Graph Capacities and Zero-Error Transmission Over Compound Channels

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

Abstract—This correspondence investigates the behavior of the compound channel under a zero-error constraint. We derive expressions for the capacity when a) neither the encoder nor decoder has side-information about the channel; b) when only the encoder has such side-information. These expressions are given in terms of capacities of appropriately defined sets of graphs. We clarify that an earlier treatment of the zero-error capacity of a compound channel corresponds to the case where the decoder has side-information about the channel. We also characterize the minimum asymptotic rate for the source coding dual of the compound channel problem, namely, source coding with compound side information. Finally, we contrast the zero-error and asymptotically vanishing error capacities of the compound channel.

Index Terms—Compound channel, graphs, side-information, Sperner capacity, Witsenhausen rate, zero-error.

I. INTRODUCTION

The compound channel is one of the simplest generalizations of the discrete memoryless channel (DMC). It is defined in terms of a set of DMCs that share their input and output alphabets. An element of this set (i.e., one of the available DMCs) is selected at random before transmission begins and this choice characterizes the channel behavior for the duration of the transmission. Both encoder and decoder are ignorant of the choice of DMC. The conventional (where an asymptotically vanishing probability of error is allowed) capacity of this channel was derived independently by Dobrushin [1], Wolfowitz [2], and Blackwell *et al.* [3] around 1960. The zero-error version of the problem, however, was only studied about three decades later.

The motivation for examining the zero-error scenario arose from certain problems in asymptotic combinatorics. Cohen *et al.* [4] showed that compound channels provided a convenient common framework for analyzing several disparate problems. By defining sets of undirected graphs appropriately, each problem could be reduced to that of finding, what Cohen *et al.* termed, the Shannon capacity of a set of graphs. They obtained an upper bound on this capacity and showed that it is tight in some cases. In a remarkable paper [5], Gargano *et al.* showed that this bound is in fact always tight. Both these papers presented an information theoretic interpretation of the Shannon capacity of a set of graphs: the zero-error capacity of a compound channel with an informed decoder[1] is the Shannon capacity of a set of graphs associated with the channel. In the follow-up study reported here, we provide expressions for the capacity in the remaining cases of interest: a) uninformed decoder and encoder; b) informed encoder and uninformed decoder.

[1]Note that [4], [5] do not make explicit the distinction between the case of the informed decoder versus the uninformed decoder. Therefore, for completeness we shall discuss the informed decoder case as well in Section IV.

In Section III, we show that the compound channel capacity is the Sperner capacity of a set of *directed* graphs. The Sperner capacity was defined in [6] as a formal generalization of the Shannon capacity to directed graphs and was not related to information transmission. In Section IV, we consider cases where either the encoder or the decoder knows the choice of DMC. In Section V, we derive the minimum asymptotic rate for the source coding problem with compound side-information at the decoder, which is the dual of the compound channel coding problem. We conclude with a comparative discussion of the asymptotically vanishing error and zero-error capacities of the compound channel in Section VI.

II. PRELIMINARIES AND NOTATION

Our analysis is grounded in graph theory and employs quantities to be defined below. All sets that arise are implicitly assumed to be finite.

A. Shannon Capacity of a Graph

A graph $G = (V, E)$ consists of a vertex set $V = \{v_i, i = 1, \ldots, |V|\}$ and a set of edges $E \subseteq V \times V$. Unless stated otherwise, all graphs are assumed to be symmetrically directed, that is $(v, v') \in E \Rightarrow (v', v) \in E$ and $(v, v) \notin E$. A symmetrically directed graph $G = (V, E)$ is for our purposes equivalent to an undirected graph $G_u = (V, E_u)$, where $\{v, v'\} \in E_u \subset \binom{V}{2} \Leftrightarrow (v, v') \in E$. Therefore, in the sequel, we shall use the terms undirected graph and symmetrically directed graph interchangeably. Given a graph G, we say that two vectors $x^n, x'^n \in V^n$ are connected with respect to G if at some coordinate $i, (x_i, x'_i) \in E$. Let $N(G, n)$ denote the size of the largest subset of V^n in which the elements are pairwise connected with respect to G. The Shannon capacity of the graph is then defined to be

$$C(G) = \lim_{n \to \infty} \frac{1}{n} \log N(G, n). \qquad (1)$$

The existence of the limit follows from the supermultiplicativity of $N(G, n)$ and Fekete's Lemma [7].

This quantity was defined by Shannon[2] [8] to characterize the zero-error capacity of a noisy channel. For a given DMC $\mathcal{C}$ with transition probability function $p(y \mid x), x \in \mathcal{X}, y \in \mathcal{Y}$, we define the fan-out sets $F(x) = \{y \in \mathcal{Y} : p(y \mid x) > 0\}$. Using these fan-out sets we can define the characteristic graph of the channel as the graph $G_X = (V_X, E_X)$ with vertex set $V_X = \mathcal{X}$ where $(x, x') \in E_X$ if and only if $x \neq x'$ and $F(x) \cap F(x') = \emptyset$.

