On The Reliability Function of Discrete Memoryless Multiple-Access Channel with Feedback

Mohsen Heidari, Achilleas Anastasopoulos, and S. Sandeep Pradhan

Abstract

We derive a lower and upper bound on the reliability function of discrete memoryless multiple-access channel (MAC) with noiseless feedback and variable-length codes (VLCs). For the upper-bound, we use proof techniques of Burnashev for the point-to-point case. Also, we adopt the techniques used to prove the converse for the feedback-capacity of MAC. For the lower-bound on the error exponent, we present a coding scheme consisting of a data and a confirmation stage. In the data stage, any arbitrary feedback capacity-achieving code is used. In the confirmation stage, each transmitter sends one bit of information to the receiver using a pair of codebooks of size two, one for each transmitter. The codewords at this stage are selected randomly according to an appropriately optimized joint probability distribution. The bounds increase linearly with respect to a specific Euclidean distance measure defined between the transmission rate pair and the capacity boundary. The lower and upper bounds match for a class of MACs.

I. INTRODUCTION

Noiseless feedback does not increase the capacity for communications over discrete memoryless channels (DMC) [1]. Furthermore, Dobrushin [4] and later Haroutunian [5] showed that feedback does not improve the error exponent of symmetric channels when fixed-length codes are used. Nevertheless, feedback can be very useful in the context of variable-length codes.

In a remarkable work, Burnashev [2] demonstrated that the error exponent improves for DMCs with feedback and variable-length codes. The error exponent has a simple form

$$E(R) = (1 - \frac{R}{C})C_1, \tag{1}$$

where R is the (average) rate of transmission, C is the capacity of the channel, and C_1 is the maximal relative entropy between conditional output distributions. Berlin et al [6] have provided a simpler derivation of the Burnashev bound that emphasizes the link between the constant C_1 and the binary hypothesis testing problem. Yamamoto and Itoh [7] introduced a coding scheme that its error exponent achieves $E(R)$ in (1). Their scheme consists of two distinct transmission phases that we called the data and the confirmation phase, respectively. In the data stage the message is encoded using a capacity achieving fixed blocklength code. During the confirmation phase, the transmitter sends one bit of information to the receiver. The decoder performs a binary hypothesis test to decide if 0 or 1 is transmitted.

In the context of communications over multi-user channels, the benefits of feedback are more prominent. For instance, Gaarder and Wolf [8] showed that feedback can expand the capacity region of discrete memoryless multiple-access channels (MAC). Willems [9] derived the feedback-capacity region for a class of MACs. Characterizing the capacity region and the error exponent for general MACs remains an open problem. Using *directed information* measures, Kramer [10] was able to characterize the feedback-capacity region of two-user MAC with feedback. However, the characterization is in the form of infinite letter directed information measures which is not computable in general. The error exponent for discrete memoryless MAC without feedback is studied in [13], [14].

The authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, 48105 USA e-mail: mohsenhd, anastas, pradhanv@umich.edu

In this paper, we study the error exponent of discrete memoryless MAC with noiseless feedback. In particular, we derive an upper-bound and a lower-bound. For that, let $(||\underline{R}||, \theta_R)$ denote the polar coordinate of (R_1, R_2) in $\mathbb{R}^2$. In this setting, the upper-bound is

$$E_u(R_1, R_2) = (1 - \frac{||\underline{R}||}{C(\theta_R)})D_u \qquad (2)$$

where $C(\theta_R)$ is the point of the capacity frontier at the angle determined by $\underline{R}$. The lower-bound is the same as E_u but with different constant D_l. The constants D_l and D_u are determined by the relative entropy between the conditional output distributions. We show that for a class of MACs the two bounds coincide.

The paper is organized as follows: In Section II, basic definitions and the problem formulation are provided. In Section III, we derive a lower-bound for the reliability function. In Section IV, we characterize an upper-bound for the reliability function. In Section V, we compare the lower and upper-bound and explore examples for the tightness of the bounds. Finally, Section VI concludes the paper.

