[21] N. K. Vereshchagin and P. M. B. Vitányi, "Kolmogorov's structure functions and model selection," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3265–3290, Dec. 2004.
[22] ——, "Arithmetic rate-distortion theory," *IEEE Trans. Inf. Theory*, submitted for publication.
[23] P. M. B. Vitányi and M. Li, "Minimum description length induction, Bayesianism, and Kolmogorov complexity," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 446–464, Mar. 2000.
[24] V. V. V'yugin, "On the defect of randomness of a finite object with respect to measures with given complexity bounds," *SIAM Theory Probab. Appl.*, vol. 32, no. 3, pp. 508–512, 1987.
[25] ——, "Algorithmic complexity and stochastic properties of finite binary sequences," *Comput. J.*, vol. 42, no. 4, pp. 294–317, 1999.

Zero-Error Source–Channel Coding With Side Information

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

***Abstract*—This correspondence presents a novel application of the theta function defined by Lovász. The problem of coding for transmission of a source through a channel without error when the receiver has side information about the source is analyzed. Using properties of the Lovász theta function, it is shown 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, it is further shown that the joint coding gain can in fact be unbounded. Since separate coding simplifies code design and use, conditions on sources and channels for the optimality of separate coding are also derived.**

***Index Terms*—Graph homomorphisms, Lovász theta function, source–channel separation, zero-error coding.**

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 nonzero 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 single-user source–channel scenarios with a few exceptions under the asymptotically vanishing error constraint described previously [2]. Note that separation theorems are asymptotic results and make no claims about the behavior at finite block lengths.

Manuscript received March 3, 2005; revised April 26, 2006. This work was supported in part by the National Science Foundation under Grant EIA-0080134, the University of California MICRO Program, Applied Signal Technology, Inc., Dolby Laboratories, Inc., Mindspeed Technologies, Inc., and Qualcomm, Inc. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Yokohama, Japan, June/July 2003.

J. Nayak is with INRIA/IRISA, Campus Universitaire de Beaulieu, 35042 Rennes, France (e-mail: jnayak@irisa.fr).
E. Tuncel is with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA (e-mail: ertem@ee.ucr.edu).
K. Rose is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: rose@ece.ucsb.edu).

Communicated by R. J. McEliece, Associate Editor for Coding Theory.
Digital Object Identifier 10.1109/TIT.2006.881718

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. 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 Section II.

Let $(\mathcal{S}_U, \mathcal{S}_V)$ be a pair of memoryless correlated sources producing realizations of 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 $\mathcal{C}$. Alice employs (m, n) codes that map m realizations of U to n-length blocks of the channel input alphabet in order to noiselessly convey U. We wish to determine the minimum amount of channel resources required for Alice to convey U to Bob. We quantify the efficiency of a code by its rate $\frac{n}{m}$ channel uses per source symbol.

Suppose we wish to design a source–channel code for the source U with side information V and channel $\mathcal{C}$. The celebrated results of Shannon [1] and Slepian and Wolf [4] imply that communication is possible using separate source and channel codes if the rate is at least $\frac{H(U|V)}{C}$. 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 correspondence, we focus on the *zero-error* setting for the problem of source–channel coding with side information. Section III presents our main results—the suboptimality of separate coding and the gains by joint coding. Our main tool in analyzing these problems is the theta function, a graph functional shown by Lovász to be an upper bound on the Shannon capacity of a graph [6]. Lovász employed the theta function to characterize the Shannon capacity of the pentagon graph, a problem that had remained open for more than two decades. To quantify the gains, we employ a graph construction by Alon that was used by him to disprove a conjecture of Shannon regarding the additivity of zero-error capacity with respect to channel sums [7]. In Section IV, we turn to the question of when separate coding is indeed optimal and present sufficient conditions on sources and channels. In Section V, we present some comments on the complexity of code design before concluding in Section VI. Since results in zero-error coding for the source and channel that we consider are not widely known, we first survey relevant aspects of this area in Section II.

II. Preliminaries and Notation

The imposition of zero-error constraints naturally leads to problem formulations in terms of graphs and we begin this section with some useful graph-theoretic definitions.

