Source Coding with Distortion
through Graph Coloring

Vishal Doshi, Devavrat Shah, and Muriel Médard
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139
Email: {vdoshi, devavrat, medard} @ mit.edu

Abstract—We consider the following rate distortion problem: given a source X and correlated, decoder side information Y, find the minimum encoding rate for X required to compute $f(X, Y)$ at the decoder within distortion D. This is a generalization of the classical Wyner-Ziv setup and was resolved by Yamamoto (1982). However, this result involved an auxiliary random variable that lacks explicit meaning.

To provide a more direct link between this variable and the function f, Orlitsky and Roche (2001) established the minimal rate required in the zero-distortion case as an extension of Körner's graph entropy. Recently, we (with Jaggi) showed that the zero-distortion rate can be achieved by minimum entropy graph coloring of an appropriate product graph. This leads to a modular architecture for functional source coding with a preprocessing "functional coding" scheme operating on top of a classical Slepian-Wolf source coding scheme.

In this paper, we give a characterization of Yamamoto's rate distortion function in terms of a reconstruction function. This (non-single-letter) characterization is an extension of our previous results as well as Orlitsky and Roche's results. We obtain a modular scheme operating with Slepian-Wolf's scheme for the problem of functional rate distortion. Further, we give an achievable rate (with single-letter characterization) utilizing this scheme that intuitively extends our previous results.

I. INTRODUCTION

A. Motivation

Consider a network of sensors in which each node transmits its measurements to a central receiver. We consider the source coding aspect of this problem with the assumption that a reduction in the source coding rate translates to a bandwidth reduction. Often, the information from each sensor node is correlated with that of other nodes. Thus, the sensors are not required to transmit all of their information. Moreover, the receiver often wishes only to compute a function of the sensed information (up to a distortion D); it has no use for the exact information. This suggests that using a clever scheme, the required transmission rate can be reduced.

Recent work in designing efficient distributed coding schemes, such as work by Pradhan and Ramachandran [1] and Coleman et al. [2], allow for a rate reduction due to correlation in the data. However, they do not account for the rate reduction possible by considering that the receiver is only required to compute some function of the transmitted information.

In this paper, the end goal will be to obtain a scheme that improves the transmission rate by first preprocessing the data

for a given function and distortion, and then using efficient distributed source coding schemes on the preprocessed information. Next, we present a motivating example.

Example 1: Consider two sources uniformly and independently producing k-bit integers X and Y; assume $k \geq 2$. The source Y is decoder side information. We assume independence to bring to focus the compression gains from using knowledge of the function.

First, suppose $f(X, Y) = (X, Y)$ is to be computed at the decoder with zero-distortion. Then, the rate at which X must encode its information is k bits per symbol (bps).

Next, suppose $f(X, Y) = X + Y \mod 4$. The value of $f(X, Y)$ depends only upon the final two bits of both X and Y. Thus, at most (and in fact, exactly) 2 bps is the encoding rate. Note that the rate gain, $k - 2$, is unbounded because we are reducing a possibly huge alphabet to one of size 4.

Finally, suppose $f(X, Y) = X + Y \mod 4$ as before, but the decoder is allowed to compute f up to a 1-bit distortion. By 1-bit distortion, we mean a Hamming distortion function. One possible coding scheme would simply encode the single least significant bit X. Then one could compute the least significant bit of $f(X, Y)$, thus achieving an encoding rate of 1 bps, further reducing the rate.

The above example describes a function that readily lends itself to further compression. In general, however, this may not hold. For example, if the function were not separable, a more complex coding scheme would be necessary, and it is not clear a priori that such a scheme even exists. Thus, we need a more general systematic method that leads to a coding scheme for any deterministic function.

