# Achievable Error Exponent of Channel Coding in Random Access Communication


Zheng Wang and Jie Luo
Electrical & Computer Engineering Department
Colorado State University, Fort Collins, CO 80523
Email: {zhwang, rockey}@engr.colostate.edu



***Abstract*— In [1], a new channel coding approach was proposed for random multiple access communication over the discrete-time memoryless channel. The coding approach allows users to choose their communication rates distributively without sharing the rate information with each other and the receiver. The receiver makes decoding and collision report decisions depending on whether reliable message recovery is possible. It was shown that, asymptotically as codeword length goes to infinity, the set of communication rates supporting reliable message recovery can be characterized by an achievable region which equals Shannon's information rate region without a convex hull operation. In this paper, we derive stronger versions of the coding theorems in [1] by characterizing how system error probability, defined as the maximum of decoding error probability and collision miss detection probability, approaches zero with respect to the codeword length.**


## I. Introduction

Classical information theoretic multiple access communication assumes each user (transmitter) is backlogged with an infinite reservoir of traffic. Users should first jointly determine their codebooks and information rates, and then send the encoded messages to the receiver continuously over a long communication duration. The only responsibility of the receiver is to decode the messages with its best effort.

In time-slotted random multiple access communication, however, data bits arrive stochastically at the users. According to data availability and the MAC layer protocol, in each time slot, each user encodes certain number of data bits into a packet (codeword) of certain slot length and transmits the packet to the receiver. The number of encoded data units per symbol is defined as the communication rate of a user in a time slot, and can vary between time slots. Due to bursty traffic arrival and the lack of instant information exchange between users, communication rates of the users are determined individually by each user, and are unknown to the receiver. Upon receiving the channel output symbols in a time slot, the receiver decodes the data if a predetermined decoding error probability requirement can be satisfied. Otherwise, the receiver outputs a "collision".

The issue of channel coding in random multiple access communication was investigated in [1], where a new coding approach was proposed for communication over a discrete time memoryless channel. Assume random channel coding (defined in [1]) is applied within each time slot. Asymptotically as the codeword length (slot length) is taken to infinity and the required decoding error probability is taken to zero, the set of communication rates supporting message decoding can be characterized by an achievable rate region, which equals the Shannon information rate region without a convex hull operation [1].

In this paper, we derive stronger versions of the coding theorems given in [1]. We consider random multiple access communication over a discrete-time memoryless channel and focus on coding *within one time slot*. Define the set of communication rate vectors under which the receiver should decode the messages as the "operation region" of the system. We show that if the operation region is strictly contained in the maximum achievable rate region given in [1], then the system error probability, defined as the maximum of the decoding error probability and the collision miss detection probability, decays exponentially in the slot length (codeword length). Close form expression of the corresponding exponent, defined as the system error exponent, is obtained.

## II. Packet Coding and Error Performance in Single-user Random Access Communication

For easy understanding, we will first consider single-user random access communication over a discrete-time memoryless channel. The channel is modeled by a conditional distribution function $P_{Y|X}$, where $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ are the channel input and output symbols, $\mathcal{X}$, $\mathcal{Y}$ are the finite input and output alphabets. Assume time is partitioned into slots each equaling $N$ symbol durations, which is also the length of a packet. We focus on coding within a time slot or a packet.

Suppose the channel is unknown at the transmitter. The significance of this assumption will become clear when we investigate multiuser systems. At the beginning of a time slot, the transmitter randomly generates a "random access codebook" [1], specified as follows. The codebook contains $M$ classes of codewords, where $M$ is a predetermined integer. The $i$th ($i \in \{1, \dots, M\}$) codeword class contains $\lfloor e^{Nr_i} \rfloor$ codewords, each of $N$ symbol length. where $r_i$ is a predetermined rate parameter, in nats per symbol. Codeword symbols are independently distributed with symbols of the $i$th codeword class being i.i.d. according to a conditional input distribution $P_{X|r_i}$. We assume $r_1 < r_2 < \dots < r_M$. According to message availability and the MAC layer protocol, the transmitter chooses a communication rate $r \in \{r_1, \dots, r_M\}$ without sharing this rate information with the receiver. Suppose $r = r_i$.

The transmitter then encodes $\lfloor Nr_i/\ln 2\rfloor$ data bits, denoted by a message $w$, into the $w$th codeword in the $i$th codeword class, denoted by $\boldsymbol{x}_{w_i}$, and sends it to the receiver. Note that this is equivalent to mapping the message and communication rate pair $(w, r_i)$ to a codeword $\boldsymbol{x}_{w_i}$ with length $N$.