II. PROBLEM FORMULATION AND DEFINITIONS

Consider a discrete memoryless MAC with input alphabets $\mathcal{X}_1, \mathcal{X}_2$, and output alphabet $\mathcal{Y}$. The channel conditional probability distribution is denoted by $Q(y|x_1, x_2)$ for all $(y, x_1, x_2) \in \mathcal{Y} \times \mathcal{X}_1 \times \mathcal{X}_2$. Such setup is denoted by $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, Q)$. Let y^t and x_i^t, $i = 1, 2$, be the channel output and the inputs sequences after t uses of the channel, respectively. Then, the following condition is satisfied:

$$P(y_t|y^{t-1}, x_1^{t-1}, x_2^{t-1}) = Q(y_t|x_{1t}, x_{2t}). \qquad (3)$$

We assume that the output of the channel as a feedback is available at the encoders with one unit of delay.

Definition 1. *An (M_1, M_2, N)- variable-length code (VLC) for a MAC $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, Q)$ with feedback is defined by*
- *A pair of messages W_1, W_2 selected randomly with uniform distribution from $\{1, 2, \ldots, M_i\}, i = 1, 2$.*
- *Two sequences of encoding functions*

$$e_{i,t} : \{1, 2, \ldots, M_i\} \times \mathcal{Y}^{t-1} \to \mathcal{X}_i, \quad t \in \mathbb{N}, \ i = 1, 2,$$

  one for each transmitter.
- *A sequence of decoding functions*

$$d_t : \mathcal{Y}^t \to \{1, 2, ..., M_1\} \times \{1, 2, ..., M_2\}, \quad t \in \mathbb{N}.$$

- *A stopping time T with respect to (w.r.t) the filtration $\mathcal{F}_t$ defined as the σ-algebra of Y^t for $t \in \mathbb{N}$. Furthermore, it is assumed that T satisfies $\mathbb{E}[T] \leq N$.*

For each $i = 1, 2$, given a message W_i, the tth output of Transmitter i is denoted by $X_{i,t} = e_{i,t}(W_i, Y^{t-1})$.

Let $(\hat{W}_{1,t}, \hat{W}_{2,t}) = d_t(Y^t)$. Then, the decoded messages at the decoder are denoted by $\hat{W}_1 = \hat{W}_{1,T}$, and $\hat{W}_2 = \hat{W}_{2,T}$. In what follows, for any (M_1, M_2, N) VLC, we define average rate-pair, error probability, and error exponent. Average rates for an (M_1, M_2, N) VLC are defined as

$$R_i \triangleq \frac{\log_2 M_i}{\mathbb{E}[T]}, \quad i = 1, 2.$$

The probability of error is defined as

$$P_e = P\left((\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\right).$$

The error exponent of a VLC with probability of error P_e and stopping time T is defined as $E \triangleq -\frac{\log_2 P_e}{\mathbb{E}[T]}$.

Definition 2. *A reliability function $E(R_1, R_2)$ is said to be achievable for a given MAC, if for any $R_1, R_2 > 0$ and $\epsilon > 0$ there exists an (M_1, M_2, N)-VLC such that*

$$-\frac{\log_2 P_e}{N} \geq E(R_1, R_2) - \epsilon, \ and \ \frac{\log_2 M_i}{N} \geq R_i - \epsilon,$$

where $i = 1, 2$, and P_e is the error probability of the VLC.

Definition 3. *The reliability function of a MAC with feedback is defined as the supremum of all achievable reliability functions $E(R_1, R_2)$.*

A. The Feedback-Capacity Region of MAC

We summarize Kramer's results presented in [10] for the feedback capacity of MAC. We use *directed information* and *conditional directed information* as defined in [10]. The normalized directed information from a sequence $\mathbf{X}^n$ to a sequence $\mathbf{Y}^n$ when causally conditioned on $\mathbf{Z}^n$ is denoted by

$$I_n(X \to Y||Z) = \frac{1}{n}I(\mathbf{X}^n \to \mathbf{Y}^n||\mathbf{Z}^n). \tag{4}$$

The feedback-capacity region of a discrete memoryless MAC with feedback $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, Q)$ is denoted by $\mathcal{C}$, and is the closure of the set of all rate-pairs (R_1, R_2) such that

$$R_1 \leq I_L(X_1 \to Y||X_2)$$
$$R_2 \leq I_L(X_2 \to Y||X_1)$$
$$R_1 + R_2 \leq I_L(X_1X_2 \to Y),$$

where L is a positive integer, and $P_{X_1^L X_2^L Y^L}$ factors as

$$\prod_{l=1}^{L} P_{1,l}(x_{1l}|x_1^{l-1}y^{l-1})P_{2,l}(x_{2l}|x_2^{l-1}y^{l-1})Q(y_l|x_{1,l}x_{2,l}). \tag{5}$$

