Opportunistic capacity and error exponent regions for compound channel with feedback

Aditya Mahajan and Sekhar Tatikonda
Department of Electrical Engineering,
Yale University, New Haven, CT 06520.
{aditya.mahajan, sekhar.tatikonda}@yale.edu

***Abstract*—Capacity of a compound channel without feedback is defined in a pessimistic manner as the maximum rate determined before the start of communication such that communication is reliable. In the presence of feedback, the transmission rate can adapt to the channel chosen by nature. Thus, capacity can be defined in an opportunistic manner as the maximum rate determined at the end of communication such that communication is reliable. Under this definition, transmission rate and error exponents are regions rather than scalars. In this paper, variable length communication over a compound channel with feedback is formulated, its opportunistic capacity region is characterized, and lower bounds for its error exponent region are provided.**

I. Introduction

A compound channel $\mathscr{Q}$ is a family of DMCs (discrete memoryless channels) defined over a common input and output alphabets $\mathscr{X}$ and $\mathscr{Y}$ (see [1], [2]). Before the start of communication, nature chooses a channel $Q_\circ$ from $\mathscr{Q}$; her choice is not revealed to the encoder or the decoder. 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. Thus, the encoder chooses a channel input P and in response nature chooses the worst $Q_\circ$ from $\mathscr{Q}$.

When channel output feedback is available to the encoder, the encoder can adapt to the choice of $Q_\circ$ by nature. Hence, the capacity is given by (see [4])

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

The above notion of feedback capacity is pessimistic. It quantifies the maximum rate *determined before the start of transmission* such that communication is reliable *over any choice of channel* $Q_\circ$. In many applications, network traffic is backlogged and we do not care about a rate guarantee before the start of transmission. We would rather communicate at the maximum rate such that communication is reliable 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, let $\mathscr{Q} = \{Q_1, \dots, Q_L\}$ be a compound channel. For any coding scheme, let P_ℓ and R_ℓ be the probability of error and transmission rate when $Q_\circ = Q_\ell$, $\ell = 1, \dots, L$. If $P_\ell < \varepsilon$ for $\ell = 1, \dots, L$ and an arbitrarily small ε, then the rate $(R_1, \dots, R_\ell)$ is *achievable*. The union of all achievable rates is called the *opportunistic capacity* $\mathscr{C}(\mathscr{Q})$ of the compound compound channel $\mathscr{Q}$ with feedback, *i.e.*,

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

Thus, in contrast to (2), the opportunistic capacity is a region rather than a scalar value. We formally define achievable rates and opportunistic capacity in Section II.

We show that the opportunistic capacity region is given by a hyper-rectangle

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

which is determined by just its upper corner $(C_{Q_1}, \dots, C_{Q_L})$. Thus, in the presence of feedback, not knowing the channel transition matrix does not result in a loss in maximum transmission rate. The same is not true for error exponents.

For error exponents of DMC, variable length coding significantly improves the reliability of communications. More importantly, this improvement comes at a very little cost: the best error exponents can be achieved by a simple coding scheme [5] that asymptotically has a constant length along almost all sample paths.

In a DMC Q with feedback, the error exponent of variable length coding at a rate $R < C_Q$ is given by (see [6])

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

where

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

$Q(\cdot|x)$ 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. $E_B(R, Q)$, also called the *Burnashev exponent* of channel Q at rate R, 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.

In a compound channel, since the opportunistic capacity $\mathscr{C}_F(\mathscr{Q})$ is a region rather than a scalar, the error exponents at any rate also belong to a region called the *error exponent region* (EER). We formally define this region in Section II.

In Section III, we propose a variable length coding scheme for communicating over a compound channel. The error exponent $(E_1, \ldots, E_L)$ of this coding scheme at any rate $(R_1, \ldots, R_L) \in \mathscr{C}_F(\mathscr{Q})$ is within a multiplicative factor of the Burnashev exponent of DMC $Q_\circ$, *i.e.*

$$E_\ell \geq \alpha B_{Q_\ell}(1 - R_\ell / C_{Q_\ell})$$

where α depends on $\mathscr{Q}$. Thus, this coding scheme retains the main advantage of communicating over of noiseless feedback—the error exponent has a non-zero slope at all rates in the capacity region, including points near the boundary.

