# On Zero-Error Source Coding With Decoder Side Information

Prashant Koulgi, *Student Member, IEEE*, Ertem Tuncel, *Student Member, IEEE*, Shankar L. Regunathan, *Member, IEEE*, and Kenneth Rose, *Fellow, IEEE*

***Abstract*—Let $(X, Y)$ be a pair of random variables distributed over a finite product set $V \times W$ according to a probability distribution $P(x, y)$. The following source coding problem is considered: the encoder knows $X$, while the decoder knows $Y$ and wants to learn $X$ *without error*. The minimum zero-error asymptotic rate of transmission is shown to be the complementary graph entropy of an associated graph. Thus, previous results in the literature provide upper and lower bounds for this minimum rate (further, these bounds are tight for the important class of perfect graphs). The algorithmic aspects of instantaneous code design are considered next. It is shown that optimal code design is $NP$-hard. An optimal code design algorithm is derived. Polynomial-time suboptimal algorithms are also presented, and their average and worst case performance guarantees are established.**

***Index Terms*—Complementary graph entropy, graph coloring, graph entropy, lossless coding, $NP$-completeness, side information, Slepian–Wolf, zero-error capacity.**

## I. Introduction

THE problem of zero-error source coding when the decoder has side information unknown to the encoder is considered. With the advent of networks (such as the Internet), distributed storage and retrieval of very large databases is seen as a promising application. Recently, this has renewed interest in multiterminal source coding frameworks such as distributed source codes (see, for example, [29], [17], [23]). The scenario of the *side-information problem*—where the encoder tries to exploit side information about the source available to the decoder but not to itself—is important both as a canonical distributed source coding system, and as a fundamental building block of more intricate real-world systems. The zero-error version of this problem, apart from its significance in practical applications, has also been studied due to its connections with basic graph-theoretic quantities.

Witsenhausen initiated the investigation of the zero-error side-information problem in [27], where he characterized the side-information structure as a confusability graph defined on the source alphabet. With this characterization, fixed-length side-information codes were equivalent to colorings of the associated graph. Alon and Orlitsky considered variable-length zero-error codes in [1]. They defined two classes of such codes, where the second class is obtained as a subclass of the first by restricting the structure of allowed codes. Along with establishing upper and lower bounds on the rates of scalar codes, they characterized the minimum asymptotic rate needed for the subclass of codes above as the graph entropy [19] of the associated graph. But a single-letter characterization of the minimum asymptotic rate for the class of all variable-length codes remained elusive.

Building on a partial characterization established in [1], in Section III we show that the minimum asymptotic variable-length coding rate for the side-information problem is the *complementary graph entropy* of the associated graph. This quantity was defined by Körner and Longo, [19], in their investigation of a two-step source coding problem. Since no formula is currently known for the complementary graph entropy, our results do not yield a single-letter characterization either. But they further strengthen the close connections between the zero-error side-information and zero-error capacity [24] problems, as we discuss next.

Associated with the zero-error versions of both the channel coding and the side-information problems are graphs defined on the corresponding input alphabets. The duality of the independence number (size of the largest edge-free induced subgraph) and the chromatic number (minimum cardinality of a partition of the graph into edge-free induced subgraphs) of a graph is well known (see [3]). Witsenhausen showed in [27] that the minimum fixed-length zero-error side-information rate (briefly, the zero-error rate) is the limit of the normalized chromatic numbers of normal powers of the underlying graph. In [24], Shannon defined the corresponding limit of normalized independence numbers as the zero-error capacity of the graph.

Suppose now that the source statistics are known, and variable-length codes tailored to the source distribution are used by the encoder. Our results show that the complementary graph entropy characterizes the minimum rate of transmission, and is thus the analog, for this problem, of the zero-error rate of Witsenhausen. Similarly, Csiszár and Körner, in [9], defined an analog of the zero-error capacity where the channel codewords are constrained to be picked according to a fixed distribution. Roughly speaking, the complementary graph entropy and the zero-error capacity within a distribution are, respectively, the limits of the normalized chromatic numbers and independence

Manuscript received July 9, 2001; revised February 18, 2002. This work was supported in part by the National Science Foundation under Grants EIA-9986057 and EIA-0080134, the University of California MICRO Program, Conexant Systems, Inc., Dolby Laboratories, Inc., Lucent Technologies, Inc., Mindspeed Technologies, and Qualcomm, Inc. The material in this paper was presented in part at the Canadian Workshop on Information Theory, Vancouver, BC, Canada, June 2001, and the IEEE International Symposium on Information Theory, Washington, DC, June 2001.

P. Koulgi, E. Tuncel, and K. Rose are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560 USA (e-mail: prashant@ece.ucsb.edu; ertem@ece.ucsb.edu; rose@ece.ucsb.edu).

S. L. Regunathan is with Microsoft Corporation, Redmond, WA 98052 USA (e-mail: shanre@microsoft.com).

Communicated by R. Zamir, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2002.806154

numbers of high-probability induced subgraphs in the normal powers of the underlying graphs. Marton showed, in [22], that the complementary graph entropy and the zero-error capacity within a distribution of an arbitrary graph sum to the entropy of the underlying probability distribution. Thus, determination of either quantity directly yields the other one. Further, better understanding of either one of the two quantities—zero-error side-information rate or zero-error capacity—may lead to new insights into the other.

While no formula is currently known for the complementary graph entropy (such a formula would also immediately yield a formula for the zero-error capacity), upper and lower bounds have been studied in [19] and [22]. Further, it was shown in [10] that these bounds are tight for the widely studied class of perfect graphs.

Complementary to the asymptotic analysis considered above is the design of optimal codes for finite block lengths. For previous work on constructive code design, see [29], [17], [23], and [28]. Our approach is different from these, in that we emphasize design complexity as well as code performance. In Section IV, we show that optimal code design is $NP$-hard for both the classes of codes introduced in [1]. Thus, if the widely held conjecture that $P \neq NP$ is true, no polynomial-time optimal code design algorithm exists.

In some applications, optimal code design may be desirable even at the cost of high design complexity. Examples include applications where the size of the source alphabet is small, and design is off-line. In Section V, we develop an optimal code design algorithm based on recursively building up the optimal code for the entire graph from optimal codes of its subgraphs. Analyzing this algorithm, we show that it has exponential (in the size of the source alphabet) worst case complexity, and derive the value of the exponent.

Polynomial-time suboptimal algorithms may be of interest when large graphs are encountered. We explore two different strategies toward the development of such algorithms in Section VI. Our first strategy is based on approximate graph coloring. Of the two such algorithms we propose, the first guarantees good worst case performance, while the second is extremely simple, and promises good performance for most inputs. As a second strategy, we develop a class of algorithms to design codes by approximating a lower bound on the optimal scalar coding rate established in [1]. Code performance can then be traded off against design complexity by suitably choosing an algorithm from this class.

We formulate the problem and establish notation in Section II. All our results are then summarized in Section III. In the concluding Section VIII, we remark on a few open problems.

## II. Preliminaries

Let $\{(X_i, Y_i)\}_{i=1}^{\infty}$ be a sequence of independent drawings of a pair of dependent random variables $X, Y$. Here $(X, Y)$ take values in the finite product set $V \times W$ according to the probability distribution $P(x, y)$. It is desired to encode the sequence $\{X_i\}$ such that the decoder can decode it *without error*. The special assumption made here is that the decoder has access to the side information $\{Y_i\}$. We will call this the *side-information problem*.

Distinct $x, x' \in V$ are *confusable* if there is a $y \in W$ such that $P(x, y) > 0$ and $P(x', y) > 0$. Two confusable letters may not be assigned the same codeword in any valid code. Thus, confusability defines a binary-symmetric relation on the letters of $V$. Witsenhausen [27] captured this confusability relation of the source pair $(X, Y)$ in the *characteristic graph* $G$. $G = (V, E)$ is defined on the vertex set $V$, and distinct $x, x' \in V$ are connected by an edge if they are confusable. The pair $(G, P)$ denotes the probabilistic graph consisting of $G = (V, E)$ together with the distribution $P$ over its vertices. (Here we denote also by $P$ the marginal distribution on $V$.)

Variable-length codes for the side-information problem were introduced by Alon and Orlitsky in [1]. They defined two families of binary variable-length codes.

1) A *restricted inputs* (RI) code for $(G, P)$ is a mapping $\phi: V \rightarrow \{0, 1\}^*$ such that if $\{x, x'\} \in E$ then $\phi(x)$ is not a prefix of $\phi(x')$.
2) An *unrestricted inputs* (UI) code for $(G, P)$ is a mapping $\phi: V \rightarrow \{0, 1\}^*$ such that, for every distinct pair $x, x' \in V$, $\phi(x)$ is not a proper prefix of $\phi(x')$, and, if $\{x, x'\} \in E$, then $\phi(x) \neq \phi(x')$.

(These definitions generalize in the obvious way to the case of $k$-ary codes, $k \geq 2$.)

UI codes, which form a subclass of the class of RI codes, may be preferred to the latter in some applications. Since the codeword set is prefix free, UI codes protect against loss of synchronization if the side information at the decoder is occasionally wrong. On the other hand, the use of an RI code in such applications may lead to catastrophic errors. The motivation behind these two classes of codes is discussed in more detail in [1] (also see [12] for a communication complexity viewpoint).

