On the Random Coding Error Exponents of the Single-User and the Multiple-Access Gel'fand-Pinsker Channels

Anelia Somekh-Baruch and Neri Merhav
Department of Electrical Engineering
Technion - I.I.T., Haifa 32000, Israel.
{anelia@tx, merhav@ee}.technion.ac.il

Abstract

The random coding error exponent of a channel with non-causal side information at the transmitter (the Gel'fand Pinsker channel) is considered. A universally attainable lower bound on the random coding error exponent is derived using a random coding scheme which improves on the one suggested by Gel'fand and Pinsker. It is shown that the bound is tight in the random coding sense for our proposed scheme. Furthermore, we extend the result to the case of a two-user multiple access channel (MAC) with non-causal side information at the transmitters, i.e., we present a universally attainable lower bound on the error exponent and, as a byproduct, an inner bound on the corresponding achievable region of rates is deduced.

1 Introduction

The study of channels with state information at the transmitter has attracted the attention of many information theorists (see e.g., [1]-[3], [5]-[12], [14],[16]-[20]). The Gel'fand-Pinsker (G-P) channel models communication scenarios where the side information is available non-causally at the transmitter. One example is a watermarking system where the host data (covertext) within which the watermark is hidden can be regarded as side information at the transmitter. When the covertext is not available at the decoder's side (or, only partial information about it is known at the decoder) the communication setup resembles that of the G-P channel (see e.g., [4], [15]). Another example relates to computer memory with defects (see e.g., [10], [11], [12], [18]), when the locations of the defects are known to the encoder but not the decoder.

In [9], a single-letter expression was provided for the capacity of the G-P channel with state information that is emitted by a stationary memoryless source. The random coding scheme that is used to prove the achievablility part in [9] involves a binning principle and a typical-set decoder. While the analysis of the average probability of error of the random coding scheme presented in [9] suffices to prove that the probability of error can be made arbitrarily small while working at rates below the capacity, it does not provide the error exponent of the G-P channel. Moreover, the resulting error exponent of the coding scheme of [9] is essentially zero because when a non-typical state sequences is encountered, an error is declared. The error exponents of channels with side information at the transmitter have not attracted much attention of information theorists. One exception is [2], where an upper bound on the error exponent of the G-P channel was presented, which turned out to be mistaken. Another exception is [8] where the error exponents of modulo-additive noise channels with side information at the transmitter were investigated.

In this work, we present a random-coding scheme for the G-P channel whose error exponent is strictly positive at all rates below capacity. The average probability of error induced by the suggested scheme is analyzed in the exponential scale, yielding a single-letter lower bound on the highest achievable error exponent. As in [9], our scheme is based on the binning principle, but it includes a universal maximum mutual information (MMI) decoder rather than a typical-set decoder. The MMI decoder is chosen not only because works well for the ordinary discrete memoryless channel, but also because it is optimal (in the exponential scale) for our encoding scheme among all the decoders that compare metrics that depend only on the joint empirical statistics of a codeword and the observed output of the channel. The universality of the MMI decoder is w.r.t. the source emitting the state sequence, and the channel transition probabilities.

Further, we extend the results of the single-user G-P channel to a two user multiple access setup. There are many applications where the G-P MAC serves as an appropriate system model. One application is a public watermarking system with several independent users embedding their watermarks within the same covertext. Another example is that of a MAC with fading coefficients known at the transmitter. Yet another application is when several users share a computer memory with defects whose locations are known to all the users. Here, we use a time sharing auxiliary vector and a universal decoder similar to the one used in [13], yielding a single-letter lower bound on the reliability function and an inner bound on the region of achievable rates.

2 Notation and Definitions

Henceforth, we adopt the following notation conventions. Random variables will be denoted by capital letters while their realizations will be denoted by the respective lower case letters. Random vectors of dimension n will be denoted by boldface letters. Thus, for example, if $\mathbf{X}$ denotes a random vector $(X_1, \ldots, X_n)$, then $\mathbf{x} = (x_1, \ldots, x_n)$ will designate a specific sample value of $\mathbf{X}$. The alphabet of a scalar random variable (RV) X will be designated by a caligraphic letter $\mathcal{X}$. The n-fold Cartesian power of a generic alphabet $\mathcal{A}$, that is, the set of all n-vectors over $\mathcal{A}$, will be denoted $\mathcal{A}^n$.