It is apparent from the definition of the fan-out sets that the elements of a scalar zero-error channel code have nonintersecting fan-out sets, that is, every pair of elements from a scalar zero-error code forms an edge in the characteristic graph. Since the channel is memoryless, the fan-out set of an $\mathcal{X}$-block of length n is the Cartesian product of the fan-out sets at each coordinate. So two blocks x^n and x'^n can both belong to a zero-error code if and only if at some coordinate their fan-out sets do not intersect; in other words x^n and x'^n must be connected with respect to G_X. Therefore the zero-error capacity of a channel is the Shannon capacity of its characteristic graph.

$$C^0(\mathcal{C}) = C(G_X).$$

Computation of the zero-error capacity for general graphs remains an open problem despite the half century that elapsed since its definition.

[2]In Shannon's definition, the right hand side in (1) would correspond to the capacity of the complement of the undirected graph corresponding to G. Similarly, the characteristic graph defined by Shannon is the complement of the undirected graph corresponding to the characteristic graph that we define later.

The concept of Shannon capacity was generalized to graphs that have a probability distribution defined on their nodes in [9]. For a graph $G = (V, E)$ and a probability distribution P on its vertices, denote by $N(G, P, n, \epsilon)$ the size of the largest subset of the ϵ-typical set $A_n^\epsilon(P)$ whose elements are pairwise connected. $A_n^\epsilon(P)$ is the set of sequences in $\mathcal{X}^n$ whose empirical distribution is close to P

$$A_n^\epsilon(P) = \left\{ x^n : \forall a \in \mathcal{X}, \left| \frac{1}{n} N(a \mid x^n) - p(a) \right| \leq \epsilon \right.$$
$$\left. p(a) = 0 \Rightarrow N(a \mid x^n) = 0 \right\}$$

where $N(a \mid x^n)$ counts the number of times letter a appears in $x^n \in \mathcal{X}^n$. The Shannon capacity of G within type P is

$$C(G, P) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log N(G, P, n, \epsilon). \tag{2}$$

For a channel with characteristic graph G_X, $C(G_X, P)$ is the capacity of the channel under the restriction that the type of every element of the codebook is approximately P. This interpretation along with results on the number of possible empirical distributions of sequences of a given length implies that [5]

$$C(G) = \max_P C(G, P).$$

Another concept that is relevant to this correspondence is the Shannon capacity of a *set of graphs* $\mathcal{G} = \{G_s, s \in \mathcal{S}\}$ that have a common vertex set V. If $N(\mathcal{G}, n)$ is the size of the largest set in V^n that is connected with respect to every graph in $\mathcal{G}$, the Shannon capacity of $\mathcal{G}$ is

$$C(\mathcal{G}) = \lim_{n \to \infty} \frac{1}{n} \log N(\mathcal{G}, n). \tag{3}$$

In [4], the following simple upper bound was presented

$$C(\mathcal{G}) \leq \max_P \min_{G \in \mathcal{G}} C(G, P).$$

Gargano *et al.* [5] showed that this bound is tight.

B. Sperner Capacity of Directed Graphs

Motivated by some open questions in extremal set theory, Gargano *et al.* [5] generalized the concept of capacity to directed graphs. A directed graph $G = (V, E)$ consists of a vertex set V and an edge set E that is an arbitrary subset of $V \times V$. Note that self loops are allowed and that both (a, b) and (b, a) may be simultaneously included in the edge set E. Given a directed graph $G = (V, E)$, two vectors $x^n, x'^n \in V^n$ are called *incomparable* with respect to G, if there exist coordinates i, i' such that $(x_i, x_i') \in E$ and $(x_{i'}', x_{i'}) \in E$, that is, there is an edge from sequence x^n to x'^n at coordinate i and there is an edge from sequence x'^n to x^n at coordinate i'. Incomparability in directed graphs corresponds to connectedness in undirected graphs. Indeed, in the special case where the directed graph degenerates to an undirected graph, the two concepts are equivalent. Hence, if we define $N_d(G, n), N_d(G, P, n, \epsilon)$ and $N_d(\mathcal{G}, n)$ similarly to $N(G, n), N(G, P, n, \epsilon)$ and $N(\mathcal{G}, n)$ in II.A by replacing connectedness with incomparability, we can define the Sperner capacity of a graph

$$\Sigma(G) = \lim_{n \to \infty} \frac{1}{n} \log N_d(G, n) \tag{4}$$

the Sperner capacity within a type

$$\Sigma(G, P) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log N_d(G, P, n, \epsilon) \tag{5}$$

and the Sperner capacity of a set of graphs

$$\Sigma(\mathcal{G}) = \lim_{n \to \infty} \frac{1}{n} \log N_d(\mathcal{G}, n). \tag{6}$$