The rate of a code $\phi$ is the expected number of bits transmitted

$$\bar{l}(\phi) = \sum_{x \in V} P(x)|\phi(x)|.$$

We denote by $\overline{L}(G, P)$ ($\overline{\mathcal{L}}(G, P)$) the minimum rate of an RI (UI) code for $(G, P)$

$$\overline{L}(G, P) = \min\left\{\bar{l}(\phi): \phi \text{ is an RI code for } (G, P)\right\}$$
$$\overline{\mathcal{L}}(G, P) = \min\left\{\bar{l}(\phi): \phi \text{ is a UI code for } (G, P)\right\}$$

and call the codes attaining these minima the optimal codes. (In general, we write $\overline{L}_k(G, P)$ and $\overline{\mathcal{L}}_k(G, P)$ for the corresponding minimum rates for $k$-ary codes ($k \geq 2$). Thus, $\overline{L}(G, P) = \overline{L}_2(G, P)$ and $\overline{\mathcal{L}}(G, P) = \overline{\mathcal{L}}_2(G, P)$.)

We have that

$$\overline{L}(G, P) \leq \overline{\mathcal{L}}(G, P).$$

To define variable-length block codes, we extend the notion of confusability to vectors. Thus, distinct

$$x^n = (x_1, x_2, \ldots, x_n) \in V^n$$
$$x'^n = (x'_1, x'_2, \ldots, x'_n) \in V^n$$

are confusable iff every distinct pair $(x_i, x'_i)$, $i = 1, 2, \ldots, n$ is confusable. The characteristic graph for $(X^n, Y^n)$ is then the so-called $n$-fold *normal* power of $G$, denoted $G^n$. $G^n =$

$(V^n, E_n)$, with, for distinct $x^n, x'^n \in V^n$, $\{x^n, x'^n\} \in E_n$ iff $\{x_i, x'_i\} \in E$ for all distinct pairs $(x_i, x'_i)$, $i = 1, 2, \ldots, n$. Note that the normal power is also referred to as the AND power in the literature. We denote by $P^n$ the product distribution induced on $V^n$ by $P$

$$P^n((x_1, x_2, \ldots, x_n)) = \prod_{i=1}^{n} P(x_i).$$

The previous definitions of RI and UI codes for $(G, P)$ may now be extended to RI and UI block codes for $(G^n, P^n)$.

We shall briefly summarize some standard notations and concepts from graph theory, which we will use extensively in the sequel (see, for example, [3]). We assume that all graphs are undirected and have no loops or multiple edges. For our purposes, these assumptions do not entail any loss of generality. Two distinct nodes are connected in $\overline{G}$—the complement of $G = (V, E)$—if they are not connected in $G$. The subgraph $G' = (V', E')$ induced in $G$ by a subset $V' \subseteq V$ is called an *induced subgraph*. A subset of the vertex set $V$ is an independent set of $G$ if it induces an edge-free subgraph in $G$. Let $\alpha(G)$—the *independence number* of $G$—be the maximum size of an independent set of $G$, and let $\chi(G)$—the *chromatic number* of $G$—be the minimum cardinality of a coloring of $G$, i.e., a partition of $V$ into independent sets. It is clear that $\alpha(G) \leq \chi(\overline{G})$ and $\alpha(G)\chi(G) \geq |V|$. $G$ is a *perfect* graph if $\chi(G') = \alpha(\overline{G'})$ for every induced subgraph of $G$. For the extensive literature on perfect graphs, see [3], [20], and the references therein.

Finally, note that all logarithms are to base two, unless mentioned otherwise.

## III. Summary of Results

### *A. Characterization of Minimum Asymptotic Rate*

Let $\overline{L}(G^n, P^n)$ denote the minimum rate of an RI code for $(G^n, P^n)$. The minimum asymptotic rate per source letter required for the side-information problem is

$$R^*(G, P) = \lim_{n\to\infty} \frac{\overline{L}(G^n, P^n)}{n}. \tag{1}$$

(Note that, by subadditivity, the limit exists.)

The characterization of $R^*(G, P)$ was first considered by Alon and Orlitsky in [1]. They defined the *chromatic entropy* of a probabilistic graph, $H_\chi(G, P)$, as the minimum entropy of its colorings. They then showed that

$$R^*(G, P) = \lim_{n\to\infty} \frac{H_\chi(G^n, P^n)}{n}$$

but a single-letter characterization of $R^*(G, P)$ remained elusive.

While we are also unable to derive such a characterization, in Section IV we build on the results of Alon and Orlitsky in [1] to equate the minimum asymptotic rate to the *complementary graph entropy*, $\overline{H}(G, P)$, of the characteristic graph $(G, P)$. In particular, we prove that

$$\lim_{n\to\infty} \frac{H_\chi(G^n, P^n)}{n} = \overline{H}(G, P). \tag{2}$$

Motivated by a two-step source coding problem, Körner and Longo considered in [19] two information-theoretic functionals on probabilistic graphs: the *graph entropy* $H(G, P)$ (previously defined in [18]) and the *complementary graph entropy* $\overline{H}(G, P)$ (this is also referred to as the co-entropy or the $\pi$-entropy in the literature). They showed that these quantities characterize the minimum asymptotic rates for the coding problems they considered. While Körner derived a formula for $H(G, P)$ in [18], no formula is currently known for $\overline{H}(G, P)$. Marton, in her investigation of the zero-error capacity of a probabilistic graph [22], revealed the close connection between the complementary graph entropy and the zero-error capacity [24]. Thus, a formula for the complementary graph entropy of an arbitrary probabilistic graph would imply, via her results, a formula for the zero-error capacity of the corresponding graph. This, in turn, would resolve a major unsolved problem of information theory and graph theory.

Upper and lower bounds for $\overline{H}(G, P)$ have been studied by Csiszár, Körner, Marton, and others. In [19], Körner and Longo established bounds for $\overline{H}(G, P)$ in terms of $H(G, P)$ and $H(\overline{G}, P)$

$$H(P) - H\left(\overline{G}, P\right) \leq \overline{H}(G, P) \leq H(G, P), \tag{3}$$

(where $H(P)$ is the Shannon entropy of $P$). We show that the lower bound above may be derived by recognizing its equivalence to a side-information problem. Csiszár *et al.*, [10], showed that both the bounds in (3) are tight for all distributions $P$ if the graph $G$ is perfect. Thus, perfect graphs satisfy an optimality condition for zero-error side-information coding, as we show. We also provide an example where neither bound in (3) is tight.

Other bounds for $\overline{H}(G, P)$ include those of Marton in [22], in terms of a generalization of the Lovász $\theta$-functional [21] to probabilistic graphs. These bounds are also tight if the underlying graph is perfect.

### *B. Design and Analysis of Codes*

In [1], it was shown that an RI code for $(G, P)$ may be interpreted as a coloring of $G$, followed by one-to-one encoding of the colors. Similarly, a UI code is a coloring of $G$ followed by prefix-free coding of the colors. But in neither of these cases does the optimal code necessarily induce a coloring of $G$ with the minimum number of colors. Consider the 3-colorable graph in Fig. 1. The optimal binary RI code (which, in this case, is the same as the optimal UI code), also shown in Fig. 1, induces a coloring with four colors.

Consider $k$-ary coding of a given probabilistic graph $(G, P)$, where $k \geq 2$ is fixed. Let $|V|$ be the cardinality of the vertex set of $G$.

An efficient algorithm for the design of optimal prefix-free codes ("Huffman codes") was discovered by Huffman in [16]. We consider the corresponding design problems for RI and UI codes. Note that RI and UI codes may be viewed as generalizations of prefix-free codes: for a complete graph (where every node is connected to every other node), the classes of RI and UI codes collapse to that of prefix-free codes, so that the Huffman algorithm may be used to design the optimal code. By contrast, for arbitrary $(G, P)$, we show that design of the optimal RI code is $NP$-hard (the various complexity classes considered in this paper are discussed in detail in [13]). Similarly, optimal UI code design is also $NP$-hard.

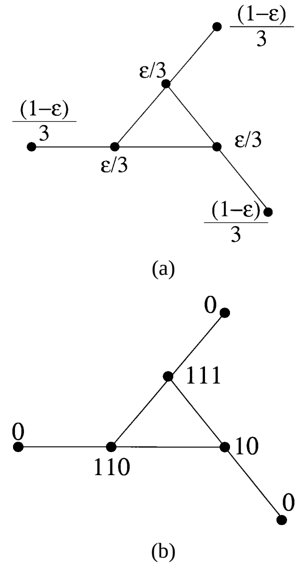


Fig. 1. The node labels in (a) indicate probabilities with $\epsilon < 1/4$. In (b) they indicate optimal codewords.

Consider the following coding problem $A$:

$$\text{Is } \overline{L}_k(G, P) \leq L?$$

$A$ is in $NP$ since, given a code for $(G, P)$, it is easy to check if it is a valid RI code, and if the rate of this code is $\leq L$. Exploiting the connections between RI coding and graph coloring, we reduce, in polynomial time, the well-known $NP$-complete problem of graph 3-colorability [13] to $A$. Thus, problem $A$ is, in fact, $NP$-complete and, correspondingly, optimal RI code design is $NP$-hard.

For the design of binary UI codes, we are able to prove a stronger hardness-of-approximability result: $\overline{\mathcal{L}}(G, P)$ cannot be approximated to within $1/3$ bits unless $P = NP$. More precisely, we show that the following problem is $NP$-hard:

Find a binary UI code of rate $\overline{\mathcal{L}}$ with

$$\overline{\mathcal{L}} - \overline{\mathcal{L}}(G, P) < 1/3 - \epsilon \text{ bits, for fixed } \epsilon > 0.$$

A similar lower bound on polynomial-time approximability can also be proved for $k$-ary UI coding where $k > 2$.

In practice, optimal codes may be desirable if the underlying graph has a small vertex set, design is off-line, and complexity is not a significant constraint. On the other hand, fast approximate (i.e., suboptimal) algorithms are needed for the design of codes for large graphs. This motivates the consideration of both optimal and approximate coding algorithms.

We develop optimal RI and UI coding algorithms via structural analysis of the respective optimal codes. These algorithms are based on efficient recursive search for optimal codes of induced subgraphs of $(G, P)$. Further, we show that these algorithms can be implemented to design the optimal $k$-ary RI/UI code in worst case time $O((k+1)^{|V|})$.

