Zero-Error Source-Channel Coding with Side Information

Jayanth Nayak, *Student Member, IEEE*, Ertem Tuncel, *Member, IEEE*, and Kenneth Rose, *Fellow, IEEE*

Abstract

We consider coding for transmission of a source through a channel without error when the receiver has side information about the source. We show that separate source and channel coding is asymptotically suboptimal in general. By contrast, in the case of vanishingly small probability of error, separate source and channel coding is known to be asymptotically optimal. For the zero-error case, we show further that the joint coding gain can in fact be unbounded. Since separate coding simplifies code design and use, we also derive conditions on sources and channels for the optimality of separate coding.

J. Nayak and K. Rose are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, California - 93106. Email: {jayanth,rose}@ece.ucsb.edu

E. Tuncel is with the Department of Electrical Engineering, University of California, Riverside, California - 92521. Email: ertem@ee.ucr.edu

Zero-Error Source-Channel Coding with Side Information

I. Introduction

An information theoretic result that has had a profound impact on practical communication system design is the separation theorem, which says that source and channel code design can be separated without any asymptotic loss of optimality. The first theorem of this kind was proved by Shannon [1] who considered the case where a discrete memoryless source needs to be communicated over a discrete memoryless channel and a non-zero reconstruction error that asymptotically vanishes as the code block length increases is allowed. This theorem has since been shown to hold for most analytically tractable singe-user source-channel scenarios with a few exceptions under the asymptotically vanishing error constraint described above [2]. Note that separation theorems are asymptotic results and make no claims about the behavior at finite block lengths.

A study of communication systems under the more stringent error free constraint was also initiated by Shannon [3]. He characterized the zero-error capacity of the discrete memoryless channel both with and without feedback and established that the zero-error regime is different from the asymptotically vanishing error regime. As we shall shortly see, for the source-channel pair of [1], the separation theorem trivially holds even under a zero-error constraint. The question of optimality of source-channel separation in the zero-error case becomes far more interesting when the decoder has access to side-information about the source. For this communication scenario we resolve the question and demonstrate that zero-error behavior and the asymptotically vanishing error behavior differ substantially.

Let $\mathcal{C}$ be a discrete memoryless channel with transition probability $p_{Y|X}(y|x), x \in \mathcal{X}, y \in \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets. With an asymptotically vanishing error requirement, the capacity of this channel is $C = \max_{p_X(x)} I(X,Y)$, where $I(X;Y)$ is the mutual information between X and Y. The zero-error capacity, C_0, which was characterized by Shannon [3], will be discussed in detail in the following section.

Let $(\mathcal{S}_U, \mathcal{S}_V)$ be a pair of memoryless correlated sources producing a pair of random variables (U, V) from a finite set $\mathcal{U} \times \mathcal{V}$ at each instant. Alice, “the sender”, has access to U while Bob, “the receiver”, has access to V. Alice and Bob are connected by the channel $\mathcal{C}$. We wish to determine the minimum amount of channel resources required for Alice to convey U to Bob. Specifically, we are interested in (m, n)-codes that convey m realizations of U in n channel uses. The ratio $\frac{n}{m}$ channel uses per source symbol is called the code rate.

Consider first the case where U needs to be conveyed across channel $\mathcal{C}$ and there is no side-information at the receiver. For the asymptotically vanishing error case, Shannon showed that source and channel coding can be done separately without any asymptotic loss of optimality: For all $\epsilon > 0$, separate source and channel coding can attain a rate less than $\frac{H(U)}{C} + \epsilon$ with probability of error less than ϵ, while no source-channel code can attain rates less than $\frac{H(U)}{C}$. Here $H(U)$ is the Shannon entropy of the random variable U. It is trivial to show that separate source and channel coding is asymptotically optimal even for the zero-error case: if $\hat{\mathcal{U}} = \{u \in \mathcal{U} : p_U(u) > 0\}$ is the support of U, separate source and channel coding can achieve a rate of $\frac{|\hat{\mathcal{U}}|}{C_0} + \epsilon$, where $|s|$ denotes the cardinality of a set s. Since there is no side-information, in any (m, n)-code the channel outputs must be sufficient to distinguish distinct members of $\hat{\mathcal{U}}^m$, which implies that there is a one-to-one mapping between the elements of $\hat{\mathcal{U}}^m$ and a channel code for $\mathcal{C}$ at block length n. Therefore the minimum rate achievable by a source-channel code is $\frac{|\hat{\mathcal{U}}|}{C_0}$ and in fact all source-channel codes correspond to separate source and channel codes.

Suppose we wish to design a source-channel code for the source U with side-information V and channel $\mathcal{C}$ with probability of error $\epsilon \in (0, 1)$. The celebrated result of Shannon [1] states that there exists a channel code for $\mathcal{C}$ at some block length n that conveys at least $C - \frac{\epsilon}{2}$ bits per channel use and result in a probability of error less than $\frac{\epsilon}{2}$. Another celebrated result of Slepian and Wolf [4] states that there exists a source code for the source-side-information pair (U, V) at some block length m with per symbol rate less than $H(U|V) + \frac{\epsilon}{2}$ bits and probability of error less than $\frac{\epsilon}{2}$. Here $H(U|V)$ is the conditional Shannon entropy of U given V. Concatenating the source and channel codes, we have a source-channel code of rate less than $\frac{H(U|V)+\frac{\epsilon}{2}}{C-\frac{\epsilon}{2}} < \frac{H(U|V)}{C} + 2\epsilon$ and probability of error less than ϵ as was desired. On the other hand, Shamai and Verdú [5] have shown that codes with rate less than $\frac{H(U|V)}{C}$ cannot exist even if joint source-channel coding is employed. Hence separate source and channel coding is asymptotically optimal when a vanishingly small probability of error is allowed.