The main contributions of this paper are threefold.

1) We define opportunistic capacity and error exponent regions of a compound channel with feedback. These notions are more realistic than the traditional worse case performance guarantees in compound channels.
2) We propose a simple and easy to implement coding scheme whose error exponents are within a multiplicative constant of the best possible error exponents.
3) We show that for variable length communication, explicitly using a training sequence can lead to reasonable error exponents. For example, the error exponent of our proposed scheme have a non-zero slope at all rates in the capacity region.[1]

II. Opportunistic capacity and error exponents

Definition 1 (Variable length coding scheme) A *variable length coding scheme* for communicating over a compound channel $\mathscr{Q} = \{Q_1, \ldots, Q_L\}$ with feedback is a tuple $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$ where

- $\mathbf{M} = (M_1, \ldots, M_L)$ is the *compound message size* where $M_\ell \in \mathbb{N}$, $\ell = 1, \ldots, L$. Define $\mathscr{M} = \prod_{\ell=1}^{L} \{1, \ldots, M_\ell\}$.
- $\mathbf{f} = (f_1, f_2, \ldots)$ is the *encoding strategy* where $f_t : \mathscr{M} \times \mathscr{Y}^{t-1} \mapsto \mathscr{X}$, $t \in \mathbb{N}$ is the encoding function used at time t.
- $\mathbf{g} = (g_1, g_2, \ldots)$ is the *decoding strategy* where $g_t : \mathscr{Y}^t \mapsto \bigcup_{\ell=1}^{L} \{(\ell, 1), (\ell, 2), \ldots, (\ell, M_\ell)\}$, $t \in \mathbb{N}$ is the decoding function at time t.
- τ is the stopping time with respect to the channel outputs Y^t. More precisely, τ 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* $\mathbf{W} = (W_1, \ldots, W_L)$ is generated such that W_ℓ is uniformly distributed in $\{1, \ldots, M_\ell\}$.[2] The transmitter uses the encoding strategy $(f_1, f_2, \ldots)$ to generate channel inputs $X_1 = f_1(\mathbf{W})$, $X_2 = f_2(\mathbf{W}, Y_1)$, … until the stopping time τ with respect to the channel outputs. (τ is known to the transmitter because of feedback.) The decoder then generates a decoding decision $(\hat{L}, \hat{W}) = g_\tau(Y_1, \ldots, Y_\tau)$. 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 $\hat{W} \neq W_{\hat{L}}$. When communication is successful one of $M_{\hat{L}}$ messages is conveyed without error. 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 $\mathbf{R} = (R_1, \ldots, R_L)$ and $\mathbf{P} = (P_1, \ldots, P_L)$, respectively. These are defined as follows.

Definition 2 (Rate) The rate $\mathbf{R} = (R_1, \ldots, R_L)$ of a coding scheme $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$ is given by $R_\ell = \mathbb{E}_\ell[\log M_{\hat{L}}] / \mathbb{E}_\ell[\tau]$ where $\mathbb{E}_\ell[\cdot]$ is a short hand notation for $\mathbb{E}[\cdot | Q_\circ = Q_\ell]$. Note that the R_ℓ component of the rate vector $\mathbf{R}$ depends on the compound message size $\mathbf{M}$ and not just its M_ℓ component. □

Definition 3 (Probability of error) The probability of error $\mathbf{P} = (P_1, \ldots, P_L)$ of a coding scheme $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$ is given by $P_\ell = \mathbb{P}_\ell(\hat{W} \neq W_{\hat{L}})$ where $\mathbb{P}_\ell(\cdot)$ is a short hand notation for $\mathbb{P}(\cdot | Q_\circ = Q_\ell)$. □

Rate and probability of error give rise to two asymptotic performance metrics, *viz.*, achievable rate and error exponents. These are defined as follows.

