribution

In this paper, we extend the results and arguments of Körner [5]. As the main result of this paper, we show that when a condition is imposed on the support of (X, Y), valid colorings completely characterize the rate region.

The condition we must impose we call the *zigzag condition* for reasons made clear by the definition. The condition requires that for any (x_1, y_1) and (x_2, y_2) in $\mathcal{X} \times \mathcal{Y}$, $p(x_1, y_1) > 0$ and $p(x_2, y_2) > 0$ imply either $p(x_1, y_2) > 0$ or $p(x_2, y_1) > 0$.

For any variables X and Y, denote by $\mathcal{R}(X, Y)$ the Slepian-Wolf achievable rate region [2]. In other words, $\mathcal{R}(X, Y)$ is the set of all (R_x, R_y) such that:

$$R_x \geq H(X|Y),$$
$$R_y \geq H(Y|X),$$
$$R_x + R_y \geq H(X, Y).$$

Also, for any n, and functions g_x and g_y defined on $\mathcal{X}^n$ and $\mathcal{Y}^n$ respectively, denote by $\mathcal{R}^n(g_x, g_y)$ the Slepian-Wolf region for the induced variables $g_x(\mathbf{X})$ and $g_y(\mathbf{Y})$ normalized by the block length. In other words, $\mathcal{R}^n(g_x, g_y)$ is the set of all (R_x, R_y) such that:

$$R_x \geq \frac{1}{n}H(g_x(\mathbf{X})|g_y(\mathbf{Y})),$$

$$R_y \geq \frac{1}{n}H(g_y(\mathbf{Y})|g_x(\mathbf{X})),$$

$$R_x + R_y \geq \frac{1}{n}H(g_x(\mathbf{X}), g_y(\mathbf{Y})).$$

Define $\mathcal{S}^\varepsilon = \bigcup_{n=1}^\infty \bigcup_{(c_x^n, c_y^n)} \mathcal{R}^n(c_x^n, c_y^n)$ where for all n, c_x^n and c_y^n are ε-colorings of G_x and G_y. We can now state our main result in the notation just given.

Theorem 1: Let $\mathcal{S}$ be the subset of $\mathcal{S}^\varepsilon$, for all $\varepsilon > 0$, such that any other subset, $\mathcal{A}$, is a subset of $\mathcal{S}$. The achievable rate region, $\mathcal{R}$, for the distributed functional source coding problem, under the zigzag condition, is the set closure of $\mathcal{S}$.

Thus, under the zigzag condition, for any n, any colorings of high probability subgraphs of G_x^n and G_y^n will allow for computation of the function. Furthermore, no other encodings can achieve smaller rates. Thus, we have decoupled the encoding such that we first color our graph, and then we apply a Slepian-Wolf code on the colors. Further, while the above characterization is not single letter, *any* coloring (except that in which the coloring is a 1-1 function of the source) necessarily does better than the Slepian-Wolf rates, by the Data Processing Inequality (cf. [3]). Also, we know that the Orlitsky-Roche bound is on the border of our region. Moreover, we can derive a characterization of the minimal joint rate, $R_x + R_y$, in terms of graph entropies.

Corollary 1: Under the zigzag condition, the joint rate $R_x + R_y = H_{G_x}(X) + H_{G_y}(Y)$ is achievable. Further, when there is a unique point that achieves this joint rate, it is the minimal joint rate.

We can approach the rates given for R_x and R_y by the previous result of Körner [5] where it requires a minimum entropy coloring of the product graph [10] (or using Körner's scheme). This implies (as shown later) that the mutual information of the minimum entropy colorings of G_x^n and G_y^n goes to zero as $n \to \infty$.

We now provide proofs of our results.

II. PROOFS

We first show that if colors are available at the decoder, the decoder can reconstruct the function. This will prove the achievability of all such rates because we know how many bits are required to send the colors using a Slepian-Wolf code. Next, we show that all valid encodings are in fact ε-colorings of the characteristic graph (and its powers). This will establish the converse.

A. Achievability

We prove the achievability of all rates in our region.