In this paper we focus on the *zero-error* setting for the problem of source-channel coding with side-information. Sections III and IV present our main results – the suboptimality of separate coding and the gains by joint coding. In Section V, we turn to the question of when separate coding is indeed optimal and present sufficient conditions on sources and channels. In Section VI we present some comments on the complexity of code design before concluding in Section VII. Since the imposition of zero-error constraints naturally leads to problem formulations in terms of graphs, we first provide useful graph theoretic definitions in Section II.

II. Preliminaries and Notation

A. Graph theoretic preliminaries

A graph $G = (V, E)$ is defined by a vertex set $V = \{v_i, i = 1, \ldots, |V|\}$ and a set of edges $E \subseteq V \times V$. We only consider undirected graphs, that is $(v, v') \in E \Rightarrow (v', v) \in E$. The subgraph induced by a set $V' \subseteq V$ is the graph $G' = (V', E')$ where $E' = (V' \times V') \cap E$. The complement of a graph $G = (V, E)$ is the graph $\bar{G} = (V, \bar{E})$ where $\bar{E} = \{(v, v') : v, v' \in V, v \neq v', (v, v') \notin E\}$.

An *independent set* in G is a set of vertices such that the graph it induces in G is edge free. The stability number $\alpha(G)$ is the cardinality of the largest independent set in G. A coloring is a mapping of vertices to “colors” such that connected vertices receive distinct colors. The *chromatic number* of G, $\chi(G)$ is the minimum number of colors that are required for coloring G. Since only independent sets can receive the same color, a coloring can also be viewed as a partitioning of V into independent sets. A complete graph is one where every pair of vertices is connected by an edge. The complete graph on m vertices will be denoted by K_m. A clique of G is a subset of the vertices that are pairwise connected, that is the subgraph induced by this subset is a complete graph. The clique number, $\omega(G)$, is the cardinality of the largest clique in G.

The normal or AND product of two graphs $G = (V, E)$ and $H = (V', E')$, denoted $G \times H = (V \times V', E_\times)$ is the graph on $V \times V'$ where $((v_1, v_1'), (v_2, v_2')) \in E_\times$ if and only if $v_1 = v_2, (v_1', v_2') \in E'$ or $(v_1, v_2) \in E, v_1' = v_2'$ or $(v_1, v_2) \in E, (v_1', v_2') \in E'$. G^n denotes the n-fold AND product of G with itself. The conormal or OR product of two graphs $G = (V, E)$ and $H = (V', E')$, denoted $G \bullet H = (V \times V', E_\bullet)$ is the graph on $V \times V'$ where $((v_1, v_1'), (v_2, v_2')) \in E_\bullet$ if and only if $(v_1, v_2) \in E$ or $(v_1', v_2') \in E'$. Observe $E_\times \subseteq E_\bullet$. $G^{(n)}$ denotes the n-fold OR product of G with itself.

Given two graphs $G = (V, E)$ and $H = (V', E')$, a *homomorphism* is said to exist from G to H, denoted $G \to H$, if there exists a mapping $\phi : V \to V'$ such that if $(v_1, v_2) \in E$, then $(\phi(v_1), \phi(v_2)) \in E'$. If there is no such mapping we denote it by $G \nrightarrow H$. The relation "$\to$" is reflexive ($G \to G$) and transitive ($G \to H$ and $H \to F \Rightarrow G \to F$) but not symmetric ($G \to H \nRightarrow H \to G$). Homomorphisms from G to H can be considered as generalizations of colorings of G since the set of vertices in G that are mapped to a given vertex in H must be an independent set. Clearly, $G \to K_m$ if and only if $m \geq \chi(G)$. Also $K_m \to G$ if and only if $1 \leq m \leq \omega(G)$.

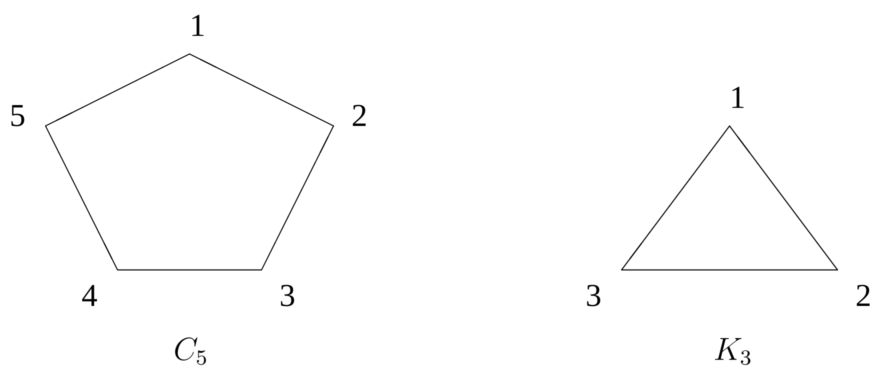

Fig. 1. The pentagon C_5 and the complete graph on 3 vertices K_3

To illustrate these graph-theoretic concepts, consider the examples in Fig. 1. $\{1, 3\}$ is one of the largest independent sets in C_5 and hence the stability number $\alpha(C_5)$ is 2. Similarly, $\{1, 2\}$ is one of the largest cliques in C_5 and hence the clique number $\omega(C_5)$ is also 2. A coloring that requires the fewest colors assigns distinct colors to the sets $\{1, 3\}, \{2, 4\}$ and $\{5\}$. The chromatic number, $\chi(C_5)$ is 3. In the second example of K_3, we have $\alpha(K_3) = 1, \omega(K_3) = \chi(K_3) = 3$.

