

Opportunistic Capacity and Error Exponent Region for the Compound Channel With Feedback

Aditya Mahajan, *Member, IEEE*, and Sekhar Tatikonda, *Member, IEEE*

Abstract—Variable length communication over a compound channel with feedback is considered. Traditionally, capacity of a compound channel without feedback is defined as the maximum rate that is determined before the start of communication such that communication is reliable. This traditional definition is pessimistic. In the presence of feedback, an opportunistic definition is given. Capacity is defined as the maximum rate that is determined at the end of communication such that communication is reliable. Thus, the transmission rate can adapt to the realized channel. Under this definition, feedback communication over a compound channel is conceptually similar to multiterminal 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. 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.

Index Terms—Burnashev exponent, error exponent region (EER), feedback error exponent, variable-length coding, variable-rate coding.

I. INTRODUCTION

THE compound channel, first considered by Wolfowitz [1] and Blackwell *et al.* [2], is one of the simplest extensions of the discrete memoryless channel (DMC). In a compound channel, the channel transition matrix Q_\circ belongs to a family \mathcal{C} that is defined over a common discrete input and discrete output alphabets \mathcal{X} and \mathcal{Y} . The transmitter and the receiver know the *compound family* \mathcal{C} but do not know the *realized channel* Q_\circ ; the realized channel Q_\circ does not change with time. We are interested in characterizing the error exponents of a compound channel used with feedback. For that purpose, we define a new notion of the capacity of the compound channel with feedback.

There have been comprehensive investigations on the capacity of compound channels, used both with and without feedback. In addition, there is some work on characterizing the error exponent of compound channels used with feedback. We briefly summarize the existing work below, focusing on finite compound families $\mathcal{C} = \{Q_1, \dots, Q_L\}$.

Given a coding scheme $S^{(n)}$ defined over a compound family \mathcal{C} , let $P_\ell^{(n)}$ and $R_\ell^{(n)}$ denote the probability of error and transmission rate when the realized channel Q_\circ is Q_ℓ , $\ell = 1, \dots, L$. The general notion of capacity of a compound channel is as follows: a rate R is said to be *achievable* if $\forall \varepsilon > 0$, \exists a sequence $S^{(n)}$ of coding schemes such that $P_\ell^{(n)} < \varepsilon$ and $R_\ell^{(n)} > R - \varepsilon$, $\ell = 1, \dots, L$. Then, the capacity is the supremum of all achievable rates. This same notion applies when the channel is used without or with feedback (the difference being in the choice of coding schemes $S^{(n)}$).

When the compound channel is used without feedback, the capacity is given by (see [3])

$$\mathcal{E}_{NF}(\mathcal{C}) = \max_{P \in \Delta(\mathcal{X})} \inf_{Q \in \mathcal{C}} I(P, Q) \quad (1)$$

where $\Delta(\mathcal{X})$ is the space of probability distributions on input alphabet \mathcal{X} and

$$I(P, Q) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x) \log \frac{Q(y|x)}{\sum_{x' \in \mathcal{X}} Q(y|x') P(x')}$$

is the mutual information between the input and output of a channel with input distribution P and channel transition matrix Q . When the compound channel is used with feedback, the capacity is given by (see [4])

$$\mathcal{E}_F(\mathcal{C}) = \inf_{Q \in \mathcal{C}} \max_{P \in \Delta(\mathcal{X})} I(P, Q). \quad (2)$$

These and other variations of the compound channel are surveyed in [5].

The aforementioned notion of capacity is pessimistic. It quantifies the maximum rate determined before the start of transmission such that communication is reliable over every realized channel Q_\circ . An opportunistic definition of feedback is possible in the presence of feedback.

For many applications, network traffic is backlogged and a rate guarantee before the start of transmission is not critical. Rather, we want to communicate at the maximum rate while ensuring that communication is reliable for the realized channel Q_\circ (even though Q_\circ is not known to the transmitter or the receiver before the start of transmission). In particular, instead of modeling achievable rate as a scalar value R that is guaranteed before the start of communication, we model achievable rate as a vector (R_1, \dots, R_L) such that the rate of communication is R_ℓ when the realized channel is Q_ℓ . In addition, communication is reliable for every realized channel. More precisely, we say that a rate vector (R_1, \dots, R_L) is *opportunistically achievable* if $\forall \varepsilon > 0$, \exists a sequence $S^{(n)}$ of coding schemes such that

Manuscript received April 06, 2010; revised June 29, 2011; accepted January 09, 2012. Date of publication March 21, 2012; date of current version June 12, 2012. This paper was presented in part at the 2010 IEEE International Symposium on Information Theory.

A. Mahajan was with the Department of Electrical Engineering, Yale University, New Haven, CT 06520 USA. He is now with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A2A7, Canada (e-mail: aditya.mahajan@mcgill.ca).

S. Tatikonda is with the Department of Electrical Engineering, Yale University, New Haven, CT 06520 USA (e-mail: sekhar.tatikonda@yale.edu).

Communicated by S. Diggavi, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2012.2191689

$P_\ell^{(n)} < \varepsilon$ and $R_\ell^{(n)} > R_\ell - \varepsilon, \ell = 1, \dots, L$. We define the union of all opportunistically achievable rates as the opportunistic capacity region $\mathcal{E}_{OF}(\mathcal{C})$, i.e.,

$$\mathcal{E}_{OF}(\mathcal{C}) = \{(R_1, \dots, R_L) : (R_1, \dots, R_L) \text{ is opportunistically achievable}\}. \quad (3)$$

We formally define opportunistically achievable rates and opportunistic capacity in Section II.

Let C_ℓ denote the capacity of DMC $Q_\ell, \ell = 1, \dots, L$. Then, it is straightforward to show (see Corollary 1) that the opportunistic capacity region is given by a hyper-rectangle

$$\mathcal{E}_{OF}(\mathcal{C}) = \{(R_1, \dots, R_L) : 0 \leq R_\ell < C_\ell, \ell = 1, \dots, L\}$$

which is determined by just its upper corner (C_1, \dots, C_L) . Thus, the capacity region $\mathcal{E}_{OF}(\mathcal{C})$ is equivalent to the capacity vector $\mathcal{C} := (C_1, \dots, C_L)$.

In this paper, we consider variable length coding schemes. For a sequence $\{S^{(n)}\}$ of coding schemes that (opportunistically) achieves a rate vector (R_1, \dots, R_L) , we define the error exponent vector (E_1, \dots, E_L) as

$$E_\ell = \lim_{n \rightarrow \infty} \frac{-\log P_\ell^{(n)}}{\mathbb{E}_\ell[\tau^{(n)}]}$$

where $\mathbb{E}_\ell[\tau^{(n)}]$ is the expected length of the coding scheme $S^{(n)}$ when the realized channel is Q_ℓ . The union of all achievable error exponent vectors is defined as the error exponent region (EER) at rate (R_1, \dots, R_L) and denoted by $\mathcal{E}(R_1, \dots, R_L)$. The formal definition is presented in Section II.

Consider a DMC Q used with feedback. Let C_Q denote its capacity. The error exponent of variable length coding scheme at rate $R < C_Q$ is given by (see [6])

$$E_B(R, Q) = B_Q(1 - R/C_Q) \quad (4)$$

where

$$B_Q = \max_{x_A, x_B \in \mathcal{X}} b_Q(x_A, x_B) \quad (5)$$

$$b_Q(x_A, x_B) = D(Q(\cdot|x_A) \| Q(\cdot|x_B)). \quad (6)$$

$Q(\cdot|x)$ is the probability distribution of the channel output when the channel input is x , and

$$D(p \| q) = \sum_{y \in \mathcal{Y}} p(y) \log \frac{p(y)}{q(y)}$$

is the Kullback–Leibler divergence between probability distributions p and q . We call $E_B(R, Q)$ as the *Burnashev exponent* of channel Q at rate R and B_Q as the zero rate Burnashev exponent.

One of the key features of the Burnashev exponent is that it has a nonzero slope at capacity. This slope captures the main advantage of feedback—by reducing the transmission rate by a small *fraction* of the capacity, we linearly increase the error exponent, and therefore, exponentially decrease the probability of error. Does feedback provide the same advantage for a compound channel?