The set of probability mass functions (pmf's), defined on a alphabet $\mathcal{X}$, will be denoted by $\mathcal{P}(\mathcal{X})$, and the set of conditional pmf's from $\mathcal{U}$ to $\mathcal{X}$ will be denoted $\mathcal{P}(\mathcal{X}|\mathcal{U})$, i.e., the transition matrices P such that $P(x|u) \geq 0$, $\sum_{x' \in \mathcal{X}} P(x'|u) = 1$, $\forall (u,x) \in \mathcal{U} \times \mathcal{X}$. The pmf of a random variable X will be denoted with an appropriate subscript, e.g., P_X, similarly, the joint pmf of the RV's X and Y will be denoted P_{XY} and the conditional pmf of Y given X will be denoted by $P_{Y|X}$. The empirical pmf induced by a vector $\mathbf{x} \in \mathcal{X}^n$ is the vector $\hat{P}_{\mathbf{x}} = \left\{\hat{P}_{\mathbf{x}}(a),\ a \in \mathcal{X}\right\}$, where $\hat{P}_{\mathbf{x}}(a)$ is the relative frequency of the letter a in the vector $\mathbf{x}$. For a pmf $P \in \mathcal{P}(\mathcal{X})$ the type class $T(P)$ is the set of n-vectors $\mathbf{x}$ such that $\hat{P}_{\mathbf{x}} = P$. Similarly, the joint empirical pmf induced by two n-vectors, $\mathbf{x}, \mathbf{y}$, is the vector $\hat{P}_{\mathbf{xy}} = \left\{\hat{P}_{\mathbf{xy}}(a,b),\ a \in \mathcal{X}, b \in \mathcal{Y}\right\}$, where $\hat{P}_{\mathbf{xy}}(a,b)$ is the relative frequency of occurrences of $(x_i, y_i) = (a,b)$. For a pmf $P \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, the type class $T(P)$ is the set of all pairs of n-vectors $\tilde{\mathbf{x}} \in \mathcal{X}^n, \tilde{\mathbf{y}} \in \mathcal{Y}^n$, such that $\hat{P}_{\tilde{\mathbf{x}}\tilde{\mathbf{y}}} = P$. The conditional type class $T_{\mathbf{y}|\mathbf{x}}$, for a given $\mathbf{x}$, is the set of all n-vectors $\tilde{\mathbf{y}} \in \mathcal{Y}^n$ such that $\hat{P}_{\mathbf{x}\tilde{\mathbf{y}}} = \hat{P}_{\mathbf{xy}}$, and the conditional empirical pmf $\hat{P}_{\mathbf{y}|\mathbf{x}}$ is defined by $\hat{P}_{\mathbf{y}|\mathbf{x}}(y|x) = \frac{\hat{P}_{\mathbf{xy}}(x,y)}{\hat{P}_{\mathbf{x}}(x)}$, $\forall x \in \mathcal{X} : \hat{P}_{\mathbf{x}}(x) > 0$. We shall also use the notation $T_{\tilde{P}}(\mathbf{x})$ for the conditional type class containing the vectors $\mathbf{y}$ such that $\hat{P}_{\mathbf{xy}} = \hat{P}_{\mathbf{x}} \times \tilde{P}$.

Information-theoretic quantities, such as the entropy of the random variable X whose pmf is P will be denoted by either $H_P(X)$ or $H(P)$. Similarly, the mutual information between X and Y, with joint pmf P, will be denoted by $I_P(X;Y)$, etc. For $P \in \mathcal{P}(\mathcal{X})$ and $Q \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$

we shall denote by PQ the induced joint pmf, i.e., $PQ(x,y) = P(x)Q(y|x)$. The divergence or Kullback-Leibler distance between two pmf's P and Q on $\mathcal{A}$, where $|\mathcal{A}| < \infty$, is defined as $D(P\|Q) = \sum_{a\in\mathcal{A}} P(a)\log\frac{P(a)}{Q(a)}$, where we use the convention that $0\log 0 = 0$ and $p\log\frac{p}{0} = \infty$, for a pmf $\mu \in \mathcal{P}(\mathcal{B})$ $|\mathcal{B}| < \infty$, the conditional divergence between two conditional pmf's P and Q on $\mathcal{A}$, where $|\mathcal{A}| < \infty$ is defined by $D(P\|Q|\mu) = \sum_{a\in\mathcal{A},\, b\in\mathcal{B}} P(b)P(a|b)\log\frac{P(a|b)}{Q(a|b)}$.