Moving to homomorphisms, consider the mapping $\phi : \{1, 2, 3, 4, 5\} \to \{1, 2, 3\}$ given in Table I. This is easily seen to be a homomorphism from C_5 to K_3. Therefore $C_5 \to K_3$, which is consistent with the coloring interpretation of homomorphisms since $\chi(C_5) = 3$. On the other hand, there is no mapping from $\{1, 2, 3\}$ to $\{1, 2, 3, 4, 5\}$ that maps connected vertices in K_3 to connected vertices in C_5. Therefore $K_3 \nrightarrow C_5$. Again this ties in with our comments above since $3 > \omega(C_5)$.

i	$\phi(i)$
1	1
2	2
3	3
4	1
5	2

TABLE I

A HOMOMORPHISM FROM C_5 TO K_3

B. Channel coding

A channel is specified by a finite input alphabet $\mathcal{X}$, an output alphabet $\mathcal{Y}$ and a transition probability function $p_{Y|X}(y|x)$. $Y^n(x^n)$ denotes the random channel output when $x^n \in \mathcal{X}^n$ is the input. The fan out set $F_x \subseteq \mathcal{Y}$ of $x \in \mathcal{X}$ is the set of output letters such that $p_{Y|X}(y|x) > 0$. With every channel, we can associate a characteristic graph G_X with vertex set $\mathcal{X}$ where two vertices $x, x' \in \mathcal{X}$ are connected by an edge if their respective fan out sets F_x and $F_{x'}$ do *not* intersect.

A channel code of block length n is a pair of mappings: an encoder $\phi_c^n : \{1, \dots, 2^{nC}\} \to \mathcal{X}^n$ and a decoder $\psi_c^n : \mathcal{Y}^n \to \{1, \dots, 2^{nC}\}$. A zero-error code is one where $\psi_c^n(Y^n(\phi_c^n(i))) = i, \forall i \in \{1, \dots, 2^{nC}\}$ with probability 1. The zero-error n-use capacity of a channel is the largest value of C for which a zero-error n-length block code exists. For a scalar (block length 1) code to be zero-error, the fan out sets of the symbols in the image of $\phi_c^1(\cdot)$ must be pairwise disjoint. This implies that these symbols form a clique in G_X. The 1-use capacity of the channel is $\log \omega(G_X)$ bits[1]. Similarly, in n uses of the channel, $\omega(G_X^{(n)})$ messages can be transmitted. $G_X^{(n)}$ generalizes the characteristic graph to block coding since we can distinguish two vectors in $\mathcal{X}^n$ on the basis of their outputs if along at least one coordinate they cannot result in the same output. We see that the zero-error capacity of the channel depends only on its characteristic graph G_X. The zero-error capacity of the graph G_X (in bits per channel use) is defined as [3]

$$C(G_X) = \lim_{n \to \infty} \frac{1}{n} \log \omega(G_X^{(n)}). \tag{1}$$

The limit in (1) exists due to the supermultiplicativity of $\omega(G_X^{(n)})$ (Fekete's Lemma [7]).

For example, the n-fold OR product of the complete graph K_m, $K_m^{(n)}$, is K_{mn}. Therefore $C(K_m) = m$. Another example, which is important for our exposition, is the pentagon C_5. Finding the capacity of this graph remained an open problem for more than two decades since the problem was first posed by Shannon [3]. Shannon showed that $\omega(C_5^{(2)}) = 5$, which implies that the capacity of C_5 is lower bounded by $\frac{1}{2} \log 5$. By using a graph functional

[1] All logarithms are to base 2

called the theta function, Lovász [11] showed that $\frac{1}{2}\log 5$ is in fact the capacity of the C_5. Lovász's theta function is a polynomially computable upper bound on the capacity of a graph. It is defined as follows:

Consider a graph $G = (V, E)$. To each vertex v_i, assign a unit vector u_i from a fixed d dimensional space such that the vectors associated with connected vertices are orthogonal. Let h, the handle, be an arbitrary unit vector from the same space. $(u_1, \dots, u_{|V|}, h)$ is called an orthonormal representation with handle. Let $\mathcal{U}_G$ be the set of all orthonormal representations with handle associated with the graph G. The Lovász theta function of G is

$$\vartheta(G) = \min_{(u_1,\dots,u_{|V|},h)\in\mathcal{U}_G} \max_{i\in\{1,\dots,|V|\}} \frac{1}{(h^T u_i)^2}. \tag{2}$$

Of the several interesting properties of the theta function, most useful for this paper are the following [11], [13]. For graphs G, H

$$\omega(G) \le \vartheta(G) \le \chi(G) \tag{3}$$

$$C(G) \le \log\vartheta(G) \tag{4}$$

$$\vartheta(G \times H) = \vartheta(G \bullet H) = \vartheta(G)\vartheta(H) \tag{5}$$