Turning then to the development of fast design algorithms, we pursue two distinct approaches. In the first approach, we consider approximately coloring the given graph, then Huffman coding the colors. The example in Fig. 1 shows that this separation is not justified in the case of optimal code design. But, by drawing on the rich literature on approximate graph coloring, we demonstrate that good suboptimal codes can indeed be designed via this approach.

Let $P_{\max}$ and $P_{\min}$ be the maximum and minimum probabilities under $P$, and define

$$d\left(\overline{G}\right) = 1 + \text{ maximum degree of } \overline{G}.$$

If $G$ can be colored with $c(G)$ colors, we show that this coloring can be translated into a UI code for $(G, P)$ of rate $\overline{\mathcal{L}}$ such that

$$\overline{\mathcal{L}} - \overline{\mathcal{L}}(G, P) \leq 2 + \log\left\{c(G) d\left(\overline{G}\right) P_{\max}\right\} \tag{4}$$

and

$$\overline{\mathcal{L}} - \overline{L}(G, P) \leq 3.4427 + \log\left\{c(G) d\left(\overline{G}\right) P_{\max}\right\} + \log\{1 - \log P_{\min}\}. \tag{5}$$

To minimize these bounds, we suggest use of the approximate graph-coloring algorithm of Halldórsson [15] which has the best currently known worst case performance guarantee. (It may be noted that the algorithm of [15] colors an arbitrary graph $G$ on $|V|$ nodes with fewer than $O(\frac{\chi(G)|V|(\log\log|V|)^2}{(\log|V|)^3})$ colors.)

In practice, worst case performance guarantees may be pessimistic, since worst cases occur infrequently. Motivated by this observation, we use a standard random graph model to analyze a simple coding algorithm, which is based on greedy coloring, and show that it produces, on average, a UI code of rate $\overline{\mathcal{L}}$ with

$$\overline{\mathcal{L}} - \overline{\mathcal{L}}(G, P) \leq 2 + \log \frac{P_{\max}}{P_{\min}} \tag{6}$$

and

$$\overline{\mathcal{L}} - \overline{L}(G, P) \leq 3.4427 + \log \frac{P_{\max}}{P_{\min}} + \log\{1 - \log P_{\min}\}. \tag{7}$$

In the second approach to the design of suboptimal codes, we consider the following inequality established in [1]:

$$H(P) - H\left(\overline{G}, P\right) \leq \overline{L}(G, P) \leq \overline{\mathcal{L}}(G, P)$$

where $H(\overline{G}, P)$ is the graph entropy of $\overline{G}$. We show that the lower bound $H(P) - H(\overline{G}, P)$ may be interpreted as a rate-distortion function [4], and thus may be calculated using the Blahut–Arimoto (BA) algorithm [5]. Since exact calculation of this quantity may be computationally intensive, we propose a class of approximating algorithms of increasing complexity. We then show that suboptimal RI codes may be designed as by-products of these algorithms.

## IV. Minimum Asymptotic Rate and the Complementary Graph Entropy

The chromatic entropy of a probabilistic graph $(G, P)$ (where $G = (V, E)$), $H_\chi(G, P)$, was defined in [1]. If $c$ is a function defined over $V$, then $c(X)$ is a random variable with entropy

$$H(c(X)) = \sum_{\gamma \in c(V)} P[c^{-1}(\gamma)] \log \frac{1}{P[c^{-1}(\gamma)]}$$

where $c^{-1}$ is the inverse of $c$, and for $U \subseteq V$

$$P(U) = \sum_{x \in U} P(x).$$

*Definition 1:* The chromatic entropy of $(G, P)$ is the lowest entropy of any coloring of $G$

$$H_\chi(G, P) = \min\{H(c(X)) \colon c \text{ is a coloring of } G\}.$$

Let $R_n$ be the minimum rate of a *uniquely decodable* (i.e., not necessarily instantaneous) code for $(G^n, P^n)$. The following lemma bounds $R_n$ in terms of $H_\chi(G^n, P^n)$.

*Lemma 1:*

$$H_\chi(G^n, P^n) - \log\{H_\chi(G^n, P^n) + 1\} - \log e \le R_n$$
$$R_n \le H_\chi(G^n, P^n) + 1. \qquad (8)$$

*Proof:* Let $\phi \colon V^n \to \{0, 1\}^*$ be a code for $(G, P)$. If distinct $x^n, x'^n \in V^n$ are confusable and, further, if $\phi(x^n) = \phi(x'^n)$, then the decoder cannot distinguish between $x^n$ and $x'^n$, and $\phi$ is not uniquely decodable. In other words, if $\phi$ is uniquely decodable, $\phi(x^n) = \phi(x'^n)$ for distinct $x^n, x'^n$ implies that $x^n$ and $x'^n$ are not connected in $G^n$. Thus, $\phi$ may be written as the composition of a coloring of $G^n$ and a one-to-one encoding of the colors. Equation (8) now follows from the upper and lower bounds established in [2] on the rates of one-to-one codes. □

Identical bounds as in (8) were proved in [1] for the restricted class of instantaneous codes, and were then used to calculate the minimum asymptotic rate of such codes. We can, therefore, parallel these calculations, to determine the minimum asymptotic rate for uniquely decodable codes.

*Lemma 2:*

$$R^*(G, P) = \lim_{n\to\infty} \frac{R_n}{n} = \lim_{n\to\infty} \frac{H_\chi(G^n, P^n)}{n}. \qquad (9)$$

*Proof:* The proof is identical to that of [1, Lemma 6]. □

Since the same asymptotic rate as in (9) is achievable with instantaneous codes, Lemma 2 shows that the possibly larger class of uniquely decodable codes offers no *asymptotic* advantage. While this situation is identical to that obtained in regular lossless source coding, we are unable to answer whether uniquely decodable codes also offer no advantage in the case of finite block lengths.

We will now derive an alternate characterization of the limit in (9). To do so, we consider the complementary graph entropy, which is an information-theoretic functional on probabilistic graphs defined in [19].

*Definition 2:* The complementary graph entropy of $(G, P)$ is the normalized logarithm of the "essential chromatic number of $G^n$ with respect to $P$," i.e., the number

$$\overline{H}(G, P) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log\left[\min_{P^n(A)\ge 1-\epsilon} \chi(G^n(A))\right] \qquad (10)$$

where $G^n(A)$ is the subgraph induced in $G^n$ by $A \subseteq V^n$.

Thus, $G^n$ has a high-probability induced subgraph which can be colored with approximately $2^{n\overline{H}(G, P)}$ colors. Körner and Longo used this fact in [19] to show that the complementary graph entropy is the rate required for the following two-step source coding problem: Consider a memoryless source emitting symbols from a finite alphabet $V$ according to a distribution $P$. Assume that some pairs of elements of the alphabet are distinguishable, while some others are not, and let $G$ be the graph on $V$ where connectedness means distinguishability (note that this graph is different from the characteristic graph defined earlier). We want to encode the $n$-length source vector $X^n$ in two steps. In the first step, an encoding function $f$ on $V^n$ is used, and it is required that, on the basis of $f(X^n)$, the decoder be able to determine a sequence $\hat{x}^n$ that is, with high probability, indistinguishable from $X^n$ in every coordinate. Call an encoder $f$ achieving this goal "$G$-faithful." In the second step, we want to encode $X^n$ by an encoding function $g$ such that the following holds: the encoded source $g(X^n)$, together with an *arbitrary* $G$-faithful encoding of $X^n$, determines $X^n$ with high probability. It was shown in [19] that the minimum asymptotic rate needed for such a "complementary encoding" in the second step is $\overline{H}(G, P)$.

We will also need the following generalization of the zero-error capacity [24] to probabilistic graphs. This quantity was introduced by Csiszár and Körner in [9] to study the capacity of an arbitrarily varying channel with maximum probability of error.

*Definition 3:* Let $\mathcal{T}^n(P, \epsilon)$ be the set of "$(P, \epsilon)$-typical" sequences in $V^n$, i.e., the set of sequences $x^n \in V^n$ for which the frequency $\pi(i|x^n)$ of each element $i \in V$ satisfies

$$|\pi(i|x^n) - P(i)| \le \epsilon.$$

Let $G^n(P, \epsilon)$ be the subgraph of $G^n$ induced by $\mathcal{T}^n(P, \epsilon)$.

*Definition 4:* The capacity of the graph $G$ relative to $P$ is

$$C(G, P) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \alpha(G^n(P, \epsilon)). \qquad (11)$$

We will need the following relation between $\overline{H}(G, P)$ and $C(G, P)$ established by Marton in [22]:

$$\overline{H}(G, P) + C(G, P) = H(P). \qquad (12)$$

Consider a fixed-length encoding function

$$f \colon V^n \to \{1, 2, \ldots, 2^{nR'}\}$$

for $G^n = (V^n, E_n)$ of which we require the following property: if $\{x^n, x'^n\} \in E_n$ then, with high probability, $f(x^n) \ne f(x'^n)$. It follows from (10) that the minimum rate required is $\overline{H}(G, P)$. In the following theorem, we show that $\overline{H}(G, P)$ is also the minimum rate required if $f$ is allowed to be a variable-length encoding function, but $\{x^n, x'^n\} \in E_n \Rightarrow f(x^n) \ne f(x'^n)$, always.

*Theorem 1:*

$$R^*(G, P) = \overline{H}(G, P) \qquad (13)$$

where $R^*(G, P)$ is defined as in (9).

