# Opportunistic capacity and error exponent regions for variable length communication over compound channel with feedback

Aditya Mahajan and Sekhar Tatikonda
Department of Electrical Engineering,
Yale University, New Haven, CT, 06520

***Abstract*—Variable length communication over a compound channel with feedback is considered. Traditionally, capacity of a compound channel is defined as the maximum rate at which reliable communication can be guaranteed before the start of communication. This is a pessimistic point of view. In this paper, we give an opportunistic definition of capacity. We define capacity as the maximum rate at which reliable communication can be guaranteed for the current choice of the channel by nature. Under this definition, a compound channel is conceptually similar to multi-terminal communication. Transmission rate is a vector rather than a scalar; channel capacity is a region rather than a scalar; error exponent is a region rather than a scalar. We formulate variable length communication over a compound channel with feedback, characterize its opportunistic capacity region, and provide lower bounds for the error exponent region.**

## I. Introduction

A compound channel, first considered by Wolfowitz [1] and Blackwell *et. al.* [2], is one of the simplest extensions of a DMC (discrete memoryless channel). In a compound channel the transmitter and the receiver do not completely know the channel transition matrix $Q_\circ$; they only that $Q_\circ$ belongs to some family $\mathscr{Q}$. All channels in $\mathscr{Q}$ are defined over a common input alphabet $\mathscr{X}$ and output alphabet $\mathscr{Y}$. The capacity of a compound channel $\mathscr{Q}$ is given by (see [3])

$$C(\mathscr{Q}) = \max_{P \in \Delta(\mathscr{X})} \inf_{Q \in \mathscr{Q}} I(P, Q) \tag{1}$$

where $\Delta(\mathscr{X})$ is the family of probability distributions on input alphabet $\mathscr{X}$ and $I(P, Q)$ is the mutual information between the input and output of a channel with input distribution $P$ and channel transition matrix $Q$.

If noiseless feedback is available at the transmitter, the capacity of the compound channel is given by (see [4])

$$C_F(\mathscr{Q}) = \inf_{Q \in \mathscr{Q}} \max_{P \in \Delta(\mathscr{X})} I(P, Q). \tag{2}$$

The above notions of capacity are pessimistic. They quantify the maximum rate at which we can guarantee reliable communication *before the start of transmission*. In many applications, we do not care about a rate guarantee before the start of transmission. We would rather like to communicate at the maximum rate for the current choice of the channel $Q_\circ$ (even though this choice is not revealed to the transmitter or the receiver before the start of transmission). For example, consider a compound channel $\mathscr{Q} = \{Q_1, \dots, Q_L\}$. Suppose a coding scheme adapts the transmission rate (based on channel feedback) in such a manner that the probability of error is arbitrarily small for all choices of $Q_\circ$. Let $R_\ell$ denote the transmission rate when $Q_\circ = Q_\ell$. Then, the rate vector $(R_1, \dots, R_L)$ is *achievable* (which is defined formally in Section II). The union of all achievable rates is called the *opportunistic capacity* $\mathscr{C}_F(\mathscr{Q})$ of the compound compound channel $\mathscr{Q}$ with feedback, *i.e.*,

$$\mathscr{C}_F(\mathscr{Q}) = \big\{(R_1, \dots, R_L) : (R_1, \dots, R_L) \text{ is achievable}\big\}.$$

In contrast to the existing notions of capacity for a compound channel, the opportunistic capacity is a region rather than a scalar value.

It is straightforward to show (see Corollary 1) that the opportunistic capacity region is given by a hyper-rectangle

$$\mathscr{C}_F(\mathscr{Q}) = \big\{(R_1, \dots, R_L) : 0 \le R_\ell < C_{Q_\ell},\ \ell = 1, \dots, L\big\}.$$

This hyper-rectangle is determined by just its upper corner $(C_{Q_1}, \dots, C_{Q_L})$. This means that, as far as the transmission rate is concerned, not knowing the channel transition matrix does not entail any loss if channel feedback is available. However, the same is not true for the corresponding error exponents. As we have to use the same communication scheme for all choices of $Q_\circ$, we lose in terms of error exponents. In this paper, we study the error exponents of the compound channel with feedback and show that the loss in error exponents for variable length communication is at most a (channel $\mathscr{Q}$ dependent) multiplicative factor.

In a DMC with feedback, *variable length coding*, *i.e.*, allowing the coding scheme to have different length along different sample paths, significantly improves the error exponent [5]. More importantly, this improvement comes at very little cost: the best error exponents can be achieved by a simple coding scheme [6]; asymptotically, the scheme has a constant length along almost all sample paths. For these reasons, we concentrate on variable length coding for a compound channel with feedback. To fix ideas, we focus on a finite family $\mathscr{Q} = \{Q_1, \dots, Q_L\}$ and restrict ourselves to rates on the principle diagonal $(\gamma C_{Q_1}, \dots, \gamma C_{Q_L})$ of the capacity region.