C. Source coding with side information at the decoder

$\mathcal{S}_\mathcal{U}$ and $\mathcal{S}_\mathcal{V}$ are a pair of memoryless correlated sources producing U_i and $V_i \sim p_{UV}(u, v), i = 1\dots\infty$ respectively. Alice, who has access to $\mathcal{S}_\mathcal{U}$ needs to transmit her information to Bob, who has access to $\mathcal{S}_\mathcal{V}$. For zero-error code design, the relevant features of the source are captured by the support set $\mathcal{S}_{UV} \subseteq \mathcal{U}\times\mathcal{V}$, which is the set of pairs $(u, v) \in \mathcal{U}\times\mathcal{V}$ such that $p_{UV}(u, v) > 0$. For encoder design and hence minimum asymptotic rate calculation, we can further reduce the source to its confusability graph G_U on $\mathcal{U}$ where $u, u' \in \mathcal{U}$ are connected if and only if there exists $v \in \mathcal{V}$ such that $p_{UV}(u, v) > 0$ and $p_{UV}(u', v) > 0$. The precise mode by which the confusability graph is utilized in coding will be made clear in the following subsections.

A source code of block length n consists of an encoder $\phi_s : \mathcal{U}^m \to \{1, \dots, 2^{mR}\}$ and a decoder $\psi_s : \{1, \dots, 2^{mR}\} \times \mathcal{V}^m \to \mathcal{U}^m$.

1) Unrestricted inputs: The unrestricted inputs (UI) version is motivated by the need to eliminate catastrophic events in the scenario where the side information can occasionally be incorrect. On Alice's side, the U_i are independent and identically distributed according to $p_U(u)$, the marginal of $p_{UV}(u, v)$. Bob observes $\hat{V}_i$, chosen from $\mathcal{V}$, which is a corrupted version of V_i and need not necessarily result in $(U_i, \hat{V}_i)$ from the support set $\mathcal{S}_{UV}$. We require Bob to decode correctly only when $(U_i, \hat{V}_i)$ is in $\mathcal{S}_{UV}$. Let $\psi_s(\phi_s(U^m), \hat{V}^m) = \hat{U}^m$. A zero-error UI code is one where $\hat{U}_i = U_i$ with probability 1 for all i such that $(U_i, \hat{V}_i) \in \mathcal{S}_{UV}$. The m-instance UI rate is the smallest value of R for which an m-length zero-error UI code exists. For $m = 1$, the zero-error constraint means that only independent sets in G_U can receive the same codeword. Hence, to code a single instance $\chi(G_U)$ codewords are necessary and sufficient. While coding blocks of length m, two vectors $\mathbf{u} = (u_1, \dots, u_m)$ and $\mathbf{u}' = (u'_1, \dots, u'_m)$ are confusable if there exists some coordinate i such that (u_i, u'_i) is an edge in G_U. This follows from the fact that if Bob's side information along coordinate i, v_i is such that $p_{UV}(u_i, v_i)p_{UV}(u'_i, v_i) > 0$ then, even if the side information along all other coordinates is incorrect, the codewords for $\mathbf{u}$ and $\mathbf{u}'$ must be different for zero-error decoding. The minimum number of codewords is therefore $\chi(G_U^{(m)})$, the chromatic number of the m-fold OR product of G_U with itself. The asymptotic rate (in bits per source symbol) is given by

$$R^*(G_U) = \lim_{m\to\infty} \frac{1}{m} \log(\chi(G_U^{(m)})). \tag{6}$$

The limit in (6) exists due to the submultiplicativity of $\chi(G_U^{(m)})$. McEliece and Posner [8] and Berge and Simonovits [9] showed that

$$\lim_{m\to\infty} \frac{1}{m} \log(\chi(G_U^{(m)})) = \log \chi^*(G_U). \tag{7}$$

$\chi^*(G_U)$, the fractional chromatic number of the graph G_U, is the solution to the following linear program:

$$\chi^*(G_U) = \min_{\mathbf{y}} \{ \sum_{i=1}^{|\mathcal{S}(G_U)|} y_i : \sum_{\substack{j\in\mathcal{S}(G_U)\\ j\ni u}} y_j \geq 1, \forall u \in \mathcal{U}, \mathbf{y} \geq 0\}, \tag{8}$$

where $\mathcal{S}(G_U)$ denotes the family of independent sets of G_U. Note that if we further require that $y_i \in \{0, 1\}$, we obtain the integer program for finding the chromatic number of G_U.

Since $K_m^{(n)} = K_{mn}$ and every vertex in a complete graph must be assigned a different color in any coloring, $R^*(K_m) = \log m$. For the pentagon C_5, in the linear program defining $\chi^*(C_5)$, consider the assignment $y_i = \frac{1}{2}$ for all i of the form $\{j, (j+2) \bmod 5\}, j \in \mathcal{U}$ and $y_i = 0$ for all other i. Since the resulting cost $\frac{5}{2}$ is also attainable by the dual of the linear program, we have that $R^*(C_5) = \log \chi^*(C_5)$ is $\log \frac{5}{2}$.