*Proof:* We will show that

$$\lim_{n\to\infty} \frac{H_\chi(G^n, P^n)}{n} = \overline{H}(G, P).$$

We claim that $\overline{H}(G, P)$ is not smaller than the limit on the left-hand side. Fix $\epsilon > 0$. Let

$$\overline{H}_\epsilon(G, P) = \limsup_{n\to\infty} \frac{1}{n} \log\left[\min_{A\subseteq V^n,\, P^n(A)\ge 1-\epsilon} \chi(G^n(A))\right].$$

Then, for any fixed $\delta > 0$, for each $n > n_0(\delta)$ there is a subset $A \subseteq V^n$ with $P^n(A) \geq 1 - \epsilon$, and a coloring $c$ of $G^n$ satisfying

$$|c(G^n(A))| \leq 2^{n\{\overline{H}_\epsilon(G, P)+\delta\}}. \tag{14}$$

For $x^n \in V^n$, define the function $\Phi: V^n \to \{0, 1\}$ as

$$\Phi(x^n) = \begin{cases} 1, & \text{if } x^n \in A \\ 0, & \text{else.} \end{cases}$$

Thus, $\Phi$ is the indicator function of $A$. Estimating the entropy of the coloring $c$

$$\begin{aligned} H(c(X^n)) &\leq H(\Phi) + H(c(X^n)|\Phi) \\ &\leq H(\Phi) + H(c(X^n)|X^n \in A) + \epsilon H(c(X^n)|X^n \notin A) \\ &\leq 1 + n\left\{\overline{H}_\epsilon(G, P) + \delta + \epsilon \log |V|\right\} \end{aligned}$$

where we used (14) in the last step. But, by the definition of the chromatic entropy

$$H_\chi(G^n, P^n) \leq H(c(X^n)).$$

Normalizing by $n$ and taking limits, the claim follows.

Now consider the reversed inequality. We lower-bound $H_\chi(G^n, P^n)$ in terms of the maximum size of an independent set induced by $\mathcal{T}^n(P, \epsilon)$ in $G^n$. But this size is related to the capacity $C(G, P)$, and the inequality will then follow from (12). Let us fill in the details. Fix $\epsilon > 0$. Let the coloring function $c$ on $G^n$ achieve $H_\chi(G^n, P^n)$, so that

$$H_\chi(G^n, P^n) = H(c(X^n)).$$

To lower-bound $H(c(X^n))$, we use the following elementary lower bound for the entropy function: if $Q$ is a probability distribution over the set $\mathcal{Q}$, and $S \subseteq \mathcal{Q}$, then

$$H(Q) \geq -\left\{\sum_{j \in S} Q(j)\right\} \log \max_{j \in S} Q(j).$$

Thus, we have the following estimate for $H_\chi(G^n, P^n)$:

$$H(c(X^n)) \geq -P^n(\mathcal{T}^n(P, \epsilon)) \log \max_{x^n \in \mathcal{T}^n(P, \epsilon)} P^n(c(x^n)). \tag{15}$$

But the set of $(P, \epsilon)$-typical sequences $\mathcal{T}^n(P, \epsilon)$ captures most of the probability [11, p. 34]

$$P^n(\mathcal{T}^n(P, \epsilon)) \geq 1 - \frac{|V|}{4n\epsilon^2}. \tag{16}$$

Further, in any coloring of $G^n$, the maximum cardinality of a single-colored subset of $\mathcal{T}^n(P, \epsilon)$ cannot exceed $\alpha(G^n(P, \epsilon))$, the size of the largest independent set induced by $\mathcal{T}^n(P, \epsilon)$ in $G^n$. Thus,

$$\begin{aligned} \max_{x^n \in \mathcal{T}^n(P, \epsilon)} P^n(c(x^n)) &\leq \alpha(G^n(P, \epsilon)) \max_{x^n \in \mathcal{T}^n(P, \epsilon)} P^n(x^n) \\ &\leq \alpha(G^n(P, \epsilon)) 2^{-n \min\{H(P')+D(P'\|P)\}} \\ &\leq \alpha(G^n(P, \epsilon)) 2^{-n\{H(P)+\epsilon|V| \log \epsilon\}} \end{aligned} \tag{17}$$

where the minimization is over the set of probability distributions $\{P': |P'(i) - P(i)| < \epsilon \,\forall i \in V\}$. We use a well-known formula for the probability of a typical sequence [11, p. 32] in the second inequality, and the uniform continuity of entropy [11, p. 33] in the third.

Substituting (16) and (17) in (15)

$$\frac{H_\chi(G^n, P^n)}{n} \geq \left(1 - \frac{|V|}{4n\epsilon^2}\right) \cdot \left\{H(P) - \frac{1}{n} \log \alpha(G^n(P, \epsilon)) + \epsilon|V| \log \epsilon\right\}.$$

Taking limits

$$\lim_{n \to \infty} \frac{H_\chi(G^n, P^n)}{n} \geq H(P) - C_\epsilon(G, P) + \epsilon|V| \log \epsilon$$

where $C_\epsilon(G, P)$ is defined as (cf. (11))

$$C_\epsilon(G, P) = \limsup_{n \to \infty} \frac{1}{n} \log \alpha(G^n(P, \epsilon)).$$

Since this is true for every $\epsilon$, the result follows by letting $\epsilon \to 0$ and using (12). □

While no formula is currently known for $\overline{H}(G, P)$, upper and lower bounds were studied in [19], [10], and [22]. We shall consider here the bounds in terms of the graph entropy in some detail.

*Definition 5:* The graph entropy of $(G, P)$ is the number

$$H(G, P) = \min\{I(X; S): X \sim P, X \in S \in \mathcal{S}(G)\}. \tag{18}$$

Here $\mathcal{S}(G)$ is the collection of independent sets of $G$. The minimum is taken over all random pairs $(X, S)$ such that $X$ has distribution $P$, $S$ takes values in $\mathcal{S}(G)$, and the random vertex $X$ belongs to the random set $S$ with probability 1. In [18], Körner provided an alternate definition of $H(G, P)$ in terms of normalized chromatic numbers of co-normal powers of $G$, analogous to the definition of $\overline{H}(G, P)$ in (10). We will not need this interpretation of $H(G, P)$ here.

Körner and Longo showed in [19] that

$$H(G, P) \geq \overline{H}(G, P) \geq H(P) - H\left(\overline{G}, P\right). \tag{19}$$

We shall give a different proof of the lower bound to bring out its simple side-information coding interpretation.

*Lemma 3:*

$$\overline{H}(G, P) \geq H(P) - H\left(\overline{G}, P\right). \tag{20}$$

*Proof:* Let $X \sim P$, and let $S$ be jointly distributed with $X$ such that $X \in S \in \mathcal{S}(\overline{G})$. Then it is clear that the characteristic graph of $(X, S)$ for any such $S$ is either $G$ or a subgraph of $G$. Consider now independent drawings of the random pair $(X, S)$. Suppose first that the decoder knows $S^n$, while the encoder knows $X^n$, and the conditional distribution of $S$ given $X$. If an eventually vanishing nonzero probability of error is permitted, Slepian and Wolf showed in [26] that the encoder must transmit at a minimum rate of $H(X|S)$ to convey $X^n$ to the decoder. Suppose next that the encoder knows only that the conditional distribution of $S$ given $X$ satisfies $X \in S \in \mathcal{S}(\overline{G})$. Then, maximizing over all possible choices of $S$, the encoder needs to transmit at a rate of not less than

$$\max\left\{H(X|S): X \sim P, X \in S \in \mathcal{S}\left(\overline{G}\right)\right\}$$

if a vanishingly small probability of error is permitted. But, from (18), this is the same as $H(P) - H(\overline{G}, P)$. Now, if zero error is required, by Theorem 1, the encoder needs to transmit at a rate of no more than $\overline{H}(G, P)$ and the lemma follows. □

In a remarkable paper [10], Csiszár *et al*. proved that equality holds in both the inequalities of (19) for all distributions $P$ if $G$ is perfect. Thus, perfectness of $G$ is sufficient to guarantee equality in (20). The necessary conditions for equality are unknown at present.

We now provide an example of a probabilistic graph $(G, P)$ for which neither bound of (19) is tight.

*Example 1:* Let $G = C_5$ be the 5-cycle, and $P = P_u$ be the uniform distribution on its nodes. Since the maximum size of a set in $\mathcal{S}(\overline{C_5})$ is 2, $H(P_u) - H(\overline{C_5}, P_u) \leq \log 2$. Choosing the distribution $q(s|x) = 1/2$ for each of the two edges $s$ that a node $x$ belongs to, we get $H(P_u) - H(\overline{C_5}, P_u) = \log 2$. Since the maximum size of a set in $\mathcal{S}(G)$ is also 2, $H(C_5, P_u) \geq \log 5 - \log 2$. Equality is achieved by setting $q(s|x) = 1/2$ for each of the two maximal independent sets $x$ belongs to. Now, let

$$C(G) = \lim_{n\to\infty} \frac{1}{n} \log \alpha(G^n)$$

be the zero-error capacity of a graph $G$ [24]. It is clear that $C(G, P_u) \leq C(G)$. Lovász showed in [21] that $C(C_5) = \frac{1}{2} \log 5$. But Shannon's construction [24], which achieves the capacity $C(C_5)$, uses each vertex the same number of times, so that $C(C_5, P_u) = C(C_5)$. Hence, $\overline{H}(C_5, P_u) = H(P_u) - C(C_5, P_u) = \frac{1}{2} \log 5$.

Putting all these numbers together,

$$\begin{aligned} \log 2 &= H(P_u) - H\left(\overline{C_5}, P_u\right) \\ &< \frac{1}{2} \log 5 = \overline{H}(C_5, P_u) \\ &< \log \frac{5}{2} = H(C_5, P_u). \end{aligned}$$ □

Bounds for $\overline{H}(G, P)$ have also been studied by Marton in [22]. She derived upper and lower bounds in terms of a generalization of the Lovász $\theta$-functional [21] to probabilistic graphs, and showed that these bounds also coincide for all distributions $P$ if $G$ is perfect.