Clearly, a particular component E_ℓ of the EER of the compound channel cannot beat the Burnashev exponent for

DMC Q_ℓ . Thus, a trivial upper bound for the EER at rate $(R_1, \dots, R_L) \in \mathcal{E}_{OF}(\mathcal{C})$ is the hyper-rectangle with upper corner

$$(B_{Q_1}(1 - R_1/C_{Q_1}), \dots, B_{Q_L}(1 - R_L/C_{Q_L})). \quad (7)$$

Tchamkerten and Telatar [7] showed that this bound is not tight by means of a simple counterexample. They considered a compound family consisting of two binary symmetric channels with complementary cross-over probabilities, p and $(1 - p)$, where p is known to the transmitter and the receiver. They showed that, even for this simple family, no coding scheme universally achieves the Burnashev exponent.

Another way to interpret that result is that the EER need not be a hyper-rectangle *i.e.*, for a fixed rate $\mathbf{R} = (R_1, \dots, R_L)$ if $(E'_1, \dots, E'_L), (E''_1, \dots, E''_L) \in \mathcal{E}(\mathbf{R})$, then it is *not necessary* that

$$(\max(E'_1, E''_1), \dots, \max(E'_L, E''_L)) \in \mathcal{E}(\mathbf{R}).$$

Thus, different sequence of coding schemes that achieve the same rate vector (R_1, \dots, R_L) may have different and *noncomparable* error exponents. Thus, in terms of error exponents, the compound channel with feedback behaves in a manner similar to multiterminal communication channels [8].

Tchamkerten and Telatar [7] also identified necessary and sufficient conditions on the compound family \mathcal{C} under which the upper bound of (7) is tight for all rates along the principle diagonal $(\gamma C_{Q_1}, \dots, \gamma C_{Q_L}), 0 \leq \gamma < 1$, of the opportunistic capacity region. For channels that do not satisfy these conditions, the EER is not characterized. Even when these conditions are satisfied, the EER is not characterized for rate vectors that are off the principle diagonal (*i.e.*, R_ℓ/C_{Q_ℓ} is not constant for all $\ell = 1, \dots, L$). In Section III, we present a coding scheme for all rates in the opportunistic capacity region. This scheme achieves an error exponent with a nonzero slope at all points in the rate region, including points near the capacity boundary. This shows that feedback provides similar advantage for a compound channel as for a DMC.

A. Notation

We use the following notation in this paper. $\Delta(\mathcal{X})$ denotes the space of probability distributions over \mathcal{X} . \mathbb{N} denotes the set of natural numbers. $\mathbb{P}(\cdot)$ denotes the probability of an event, $\mathbb{E}[\cdot]$ denotes the expectation of a random variable, and $\mathbf{1}\{\cdot\}$ denotes the indicator function. All logarithms are to the base 2, and $\exp_2(\cdot)$ denotes $2^{(\cdot)}$.

C_Q denotes the capacity of a DMC with transition matrix Q ; B_Q denotes its zero-rate Burnashev exponent. Given a compound family $\mathcal{C} = \{Q_1, \dots, Q_L\}$, Q_o denotes the realized channel; C_ℓ denotes the capacity C_{Q_ℓ} of DMC Q_ℓ ; B_ℓ denotes the zero-rate Burnashev exponent of DMC Q_ℓ . $\mathbb{P}_\ell(\cdot)$ is short hand for $\mathbb{P}(\cdot|Q_o = Q_\ell)$; and $\mathbb{E}_\ell[\cdot]$ is a short hand for $\mathbb{E}_\ell[\cdot|Q_o = Q_\ell]$.

II. OPPORTUNISTIC CAPACITY AND ERROR EXPONENTS

In this section, we formally define opportunistic capacity and EERs for a compound channel with feedback. Conceptually, it is easier to first define achievable rate vector for fixed length communication and then extend that definition to variable

length communication. However, for succinctness, we only define achievable rate vector for variable length communication.

Definition 1 (Variable-Rate Variable-Length Coding Scheme): A variable-rate variable-length coding scheme for communicating over a compound channel $\mathcal{C} = \{Q_1, \dots, Q_L\}$ with feedback is a tuple $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$ where

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

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

is the encoding function used at time t .

- 3) $\mathbf{g} = (g_1, g_2, \dots)$ is the *decoding strategy* where

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

is the decoding function at time t .

- 4) 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 $\{2^{\mathcal{Y}^t}, t \in \mathbb{N}\}$.

□

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_ℓ is uniformly distributed in $\{1, \dots, M_\ell\}$.¹ The transmitter uses the encoding strategy (f_1, f_2, \dots) to generate channel inputs

$$X_1 = f_1(\mathbf{W}), \quad X_2 = f_2(\mathbf{W}, Y_1), \quad \dots$$

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

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

The decoding decision consists of two components: the index \hat{L} of decoded component and an estimate \hat{W} of the \hat{L} -component of the compound message \mathbf{W} . A communication error occurs if $\hat{W} \neq W_{\hat{L}}$.

Remark 1: The aforementioned scheme is a variable-rate variable-length coding scheme. The transmitter and receiver agree upon the set of rates $\{R_1, \dots, R_L\}$ before the start of communication. The transmitter chooses L different messages, one message for each rate. At the end of communication, the receiver decides the message \hat{L} it wants to decode and generates an estimate \hat{W} for that message. Because of noiseless feedback, the encoder knows what the decoder decoded. In principle, the index \hat{L} need not be the same as the index ℓ of the realized channel. For that reason, $\{\hat{L} \neq \ell\}$ is not considered a communication error.

□

The two main performance metrics of a coding scheme are its error probability and rate, both of which are vectors (rather

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

than scalars), and denoted by $\mathbf{P} = (P_1, \dots, P_L)$ and $\mathbf{R} = (R_1, \dots, R_L)$, respectively. These are defined as follows.

Definition 2 (Probability of Error): A communication error occurs when $\hat{W} \neq W_{\hat{L}}$. The probability of error $\mathbf{P} = (P_1, \dots, P_L)$ 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 $\mathbb{P}(\cdot | Q_o = Q_\ell)$.

□

Definition 3 (Rate): The rate $\mathbf{R} = (R_1, \dots, R_L)$ of a coding scheme $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$ is given by

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

where $\mathbb{E}_\ell[\cdot]$ is a short hand notation for $\mathbb{E}[\cdot | Q_o = Q_\ell]$.

□

Remark 2: The aforementioned scheme is a variable rate communication scheme. The size $M_{\hat{L}}$ of the communicated message $W_{\hat{L}}$ is a random variable taking values in $\{M_1, \dots, M_L\}$. For that reason, we define the rate as $\mathbb{E}_\ell[\log M_{\hat{L}}] / \mathbb{E}_\ell[\tau]$. When all rates $\{R_1, \dots, R_L\}$ are equal, the above scheme reduces to a fixed-rate variable-length coding scheme and the definition of rate in Definition 4 collapses to the traditional definition of fixed-rate variable-length coding.

□

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

Definition 4 (Opportunistically Achievable Rate): A rate vector $\mathbf{R} = (R_1, \dots, R_L)$ is said to be *opportunistically achievable* if there exists a sequence of variable-rate variable-length coding schemes $(\mathbf{M}^{(n)}, \mathbf{f}^{(n)}, \mathbf{g}^{(n)}, \tau^{(n)})$, $n \in \mathbb{N}$ such that:

- 1) $\lim_{n \rightarrow \infty} \mathbb{E}_\ell[\tau^{(n)}] = \infty$ for $\ell = 1, \dots, L$.
- 2) For every $\varepsilon > 0$, there exists a $n_o(\varepsilon)$ so that for every $n \geq n_o(\varepsilon)$, we have

$$P_\ell^{(n)} < \varepsilon \quad \text{and} \quad R_\ell^{(n)} > R_\ell - \varepsilon, \quad \text{for all } \ell = 1, \dots, L;$$

or equivalently

$$\lim_{n \rightarrow \infty} P_\ell^{(n)} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} R_\ell^{(n)} = R_\ell.$$

□

Definition 5 (Opportunistic Capacity): The union of all opportunistically achievable rates is called the opportunistic capacity region of the compound channel \mathcal{C} with feedback and denoted by $\mathcal{E}_{OF}(\mathcal{C})$.

□

In Corollary 1, we show that $\mathcal{E}_{OF}(\mathcal{C})$ is given by a hyperrectangle with upper corner $(C_{Q_1}, \dots, C_{Q_L})$. For that reason, we call $\mathcal{C}_\mathcal{C} := (C_{Q_1}, \dots, C_{Q_L})$ as the *capacity vector* of the compound channel \mathcal{C} .