2) Restricted Inputs: In the restricted inputs (RI) scenario, Alice again observes the U_i, while Bob's observation $\hat{V}_i$ is identical to V_i. A zero-error RI code is one where $\psi_s(\phi_s(U^m), V^m) = U^m$ with probability 1. The m-instance RI rate is the smallest value of R for which an m-length zero-error RI code exists. Again, for scalar coding only independent sets in G_U can be assigned the same codeword. So the minimum number of codewords required is the chromatic number of the graph $\chi(G_U)$. Suppose we wish to encode a block of length m. Two realizations of U^m, $\mathbf{u} = (u_1, \ldots, u_m)$ and $\mathbf{u}' = (u'_1, \ldots, u'_m)$, are confusable given the side information if they are confusable in every coordinate. To see this consider the case when u_i and u'_i are distinguishable. If the two vectors $\mathbf{u}$ and $\mathbf{u}'$ are assigned the same codeword, they can still be decoded correctly on the basis of v_i, the side information in the i^{th} coordinate. Therefore the confusability graph for m instances is the m-fold AND power of G_U, G_U^m. The minimum number of codewords is the chromatic number $\chi(G_U^m)$. The asymptotic rate (in bits per source symbol), called the Witsenhausen rate of a graph, is given by [10]

$$R_w(G_U) = \lim_{m\to\infty} \frac{1}{m} \log(\chi(G_U^m)). \tag{9}$$

The limit in (9) exists due to the submultiplicativity of $\chi(G_U^m)$.

K_m^n is also K_{mn} and hence $R_w(K_m)$ is $\log m$ as well. For the pentagon example, the results of Lovász [11] on its capacity combined with the results of Witsenhausen [10] imply that $R_w(C_5) = \frac{1}{2}\log 5$.

We note in passing that computation of channel capacity and Witsenhausen rate for an arbitrary graph remains an open problem.

III. Source-Channel Coding – Main Results

In this section, we define zero-error source-channel codes and present our results.

A source-channel (m, n)-code is again a pair of mappings: the encoder $\phi_{sc}^{mn} : \mathcal{U}^m \to \mathcal{X}^n$ and the decoder $\psi_{sc}^{mn} : \mathcal{Y}^n \times \mathcal{V}^m \to \mathcal{U}^m$. Let $\hat{V}^n$ be the side-information observed by the decoder. In the RI case, this is the same as V^n. Using the side-information and the output of the encoder, the decoder produces $\psi_{sc}^{mn}(Y^n(\phi_{sc}^{mn}(U^m)), \hat{V}^m) = \hat{U}^m$. Of a zero-error UI code we require $\hat{U}_i = U_i$ with probability 1 for all i such that $(U_i, \hat{V}_i) \in \mathcal{S}_{UV}$. Of a zero-error RI code $\hat{U}_i = U_i$ with probability $1, \forall i$. In both cases this means that m-length vectors of the source alphabet that are not distinguishable on the basis of the side information, must be distinguishable through the channel outputs they induce. For the case where $m = n = 1$ this implies that if two nodes are connected in G_U, their images under ϕ_{sc}^{11} must also be connected in G_X. In other words, we seek homomorphisms from G_U to G_X. The source confusability graph and the channel characteristic graph capture all the information in a source-channel pair required for zero-error source-channel coding.

In the sequel, we represent the source-channel pair by the corresponding graph pair. Further, in this section we shall restrict our attention to rate 1 codes, which we denote $\mathbb{1}$-codes. A $\mathbb{1}$-code of block length n is a (n, n)-code. Separate $\mathbb{1}$-codes are $\mathbb{1}$-codes where source and channel coding are performed independent of each other. A source-channel pair is called UI (RI) compatible if there exists a UI (RI) $\mathbb{1}$-code (for some block length n). For example, (G, G) is both UI and RI compatible – any source with confusability graph G can be transmitted through a channel whose characteristic graph is also G using $(1, 1)$-codes. To show the suboptimality of separate coding, we present cases where joint $\mathbb{1}$-codes exist at block length 1, but no separate $\mathbb{1}$-codes exist at any block length.

If source and channel coding are done separately, we have a composition of maps $\phi_s^m : \mathcal{U}^m \to \{1, \ldots, 2^{mR}\}$ and $\phi_c^n : \{1, \ldots, 2^{nC}\} \to \mathcal{X}^n$ at the encoder. The zero-error constraint implies that a one-to-one mapping should exist between source encoder output and channel encoder input which in turn implies $mR \leq nC$. Note that in the UI case $R \geq R^*(G_U)$, while in the RI case $R \geq R_w(G_U)$. In both cases $C \leq C(G_X)$. Therefore for a separate $\mathbb{1}$-code to exist at some block length, the minimum asymptotic rate for the source should not be greater than the capacity of the channel.

A. Separate Coding is Asymptotically Suboptimal: Unrestricted Inputs

With unrestricted inputs, an (m, n)-code is a homomorphism $\phi_{sc}^{mn} : \mathcal{U}^n \to \mathcal{X}^n$ from the m-fold OR product of the source confusability graph $G_U^{(m)}$ to the n-fold OR product of the channel characteristic graph $G_X^{(n)}$.

In the separate source and channel coding case, the source can be encoded using block coding to any rate greater than $R^*(G_U)$. Therefore separate $\mathbb{1}$-codes cannot exist if $C(G_X) < R^*(G_U)$.

Consider a source-channel pair such that $G_X = G_U = C_5$, the pentagon. The capacity of the channel [11] is $\frac{1}{2} \log 5$, while the minimum asymptotic rate of the source is $\log \frac{5}{2}$. $C(G_X) < R^*(G_U)$ and communication with separate source and channel coding is not possible at any block length.

However, the source and channel graph are identical. Therefore the pair is UI compatible and the identity mapping $\phi_{sc}^{11}(i) = i, \forall i \in \{1, 2, 3, 4, 5\}$ is a valid zero-error $(1, 1)$-code.

We thus prove the asymptotic suboptimality of separate source and channel coding in the UI scenario.