## V. Hardness of Optimal Code Design

The complexity classes $P$ and $NP$ are well known (for an in-depth discussion see [13]), so we will confine ourselves to their rough, working descriptions. A problem belongs to the class $P$ if it can be solved in time bounded by a polynomial in the size of the input. It is in $NP$ if a guessed solution can be verified in polynomial time. Clearly, $P \subset NP$. But proof of the long-standing conjecture that the inclusion is strict has not been forthcoming; in fact, this problem remains a major challenge in computer science.

$NP$-complete problems are the "hardest" problems in $NP$ in the following sense: if any single $NP$-complete problem can be solved in polynomial time, then *all* problems in $NP$ can so be solved. A problem, not necessarily in $NP$, is $NP$-hard if some $NP$-complete problem can be transformed to it in polynomial time. An $NP$-hard problem is "at least as hard" as the $NP$-complete problems, since it cannot be solved in polynomial time unless $P = NP$.

Fix $k \geq 2$. We consider the complexity of the following coding problems.

$(A)$ INSTANCE: Graph $G = (V, E)$, distribution $P$ on $V$, and a positive real number $L$.
QUESTION: Is there a $k$-ary RI code for $(G, P)$ of rate $\leq L$?

$(B)$ INSTANCE: Graph $G = (V, E)$, distribution $P$ on $V$, and a positive real number $L$.
QUESTION: Is there a $k$-ary UI code for $(G, P)$ of rate $\leq L$?

We will show that both problems $A$ and $B$ are $NP$-complete. It follows, as a simple corollary, that the design of optimal $k$-ary RI and UI codes is $NP$-hard.

For the case of $k > 2$, it turns out that a simple observation proves both $A$ and $B$ to be $NP$-complete. After disposing this case in Theorem 2, we treat the binary coding cases separately in Theorems 3 and 4. The proof for the binary UI coding case yields a slightly stronger result, in that it brings out the $NP$-hardness of finding even "good" suboptimal codes. This statement is made precise in Corollary 1. Throughout, we will consider polynomial-time reductions from and to $A$ and $B$ of the well-known $NP$-complete problems of graph $N$-colorability, $(GNC)$, for $N \geq 3$ [13]:

$(GNC)$ INSTANCE: Graph $G = (V, E)$.
QUESTION: Is $G$ colorable with $N$ colors?

We will assume throughout this section, without loss of generality, that the graphs under consideration do not have isolated vertices.

*Theorem 2:* Problems $A$ and $B$ are $NP$-complete for fixed (i.e., prespecified) $k > 2$.

*Proof:* Note that, in a $k$-ary (UI or RI) code of rate $\log_k k = 1$, each codeword is composed of a single letter from $\{0, 1, \ldots, k-1\}$, and no codeword composed of more than one letter exists. This observation leads to a simple proof. Thus, let $G = (V, E)$ be given as an instance of $GNC$. Consider $(G, P)$, where $P$ is an arbitrary distribution on $V$. Further, set $k = N$, and $L = \log N$. If $G$ is $k$-colorable, then successive color classes may be assigned codewords from among $\{0, 1, \ldots, k-1\}$ to obtain an RI code of rate $\log k = \log N$ bits. Conversely, if an RI code for $(G, P)$ of rate $\log k = \log N$ bits exists, then distinct codewords can be identified as colors to obtain a coloring of $G$ with $k = N$ or fewer colors. But graph $N$-colorability is $NP$-complete for every fixed $N > 2$.

The argument is identical in the case of UI coding. □

Consider now the case $k = 2$. Note that the problem $GNC$ is equivalent to asking if binary codewords can be assigned to the vertices of $G$, such that connected vertices get different codewords, with maximum length less than $\log N$. On the other hand, problems $A$ and $B$ consider expected lengths of the assigned codewords. These quantities coincide if all the lengths are 1, but may not coincide otherwise. Therefore, since graph 2-colorability can be easily checked in polynomial time, the previous proof cannot be extended to the case $k = 2$. Instead, in the following theorems, we will construct from the instance $G$ of $GNC$ auxiliary probabilistic graphs $(G', P)$ such that a binary RI (UI) code of a certain rate for $(G', P)$ exists iff $G$ is $N$-colorable.

*Theorem 3:* Problem $A$ is $NP$-complete for $k = 2$.

*Proof:* Let $G = (V, E)$ be an instance of $GNC$, with $N = 4$. Write $|V| = n$. Construct the auxiliary graph $G' = (V', E')$ as follows:

$$\begin{aligned} V' &= V \cup \{i_1, i_2, i_3, i_4\} \cup \{j_1, j_2, j_3, j_4\} \cup \{k_1, k_2, k_3, k_4\} \\ E' &= E \cup \{\{i_a, i_b\}, \{j_a, j_b\}, \{k_a, k_b\}: 1 \le a \ne b \le 4\} \\ &\quad \cup \{\{i_a, j_b\}, \{i_a, k_b\}: 1 \le a \ne b \le 4\} \\ &\quad \cup \{\{j_a, k_b\}: 1 \le a, b \le 4\} \\ &\quad \cup \{\{j_a, v\}: 1 \le a \le 4, v \in V\} \end{aligned}$$

where by $1 \le a \ne b \le 4$ we mean $1 \le a \le 4$, $1 \le b \le 4$, and $a \ne b$. Thus, the subgraph induced in $G'$ by

$$\{i_1, i_2, i_3, i_4\} \cup \{j_1, j_2, j_3, j_4\} \cup \{k_1, k_2, k_3, k_4\}$$

is obtained by removing the edges $\{i_a, j_a\}$ and $\{i_a, k_a\}$, $1 \le a \le 4$, from the complete graph on those 12 nodes.

Next, assign the probability distribution $P$ on $V'$ such that, for $v \in V'$

$$P(v) = \begin{cases} \frac{\epsilon}{n}, & \text{if } v \in V \\ \frac{1-\epsilon}{12}, & \text{else.} \end{cases}$$

Let $\phi: V' \to \{0, 1\}^*$ be the optimal binary RI code for $(G', P)$. We hope that $\phi$ also minimizes $\sum_{v \in V'-V} |\phi'(v)|$ over all possible binary RI codes $\phi'$ for $(G', P)$. This can be guaranteed by choosing $\epsilon < \frac{1}{12n+25}$, since for such a choice of $\epsilon$ and any binary RI code $\phi'$, if

$$\sum_{v \in V'-V} |\phi(v)| > \sum_{v \in V'-V} |\phi'(v)|$$

then

$$\begin{aligned} &\left(\frac{1-\epsilon}{12} \sum_{v \in V'-V} |\phi(v)| + \frac{\epsilon}{n} \sum_{v \in V} |\phi(v)|\right) \\ &\quad - \left(\frac{1-\epsilon}{12} \sum_{v \in V'-V} |\phi'(v)| + \frac{\epsilon}{n} \sum_{v \in V} |\phi'(v)|\right) \\ &= \frac{1-\epsilon}{12}\left(\sum_{v \in V'-V} |\phi(v)| - \sum_{v \in V'-V} |\phi'(v)|\right) \\ &\quad - \frac{\epsilon}{n}\left(\sum_{v \in V} |\phi'(v)| - \sum_{v \in V} |\phi(v)|\right) \\ &\ge \frac{1-\epsilon}{12} - \frac{\epsilon}{n} n((n+3)-1) \\ &\quad \text{(in the impossible worst case when } |\phi'(v)| = n+3 \\ &\quad \text{and } |\phi(v)| = 1 \text{ for all } v \in V) \\ &> 0 \end{aligned}$$

and we have a contradiction.

The edge structure of the subgraph induced by $V' - V$ in $G'$ enforces the following constraints on the codewords $\phi(v)$, $v \in V' - V$:

1) $\phi(i_a)$, $\phi(j_a)$, and $\phi(k_a)$ may not prefix $\phi(i_b)$, $\phi(j_b)$, and $\phi(k_b)$, respectively, for $1 \le a \ne b \le 4$.
2) $\phi(j_a)$ may not prefix $\phi(k_b)$ for $1 \le a, b \le 4$, and *vice versa*.
3) $\phi(i_a)$ may not prefix $\phi(j_b)$ or $\phi(k_b)$ for $1 \le a \ne b \le 4$, and *vice versa*.

It is easy to check that $\sum_{v \in V'-V} |\phi(v)|$ is minimized by choosing

$$\begin{aligned} &\phi(i_1) = 00,\ \phi(i_2) = 01,\ \phi(i_3) = 10,\ \phi(i_4) = 11 \\ &\phi(j_1) = 000,\ \phi(j_2) = 010,\ \phi(j_3) = 100,\ \phi(j_4) = 110 \\ &\phi(k_1) = 001,\ \phi(k_2) = 011,\ \phi(k_3) = 101 \text{ and } \phi(k_4) = 111. \end{aligned}$$

Correspondingly, $\bar{l}(\phi)$ may be calculated as

$$\begin{aligned} \bar{l}(\phi) &= \frac{\epsilon}{n} \sum_{v \in V} |\phi(v)| + \frac{1-\epsilon}{12}\{4 \cdot 2 + 8 \cdot 3\} \\ &= \frac{8(1-\epsilon)}{3} + \frac{\epsilon}{n} \sum_{v \in V} |\phi(v)|. \end{aligned}$$

Now consider assignment of codewords to nodes in $V$. Since every $v \in V$ is connected to $\phi(j_a)$, the pair $\{\phi(j_a), \phi(v)\}$ should be prefix free for $1 \le a \le 4$ and $v \in V$. On the other hand, the pairs $\{\phi(i_a), \phi(v)\}$ and $\{\phi(k_a), \phi(v)\}$, $1 \le a \le 4$, $v \in V$, do not need to be prefix free.