Lemma 1: Suppose we have a joint source (X, Y) drawn i.i.d. Then for any colorings c_x and c_y of G_x and G_y, respectively, there exists

$$\hat{f} : c_x(\mathcal{X}) \times c_y(\mathcal{Y}) \to \mathcal{Z}$$

such that $\hat{f}(c_x(x), c_y(y)) = f(x, y)$ for all (x, y) such that $p(x, y) > 0$.

Proof: We proceed by constructing such a $\hat{f}$. Suppose we have two colors, $\gamma \in c_x(\mathcal{X})$ and $\sigma \in c_y(\mathcal{Y})$, respectively. Then, let $\hat{x}$ and $\hat{y}$ be any (say the first) elements of the color classes $c_x^{-1}(\gamma)$ and $c_y^{-1}(\sigma)$, and let $\hat{f}(\gamma, \sigma) = f(\hat{x}, \hat{y})$.

We now show that this function is well-defined on the support set of (X, Y). Suppose (x_1, y_1) and (x_2, y_2) are in $\mathcal{X} \times \mathcal{Y}$ and both pairs have positive probability. Suppose further that $c_x(x_1) = c_x(x_2)$ and $c_y(y_1) = c_y(y_2)$. Then we know that there is no edge (x_1, x_2) in G_x or (y_1, y_2) in G_y.

By the zigzag condition, $p(x_1, y_2) > 0$ or $p(x_2, y_1) > 0$. By the definition of G_x, for all y such that $p(x_1, y)p(x_2, y) > 0$, (a) $f(x_1, y) = f(x_2, y)$. By definition of G_y, for all x such that $p(x, y_1)p(x, y_2) > 0$, (b) $f(x, y_1) = f(x, y_2)$.

If $p(x_1, y_2) > 0$, (a) would imply that $f(x_1, y_2) = f(x_2, y_2)$. Further, (b) would imply $f(x_1, y_1) = f(x_1, y_2)$. Thus, $f(x_1, y_1) = f(x_2, y_2)$.

If $p(x_2, y_1) > 0$, a similar argument applies. Thus, our function $\hat{f}$ is well-defined and has the desired property. ∎

This argument can be clearly extended to valid colorings of G_x^n and G_y^n where the function is now the n-fold Cartesian product of the original function. Thus, Lemma 1 implies that we can successfully recover our function f from colors of the characteristic graphs and their powers.

Thus, if the receiver is sent colors, it can look up f based on its table of $\hat{f}$. The question is now of faithfully (with probability of error less than ε) transmitting these colors to the receiver. However, when we consider the colors as sources, we know the achievable rates.

Lemma 2: For any n, colorings c_x^n and c_y^n of G_x^n and G_y^n, respectively, the following rates, (R_x^c, R_y^c), are achievable:

$$R_x^c \geq H(c_x^n(\mathbf{X})|c_y^n(\mathbf{Y})),$$
$$R_y^c \geq H(c_y^n(\mathbf{Y})|c_x^n(\mathbf{X})),$$
$$R_x^c + R_y^c \geq H(c_x^n(\mathbf{X}), c_y^n(\mathbf{Y})).$$

Proof: This follows directly from the Slepian-Wolf Theorem [2] for the separate encoding of correlated sources. ∎

Lemma 3: Lemma 2 applies when the colorings are ε-colorings.

Proof: For all elements of the high probability subset that the coloring covers, the above proof works. Everything else, by definition, has probability less than ε. Declare an error when the elements to color are not in the high probability subset. Thus, we have an error if the elements we want to color are, in this sense, atypical, or if they are typical and the Slepian-Wolf code fails. The probability of this event is less than 2ε when the probability of a Slepian-Wolf code failure is less than ε. Thus, the rates for successful transmission of the coloring, given in Lemma 2, are achievable when c_x^n and c_y^n are ε-colorings. ∎

Finally, in light of the fact that n source symbols are encoded for each color, the achievable rates for X and Y are: for any ε, all rates (R_x, R_y) such that

$$R_x \geq \frac{1}{n} H(c_x^n(\mathbf{X})|c_y^n(\mathbf{Y})),$$

$$R_y \geq \frac{1}{n} H(c_y^n(\mathbf{Y})|c_x^n(\mathbf{X})),$$

$$R_x + R_y \geq \frac{1}{n} H(c_x^n(\mathbf{X}), c_y^n(\mathbf{Y})).$$

where c_x^n and c_y^n are achievable ε-colorings. Thus every $(R_x, R_y) \in \mathcal{S}^\varepsilon$ is achievable for all $\varepsilon > 0$. Therefore, every $(R_x, R_y) \in \mathcal{S}$ is achievable.