B. Separate Coding is Asymptotically Suboptimal: Restricted Inputs

With restricted inputs, an (m, n)-code is a homomorphism $\phi_{sc}^{mn} : \mathcal{U}^n \to \mathcal{X}^n$ from the m-fold AND product of the source confusability graph, G_U^m to the n-fold OR product of the channel characteristic graph $G_X^{(n)}$.

If source and channel coding are done separately, the source can be encoded using block coding to any rate greater than $R_w(G_U)$. Therefore separate $\mathbb{1}$-codes cannot exist if $C(G_X) < R_w(G_U)$.

The example in the previous subsection is no longer useful as a counterexample since by [10] $R_w(G_U) = \log\sqrt{5} = C(G_X)$.

Our approach here is to find an appropriate graph G and let $G_U = G_X = G$. We employ an upper bound on the capacity of a graph which is also a lower bound on its Witsenhausen rate. Existence of a graph G for which either bound is not tight will immediately prove that separate source and channel coding is asymptotically suboptimal and outperformed by joint source-channel coding. The upper bound on the capacity is the one discovered by Lovász [11]: the theta function $\vartheta(G)$. The fact that it is also a lower bound on $R_w(G)$ is implicit in a paper by Marton [12]. Our next theorem makes this relation explicit.

Theorem 1: For any graph $G = (V, E)$

$$\log\vartheta(G) \leq R_w(G) \tag{10}$$

Proof:

$$\begin{aligned} \vartheta(G^n) &\overset{(a)}{\leq} \chi(G^n) \\ (\vartheta(G))^n &\overset{(b)}{\leq} \chi(G^n) \\ \vartheta(G) &\leq (\chi(G^n))^{\frac{1}{n}} \end{aligned}$$

Here (a) and (b) follow from (3) and (5) respectively. Taking logarithms and the limit as $n \to \infty$,

$$\log\vartheta(G) \leq \lim_{n\to\infty} \frac{1}{n}\log(\chi(G^n)) = R_w(G).$$

■

Haemers [14] presents an example where $C(G) < \vartheta(G)$ - the Schläfli graph. The Schläfli graph G_{27} is a strongly regular graph on 27 vertices. A strongly regular graph with parameters (n, k, λ, μ) is a graph G with n vertices, not complete or null, in which the number of common neighbors of vertices v and v' is k, λ or μ according to whether v and v' are equal, adjacent or non-adjacent respectively. The parameter set of G_{27} is $(27, 10, 1, 5)$. Let both the source and channel graphs, G_U and G_X, be G_{27}. Haemers proved that $C(G_X) \leq \log 7 < \vartheta(G_X) = \log 9$. Therefore, by Theorem 1, $C(G_X) < R_w(G_U)$, although transmission is possible using scalar joint source-channel coding[2].

[2]This example can also be used to show that joint source-channel coding helps in the unrestricted inputs scenario since $R_w(G_U) \leq R^*(G_U)$ (The edges in the AND product are a subset of the edges in the OR product). We provided a separate example to highlight the difference between the two cases.

IV. How large is the joint source-channel coding gain?

Given a source-channel pair (G_U, G_X), let us rephrase the problem as: what is the minimum rate required for zero-error communication? With separate coding, the minimum rate in the UI case is $r^*(G_U, G_X) = \frac{R^*(G_U)}{C(G_X)}$ while in the RI case it is $r_w(G_U, G_X) = \frac{R_w(G_U)}{C(G_X)}$. However, there is no expression known for the corresponding quantity in the joint coding case. Let us focus on the special case where $G_U = G_X$, where we are guaranteed that the minimum rate is at most 1 channel use per source symbol. Using a recent result by Alon [15], we show in Theorem 2 that both $r^*(G_U, G_X)$ and $r_w(G_U, G_X)$ can be *arbitrarily large* even in this restricted scenario. Hence, the joint coding gain is generally unbounded.

Lemma 1: For every k, there exists a graph G such that $C(G) < \log k$ and

$$\vartheta(G) \geq k^{(1+o(1))\frac{\log k}{8 \log\log k}} \tag{11}$$

with the $o(1)$-term tending to zero as k tends to infinity.

Proof: This result is essentially due to Alon [15]. For completeness, we present his construction and an outline of the proof. The graph family is constructed as follows: Let p and q be a pair of primes, $s = pq - 1$ and $r > s$ an integer. Let $H = H(p, q, r)$ be a graph whose vertices are all the subsets of $\{1, \ldots, r\}$ that are of cardinality s. Two vertices are connected if the intersection of the associated sets has cardinality $-1 \pmod p$. H has $m = \binom{r}{s}$ vertices. For the choice of parameters $q < p < q + O(q^{\frac{2}{3}})$ and $r = p^3$, Alon showed that $C(H) < 2\binom{p^3}{p-1}$ and $C(\bar{H}) < 2\binom{p^3}{q-1}$. The existence of such pairs of primes follows from standard number theoretic results [16]. Note that if G is any graph on n vertices, $C(G \times \bar{G}) \geq n$ (the set of nodes $\{(i,i), i \in V(G)\}$ will always form a clique in $G \times \bar{G}$) . Therefore $C(H \times \bar{H}) \geq m = \binom{p^3}{pq-1}$.

Finally, for a given k, choose primes p and q as follows

i) p is the largest prime such that $2\binom{p^3}{p-1} < k$.

ii) q is the largest prime less than p.