We assume the receiver knows the channel $P_{Y|X}$ and the randomly generated codebook. Based on these information, the receiver chooses an integer $H \leq M$. According to the channel output symbol vector $\boldsymbol{y}$, the receiver outputs an estimated message and rate pair $(\hat{w}, \hat{r})$ only if $\hat{r} \in \{r_1, \dots, r_H\}$ and a predetermined decoding error probability requirement can be satisfied. Otherwise the receiver outputs "$\hat{r}$ = collision". Note that the term "collision" here is used to maintain consistency with the networking terminology. Throughout the paper, collision means outage, irrespective whether it is caused by interference due to multiuser transmissions.

Since the receiver *intends* to decode all messages with $r \leq r_H$, we say $R = \{r|r \leq r_H\}$ is the "operation region" of the system. For all messages with $r \leq r_H$, we define the decoding error probability as

$$P_e(w, r) = Pr\{(\hat{w}, \hat{r}) \neq (w, r)\}, \qquad \forall (w, r), r \leq r_H. \quad (1)$$

For all messages with $r > r_H$, we define the collision miss detection probability as

$$\bar{P}_c(w, r) = Pr\{\hat{r} \neq \text{collision}\}, \qquad \forall (w, r), r > r_H. \quad (2)$$

The system error probability is defined as the maximum of (1) and (2).

Assume $r_i < I_{P_{X|r_i}}(X;Y)$ for all $i \leq H$, where $I_{P_{X|r_i}}(X;Y)$ is the mutual information between $X$ and $Y$ computed based on input distribution $P_{X|r_i}$. According to [1], we have the following asymptotic results.

$$\begin{aligned}
&\lim_{N\to\infty} P_e(w, r) = 0, \quad \forall (w, r), r \in \{r_1, \dots, r_H\}, \\
&\lim_{N\to\infty} \bar{P}_c(w, r) = 0, \quad \forall (w, r), r \notin \{r_1, \dots, r_H\}. \quad (3)
\end{aligned}$$

In other words, asymptotically, the receiver can reliably decode the message if the random communication rate $r$ is inside the operation region; the receiver can reliably report a collision if $r$ is outside the operation region.

In the rest of this section, we will analyze the convergence of the system error probability with respect to $N$. Assume $(w, r)$ is encoded and transmitted. The receiver outputs an estimate $(\hat{w} = m, \hat{r} = r_h)$ if both the following conditions are satisfied,

$$\begin{aligned}
&\text{C1: } -\frac{1}{N}\ln Pr\{\boldsymbol{y}|\boldsymbol{x}_{m_h}\} < -\frac{1}{N}\ln Pr\{\boldsymbol{y}|\boldsymbol{x}_{\tilde{m}_{\tilde{h}}}\}, \\
&\qquad \text{for all } (\tilde{m}, \tilde{h}) \neq (m, h), h, \tilde{h} \leq H; \\
&\text{C2: } -\frac{1}{N}\ln Pr\{\boldsymbol{y}|\boldsymbol{x}_{m_h}\} < \tau_h(\boldsymbol{y}), \quad (4)
\end{aligned}$$

where $\tau_h(\boldsymbol{y})$, as a function of $\boldsymbol{y}$, is a predetermined threshold corresponding to the codewords of the $h$th class. If there is no codeword satisfying (4), the receiver reports $\hat{r}$ = collision. In other words, the receiver decodes only if the negative log-likelihood of the maximum likelihood estimation exceeds certain threshold.

For $r \in \{r_1, \dots, r_H\}$, we define the decoding error exponent $E_d$ as

$$E_d = \min_{(w,r), r\in\{r_1,\dots,r_H\}} \lim_{N\to\infty} -\frac{1}{N}\ln P_e(w, r). \quad (5)$$

For $r \in \{r_{H+1}, \dots, r_M\}$, we define the collision miss detection exponent $E_c$ as

$$E_c = \min_{(w,r), r\in\{r_{H+1},\dots,r_M\}} \lim_{N\to\infty} -\frac{1}{N}\ln \bar{P}_c(w, r). \quad (6)$$

The system error exponent $E_s$ is defined by $E_s = \min\{E_d, E_c\}$.

The following theorem gives a lower bound on the system error exponent.

**Theorem 1:** For single-user random access communication over a discrete-time memoryless channel $P_{Y|X}$. Given the conditional input distributions $P_{X|r_h}$ for all $1 \leq h \leq M$, the system error exponent is bounded by,

$$\begin{aligned}
E_s \geq \min\Bigg\{ &\min_{h,\tilde{h}\leq H} \max_{0\leq\rho\leq 1} \left[-\rho r_{\tilde{h}} + E_0(\rho, P_{X|r_h}, P_{X|r_{\tilde{h}}})\right], \\
&\min_{h>H,\tilde{h}\leq H} \max_{0\leq\rho\leq 1} \left[-\rho r_{\tilde{h}} + E_1(\rho, P_{X|r_h}, P_{X|r_{\tilde{h}}})\right] \Bigg\}, \quad (7)
\end{aligned}$$