Definition 4. *Let $\lambda_1, \lambda_2, \lambda_3 \geq 0$, and $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Define*

$$C_{\underline{\lambda}} = \sup_{L \in \mathbb{N}} \sup_{P_{X_1^L X_2^L Y^L}} \lambda_1 I_L(X_1 \to Y|X_2) + \lambda_2 I_L(X_2 \to Y|X_1)$$
$$+ \lambda_3 I_L(X_1X_2 \to Y),$$

where $P_{X_1^L X_2^L Y^L}$ factors as in (5).

Fact 1. The feedback-capacity of a discrete memoryless MAC with feedback is the same as the closure of the set of rate-pairs (R_1, R_2) such that the inequality

$$\lambda_1 R_1 + \lambda_2 R_2 + \lambda_3(R_1 + R_2) \leq C_{\underline{\lambda}}$$

holds for all $\lambda_1, \lambda_2, \lambda_3 \geq 0$, with $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

B. Notational Conventions

For more convenience, we denote a rate-pair (R_1, R_2) by (R_1, R_2, R_3), where $R_3 = R_1 + R_2$. For a $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, Q)$ MAC we use the following notational convenience

$$I_L^1 \triangleq I_L(X_1 \to Y||X_2), \tag{6}$$
$$I_L^2 \triangleq I_L(X_2 \to Y||X_1), \tag{7}$$
$$I_L^3 \triangleq I_L(X_1X_2 \to Y). \tag{8}$$

The Kullback–Leibler divergence for the MAC with transition probability matrix Q is defined as

$$D_Q(x_1, x_2||z_1, z_2) = \sum_{y \in \mathcal{Y}} Q(y|x_1, x_2) \log_2 \frac{Q(y|x_1, x_2)}{Q(y|z_1, z_2)},$$

where $(x_1, x_2), (z_1, z_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. For notational convenience we denote

$$D_1(x_1, x_2||z_1, z_2) = D_Q(x_1, x_2||z_1, x_2)$$
$$D_2(x_1, x_2||z_1, z_2) = D_Q(x_1, x_2||x_1, z_2)$$
$$D_3(x_1, x_2||z_1, z_2) = D_Q(x_1, x_2||z_1, z_2).$$

III. A LOWER-BOUND FOR THE RELIABILITY FUNCTION

We build upon Yamamoto-Itoh transmission scheme for point-to-point (ptp) channel coding with feedback [7]. The scheme sends the messages W_1, W_2 through blocks of length n. The transmission process is performed in two stages: 1) The "data transmission" stage taking up to $n(1 - \gamma)$ channel uses, 2) The "confirmation" stage taking up to $n\gamma$ channel uses, where γ is a design parameter taking values from $[0, 1]$.

Stage 1: For the first stage, we use any coding scheme that achieves the feedback-capacity of the MAC. The length of this coding scheme is at most $n(1 - \gamma)$. Let $\hat{W}_1, \hat{W}_2$ denote the decoder's estimation of the messages at the end of the first stage. Define the following random variables:

$$H_i = 1\{\hat{W}_i \neq W_i\}, \quad i = 1, 2.$$

Because of the feedback, $\hat{W}_1$ and $\hat{W}_2$ are known at each transmitter. Therefore, at the end of the first stage, transmitter i has access to $W_i, \hat{W}_1, \hat{W}_2$, and H_i, where $i = 1, 2$.

Stage 2: The objective of the second stage is to inform the receiver whether the hypothesis $\Theta_0 : (\hat{W}_1, \hat{W}_2) = (W_1, W_2)$ or $\Theta_1 : (\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)$ is correct. For that, each transmitter employs a code of size two and length γn. The codewords of such codebooks are denoted by two pairs of sequences $(\underline{x}_1(0), \underline{x}_2(0))$ and $(\underline{x}_1(1), \underline{x}_2(1))$ each with elements belonging to $\mathcal{X}_1 \times \mathcal{X}_2$. Fix a joint-type P_n defined over the set $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_1 \times \mathcal{X}_2$ and for sequences of length γn. The sequences $(\underline{x}_1(0), \underline{x}_2(0), \underline{x}_1(1), \underline{x}_2(1))$ are selected randomly among all the sequences with joint-type P_n. During this stage and given H_1, Transmitter 1 sends $\underline{x}_1(H_1)$. Similarly, Transmitter 2 sends $\underline{x}_2(H_2)$.