B. Converse

Next, we prove that any distributed functional source code with small probability of error induces a coloring.

Suppose $\varepsilon > 0$. Define for all (n, ε),

$$\mathcal{F}_\varepsilon^n = \{\hat{f} : \Pr[\hat{f}(\mathbf{X}, \mathbf{Y}) \neq f(\mathbf{X}, \mathbf{Y})] < \varepsilon\}.$$

This is the set of all functions that equal f to within ε probability of error. (Note that all achievable distributed functional source codes are in $\mathcal{F}_\varepsilon^n$ for large enough n.)

Lemma 4: Consider some function $g : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. Any distributed functional code that reconstructs g with zero-error (with respect to a distribution $p(x, y)$) induces colorings on the characteristic graphs of X and Y with respect to g, $p(x, y)$, and Y and X, respectively.

Proof: Suppose we have encoders e_x and e_y, decoder d, and characteristic graphs G_x^n and G_y^n. Then by definitions, a zero-error reconstruction implies that for any $(\mathbf{x}_1, \mathbf{y}_2)$, $(\mathbf{x}_2, \mathbf{y}_2)$ such that if $p(\mathbf{x}_1, \mathbf{y}_1) > 0$, $p(\mathbf{x}_2, \mathbf{y}_2) > 0$, $e_x(\mathbf{x}_1) = e_x(\mathbf{x}_2)$, and $e_y(\mathbf{y}_1) = e_y(\mathbf{y}_2)$, then

$$f(\mathbf{x}_1, \mathbf{y}_1) = f(\mathbf{x}_2, \mathbf{y}_2) = d(e_x(\mathbf{x}_1), e_y(\mathbf{y}_1)). \tag{1}$$

We now show that e_x and e_y are valid colorings of G_x^n and G_y^n. We demonstrate the argument for X. The argument for Y is analogous. We proceed by contradiction. There must be some edge with both vertices with the same color. In other words, there must exist $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$ such that $p(\mathbf{x}_1, \mathbf{y})p(\mathbf{x}_2, \mathbf{y}) > 0$, $e_x(\mathbf{x}_1) = e_x(\mathbf{x}_2)$, and $f(\mathbf{x}_1, \mathbf{y}) \neq f(\mathbf{x}_2, \mathbf{y})$. This is impossible (by taking $\mathbf{y}_1 = \mathbf{y}_2 = \mathbf{y}$ in equation (1)). Hence, we have induced colorings of the characteristic graphs. ∎

We now show that any achievable distributed functional code also induces an ε-coloring of the characteristic graphs.

Lemma 5: All achievable distributed functional codes induce ε-colorings of the characteristic graphs.

Proof: Let $g(\mathbf{x}, \mathbf{y}) = d(e_x(\mathbf{x}), e_y(\mathbf{y})) \in \mathcal{F}_\varepsilon^n$ be such a code. Then, we know that a zero-error reconstruction (with respect to p) of g induces colorings, e_x and e_y, of the characteristic graphs with respect to g and p by Lemma 4. Let the set of all $(\mathbf{x}, \mathbf{y})$ such that $g(\mathbf{x}, \mathbf{y}) \neq f(\mathbf{x}, \mathbf{y})$ be denoted as $\mathcal{C}$. Then because $g \in \mathcal{F}_\varepsilon^n$, we know that $\Pr[\mathcal{C}] < \varepsilon$. Therefore, the functions e_x and e_y restricted to $\mathcal{C}$ are ε-colorings of G_x and G_y (by definition). ∎