where

$$\begin{aligned}
E_0(\rho, P_{X|r_h}, P_{X|r_{\tilde{h}}}) = &\max_{0\leq s\leq 1} -\ln \sum_y \left(\sum_x P_{X|r_h}(x)\right. \\
&\left.\times P_{Y|X}(y|x)^{1-s}\right) \left(\sum_x P_{X|r_{\tilde{h}}}(x) P_{Y|X}(y|x)^{\frac{s}{\rho}}\right)^{\rho}, \\
E_1(\rho, P_{X|r_h}, P_{X|r_{\tilde{h}}}) = &\max_{0\leq s\leq 1-\rho} -\ln \sum_y \left(\sum_x P_{X|r_h}(x)\right. \\
&\left.\times P_{Y|X}(y|x)\right)^{1-s} \left(\sum_x P_{X|r_{\tilde{h}}}(x) P_{Y|X}(y|x)^{\frac{s}{s+\rho}}\right)^{s+\rho}. \quad (8)
\end{aligned}$$

■

The proof of Theorem 1 is given in the Appendix. Note that the first term on the right hand side (RHS) of (7) is due to condition C1 in (4), which corresponds to the maximum likelihood decoding criterion. This term becomes Gallager's random-coding exponent [2] if the system has only one class of codewords with the same input distribution. The second term is due to condition C2, which corresponds to a typical sequence decoding criterion.

## III. Packet Coding and Error Performance in Multi-user Random Access Communication

In this section, we consider a $K$-user time-slotted random access communication system over a discrete-time memoryless channel. The channel is modeled by a conditional distribution $P_{Y|X_1,\dots,X_K}$, where $X_k \in \mathcal{X}_k$, $k \in \{1, \dots, K\}$, is the channel input symbol of user $k$ with $\mathcal{X}_k$ being the the finite input alphabet, $Y \in \mathcal{Y}$ is the channel output symbol with $\mathcal{Y}$ being the finite output alphabet. Assume a slot length equals $N$ symbol durations, which is also the length of a packet. We again focus on coding within a time slot.

Assume at the beginning of a time slot, user $k$, $k \in \{1, \ldots, K\}$, generates a "random access codebook" [1] that has $M$ classes of codewords. The $i$th codeword class contains $\lfloor e^{Nr_{i_k}} \rfloor$ codewords, each of $N$ symbol length, with $r_{i_k}$ being a predetermined rate parameter, in nats per symbol. We assume all codeword symbols are generated independently. Codeword symbols of user $k$ in the $i$th class are i.i.d. according to a conditional input distribution $P_{X|r_{i_k}}$. We assume $r_{1_k} < \ldots < r_{M_k}$. In each time slot, according to data availability and the MAC protocol, user $k$ chooses a communication rate $r_{i_k} \in \{r_{1_k}, \ldots, r_{M_k}\}$ without sharing this rate information with the receiver or with other users. User $k$ then encodes $\lfloor Nr_{i_k}/\ln 2 \rfloor$ data bits, denoted by a message $w_k$, into the $w_k$th codeword in the $i_k$th codeword class, denoted by $\boldsymbol{x}_{w_k i_k}$, and sends it to the receiver.

Assume the receiver knows the channel and the randomly generated codebooks. Note that this assumption can be satisfied by pre-sharing the random codebook generation algorithms with the receiver. Based on the channel and the codebook information, the receiver predetermines an "operation region" $\mathcal{R}$, which is a set of communication rate vectors under which the receiver *intends* to decode the messages. In each time slot, upon receiving the channel output symbol vector $\boldsymbol{y}$, the receiver outputs the estimated message and rate vector pair $(\hat{\boldsymbol{w}}, \hat{\boldsymbol{r}})$ (that contains the estimates for all users) only if $\hat{\boldsymbol{r}} \in \mathcal{R}$ and a predetermined decoding error probability requirement can be satisfied. Otherwise the receiver outputs "$\hat{\boldsymbol{r}} =$ collision".

To simplify the notations, we will use bold font vector (or matrix) variables to denote the corresponding variables of multiple users. For example, $\hat{\boldsymbol{w}}$ denotes the message estimates of all users, $\boldsymbol{r}$ denotes the communication rates of all users, $\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}$ denotes the input distributions conditioned on communication rates $\boldsymbol{r}$, etc. Let $\mathcal{S} \subset \{1, \ldots, K\}$ be an arbitrary subset of users. We will use $\boldsymbol{r}_{\mathcal{S}}$ to denote the communication rates of users in $\mathcal{S}$, and will use $\boldsymbol{w}_{\bar{\mathcal{S}}}$ to denote the messages of users not in $\mathcal{S}$, etc.