B. Setup and Problem Statement

As the first step towards finding explicit schemes that require minimal transmission rate for computing functions within a certain distortion fidelity criterion, we consider the case of a single source transmitting to a receiver (with side information) that wishes to compute a function of the source information and the side information. We are given a random variable X taking values in a finite set $\mathcal{X}$. The side information available at the receiver is denoted by the random variable Y taking value in the finite set $\mathcal{Y}$. Both X and Y are discrete memoryless sources with joint distribution $p(x, y)$, with $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Denote n-sequences of the random
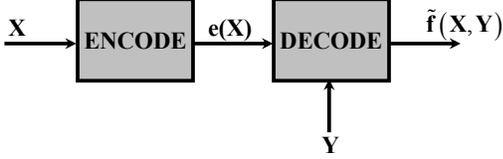
Fig. 1. **X** is encoded such that $\tilde{f}(\mathbf{X}, \mathbf{Y})$, a representation of $f(\mathbf{X}, \mathbf{Y})$ to within distortion D, can be computed using $e(\mathbf{X})$ and side information **Y**.

variables as **X** and **Y** where $(\mathbf{X}, \mathbf{Y}) = \{X_i, Y_i\}_{i=1}^n$ and n will be clear from context.

The receiver's function of interest is the deterministic function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, or its natural vector extension $f : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}^n$. The receiver wishes to compute f within distortion D with respect to a given distortion function $d : \mathcal{Z} \times \mathcal{Z} \to [0, \infty)$. The natural vector extension of the distortion function is

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

where $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{Z}^n$. As in [3], we assume that the distortion function satisfies $d(z_1, z_2) = 0$ if and only if $z_1 = z_2$. (Otherwise, one can redefine the equivalence classes of the function values to make this condition hold.) This restriction forces the vector extension to satisfy the same property.

We say that a code with parameters (n, R, D) produces f within distortion D if for every $\varepsilon > 0$ and sufficiently large $n = n_\varepsilon$, there exists a source encoder $e_1^n : \mathcal{X}^n \to \{1, \ldots, 2^{n(R+\varepsilon)}\}$, and decoder $e_2^n : \{1, \ldots, 2^{n(R+\varepsilon)}\} \times \mathcal{Y}^n \to \mathcal{Z}^n$, such that

$$\mathbb{E}\left[d(f(\mathbf{X}, \mathbf{Y}), e_2(e_1(\mathbf{X}), \mathbf{Y}))\right] \leq D + \varepsilon.$$

We wish to find efficient coding-decoding schemes with improved rates R and a characterization of the minimal R. See Figure 1 for an illustration of the question of interest.

C. Relevant Previous Work

We introduce the necessary definitions and notation to explain previous work. Given the source X, the *characteristic graph* of X with respect to Y, f and $p(x, y)$ is a graph $G = (V, E)$ with vertex set $V = \mathcal{X}$ and edge set E defined as follows: $(x_1, x_2) \in E$ if there exists some $y \in \mathcal{Y}$ such that $p(x_1, y)p(x_2, y) > 0$ and $f(x_1, y) \neq f(x_2, y)$. Effectively, G describes the *confusability* of various values of X and captures the essence of f. This was first defined by Shannon [4].

We extend the notion of the confusability graph to include distortion and define the *D-characteristic graph* of X with respect to Y, f, and $p(x, y)$. The vertex set is again $V = \mathcal{X}$. The edge set is now E^D where a pair $(x_1, x_2) \in E$ if there exists some $y \in \mathcal{Y}$ such that $p(x_1, y)p(x_2, y) > 0$ and $d(f(x_1, y), f(x_2, y)) > D$. Denote this graph as G^D. Because the $d(z_1, z_2) = 0$ iff $z_1 = z_2$, the 0-characteristic graph, G^0, is the characteristic graph G.