The variable-rate variable-length coding scheme defined above is related to the notion of rateless codes used in fountain codes [9]–[11] for BER (binary erasure channel).

Definition 6 (Error Exponent): 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

$$E_\ell = \lim_{n \rightarrow \infty} -\frac{\log P_\ell^{(n)}}{\mathbb{E}_\ell[\tau^{(n)}]}.$$

Then, $\mathbf{E} = (E_1, \dots, E_L)$ is the error exponent of the sequence of coding schemes $(\mathbf{M}^{(n)}, \mathbf{f}^{(n)}, \mathbf{g}^{(n)}, \tau^{(n)})$, $n \in \mathbb{N}$. \square

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

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

The aforementioned scheme describes a variable-rate variable-length coding scheme; varying the rate of the coding scheme allows for an additional degree of freedom. This additional freedom does not affect the opportunistic capacity region of compound channel; all rates within $\mathcal{E}_{OF}(\mathcal{C})$ defined above can be achieved using a fixed-rate variable length coding scheme. We do not know if this additional degree of freedom improves the EER since the EER of a compound channel has not been investigated using the traditional fixed-rate variable-length coding scheme. The reason that we chose a variable-rate coding scheme is that this additional degree of freedom significantly simplifies the coding scheme.

A. 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-rate variable-length coding scheme $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$. For ease of exposition, assume that every M_ℓ , $\ell = 1, \dots, L$, is a power of 2 so that $\log M_\ell$ is an integer. Let $M^* = \max\{M_1, \dots, M_L\}$ and $M_* = \min\{M_1, \dots, M_L\}$. The transmitter picks $\log M^*$ bits from the bit stream. The decimal expansion of the first $\log M_\ell$ of these bits determine the component W_ℓ of \mathbf{W} . The message \mathbf{W} is transmitted as described previously. At stopping time τ the receiver passes (\hat{L}, \hat{W}) to a higher layer application (which then converts \hat{W} to bits) and the transmitter removes the first $\log M_{\hat{L}}$ bits from the $\log M^*$ initially chosen bits and return the remaining $\log M^* - \log M_{\hat{L}}$ bits to the bit stream. Then, the aforementioned process is repeated.

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

B. Trivial Outer Bound on Error Exponents

Any coding scheme $(\mathbf{M}, \mathbf{f}, \mathbf{g}, \tau)$ for communicating over a compound channel \mathcal{C} 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-rate variable-length coding scheme for communicating over \mathcal{C} at rate (R_1, \dots, R_L) , each component of the EER is bounded by the Burnashev exponent of channel Q_ℓ , i.e.,

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

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

III. CODING SCHEME

In this section, we define a family of variable-rate variable-length coding schemes indexed by $n \in \mathbb{N}$. As $n \rightarrow \infty$, the scheme opportunistically achieves a rate vector (R_1, \dots, R_L) . This coding scheme is based on the Yamamoto–Itoh [12] scheme that achieves the Burnashev exponent for DMC.

A. Parameters of the Coding Scheme

For each $n \in \mathbb{N}$, the scheme is parameterized by the following nonnegative real constants:²

$$\alpha_m^{(n)}, \alpha_c^{(n)}, \text{ and } \beta_{m,\ell}^{(n)}, \beta_{c,\ell}^{(n)}, \xi_\ell^{(n)}, \quad \ell = 1, \dots, L.$$

We will explain the purpose and choice of these constants later. For now, we assume that $\alpha_m^{(n)}$, $\alpha_c^{(n)}$, $\beta_{m,\ell}^{(n)}$, and $\beta_{c,\ell}^{(n)}$ are chosen such that $\alpha_m^{(n)}n$, $\alpha_c^{(n)}n$, $\beta_{m,\ell}^{(n)}n$, and $\beta_{c,\ell}^{(n)}n$ are integers. When there is no ambiguity, we will not explicitly show the dependence on n and drop the superscripts $^{(n)}$.

For each n , the encoder and the decoder agree upon the following:

- 1) *Two training sequences* σ_m and σ_c of lengths $\alpha_m n$ and $\alpha_c n$ and *corresponding channel estimation rules* $\hat{\theta}_m$ and $\hat{\theta}_c$.
- 2) *L codebooks*; one for each Q_ℓ , $\ell = 1, \dots, L$. Codebook ℓ has rate $\xi_\ell R_\ell / \beta_{m,\ell}$ and length $\beta_{m,\ell} n$.
- 3) *2L control sequences*; two for each Q_ℓ , $\ell = 1, \dots, L$, viz. $\sigma_{A,\ell}$ and $\sigma_{R,\ell}$, both of length $\beta_{c,\ell} n$ and *corresponding hypothesis testing rules* $\hat{\theta}_{H,\ell}$ for disambiguating $\sigma_{A,\ell}$ and $\sigma_{R,\ell}$ over DMC Q_ℓ .

A compound message $\mathbf{W}^{(n)}$ is chosen at random such that component $W_\ell^{(n)}$, $\ell = 1, \dots, L$, is uniformly distributed over $\{1, \dots, \exp_2(n\xi_\ell R_\ell)\}$.⁴

B. Operation of the Coding Scheme

The coding scheme transmits in multiple epochs indexed by $k \in \mathbb{N}$. Each epoch consists of four phases.

- 1) *A fixed length training phase* of length $\alpha_m n$. During this phase the transmitter sends the training sequence σ_m ; both the transmitter and the receiver use the estimation rule $\hat{\theta}_m$ to determine a channel estimate $\hat{L}_m = \hat{L}_m(k)$.
- 2) *A variable length message phase* of length $\beta_{m,\hat{L}_m} n$. The transmitter and receiver use codebook \hat{L}_m to send component \hat{L}_m of the compound message $\mathbf{W}^{(n)}$. Let $W_m(k)$

²The subscripts stand for *message* and *control*.

³The subscripts stand for *accept* and *reject*.

⁴The joint distribution of $(W_1^{(n)}, \dots, W_L^{(n)})$ does not matter.

denote the transmitted message and $\hat{W}_m(k)$ the decoded message.

- 3) A *fixed length re-training phase* of length $\alpha_c n$. During this phase the transmitter sends the training sequence σ_c ; both the transmitter and the receiver use the estimation rule $\hat{\theta}_c$ to determine a channel estimate $\hat{L}_c = \hat{L}_c(k)$.
- 4) A *variable length control phase* of length $\beta_{c, \hat{L}_c} n$. If $W_m(k) = \hat{W}_m(k)$, the transmitter sends a control message $W_c(k) = \sigma_{A, \hat{L}_c}$; otherwise it sends $W_c(k) = \sigma_{R, \hat{L}_c}$. The receiver decodes the *control message* using $\hat{\theta}_{H, \hat{L}_c}$.

Let $\hat{W}_c(k)$ denote the estimated control message.

If $\hat{W}_c(k) = \sigma_{A, \hat{L}_c}$, then transmission stops and the receiver declares $(\hat{L}_m(k), \hat{W}_m(k))$ as its final decision; otherwise, the compound message is retransmitted in the next epoch. Let $K^{(n)}$ denote the epoch when communication stops, *i.e.*,

$$K^{(n)} = \inf\{k \in \mathbb{N} : \hat{W}_c(k) = \sigma_{A, \hat{L}_c(k)}\}.$$

Let the length of epoch k be $\Lambda^{(n)}(k)n$, *i.e.*,

$$\Lambda^{(n)}(k)n = \alpha_m^{(n)}n + \beta_{m, \hat{L}_m(k)}^{(n)} + \alpha_c^{(n)}n + \beta_{c, \hat{L}_c(k)}^{(n)}.$$

Hence, the length of communication is

$$\tau^{(n)} = \sum_{k=1}^{K^{(n)}} \Lambda^{(n)}(k)n.$$

C. Choice of Training Sequences

As described earlier, the transmitter and the receiver agree upon two training sequences, σ_m and σ_c , of lengths $\alpha_m n$ and $\alpha_c n$, respectively. The optimal choice of such training sequences falls under the domain of experiment design for estimating unknown parameters. We assume that we can find good training sequences for \mathcal{C} ; if not, we choose a simple training sequence that cycles through all the channel inputs one-by-one.