Definition 4 (Achievable rate) A rate vector $\mathbf{R} = (R_1, \ldots, R_L)$ 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 $\ell = 1, \ldots, L$.
2) For any $\varepsilon > 0$, there exists a $n_\circ(\varepsilon)$ so that for all $n \geq n_\circ(\varepsilon)$, we have $P_\ell^{(n)} < \varepsilon$ and $R_\ell^{(n)} \geq R_\ell - \varepsilon$, for all $\ell = 1, \ldots, L$. □

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.

Definition 5 (Opportunistic Capacity) The union of all achievable compound rates is called the *opportunistic capacity region* of channel $\mathscr{Q}$ with feedback and denoted by $\mathscr{C}_F(\mathscr{Q})$. □

In Corollary 1, we show that $\mathscr{C}_F(\mathscr{Q})$ is given by a hyper-rectangle with upper corner $(C_{Q_1}, \ldots, C_{Q_L})$.

Definition 6 (Error exponents) 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_ℓ of error probability P_ℓ is given by $E_\ell = \lim_{n \to \infty} -\log P_\ell^{(n)} / \mathbb{E}_\ell[\tau^{(n)}]$. Then $\mathbf{E} = (E_1, \ldots, E_L)$ is the error exponent of sequence of coding schemes $(\mathbf{M}^{(n)}, \mathbf{f}^{(n)}, \mathbf{g}^{(n)}, \tau^{(n)})$, $n \in \mathbb{N}$. □

Different sequence of coding schemes that achieve the same rate can have different exponents. Thus, the error exponents of a compound channel with feedback lie in a region, just like the error exponents of multi-terminal communication [8].

[1] This is not a contradiction of the results of [7]. Unlike [7], we allow the coding scheme to vary the number of messages depending on $Q_\circ$.

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

Definition 7 (Error exponent region) For a particular rate $\mathbf{R}$, the union of all possible error exponents is called the *the error exponent region* (EER) of a compound channel with feedback and denoted by $\mathscr{E}(\mathbf{R})$. $\square$

In this paper, we study the EER for all rate of the opportunistic capacity region and present lower bounds on the EER.

The above definitions of variable length communication is different from the traditional definition where M_ℓ is a constant and a single message is communicated (see, for example, [7]). We allow the actual message $\mathbf{W}$ to depend on $Q_\circ$ (even though $Q_\circ$ is not known at the transmitter or the receiver). This allows for an additional degree of freedom in the choice of the coding scheme. However, this additional degree of freedom does not affect the opportunistic capacity region of compound channel; all rates within $\mathscr{C}_F(\mathscr{Q})$ defined above can be achieved using the traditional variable length coding schemes. Neither do we know if this additional degree of freedom improves the EER since the EER of a compound channel has not been investigated using the traditional variable length coding scheme. The main advantage of this additional degree of freedom is that it significantly simplifies the coding scheme.

Operational interpretation

A transmitter has to reliably communicate an infinite bit stream, which is generated by a higher-layer application, to a receiver over a compound channel with feedback. The transmitter uses a variable length coding scheme $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$. For ease of exposition, assume that all M_ℓ, $\ell = 1, \dots, L$, 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. The decimal expansion of the first $\log_2 M_\ell$ of these bits determine the component W_ℓ of $\mathbf{W}$. The message $\mathbf{W}$ is transmitted as described above. At stopping time τ the receiver passes $(\hat{W}, \hat{L})$ to a higher-layer application (which then converts $\hat{W}$ 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.

If the traditional pessimistic approach is followed, only $\log_2 M_*$ bits are removed from the bit stream at each stage. By following the opportunistic approach, with high probability $\log_2 M_\ell$ bits are removed from the bit stream when the channel $Q_\circ = Q_\ell$. By definition, $M_\ell \geq M_*$. Thus, by defining capacity in an opportunistic manner, an additional $\log_2 M_\ell - \log_2 M_*$ bits are removed at each step.

A. A trivial outer bound on error exponents

Any coding scheme $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$ for communicating over a compound channel $\mathscr{Q}$ can also be used to communicate over DMC Q_ℓ. Hence, we have the following trivial upper bound on the EER.