Consider a graph $G = (V, E)$ with the distribution, $p(x)$, over its vertices V. Let X denote the random variable over the vertices with that distribution. The graph entropy of G with respect to distribution of X, first introduced by Körner [5], is defined as

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

where $\Gamma(G)$ is the collection of all independent sets[1] of G. To clarify, $X \in W$ means that the joint distribution $p(w, x)$ on $\Gamma(G) \times \mathcal{X}$ is such that $p(w, x) > 0$ implies $x \in w$. This graph entropy can be shown to be the minimal rate at which X must be sent to compute a function $f(X)$ at the receiver [6] with vanishing error. Equivalently, this is the characterization of the minimal rate requirement for our problem of interest in the special case when cardinality of $\mathcal{Y}$ is 1 and the distortion is $D = 0$.

This was further improved by Orlitsky and Roche with the restriction $D = 0$ of our problem of interest. They extended the definition of graph entropy to *conditional graph entropy* defined as

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

As in the definition of graph entropy, W is a random variable over space of independent sets $\Gamma(G)$ satisfying Markov property $W - X - Y$ and $X \in W$ as defined earlier. They showed [7] that $H_G(X|Y)$ is the minimum rate at which X must be encoded in order for a receiver with side information Y to compute a function $f(X, Y)$ at zero-distortion (i.e. $D = 0$) with vanishing probability of error.

The rate distortion function, $R(D)$, for the functional source coding problem with side information evaluated at D is defined as the *minimum* R such that there is a code with parameters (n, R, D) that produces f within distortion D. In other words, we can equivalently say that there exists a sequence (in n) of codes such that $\lim_{n \to \infty} E[d(f(\mathbf{X}, \mathbf{Y}), e_2(e_1(\mathbf{X}), \mathbf{Y}))] \leq D$ because as $n \to \infty$, $\varepsilon \to 0$.

As stated earlier, Yamamoto fully characterized the rate distortion function [8] in terms of auxiliary random variable W. The rate distortion function is:

$$R(D) = \min_{p \in \mathcal{P}(D)} I(W; X|Y)$$

where $\mathcal{P}(D)$ is the collection of distributions on (W, X, Y) such that $W - X - Y$ forms a Markov chain and such that there exists a $g : \mathcal{W} \times \mathcal{Y} \to \mathcal{Z}$, where $\mathcal{W}$ is the alphabet of W, with $E[d(f(X, Y), g(W, Y))] \leq D$. The cardinality of W can be bounded as in the Wyner-Ziv result ($|\mathcal{W}| \leq |\mathcal{Y}| + 1$). Feng, Effros, and Sevari [9] considered the same problem with $\tilde{X}$ and $\tilde{Y}$, noisy versions of X and Y, available instead of the sources.

---

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

D. Our Contribution

When $D = 0$, any distribution over independent sets of the characteristic graph (with the Markov property $W - X - Y$) is also in $\mathcal{P}(0)$. Further, any distribution in $\mathcal{P}(0)$ can be thought of as a distribution on independent sets of a graph G. When $D > 0$, is the same true? Can we parameterize the graph G with D and improve on the rate?

First, we show that finding g in the Yamamoto result is equivalent to finding a suitable reconstruction function, $\hat{f}$. Let $\mathcal{F}_m(D)$ denote the set of all functions $\hat{f}_m : \mathcal{X}^m \times \mathcal{Y}^m \to \mathcal{Z}^m$ for any m with the property that

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

Above, $n \to \infty$ refers to the block length of $\hat{f}$-values increasing without bound. In other words, the functions in the expectation above are actually functions on $\mathcal{X}^{mn} \times \mathcal{Y}^{mn}$.