The transmitter and the receiver also agree upon two estimating rules, $\hat{\theta}_m$ and $\hat{\theta}_c$. For a training sequence σ of size n and an estimation rule $\hat{\theta}$, define the estimation error exponent as follows. For $k, \ell = 1, \dots, L$,

$$T^{\ell, k} = \lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}_\ell(\hat{\theta}(Y^n) = k \mid X^n = \sigma) \quad (8)$$

and for $\ell = 1, \dots, L$

$$\begin{aligned} T_\ell &= \lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}_\ell(\hat{\theta}(Y^n) \neq \ell \mid X^n = \sigma) \\ &= \min\{T^{\ell, k} : k = 1, \dots, L, k \neq \ell\} \end{aligned} \quad (9)$$

where X^n and Y^n are the channel inputs and outputs, respectively. We are interested in characterizing the union of (T_1, \dots, T_L) for all choices of estimation rule $\hat{\theta}$. We call this region the estimation EER and denote it by \mathcal{F} . Instead of directly characterizing estimation EER, it is easier to first characterize pairwise estimation EER—the union of $(T^{\ell, k}; \ell, k = 1, \dots, L;$

$k \neq \ell)$ for all choices of estimation rule $\hat{\theta}$; this region is denoted by \mathcal{F}^* —and then obtain the estimation error estimation region \mathcal{F} using (9).

Characterizing the pairwise estimation error exponent is equivalent to characterizing the pairwise hypothesis testing exponent for multiple hypothesis testing. The latter was characterized by Tuncel [13] for L -ary hypothesis testing with independent and identically distributed observations. Let p_ℓ be the probability distribution of the observations under hypothesis ℓ . Then

$$\begin{aligned} \mathcal{F}^* &= \{(T^{\ell, k}, \ell \neq k) : \forall p \in \Delta(Y), \exists k \text{ such that} \\ &\quad D(p \| p_\ell) \geq T^{\ell, k} \text{ for all } \ell \neq k\}. \end{aligned}$$

For our setup, the observations at the receiver need not be identically distributed. Nonetheless, the observations are independent across time, and it is easy to generalize the above region to the case of independent (but not identically distributed) observations. We then use (9) to obtain the desired region \mathcal{F} as follows:

$$\begin{aligned} \mathcal{F} &= \{(T_1, \dots, T_L) : \exists (T^{\ell, k}, \ell \neq k) \in \mathcal{F}^* \text{ such that} \\ &\quad \forall \ell, T_\ell = \min_{k \neq \ell} T_{\ell k}\}. \end{aligned}$$

The estimation rules $\hat{\theta}_m$ and $\hat{\theta}_c$ attain particular points in \mathcal{F} ; denote these by $(T_{m,1}, \dots, T_{m,L})$ and $(T_{c,1}, \dots, T_{c,L})$, respectively. Recall that the training sequences σ_m and σ_c are of length $\alpha_m n$ and $\alpha_c n$, respectively. Thus, for any epoch k ,

$$\lim_{n \rightarrow \infty} -\frac{1}{\alpha_m n} \log \mathbb{P}_\ell(\hat{L}_m \neq \ell) = T_{m, \ell}, \quad \ell = 1, \dots, L \quad (10)$$

and

$$\lim_{n \rightarrow \infty} -\frac{1}{\alpha_c n} \log \mathbb{P}_\ell(\hat{L}_c \neq \ell) = T_{c, \ell}, \quad \ell = 1, \dots, L. \quad (11)$$

Choose $\hat{\theta}_m$ and $\hat{\theta}_c$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P}_\ell(\hat{L}_m \neq \ell) = 0, \quad \lim_{n \rightarrow \infty} \mathbb{P}_\ell(\hat{L}_c \neq \ell) = 0 \quad (12)$$

and

$$T_{m, \ell} \geq 0, \quad T_{c, \ell} > 0, \quad \ell = 1, \dots, L. \quad (13)$$

D. Choice of Codebooks

As described earlier, the transmitter and receiver agree upon L codebooks. Codebook ℓ is a fixed length codebook for DMC Q_ℓ , $\ell = 1, \dots, L$, with rate $\xi_\ell R_\ell / \beta_{m, \ell}$ and length $\beta_{m, \ell} n$. Choose codebook ℓ such that the error exponent is positive for all rates below capacity, *i.e.*, if $\xi_\ell R_\ell / \beta_{m, \ell} < C_\ell$, then

$$\lim_{n \rightarrow \infty} -\frac{1}{\beta_{m, \ell} n} \log \mathbb{P}_\ell(W_m(k) \neq \hat{W}_m(k)) > 0. \quad (14)$$

The actual form of the codebook does not matter; for example, it could be a linear code, or a convolutional code, or a LDPC code, or a polar code, or a posterior matching code that uses feedback.

E. Choice of Control Sequences

As described earlier, the transmitter and the receiver agree upon two control sequences, $\sigma_{A,\ell}$ and $\sigma_{R,\ell}$ of length $\beta_{c,\ell}$, for signaling ACCEPT (when $W_m = \hat{W}_m$) and REJECT (when $W_m \neq \hat{W}_m$). Choose these sequences as repetitions of $x_{A,\ell}$ and $x_{R,\ell}$, the maximally separated input symbols for Q_ℓ , i.e., the $\arg \max$ in (5) for B_{Q_ℓ} .

The transmitter and the receiver also agree upon a hypothesis testing rule $\hat{\theta}_{H,\ell}$ for disambiguating $\sigma_{A,\ell}$ and $\sigma_{R,\ell}$. Let $H_{A,\ell}$ and $H_{R,\ell}$ denote the error exponents of this rule, i.e.,

$$H_{A,\ell} = \lim_{n \rightarrow \infty} -\frac{1}{\beta_{c,\ell} n} \log \mathbb{P}_\ell(\hat{W}_c(k) \neq W_c(k) \mid W_c(k) = \sigma_{A,\ell}) \quad (15)$$

and

$$H_{R,\ell} = \lim_{n \rightarrow \infty} -\frac{1}{\beta_{c,\ell} n} \log \mathbb{P}_\ell(\hat{W}_c(k) \neq W_c(k) \mid W_c(k) = \sigma_{R,\ell}). \quad (16)$$

Choose $\hat{\theta}_{H,\ell}$ such that

$$H_{A,\ell} = 0 \quad \text{and} \quad H_{R,\ell} = B_\ell \quad (17)$$

while

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}_\ell(\hat{W}_c(k) \neq W_c(k) \mid W_c(k) = \sigma_{A,\ell}) \\ = \lim_{n \rightarrow \infty} \mathbb{P}_\ell(\hat{W}_c(k) \neq W_c(k) \mid W_c(k) = \sigma_{R,\ell}) \\ = 0. \end{aligned} \quad (18)$$

Such a choice of $\hat{\theta}_{H,\ell}$ is always possible (see [14]).

F. Choice of Parameters

The first and second phases of the proposed scheme correspond to the *message mode* of the Yamamoto–Itoh [12] scheme, while the third and fourth phases correspond to the *control mode*. In the Yamamoto–Itoh scheme, the ratio of the lengths of the message and control modes is $\gamma/(1-\gamma)$ where $\gamma = R/C$. We choose the parameters such that a similar relation holds for the proposed scheme. In particular, let $\gamma_\ell = R_\ell/C_\ell$; then, we want

$$\lim_{n \rightarrow \infty} \frac{\alpha_m + \beta_{m,\ell}}{\alpha_c + \beta_{c,\ell}} = \frac{\gamma_\ell}{1 - \gamma_\ell}.$$

The parameter ξ_ℓ is the proportionality constant, i.e.,

$$\lim_{n \rightarrow \infty} \alpha_m + \beta_{m,\ell} = \xi_\ell \gamma_\ell \quad \text{and} \quad \lim_{n \rightarrow \infty} \alpha_c + \beta_{c,\ell} = \xi_\ell (1 - \gamma_\ell).$$

We let one of these proportionality constants to be one and call that channel the *reference channel* Q_* .

In the Burnashev exponent, the slope (i.e., the B_Q term in (4)) is determined by the “signaling exponent” in the control mode. As will become apparent in the proof of Proposition 6, to maximize the slope of our exponent, we need to choose the parameters such that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\ell(\hat{W}_c(1) \neq W_c(1) \mid W_c(1) = \sigma_{R,\ell}, \hat{L}_c(1) = \ell) \\ = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_\ell(\hat{L}_c(1) \neq \ell). \end{aligned}$$

We choose the parameters that satisfy the above properties as follows. For $\ell = 1, \dots, L$, define constants

$$\kappa_\ell = \frac{T_{c,\ell}}{B_\ell}, \quad \gamma_\ell = \frac{R_\ell}{C_\ell}, \quad \zeta_\ell = \frac{(1 - \gamma_\ell)}{(1 + \kappa_\ell)}. \quad (19)$$

Let κ_* , γ_* , and ζ_* be the κ , γ , and ζ parameters corresponding to the reference channel Q_* . Then, choose the parameters of the coding scheme as follows.