It turns out that the error exponent of a compound channel is conceptually similar to the error exponents of multi-terminal channels [7]—for any achievable rate $(R_1, \dots, R_L)$, the error exponent is given by a region, called the *error exponent region* (EER).

In a DMC $Q$ with feedback, variable length coding at a rate $R < C_Q$ gives an error exponent (see [5])

$$E_B(R, Q) = B_Q \left(1 - R/C_Q\right), \tag{3}$$

where

$$B_Q = \max_{x_A, x_N \in \mathscr{X}} D\big(Q(\cdot|x_A) \| Q(\cdot|x_B)\big), \tag{4}$$

$Q(\cdot|x)$ is the probability distribution of the channel output when the channel input is $x$, and $D(p\|q)$ is the Kullback-Leibler divergence between probability distributions $p$ and $q$. We call $E_B(R, Q)$ the *Burnashev exponent* of channel $Q$ at rate $R$. The Burnashev exponent has a non-zero slope at capacity. This slope captures the main advantage of noiseless feedback—by reducing the transmission rate by a small *fraction* of the capacity, we can linearly increase the error exponent, and therefore, exponentially decrease the probability of error. We propose a coding scheme that retains this feature when communicating over a compound channel with feedback.

A trivial upper bound for the EER at rate $(R_1, \dots, R_L)$ inside the opportunistic capacity region $C(\mathscr{Q})$ is a hyper-rectangle with upper corner

$$\left(B_{Q_1}\left(1 - R_1/C_{Q_1}\right), \dots, B_{Q_L}\left(1 - R_L/C_{Q_L}\right)\right) \tag{5}$$

Tchamkerten and Telatar [8] identified necessary and sufficient conditions under which the above upper bound is tight for all rates along the principle diagonal $(\gamma C_{Q_1}, \dots, \gamma C_{Q_L})$, $0 \le \gamma < 1$, of the opportunistic capacity region. No lower bounds for the error exponents are known for channels that do not satisfy the conditions of [8]. In this paper, we present an achievable coding scheme for all rates along the principle diagonal of the opportunistic capacity region. The error exponents of this scheme have non-zero slope at capacity.

## II. Opportunistic capacity and error exponents

In this section we formally define opportunistic capacity and error exponent regions for a compound channel with feedback.

A *variable length coding scheme* for communicating over a compound channel $\mathscr{Q} = \{Q_1, \dots, Q_L\}$ with feedback is a tuple $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$ where

- $\mathbf{M} = (M_1, \dots, M_L)$ is the *compound message size* where $M_\ell \in \mathbb{N}$, $\ell = 1, \dots, L$. Define $\mathscr{M} = \prod_{\ell=1}^{L}\{1, \dots, M_\ell\}$.
- $\mathbf{f} = (f_1, f_2, \dots)$ is the *encoding strategy* where

$$f_t : \mathscr{M} \times \mathscr{Y}^{t-1} \mapsto \mathscr{X}, \quad t \in \mathbb{N}$$

  is the encoding function used at time $t$.
- $\mathbf{g} = (g_1, g_2, \dots)$ is the *decoding strategy* where

$$g_t : \mathscr{Y}^t \mapsto \bigcup_{\ell=1}^{K}\{(\ell,1), (\ell,2), \dots, (\ell, M_\ell)\}, \quad t \in \mathbb{N}$$

  is the decoding function at time $t$.
- $\tau$ is the stopping time with respect to the channel outputs $Y^t$. More precisely, $\tau$ is a stopping time with respect to the filtration $\{2^{\mathscr{Y}^t}, t \in \mathbb{N}\}$.

The coding scheme is known to both the transmitter and the receiver. Variable length communication takes place as follows. A *compound message* $\mathbf{W} = (W_1, \dots, W_L)$ is generated such that $W_\ell$ is uniformly distributed in $\{1, \dots, M_\ell\}$.[1] The transmitter uses the encoding strategy $(f_1, f_2, \dots)$ to generate channel inputs

$$X_1 = f_1(W), \quad X_2 = f_2(W, Y_1), \quad \cdots$$

until the stopping time $\tau$ with respect to the channel outputs. ($\tau$ is known to the transmitter because of feedback.) The decoder then generates a decoding decision

$$(\hat{L}, \hat{W}) = g_\tau(Y_1, \dots, Y_\tau).$$

The decoding decision consists of two components: an estimate $\hat{L}$ of the channel, and an estimate $\hat{W}$ for the $\hat{L}$-component of $\mathbf{W}$. A communication error occurs if $\hat{W} \neq W_{\hat{L}}$. Note that successful communication does not require $\hat{L}$ to be the equal to the index of the true channel.

The two main performance metrics of a coding scheme are its rate and error probabilities. Both the rate and error probabilities are vectors (rather than scalars) and denoted by $\mathbf{R} = (R_1, \dots, R_L)$ and $\mathbf{P} = (P_1, \dots, P_L)$, respectively.