In [5], it was also shown that

$$\Sigma(G) = \max_P \Sigma(G, P)$$
$$\Sigma(\mathcal{G}) = \max_P \min_{G \in \mathcal{G}} \Sigma(G, P).$$

All the above Sperner capacities reduce to the corresponding Shannon capacities for undirected graphs. In Section III we will show that the capacity of the compound channel is the Sperner capacity of a certain set of graphs (see [10] for a different information theoretic interpretation of Sperner capacity).

III. CAPACITY OF THE COMPOUND CHANNEL

A compound channel $\mathfrak{C}$ is a set of DMCs $\{p_s(y \mid x): x \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S}\}$, where $\mathcal{X}, \mathcal{Y}$ and $\mathcal{S}$, which are the channel input, output, and index alphabet, respectively, are finite sets. When a codeword is input into this channel, the compound channel behaves like one of its constituent channels. The transition probability distribution remains constant for the duration of the codeword. Let $F_s(x)$ denote the fan-out set of $x \in \mathcal{X}$ with respect to channel s. The fan-out set of $x^n \in \mathcal{X}^n$ with respect to a channel s, denoted $F_s(x^n)$ is the Cartesian product of the fan-out sets at each coordinate of x^n. The fan-out set of x^n with respect to $\mathfrak{C}$ is defined as $F_{\mathfrak{C}}(x^n) = \cup_{s \in \mathcal{S}} F_s(x^n)$. $F_{\mathfrak{C}}(x^n)$ is the set of all possible channel outputs when x^n is the input to the channel. A zero-error n-code for $\mathfrak{C}$ is a set $\{x^n(j) \in \mathcal{X}^n, j = 1 \ldots N\}$ such that $F_{\mathfrak{C}}(x^n(j)) \cap F_{\mathfrak{C}}(x^n(j')) = \emptyset, \forall j \neq j'$. Note that the code does not depend on the constituent channel actually in operation. N is the number of messages that can be sent using this code. If the message to be sent is $j \in \{1, \ldots, N\}$, the encoder chooses $x^n(j)$ as the input to the channel. If the channel output belongs to the corresponding decoding region $F_{\mathfrak{C}}(x^n(j))$, the decoder declares that message j was sent. The conditions on the code guarantee that the probability of decoding error is zero. If $N(\mathfrak{C}, n)$ denotes the size of the largest n-code, the zero-error capacity of the compound channel is $C^0(\mathfrak{C}) = \lim_{n \to \infty} \frac{1}{n} \log N(\mathfrak{C}, n)$. It is the maximum number of bits that can be reliably transmitted in one channel use.

The condition that the fan-out sets of x^n and x'^n with respect to $\mathfrak{C}$ should not intersect can be rewritten as

$$F_{\mathfrak{C}}(x^n) \cap F_{\mathfrak{C}}(x'^n) = [\cup_{s \in \mathcal{S}} F_s(x^n)] \cap [\cup_{s \in \mathcal{S}} F_s(x'^n)]$$
$$= \cup_{s \in \mathcal{S}} \cup_{s' \in \mathcal{S}} [F_s(x^n) \cap F_{s'}(x'^n)]$$
$$= \emptyset.$$

This condition is equivalent to the set of conditions

$$\forall s, \quad s' \in \mathcal{S}: [F_s(x^n) \cap F_{s'}(x'^n)] = \emptyset$$

and leads us to our main result, which is described here.

To develop an expression for the capacity of the compound channel $\mathfrak{C}$, we define the characteristic set of (directed) graphs $\mathcal{G}(\mathfrak{C}) = \{G_{ss'}, s, s' \in \mathcal{S}\}$: For every pair $s, s' \in \mathcal{S}$, $G_{ss'}$ is the directed graph on $\mathcal{X}$ such that $(x, x') \in E_{ss'} \Leftrightarrow F_s(x) \cap F_{s'}(x') = \emptyset$. $G_{ss'}$ tells us whether we can distinguish two letters when the first letter is sent through channel s and the second through channel s'. G_{ss}

is the symmetrically directed graph corresponding to the characteristic graph G_s of channel s. Also $G_{ss'}$ and $G_{s's}$ can be obtained from each other by reversing edge directions.

Just as every compound channel corresponds to a set of graphs, every set of graphs $\mathcal{G} = \{G_{ss'}, s, s' \in \mathcal{S}\}$, where the $G_{ss}, s \in \mathcal{S}$ are undirected and $G_{ss'}$ and $G_{s's}$ can be obtained from each other by reversing edge directions, is equivalent to some compound channel. One way to obtain the compound channel from the set of graphs is to define the fan-out sets as $F_s(x) \triangleq \{\{(s, x), (s', x')\}: (x, x') \notin G_{ss'}\}$. The output alphabet is $\cup_s \cup_x F_s(x)$.