Proposition 1 *For any variable length coding scheme for communicating over $\mathscr{Q}$ at rate $(R_1, \dots, R_L)$, each component of the error exponent region is bounded by the Burnashev exponent of channel Q_ℓ,* i.e.,

$$E_\ell \leq B_{Q_\ell}(1 - R_\ell / C_{Q_\ell}) \qquad \square$$

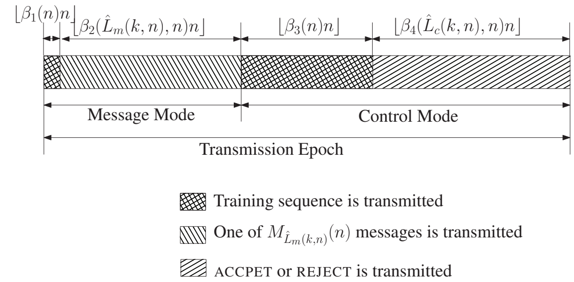

Fig. 1. The four phases of each transmission epoch.

In the remainder of the paper, we try to derive a reasonable lower bound on the EER.

III. The coding scheme

A. The coding scheme

We now describe a family of coding schemes to transmit at a rate vector $(R_1, \dots, R_L)$. Let C_ℓ denote the capacity C_{Q_ℓ} of channel Q_ℓ, and let $\gamma_\ell = R_\ell / C_\ell$. The proposed coding scheme transmits for multiple epochs, where each epoch consists of four phases, two of which are variable length. The number of epochs is a stopping time. The sequence of coding schemes is parameterized by n; for a particular value of n, the scheme is parameterized by constants $\beta_1(n)$, $\beta_2(\ell, n)$, $\beta_3(n)$, $\beta_4(\ell, n)$, $\ell = 1, \dots, L$, and channel estimation rules $\hat{\theta}_m(n)$ and $\hat{\theta}_c(n)$.

The length of the scheme depends on $\hat{\theta}_m(n)$ and $\hat{\theta}_c(n)$. Let $(T_1^m, \dots, T_L^m)$ and $(T_1^c, \dots, T_L^c)$ be the exponents of channel estimation errors 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 $T_\ell^c > 0$, $\ell = 1, \dots, L$.

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

$$\alpha_\ell = \frac{(1 + \kappa_\ell)}{(1 - \gamma_\ell)} \cdot \frac{(1 - \gamma^*)}{(1 + \kappa^*)}.$$

The β 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 scheme under the reference channel is n. We choose $\beta_1(n)$, $\beta_2(\ell, n)$, $\beta_3(n)$, and $\beta_4(\ell, n)$ such that

1) $\beta_1(n) > 0$, $\lim_{n\to\infty} \beta_1(n) = 0$, and $\lim_{n\to\infty} \beta_1(n)n = \infty$;
2) $\beta_2(\ell, n) > \alpha_\ell \gamma_\ell$ and $\lim_{n\to\infty} \beta_2(\ell, n) = \alpha_\ell \gamma_\ell$, for all $\ell = 1, \dots, L$;
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 $\ell = 1, \dots, L$.

When there is no ambiguity, we will drop the dependence on n and denote $\beta_1(n)$ by β_1, $\beta_2(\ell,n)$ by $\beta_2(\ell)$, $\beta_3(n)$ by β_3 and $\beta_4(\ell,n)$ by $\beta_4(\ell)$. We assume that the n is large enough so that $\lfloor \beta_i n \rfloor \approx \beta_i n$, $i = 1,2,3,4$.

Epoch k, $k \in \mathbb{N}$, of the scheme consists of four phases (see Figure 1):

1) *Training phase*: The transmitter sends a training sequence $z^{\beta_1 n}$. The transmitter and the receiver use an estimation rule $\hat{\theta}_m(n)$ with the corresponding hypothesis testing exponent $(T_1^m,\dots,T_L^m)$. Let $\hat{L}_m(k,n)$ denote the channel estimate at the end of the training phase. We have that

$$\mathbb{P}_\ell(\hat{L}_m(k,n) \neq \ell) \leq 2^{-\beta_1 n T_\ell^m}, \quad \ell = 1,\dots,L. \quad (5)$$