Similar to the single-user case, conditioned on $(\boldsymbol{w}, \boldsymbol{r})$ is transmitted, we define the decoding error probability for $(\boldsymbol{w}, \boldsymbol{r})$ with $\boldsymbol{r} \in \mathcal{R}$ as

$$P_e(\boldsymbol{w}, \boldsymbol{r}) = Pr\{(\hat{\boldsymbol{w}}, \hat{\boldsymbol{r}}) \neq (\boldsymbol{w}, \boldsymbol{r}) | (\boldsymbol{w}, \boldsymbol{r})\}, \forall (\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}. \tag{9}$$

We define the collision miss detection probability for $(\boldsymbol{w}, \boldsymbol{r})$ with $\boldsymbol{r} \notin \mathcal{R}$ as

$$\bar{P}_c(\boldsymbol{w}, \boldsymbol{r}) = Pr\{\hat{\boldsymbol{r}} \neq \text{collision} | (\boldsymbol{w}, \boldsymbol{r})\}, \forall (\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \notin \mathcal{R}. \tag{10}$$

The system error probability is defined as the maximum of (9) and (10).

Assume for all $\boldsymbol{r} \in R$ and for all user subset $\mathcal{S} \subset \{1, \ldots, K\}$, we have $\sum_{k \in \bar{\mathcal{S}}} r_k < I_{\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}}}(\boldsymbol{X}_{\bar{\mathcal{S}}}; Y | \boldsymbol{X}_{\mathcal{S}})$. According to the achievable region result given in [1], asymptotically, the receiver can reliably decode the messages for all rate vectors in $\mathcal{R}$ and can reliably report a collision for all rate vectors outside $\mathcal{R}$. In other words,

$$\begin{aligned} &\lim_{N \to \infty} P_e(\boldsymbol{w}, \boldsymbol{r}) = 0, \qquad \forall (\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}, \\ &\lim_{N \to \infty} \bar{P}_c(\boldsymbol{w}, \boldsymbol{r}) = 0, \qquad \forall (\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \notin \mathcal{R}. \end{aligned} \tag{11}$$

To analyze the convergence of the system error probability, we again specify the detailed decoding operation in the following. Assume $(\boldsymbol{w}, \boldsymbol{r})$ is encoded and transmitted. The receiver outputs an estimate $(\hat{\boldsymbol{w}} = \boldsymbol{m}, \hat{\boldsymbol{r}} = \boldsymbol{r}_{\boldsymbol{h}})$ if both of the following conditions are satisfied,

$$\begin{aligned} &\text{C1: } -\frac{1}{N} \ln Pr\{\boldsymbol{y} | \boldsymbol{x}_{\boldsymbol{m}_{\boldsymbol{h}}}\} < -\frac{1}{N} \ln Pr\{\boldsymbol{y} | \boldsymbol{x}_{\tilde{\boldsymbol{m}}_{\tilde{\boldsymbol{h}}}}\}, \\ &\quad \text{for all } \boldsymbol{x}_{\tilde{\boldsymbol{m}}_{\tilde{\boldsymbol{h}}}} \neq \boldsymbol{x}_{\boldsymbol{m}_{\boldsymbol{h}}}, \boldsymbol{r}_{\boldsymbol{h}}, \boldsymbol{r}_{\tilde{\boldsymbol{h}}} \in \mathcal{R}; \\ &\text{C2: } -\frac{1}{N} \ln Pr\{\boldsymbol{y} | \boldsymbol{x}_{\boldsymbol{m}_{\boldsymbol{h}}}\} < \tau_{\boldsymbol{h}}(\boldsymbol{y}), \end{aligned} \tag{12}$$

where $\tau_{\boldsymbol{h}}(\boldsymbol{y})$, as a function of $\boldsymbol{y}$, is a predetermined threshold corresponding to the codeword vectors with rate $\boldsymbol{r}_{\boldsymbol{h}}$. If there is no codeword vector satisfying (12), the receiver reports $\hat{\boldsymbol{r}} =$ collision.

Define the decoding error exponent $E_d$ and the collision miss detection exponent $E_c$ as

$$\begin{aligned} E_d &= \min_{(\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \in \mathcal{R}} \lim_{N \to \infty} -\frac{1}{N} \ln P_e(\boldsymbol{w}, \boldsymbol{r}) \\ E_c &= \min_{(\boldsymbol{w}, \boldsymbol{r}), \boldsymbol{r} \notin \mathcal{R}} \lim_{N \to \infty} -\frac{1}{N} \ln \bar{P}_c(\boldsymbol{w}, \boldsymbol{r}). \end{aligned} \tag{13}$$

The system error exponent is defined by $E_s = \min\{E_d, E_c\}$.

The following theorem gives a lower bound on $E_s$. Due to the page limitation, proof of the theorem is skipped.