Theorem 1: The zero-error capacity of the compound channel $\mathfrak{C}$ is

$$C^0(\mathfrak{C}) = \Sigma(\mathcal{G}(\mathfrak{C})). \tag{7}$$

Proof: Let x^n and x'^n be distinct elements of a block length n codebook $\mathcal{C}$ for the compound channel. Since the decoding regions are fixed, for every pair $s, s' \in \mathcal{S}$, the fan-out set of x^n with respect to channel s is disjoint from the fan-out set of x'^n with respect to channel s', that is, there exists $1 \leq i \leq n$ such that $F_s(x_i) \cap F_{s'}(x_i') = \emptyset$. Swapping the roles of s and s', there exists $1 \leq i' \leq n$ such that $F_s(x_{i'}') \cap F_{s'}(x_{i'}) = \emptyset$. This means that the two sequences are incomparable with respect to $G_{ss'}$. Since this holds for all s, s', a codebook is an incomparable set with respect to $\mathcal{G}(\mathfrak{C})$. Conversely, we can show that every incomparable set is a valid zero-error code for the compound channel. It follows by standard asymptotic arguments that the capacity of the compound channel is the Sperner capacity of the characteristic set of graphs. $\square$

IV. COMPOUND CHANNEL WITH SIDE-INFORMATION

We now study the effect of knowledge of the choice of the constituent DMC at either the encoder or the decoder on the capacity of the compound channel. We evaluate an example in Section IV.C to contrast the various capacities.

A. Capacity of the Compound Channel With an Informed Encoder

When the encoder, but not the decoder, knows which of the constituent channels will be in operation, the codewords can depend on the channel, but the decoding regions cannot. The following lemma leads to an expression for the capacity.

Lemma 1: $\min_{i \in \mathcal{I}} \Sigma(G_i) > 0$ if and only if

$$\max_P \min_{i \in \mathcal{I}} \Sigma(G_i, P) > 0.$$

Proof: The "if" part follows from basic min-max results

$$\min_{i \in \mathcal{I}} \Sigma(G_i) = \min_{i \in \mathcal{I}} \max_P \Sigma(G_i, P) \geq \max_P \min_{i \in \mathcal{I}} \Sigma(G_i, P).$$

To show the converse, we prove that

$$\max_P \min_{i \in \mathcal{I}} \Sigma(G_i, P) = 0 \Rightarrow \min_{i \in \mathcal{I}} \Sigma(G_i) = 0.$$

If $\max_P \min_{i \in \mathcal{I}} \Sigma(G_i, P) = 0, \min_{i \in \mathcal{I}} \Sigma(G_i, P) = 0$ for every P, in particular when P is the uniform distribution over the vertices. Let G be the graph that achieves the minimum for the uniform distribution. Since the minimum is zero, for all n, no pair of n-vectors is incomparable with respect to G. Without loss of generality assume that $\mathcal{X}$ is of the form $\{1, 2, \ldots, k\}$, where $k = |\mathcal{X}|$. For $n = k^2$, consider the following sequences in $\mathcal{X}^n: x^n = 11 \ldots 122 \ldots 2 \ldots kk \ldots k$, where each letter is repeated k times and $x'^n = 123 \ldots k123 \ldots k \ldots 123 \ldots k$, where the alphabet is repeated k times. Both sequences belong to the typical set for the uniform distribution on $\mathcal{X}$. Let $c, d \in \mathcal{X}$ be a pair of distinct vertices.

Both (c, d) and (d, c) appear as (x_i, x_i') and $(x_{i'}, x_{i'}')$ at certain coordinates i, i' in the above sequences. If (c, d) were an edge in G, the two sequences would be incomparable and violate the hypothesis. Moreover, G has no self-loops since that would make the sequences incomparable as well. Therefore, G is an edge-free graph. Since distinct sequences are never incomparable with respect to such a graph $\min_{i \in \mathcal{I}} \Sigma(G_i) = 0$. $\square$

The above lemma is used in the next theorem, which specifies the capacity.