Suppose now that $G$ is 4-colorable. Then the four color classes may be assigned codewords 001, 011, 101, and 111, respectively, so that

$$\bar{l}(\phi) = \frac{8(1-\epsilon)}{3} + \frac{\epsilon}{n} \cdot 3n = \frac{8+\epsilon}{3}.$$

Conversely, suppose $G$ is not 4-colorable. Then $|\phi(v)| > 3$ for at least one node $v \in V$, so that

$$\bar{l}(\phi) \ge \frac{8(1-\epsilon)}{3} + \frac{\epsilon}{n} \cdot 3n + \frac{\epsilon}{n} > \frac{8+\epsilon}{3}.$$

Thus, $G$ is 4-colorable iff there exists a binary RI code for $(G', P)$ of rate $\frac{8+\epsilon}{3}$ bits. □

Construction of the auxiliary graph in the previous theorem is slightly involved since, in an RI code, some codewords may be proper prefixes of others. In contrast, the construction for the binary UI coding case is simpler, as shown in the proof of the following theorem.

*Theorem 4:* Problem $B$ is $NP$-complete for $k = 2$.

*Proof:* Let $G = (V, E)$ be an instance of $GNC$, with $N = 3$. Construct $G' = (V', E')$ as follows:

$$\begin{aligned} V' &= V \cup \{i_1, i_2, i_3\} \\ E' &= E \cup \{\{i_1, i_2\}, \{i_1, i_3\}, \{i_2, i_3\}\}. \end{aligned}$$

Let $|V| = n$. Assign the probability distribution $P$ on $V'$ such that, for $v \in V'$

$$P(v) = \begin{cases} \frac{\epsilon}{n}, & \text{if } v \in V \\ \frac{1-\epsilon}{3}, & \text{else.} \end{cases}$$

Let $\phi$: $V' \to \{0, 1\}^*$ be the optimal binary UI code for $(G', P)$. We hope that $\phi$ also minimizes $\sum_{v\in V'-V} |\phi'(v)|$ over all possible binary UI codes $\phi'$ for $(G', P)$. This can be guaranteed by choosing $\epsilon < \frac{1}{3n+1}$, since for such a choice of $\epsilon$ and any binary UI code $\phi'$, if

$$\sum_{v\in V'-V} |\phi(v)| > \sum_{v\in V'-V} |\phi'(v)|$$

then

$$\begin{aligned}
&\left(\frac{1-\epsilon}{3}\sum_{v\in V'-V} |\phi(v)| + \frac{\epsilon}{n}\sum_{v\in V} |\phi(v)|\right)\\
&\quad - \left(\frac{1-\epsilon}{3}\sum_{v\in V'-V} |\phi'(v)| + \frac{\epsilon}{n}\sum_{v\in V} |\phi'(v)|\right)\\
&= \frac{1-\epsilon}{3}\left(\sum_{v\in V'-V} |\phi(v)| - \sum_{v\in V'-V} |\phi'(v)|\right)\\
&\quad - \frac{\epsilon}{n}\left(\sum_{v\in V} |\phi'(v)| - \sum_{v\in V} |\phi(v)|\right)\\
&\geq \frac{1-\epsilon}{3} - \frac{\epsilon}{n} n((n+3)-1)\\
&\quad \text{(in the impossible worst case when } |\phi'(v)| = n+3\\
&\quad \text{and } |\phi(v)| = 1 \text{ for all } v \in V)\\
&> 0
\end{aligned}$$

and we have a contradiction.

Suppose that $G$ is 3-colorable. Then $\phi$ has only three distinct codewords, with (say) $\phi(i_1) = 0$, $\phi(i_2) = 10$, and $\phi(i_3) = 11$. Further

$$\begin{aligned}
\bar{l}(\phi) &\leq \frac{1-\epsilon}{3}(1 + 2\cdot 2) + \frac{\epsilon}{n}\cdot 2n - \frac{\epsilon}{n}\\
&= \frac{5+\epsilon}{3} - \frac{\epsilon}{n}.
\end{aligned} \tag{21}$$

Conversely, suppose that $G$ is not 3-colorable. Then $\phi$ has at least four distinct (prefix-free) codewords. Also, crucially, the assignment above of codewords of lengths $1, 2, 2$ to $i_1$, $i_2$ and $i_3$ is no longer possible. Therefore, the rate $\bar{l}(\phi)$ is bounded from below as

$$\begin{aligned}
\bar{l}(\phi) &> \frac{1-\epsilon}{3}(2\cdot 3) + \frac{\epsilon}{n}\cdot n\cdot 1\\
&= 2-\epsilon.
\end{aligned} \tag{22}$$

Thus, $G$ is 3-colorable iff there exists a binary UI code for $(G', P)$ of rate not above $\frac{5+\epsilon}{3} - \frac{\epsilon}{n}$. □

Note that, from (21) and (22) in the above proof, if $G$ is 3-colorable, then $\bar{l}(\phi) \leq 5/3 + \epsilon(1/3 - 1/n)$, and, if $G$ is not 3-colorable, then $\bar{l}(\phi) > 2 - \epsilon$. The difference in rates between these two cases is not less than

$$2 - \epsilon - \frac{5}{3} - \epsilon\left(\frac{1}{3} - \frac{1}{n}\right) = \frac{1}{3} - \epsilon\left(\frac{4}{3} - \frac{1}{n}\right) \text{ bits.}$$

Now, suppose there exists a polynomial-time algorithm which is guaranteed to design a UI code of rate within $1/3 - \epsilon'$ bits of the minimum rate ($\epsilon' > 0$). After suitably picking $\epsilon$, this algorithm can be executed on $(G', P)$ to decide if $G$ is 3-colorable. Since, from the above theorem this is not possible unless $P = NP$, we have the following corollary.

*Corollary 1:* Let $(G, P)$ be the input. Finding a UI code $\phi$ for $(G, P)$ with rate $\bar{l}(\phi)$ such that

$$\bar{l}(\phi) - \overline{\mathcal{L}}(G, P) < \frac{1}{3} - \epsilon$$

for any fixed $\epsilon > 0$ is $NP$-hard.

This section focused on the complexity of scalar coding. One might be concerned about the complexity of block coding, where the corresponding graph is an AND power graph, and has a special structure. However, it is trivial to show that optimal design is still $NP$-hard in the size of the alphabet, and the complexity grows at least exponentially with the block length.

## VI. Optimal Coding Algorithms

Optimal coding algorithms for RI codes have previously also been considered by Zhao and Effros in [29]. Here, using the language of partition trees, the problem is separated into optimal code design for a given partition tree, and search for the optimal partition tree. The search itself is simplified by the derivation of necessary conditions for optimality of a partition tree. On the other hand, our approach is again graph theoretic, and is apparently simpler. Unlike the algorithm in [29], our algorithms are amenable to complexity analysis.

Let a probabilistic graph $(G, P)$ be given, and let $G' = (V', E')$ be the subgraph induced in $G$ by $V' \subseteq V$. We write

$$P(G') = \sum_{v\in V'} P(v)$$

for the total probability of the set $V'$. Let the distribution $P_{G'}$ denote the restriction of $P$ to $V'$, i.e., for $v \in V'$

$$P_{G'}(v) = \frac{P(v)}{P(G')}.$$

We then say that the probabilistic graph $(G', P_{G'})$ is induced in $(G, P)$ by $V' \subseteq V$.

We will illustrate the derivation of the optimal algorithms for binary codes. The algorithms can be naturally generalized to $k$-ary code design, $k \geq 2$, and we omit the details.

Throughout this section, we will use the following compact notation for the weighted codeword length of the subgraph $G'$:

$$\begin{aligned}
L(G') &= P(G')\overline{L}(G', P_{G'})\\
\mathcal{L}(G') &= P(G')\overline{\mathcal{L}}(G', P_{G'}).
\end{aligned}$$

Consider optimal RI code design. Let $\phi$: $V \to \{0, 1\}^*$ be the optimal RI code for $(G, P)$. If $i$ is an arbitrary intermediate (i.e., nonleaf) node of the code tree corresponding to $\phi$, we define the sets

$$\begin{aligned}
\phi^{-1}(i) &= \{v \in V'\colon \phi(v) = i\}\\
\phi^{-1}(i*) &= \{v \in V'\colon i \text{ is a prefix of } \phi(v)\}
\end{aligned}$$

and write $(G_{i*}, P_{i*})$ for the subgraph induced in $(G, P)$ by $\phi^{-1}(i*)$. Let $i0$ and $i1$ denote the two children of $i$. Then we have

$$L(G_{i*}) = L(G_{i0*}) + L(G_{i1*}) + P(G_{i*}) - P(\phi^{-1}(i)).$$

Since $\phi$ is the *optimal* code, this may be recast as

$$L(G_{i*}) = \min_{D \subseteq G_{i*} - \phi^{-1}(i)} \{L(D) + L(G_{i*} - \phi^{-1}(i) - D)\} + P(G_{i*}) - P(\phi^{-1}(i)). \quad (23)$$

(The minimization is over all induced subgraphs $D = (V_D, E_D)$ of $G_{i*} - \phi^{-1}(i)$. $G_{i*} - \phi^{-1}(i) - D$ is the subgraph induced by $\phi^{-1}(i*) - \phi^{-1}(i) - D$ in $G_{i*}$.)

This suggests a recursive algorithm to find $\overline{L}(G, P)$, and the corresponding optimal RI code. Let $G' = (V', E')$ be an induced subgraph of $G$, and let $I$ be the set of isolated nodes in $G'$. Then, as in (23), we have

$$L(G') = \min_{D \subseteq G' - I} \{L(D) + L(G' - I - D)\} + P(G') - P(I) \quad (24)$$

with the terminating condition

$$L(G') = 0 \quad (25)$$

when $G' = I$.