A. Graph-Theoretic Preliminaries

An (undirected) graph $G = (V, E)$ is defined by a vertex set $V = \{v_i, i = 1, \dots, |V|\}$ and a set of edges $E \subseteq \binom{V}{2}$. The complement of a graph $G = (V, E)$ is the graph $\bar{G} = (V, \bar{E})$ where $\bar{E} = \binom{V}{2} \setminus E$.

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. 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 complete subgraph of G. The clique number $\omega(G)$ is the cardinality of the largest clique in G. An independent set of G is a subgraph whose complement is a clique.

We now define the various products for a pair of graphs $G = (V, E)$ and $H = (V', E')$. A pair of distinct vertices (v_1, v_1') and (v_2, v_2') from $V \times V'$ form an edge in the OR product $G \circ H$ if either $(v_1, v_2) \in E$ or $(v_1', v_2') \in E'$. The AND product of the two graphs, denoted $G \times H$ is defined as the complement of $\bar{G} \circ \bar{H}$. The n-fold OR and AND products of a graph G with itself are denoted $G^{(n)}$ and G^n, respectively.

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 $\{\phi(v_1), \phi(v_2)\} \in E'$. 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 $1 \leq m \leq \omega(G)$.

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 fanout 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 fanout 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

$$\psi_c^n(Y^n(\phi_c^n(i))) = i, \qquad \forall i \in \{1, \dots, 2^{nC}\}$$

with probability 1. For a scalar (block length 1) code to be zero-error, the fanout 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 and $\log \omega(G_X)$ bits[1] can be transmitted in one channel use. 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 and only 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 Shannon 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 $\omega(G_X^{(n)})$ (using Fekete's lemma [8]).

For example, the n-fold OR product of the complete graph K_m, $K_m^{(n)}$, is K_{m^n}. Therefore, $C(K_m) = \log m$. Another interesting example 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 called the theta function, Lovász [6] showed that $\frac{1}{2}\log 5$ is in fact the capacity of the C_5. Lovász's theta function gives 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 unconnected 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 correspondence are the following [6], [9]. For graphs G, H

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

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

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

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 according to the joint distribution $p_{UV}(u, v)$. 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 encoder design and, hence, minimum asymptotic rate calculation, we can 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 $p_{UV}(u', v) > 0$.

A source code of block length m consists of an encoder $\phi_s^m : \mathcal{U}^m \to \{1, \dots, 2^{mR}\}$ and a decoder $\psi_s^m : \{1, \dots, 2^{mR}\} \times \mathcal{V}^m \to \mathcal{U}^m$. A zero-error code is one where $\psi_s^m(\phi_s^m(U^m), V^m) = U^m$ with probability 1. For scalar coding, only independent sets in G_U can be assigned the same codeword and the minimum rate of a scalar code is $\chi(G_U)$. While encoding a block of length m, two realizations of U^m, $\boldsymbol{u} = (u_1, \dots, u_m)$ and $\boldsymbol{u}' = (u_1', \dots, u_m')$, are confusable given the side information if they are confusable in every coordinate. Therefore, the confusability graph for m instances is the m-fold AND power of G_U, G_U^m. The asymptotic rate (in bits per source symbol), called the Witsenhausen rate of a graph, is given by [10]

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

The limit in (6) exists due to the submultiplicativity of $\chi(G_U^m)$.

K_m^n is also K_{m^n} and, hence, $R_{\mathrm{w}}(K_m)$ is $\log m$ as well. For the pentagon example, the results of Lovász [6] on its capacity combined with the results of Witsenhausen [10] imply that $R_{\mathrm{w}}(C_5) = \frac{1}{2}\log 5$.

We note in passing that computation of Shannon 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 $\phi_{sc}^{(m,n)} : \mathcal{U}^m \to \mathcal{X}^n$ and the decoder $\psi_{sc}^{(m,n)} : \mathcal{Y}^n \times \mathcal{V}^m \to \mathcal{U}^m$. Using the side information and the output of the encoder, the decoder produces $\psi_{sc}^{(m,n)}(Y^n(\phi_{sc}^{(m,n)}(U^m)), V^m) = \hat{U}^m$. Of a zero-error

[1] All logarithms are to base 2.

code, we require $\hat{U}_i = U_i$ with probability 1, $\forall i$. 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 cases where the source and channel graphs are identical. By doing so, we ensure that a joint code at rate 1, namely the identity mapping, always exists. To show the suboptimality of separate coding, we present cases where joint rate 1 codes exist at block length 1, but no separate rate–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 $R \geq R_{\mathrm{w}}(G_U)$ and $C \leq C(G_X)$. Therefore, for a separate rate-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

An (m,n)-code is a homomorphism $\phi_{sc}^{(m,n)}: \mathcal{U}^m \to \mathcal{X}^n$ 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 $R_{\mathrm{w}}(G_U)$. Therefore, separate rate-1 codes cannot exist if $C(G_X) < R_{\mathrm{w}}(G_U)$.