During a particular transmission instance, the error event is $\{\hat{W} \neq W_{\hat{L}}\}$. Communication takes $\tau$ units of time, and if a communication error does not occur, $M_{\hat{L}}$ messages are communicated. Once $Q_\circ$ is specified, $\hat{W}$, $W_{\hat{L}}$, $M_{\hat{L}}$ and $\tau$ become random variables. Then, rate and error probabilities of the coding scheme $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$ can be defined as follows.

The component $R_\ell$ of the rate vector $\mathbf{R}$ is the ratio of the expected value of $\log M_{\hat{L}}$ and the expected value of $\tau$, where the expectations are assuming that $Q_\circ = Q_\ell$, *i.e.*,

$$R_\ell = \frac{\mathbb{E}_\ell[\log M_{\hat{L}}]}{\mathbb{E}_\ell[\tau]}$$

where $\mathbb{E}_\ell[\cdot]$ is a short hand notation for $\mathbb{E}[\cdot|Q_\circ = Q_\ell]$. Note that the $R_\ell$ component of the rate vector $\mathbf{R}$ depends on the compound message size $\mathbf{M}$ and not just its $M_\ell$ component.

The components $P_\ell$ of the probability of error vector $\mathbf{P}$ is the probability of the error event $\{\hat{W} \neq W_{\hat{L}}\}$ when $Q_\circ = Q_\ell$, *i.e.*,

$$P_\ell = \mathbb{P}_\ell(\hat{W} \neq W_{\hat{L}})$$

where $\mathbb{P}_\ell(\cdot)$ is a short hand notation for $\mathbb{P}(\cdot|Q_\circ = Q_\ell)$.

A rate vector $\mathbf{R} = (R_1, \dots, R_L)$ is said to be *achievable* if there exists a sequence of variable length coding schemes $(\mathbf{M}^{(n)}, \mathbf{f}^{(n)}, \mathbf{g}^{(n)}, \tau^{(n)})$, $n \in \mathbb{N}$ such that:

1) $\lim_{n\to\infty} \mathbb{E}_\ell[\tau^{(n)}] = \infty$ for $\ell = 1, \dots, L$.

[1] All the probabilities of interest only depend on the marginal distributions of $W_1$, …, $W_L$. So, the joint distribution of $(W_1, \dots, W_L)$ need not be specified.

2) For any $\varepsilon > 0$, there exists a $n_\circ(\varepsilon)$ so that for all $n \geq n_\circ(\varepsilon)$, we have
$$P_\ell^{(n)} < \varepsilon \quad \text{and} \quad R_\ell^{(n)} \geq R_\ell - \varepsilon, \quad \text{for all } \ell = 1, \dots, L.$$

Note that our definition does not require $\lim_{n\to\infty} \mathbb{E}_\ell[\hat{L}^{(n)}] = \ell$, although we expect that any reasonable coding scheme will achieve that.

The union of all achievable compound rates is called the *opportunistic capacity region* of channel $\mathscr{Q}$ with feedback and denoted by $\mathscr{C}_F(\mathscr{Q})$. It can be shown that (see Corollary 1) $\mathscr{C}_F(\mathscr{Q})$ is given by a hyper-rectangle with upper corner $(C_{Q_1}, \dots, C_{Q_L})$.

Given a sequence of coding schemes $(\mathbf{M}^{(n)}, \mathbf{f}^{(n)}, \mathbf{g}^{(n)}, \tau^{(n)})$, $n \in \mathbb{N}$, that achieve a rate vector $\mathbf{R}$, the asymptotic exponent $E_\ell$ of error probability $P_\ell$ is given by
$$E_\ell = \lim_{n\to\infty} -\frac{\log P_\ell^{(n)}}{\mathbb{E}_\ell[\tau^{(n)}]}.$$
Then $\mathbf{E} = (E_1, \dots, E_L)$ is the error exponent of sequence of coding schemes $(\mathbf{M}^{(n)}, \mathbf{f}^{(n)}, \mathbf{g}^{(n)}, \tau^{(n)})$, $n \in \mathbb{N}$.

It turns out that we can have multiple sequences of coding schemes that achieve the same rate but have different error exponents. Thus, similar to the error exponents of multi-terminal communication [7], the error exponents of a compound channel with feedback are given by a region. We call this region, *the error exponent region* (EER), and denote it by $\mathscr{E}(\mathbf{R})$. In this paper, we study the EER for all rate of the opportunistic capacity region and present lower bounds on the EER.

*Operational interpretation*

A compound channel is a point-to-point channel, so at first glance, our definitions of transmitted message $\mathbf{W} = (W_1, \dots, W_L)$ and decoding error $\{\hat{W} \neq W_{\hat{L}}\}$ may seem a bit strange. Below, we provide an operational interpretation of the compound message.