Let $\mathcal{F}(D) = \bigcup_{m \in \mathbb{N}} \mathcal{F}_m(D)$. Let $G(\hat{f})$ denote the characteristic graph of $\mathbf{X}$ with respect to $\hat{f}$, $\mathbf{Y}$ and $p(\mathbf{x}, \mathbf{y})$ for any $\hat{f} \in \mathcal{F}(D)$. Note that this must be a subgraph of the characteristic graph G^m (for the appropriate m). Because G^m has finitely many edges, there are only finitely many subgraphs. Thus, for each m, there are finitely many graphs $G(\hat{f}_m)$ to consider and the total number of *graphs* (as opposed to functions) to consider is countable. For each m and all functions in $\mathcal{F}_m(D)$, denote, for brevity, the normalized graph entropy $\frac{1}{m} H_{G(\hat{f})}(\mathbf{X}|\mathbf{Y})$ as $H_{G(\hat{f})}(X|Y)$. Then,

Theorem 1:

$$R(D) = \min_{\hat{f} \in \mathcal{F}(D)} H_{G(\hat{f})}(X|Y)$$

Theorem 1 gives meaning to the auxiliary random variable from Yamamoto's result. The auxiliary random variable is over the independent sets of the characteristic graph with respect to a distortion-D approximation $\hat{f}$ of f. The above is not single-letter, but motivates the single-letter achievable rate we describe next.

We parameterize the characteristic graph G on D (specifically, we use G^D) and implicitly find a class of $\hat{f}$ (up to graph equivalence), though the rate given by this graph is perhaps not tight. By restricting to a subset of $\mathcal{F}_1(D)$, we make the optimization defined by Theorem 1 tractable. In fact, this is now a finite optimization. The modular architecture implied by the following theorem is the main contribution of this paper.

Theorem 2: The rate $H_{G^D}(X|Y)$ is achievable, where G^D is the D-characteristic graph of X with respect to Y, f, and $p(x, y)$.

E. Implications

A coding scheme that employs a modular architecture (preprocessing followed by a Slepian-Wolf [15] code) is an immediate corollary. See Figure 2. The preprocessing is coloring of the data with respect to the graph G^D. This is followed by a Slepian-Wolf code. At the decoder side the reconstructed colors can be matched up to a particular $\hat{f}$ value,
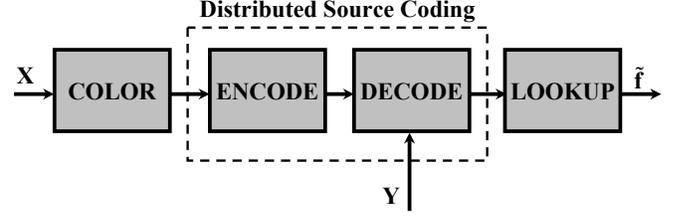

Distributed Source Coding

Fig. 2. Coloring based coding allows separation between functional coding and distributed source coding to compute a function of the sources within a given expected distortion as shown above.

which is within D of f. This follows by our previous work (with Jaggi) [10]. The details are given in section III.

II. PROOFS

A. Proof of Theorem 1

We restate Theorem 1 for completeness:

$$R(D) = \min_{\hat{f} \in \mathcal{F}(D)} H_{G(\hat{f})}(X|Y)$$

Proof: We prove that the given characterization is valid by first showing the rate $H_{G(\hat{f})}(X|Y)$ is achievable for any $\hat{f} \in \mathcal{F}(D)$, and next showing that every achievability scheme must be in $\mathcal{F}(D)$.

By Orlitsky and Roche [7], we know that the rate $H_{G(\hat{f})}(X|Y)$ is sufficient to determine the function $\hat{f}(\mathbf{X}, \mathbf{Y})$ at the receiver. By definition, $\lim_{n \to \infty} E[d(f(\mathbf{X}, \mathbf{Y}), \hat{f}(\mathbf{X}, \mathbf{Y}))] \leq D$. Thus, the rate $H_{G(\hat{f})}(X|Y)$ is achievable.