Information-theoretic quantities governed by empirical measures induced by the n-vectors $\mathbf{u},\mathbf{x},\mathbf{y}$, will have a shorthand notation, e.g., $\hat{H}(\mathbf{x}) \triangleq H_{\hat{P}_{\mathbf{x}}}(X)$, $\hat{I}(\mathbf{x};\mathbf{y}|\mathbf{u}) \triangleq I_{\hat{P}_{\mathbf{uxy}}}(X;Y|U)$. The notation $U \leftrightarrow X \leftrightarrow Y$ will signify that the RV's U,X,Y, in this order, form a Markov chain. For two functions $f_i : \mathbb{R}\to\mathbb{R}$, $i=1,2$, the notation $f_1(n) = O(f_2(n))$ will express the fact that the functions are of the same asymptotic order, i.e., $0 < \lim_{n\to\infty}\frac{f_1(n)}{f_2(n)} < \infty$. The notation $|t|^+$, where $t\in\mathbb{R}$, will designate $\max\{0,t\}$.

3 Statement of the Problem

Consider the discrete state-dependent channel with input alphabet $\mathcal{S}\times\mathcal{X}$ and output alphabet $\mathcal{Y}$ such that $|\mathcal{S}\times\mathcal{X}| < \infty$, $|\mathcal{Y}| < \infty$,

$$W_0^n(\mathbf{y}|\mathbf{s},\mathbf{x}) \;=\; \prod_{i=1}^{n} W_0(y_i|s_i,x_i), \tag{1}$$

where $\mathbf{s}$ is the state sequence emitted from a stationary memoryless source $Q_0\in\mathcal{P}(\mathcal{S})$, i.e., $Q_0^n(\mathbf{s}) = \prod_{i=1}^n Q_0(s_i)$. We shall refer to this channel as *a G-P channel*. The transmitter observes $\mathbf{s}$ and wishes to send a message index taking values in $\mathcal{M}_{n,R} \triangleq \{1,...,e^{nR}\}$ to the receiver, which in turn, does not observe the state sequence. It is assumed that the message is distributed uniformly over $\mathcal{M}_{n,R}$. A block (n,M)-code denoted $\mathcal{C}_{n,M}$ (where $M = e^{nR}$) for the channel is defined by the encoder $\varphi_n : \mathcal{M}_{n,R}\times\mathcal{S}^n\to\mathcal{X}^n$, and the decoder $\phi_n : \mathcal{Y}^n\to\mathcal{M}_{n,R}$. The average probability of error is induced by Q_0, W_0, φ_n and ϕ_n, and is given by $P_e(\mathcal{C}_{n,M},Q_0,W_0) = \frac{1}{M}\cdot\Pr\{\phi_n(\mathbf{Y})\neq m|m \text{ transmitted}\}$, where $\mathbf{Y}$ is the RV designating the received n-vector. A rate $R>0$ is called achievable if there exists a sequence of $(n,\exp\{nR\})$-codes, $\{\mathcal{C}_{n,M}\}_{n\geq 1}$, such that $P_e(\mathcal{C}_{n,M},Q_0,W_0)\to 0$. The capacity is defined as the supremum of all achievable rates. A number $E\geq 0$ is called *an achievable error exponent at rate R for the G-P channel W_0 with state distribution Q_0*, if for every $\epsilon > 0$ and sufficiently large n, there exists a code $(\mathcal{C}_{n,M})$ such that $M = \exp\{nR\}$ and $P_e(\mathcal{C}_{n,M},Q_0,W_0)\leq \exp\{-n(E-\epsilon)\}$. The supremum of all achievable exponents at rate R is referred to as *the reliability function* of the G-P channel.