Suppose a higher-layer application generates an infinite bit stream that the transmitter wants to communicate reliably to a receiver over a compound channel with feedback. Furthermore assume that the transmitter uses a variable length coding scheme $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$ for that purpose. For ease of exposition, assume that all $M_\ell$, $\ell = 1, \dots, L$, are powers of 2 so that $\log_2 M_\ell$ is an integer. Let $M^* = \max\{M_1, \dots, M_L\}$ and $M_* = \min\{M_1, \dots, M_L\}$. The transmitter picks $\log_2 M^*$ bits from the bit stream. Component $W_\ell$ of the compound message $\mathbf{W}$ corresponds to the decimal expansion of the first $\log_2 M_\ell$ bits from the $\log_2 M^*$ chosen bits. The variable length coding scheme operates as described above. At stopping time $\tau$ the receiver passes $(\hat{W}, \hat{L})$ to the higher-layer application (which can then convert $\hat{W}$ to bits) and the transmitter removes the first $\log_2 M_{\hat{L}}$ bits from the $\log_2 M^*$ initially chosen bits and return the remaining $\log_2 M^* - \log_2 M_{\hat{L}}$ bits to the bit stream. Then, the above process is repeated.

Had we taken the traditional pessimistic point of view, at each state only $\log_2 M_*$ bits would be removed from the bit stream. By taking the opportunistic point of view, when the channel $Q_\circ = Q_\ell$, with high probability we remove $\log_2 M_\ell$ bits from the bit stream. By definition, $M_\ell \geq M_*$. The additional $\log_2 M_\ell - \log_2 M_*$ bits removed at each step quantify the advantage of defining capacity in an opportunistic manner.

The notion of opportunistic capacity is similar to the notion of rateless codes used in fountain codes [9]–[11]. The advantage of modeling capacity as a rate region is that we can more easily talk about the error exponent region of an achievable rate vector, which is the main focus of this paper.

## III. Coding scheme and the main result

Our scheme is based on the Yamamoto-Itoh scheme [6], which asymptotically achieves the Burnashev's exponent for a DMC. We use a training sequence in each phase of Yamamoto-Itoh's scheme to estimate the channel. The design of such a training sequence and the corresponding channel estimate rule falls under the domain of experiment design for parameter estimation. Optimal choice of a training sequence for the channel $\mathscr{Q}$ is beyond the scope of this paper. We assume that such a training sequence can be found easily; if not, we choose a simple training sequence that cycles through all the channel inputs one-by-one. We represent such a sequence of size $n$ by $t^n$.

Given the training sequence, the channel estimation is a multiple hypothesis testing problem. For the analysis presented in this paper, we only care about the asymptotic exponent of the error under each hypothesis. This asymptotic performance has been studied in detail using various approaches but the most relevant for our analysis is the generalization of Blahut's [12] binary hypothesis exponents to multiple hypothesis testing [13].

Consider a multiple hypothesis testing problem where the observations are in $\mathscr{Y}$. According to [13], the error exponents of hypothesis testing lie in a region $\mathscr{T}$ that is given by
$$\mathscr{T} = \{(T_{\ell k}, \ell \neq k) : \forall p \in \Delta(Y), \\ \exists k \text{ such that } \forall \ell \neq k, D(p \| p_\ell) \geq T_{\ell k}\}$$
where $T_{\ell k}$ is the error exponent of estimating hypothesis $\ell$ as hypothesis $k$ and $p_\ell$ is the probability distribution of the observations under hypothesis $\ell$. In the multiple hypothesis testing problem, the observations are assumed to be independent and identically distributed. This is not the case for channel estimation, because the input symbols vary according to $t^n$. Nonetheless, the channel outputs are independent, and it is easy to generalize the above region to the case of independent (but not identically distributed) observations. We are interested in a projection $\mathscr{T}^*$ of the region $\mathscr{T}$ given by
$$\mathscr{T}^* = \{(T_1, \dots, T_L) : \exists (T_{\ell k}, \ell \neq k) \in \mathscr{T} \\ \text{such that } \forall \ell, T_\ell = \min_{k \neq \ell} T_{\ell k}\}$$
For any $(T_1, \dots, T_L)$ in $\mathscr{T}^*$, we can find an estimation rule $\hat{\theta}$ such that for all $n$
$$\mathbb{P}_\ell(\hat{\theta}(Y^n) \neq \ell \mid X^n = t^n) \leq 2^{-nT_\ell}, \quad \ell = 1, \dots, L, \quad (6)$$

where $X^n$ and $Y^n$ denote the channel inputs and outputs respectively.

### A. The coding scheme

In this paper, we only restrict attention to rates on the principle diagonal of the capacity region, *i.e.*, rate vectors of the form $(\gamma C_{Q_1}, \ldots, \gamma C_{Q_L})$, for a given $\gamma \in [0, 1)$. See [14] for a generalization of this coding scheme to all points in the rate region. Let $C_\ell$ denote the capacity $C_{Q_\ell}$ of channel $Q_\ell$. We use a variable length communication scheme that transmits for multiple epochs, where each epoch consists of four phases. The number of epochs is a stopping time.