**Theorem 2:** For $K$-user random access system over a discrete time memoryless channel $P_{Y|\boldsymbol{X}}$. Given the conditional input distributions $\boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}_{\boldsymbol{h}}} = \left[P_{X|r_{h_1}}, \cdots, P_{X|r_{h_K}}\right]^T$ for all $\boldsymbol{h} = [h_1, \cdots, h_M]^T$. The system error exponent is bounded as follows.

$$\begin{aligned} E_s \geq \min_{\mathcal{S}} \min \Bigg\{ &\min_{\boldsymbol{h}, \tilde{\boldsymbol{h}}: \boldsymbol{r}_{\boldsymbol{h}}, \boldsymbol{r}_{\tilde{\boldsymbol{h}}} \in \mathcal{R}} \max_{0 \leq \rho \leq 1} -\rho \sum_{i \in \bar{\mathcal{S}}} r_{\tilde{h}_i} \\ &\qquad + E_0(\rho, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}_{\boldsymbol{h}}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}_{\tilde{\boldsymbol{h}}}}), \\ &\min_{\boldsymbol{h}, \tilde{\boldsymbol{h}}: \boldsymbol{r}_{\boldsymbol{h}} \notin \mathcal{R}, \boldsymbol{r}_{\tilde{\boldsymbol{h}}} \in \mathcal{R}} \max_{0 \leq \rho \leq 1} -\rho \sum_{i \in \bar{\mathcal{S}}} r_{\tilde{h}_i} \\ &\qquad + E_1(\rho, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}_{\boldsymbol{h}}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}_{\tilde{\boldsymbol{h}}}}) \Bigg\}, \end{aligned} \tag{14}$$

where,

$$\begin{aligned} E_0(\rho, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}_{\boldsymbol{h}}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r}_{\tilde{\boldsymbol{h}}}}) = \max_{0 \leq s \leq 1} -\ln \sum_{\boldsymbol{x}_{\mathcal{S}}} \prod_{k \in \mathcal{S}} P_{X|r_{h_k}}(x_k) \\ \times \sum_{y} \left( \sum_{\boldsymbol{x}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|r_{h_k}}(x_k) P_{Y|\boldsymbol{X}}(y|\boldsymbol{x})^{1-s} \right) \\ \times \left( \sum_{\boldsymbol{x}_{\bar{\mathcal{S}}}} \prod_{k \notin \mathcal{S}} P_{X|r_{\tilde{h}_k}}(x_k) P_{Y|\boldsymbol{X}}(y|\boldsymbol{x})^{\frac{s}{\rho}} \right)^{\rho}, \end{aligned}$$

$$E_1(\rho, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r_h}}, \boldsymbol{P}_{\boldsymbol{X}|\boldsymbol{r_{\tilde{h}}}}) = \max_{0\le s\le 1-\rho} -\ln \sum_{y}\sum_{\boldsymbol{x}_{\mathcal{S}}} \prod_{k\in\mathcal{S}} P_{X|r_{h_k}}(x_k) P_{Y|\boldsymbol{X}}(y|\ \boldsymbol{x}_{\mathcal{S}})^{1-s} \times \left(\sum_{\boldsymbol{x}_{\bar{\mathcal{S}}}} \prod_{k\notin\mathcal{S}} P_{X|r_{\tilde{h}_k}}(x_k) P_{Y|\boldsymbol{X}}(y|\boldsymbol{x})^{\frac{s}{s+\rho}}\right)^{s+\rho}. \quad (15)$$

■

## APPENDIX

*Proof of Theorem 1*: The main technique used in this proof is motivated by [4].

Let $q(x, y)$ be the empirical distribution of the symbol pairs $(x, y)$, derived from a specific sequence pair $(\boldsymbol{x}, \boldsymbol{y})$. Similarly, let $q(x)$ and $q(y)$ be the empirical marginal distributions, and $q(x|y)$ and $q(y|x)$ be the empirical conditional distributions. When ever empirical distributions are used, we will carefully make sure that the sequence pair $(\boldsymbol{x}, \boldsymbol{y})$ used to derive the empirical distributions is always specified. However, we choose to skip the sequence pair in the notations of the empirical distributions.

Assume the actual transmitted message and rate pair is $(w, r_h)$, which is mapped to codeword $\boldsymbol{x}_{w_h}$. Let $\boldsymbol{y}$ be the channel output sequence, whose empirical distribution is $q(y)$. For each codeword $\boldsymbol{x}_{\tilde{m}_{\tilde{h}}}$, we define $G(\boldsymbol{y}, \boldsymbol{x}_{\tilde{m}_{\tilde{h}}}) = -\frac{1}{N}\ln Pr\{\boldsymbol{y}|\boldsymbol{x}_{\tilde{m}_{\tilde{h}}}\}$, which is the normalized negative log likelihood of $\boldsymbol{x}_{\tilde{m}_{\tilde{h}}}$.