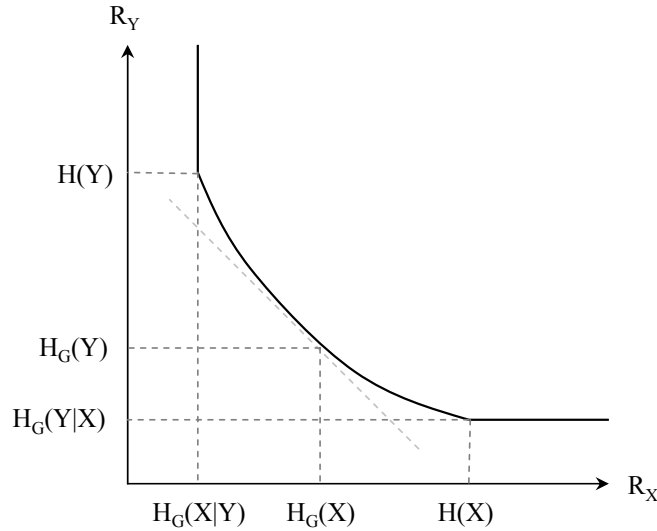Thus, the Lemma 3 and Lemma 5 establish Theorem 1 in full.

Fig. 2. The rate region when there is an extreme point. The Orlitsky-Roche points are also shown.

C. Rate region

The region given in Theorem 1 has several interesting properties. First, it is convex. This can be seen using time-sharing arguments for any two points in the region (as is done for all rate region problems). Second, there is a unique minimal joint rate $R_x + R_y$. This can be seen from Figure 2; note that G is assumed to be G_x for points on the R_x axis and G_y for points on the R_y axis.

This rate region has (on its "corners") the Orlitsky-Roche points, $(H_{G_x}(X|Y), H(Y))$ and $(H(X), H_{G_y}(Y|X))$, which can be achieved with graph coloring, in the limit sense, as shown in [11]. For any rate $R_x \in (H_{G_x}(X|Y), H(X))$, the joint rate required is less than or equal to the joint rate required by a time-sharing of the Orlitsky-Roche scheme.

We can give another point on the "edge" of the region that can be characterized with graph entropies when there is a unique rate pair that achieves the minimal joint rate. We next prove Corollary 1 and describe this rate pair.

D. Corollary

First we note that the rate pair $(H_{G_x}(X), H_{G_y}(Y))$ can be achieved. This is true by Körner's result showing that graph colorings can achieve these rates. Because they are colorings, they must be achievable rates.

Next, suppose there is a pair that achieves the minimal joint rate. On the part of the rate region where $R_x < H(X)$ and $R_y < H(Y)$, we can describe all edge points as close to the corner points of $\mathcal{R}^n(c_x, c_y)$ for some ε-colorings c_x and c_y. Further, this implies that the corner points, as one moves further away from $R_x = H(X)$ and $R_Y = H(Y)$ must converge. Suppose this point of convergence is (R_1, R_2).

Thus, for any $\varepsilon > 0$, there exists an N such that for all $n > N$, $\left|R_1 - \frac{1}{n}H(c_x^n|c_y^n)\right| < \varepsilon$, $\left|R_2 - \frac{1}{n}H(c_y^n|c_n^n)\right| < \varepsilon$, and $\left|R_1 + R_2 - \frac{1}{n}H(c_x^n, c_y^n)\right| < \varepsilon$, for some c_x^n and c_y^n, ε-colorings of G_x^n and G_y^n, respectively. We can combine the above inequalities to show $\left|R_1 - \frac{1}{n}H(c_x^n)\right| < 2\varepsilon$, $\left|R_2 - \frac{1}{n}H(c_y^n)\right| < 2\varepsilon$, and $\left|R_1 + R_2 - \left(\frac{1}{n}H(c_x^n) + \frac{1}{n}H(c_y^n)\right)\right| < 4\varepsilon$. This, the joint rate is within 4ε of $\frac{1}{n}H(c_x^n) + \frac{1}{n}H(c_y^n)$, and it is minimal. We know that

those rates are minimized in the limit as n increases without bound to $H_{G_x}(X) + H_{G_y}(Y)$ (by Körner's result [5]). Thus, $R_x = H_{G_x}(X) + \varepsilon$ and $R_y = H_{G_y}(Y) + \varepsilon$ will achieve the minimum joint (achievable) rate. Thus, Corollary 1 is established.

This corollary implies that minimum entropy colorings have decreasing mutual information as n increases. Thus, the Slepian-Wolf codes are unnecessary when achieving the minimal joint rate! (Nevertheless, finding the minimum entropy colorings is, again, NP-hard.)