Below we describe a family of such coding schemes parameterized by $n$, which roughly corresponds to the average length of scheme. We will analyze the performance of this family of schemes for large values of $n$. For a particular value of $n$, the scheme is parameterized by constants $\beta_1(n)$, $\beta_2(\ell, n)$, $\beta_3(n)$, $\beta_4(\ell, n)$, $\ell = 1, \ldots, L$, and channel estimation rules $\hat{\theta}_m(n)$ and $\hat{\theta}_c(n)$. Epoch $k$, $k \in \mathbb{N}$, of the scheme consists of four phases:

1) A fixed length ***training phase*** of length $\lfloor \beta_1(n)n \rfloor$. At the end of this phase both the transmitter and the receiver generate a channel estimate $\hat{L}_m(k, n)$ using rule $\hat{\theta}_m(n)$.
2) A variable length ***message phase*** of length $\lfloor \beta_2(\hat{L}_m(k, n), n)n \rfloor$. Since $\hat{L}_m(k, n)$ is random, this phase is of variable length.
3) A fixed length ***retraining phase*** of length $\lfloor \beta_3(n)n \rfloor$. At the end of this phase both the transmitter and the receiver generate a channel estimate $\hat{L}_c(k, n)$ using rule $\hat{\theta}_c(n)$.
4) A variable length ***control phase*** of length $\lfloor \beta_4(\hat{L}_c(k, n), n) \rfloor$. Since $\hat{L}_c(k, n)$ is random, this phase is of variable length.

Phases one and two of the above scheme correspond to the *message mode* of Yamamoto-Itoh's scheme; phases three and four correspond to the *control mode*.

The length of the scheme depends on $\hat{\theta}_m(n)$ and $\hat{\theta}_c(n)$. We assume that $(T_1^m, \ldots, T_L^m)$ and $(T_1^c, \ldots, T_L^c)$ are the channel estimation exponents of rules $\hat{\theta}_m(n)$ and $\hat{\theta}_c(n)$, respectively. We assume that $\hat{\theta}_m(n)$ and $\hat{\theta}_c(n)$ are chosen such that $T_\ell^m > 0$ and $T_\ell^c > 0$, $\ell = 1, \ldots, L$.

Let $\kappa_\ell = T_\ell^c / B_{Q_\ell}$. Before communication starts, the encoder and the receiver agree upon a reference channel $Q^*$. Let $\kappa^*$ denote the $\kappa$ corresponding to $Q^*$. Now define,

$$\alpha_\ell = (1 + \kappa_\ell)/(1 + \kappa^*).$$

The $\beta$ parameters are chosen such that the expected length of the coding scheme when $Q_\circ = Q_\ell$ is $\alpha_\ell n$. This means that the expected length of the coding sheme under the reference channel $Q^*$ is $n$. For that matter, we choose

1) $\beta_1(n) > 0$, $\lim_{n\to\infty} \beta_1(n) = 0$, and $\lim_{n\to\infty} \beta_1(n)n = \infty$;
2) $\beta_2(\ell, n) > \alpha_\ell \gamma$ and $\lim_{n\to\infty} \beta_2(\ell, n) = \alpha_\ell \gamma$, for all $\ell = 1, \ldots, L$;
3) $\beta_3(n) > 0$ and $\lim_{n\to\infty} \beta_3(n) = \dfrac{(1-\gamma)}{(1+\kappa^*)}$; and
4) $\beta_4(\ell, n) > 0$ and $\lim_{n\to\infty} \beta_4(\ell, n) = \kappa_\ell \dfrac{(1-\gamma)}{(1+\kappa^*)}$, for all $\ell = 1, \ldots, L$.

When there is no ambiguity, we will drop the dependence on $n$ and denote $\beta_1(n)$ by $\beta_1$, $\beta_2(\ell, n)$ by $\beta_2(\ell)$, $\beta_3(n)$ by $\beta_3$ and $\beta_4(\ell, n)$ by $\beta_4(\ell)$. We assume that the $n$ is large enough so that $\lfloor \beta_i n \rfloor \approx \beta_i n$, $i = 1, 2, 3, 4$.

Next we describe each phase of epoch $k$, $k \in \mathbb{N}$, in detail.