Given $q(y)$, we define the following two exponents, as functions of a threshold variable $\tau_0$,

$$E_{th}(\tau_0) = \lim_{N\to\infty} -\frac{1}{N}\ln Pr\left\{G(\boldsymbol{y}, \boldsymbol{x}_{w_h}) \ge \tau_0\right\},$$
$$E_i^{[\tilde{h},h]}(\tau_0) = \lim_{N\to\infty} -\frac{1}{N}\ln Pr\left\{G(\boldsymbol{y}, \boldsymbol{x}_{\tilde{m}_{\tilde{h}}}) < \tau_0\right\} (16)$$

$E_{th}(\tau_0)$ characterizes the probability that the negative log likelihood of the *actual* transmitted codeword is above a threshold. $E_i^{[\tilde{h},h]}(\tau_0)$ characterizes the probability that the negative log likelihood of an arbitrary codeword (other than the transmitted codeword) in the $\tilde{h}$th class is below the threshold. We use a superscript $[\tilde{h}, h]$ here to indicate that the actual transmitted codeword is in the $h$th class and the codeword considered in the $G(\boldsymbol{y}, \boldsymbol{x}_{\tilde{m}_{\tilde{h}}})$ function is in the $\tilde{h}$th class.

**Step I:** Let $q_h(x|y)$ be the conditional empirical distribution derived from the sequence pair $(\boldsymbol{x}_{w_h}, \boldsymbol{y})$. Let $Pr\{q_h(x|y)\}$ be the total probability of all sequence pairs having conditional empirical distribution $q_h(x|y)$. We have $Pr\left\{G(\boldsymbol{y}, \boldsymbol{x}_{w_h}) \ge \tau_0\right\} = \sum_{q_h(x|y):G(\boldsymbol{y},\boldsymbol{x}_{w_h})\ge\tau_0} Pr\{q_h(x|y)\}$.

Given input distribution $P_{X|r_h}$ and the channel model $P_{Y|X}$, we define $p_h(y) = \sum_x P_{Y|X}(y|x)P_{X|r_h}(x)$ as the channel output symbol distribution, and define $p_h(x|y) = P_{Y|X}(y|x)P_{X|r_h}(x)/p_h(y)$ as the posterior distribution of the symbols of the actual transmitted codeword. Due to large deviation theory, the exponent of $Pr\{q_h(x|y)\}$ is given as in [4] by

$$E(q_h(x|y)) = D(q_h(x|y)||p_h(x|y)) = \sum_{x,y} q(y)q_h(x|y)\ln\frac{q_h(x|y)}{p_h(x|y)}, \quad (17)$$

where $D(\cdot||\cdot)$ is the Kullback-Leibler distance.

Consequently,

$$E_{th}(\tau_0) = \min_{q_h(x|y):G(\boldsymbol{y},\boldsymbol{x}_{w_h})\ge\tau_0} E(q_h(x|y)) = \max_{s_1\ge 0}\min_{q_h(x|y)} E(q_h(x|y)) - s_1\left(G(\boldsymbol{y}, \boldsymbol{x}_{w_h}) - \tau_0\right) (18)$$

According to the definition, $G(\boldsymbol{y}, \boldsymbol{x}_{w_h})$ can be written as

$$G(\boldsymbol{y}, \boldsymbol{x}_{m_h}) = -\sum_{x,y} q(y)q_h(x|y)\ln P_{Y|X}(y|x). \quad (19)$$

Substitute (17) and (19) into (18). We get

$$E_{th}(\tau_0) = \max_{s_1\ge 0}\min_{q_h(x|y)} s_1\tau_0 + \sum_y q(y)\sum_x q_h(x|y)\ln\frac{q_h(x|y)P_{Y|X}(y|x)^{s_1}}{p_h(x|y)} = \max_{s_1\ge 0} s_1\tau_0 - \sum_y q(y)\ln\sum_x p_h(x|y)P_{Y|X}(y|x)^{-s_1}. \quad (20)$$

By following a similarly derivation, we can also write $E_i^{[\tilde{h},h]}(\tau_0)$ as

$$E_i^{[\tilde{h},h]}(\tau_0) = \max_{s_2\le 0} s_2\tau_0 - \sum_y q(y)\ln\sum_x P_{X|r_{\tilde{h}}}(x)P_{Y|X}(y|x)^{-s_2}. \quad (21)$$

**Step II:** Conditioned on that the transmitted communication rate $r_h$ is within the operation region, i.e. $h \le H$, let $E_{dh}(q(y))$ be the decoding error exponent given the empirical distribution $q(y)$ of sequence $\boldsymbol{y}$. A decoding error occurs if either of the conditions in (4) is violated. Let $E_{Ci}(q(y))$ be the exponent of the probability that condition $Ci$ is violated, $i \in \{1, 2\}$. We have $E_{dh}(q(y)) = \min\{E_{C1}(q(y)), E_{C2}(q(y))\}$.