- 1) Choose $\xi_\ell = \zeta_*/\zeta_\ell$.
- 2) Choose $\alpha_m^{(n)} > 0$ such that $\alpha_m^{(n)} n$ is an integer, $\lim_{n \rightarrow \infty} \alpha_m^{(n)} = 0$ while $\lim_{n \rightarrow \infty} \alpha_m^{(n)} n = \infty$. An example for such a choice is $\alpha_m^{(n)} = \lfloor n/\log n \rfloor/n$.
- 3) Choose $\beta_{m,\ell}^{(n)} > \xi_\ell \gamma_\ell$ such that $\beta_{m,\ell}^{(n)} n$ is an integer, $\lim_{n \rightarrow \infty} \beta_{m,\ell}^{(n)} = \xi_\ell \gamma_\ell$.
- 4) Choose $\alpha_c^{(n)} > 0$ such that $\alpha_c^{(n)} n$ is an integer, $\lim_{n \rightarrow \infty} \alpha_c^{(n)} = \zeta_*$.
- 5) Choose $\beta_{c,\ell}^{(n)} > 0$ such that $\beta_{c,\ell}^{(n)}$ is an integer, $\lim_{n \rightarrow \infty} \beta_{c,\ell}^{(n)} = \kappa_\ell \zeta_*$.

G. Consequences of the Choice of Parameters

The choice of the parameters α_m , α_c , $\beta_{m,\ell}$, $\beta_{c,\ell}$, and ξ_ℓ , $\ell = 1, \dots, L$ implies the following:

Lemma 1 (Length of Message and Control Phases): For every $\ell = 1, \dots, L$, we have that

- 1) $\lim_{n \rightarrow \infty} (\alpha_m + \beta_{m,\ell}) = \xi_\ell \gamma_\ell$;
- 2) $\lim_{n \rightarrow \infty} (\alpha_c + \beta_{c,\ell}) = \xi_\ell (1 - \gamma_\ell)$.

□

The choice of the estimation rules $\hat{\theta}_m$, $\hat{\theta}_c$, the codebooks, and the hypothesis testing rules $\hat{\theta}_{H,\ell}$, $\ell = 1, \dots, L$, implies the following properties:

Lemma 2: For every $k \in \mathbb{N}$ and $\ell = 1, \dots, L$, we have that

- 1) $\lim_{n \rightarrow \infty} -\log \mathbb{P}_\ell(\hat{L}_m(k) \neq \ell \mid K \geq k)/\alpha_m n = T_{m,\ell}$;
- 2) $\lim_{n \rightarrow \infty} -\log \mathbb{P}_\ell(\hat{L}_c(k) \neq \ell \mid K \geq k)/\alpha_c n = T_{c,\ell}$;
- 3) $\lim_{n \rightarrow \infty} -\log \mathbb{P}_\ell(\hat{W}_m(k) \neq W_m(k) \mid \hat{L}_m(k) = \ell, K \geq k)/\beta_{m,\ell} n > 0$;
- 4) $\lim_{n \rightarrow \infty} -\log \mathbb{P}_\ell(\hat{W}_c(k) \neq W_c(k) \mid \hat{L}_c(k) = \ell, W_c(k) = \sigma_{A,\ell}, K \geq k)/\beta_{c,\ell} n = H_{A,\ell} = 0$;
- 5) $\lim_{n \rightarrow \infty} -\log \mathbb{P}_\ell(\hat{W}_c(k) \neq W_c(k) \mid \hat{L}_c(k) = \ell, W_c(k) = \sigma_{R,\ell}, K \geq k)/\beta_{c,\ell} n = H_{R,\ell} = B_\ell$.

□

An immediate consequence of the above is that each of the error probabilities approach zero as $n \rightarrow \infty$. Specifically,

Lemma 3: For every $k \in \mathbb{N}$ and $\ell = 1, \dots, L$, we have that

- 1) $\lim_{n \rightarrow \infty} \mathbb{P}_\ell(\hat{L}_m(k) \neq \ell \mid K \geq k) = 0$;
- 2) $\lim_{n \rightarrow \infty} \mathbb{P}_\ell(\hat{L}_c(k) \neq \ell \mid K \geq k) = 0$;
- 3) $\lim_{n \rightarrow \infty} \mathbb{P}_\ell(\hat{W}_m(k) \neq W_m(k) \mid \hat{L}_m(k) = \ell, K \geq k) = 0$;
- 4) $\lim_{n \rightarrow \infty} \mathbb{P}_\ell(\hat{W}_c(k) \neq W_c(k) \mid \hat{L}_c(k) = \ell, W_c(k) = \sigma_{A,\ell}, K \geq k) = 0$;
- 5) $\lim_{n \rightarrow \infty} \mathbb{P}_\ell(\hat{W}_c(k) \neq W_c(k) \mid \hat{L}_c(k) = \ell, W_c(k) = \sigma_{R,\ell}, K \geq k) = 0$.

□

IV. PERFORMANCE ANALYSIS

A. Some Preliminary Results

Recall that the length of epoch $k \in \mathbb{N}$ is $\Lambda(k)n$. Thus

$$\Lambda(k) = \alpha_m + \beta_{m, \hat{L}_m(k)} + \alpha_c + \beta_{c, \hat{L}_c(k)}.$$

Combining Lemmas 1 and 3, we get the following:

Lemma 4: For every $k \in \mathbb{N}$ and $\ell = 1, \dots, L$, we have that

$$\lim_{n \rightarrow \infty} \mathbb{E}_\ell[\Lambda(k) \mid K = k] = \lim_{n \rightarrow \infty} \mathbb{E}_\ell[\Lambda(k)] = \xi_\ell. \quad (20)$$

Thus, for large n and realized channel Q_ℓ , the expected length of each epoch is $\xi_\ell n$. \square

Proof: The event $\{K = k\}$ depends on the third and the fourth phase of the coding scheme and is, therefore, independent of the first phase; hence

$$\mathbb{P}_\ell(\hat{L}_m(k) \mid K = k) = \mathbb{P}_\ell(\hat{L}_m(k)).$$

Moreover, Lemma 2 also implies that

$$\lim_{n \rightarrow \infty} \mathbb{P}_\ell(\hat{L}_c(k) \neq \ell \mid K = k) = 0.$$

Combining the above two relations with Lemmas 1 and 3 give the result. \blacksquare

Let

$$\Sigma_R = \{\sigma_{R,\ell}, \ell = 1, \dots, L\}$$

denote the set of all REJECT control signals and let $\rho_\ell^{(n)}$ denote the probability that the estimated control sequence in epoch k is in Σ_R , i.e.,

$$\rho_\ell = \mathbb{P}_\ell(\hat{W}_c(k) \in \Sigma_R \mid K \geq k). \quad (21)$$

Due to symmetry across each epoch, ρ_ℓ does not depend on k .

Conditioned on the event that $K \geq k$, communication stops at epoch k if the estimated control sequence $\hat{W}_c(k)$ is REJECT. Hence

$$\mathbb{P}_\ell(K = k \mid K \geq k) = \rho_\ell.$$

Consequently, we have the following.