1) *Training phase*: The transmitter sends a training sequence $t^{\beta_1 n}$. The transmitter and the receiver use an estimation rule $\hat{\theta}_m(n)$ with the corresponding hypothesis testing exponent $(T_1^m, \ldots, T_L^m)$. Let $\hat{L}_m(k, n)$ denote the channel estimate at the end of the training phase. From (6), we have that
$$\mathbb{P}_\ell(\hat{L}_m(k, n) \neq \ell) \leq 2^{-\beta_1 n T_\ell^m}, \quad \ell = 1, \ldots, L. \quad (7)$$
Note that the channel estimate $\hat{L}_m(k, n)$ depends on only the training sequence of the first phase of epoch $k$; it does not depend on the training sequences of previous epochs.
2) *Message phase*: The transmitter and the receiver agree upon $L$ codebooks. Codebook $\ell$ is of length $\beta_2(\ell)n$ and designed for optimally transmitting $M_\ell(n) = \lfloor 2^{n\alpha_\ell \gamma C_\ell} \rfloor$ messages over channel $Q_\ell$ without feedback, $\ell = 1, \ldots, L$. At the beginning of the second phase, the transmitter uses codebook $\hat{L}_m(k, n)$ to transmit one of $M_{\hat{L}_m(k,n)}(n)$ messages; the receiver decodes according to the same codebook. Let $D(k, n)$ be the indicator function of the event that the decoded message is in error. Then, if the estimation of the first phase is correct, the probability of decoding error is given by
$$\mathbb{E}_\ell[D(k, n) \mid \hat{L}_m(k, n) = \ell] \leq 2^{-\beta_2(\ell)nE_G\left(\alpha_\ell \gamma C_\ell / \beta_2(\ell), Q_\ell\right)} \quad (8)$$
where $E_G(R, Q)$ is Gallager's random coding exponent [15, Theorem 5.6.2] for communicating at rate $R$ over DMC $Q$. Since $\beta_2(\ell) > \alpha_\ell \gamma$, the transmission rate $\alpha_\ell \gamma C_\ell / \beta_2(\ell)$ is less than the capacity $C_\ell$ of the channel $Q_\ell$. So we have
$$E_G(\alpha_\ell \gamma C_\ell / \beta_2(\ell), Q_\ell) > 0. \quad (9)$$
3) *Retraining phase*: The transmitter sends another training sequence $t^{\beta_3 n}$. The transmitter and the receiver use an estimation rule $\hat{\theta}_c(n)$ with the corresponding hypothesis testing exponent $(T_1^c, \ldots, T_L^c)$. Let $L_c(k, n)$ denote the channel estimate at the end of this training phase. From (6), we have that
$$\mathbb{P}_\ell(\hat{L}_c(k, n) \neq \ell) \leq 2^{-\beta_1 n T_\ell^c}, \quad \ell = 1, \ldots, L. \quad (10)$$
Note that the channel estimate $\hat{L}_c(k, n)$ only depends on the training sequence of the third phase of epoch $k$; it does not depend on the training sequence of previous epochs or the training sequence of the first phase of epoch $k$.

4) *Control phase*: Let $x_A(\ell)$ and $x_N(\ell)$ denote the maximally separated input symbols for channel $Q_\ell$, *i.e.*, the arg max in (4) for $B_{Q_\ell}$. From channel feedback, the transmitter knows whether the decoding in the second phase was correct or not. If the decoding was correct, the transmitter sends an ACCEPT consisting of $\beta_4(\hat{L}_c(k,n))n$ repetitions of $x_A(\hat{L}_c(k,n))$; otherwise it sends a REJECT consisting of $\beta_4(\hat{L}_c(k,n))n$ repetitions of $x_N(\hat{L}_c(k,n))$. The decoder assumes that the channel is $\hat{L}_c(k,n)$ and treats detecting an ACCEPT or a REJECT as a binary hypothesis testing problem (with REJECT as the null hypothesis). Let $A(k,n)$ and $N(k,n)$ denote the indicators for whether ACCEPT or REJECT is transmitted, and let $H(k,n)$ denote the indicator that the hypothesis testing is in error. Then, according to [12], there exist estimation regions at the receiver such that

$$\mathbb{E}_\ell[H(k,n) \mid \hat{L}_c(k,n) = \ell, A(k,n) = 1] \leq 2^{-\beta_4 n H_\ell^A(\beta_4 n)} \quad (11)$$

$$\mathbb{E}_\ell[H(k,n) \mid \hat{L}_c(k,n) = \ell, N(k,n) = 1] \leq 2^{-\beta_4 n H_\ell^N(\beta_4 n)} \quad (12)$$

where

$$\lim_{n\to\infty} H_\ell^N(n) = B_{Q_\ell} \quad \text{and} \quad \lim_{n\to\infty} H_\ell^A(n) = 0. \quad (13)$$

To describe the decoding operation, we need two definitions:

**Definition 1** Let $K(n)$ be the epoch when communication stops, *i.e.*, the epoch when the receiver decodes an ACCEPT. Thus,

$$K(n) = \{\inf k \in \mathbb{N} : A(k,n)[1 - H(k,n)] + N(k,n)H(k,n) = 1\}. \quad \square$$

**Definition 2** Let $\Lambda(k,n)$ denote the ratio of the length of phase $k$ and parameter $n$, *i.e.*,

$$\Lambda(k,n) = \beta_1(n) + \beta_2(\hat{L}_m(k,n),n) + \beta_3(n) + \beta_4(\hat{L}_c(k,n),n). \quad \square$$

The final decoding decision at the receiver is $(\hat{L}_m(K(n),n), \hat{W}(K(n),n))$, where $\hat{W}(k,n)$ is the decoding decision at the end of the second phase for epoch $k$.