2) *Message phase*: The transmitter and the receiver agree upon L codebooks. Codebook ℓ is of length $\beta_2(\ell)n$ and designed for optimally transmitting $M_\ell(n) = \lfloor 2^{n\alpha_\ell \gamma_\ell C_\ell} \rfloor$ messages over channel Q_ℓ without feedback, $\ell = 1,\dots,L$. 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

$$\mathbb{E}_\ell[D(k,n) \mid \hat{L}_m(k,n) = \ell] \leq 2^{-\beta_2(\ell) n E_G\left(\alpha_\ell \gamma_\ell C_\ell / \beta_2(\ell), Q_\ell\right)} \quad (6)$$

where $E_G(R,Q)$ is Gallager's random coding exponent [9, Theorem 5.6.2] for communicating at rate R over DMC Q. Since $\beta_2(\ell) > \alpha_\ell\gamma_\ell$, the transmission rate $\alpha_\ell\gamma_\ell C_\ell/\beta_2(\ell)$ is less than the capacity C_ℓ of the channel Q_ℓ. So we have

$$E_G(\alpha_\ell\gamma_\ell C_\ell/\beta_2(\ell), Q_\ell) > 0. \quad (7)$$

3) *Retraining phase*: The transmitter sends another training sequence $z^{\beta_3 n}$. The transmitter and the receiver use an estimation rule $\hat{\theta}_c(n)$ with the corresponding hypothesis testing exponent $(T_1^c,\dots,T_L^c)$. Let $L_c(k,n)$ denote the channel estimate at the end of this training phase. We have that

$$\mathbb{P}_\ell(\hat{L}_c(k,n) \neq \ell) \leq 2^{-\beta_3 n T_\ell^c}, \quad \ell = 1,\dots,L. \quad (8)$$

4) *Control phase*: Let $x_A(\ell)$ and $x_R(\ell)$ denote the maximally separated input symbols for channel Q_ℓ, *i.e.*, the arg max in (4) for B_{Q_ℓ}. 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 $x_R(\hat{L}_c(k,n))$. 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 $N_A(k,n)$ and $N_R(k,n)$ 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 [10], there exist estimation regions at the receiver such that

$$\mathbb{E}_\ell[H(k,n) \mid \hat{L}_c(k,n) = \ell, N_A(k,n) = 1] \leq 2^{-\beta_4 n H_\ell^A(\beta_4 n)} \quad (9)$$

$$\mathbb{E}_\ell[H(k,n) \mid \hat{L}_c(k,n) = \ell, N_R(k,n) = 1] \leq 2^{-\beta_4 n H_\ell^R(\beta_4 n)} \quad (10)$$

where

$$\lim_{n\to\infty} H_\ell^R(n) = B_{Q_\ell} \quad \text{and} \quad \lim_{n\to\infty} H_\ell^A(n) = 0 \quad (11)$$

To describe the decoding operation, we need two definitions:

Definition 8 Let $K(n)$ be the epoch when communication stops, *i.e.*, the epoch when the receiver decodes an ACCEPT. Thus,

$$K(n) = \{\inf k : N_A(k,n)[1 - H(k,n)] + N_R(k,n)H(k,n) = 1\}. \quad \square$$

Definition 9 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.

IV. Performance analysis

Due to lack of space, we state the simple state the results here without proofs. See [11] for proofs.

A. Some preliminary results

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

Lemma 1 *When $Q_\circ = Q_\ell$, $\ell = 1,\dots,L$,*

$$\mathbb{E}_\ell[\mathbb{1}\{K(n) = k\}] = p_\ell(n)(1 - p_\ell(n))^{k-1}, \quad k \in \mathbb{N} \quad (12)$$

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

$$\lim_{n\to\infty} \mathbb{E}_\ell[K(n)] = 1. \quad (13)$$

$\square$

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*

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

B. Achievability results

The proposed scheme achieves the rate vector $(\gamma_1 C_1, \dots, \gamma_L C_L)$.