Proposition 2: For any $n \in \mathbb{N}$ and $\ell = 1, \dots, L$, the number of retransmissions has a geometric distribution; in particular

$$\mathbb{P}_\ell(K = k) = \rho_\ell(1 - \rho_\ell)^{k-1}, \quad k \in \mathbb{N}. \quad (22)$$

Furthermore, Lemma 3 implies that

$$\lim_{n \rightarrow \infty} \rho_\ell^{(n)} = 1. \quad (23)$$

Hence

$$\lim_{n \rightarrow \infty} \mathbb{P}_\ell(K^{(n)} = 1) = 1. \quad (24)$$

Thus, for large n and irrespective of the realized channel, the expected number of transmission epochs is one. \square

B. Expected Length of Communication

Proposition 3: For every $\ell = 1, \dots, L$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_\ell[\tau^{(n)}] = \xi_\ell. \quad (25)$$

Proof: Since $\tau = \sum_{k=1}^K \Lambda(k)n$, we get

$$\begin{aligned} \frac{1}{n} \mathbb{E}_\ell[\tau] &= \mathbb{E}_\ell \left[\sum_{k=1}^K \Lambda(k) \right] = \mathbb{P}_\ell(K = 1) \mathbb{E}_\ell[\Lambda(1) \mid K = 1] \\ &\quad + \mathbb{P}_\ell(K > 1) \mathbb{E}_\ell \left[\sum_{k=1}^K \Lambda(k) \mid K > 1 \right]. \end{aligned}$$

Proposition 2 implies that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_\ell[\tau] = \lim_{n \rightarrow \infty} \mathbb{E}_\ell[\Lambda(1)] = \xi_\ell$$

where the last equality follows from Lemma 4. \blacksquare

C. Probability of Error

Proposition 4: For any $n \in \mathbb{N}$ and $\ell = 1, \dots, L$, the probability of error is given by

$$\begin{aligned} P_\ell^{(n)} &= \frac{1}{\rho_\ell^{(n)}} \mathbb{P}_\ell(\hat{W}_m(1) \neq W_m(1)) \\ &\quad \times \mathbb{P}_\ell(\hat{W}_c(1) \neq W_c(1) \mid W_c(1) \in \Sigma_R). \quad (26) \end{aligned}$$

Proof: The error event is $\{\hat{W}_m(K) \neq W_{\hat{L}_m(K)}\}$. For each $k \in \mathbb{N}$, $W_m(k) = W_{\hat{L}_m(k)}$. Using this to simplify the probability of error, we get that

$$\begin{aligned} P_\ell^{(n)} &= \mathbb{P}_\ell(\hat{W}_m(K) \neq W_m(K)) \\ &= \sum_{k=1}^{\infty} \mathbb{P}_\ell(\hat{W}_m(K) \neq W_m(K), K = k) \\ &= \sum_{k=1}^{\infty} \mathbb{P}_\ell(K = k) \mathbb{P}_\ell(\hat{W}_m(k) \neq W_m(k) \mid K = k) \\ &\stackrel{(a)}{=} \mathbb{P}_\ell(\hat{W}_m(1) \neq W_m(1) \mid K = 1) \sum_{k=1}^{\infty} \mathbb{P}_\ell(K = k) \\ &\stackrel{(b)}{=} \frac{1}{\mathbb{P}_\ell(K = 1)} \mathbb{P}_\ell(\hat{W}_m(1) \neq W_m(1)) \\ &\quad \times \mathbb{P}_\ell(K = 1 \mid \hat{W}_m(1) \neq W_m(1)) \\ &= \frac{1}{\mathbb{P}_\ell(K = 1)} \mathbb{P}_\ell(\hat{W}_m(1) \neq W_m(1)) \\ &\quad \times \mathbb{P}_\ell(\hat{W}_c(1) \neq W_c(1) \mid W_c(1) \in \Sigma_R) \end{aligned}$$

where (a) follows from the symmetry across epochs and (b) follows from the Bayes rule. \blacksquare

D. Opportunistically Achievable Rate

Proposition 5: The coding scheme of Section III opportunistically achieves the rate vector (R_1, \dots, R_L) . \square

Proof: To prove the result, we need to show the proposed scheme satisfies the properties described in Definition 4. Specifically

$$\lim_{n \rightarrow \infty} \mathbb{E}_\ell[\tau] = \infty \quad (27)$$

along with

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}_\ell[\log M_{\hat{L}_m(K)}]}{\mathbb{E}_\ell[\tau]} = R_\ell \quad (28)$$

and

$$\lim_{n \rightarrow \infty} P_\ell^{(n)} = 0. \quad (29)$$

We prove these separately.

(a) Property (27) follows from Proposition 3.

(b) Recall that $M_\ell = \exp_2(n\xi_\ell R_\ell)$. Hence

$$\frac{\mathbb{E}_\ell[\log M_{\hat{L}_m(K)}]}{\mathbb{E}_\ell[\tau]} = \mathbb{E}_\ell[\xi_{\hat{L}_m(K)} R_{\hat{L}_m(K)}] \frac{n}{\mathbb{E}_\ell[\tau]}.$$

Proposition 3 implies that

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}_\ell[\log M_{\hat{L}_m(K)}]}{\mathbb{E}_\ell[\tau]} = \frac{1}{\xi_\ell} \lim_{n \rightarrow \infty} \mathbb{E}_\ell[\xi_{\hat{L}_m(K)} R_{\hat{L}_m(K)}]. \quad (30)$$

Now

$$\begin{aligned} \mathbb{E}_\ell[\xi_{\hat{L}_m(K)} R_{\hat{L}_m(K)}] &= \mathbb{P}_\ell(K = 1) \mathbb{E}_\ell[\xi_{\hat{L}_m(1)} R_{\hat{L}_m(1)}] \\ &\quad + \mathbb{P}_\ell(K > 1) \mathbb{E}_\ell[\xi_{\hat{L}_m(K)} R_{\hat{L}_m(K)} \mid K > 1] \end{aligned}$$

where we do not condition on $K = 1$ in the first term of the RHS because the event $K = 1$ depends only on the events of the third and the fourth phases, which are independent of the first and the second phases, and hence independent of $\hat{L}_m(1)$. Using Proposition 2, we get that

$$\lim_{n \rightarrow \infty} \mathbb{E}_\ell[\xi_{\hat{L}_m(K)} R_{\hat{L}_m(K)}] = \lim_{n \rightarrow \infty} \mathbb{E}_\ell[\xi_{\hat{L}_m(1)} R_{\hat{L}_m(1)}]. \quad (31)$$

Now

$$\begin{aligned} \mathbb{E}_\ell[\xi_{\hat{L}_m(1)} R_{\hat{L}_m(1)}] &= \mathbb{P}_\ell(\hat{L}_m(1) = \ell) \mathbb{E}_\ell[\xi_{\hat{L}_m(1)} R_{\hat{L}_m(1)} \mid \hat{L}_m(1) = \ell] \\ &\quad + \mathbb{P}_\ell(\hat{L}_m(1) \neq \ell) \mathbb{E}_\ell[\xi_{\hat{L}_m(1)} R_{\hat{L}_m(1)} \mid \hat{L}_m(1) \neq \ell]. \end{aligned}$$

Using Lemma 3, we get that

$$\lim_{n \rightarrow \infty} \mathbb{E}_\ell[\xi_{\hat{L}_m(1)} R_{\hat{L}_m(1)}] = \xi_\ell R_\ell. \quad (32)$$

Substituting (31) and (32) into (30) gives (28).

(c) Property (29) follows substituting the results of Proposition 2 and Lemma 3 in Proposition 4. \blacksquare

E. EER

Proposition 6: For a particular choice of estimation rule $\hat{\theta}_c$, the ℓ -component of the error exponent (E_1, \dots, E_L) of the coding scheme of Section III is bounded by

$$E_\ell \geq \frac{\kappa_\ell}{1 + \kappa_\ell} B_\ell (1 - \gamma_\ell) = \frac{T_{c,\ell}}{T_{c,\ell} + B_\ell} B_\ell \left(1 - \frac{R_\ell}{C_\ell}\right). \quad (33)$$

By varying the choice of $\hat{\theta}_c$, we get

$$\begin{aligned} \mathcal{E}(R_1, \dots, R_L) &\supseteq \\ &\bigcup_{(T_{c,1}, \dots, T_{c,L}) \in \mathcal{F}} \left(\frac{T_{c,1}}{T_{c,1} + B_1} B_1 \left(1 - \frac{R_1}{C_1}\right), \dots, \right. \\ &\quad \left. \frac{T_{c,L}}{T_{c,L} + B_L} B_L \left(1 - \frac{R_L}{C_L}\right) \right). \quad (34) \end{aligned}$$

Proof: Consider the expression for P_ℓ in Proposition 4. Taking logs, we get

$$\begin{aligned} -\frac{1}{n} \log P_\ell &= -\frac{1}{n} \log \mathbb{P}_\ell(\hat{W}_m(1) \neq W_m(1)) \\ &\quad - \frac{1}{n} \log \mathbb{P}_\ell(\hat{W}_c(1) \neq W_c(1) \mid W_c(1) \in \Sigma_R) \\ &\quad + \frac{\log \rho_\ell}{n}. \quad (35) \end{aligned}$$

Consider the three summands in the RHS of (35). First consider the first term of the RHS of (35). From Lemma 3, we have that

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}_\ell(\hat{W}_m(1) \neq W_m(1)) > 0. \quad (36)$$

Next consider the second term of the RHS of (35).