As in Yamamoto-Itoh's scheme, a decoding error occurs if the decoding in the first phase is incorrect and the subsequent REJECT is decoded as an ACCEPT. All other erroneous situations are corrected by retransmission and increase the communication duration.

### B. Performance Analysis

In this section we present the rate and error exponent of the above scheme. Due to lack of space, the proofs are omitted. See [14] for detailed proofs.

Asymptotically, the number of retransmissions go to zero. Specifically, we have the following.

**Lemma 1** *When $Q_\circ = Q_\ell$, $\ell = 1, \dots, L$, the number $K(n)$ of retransmissions is geometrically distributed with a vanishingly small parameter. Specifically,*

$$\mathbb{E}_\ell[\mathbb{1}\{K(n) = k\}] = (1 - p_\ell(n))p_\ell(n)^{k-1}, \quad k \in \mathbb{N} \quad (14)$$

*where $\lim_{n\to\infty} p_\ell(n) = 0$, $\ell = 1, \dots, L$. Consequently, for asymptotically large values of $n$, there is only one transmission,* i.e.*,*

$$\lim_{n\to\infty} \mathbb{E}_\ell[K(n)] = 1. \quad (15)$$

$\square$

Furthermore, along each sample path, the expected length of phase $k$ is proportional to $n$. Specifically, we have the following.

**Lemma 2** *For all $n \in \mathbb{N}$ and any $k \in \mathbb{N}$, we have that $\mathbb{E}_\ell[\Lambda(k,n)] = \mathbb{E}_\ell[\Lambda(1,n)]$ and*

$$\lim_{n\to\infty} \mathbb{E}_\ell[\Lambda(1,n)] = \alpha_\ell. \quad \square$$

The proposed scheme achieves the rate vector $(\gamma C_1, \dots, \gamma C_L)$. Specifically, we have the following.

**Proposition 1** *The rate of transmission is*

$$\lim_{n\to\infty} \frac{\mathbb{E}_\ell[\log M_{\hat{L}_m(k,n)}(n)]}{\mathbb{E}_\ell[K(n)\Lambda(K(n),n)n]} = \gamma C_\ell \quad (16)$$

$\square$

The above result implies that the rate point $(C_1, \dots, C_L)$ is achievable. Using time sharing, we can achieve all points in the hyper-rectangle with the upper corner $(C_1, \dots, C_L)$. Since, we cannot communicate over channel $Q_\ell$ at rates larger than $C_\ell$, this hyper-rectangle is the opportunistic capacity of the compound channel.

**Corollary 1** *The opportunistic capacity opportunistic capacity region is given by a hyper-rectangle*

$$\mathscr{C}_F(\mathscr{Q}) = \{(R_1, \dots, R_L) : 0 \leq R_\ell < C_\ell,\ \ell = 1, \dots, L\}. \quad \square$$

The error exponent of this scheme is within a constant factor of the Burnashev's exponent when $Q_\circ$ is known.

**Proposition 2** *The error exponent region at rate $(\gamma C_1, \dots, \gamma C_L)$ is given by $(E_1, \dots, E_L)$ such that*

$$E_\ell \geq \frac{T_\ell^c}{T_\ell^c + B_{Q_\ell}} B_{Q_\ell}(1 - \gamma) \quad \square$$

Compare the above error exponent with the trivial upper bound on the error exponent region (5). The exponents of the proposed scheme are at least within a constant factor of the best possible error exponent for the channel.

## IV. An Example

Consider a compound channel consisting of two BSCs with complementary crossover probabilities, $p$ and $(1 - p)$, where $0 < p < 1/2$ and $p$ is known to the transmitter and the receiver. We denote this compound channel by

$$\mathscr{Q}_p := \{BSC_p, BSC_{1-p}\}$$

where $BSC_p$ denotes a binary symmetric channel with crossover probability $p$. For convenience, we will index all variables by $p$ and $(1-p)$ rather than by 1 and 2. For binary symmetric channel, the capacity and $B_Q$ term of Burnashev exponent are given by

$$C_p = C_{1-p} = 1 - h(p) \quad \text{and} \quad B_p = B_{1-p} = D(p\|1-p).$$

where $h(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function and $D(p\|q) = -p\log(p/q) - (1-p)\log((1-p)/(1-q))$ is the binary Kullback-Leibler function.