Our approach here is to find an appropriate graph G and let $G_U = G_X = G$. We employ the result that the logarithm of the Lovász theta function $\vartheta(\bar{G})$ is sandwiched between the Shannon capacity of G and the Witsenhausen rate of G. The fact that $\log \vartheta(\bar{G})$ is a lower bound on $R_{\mathrm{w}}(G)$ is implicit in a paper by Marton [11]. Our next lemma makes this relation explicit.

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

$$\log \vartheta(\bar{G}) \leq R_{\mathrm{w}}(G) \tag{7}$$

Proof:

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

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

$$\log \vartheta(\bar{G}) \leq \lim_{n\to\infty} \frac{1}{n} \log(\chi(G^n)) = R_{\mathrm{w}}(G). \qquad \square$$

Haemers [12] presents an example where $C(G) < \log \vartheta(\bar{G})$—the Schläfli graph. The Schläfli graph G_{27} is a strongly regular graph on 27 vertices. Let both the source and channel graphs G_U and G_X be G_{27}. Haemers proved that $C(G_X) \leq \log 7 < \log \vartheta(\overline{G_X}) = \log 9$. Therefore, by Lemma 1, $C(G_X) < R_{\mathrm{w}}(G_U)$, although transmission is possible using scalar joint source–channel coding.

B. 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 is $r_{\mathrm{w}}(G_U, G_X) = \frac{R_{\mathrm{w}}(G_U)}{C(G_X)}$. 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 one channel use per source symbol. Using a recent result by Alon [7], we show in Corollary 1 that $r_{\mathrm{w}}(G_U, G_X)$ can be *arbitrarily large* even in this restricted scenario. Hence, the joint coding gain is generally unbounded.

Lemma 2: For every k, there exists a graph G such that $C(G) < \log k$ and

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

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

A formal proof of Lemma 2 is omitted since it uses the same construction as in the proof of Theorem 1.1 in [7]. For every k, that construction, which is based on algebraic and number-theoretic arguments yields a graph H on n vertices such that $C(H)$ and $C(\bar{H})$ are both less than $\log k$ while $n \geq k^{(1+o(1))\frac{\log k}{4\log\log k}}$. Since for any graph $F = (V, E), \omega(F \circ \bar{F}) \geq |V|$, the standard properties of the θ function can be used to show that at least one of the graphs H and $\bar{H}$ satisfies (8).

Corollary 1: Given any l, we can find a graph G such that

$$\frac{R_{\mathrm{w}}(G)}{C(G)} \geq l.$$

Proof: Fix k and let G be as in Lemma 2. Then

$$\frac{R_{\mathrm{w}}(G)}{C(G)} \overset{(\mathrm{a})}{\geq} \frac{\log \vartheta(\bar{G})}{C(G)} \overset{(\mathrm{b})}{\geq} \frac{(1+o(1))\log k}{8 \log\log k}$$

where (a) follows from Lemma 1, and (b) follows from Lemma 2. Examining the right-hand side of (b), we see that we can always find a k that makes it greater than the given l. $\square$

IV. 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. The following proposition specifies conditions on graphs that are sufficient for optimality of source–channel separation.

Proposition 1: Asymptotic optimality is achievable by separate 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

$$\begin{aligned}\omega(G)^n &= \omega(G^n) \leq \omega(G^{(n)}) \leq \vartheta(\bar{G})^n\\ &\leq \chi(G^n) \leq \chi(G^{(n)}) \leq \chi(G)^n.\end{aligned} \tag{9}$$

Let (G_U, G_X) be an arbitrary source–channel pair such that G_X satisfies condition i) of the theorem. If an (m,n)-code exists, $G_U^m \to G_X^{(n)}$.