Next, suppose we have any achievable rate R, with corresponding *sequence* of encoding and decoding functions e_1^n and e_2^n respectively. Then the function $\hat{f}(\cdot, \cdot) = e_2^n(e_1^n(\cdot), \cdot)$ is a function $\hat{f} : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}^n$ with the property (by achievability) that $\lim_{n \to \infty} E[d(f(\mathbf{X}, \mathbf{Y}), \hat{f}(\mathbf{X}, \mathbf{Y}))] \leq D$ (again because as $n \to \infty$, ε is driven to 0). Thus, $\hat{f} \in \mathcal{F}(D)$, completing the proof of Theorem 1. ∎

This characterization is mostly illustrative. Indeed, $\mathcal{F}(D)$ is an (uncountably) infinite set, but as stated before, the set of graphs associated with these functions is countably infinite. Moreover, any allowable graph dictates an ordinal function, but it has no meaning in terms of distortion. Given the ordinal function $\hat{f}$, choosing the cardinal values that minimize expected distortion is a tractable optimization problem. Further, this shows that if one could find an approximation function $\hat{f}$, the compression rate will improve (even when $\hat{f}$ is not optimal).

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. This motivates Theorem 2 where we use the D-characteristic graph to look at a subset of $\mathcal{F}(D)$.

B. Proof of Theorem 2

In this section, we prove Theorem 2, which states that $H_{G^D}(X|Y)$ is an achievable rate.

Proof: We show that $H_{G_D}(X|Y)$ is achievable by demonstrating that any distribution on (W, X, Y) satisfying $W - X - Y$ and $X \in W \in \Gamma(G^D)$ also satisfies the Yamamoto requirement (i.e. is also in $\mathcal{P}(D)$).

Suppose $p(w, x, y)$ is such that $p(w, x, y) = p(w|x)p(x, y)$, or $W - X - Y$ is a Markov chain. Further suppose that $X \in W \in \Gamma(G^D)$. Then define $g(w, y) = f(x^*, y)$ where x^* is any (say, the first) $x \in w$ with $p(x^*, y) > 0$. This is well-defined because the nonexistence of x such that $p(x, y) > 0$ is a zero probability event, and $x \in w$ occurs with probability one by assumption.

Further, because w is an independent set, for any $x_1, x_2 \in w$, one must have $(x_1, x_2) \notin E^D$, the edge set of G^D. By definition of G^D, this means that for all $y \in \mathcal{Y}$ such that $p(x_1, y)p(x_2, y) > 0$, it must be the case that $d(f(x_1, y), f(x_2, y)) \leq D$.

Therefore,

$$E[d(f(X, Y), g(W, Y))] = E[d(f(X, Y), f(X^*, Y))] \leq D$$

because both $X \in W$ and $X^* \in W$ are probability 1 events.

We have shown that for a given distribution achieving the conditional graph entropy, there is a function g on $\mathcal{W} \times \mathcal{Y}$ that has expected distortion less than D. In other words, any distribution satisfying $W - X - Y$ and $X \in W \in \Gamma(G^D)$ is also in $\mathcal{P}(D)$. Further, any such distribution can be associated with a coding scheme, by Orlitsky and Roche's work [7], that achieves the rate $I(W; X|Y)$. When the distribution is chosen such that $I(W; X|Y)$ is minimized, this is by definition equal to $H_{G^D}(X|Y)$. Thus, the rate $H_{G^D}(X|Y)$ is achievable, proving Theorem 2 and providing a single-letter upper bound for $R(D)$. ∎

III. Coding Scheme

To describe why a modular coding scheme is implied by Theorems 1 and 2, we restate some past work in this area.

The *OR-product graph* of $G = (V, E)$, denoted $G^n = (V_n, E_n)$, is defined as $V_n = V^n$, and two vertices $(\mathbf{x}_1, \mathbf{x}_2) \in E_n$ if any component $(x_{1i}, x_{2i}) \in E$.

A *graph coloring* is a function $c : V \to \mathbb{N}$ with the property that for any two $x_1, x_2 \in V$ with $c(x_1) = c(x_2)$, it must be the case that $(x_1, x_2) \notin E$. The entropy of any coloring is given by the distribution induced on the colors ($p(c(x)) = p(c^{-1}(c(x)))$) where $c^{-1}(x) = \{\bar{x} : c(\bar{x}) = c(x)\}$.