$$\begin{aligned} \mathbb{P}_\ell(\hat{W}_c(1) \neq W_c(1) \mid W_c(1) \in \Sigma_R) &< \mathbb{P}_\ell(\hat{W}_c(1) \neq W_c(1) \mid W_c(1) = \sigma_{R,\ell}, \hat{L}_c(1) = \ell) \\ &\quad + \mathbb{P}_\ell(\hat{L}_c(1) \neq \ell). \quad (37) \end{aligned}$$

From Lemma 3, we have that

$$\begin{aligned} \lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}_\ell(\hat{W}_c(1) \neq W_c(1) \mid W_c(1) = \sigma_{R,\ell}, \hat{L}_c(1) = \ell) &= \kappa_\ell \zeta_* B_\ell \quad (38) \end{aligned}$$

and

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}_\ell(\hat{L}_c(1) \neq \ell) = \zeta_* T_{c,\ell} = \zeta_* \kappa_\ell B_\ell. \quad (39)$$

where the last equality follows because $\kappa_\ell = T_{c,\ell}/B_\ell$. Substituting (38) and (39) into (37), and taking logarithms and limits, we get

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}_\ell(\hat{W}_c(1) \neq W_c(1) \mid W_c(1) \in \Sigma_R) > \zeta_* \kappa_\ell B_\ell. \quad (40)$$

Next consider the third term of the RHS of (35). From Proposition 2, it follows that

$$\lim_{n \rightarrow \infty} \frac{\log \rho_\ell}{n} = 0. \quad (41)$$

Substituting the result of (36), (40), and (41) into (35), we get

$$-\frac{1}{n} \log P_\ell > \zeta_* \kappa_\ell B_\ell. \quad (42)$$

Combining this with Proposition 3, we get

$$E_\ell = -\frac{1}{\mathbb{E}_\ell[\tau]} \log P_\ell > \zeta_* \kappa_\ell B_\ell / \xi_\ell. \quad (43)$$

The result follows by observing that $\xi_\ell = \zeta_* / \zeta_\ell$, and substituting the value of κ_ℓ and ζ_ℓ in (43). ■

The choice of operating point on the EER boundary depends on the objective. For given positive constants w_1, \dots, w_L , two possible objectives are to minimize the weighted probability of error

$$\bar{P} := (w_1 P_1 + \dots + w_L P_L) / (w_1 + \dots + w_L)$$

or maximize the weighted error exponent

$$\bar{E} := (w_1 E_1 + \dots + w_L E_L) / (w_1 + \dots + w_L).$$

As $n \rightarrow \infty$, each of P_1, \dots, P_L decay to zero exponentially. Thus, minimizing \bar{P} is equivalent to maximizing $\min\{E_1, \dots, E_L\}$. The choice of the operating point (E_1, \dots, E_L) , and hence the choice of $\hat{\theta}_c$, depends on the objective.

F. Capacity

Proposition 6 implies that for any rate vector (R_1, \dots, R_L) such that $R_\ell < C_\ell, \ell = 1, \dots, L$, each component of the probability of error (P_1, \dots, P_L) goes to zero as $n \rightarrow \infty$. Thus

$$\mathcal{E}_{OF}(\mathcal{C}) \supseteq \{(R_1, \dots, R_L) : 0 \leq R_\ell < C_\ell, \ell = 1, \dots, L\}.$$

Furthermore, if a coding scheme (opportunistically) achieves rate R_ℓ when the realized channel $Q_o = Q_\ell$, then the same scheme will also achieve rate R_ℓ when used over DMC Q_ℓ . Thus

$$\mathcal{E}_{OF}(\mathcal{C}) \subseteq \{(R_1, \dots, R_L) : 0 \leq R_\ell < C_\ell, \ell = 1, \dots, L\}.$$

Combining these two bounds, we get

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

$$\mathcal{E}_{OF}(\mathcal{C}) = \{(R_1, \dots, R_L) : 0 \leq R_\ell < C_\ell, \ell = 1, \dots, L\}.$$

We call $\mathcal{E}_{\mathcal{C}} := (C_1, \dots, C_L)$ as called the *capacity vector* of the compound channel \mathcal{C} . □

V. EXAMPLE

Consider a compound channel consisting of two BSCs with complementary crossover probabilities, p and $(1-p)$, where

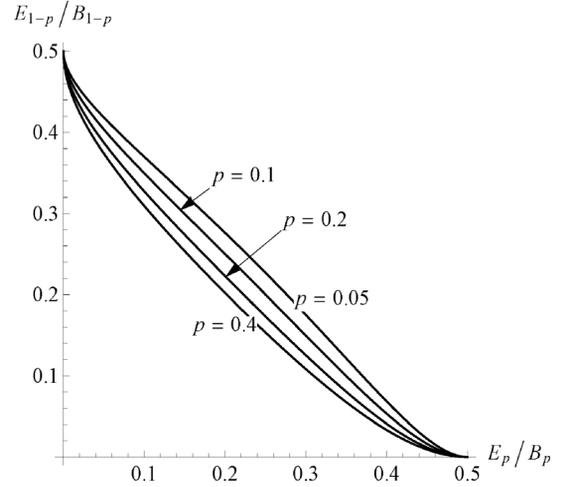


Fig. 1. Scaled EER region $\Phi(p) = \{(E_p/B_p, E_{1-p}/B_{1-p}) : (E_p, E_{1-p}) \in \mathcal{E}(R_p, R_{1-p})\}$ for different instances of the compound channel $\mathcal{C}_p := \{\text{BSC}_p, \text{BSC}_{1-p}\}$.

$0 < p < 1/2$ and p is known to the transmitter and the receiver. Denote this compound channel by

$$\mathcal{C}_p := \{\text{BSC}_p, \text{BSC}_{1-p}\}$$

where BSC_p denotes a binary symmetric channel with crossover probability p . For convenience, we index all variables by p and $(1-p)$ rather than by 1 and 2. For binary symmetric channel, the capacity and the zero-rate 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. Assume that the desired communication rate is (R_p, R_{1-p}) , where $R_p < C_p$ and $R_{1-p} < C_{1-p}$.

Choose the training sequences σ_m and σ_c as all zero sequences of length $\alpha_m n$ and $\alpha_c n$. Choose the channel estimation rules $\hat{\theta}_m$ and $\hat{\theta}_c$ as the threshold tests: if the empirical frequency of ones in the output is less than q , $p < q < 1-p$, estimate the channel as BSC_p ; otherwise, estimate the channel as BSC_{1-p} . The thresholds for $\hat{\theta}_m$ and $\hat{\theta}_c$ are q_m and q_c , respectively. For such a threshold test, the probability of estimation error is bound by the tail probability of a sum of independent random variables. From Hoeffding's inequality [15, Th. 1], the exponents of the estimation errors are

$$\begin{aligned} T_{m,p} &= D(q_m||p), & T_{m,1-p} &= D(q_m||1-p), \\ T_{c,p} &= D(q_c||p), & T_{c,1-p} &= D(q_c||1-p). \end{aligned}$$

Choose the two codebooks as any codebooks for BSC_p and BSC_{1-p} that have positive error exponents.

Choose the control sequences $\sigma_{A,p}$ and $\sigma_{R,p}$ as $\beta_{c,p} n$ repetitions of zeros and ones, respectively. Similarly, choose the

control sequences $\sigma_{A,1-p}$ and $\sigma_{R,1-p}$ as $\beta_{c,1-p}n$ repetitions of ones and zeros, respectively. The hypothesis testing rules $\hat{\theta}_{H,p}$ and $\hat{\theta}_{H,1-p}$ are chosen as described in Section III-E.

Proposition 6 implies that for any rate vector (R_p, R_{1-p}) and a particular choice of the estimation threshold q_c ,⁵ the above scheme achieves an error exponent (E_p, E_{1-p}) such that

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

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

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

There are no known upper bounds on the EER. Hence, we compare with the trivial upper bound of the Burnashev exponent of BSC_p and BSC_{1-p}.

$$\frac{E_p}{B_p} \geq \frac{D(q_c||p)}{D(q_c||p) + D(p||1-p)}$$

$$\frac{E_{1-p}}{B_{1-p}} \geq \frac{D(q_c||1-p)}{D(q_c||1-p) + D(p||1-p)}.$$

Let

$$\varphi(p, q_c) := \left(\frac{E_p}{B_p}, \frac{E_{1-p}}{B_{1-p}} \right)$$

$$\geq \left(\frac{D(q_c||p)}{D(q_c||p) + D(p||1-p)}, \frac{D(q_c||1-p)}{D(q_c||1-p) + D(p||1-p)} \right)$$

and

$$\Phi(p) := \left\{ \left(\frac{E_p}{B_p}, \frac{E_{1-p}}{B_{1-p}} \right) : (E_p, E_{1-p}) \in \mathcal{A}(R_p, R_{1-p}) \right\}$$

$$= \{ \varphi(p, q_c) : p < q_c < 1-p \}.$$

For the scheme proposed in Section III, $\Phi(p)$ does not depend on the transmission rate (R_p, R_{1-p}) . We plot $\Phi(p)$ for different values of p in Fig. 1.