E. Support issues

Throughout our analysis, we have assumed that the zigzag condition holds. Practically, this assumption is quite reasonable as there is seldom the need to account for zero probability events. In its absence, our simple analysis that colors are sufficient to compute the function fails. Nevertheless, it is unclear whether a more robust proof exists that does not require this condition. The next section shows some of our simulation results using real data that shows that indeed that condition is not unreasonable.

III. SIMULATION RESULTS

We obtained tracking data[2] from SRI. This data represents GPS location data (for so-called "Blue Force Tracking"). It includes information on various mobiles, including latitude and longitude coordinates. We ignored the other information for the purpose of this simulation.

We focused on two mobiles, our sources. We assume that our sources are the positional differences (i.e. Δ-encoding), $\mathbf{X}_1$ and $\mathbf{X}_2$, where $\mathbf{X}_i$ is actually a pair, (X_{1a}, X_{1b}) of the latitude and longitude data. The empirical distributions for these suggest that the latitude and longitude data for each source is uncorrelated, and across sources, the information is also uncorrelated. This means that any gains will be because the complexity of the function has decreased from $f(X, Y) = (X, Y) \in \mathcal{X} \times \mathcal{Y}$ to recovering the function described next. We use $f(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{1}_{|\mathbf{x}_1 - \mathbf{x}_2| < D}$ where $D = 1.1 \cdot 10^{-5}$ to construct our characteristic graphs. Thus, this is a proximity function that equals 1 when the two mobiles are within range of each other (and 0 otherwise).

We chose this value of D because it occurred with empirical probability 0.437. If we chose a higher probability event, we would expect higher gains. We use $n = 1$ for the graphs and do not consider power graphs. Had we, our rate gains would be higher, though the computational complexity would increase exponentially with n. Our coloring algorithm was a simple greedy algorithm that did not use any of the probability information nor was it an ε-coloring. We expect better gains with more advanced graph coloring schemes.

We do not actually simulate the Slepian-Wolf coder, but rather assume the Slepian-Wolf code will introduce no more error to the colors than it will to the original sources. Thus, the entropy provides a good comparison point. All units are bits. The final column represents the percent reduction of the second column with respect to the first column. We find:

$H(X_1) = 3.94$	$H(c(X_1)) = 3.51$	10.9%
$H(X_2) = 3.99$	$H(c(X_2)) = 2.56$	35.8%

[2]Data available at: `http://www.ai.sri.com/ajh/isat`.

Thus, the joint rate reduction is 23.4%.

IV. Conclusion

We considered the distributed functional source coding problem. We found it was possible to decouple the correlation coding from the functional coding under the zigzag condition. Thus, we found that a layering of the functional compression problem is possible. Further, the entire rate region can be accounted for in the layering. We have not found a single-letter characterization for the rate region, but we have shown three points in the rate region that can be described in terms of Körner's graph entropy and Orlitsky and Roche's conditional graph entropy. Further, we have shown that in order to achieve the minimal joint rate, we need to find minimum entropy colorings of the characteristic graphs. In this case, the Slepian-Wolf code is unnecessary because the colorings have very small mutual information. Moreover, we have shown, through simulation, that graph coloring algorithms need not be complex to lead to large compression gains.

Future work includes a characterization of the boundary points for the region (not just the three points we have computed). Further, we want to relax the zigzag condition. While this condition is rather mild, we want to see if lower rates could be achieved as a result of such a relaxation. Further, we want to relax the fidelity criterion, i.e. examine the rate distortion problem. Yamamoto [7] and Feng, Effros, and Sevari [9] have demonstrated rate distortion functions, and we wish to see if a graph coloring approach could be used to achieve those rates.

Acknowledgment

References

[1] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "Networked Slepian-Wolf: Theory, Algorithms and Scaling Laws," *IEEE Trans. Inform. Theory*, 2005, to appear.

[2] 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.

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

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

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

[11] V. Doshi, D. Shah, M. Médard, and S. Jaggi, "Graph coloring and conditional graph entropy," in *Asilomar Conference on Signals, Systems, and Computers*, Nov. 2006.

[12] J. Cardinal, S. Fiorini, and G. V. Assche, "On minimum entropy graph colorings," ISIT 2004, June 27–July 2 2004, p. 43.