Theorem 2: The capacity of a compound channel with an informed encoder (i.e., whose encoder knows which constituent DMC is in use) is

$$C_{\text{enc}}^0(\mathfrak{C}) = \begin{cases} 0, & \text{if } C^0(\mathfrak{C}) = 0 \\ \min_{s \in \mathcal{S}} C(G_s), & \text{otherwise.} \end{cases} \tag{8}$$

Proof: G_s denotes the (undirected) characteristic graph of channel s. Consider the case where $C^0(\mathfrak{C}) = 0$. Let $\{C_s = \{x^n(s, j): j = 1, \ldots, N\}, s \in \mathcal{S}\}$ be the set of codebooks, one for each possible DMC. The zero-error constraint requires that the decoding region for message j (common for all $s \in \mathcal{S}$) be $\cup_{s \in \mathcal{S}} F_s(x^n(s, j))$ and that the decoding regions be mutually exclusive. Hence, $F_s(x^n(s, j)) \cap F_{s'}(x^n(s', j')) = \emptyset$ for all $1 \leq j \neq j' \leq N$ and $s, s' \in \mathcal{S}$.

Since $C^0(\mathfrak{C}) = 0$, the Proof of lemma 1 above implies that one of the graphs in $\mathcal{G}(\mathfrak{C})$, say $G_{ss'}$, is an edge-free graph. This implies that for all $j, j', F_s(x^n(s, j)) \cap F_{s'}(x^n(s', j'))$ is nonempty. So the channel code cannot contain more than one codeword, and the capacity is zero.

When $C^0(\mathfrak{C}) > 0$, the encoding can be done in two steps. Using a code for the compound channel (that does not exploit side-information), the encoder conveys its knowledge about the channel to the decoder. Since the set of DMCs is finite, this procedure requires a constant number of channel uses. After the first step, both encoder and decoder know the channel in operation. Hence they can transmit at least at the rate $\min_s C(G_s)$. Since this is also the maximum rate at which zero-error transmission can be guaranteed (it cannot be exceeded when the channel that minimizes $C(G_s)$ is the channel in operation), the capacity of the compound channel is as given above. $\square$

B. Capacity of the Compound Channel With an Informed Decoder

When the decoder knows the channel in operation, the code is a set $\{x^n(j), j = 1 \ldots N\}$. The decoding region for message j when s is the channel in operation is $F_s(x^n(j))$. For every $s \in \mathcal{S}$, the decoding regions for distinct messages should not overlap, that is, $F_s(x^n(j)) \cap F_s(x^n(j')) = \emptyset$ for all $1 \leq j \neq j' \leq N$, or, the codewords should be connected with respect to G_s for all $s \in \mathcal{S}$. Therefore, the zero-error capacity of a compound channel with an informed decoder, $C_{\text{dec}}^0(\mathfrak{C})$, is the capacity of the set of characteristic graphs $\mathcal{G}_{\mathcal{S}} = \{G_s, s \in \mathcal{S}\}$

$$C_{\text{dec}}^0(\mathfrak{C}) = C(\mathcal{G}_{\mathcal{S}}). \tag{9}$$

Using a definition of a valid zero-error code slightly different from ours, the above result was presented in [4].

C. Example: $C_{\text{dec}}^0(\mathfrak{C}) \neq C^0(\mathfrak{C})$

Consider the compound channel $\mathfrak{C}_1$ defined by $\mathcal{X} = \{1, 2\}, \mathcal{Y} = \{1, 2, 3, 4\}, \mathcal{S} = \{1, 2\}$. The fan out sets for the two constituent channels are as in Table I. Observe that $F_1(a) \cap F_2(b) \neq \emptyset, \forall a, b \in \mathcal{X}$. Given any two sequences x^n and x'^n, there exists a possible output sequence y^n such that either input sequence could have produced it: at every coordinate i choose y_i from $F_1(x_i) \cap F_2(x_i')$. Thus we can never reliably distinguish any pair of sequences, which implies that $C^0(\mathfrak{C}_1)$

TABLE I
FAN-OUT SETS FOR THE COMPOUND CHANNEL $\mathfrak{C}_1$/SI SETS FOR THE SOURCE
WITH COMPOUND SIDE-INFORMATION $\mathfrak{S}_1$

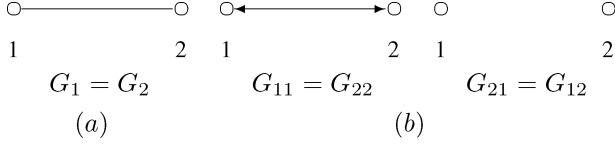| Channel Input/ Source Letter	Fan-out Sets/ SI Sets	
	Channel 1: $F_1(x)$/ (U, W_1): $S_1(u)$	Channel 2: $F_2(x)$/ (U, W_2): $S_2(u)$
1	$\{1, 2\}$	$\{1, 3\}$
2	$\{3, 4\}$	$\{2, 4\}$

Fig. 1. (a) Characteristic graphs for the channels in $\mathfrak{C}_1$ (Distinguishability graphs for the source-side-information pairs in $\mathfrak{S}_1$). (b) Characteristic set of graphs for $\mathfrak{C}_1$ (Graphs in the distinguishability set of $\mathfrak{S}_1$).

is zero, which is consistent with Theorem 1 since G_{12} and G_{21} are edge free graphs.