If $a = \chi(G_X)$, (9) 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 $G_U^m \to K_{a^n}$. Therefore, $\chi(G_U^{(m)}) \le a^n$ that is, $\chi(G_U^m) \le \omega(G_X^{(n)})$ implying that a separate code exists.

Now, we consider the case where G_X is arbitrary and G_U satisfies condition ii). If an (m, n) code exists, $G_U^m \to G_X^{(n)}$. If $b = \chi(G_U)$, (9) 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 $\omega(G_X^{(n)}) \ge b^m$ that is, $\omega(G_X^{(n)}) \ge \chi(G_U^m)$ implying that a separate code exists. □

The conditions of Proposition 1 are satisfied by an important class of graphs, called perfect graphs [13]. Proposition 1 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 source coding with decoder side-information 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) *Source side information available at both the encoder and the decoder:* This case is equivalent to the case where the source that Alice needs to communicate to Bob is the pair $(\mathcal{S}_\mathcal{U}, \mathcal{S}_\mathcal{V})$ and Bob has access to the side information $\mathcal{S}_\mathcal{V}$. 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.

If both the conditions in Proposition 1 are satisfied, then not only is separate coding optimal, but separate scalar coding (possibly followed by a mapping between blocks of the intermediate indices) achieves the optimal rate.

V. 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 [14] 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 over this channel using a scalar source–channel code, i.e., the decision problem is the following.

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 [15] showed that deciding whether there is a graph homomorphism from a given G to a fixed H is polynomial if H is bipartite[2] 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.

[2]A graph $G = (V, E)$ is bipartite if there exist disjoint sets V_1 and V_2 such that $V = V_1 \cup V_2$ and $E \subseteq \{\{v_1, v_2\} : v_1 \in V_1, v_2 \in V_2\}$, that is, all edges in G have one end in V_1 and the other in V_2. Another characteristic of bipartite graphs is that their chromatic number is at most 2.

VI. Summary and Conclusion

The main objective of this correspondence 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. 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 surprising in two respects: 1) separate source and channel coding is optimal in most other point-to-point communication scenarios that have been studied; 2) for the very same setup as the one we considered in this correspondence, if an asymptotically vanishing error is allowed, separate source and channel coding is again 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 Proposition 1 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 (9). Finally, the equivalence between joint source–channel codes and graph homomorphisms led to the following code design complexity result: for all nonbipartite channel graphs, scalar source–channel code design is NP-hard even for a fixed channel graph.

References

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.

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

[3] C. E. Shannon, “The zero error capacity of a noisy channel,” *IRE Trans. Inf. Theory*, vol. PGIT-2, no. 3, pp. 8–19, Sep. 1956.

[4] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[5] S. Shamai (Shitz) and S. Verdú, “Capacity of channels with side information,” *Europ. Trans. Telecommun.*, vol. 6, no. 5, pp. 587–600, Sep./Oct. 1995.

[6] L. Lovász, “On the Shannon capacity of a graph,” *IEEE Trans. Inf. Theory*, vol. IT-25, no. 1, pp. 1–7, Jan. 1979.

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

[8] J. H. van Lint and R. M. Wilson, *A Course in Combinatorics*. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[9] D. E. Knuth, “The sandwich theorem,” *Electron. J. Combin.*, vol. 1, 1994.

[10] H. S. Witsenhausen, “The zero-error side information problem and chromatic numbers,” *IEEE Trans. Inf. Theory*, vol. IT-22, no. 6, pp. 592–593, Nov. 1976.

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

[12] W. Haemers, “On some problems of Lovász concerning the Shannon capacity of a graph,” *IEEE Trans. Inf. Theory*, vol. IT-25, no. 2, pp. 231–232, Mar. 1979.

[13] J. L. R. Alfonsín and B. A. Reed, Eds., *Perfect Graphs*. West Sussex, England: Wiley, 2001.

[14] M. R. Garey and D. S. Johnson, *Computers and Intractability*. San Francisco, CA: Freeman, 1979.

[15] P. Hell and J. Nešetřil, “On the complexity of H-coloring,” *J. Comb. Theory, Ser. B*, vol. 48, pp. 92–110, 1990.