Due to the union bound, the exponent of the probability that any codeword $\boldsymbol{x}_{\tilde{m}_{\tilde{h}}} \ne \boldsymbol{x}_{w_h}$ satisfies $G(\boldsymbol{y}, \boldsymbol{x}_{\tilde{m}_{\tilde{h}}}) < \tau_0$ is given by $\max\left\{0, \min_{\tilde{h}\le H}\left(E_i^{[\tilde{h},h]}(\tau_0) - r_{\tilde{h}}\right)\right\}$. This yields

$$E_{C1}(q(y)) = \min_{\tau_0,\tilde{h}\le H}\max_{0\le\rho\le 1} E_{th}(\tau_0) + \rho\left(E_i^{[\tilde{h},h]}(\tau_0) - r_{\tilde{h}}\right), \quad (22)$$

where $\rho$ is a Lagrange multiplier.

Substitute (20), (21) into (22), and let $s = s_1 + \rho s_2$ to obtain

$$E_{C1}(q(y)) = \min_{\tilde{h}\le H}\max_{0\le\rho\le 1, s_1\ge 0, s}\min_{\tau_0} -\rho r_{\tilde{h}} + s\tau_0 - \sum_y q(y)\ln\left[\left(\sum_x p_h(x|y)P_{Y|X}(y|x)^{-s_1}\right) \times \left(\sum_x P_{X|r_{\tilde{h}}}(x)P_{Y|X}(y|x)^{\frac{s_1-s}{\rho}}\right)^{\rho}\right]. \quad (23)$$

Note that the objective function in (23) increases monotonically in $\tau_0$ and decreases monotonically in $s$. Since the optimal $\tau_0$ is positive and finite, the optimal $s$ should be nonnegative. Consequently, (23) must be optimized at $s = 0$. This yields

$$E_{C1}(q(y)) = \min_{\tilde{h}\le H} \max_{0\le\rho\le 1, s_1\ge 0} -\rho r_{\tilde{h}} - \sum_y q(y) \times \ln\left[\left(\sum_x \frac{P_{X|r_h(x)}}{p_h(y)} P_{Y|X}(y|x)^{1-s_1}\right) \times \left(\sum_x P_{X|r_{\tilde{h}}}(x) P_{Y|X}(y|x)^{\frac{s_1}{\rho}}\right)^{\rho}\right]. \quad (24)$$

Furthermore, we have $E_{C2}(q(y)) = E_{th}(\tau_h(\boldsymbol{y}))$.

Let $E_{dh}$ be the decoding error exponent *without* providing the empirical distribution $q(y)$. We have

$$E_{dh} = \min_{q(y)} \left(E_{dh}(q(y)) + D(q(y)||p_h(y))\right). \quad (25)$$

Define $E_{ih} = \min_{q(y)} \left(E_{Ci}(q(y)) + D(q(y)||p_h(y))\right)$ for $i \in \{1, 2\}$. We have $E_{dh} = \min\{E_{1h}, E_{2h}\}$.

To derive the expression of $E_{1h}$, we have

$$E_{1h} = \min_{\tilde{h}\le H} \max_{0\le\rho\le 1, s_1\ge 0} \min_{q(y)} -\rho r_{\tilde{h}} - \sum_y q(y) \ln\left[q(y)^{-1} \times \left(\sum_x P_{X|r_h}(x) P_{Y|X}(y|x)^{1-s_1}\right) \times \left(\sum_x P_{X|r_{\tilde{h}}}(x) P_{Y|X}(y|x)^{\frac{s_1}{\rho}}\right)^{\rho}\right]. \quad (26)$$

It can be easily verified that $E_{1h} \le 0$ for $s_1 \ge 1$. Therefore, $0 \le s_1 \le 1$ must hold. This consequently leads to

$$E_{1h} = \min_{\tilde{h}\le H} \max_{0\le\rho\le 1} -\rho r_{\tilde{h}} + E_0(\rho, P_{X|r_h}, P_{X|r_{\tilde{h}}}), \quad (27)$$

where $E_0(\rho, P_{X|r_h}, P_{X|r_{\tilde{h}}})$ is defined in (8).

**Step III:** Conditioned on that the transmitted communication rate $r_h$ is outside the operation region, i.e. $h > H$, let $E_{ch}(q(y))$ be the collision miss detection exponent given the empirical distribution $q(y)$. Note that the collision miss detection probability is upper bounded by the probability that any codeword in the operation region satisfies condition C2 in (4). Due to the union bound, the corresponding exponent is lower-bounded by

$$E_{ch}(q(y)) \ge \max_{0\le\tilde{\rho}\le 1} \min_{\tilde{h}\le H} \tilde{\rho}\left(E_i^{[\tilde{h},h]}(\tau_{\tilde{h}}(\boldsymbol{y})) - r_{\tilde{h}}\right) \quad (28)$$