The characteristic graph for both channels is identical and is the complete graph on $\mathcal{X}$, shown in Fig. 1(a), which we denote by G. $C_{\text{dec}}^0(\mathfrak{C}_1) = C(\mathcal{G}_S) = C(G) = 1$ bit/channel use. Since, $C^0(\mathfrak{C}_1) = 0$, $C_{\text{enc}}^0(\mathfrak{C}_1)$ is also zero. Therefore, for this example, we have

$$0 = C^0(\mathfrak{C}_1) = C_{\text{enc}}^0(\mathfrak{C}_1) < C_{\text{dec}}^0(\mathfrak{C}_1) = 1.$$

V. SOURCE CODING WITH COMPOUND SIDE-INFORMATION AT THE DECODER

We study the problem of source coding with compound side-information, which is defined as follows: Let $(U, W_1, \ldots, W_M) \sim P_{UW_1 \ldots W_M}$ be a set of random variables defined over a finite set $\mathcal{U} \times \mathcal{W} \times \cdots \times \mathcal{W}$. The random variables represent the values taken at any instant by a discrete time stationary memoryless random process. Alice, who has access to the U sequence wishes to convey her information using fixed length codes to Bob, who has access to *one* of the M random sequences corresponding to $W_1, \ldots, W_M$. Neither Alice nor Bob knows which of the M random sequences is available to Bob. Such a situation can be motivated as follows: Alice and Bob are observing the samples of some underlying continuous time stationary process $\{(\hat{U}_t, \hat{W}_t)\}_{t \in [0, \infty)}$. However the clocks used for sampling by Alice and Bob are not synchronized. Although there is drift between the clocks, we assume that it is slow enough for the shift between sampling times to be constant over a block length. However, because of this drift, synchronizing would require a large overhead. Therefore, while Alice and Bob know that the joint distribution of their random sequences is one of a set of distributions, neither knows the exact distribution. Since the continuous time process is stationary, the coding process can assume that the U marginal is fixed and we have the problem described earlier. We characterize the minimum asymptotic rate $R(P_{UW_1 \ldots W_M})$ required for Alice to convey her information without any error by recasting the problem in terms of graphs. In the graph theoretic framework, this problem is seen to be the dual of the compound channel coding problem. The simpler case where Bob knows which of the possible sequences he has access to is equivalent to the scenario where Alice has to broadcast her information to a set of users each with his own side-information. Here, the zero-error minimum asymptotic average rate has been characterized earlier for both fixed-length [11], [12] and variable length coding [13].

Consider first an even simpler scenario where there is only one source of side-information. For this setup, Witsenhausen developed

a graph theoretic framework which he used to express the minimum asymptotic rate. If $(U, W) \sim P_{UW}(u, w), (u, w) \in \mathcal{U} \times \mathcal{W}$ is the source-side-information pair, for each $u \in \mathcal{U}$, define the SI set $S(u) = \{w \in \mathcal{W}: p(u, w) > 0\}$. The SI set for a vector in $\mathcal{U}^n$ is the Cartesian product of the SI sets at each coordinate. For the pair (U, W), the distinguishability graph $G_{UW} = (V_{UW}, E_{UW})$ is a graph with $V_{UW} = \mathcal{U}$ and $(u_1, u_2) \in E_{UW} \Leftrightarrow S(u_1) \cap S(u_2) = \emptyset$. Under any valid coding scheme, only vectors whose SI sets are nonintersecting can receive the same codeword. In other words, a valid code at block length n is equivalent to a partition of $\mathcal{U}^n$ into sets whose elements are pairwise connected with respect to the distinguishability graph. The minimum asymptotic rate $R(P_{UW})$ is the Witsenhausen rate of the distinguishability graph $R_{\text{w}}(G_{UW})$ defined next. For a graph $G = (V, E)$, if $M(G, n)$ denotes the size of the smallest partition of V^n using sets of elements that are pairwise connected with respect to G, the Witsenhausen rate is [14]

$$R_{\text{w}}(G) = \lim_{n \to \infty} \frac{1}{n} \log M(G, n). \qquad (10)$$

The limit in (10) exists from the submultiplicativity of $M(G, n)$ and Fekete's Lemma [7]. Note that the Shannon capacity and the Witsenhausen rate of a graph are duals in the sense that finding the capacity corresponds to finding the largest among the sets with pairwise connected elements while finding the Witsenhausen rate corresponds to finding the smallest partition into sets of the above type.