For the corresponding $H = H(p, q, p^3)$, we have:

$$(\max[\vartheta(H), \vartheta(\bar{H})])^2 \geq \vartheta(H)\vartheta(\bar{H}) \overset{(a)}{=} \vartheta(H \times \bar{H}) \overset{(b)}{\geq} 2^{C(H \times \bar{H})} \overset{(c)}{\geq} k^{(1+o(1))\frac{\log k}{4 \log\log k}}$$

where (a) follows from (5), (b) follows from (4) and (c) follows from the relations between p, q and k, the standard bounds on binomial coefficients and results on the distribution of primes [16]. Therefore at least one of H and $\bar{H}$ satisfies (11). ■

Theorem 2: Given any l, we can find a graph G such that

$$\frac{R^*(G)}{C(G)} \geq \frac{R_w(G)}{C(G)} \geq l.$$

Proof: Fix k and let G be as in Lemma 1,

$$\frac{R^*(G)}{C(G)} \overset{(a)}{\geq} \frac{R_w(G)}{C(G)} \overset{(b)}{\geq} \frac{\log \vartheta(G)}{C(G)} \overset{(c)}{\geq} \frac{(1+o(1))\log k}{8 \log\log k},$$

where (a) follows from $R^*(G) \geq R_w(G)$, (b) follows from Theorem 1, and (c) follows from Lemma 1. Examining the right hand side of (c), we see that we can always find a k that makes it greater than the given l. ■

V. When is separate coding optimal?

Our main result was that separating source and channel coding was suboptimal. However, separate coding offers the advantage of reusing the code design. One could design the source code and use the same code for more than one channel. One could similarly reuse the channel code for various sources. So it is of interest to characterize source (channel) graphs such that separate coding is optimal for all channel (source) graphs. We were able to provide conditions on graphs that are sufficient for optimal source-channel separation.

Theorem 3: Asymptotic optimality is achievable by separate UI and RI coding if one of the following two conditions is satisfied:

i) the channel graph G_X satisfies $\chi(G_X) = \omega(G_X)$,

ii) the source graph G_U satisfies $\chi(G_U) = \omega(G_U)$.

Proof: We prove that the existence of a source-channel code implies the existence of separate source and channel codes. The analysis is most direct in terms of graph homomorphisms. We rely on the following string of inequalities: For any graph G,

$$\omega(G)^n = \omega(G^n) \leq \omega(G^{(n)}) \leq \vartheta(G)^n \leq \chi(G^n) \leq \chi(G^{(n)}) \leq \chi(G)^n. \tag{12}$$

Let (G_U, G_X) be an arbitrary source-channel pair such that G_X satisfies condition i of the theorem. We prove the theorem for the UI case. The proof for the RI case is similar. If a UI (m,n)-code exists, $G_U^{(m)} \to G_X^{(n)}$. If $a = \chi(G_X)$, (12) with $G = G_X$ implies $\chi(G_X^{(n)}) = \omega(G_X^{(n)}) = a^n$. Now $G_X^{(n)} \to K_{a^n}$, which implies $G_U^{(m)} \to K_{a^n}$ which implies $\chi(G_U^{(m)}) \leq a^n$ that is $\chi(G_U^{(m)}) \leq \omega(G_X^{(n)})$ or a separate code exists.

Now, we consider the case where G_X is arbitrary and G_U satisfies condition ii. We first note that in this case $R^*(G_U) = R_w(G_U)$ and the existence of separate codes in the UI case is equivalent to the existence of separate codes in the RI case. If a UI (m,n) code exists, $G_U^{(m)} \to G_X^{(n)}$. If $b = \chi(G_U)$, (12) with $G = G_U$ implies $\chi(G_U^{(m)}) = \omega(G_U^{(m)}) = b^m$. Now $K_{b^m} \to G_U^{(m)}$, which implies $K_{b^m} \to G_X^{(n)}$ which implies $\omega(G_X^{(n)}) \geq b^m$ that is $\omega(G_X^{(n)}) \geq \chi(G_U^{(m)})$ or a separate code exists. ■

The conditions of Theorem 3 are satisfied by an important class of graphs, called perfect graphs [17]. G is a perfect graphs if it satisfies $\chi(G') = \omega(G')$ for all induced subgraphs G' of G. Interval graphs and comparability graphs are examples of perfect graphs. Theorem 3 implies that source-channel separation is optimal for the following point-to-point communication scenarios:

a) *No side-information available at either end:* This case can be considered as a special case of the RI scenario where the confusability graph is the complete graph. Since complete graphs are perfect, source-channel separation is optimal as mentioned in Section I.

b) *Accurate source side-information available at both the encoder and the decoder:* This case is equivalent to the RI problem where both U and V need to be reconstructed at the decoder. The source confusability graph on the vertex set $\mathcal{U}^n \times \mathcal{V}^n$ is the vertex disjoint union of $|\mathcal{V}^n|$ complete graphs: to each element $v^n \in \mathcal{V}^n$ corresponds the clique $\{u^n \times v^n : u^n \in \mathcal{U}^n, p_{UV}(u^n, v^n) > 0\}$. Since the disjoint union of complete graphs is a perfect graph, source-channel separation is asymptotically optimal.

VI. Complexity of Scalar Code Design

Consider the following decision problem

Instance: Graphs G and H

Question: Is there a zero-error source-channel $(1,1)$-code from source G to channel H?