A possible generalization of this channel for the two user multiple access case is as follows. The multiple access channel (MAC) with input alphabet $\mathcal{S}\times\mathcal{X}_1\times\mathcal{X}_2$ and output alphabet $\mathcal{Y}$ such that $|\mathcal{S}\times\mathcal{X}_1\times\mathcal{X}_2| < \infty$, $|\mathcal{Y}| < \infty$, is given by

$$W_0^n(\mathbf{y}|\mathbf{s},\mathbf{x},\tilde{\mathbf{x}}) \;=\; \prod_{i=1}^{n} W_0(y_i|s_i,x_i,\tilde{x}_i), \tag{2}$$

where $\mathbf{s}$ is the state sequence, generated by Q_0 as above and available to the two transmitters. A code $\mathcal{C}_{n,M_1,M_2}$ for the G-P MAC is composed of the message sets $\mathcal{M}_{n,R_1}$ and $\mathcal{M}_{n,R_2}$, the encoders $\varphi_n^{(1)} : \mathcal{M}_{n,R_1}\times\mathcal{S}^n\to\mathcal{X}^n$, $\varphi_n^{(2)} : \mathcal{M}_{n,R_2}\times\mathcal{S}^n\to\mathcal{X}^n$ and the decoder $\phi_n : \mathcal{Y}^n\to\mathcal{M}_{n,R_1}\times\mathcal{M}_{n,R_2}$. We assume that the messages are independent RV's, each distributed uniformly over its corresponding message set. The average probability of error induced by $Q_0, W_0, \mathcal{C}_{n,M_1,M_2}$ is given by $P_e(\mathcal{C}_{n,M_1,M_2},Q_0,W_0) = \frac{1}{M_1\cdot M_2}\cdot\Pr\{\phi_n(\mathbf{Y})\neq(m_1,m_2)|(m_1,m_2) \text{ transmitted}\}$. A rate pair (R_1,R_2)

is called achievable if there exists a sequence of $(n, \exp\{nR_1\}, \exp\{nR_2\})$-codes, $\{\mathcal{C}_{n,M_1,M_2}\}_{n\geq 1}$, such that $P_e(\mathcal{C}_{n,M_1,M_2}, Q_0, W_0) \to 0$. The capacity region is defined as the closure of the set of achievable rates. The definitions of an achievable error exponent and the reliability function at rate pair (R_1, R_2) are analogous to those of the single-user case.

4 Upper Bound on the Error Exponent for the Single-User Channel

In this section, we present two random coding schemes (Scheme A and Scheme B) and analyze their resulting error exponents. Scheme B has an error exponent at least as large as that of Scheme A (and hence, yields a tighter lower bound on the reliability function of the G-P channel), but it is composed of two stages, with the first stage being implementation of the Scheme A, thus, the two schemes are described below.

Scheme A: The coding scheme is as follows. Let $\mathcal{U}$ be a finite set. For given $R > 0$ and a constant $K > 0$, the encoder draws $e^{n(R+K)}$ independent random variables, uniformly distributed within a certain type class $T(\tilde{P})$ of n-sequences corresponding to $U \sim \tilde{P}$, $\tilde{P} \in \mathcal{P}(\mathcal{U})$. This set of $e^{n(R+K)}$ vectors is partitioned into M bins each of size $J = e^{nK}$ in an arbitrary manner.

The encoder selects a mapping $\mu : \mathcal{P}(\mathcal{S}) \to \mathcal{P}(\mathcal{U}|\mathcal{S})$ that satisfies for every $Q \in \mathcal{P}(\mathcal{S})$

$$I_{Q\times\mu(Q)}(S;U) + \epsilon_n \leq K, \tag{3}$$

where $\epsilon_n = O(n^{-\gamma})$ for some $\gamma < 1$, such that the marginal of the joint pmf defined by $Q \times \mu(Q)$ is always $\tilde{P}$ (obviously, such a mapping always exists, e.g., $\mu(Q) = \tilde{P}$ and consequently $I_{Q\times\mu Q}(S;U) = 0$ hence, assuming $\epsilon_n \to 0$, it is guaranteed that (3) is satisfied for sufficiently large n).

Given a message index $0 \leq m \leq e^{nR}$ and $\mathbf{s}$, let $\mathbf{u}(\mathbf{s}, m)$ be the first randomly drawn vector $\mathbf{u}$ within bin number m such that $(\mathbf{s}, \mathbf{u})$ is strongly jointly typical w.r.t. $\hat{P}_{\mathbf{s}} \times \mu(\hat{P}_{\mathbf{s}})$. If no such a $\mathbf{u}$ vector exists within bin number m, then an error is declared by the encoder.