When there is compound side-information, for a given source vector u^n, the set of possible side-information sequences, denoted $S_{\mathfrak{S}}(u^n)$, is $\cup_k S_k(u^n)$, where $S_k(u^n)$ is the SI set for u^n with respect to the source-side-information pair (U, W_k). Two sequences in $\mathcal{U}^n$ can receive the same codeword only if their compound SI sets do not intersect. The distinguishability set (of directed graphs) for the source is defined as follows: $\mathcal{G}_{\mathfrak{S}} = \{G_{kk'} = (\mathcal{U}, E_{kk'}): k, k' = 1, \ldots, M\}$, where $(u_1, u_2) \in E_{kk'} \Leftrightarrow S_k(u_1) \cap S_{k'}(u_2) = \emptyset$. A necessary and sufficient condition for two vectors to be able to receive the same codeword is that they be incomparable with respect to every graph in the set $\mathcal{G}_{\mathfrak{S}}$. Motivated by this observation, we define the Witsenhausen rate of a set of directed graphs as follows. For a set of directed graphs $\mathcal{G}$ with a common vertex set V, if $M(\mathcal{G}, n)$ denotes the size of the smallest partition of V^n into sets whose elements are pairwise incomparable with respect to every graph in $\mathcal{G}$, the Witsenhausen rate of $\mathcal{G}$ is

$$\rho_{\text{w}}(\mathcal{G}) = \lim_{n \to \infty} \frac{1}{n} \log M(\mathcal{G}, n). \qquad (11)$$

Now we have the following.

Theorem 3: Consider a source with compound side-information $\mathfrak{S}$ whose joint probability distribution is $P_{U V_1 \ldots V_M}$. Let $\mathcal{G}_{\mathfrak{S}}$ denote the corresponding distinguishability set. For this source

$$R(P_{UW_1 \ldots W_M}) = \rho_{\text{w}}(\mathcal{G}_{\mathfrak{S}}). \qquad (12)$$

Using essentially the same arguments as in [11], [12], we can show that for a set of directed graphs $\mathcal{G}$

$$\rho_{\text{w}}(\mathcal{G}) = \max_{G \in \mathcal{G}} \rho_{\text{w}}(\{G\}). \qquad (13)$$

For the sake of brevity, we omit the proof.

If two vectors are incomparable with respect to G_{kk}, they are also connected with respected to G_k, the (undirected) distinguishability graph corresponding to the source-side-information pair (U, W_k). Therefore the minimum asymptotic rate achievable for the source coding with compound side-information problem is at least as large as $R(P_{UW_k}) = R_{\text{w}}(G_k)$, the minimum asymptotic rate when both Alice and Bob know that Bob is observing the W_k sequence. However, as is seen in the following example, $R(P_{UW_1 \ldots W_M})$ can be strictly greater than $\max_k R(P_{UW_k})$.

Consider the problem of coding for a source with compound side-information $\mathfrak{S}_1 = (U, W_1, W_2)$ where $\mathcal{U} = \{1, 2\}$, $\mathcal{W} = \{1, 2, 3, 4\}$. The SI sets for the two constituent source-side-information pairs of random variables are as in Table I. This source is the dual of the channel considered in Section IV-C.

Observe that $S_1(a) \cap S_2(b) \neq \emptyset, \forall a, b \in \mathcal{U}$. The distinguishability graph for both source-side-information pairs is identical and is the complete graph on $\mathcal{U}$, shown in Fig. 1(a), which we denote by G. $\max_k R_w(G_k) = R_w(G) = 0$ bit/channel use. However, given any two sequences u^n and u'^n, there exists a possible side-information sequence w^n such that both (u^n, w^n) and (u'^n, w^n) are observable with probability greater than zero: at every coordinate i choose w_i from $S_1(u_i) \cap S_2(u'_i)$. Thus, we can never reliably distinguish any pair of source sequences based on the side-information, which implies that the minimum asymptotic rate for this source is $\log 2 = 1$ bit/sample, which is consistent with Theorem 3 since G_{12} and G_{21} are edge free graphs. Therefore, in general, $R(P_{UW_1 \dots W_M})$ is strictly greater than $\max_k R(P_{UW_k})$. However, the minimum asymptotic rate is, somewhat surprisingly, indeed $\max_k R(P_{UW_k})$ [11], [12] if Bob (but not Alice) knows which of the M side-information sequences he is observing.

VI. THE COMPOUND CHANNEL: ASYMPTOTICALLY VANISHING ERROR V. ZERO ERROR

We observe that there are a number of differences between the conventional (asymptotically vanishing error) and zero-error scenarios. First, the conventional capacity does not increase if the decoder has side-information about the channel while the zero-error capacity does. The difference arises because, even without side-information, the decoder can almost reliably identify the channel in operation based on the channel output, which yields $C(\mathfrak{C}) = C_{\text{dec}}(\mathfrak{C})$ in the conventional case. However, if we require zero-error, *almost* reliable identification is not sufficient and $C^0(\mathfrak{C}) \leq C^0_{\text{dec}}(\mathfrak{C})$, possibly with strict inequality.