This definition can be extended to colorings of high probability subgraphs of the original graph. For any $\varepsilon > 0$, an ε-*coloring* of a graph G is defined next. Let $\mathcal{A} \subseteq \mathcal{X} \times \mathcal{Y}$ be such that $p(\mathcal{A}) \geq 1 - \varepsilon$. Let $\hat{p}(x, y) = p(x, y|(x, y) \in \mathcal{A})$ for $(x, y) \in \mathcal{A}$, and 0 otherwise. Let $\hat{G}$ denote the characteristic graph of X with respect to Y, f, and $\hat{p}$. Note that the edge set of $\hat{G}$ is a subset of the edge set of G. Any coloring c of $\hat{G}$ is an ε-coloring of G.

Let the *conditional chromatic entropy* of a graph G on vertices $\mathcal{X}$ with respect to a distribution $p(x, y)$ be defined as:

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

By our previous work (with Jaggi) [10],

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

In other words, coloring high probability subsets of sufficiently large power graphs of the characteristic graph is sufficient (and in the limit, necessary) to compute a function at zero distortion with arbitrary probability of error.

For Theorem 1, this implies that given any reconstruction function $\hat{f} \in \mathcal{F}(D)$, the minimum entropy graph coloring of $G(\hat{f})$ is necessary and sufficient to compute $\hat{f}$. Computation of $\hat{f}$ means a distortion-D reconstruction of f. Thus, a graph coloring scheme as in Figure 2 is achievable for any valid reconstruction.

This directly applies to the result of Theorem 2, where the graph is G^D; coloring it and its powers is sufficient to compute the function $f(\mathbf{X}, \mathbf{Y})$ to within a distortion D. This gives the modular scheme we have sought. We now describe the particulars of a coding scheme like that shown in Figure 2.

First, all atypical elements would be removed from consideration. Then, the D-characteristic graph would be constructed based on the function and the distribution (conditioned on typicality). Finally, the graph would be colored, and the colors are sent over the channel using a Slepian-Wolf code (at rate $H(c(X)|Y)$). These colorings are then ε-colorings of the original graph G^D. At the receiver, the color is recovered. And from the color, the function can be computed (to within distortion D).

Finding minimum-entropy graph colorings (i.e. graph colorings that achieve the (conditional) chromatic entropy) is NP–complete [11]. Nevertheless, the scheme provided is layered. The distributed source coding module is well-understood (e.g. [1], [2]). The graph coloring module has been well-studied; algorithms and heuristics that perform well exist (e.g. [12], [13]).

IV. Conclusion

In this paper, we have shown that Yamamoto's rate distortion function can be achieved by finding the conditional graph entropy minimizing approximation (to within distortion D) of the function f. This is a hard optimization problem. Nevertheless, the insight gained from this perspective led to a simple modular achievability scheme.

The modules in the scheme are a graph coloring component followed by a distributed source coding component. This separation into a well-understood problem (distributed source coding) and a NP–complete problem (minimum entropy graph coloring) is beneficial because it allows for the use of the many heuristics available in the graph coloring literature. This scheme is likely suboptimal, but because $E^D \subseteq E$, we must

have $H_{G^D}(X|Y) \leq H_G(X|Y) \leq H(X|Y)$, and thus there is guaranteed (weak) improvement.

The authors (with Jaggi) have examined the case where Y is not side information in [14], completely characterizing the region when a restriction is placed on the source distribution. This extended the work of Slepian and Wolf [15]. The authors intend to similarly extending the results for side information from this paper to the case where Y is not side information.

One can approach many functional source coding problems as graph coloring problems. This approach is attractive because it reduces the problems into smaller problems that are well-studied, if not completely solved.

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] 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.

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

[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] 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.

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

[8] 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.

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

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

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

[12] 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.

[13] 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.

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

[15] 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.