Let $E_{ch}$ be the collision miss detection exponent *without* providing the empirical distribution $q(y)$. We have

$$E_{ch} = \min_{q(y)} \left(E_{ch}(q(y)) + D(q(y)||p_h(y))\right). \quad (29)$$

**Step IV:** The system error exponent is given by

$$E_s = \min\left\{\min_{h\le H} E_{1h}, \max_{\tau_{\tilde{h}}(\boldsymbol{y})} \min\left\{\min_{\tilde{h}\le H} E_{2\tilde{h}}, \min_{h>H} E_{\text{c}h}\right\}\right\}. \quad (30)$$

The first terms on the RHS of (30) and (7) are equal. The second term is lowered bounded by $E_T$, which is defined below.

$$E_T = \max_{\tau_{\tilde{h}}(\boldsymbol{y})} \min\{\min_{\tilde{h}\le H} E_{2\tilde{h}}, \min_{h>H} E_{\text{c}h}\} = \min_{h>H,\tilde{h}\le H} \max_{0\le\tilde{\rho}\le 1} \min_{q(y)} E_T(q(y)), \quad (31)$$

where

$$E_T(q(y)) = \max_{s_1\ge 0, s_2\le 0} \max_{\tau_{\tilde{h}}(\boldsymbol{y})} \min\{Q_1, Q_2\},$$
$$Q_1 = s_1\tau_{\tilde{h}}(\boldsymbol{y}) + \sum_y q(y)\ln q(y) - \sum_y q(y)\ln\sum_x P_{X|r_{\tilde{h}}}(x)P_{Y|X}(y|x)^{1-s_1},$$
$$Q_2 = -\tilde{\rho} r_{\tilde{h}} + \tilde{\rho} s_2 \tau_{\tilde{h}}(\boldsymbol{y}) + \sum_y q(y)\ln q(y) - \sum_y q(y)\ln p_h(y)\left(\sum_x P_{X|r_h}(x)P_{Y|X}(y|x)^{-s_2}\right)^{\tilde{\rho}}. \quad (32)$$

Note that $Q_1$ is monotonically increasing in $\tau_{\tilde{h}}(\boldsymbol{y})$, while $Q_2$ is monotonically decreasing in $\tau_{\tilde{h}}(\boldsymbol{y})$. $\tau_{\tilde{h}}(\boldsymbol{y})$ should be chosen such that $Q_1 = Q_2$. Consequently, we can change the optimization term $\max_{\tau_{\tilde{h}}(\boldsymbol{y})} \min\{\}$ in (32) to $\min_{\tau_{\tilde{h}}(\boldsymbol{y})} \max\{\}$. This gives,

$$E_T(q(y)) = \max_{s_1\ge 0, s_2\le 0} \min_{\tau_{\tilde{h}}(\boldsymbol{y})} \max\{Q_1, Q_2\} = \max_{s_1\ge 0, s_2\le 0} \min_{\tau_{\tilde{h}}(\boldsymbol{y})} \max_{0\le\lambda\le 1} (1-\lambda)Q_1 + \lambda Q_2. \quad (33)$$

Similar to Step **II**, (33) is optimized at $(1-\lambda)s_1 + \lambda\tilde{\rho}s_2 = 0$. Consequently,

$$E_T(q(y)) = \min_{0\le\lambda\le 1} -\lambda r_{\tilde{h}} + \sum_y q(y)\ln\frac{q(y)}{p_h(y)^{\lambda}} - \sum_y q(y)\ln\left(\sum_x P_{X|r_{\tilde{h}}}(x)P_{Y|X}(y|x)^{\frac{1-\lambda}{1-\lambda+\lambda\tilde{\rho}}}\right)^{1-\lambda+\lambda\tilde{\rho}}. \quad (34)$$

Given $\lambda$ and $\tilde{\rho}$, we first optimize (34) over $q(y)$ to obtain $E_T$. Then substitute $\rho = \lambda\tilde{\rho} \in [0, 1]$, $s = 1-\lambda \in [0, 1-\rho]$ into the final solution to show the equality of $E_T$ and the second term on the RHS of (7).

This completes the proof. ■

## References


[1] J. Luo, A. Ephremides "A New Channel Coding Approach for Random Access with Bursty Traffic," submitted to *IEEE Trans. on Inform. Theory*.

[2] R. Gallager, " A Simple Derivation of The Coding Theorem and Some Applications", IEEE Trans. Inform. Theory, Vol. 11, pp. 3-18, Jan. 1965.

[3] L. Weng, S. Pradhan, and A. Anastasopoulos, "Error Exponent Regions for Gaussian Broadcast and Multiple Acsess Channels," IEEE Trans. Inform. Theory, Vol. 54, pp. 2919-2942, Jul. 2008.

[4] A. Montanari, G. Forney, "On Exponential Error Bounds for Random Codes on the DMC," http://www.stanford.edu/ montanar/PAPERS/FILEPAP/dmc.ps, unpublished.