Next, the encoder selects a mapping $\zeta : \mathcal{P}(\mathcal{S} \times \mathcal{U}) \to \mathcal{P}(\mathcal{X}|\mathcal{S} \times \mathcal{U})$ and generates $\mathbf{x}$ from $\mathbf{s}$ and $\mathbf{u} = \mathbf{u}(\mathbf{s}, m)$ using a pmf that is uniform over the type class $T_\zeta(\mathbf{s}, \mathbf{u}) \triangleq \{\mathbf{x}' : \hat{P}_{\mathbf{s},\mathbf{u},\mathbf{x}'} = \hat{P}_{\mathbf{s},\mathbf{u}} \times \zeta(\hat{P}_{\mathbf{s},\mathbf{u}})\}$ and transmits $\mathbf{x}$ across the channel.

The decoder's estimation of the message index, $\hat{m}$, is the index of the bin that contains the vector $\mathbf{u}$ that maximizes the empirical mutual information $\hat{I}(\mathbf{u};\mathbf{y})$ where ties are broken arbitrarily.

The main differences between our proposed scheme and the one used in [9] are (a) the use of the MMI decoder rather than the typical-set decoder and (b) the treatment of non-typical state sequences: instead of declaring an error when a non typical state sequence is encountered, we search for a codeword whose joint empirical statistics with the state sequence is typical w.r.t. some distribution $(\hat{P}_{\mathbf{s}} \times \mu(\hat{P}_{\mathbf{s}}))$ that can be optimized.

The average probability of error attained by this scheme depends on $\mu, K, \tilde{P}$ and ζ.

Proposition 1. *(Random Coding Bound for the Single-User Channel) For every G-P channel $W_0 : \mathcal{S} \times \mathcal{X} \to \mathcal{Y}$ with state information distribution $Q_0 \in \mathcal{P}(\mathcal{S})$, $\epsilon > 0$, $R \geq 0, K > 0$, every finite set $\mathcal{U}$, $\tilde{P} \in \mathcal{P}(\mathcal{U})$ and sufficiently large n there exists a code $\mathcal{C}_{n,M}$ with $M \geq \exp\{nR\}$ such that*

$$P_e(\mathcal{C}_{n,M}, Q_0, W_0) \leq \exp\left\{-n\left[E_1(\tilde{P}, K, R) - \epsilon\right]\right\}, \tag{4}$$

where

$$E_1(\tilde{P}, K, R) = \inf_{Q} \sup_{F,V} \inf_{W} \left[D(QFVW\|Q_0FVW_0) + |I_{QFVW}(U;Y) - K - R|^+\right], \tag{5}$$

with the left-most infimum carried out over $Q \in \mathcal{P}(\mathcal{S})$, the supremum is over[1] $\{F \in \mathcal{P}(\mathcal{U}|\mathcal{S}) : I_{QF}(U;S) \leq K$, *with marginal* $\tilde{P}$, $V \in \mathcal{P}(\mathcal{X}|\mathcal{S}\times\mathcal{U})\}$, *and the right-most infimum is over the set of* $W \in \mathcal{P}(\mathcal{Y}|\mathcal{S}\times\mathcal{U}\times\mathcal{X})$.

Since both $\tilde{P}$ and K can be optimized by the encoder, we get the following Corollary.

Corollary 1. *The reliability function of the G-P channel W_0 with state information distribution Q_0 at rate R is lower bounded by*

$$E_1(R) = \sup_{\mathcal{U},\tilde{P}\in\mathcal{P}(\mathcal{U}),K>0} E_1(\tilde{P},K,R). \tag{6}$$

The proof of Proposition 1 is based on the observation that since our choice of the mapping μ satisfies (3) by definition, a random selection of e^{nK} random variables in $T(\tilde{P})$ fails to cover $\hat{P}_{\mathbf{s}} \times \mu(\hat{P}_{\mathbf{s}})$ in the sense of joint typicality with probability that vanishes faster than exponentially. It should be noted that the bound is tight (in the exponential scale) in the sense that it reflects the error exponent of the proposed Scheme A, i.e., for given $\tilde{P}, K, R$ the best attainable error exponent (corresponding to the optimal choice of μ and ν) is given by $E_1(\tilde{P},K,R)$.