Proposition 2 *The rate of transmission is*

$$\lim_{n\to\infty} \frac{\mathbb{E}_\ell[\log M_{\hat{L}_m(k,n)}(n)]}{\mathbb{E}_\ell[K(n)\Lambda(K(n),n)n]} = \gamma_\ell C_\ell \qquad (14)$$

□

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

Proposition 3 *The error exponent region at rate $(\gamma_1 C_1, \dots, \gamma_L C_L)$ is given by $(E_1, \dots, E_L)$ such that*

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

□

The above result implies that the rate point $(C_1, \dots, C_L)$ is achievable.

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

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

□

V. An example

Consider the compound channel $\mathscr{Q} := \{BSC_p, BSC_{1-p}\}$, where BSC_p denotes a binary symmetric channel with crossover probability p. Assume that p is known to the encoder and the decoder. For convinience, we index all variables by p and $1-p$ rather than 1 and 2. The capacity and B_Q term of Burnashev exponent are given by $C_p = C_{1-p} = 1 - h(p)$ and $B_p = B_{1-p} = D(p\|1-p)$ 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 choose the all zero sequence as a training sequence and estimate the channel based on the type of the output sequence. If the empirical frequency of ones in the output is less than q, $p < q < 1-p$, the channel is estimated as BSC_p; otherwise the channel is estimated as BSC_{1-p}. For this class of channel estimation rules, the estimation error probability is bounded by the tail of the probability of the sum of independent random variables. From Hoeffding's inequality [12, Theorem 1], the exponents of the estimation errors are given by $T_p = D(q\|p)$ and $T_{1-p} = D(q\|1-p)$.

Suppose we want to communicate at rate (R_p, R_{1-p}), $R_p < C_p$ and $R_{1-p} < C_{1-p}$, using the coding scheme of Section III. Let q_m and q_c be the estimation thresholds for the message and control mode. The lower bound of Proposition 3 simplifies to

$$E_p \geq \frac{D(q_c\|p)D(p\|1-p)}{D(q_c\|p) + D(p\|1-p)}(1-\gamma_p),$$
$$E_{1-p} \geq \frac{D(q_c\|1-p)D(p\|1-p)}{D(q_c\|1-p) + D(p\|1-p)}(1-\gamma_{1-p})$$

where $\gamma_p = R_p/C_p$ and $\gamma_{1-p} = R_{1-p}/C_{1-p}$.

The choice of q_m does not affect the values of E_p and E_{1-p} as long as $T_m > 0$. For that, we require only that $p < q_m < 1-p$. Choosing $q_m = 0.5$ ensures that.

We want to choose q_c such that $E_p = E_{1-p}$ Thus, choosing $q_c = 0.5$, which maximally distinguishes between BSC_p and BSC_{1-p}, is optimal only when $\gamma_p = \gamma_{1-p}$. For other values of γ_p and γ_{1-p}, the optimal value of q_c is determined by inverting $\varphi(q_c, p)$, where

$$\varphi(q,p) = \frac{1 + D(p\|1-p)/D(q\|p)}{1 + D(p\|1-p)/D(q\|1-p)} \qquad (15)$$

VI. Conclusion

In the presence of feedback, not knowing the exact channel transition matrix does not result in a loss in capacity. As a result, we can provide an optimistic rate guarantee for any rate less than the capacity of the actual channel, even though we do not know the actual channel before the start of communication. This is in contrast to the pessimistic rate guarantees in compound channel without feedback. More importantly, any rate vector in the optimistic capacity region can be achieved using a simple, training- based coding scheme. The error exponent of this scheme has a negative slope at all rates in the capacity region, even at rates near the boundary of the capacity region.

References

[1] J. Wolfowitz, "Simultaneous channels," *Archive for Rational 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] 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.

[6] 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.

[7] A. Tchamkerten and I. E. Telatar, "On the use of training sequences for channel estimation," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1171–1176, Mar. 2006.

[8] 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.

[9] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.

[10] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, no. 4, Jul. 1974.

[11] A. Mahajan and S. Tatikonda, "Opportunistic capacity and error exponent regions for compound channel with feedback," 2010, submitted to IEEE Trans. Inf. Theory. [Online]. Available: http://arxiv.org/abs/0911.2023

[12] 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.