The aforementioned phenomenon also leads to another difference: in the conventional case, since the decoder can always effectively (almost reliably) know the channel in operation, $C_{\text{dec}}(\mathfrak{C}) \leq C_{\text{enc}}(\mathfrak{C})$. In the zero-error case, this inequality holds often but not always. The exception is encountered when some $G_{ss'}, s \neq s'$ in $G(\mathfrak{C})$ is the empty graph, and no $G_{ss}, s \in S$ is empty. In this case, the encoder cannot convey its knowledge to the decoder and $C^0_{\text{enc}}(\mathfrak{C})$ is zero while if the decoder knew the channel in operation, transmission at nonzero rates would be possible. Such a case arose in the example considered in Section IV. Although $C^0_{\text{enc}} < C^0_{\text{dec}}$ implies that $C^0_{\text{enc}} = 0$, in general, we could have $C^0(\mathfrak{C}) < C^0_{\text{dec}}(\mathfrak{C})$ with neither quantity being zero. For example, if $\mathcal{G} = \{G_{11}, G_{12}, G_{21}, G_{22}\}$ where G_{11} and G_{22} are both symmetrically directed complete graphs (edge set contains all ordered pairs of distinct vertices) on K vertices and G_{12} and G_{21} are identical symmetrically directed graphs on K vertices with a single pair of edges. From the discussion in Section III, there exists a compound channel $\mathfrak{C}$ whose characteristic set of graphs $\mathcal{G}(\mathfrak{C}) = \mathcal{G}$. For this channel, $C^0(\mathfrak{C}) = 1$ bit/channel use while $C^0_{\text{dec}}(\mathfrak{C}) = \log_2 K$ bits/channel use.

REFERENCES

[1] R. L. Dobrushin, "Optimal transmission of information in a channel with unknown parameters," *Radiotech. Electron.*, vol. 4, pp. 1951–1956, 1959.

[2] J. Wolfowitz, "Simultaneous channels," *Arch. Rational Mech. Anal.*, vol. 4, no. 4, pp. 371–386, 1960.

[3] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, 1959.

[4] G. Cohen, J. Körner, and G. Simonyi, "Zero-error capacities and very different sequences," in *Sequences: Combinatorics, Compression, Security and Transmission*, R. Capocelli, Ed.   New York: Springer-Verlag, 1988, pp. 144–155.

[5] L. Gargano, J. Körner, and U. Vaccaro, "Capacities: From information theory to extremal set theory," *J. Comb. Theory*, ser. A, vol. 68, no. 2, pp. 296–316, 1994.

[6] ——, "Qualitative independence and Sperner problems for directed graphs," *J. Comb. Theory, Ser. A*, vol. 61, no. 2, pp. 173–192, 1992.

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

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

[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 and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.

[11] G. Simonyi, "On Witsenhausen's zero-error rate for multiple sources," in *Proc. IEEE Int. Symp. Infrmation Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 363.

[12] ——, "On Witsenhausen's zero-error rate for multiple sources," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3258–3260, Dec. 2003.

[13] E. Tuncel and K. Rose, "On variable length coding of sources with side information at multiple decoders," in *Proc. IEEE Int. Symp. Inf.ormation Theory*, Yokohama, Japan, Jun./Jul. 2003, p. 140.

[14] H. S. Witsenhausen, "The zero-error side information problem and chromatic numbers," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 5, pp. 592–593, Nov. 1976.

Locally Optimal (Nonshortening) Linear Covering Codes and Minimal Saturating Sets in Projective Spaces

Alexander A. Davydov, Giorgio Faina, Stefano Marcugini, and Fernanda Pambianco

Abstract—A concept of locally optimal (LO) linear covering codes is introduced in accordance with the concept of minimal saturating sets in projective spaces over finite fields. An LO code is nonshortening in the sense that one cannot remove any column from a parity-check matrix without increasing the code covering radius. Several q^m-concatenating constructions of LO covering codes are described. Taking a starting LO code as a "seed", such constructions produce an infinite family of LO codes with the same covering radius. The infinite families of LO codes are designed using minimal saturating sets as starting codes. New upper bounds on the length function are given. New extremal and classification problems for linear covering codes are formulated and investigated, in particular, the spectrum of possible lengths of LO codes including the greatest possible length. The complete computer classification of the minimal saturating sets in small geometries and of the corresponding LO codes is obtained.

Index Terms—Covering codes, covering density, covering radius, minimal saturating sets in projective geometry, nonshortening covering codes.

I. INTRODUCTION

We consider linear covering codes, saturating sets in the projective spaces over finite fields, and connections between these objects. We