It is not necessary to search over all possible induced subgraphs in the minimization of (24).

*Definition 6:* A 2-partition $(D, V - D)$ of the vertex set of a graph $G = (V, E)$ is called a *dominating* 2*-partition*, [7], if every node in $V - D$ is connected to some node in $D$, and *vice versa*.

*Lemma 4:* Let $\phi$: $V \to \{0, 1\}^*$ be the optimal RI code for $(G, P)$, and let $I$ be the set of all isolated nodes of $G$. Then $(\phi^{-1}(0*), \phi^{-1}(1*))$ form a dominating 2-partition of $G - I$.

*Proof:* We will show that every node in $\phi^{-1}(0*)$ is connected to some node in $\phi^{-1}(1*)$. The lemma will then follow by reversing the roles of 0 and 1 in the subsequent arguments.

Note that an isolated node need not be assigned a codeword. A node in $\phi^{-1}(0)$ is not connected to any other node in $\phi^{-1}(0*)$. It should be connected to some node in $\phi^{-1}(1*)$; otherwise, it is isolated in $G - I$, and the rate can be reduced by moving it to $I$. If a node in $\phi^{-1}(0*) - \phi^{-1}(0)$ is not connected to any node in $\phi^{-1}(1*)$, the rate can again be reduced by assigning it the codeword 1. □

Thus, the minimization in (24) can be restricted to the dominating 2-partitions of $G' - I$.

Here, we outline a possible implementation of the optimal binary RI coding algorithm.

Input: $(G, P)$
1. for $r = 1$: $|V|$,
2. for $i = 1$: $\binom{|V|}{r}$,
3. calculate $L(G_{i,r})$ from (24) and (25), where $G_{i,r}$ is the $i$th $r$-node induced subgraph of $G$.
4. end.
5. end.
6. Calculate $\overline{L}(G, P)$ from (24).

Note that the previously calculated optimal codes for smaller subgraphs may be used in the minimization of Step 3. The worst case complexity, $C$, of the algorithm is determined as follows. In Step 3, in the worst case, an exhaustive search over all possible smaller subgraphs would have complexity $O(2^r)$. Thus,

$$C = \sum_{r=1}^{|V|} \binom{|V|}{r} O(2^r) = O(3^{|V|}).$$

A similar recursive algorithm can be derived for the design of optimal binary UI codes also. The recursive relation (24) is modified in this case to

$$\mathcal{L}(G') = \min_{D \subseteq G'} \{\mathcal{L}(D) + \mathcal{L}(G' - D)\} + P(G')$$

since the codeword set is required to be prefix free. The terminating condition remains unchanged: if $G' = I$

$$\mathcal{L}(G') = 0.$$

Clearly, the algorithm derived from these recursions again has a worst case complexity of $O(3^{|V|})$.

## VII. Fast (Suboptimal) Design Algorithms

### A. Design Algorithms Based on Approximate Coloring

Let $(G, P)$ be given. Let $P_{\max}$ and $P_{\min}$ denote the maximum and minimum node probabilities under $P$. In preparation for the performance analysis of suboptimal coding algorithms, we prove the following lemma. (Recall that $\alpha(G)$ is the size of the largest independent set in $G$, while $d(\overline{G}) = 1 +$ the maximum degree of $\overline{G}$. Clearly, $\alpha(G) \leq d(\overline{G})$.)

*Lemma 5:* Let $\phi$ be a UI code for $(G, P)$, of rate $\bar{l}(\phi)$. Let $c(G)$ be the cardinality of the coloring $c_\phi$ of $G$ induced by $\phi$, and $a(G)$ be the size of the smallest color class in this coloring. Then we have the following estimates for the suboptimality of $\phi$.

1) In terms of $\alpha(G)$ and $a(G)$

$$\bar{l}(\phi) - \overline{\mathcal{L}}(G, P) \leq 1 + \log \frac{\alpha(G)}{a(G)} + \log \frac{P_{\max}}{P_{\min}} \quad (26)$$

$$\bar{l}(\phi) - \overline{L}(G, P) \leq 1 + \log e + \log \frac{\alpha(G)}{a(G)} + \log \frac{P_{\max}}{P_{\min}} + \log\{1 - \log P_{\min}\}. \quad (27)$$

2) In terms of $c(G)$ and $d(\overline{G})$

$$\bar{l}(\phi) - \overline{\mathcal{L}}(G, P) \leq 2 + \log\left\{c(G) d\left(\overline{G}\right) P_{\max}\right\} - P_{\max} \log c(G) \quad (28)$$

$$\bar{l}(\phi) - \overline{L}(G, P) \leq 2 + \log e + \log\left\{c(G) d\left(\overline{G}\right) P_{\max}\right\} - P_{\max} \log c(G) + \log\{1 - \log P_{\min}\}. \quad (29)$$

*Proof:* We use the lower bounds for $\overline{L}(G, P)$ and $\overline{\mathcal{L}}(G, P)$ established in [1]

$$H_\chi(G, P) \leq \overline{\mathcal{L}}(G, P) \quad (30)$$

$$H_\chi(G, P) - \log\{H_\chi(G, P) + 1\} - \log e \leq \overline{L}(G, P). \quad (31)$$

Via elementary arguments, the chromatic entropy may itself be bounded in terms of $\alpha(G)$ as

$$H_\chi(G, P) \geq \log \frac{1}{P_{\max} \alpha(G)}. \quad (32)$$

1) The following upper bound for $H(c_\phi(X))$ is clear:

$$H(c_\phi(X)) \le \log \frac{1}{P_{\min} a(G)}.$$

Since $\phi$ is a prefix-free code for $c_\phi(V)$, we obtain (26) using (30) and (32):

$$\begin{aligned}\bar{l}(\phi) - \overline{\mathcal{L}}(G, P) &\le H(c_\phi(X)) + 1 - H_\chi(G, P)\\ &\le 1 + \log \frac{P_{\max}}{P_{\min}} + \log \frac{\alpha(G)}{a(G)}.\end{aligned}$$

Equation (27) similarly follows from (31) and (32).

2) Let $Q = (q_1, q_2, \ldots, q_m)$ be a probability distribution with $q_1 \ge q_2 \ge \cdots q_m$. The following upper bound on $H(Q)$ is easily derived:

$$H(Q) \le 1 + (1 - q_1)\log m.$$

Thus, we have

$$H(c_\phi(X)) \le 1 + \{1 - P_{\max}\} \log c(G).$$

Equations (28) and (29) now follow by using this estimate with (30) and (31) respectively, and paralleling the calculations made in proving 1). □

Consider the following generic algorithm for the design of codes for $(G, P)$.

1. Color $G$ using an approximate graph coloring algorithm $A$.
2. Find the Huffman code for the color probabilities. (Here, the probability of a color is the sum of probabilities of the nodes assigned that color.)
3. To each node assign the Huffman codeword of its color.

Since the codeword set is prefix free, a UI code is produced by this approach. We analyze two possible choices for the algorithm $A$ in Step 1. The first choice is an algorithm proposed by Halldórsson in [15]. This algorithm is guaranteed to color the graph $G$ with fewer than

$$c_{\max}(G) = O\left(\frac{|V|(\log\log|V|)^2 \chi(G)}{(\log|V|)^3}\right)$$

colors. The worst case performance guarantee of the corresponding coding algorithm may be calculated by substituting $c_{\max}(G)$ for $c(G)$ in (28) and (29).

In practice, worst case performance guarantees may be pessimistic, since worst cases occur infrequently. Thus, as a second choice for $A$, we consider an algorithm which provides good average-case performance. The following greedy graph coloring algorithm was analyzed by Grimmett and McDiarmid in [14]: Index the nodes of $G = (V, E)$ as $\{1, 2, \ldots, |V|\}$, and let $\{c_1, c_2, \ldots\}$ be the set of allowed colors. Color node 1 with $c_1$. Color node $i$ $(i > 1)$ with $c_j$ if $i$ has a neighbor already colored with $c_l$ for all $l = 1, 2, \ldots, j-1$, but not one with $c_j$. Our coding algorithm consists simply of greedily coloring the graph, then Huffman coding the colors.

Consider the following probability distribution on graphs with $n$ vertices: the probability that an edge joins any two given vertices of a randomly chosen graph is a fixed number $1 - p$ $(0 < p < 1)$ independently of any set of information about the presence or absence of other edges. Note that this is one of the standard random graph models; see, e.g., [6]. Let $G_n$ denote a graph drawn from this distribution. Then $\alpha(G_n)$ and $a(G_n)$ are, respectively, the size of the largest independent set in $G_n$ and the smallest number of nodes of $G_n$ assigned the same color by the greedy coloring algorithm.

*Lemma 6:* The sequence of random variables $\{\alpha(G_n)\}$ and $\{a(G_n)\}$ of random variables satisfy

$$\frac{\alpha(G_n)}{a(G_n)} \to 2 \text{ as } n \to \infty \tag{33}$$

in probability.

*Proof:* Let $T_n(G)$ denote the number of colors used by the greedy coloring algorithm to color $G_n$. Let $A_n^j(G)$ denote the number of nodes assigned color $c_j$, $j = 1, 2, \ldots$. The following results are proved in [14]:

$$\frac{\alpha(G_n)}{\log n} \to \frac{2}{\log 1/p}, \qquad \text{as } n \to \infty, \tag{34}$$

and

$$\frac{A_n^1(G_n)}{\log n} \to \frac{1}{\log 1/p}, \qquad \text{as } n \to \infty \tag{35}$$

in both cases almost surely (a.s.). Further

$$T_n \frac{\log n}{n} \to \log 1/p, \qquad \text{as } n \to \infty \tag{36}$$

in mean, and hence in probability.