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: any rate less than the capacity of the realized channel is opportunistically achievable, even though we do not know the realized 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.

Our proposed training-based scheme is conceptually similar to Yamamoto-Itoh's scheme. It operates in multiple epochs; each epoch is divided into a message mode and a control mode. A training sequence is transmitted at the beginning of each

⁵The choice of q_m does not affect the values of E_p and E_{1-p} as long as $\mathbb{P}_\ell(\hat{L}_m \neq \ell) \rightarrow 0$. For that, we require only that $p < q_m < 1-p$. Choosing $q_m = 0.5$ ensures that.

mode, and the corresponding channel estimate determines the operation during the remainder of the mode.

It may appear that the proposed scheme can be simplified by combining the training phases in each epoch, i.e., have a training phase followed by message and control modes. However, as argued by Tchamkerten and Telatar [16], such a simplification will lead to error exponents that have zero-slope near capacity. Our results do not contradict the results of [16] because we allow for more sophisticated training. Re-training in the control mode ensures that the error events $\{\hat{W}_m(k) \neq W_m(k)\}$ and $\{\hat{W}_c(k) \neq W_c(k)\}$ are independent, which, in turn, is essential to obtain an error exponent of the form $B_\ell(1 - \gamma_\ell)$.

One possible way to make the scheme more efficient is to accumulate the training sequences for each phase, i.e., the channel estimation for the message mode and the control mode is based on all past training sequences for that mode. Such an accumulation will improve the finite length performance of the scheme, but does not affect the asymptotic performance because, in the limit, the communication lasts for only one epoch with high probability.

Another possibility to improve the performance of the coding scheme is to use a universal coding scheme for the control mode rather than a training-based scheme. This motivates the study of the following communication problem.

Open Problem: Consider the communication of a binary valued message over a compound channel with feedback. Let $\mathcal{C} = \{Q_1, \dots, Q_L\}$ denote the compound channel, $W \in \{\theta_0, \theta_1\}$ denote the message, X_t and Y_t denote the channel inputs and output at time t , and \hat{W} denotes the decoded message. Consider a variable length coding scheme $(\mathbf{c}, \mathbf{g}, \tau)$, where c_t is the encoding function at time t , g_t is the decoding function at time t , and τ is a Y^t -measurable stopping time. The decoded message is

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

Let $a_{\ell n}$ and $b_{\ell n}$ denote the exponent of the two types of errors, i.e.,

$$a_{\ell n}(\mathbf{c}, \mathbf{g}, \tau) = \frac{-\log \mathbb{P}_\ell(\hat{W} = \theta_1 | W = \theta_0)}{\mathbb{E}_\ell[\tau | W = \theta_0]}, \quad \ell = 1, \dots, L$$

$$b_{\ell n}(\mathbf{c}, \mathbf{g}, \tau) = \frac{-\log \mathbb{P}_\ell(\hat{W} = \theta_0 | W = \theta_1)}{\mathbb{E}_\ell[\tau | W = \theta_1]}, \quad \ell = 1, \dots, L.$$

where \mathbb{P}_ℓ is the induced probability measure when the true channel equals Q_ℓ .

For a sequence $S = \{\mathbf{c}^{(n)}, \mathbf{g}^{(n)}, \tau^{(n)}\}_{n=1}^\infty$ of coding schemes such that

$$\lim_{n \rightarrow \infty} \mathbb{E}_\ell[\tau^{(n)} | W = \theta_i] = \infty, \quad i = 0, 1, \quad \ell = 1, \dots, L.$$

define the type-I and type-II error exponents of S as

$$a_{\ell*} = \lim_{n \rightarrow \infty} a_{\ell n}(\mathbf{c}^{(n)}, \mathbf{g}^{(n)}, \tau^{(n)})$$

$$b_{\ell*} = \lim_{n \rightarrow \infty} b_{\ell n}(\mathbf{c}^{(n)}, \mathbf{g}^{(n)}, \tau^{(n)}).$$

Furthermore, define

$$b_{\ell}^* = \max_{S: a_{\ell^*}(S)=0} b_{\ell^*}(S).$$

What is the best type-II exponent (b_1^*, \dots, b_L^*) ?

Tchamkerten and Telatar [17] studied a similar problem and identified necessary and sufficient conditions under which

$$b_{\ell}^* = B_{Q_{\ell}}, \quad \ell = 1, \dots, L.$$

We are not aware of the solution to the above problem when the conditions in [17] are not satisfied.

Given any sequence S of coding schemes for Problem 1, we can replace the control mode (phases three and four) of the proposed coding scheme by S and achieve an error exponent of

$$(b_{1^*}(S)(1 - \gamma_1), \dots, b_{L^*}(S)(1 - \gamma_L)).$$

If S is optimal, the error exponent is

$$(b_1^*(S)(1 - \gamma_1), \dots, b_L^*(S)(1 - \gamma_L)). \quad (46)$$

We conjecture that no coding scheme can achieve a better error exponent, i.e., (46) is the Pareto frontier of the EER.

When the conditions in [17] are satisfied, we can replace the control mode by the variable length coding scheme proposed in [17], and thereby recover the result of [7]. In fact, in that case, our modified scheme is exactly the same as the variation proposed in [7, Sec. IV-B]. When the conditions in [17] are not satisfied, the scheme proposed in this paper provide an inner bound on the EER. To find the best error exponents, we need to solve Problem 1.

In this paper, we presented an inner bound on the EER when the compound channel is defined over a finite family. Generalization of the coding scheme to compound channels defined over continuous families is an important and interesting future direction. We believe that solving Problem 1 is a critical step in that direction.

ACKNOWLEDGMENT

The authors are grateful to A. Tchamkerten for helpful feedback and to the anonymous reviewers whose suggestions helped to improve the presentation of the paper.

REFERENCES

- [1] J. Wolfowitz, "Simultaneous channels," *Archive for Rational Mech. Anal.*, vol. 4, pp. 371–386, Nov. 1959.
- [2] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *The Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.

- [3] J. Wolfowitz, *Coding Theorems of Information Theory*. New York: Springer-Verlag, 1964.
- [4] B. Shrader and H. Permuter, "Feedback capacity of the compound channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3629–3644, Aug. 2009.
- [5] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.
- [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, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2126–2145, May 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] 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 *Proc. IEEE Symp. Found. Comput. Sci.*, 2002, pp. 271–280.
- [11] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 6, no. 52, pp. 2551–2567, Jun. 2006.
- [12] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 6, pp. 729–733, Nov. 1979.
- [13] E. Tuncel, "On error exponents in hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2945–2950, Aug. 2005.
- [14] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, Jul. 1974.
- [15] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Assoc.*, vol. 58, no. 301, pp. 13–30, Mar. 1963.
- [16] 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.
- [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.

Aditya Mahajan (S'09–M'09) received the B.Tech degree in Electrical Engineering from the Indian Institute of Technology, Kanpur, India in 2003 and the MS and PhD degrees in Electrical Engineering and Computer Science from the University of Michigan, Ann Arbor, MI, USA in 2006 and 2008. He is currently an Assistant Professor of Electrical and Computer Engineering, McGill University, QC, Canada. From 2008 to 2010, he was a Postdoctoral Researcher in the department of Electrical Engineering at Yale University, New Haven, CT. His research interests include decentralized stochastic control, team theory, real-time communication, information theory, and discrete event systems.

Sekhar Tatikonda (S'92–M'00) received the PhD degree in Electrical Engineering and Computer Science from Massachusetts Institute of Technology, Cambridge, in 2000. He is currently an Associate Professor of Electrical Engineering in the Department of Electrical Engineering, Yale University, New Haven, CT. From 2000 to 2002, he was a Post-Doctoral Fellow in the Department of Computer Science, University of California, Berkeley. His current research interests include communication theory, information theory, stochastic control, distributed estimation and control, statistical machine learning, and inference.