This problem is easily shown to be NP-complete, since the K-coloring problem [18] reduces to it when the channel graph H is the complete graph on K vertices. However in typical applications, the channel is fixed and the question that needs to be answered is whether some given source can be transmitted through this channel using a scalar source-channel code, i.e., the decision problem is:

Instance: Graph G

Question: Is there a zero-error source-channel $(1,1)$-code from source G to channel H?

This problem is much harder to classify. However, scalar code design is equivalent to finding a homomorphism from G to H. In this guise the above problem has been extensively studied by graph theorists. In 1990, Hell and Nešetřil [19] showed that deciding whether there is a graph homomorphism from a given G to a fixed H is polynomial if H is bipartite[3] and is NP-complete for all other H. Therefore, if the widely held conjecture $P \neq NP$ is true, no polynomial time optimal code design algorithm exists for most channels and we can only hope for efficient approximate algorithms.

VII. Summary and Conclusions

The main objective of this paper was to show the asymptotic suboptimality of separate source and channel coding for zero-error transmission of a discrete memoryless source through a discrete memoryless channel when there is source side-information solely at the decoder. For both UI and RI cases, we observed that not only is separate source and channel coding suboptimal, the gains from joint coding can be unbounded in the following sense: There exists a sequence $\{(G_{Ui}, G_{Xi})\}_{i=1}^{\infty}$ of source-channel pairs such that there exist joint source-channel codes of rate less than 1 for all i while the minimum rate for separate source and channel coding tends to ∞ as i tends to ∞. This is rather surprising, especially in view of the fact that if an asymptotically vanishing error is allowed, separate source and channel coding is optimal. The interesting, if challenging, problem that our result opens up, namely, finding the minimum rate necessary for zero-error joint source-channel coding is left for future work.

The convenience that separate coding affords led us to investigate conditions for optimality of separate coding. The sufficient conditions that we arrived at in Theorem 3 are identical to those for the achievability of channel capacity (condition i) or the Witsenhausen rate (condition ii) by scalar codes. All these results are a consequence of the basic inequality chain (12).

Finally, the equivalence between joint source-channel codes and graph homomorphisms led to the following code design complexity result: for all non-bipartite channel graphs, source-channel code design is NP-hard even for a fixed channel graph.

[3] A graph $G = (V, E)$ is a bipartite if there exist disjoint sets V_1 and V_2 such that $V = V_1 \cup V_2$ and $E \subseteq (V_1 \times V_2) \cup (V_2 \times V_1)$ that is the only edges in G have one end in V_1 and the other in V_2.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, Jul. and Oct. 1948.

[2] S. Vembu, S. Verdú,and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. on Information Theory*, vol. 41, no. 1, pp. 44–54, Jan. 1995.

[3] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. on Information Theory*, vol. 2, pp. 8–19, Sep. 1956.

[4] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Information Theory*, vol. 19, no. 4, pp. 471–80, Jul. 1973.

[5] S. Shamai (Shitz) and S. Verdú, "Capacity of channels with side information," *European Trans. on Telecommunications*, vol. 6, no. 5, pp. 587–600, Sep.-Oct. 1995

[6] N. Alon and A. Orlitsky, "Source coding and graph entropies,"*IEEE Trans. on Information Theory*, vol. 42, no. 5, pp. 1329–39, Sep. 1996

[7] J. H. .Van Lint and R. M. Wilson, *A course in combinatorics*, Cambridge Univ. Press, Cambridge, U. K., 1992.

[8] R. J. McEliece and E. C. Posner, "Hide and seek, data storage, and entropy," *Annals Math. Statist.*, vol. 42, no. 5, pp 1706–1716, 1971.

[9] C. Berge and M. Simonovits, "The coloring numbers of direct product of two hypergraphs," in C. Berge and D. Ray-Chaudhuri, Eds., *Hypergraph Seminar* (Lecture Notes on Mathematics, 411). New York: Springer Verlag, 1974.

[10] H. Witsenhausen, "The zero-error side information problem and chromatic numbers," *IEEE Trans. on Information Theory*, vol. 22, no. 5, pp. 592–593, Sep. 1976.

[11] L. Lovász, " On the Shannon capacity of a graph," *IEEE Trans. on Information Theory*, vol. 25, no. 1, pp. 1–7, Jan. 1979.

[12] K. Marton, "On the Shannon capacity of probabilistic graphs," *J. of Comb. Theory*, Series B, vol. 57, no. 2, pp. 183–195, Mar. 1993.

[13] D. E. Knuth, "The sandwich theorem," *Electron. J. Combin.* 1:Article 1, approx. 48 pp. (electronic), 1994.

[14] W. Haemers, "On some problems of Lovász concerning the Shannon capacity of a graph," *IEEE Trans. on Information Theory*, vol. 25, no. 2, pp. 231–232, Mar. 1979.

[15] N. Alon, "The Shannon capacity of a union," *Combinatorica*, vol. 18, no. 3, pp. 301–310, 1998.

[16] M. N. Huxley, *The distribution of prime numbers*, Oxford University Press, Oxford, 1972.

[17] J. L. Ramírez Alfonsín and B. A. Reed eds., *Perfect graphs*, John Wiley and Sons, West Sussex, England, 2001.

[18] M. R. Garey and D. S. Johnson *Computers and intractability: A guide to the theory of NP-completeness*, Freeman, San Francisco, 1979.

[19] P. Hell and J. Nešetřil, "On the complexity of H-coloring," *J. of Combinatorial Theory*, Series B, vol. 48, pp. 92–110, 1990.