Scheme B: The second scheme is composed of two stages. At the first stage, the encoder indexes the empirical measures of order n and transmits over the channel the index of the type class of the observed state sequence $\mathbf{s}$ using Scheme A. As the information conveyed in the index of the type class grows logarithmically with n this can be achieved using Scheme A at a rate $R = 0$ with error exponent given by $E_1(0)$. If the transmission of the type class fails, we consider it as an error event in the analysis of the probability of error.

Further, a random code that assumes knowledge of $\hat{P}_{\mathbf{s}}$ at the receiver can be designed. For each empirical measure $\hat{P} \in \mathcal{P}(\mathcal{S})$ of order n, a separate codebook is drawn similarly to the method described in the first scheme, where now, $\mathcal{U}$, $\tilde{P}$ and K can also depend on $\hat{P}_{\mathbf{s}}$. Again, given a message index $0 \leq m \leq e^{nR}$ and $\mathbf{s} \in T(\hat{P})$, the encoder searches for $\mathbf{u}(\mathbf{s},m)$, the first randomly drawn vector $\mathbf{u}$ within bin number m of the codebook corresponding to $\hat{P}$ such that $(\mathbf{s},\mathbf{u})$ is jointly typical w.r.t. $\hat{P}\times\mu(\hat{P})$, and generates $\mathbf{x}$ from $\mathbf{s}$ and $\mathbf{u}(\mathbf{s},m)$ using a pmf that is uniform over the conditional type class $T_\nu(\mathbf{s},\mathbf{u}(\mathbf{s},m))$. The decoder's estimation of the message index, $\hat{m}$, is the number of the bin that contains the vector $\mathbf{u}$ that maximizes the empirical mutual information $\hat{I}(\mathbf{u};\mathbf{y})$ within the codebook corresponding to $\hat{P}$.

The fact that $K,\mathcal{U},\tilde{P}$ may depend on the type class of the state sequence, enables us to replace the order of the supremum in (6) with the infimum over Q in $E_1(\tilde{P},K,R)$ (5). Next, it is easily verified that the optimal (K,μ) satisfying (3) are such that $K = I_{Q\times\mu}(S;U) + \epsilon_n$. Following the same line of proof as in the first bound, this yields the following lower bound on the reliability function *assuming knowledge of the type class of the states sequence at the receiver*

$$\tilde{E}_2(R) \triangleq \inf_{Q} \sup_{(\mathcal{U},F,V)} \inf_{W} \left[D(QFVW\|Q_0FVW_0) + |I_{QFVW}(U;Y) - I_{QF}(S;U) - R|^+\right], \tag{7}$$

where optimizations are over $Q \in \mathcal{P}(\mathcal{S})$, finite set $\mathcal{U}$, $F \in \mathcal{P}(\mathcal{U}|\mathcal{S})$, $V \in \mathcal{P}(\mathcal{X}|\mathcal{S}\times\mathcal{U})$ and $W \in \mathcal{P}(\mathcal{Y}|\mathcal{S}\times\mathcal{U}\times\mathcal{X})$. The resulting error exponent of the two-stage scheme is presented in the following theorem.

Theorem 1. *The reliability function of the G-P channel W_0 with state information distribution Q_0 at rate R is lower bounded by*

$$E_2(R) = \min\{E_1(0), \tilde{E}_2(R)\}. \tag{8}$$

[1] In fact, $F = \mu(Q)$ and $V = \nu(Q\times F)$.

Obviously, $E_2(R) \geq E_1(R)$, because $\tilde{E}_2(R) \geq E_1(R)$ since $\tilde{E}_2(R)$ expresses genie-aided error exponent where the type class of the state sequence is known at the receiver (mathematically speaking, this is also easily verified by inspecting the order of the infimum and supremum which appear in $\tilde{E}_2(R)$ and $E_1(R)$). It should be noted that while the bound in Corollary 1 is tight for Scheme A, it remains a question whether $E_2(R)$ is a tight bound on the exponent of Scheme B.