We use the all zero sequence as a training sequence and a channel estimation rule that chooses $BSC_p$ when the empirical frequency of ones in the channel output is less than 0.5, otherwise it chooses $BSC_{1-p}$. The estimation error probability is bounded by the tail probability of the sum of independent random variables. From Hoeffding's inequality [16, Theorem 1], the exponents of the estimation errors are given by

$$T_p^c = T_{1-p}^c = D(0.5\|p).$$

Suppose we want to communicate at rate $(\gamma C_p, \gamma C_{1-p})$. Due to symmetry, $\kappa_p = \kappa_{1-p}$. Thus, $\alpha_1 = \alpha_2 = 1$, and consequently, $\beta_2(BSC_p, n) = \beta_2(BSC_{1-p}, n)$. Similarly, $\beta_4(BSC_p, n) = \beta_4(BSC_{1-p}, n)$. Thus, the second and forth phase of the coding scheme are of fixed rather than variable length.

If we use the coding scheme of Section III, the lower bound on the error exponent region given by Proposition 2 simplifies to

$$E_p = E_{1-p} \geq \frac{D(0.5\|p)}{D(0.5\|p) + D(p\|1-p)} B_p(1-\gamma) \qquad (17)$$

The best possible error exponent for this channel is not known. Nonetheless, we can compare the error exponent of our scheme with two simple upper bounds. The first upper bound is when the channel is known at the transmitter and the receiver. In this case the error exponent is given by (3). Thus, our exponent is within a factor of

$$\lambda_p = D(0.5\|p)/(D(0.5\|p) + D(p\|1-p))$$

of this upper bound. The second upper bound is given by the zero-rate error exponent for communicating over unknown DMCs (discrete memoryless channel), given by [17]. For $\mathcal{Q}_p$, this upper bound evaluates to $E_0 = \frac{1}{2}B_p$. Since increasing the rate of transmission cannot improve the error exponent, $E_0$ is another upper bound for error exponent.

Combine these two upper bounds to obtain a unified upper bound

$$E_U(\gamma) = B_p\left(1 - \max\left(\tfrac{1}{2}, \gamma\right)\right).$$

Thus, our scheme is at least within a fraction

$$\lambda_p \big/ \left(1 - \max\left(\tfrac{1}{2}, \gamma\right)\right) \qquad (18)$$

of the best possible error exponent.

Both of the above upper bounds bounds are loose. For the channel considered in this paper, no coding scheme can universally achieve Burnashev's exponent (see [8]). Hence, (3) is a loose upper bound. Furthermore, at rates below capacity, the slope of the error exponent is always negative; so, $E_0$ is also a loose upper bound. Thus, the error exponent of our scheme is closer than the fraction (18) of the best possible error exponent.

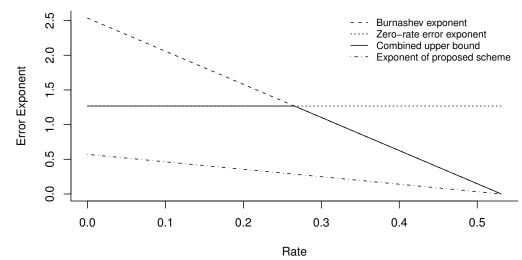


Fig. 1. Performance of the coding scheme for communicating over channel $\mathcal{Q}_{0.1}$. For comparison, different upper bounds on the error exponent are also shown.

## References

[1] J. Wolfowitz, "Simultaneous channels," *Archive for Rantional Mechanics and Analysis*, pp. 371–386, Nov. 1959.
[2] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.
[3] J. Wolfowitz, *Coding Theorems of Information Theory*. Springer Verlag, 1964.
[4] B. Shrader and H. Permuter, "Feedback capacity of the compound channel," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3629–3644, Aug. 2009.
[5] M. V. Burnashev, "Data transmission over a discrete channel with feedback. Random transmission time," *Problemy peredachi informatsii*, vol. 12, no. 4, pp. 10–30, 1976.
[6] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, pp. 729–733, Nov. 1979.
[7] L. Weng, S. S. Pradhan, and A. Anastasopoulos, "Error exponent regions for Gaussian broadcast and multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 2919–2942, Jul. 2008.
[8] A. Tchamkerten and I. E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2126–2145, May 2006.
[9] J. W. Byers, M. Luby, M. Mitzenmacher, and A. Rege, "A digital fountain approach to reliable distribution of bulk data," *SIGCOMM Comput. Commun. Rev.*, vol. 28, no. 4, pp. 56–67, 1998.
[10] M. Luby, "LT codes," in *Proceedings of the IEEE Symposium on Foundations of Computer Science*, 2002, pp. 271–280.
[11] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 6, no. 52, pp. 2551–2567, Jun. 2006.
[12] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, no. 4, Jul. 1974.
[13] E. Tuncel, "On error exponents in hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2945–2950, Aug. 2005.
[14] A. Mahajan and S. Tatikonda, "Opportunistic capacity and error exponent regions for compound channel with feedback," 2010, in preparation.
[15] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.
[16] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, Mar. 1963.
[17] A. Tchamkerten and I. E. Telatar, "On the universality of Burnashev's exponent," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2940–2944, Aug. 2005.