By considering the effect of renumbering the nodes, it is clear that, a.s.

$$A_n^j(G) \le A_n^1(G), \qquad \text{for } j = 2, 3, \ldots.$$

Therefore, since $n = \sum_{j=1}^{T_n(G)} A_n^j(G)$, a.s.

$$n \le A_n^1(G) \cdot T_n(G).$$

But from (35) and (36)

$$\frac{A_n T_n}{n} \to 1, \qquad \text{as } n \to \infty$$

in probability. Thus, the $A_n^j(G), j = 1, 2, \ldots$ are equal in probability, and the claim of the lemma follows on again using (34) and (35). □

The result of this lemma, combined with (26) and (27), then proves the average-case performance guarantee claimed in (6) and (7) for the coding algorithm consisting of greedy graph coloring followed by Huffman coding of the colors. Note that the complexity of greedy coloring is $O(|V|)$, while that of Huffman coding is $O(|V|\log|V|)$. So the resultant coding algorithm only has complexity of $O(|V|\log|V|)$.

### *B. Design Algorithms Based on Graph Entropy Approximation*

We turn now to a second approach to the development of suboptimal design algorithms. The following lower bound for the minimum coding rates in terms of the graph entropy is from [1]

$$H(P) - H(\overline{G}, P) \le \overline{L}(G, P) \le \overline{\mathcal{L}}(G, P). \tag{37}$$

Note that exact calculation of the graph entropy of an arbitrary graph may be computationally expensive [25]. Instead, we propose an algorithm for approximating this quantity. We then show that this approximation may be used in the design of an RI code for $(G, P)$.

Graph entropy is defined by the following formula (see Definition 5):

$$H(G, P) = \min_{\substack{X \sim P \\ X \in S \in \mathcal{S}(G)}} I(X; S) \tag{38}$$

where $\mathcal{S}(G)$ is the collection of all independent sets in $G$.

We convert this minimization into a rate-distortion problem [4] by introducing the distortion function

$$d(x, s) = \begin{cases} 0, & \text{if } x \in s \in \mathcal{S}(G) \\ 1, & \text{otherwise} \end{cases}$$

and rephrasing (38) as

$$H(G, P) = \min_{\substack{X \sim P \\ E[d(X, S)]=0}} I(X; S).$$

The BA algorithm [5] for a slope $\beta \to \infty$ may now be used. The iterations of the BA algorithm become

$$p(s|x) = \begin{cases} \frac{q(s)}{\sum\limits_{s' \ni x} q(s')}, & \text{if } x \in s \in \mathcal{S}(G) \\ 0, & \text{otherwise} \end{cases}$$

$$q(s) = \sum_{x \in s} p(s|x)p(x)$$

and the corresponding optimality conditions are

$$\sum_{x \in s} \frac{p(x)}{\sum\limits_{s' \ni x} q(s')} \le 1, \qquad \forall\, s.$$

In order to reduce the computational burden, the collection $\mathcal{S}(G)$ may be restricted to, say, $\mathcal{S}_r(G)$ = collection of all independent sets of size $\le r$ for some fixed $r$. The approximation of $H(G, P)$ is then improved by increasing $r$.

RI code design may be based on approximate calculation of the lower bound $H(P) - H(\overline{G}, P)$ of (37). We shall illustrate the idea via an example. Let $\mathcal{S}_2(\overline{G})$ be the collection of all independent sets of size 2 in $\overline{G}$, i.e., of all edges of $G$. Note that the rate of an RI code $\phi$ may be expressed as

$$\sum_{x \in V} P(x)|\phi(x)| = \sum_{s \in \mathcal{S}(\overline{G})} q(s) \sum_{x \in s} p(x|s)|\phi(x)|. \tag{39}$$

The BA algorithm is used to calculate $q(s)$ for each edge $s$ of $G$, and these edges are ordered by decreasing probabilities. The edges are then traversed in order, resolving the prefix violation at that edge (if any) by extending the current codewords at the two ends of the edge by the minimum possible number of bits. The heuristic behind the algorithm is based on (39): if the higher probability edges are visited earlier, it is more likely that the codewords assigned to the corresponding vertices are shorter, thus reducing the rate of the designed code.

## VIII. Conclusion

In this paper, we considered zero-error source coding with side information at the decoder. We showed that the minimum asymptotic rate required is the complementary graph entropy of a graph associated with the problem. Previous studies of the complementary graph entropy had been motivated by its connections to the zero-error capacity and graph entropy.

Slepian and Wolf showed in [26] that knowledge at the encoder of decoder side information affords no advantage in terms of the asymptotic rate if occasional errors are tolerated. Lemma 3 shows that this is no longer the case in general if zero errors are required. But if the underlying graph is perfect, no loss of optimality may be entailed; this is yet another instance of the striking information-theoretic properties of perfect graphs. While perfectness of the graph is sufficient, it may be of interest to determine whether perfectness is also necessary for this optimality. This would also answer a question of Körner and Longo, raised in a different context in [19], about necessary conditions for $\overline{H}(G, P) + H(\overline{G}, P) = H(P)$.

Turning then to the design of optimal codes, we showed that this problem is $NP$-hard. Further, even suboptimal design (to within $1/3$ bits) is $NP$-hard for a particular class of codes where the codeword set is required to be prefix free. Investigation of hardness of approximability for the class of all codes remains essentially open.

We developed optimal and suboptimal algorithms for code design. While polynomial-time approximate algorithms represent one line of attack, another popular approach to $NP$-hard graph-theoretic problems has been the consideration of restricted classes of graphs. It may be of interest to identify classes of graphs more likely to be encountered in practical side-information coding scenarios. The existence of polynomial-time optimal coding algorithms for these graphs could then be investigated.

## Acknowledgment

The authors wish to thank the reviewers for their constructive comments. In particular, the suggestions of one of the reviewers helped to clarify the proof of Lemma 3, and the exposition of Example 1, while the comments of another reviewer resulted in simplification of the proofs of Theorems 3 and 4.

## References

[1] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1329–1339, Sept. 1996.

[2] ——, "A lower bound on the expected length of one-to-one codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1670–1672, Sept. 1994.

[3] C. Berge, *Graphs and Hypergraphs*. Amsterdam, The Netherlands: North-Holland, 1973.

[4] T. Berger, *Rate Distortion Theory—A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[5] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 460–473, July 1972.

[6] B. Bollobás, *Graph Theory. An Introductory Course*. Berlin/Heidelberg/New York: Springer-Verlag, 1979.

[7] E. J. Cockayne and S. T. Hedetniemi, "Toward a theory of domination in graphs," *Networks*, vol. 7, no. 3, pp. 247–261, Fall 1977.

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

[9] I. Csiszár and J. Körner, "On the capacity of the arbitrarily varying channel for maximum probability of error," *Z. Wahrsch. Verw. Gebiete*, vol. 57, pp. 87–101, 1981.

[10] I. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simonyi, "Entropy splitting for antiblocking corners and perfect graphs," *Combinatorica*, vol. 10, no. 1, 1990.

[11] I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Sources*. New York: Academic, 1982.

[12] T. Feder, E. Kushilevitz, M. Naor, and N. Nisan, “Amortized communication complexity,” *SIAM J. Comput.*, vol. 24, no. 4, pp. 736–750, Aug. 1995.

[13] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. San Francisco, CA: Freeman, 1979.

[14] G. R. Grimmet and C. J. H. McDiarmid, “On coloring random graphs,” *Math. Proc. Cambr. Philos. Soc.*, vol. 77, pp. 313–324, 1975.

[15] M. M. Halldórsson, “A still better performance guarantee for approximate graph coloring,” *Inform. Processing Lett.*, vol. 45, pp. 19–23, 1993.

[16] D. Huffman, “A method for the construction of minimum redundancy codes,” *Proc. IRE*, vol. 40, pp. 1098–1101, Sept. 1952.

[17] A. Kh. Al Jabri and S. Al-Issa, “Zero-error codes for correlated information sources,” in *Proc. Cryptography*, Cirencester, U.K., Dec. 1997, pp. 17–22.

[18] J. Körner, “Coding of an information source having ambiguous alphabet and the entropy of graphs,” in *Proc. 6th Prague Conf. Information Theory*. Prague, Czechoslovakia: Academia, 1973, pp. 411–425.

[19] J. Körner and G. Longo, “Two-step encoding of finite memoryless sources,” *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 778–782, Nov. 1973.

[20] L. Lovász, “Perfect graphs,” in *More Selected Topics in Graph Theory*, L. W. Beineke and R. J. Wilson, Eds. New York/London: Academic, 1983, pp. 55–87.

[21] ——, “On the Shannon capacity of a graph,” *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 1–7, Jan. 1979.

[22] K. Marton, “On the Shannon capacity of probabilistic graphs,” *J. Comb. Theory*, ser. B, vol. 57, no. 2, Mar. 1993.

[23] S. S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): Design and construction,” in *Proc. Data Compr. Conf.*, Snowbird, UT, Mar. 1999, pp. 158–167.

[24] C. E. Shannon, “The zero-error capacity of a noisy channel,” *IRE Trans. Inform. Theory*, vol. IT-2, pp. 8–19, Sept. 1956.

[25] G. Simonyi, “Graph entropy: A survey,” in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 1995, vol. 20, pp. 399–441.

[26] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.

[27] H. S. Witsenhausen, “The zero-error side information problem and chromatic numbers,” *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 592–593, Sept. 1976.

[28] Y. Yan and T. Berger, “On instantaneous codes for zero-error coding of two correlated sources,” in *Proc. IEEE Int. Symp. Information Theory*, Sorrento, Italy, June 2000, p. 344.

[29] Q. Zhao and M. Effros, “Optimal code design for lossless and near lossless source coding in multiple access networks,” in *Proc. Data Compression Conf.*, Snowbird, UT, Mar. 2001, pp. 263–272.