5 Upper Bound on the Error Exponent for the Multiple Access Channel

The suggested scheme for the two-user multiple-access channel is a generalization of the single user's Scheme B. Given rates $R_1 \geq 0$, $R_2 \geq 0$, a finite set $\mathcal{L}$ and a pmf $P_L \in \mathcal{P}(\mathcal{L})$, the encoders and decoder set a vector $\mathbf{l} \in T(P_L)$ to be used for time-sharing. Then, each user $i = 1, 2$ chooses a constant $K_i > 0$, a finite set $\mathcal{U}_i$, a conditional probability distribution $\tilde{P}_i \in \mathcal{P}(\mathcal{U}_i|\mathcal{L})$, a mapping $\mu_i \in \mathcal{P}(\mathcal{U}_i|\mathcal{L} \times \mathcal{S})$ that satisfies for every $P \in \mathcal{P}(\mathcal{L} \times \mathcal{S})$

$$I_{P \times \mu_i(P)}(S; U|L) + \epsilon_n \leq K_i, \tag{9}$$

where $\epsilon_n = O(n^{-\gamma})$ for some $\gamma < 1$, and a mapping $\zeta_i : \mathcal{P}(\mathcal{L} \times \mathcal{S} \times \mathcal{U}_i) \to \mathcal{P}(\mathcal{X}|\mathcal{L} \times \mathcal{S} \times \mathcal{U}_i)$.

The users generate independently separate codebooks, as described in Scheme A, and given the message index m_i and the states sequence $\mathbf{s}$, user i searches for the first randomly drawn vector $\mathbf{u}$ within bin number m_i in its codebook, denoted $\mathbf{u}_i(\mathbf{l}, \mathbf{s}, m_i)$, such that $(\mathbf{l}, \mathbf{s}, \mathbf{u})$ is strongly jointly typical w.r.t. $\hat{P}_{\mathbf{ls}} \times \mu_i(\hat{P}_{\mathbf{ls}})$, and then draws a vector $\mathbf{x}_i$ to be transmitted uniformly over $T_{\zeta_i}(\mathbf{l}, \mathbf{s}, \mathbf{u}_i(\mathbf{l}, \mathbf{s}, m_i))$, i=1,2.

Next we turn to the description of the decoding rule. In [13], Liu and Hughes introduced a universal decoding rule for a multiple access channel which minimizes the empirical conditional entropy $\hat{H}(\mathbf{x}_1, \mathbf{x}_2|\mathbf{l}, \mathbf{y})$ among the transmitted codewords $\mathbf{x}_1, \mathbf{x}_2$. Here, we consider an extension of this decoder to the case of SI at the transmitter. The estimations of the message indices, $\hat{m}_1$ and $\hat{m}_2$, are the indices of the bins that contain the vectors $\mathbf{u}_1, \mathbf{u}_2$ (belonging to the first and second codebooks, respectively) that minimize the empirical conditional entropy $\hat{H}(\mathbf{u}_1, \mathbf{u}_2|\mathbf{l}, \mathbf{y})$. As explained in [13], since $\hat{H}(\mathbf{x}|\mathbf{y}) = \hat{H}(\mathbf{x}) - \hat{H}(\mathbf{x}|\mathbf{y})$, a minimum empirical entropy decoder coincides with the MMI decoder for single user channel and constant composition codes, however, in the multiple access channel $\hat{H}(\mathbf{u}_1, \mathbf{u}_2|\mathbf{l}, \mathbf{y}) = \hat{H}(\mathbf{u}_1, \mathbf{u}_2|\mathbf{l}) - \hat{I}(\mathbf{u}_1\mathbf{u}_2; \mathbf{y}|\mathbf{l})$, hence the minimum empirical decoder does not coincide with the MMI decoder. Finally, one can use the above described scheme to first transmit the joint type class of $\mathbf{s}, \mathbf{l}$ at rate $R_1 = 0, R_2 = 0$, and then, as in Scheme B, implement a similar scheme where the knowledge of the type class is assumed in the receiver. The error exponent of the resulting scheme yields a lower bound on the reliability function of the G-P MAC, in the spirit of the expression attained in [13]. For the sake of brevity, we do not present the expression here, but it is easily verified that it is equal zero if at least one of the following inequalities is violated

$$\begin{aligned} R_1 &\leq I(U_1; Y|U_2 L) - I(U_1; S|U_2 L) \\ R_2 &\leq I(U_2; Y|U_1 L) - I(U_2; S|U_1 L) \\ R_1 + R_2 &\leq I(U_1 U_2; Y|L) - I(U_1 U_2; S|L) \end{aligned} \tag{10}$$

for every seven RV's $(L, S, U_1, U_2, X_1, X_2, Y)$ having joint pmf $P_L \times Q_0 \times P_{U_1,X_1|L,S} \times P_{U_2,X_2|L,S} \times W_0$ where L, U_1, U_2 take values in some finite sets $\mathcal{L}, \mathcal{U}_1, \mathcal{U}_2$, respectively. Hence, the set of achievable rates is given by the closure of the set of all (R_1, R_2) pairs satisfying these three inequalities for some choice of joint distribution $P_L \times Q_0 \times P_{U_1,X_1|L,S} \times P_{U_2,X_2|L,S} \times W_0$.

References

[1] R. Ahlswede, "Arbitrarily varying channels with states sequence known to the sender," *IEEE Trans. Inform. Theory*, vol. IT-32, no. 5, pp. 621–629, September 1986.

[2] E. A. Aroutunian and M. E. Aroutunian, "E-capacity upper bound for a channel with random parameter," *Probl. Contr. Inform. Theory*, vol. 17, no. 2, pp. 99–105, 1988.

[3] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading Channels: Information- Theoretic and Communications Aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, October 1998 (special commemorative issue 1948–1998).

[4] A. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1639-1667, June 2002.

[5] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[6] A. Das and P. Narayan, "Capacities of time-varying multiple-access channels with side information," *IEEE Trans. Inform. Theory*, vol. 48, no. 1, pp. 4–25, January 2002.

[7] U. Erez and R. Zamir, "Noise prediction for channel coding with side information at the transmitter," *IEEE Trans. Inform. Theory*, vol. 46, no. 1, pp. 1610–1617, July 2000.

[8] U. Erez and R. Zamir, "Error exponents of modulo-additive noise channels with side information at the transmitter," *IEEE Trans. Inform. Theory*, vol. 47, no. 1, pp. 210–218, January 2001.

[9] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Information and Control*, pp. 19-31, 1980.

[10] C. Heegard, "On the capacity of permanent memory," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 1, pp. 34–42, January 1985.

[11] C. Heegard and A. A. El-Gammal, "On the capacity of commuter memory with defects," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 5, pp. 731–739, September 1983.

[12] A. V. Kuznetsov and A. J. Han Vinck, "On the general defective channel with informed encoder and capacities of some constrained memories," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1866–1871, November 1994.

[13] Y. S. Liu and B. L. Hughes, "A new universal random coding bound for hte multiple access channel," *IEEE Trans. Inform. Theory*, vol. 42, no. 2, pp. 376–386, March 1996.

[14] C. E. Shannon, "Channels with side information at the transmitter," IBM J. Res. Develop., vol. 2, pp. 289–293, October 1958.

[15] A. Somekh-Baruch and N. Merhav, "On the capacity game of public watermarking systems," to appear in *IEEE Trans. Inform. Theory.*

[16] A. Sutivong, T. M. Cover, and M. Chiang, "Tradeoff between message and state information rates," *Proc. ISIT 2001*, p. 303, Washington, D.C., June 2001.

[17] A. Sutivong, T. M. Cover, M. Chiang, and Y.-H. Kim, "Rate vs. distortion trade-off for channels with state information," *Proc. ISIT 2002*, p. 226, Lausanne, Switzerland, June-July, 2002 (full version to appear in *IEEE Trans. Inform. Theory*).

[18] B. van Thanh, "Storage capacity of computer memories with defects," *Probl. Contr. Inform. Theory*, vol. 19, no. 5-6, pp. 423–434, 1990.

[19] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inform. Theory*, vol. 45, no. 2 pp. 761–771, March 1999.

[20] W. Yu, A. Sutivong, D. Julian, T. M. Cover, and M. Chiang, "Writing on colored paper," *Proc. ISIT 2001*, p. 302, Washington, D.C., June 2001.