Decoding: Upon receiving the channel output, the receiver estimates H_1, H_2. Denote this estimation by $\hat{H}_1, \hat{H}_2$. If $(\hat{H}_1, \hat{H}_2) = (0, 0)$, then the hypothesis $\hat{\Theta} = \Theta_0$ is declared. Otherwise, $\hat{\Theta} = \Theta_1$ is declared. Because of the feedback, $\hat{\Theta}$ is also available at each encoders. If $\hat{\Theta} = \Theta_0$, then transmission stops and a new data packet is transmitted at the next block. Otherwise, the message is transmitted again at the next block. The process continues until $\hat{\Theta} = \Theta_0$ occurs.

The confirmation stage in the proposed scheme can be viewed as a decentralized binary hypothesis problem in which a binary hypothesis $\{\Theta_0, \Theta_1\}$ is observed partially by two distributed agents and the objective is to convey the true hypothesis to a central receiver. This problem is qualitatively different from the sequential binary hypothesis testing problem as identified in [6] for ptp channel. Note also that in the confirmation stage we use a different coding strategy than the one used in Yamamoto-Itoh scheme [7]. Here, all four codewords have a joint-type P_n. It can be shown that repetition codes, and more generally, constant composition codes are strictly suboptimal in this problem.

Theorem 1. *The following is a lower-bound for the reliability function of any discrete memoryless MAC:*

$$E_l(R_1, R_2) = \min_{\substack{\lambda_1, \lambda_2, \lambda_3 \geq 0 \\ \lambda_1 + \lambda_2 + \lambda_3 = 1}} D_l(1 - \frac{\sum_i \lambda_i R_i}{C_{\underline{\lambda}}}), \tag{9}$$

where,

$$D_l \triangleq \sup_{P_{X_1 X_2 Z_1 Z_2}} \min_{i=1,2,3} \mathbb{E}\left[D_i(X_1, X_2 || Z_1, Z_2)\right], \tag{10}$$

and the supremum is taken over all probability distributions $P_{X_1 X_2 Z_1 Z_2}$ defined over $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_1 \times \mathcal{X}_2$.

Proof: The proof is given in Appendix A. ∎

IV. AN UPPER-BOUND FOR THE RELIABILITY FUNCTION

In this part of the paper, we establish an upper-bound for the reliability function of any discrete memoryless MAC. Define

$$D_i \triangleq \max_{\substack{x_1, z_1 \in \mathcal{X}_1, \\ x_2, z_2 \in \mathcal{X}_2}} D_i(x_1, x_2 || z_1, z_2), \quad i = 1, 2, 3. \tag{11}$$

Theorem 2 (Upper-bound). *For any (N, M_1, M_2) VLC with probability of error P_e, and any $\epsilon > 0$, there exists a function δ such that the following is an upper-bound for the reliability function of the VLC*

$$E(R_1, R_2) \leq \min_{\substack{\lambda_1, \lambda_2, \lambda_3 \geq 0 \\ \lambda_1 + \lambda_2 + \lambda_3 = 1}} \min_{j \in \{1,2,3\}} D_j \left(1 - \frac{\lambda_j R_j}{C_\lambda} \right)$$
$$+ \delta(P_e, M_1 M_2, \epsilon), \tag{12}$$

where (R_1, R_2) is the rate pair of the VLC and δ satisfies

$$\lim_{\epsilon \to 0} \lim_{P_e \to 0} \lim_{M_1 M_2 \to \infty} \delta(P_e, M_1 M_2, \epsilon) = 0.$$

Corollary 1. *From Theorem 2, the following is an upper-bound for the error exponent of a MAC:*

$$E_u(R_1, R_2) = \min_{\substack{\lambda_1, \lambda_2, \lambda_3 \geq 0 \\ \lambda_1 + \lambda_2 + \lambda_3 = 1}} D_u \left(1 - \frac{\sum_{i=1}^3 \lambda_i R_i}{C_\lambda} \right) + \delta,$$

where $D_u = \max\{D_1, D_2, D_3\}$, and δ is as in Theorem 2.

Proof: The proof is given in Appendix E. ∎

A. Proof of the Upper-Bound

Consider any (N, M_1, M_2) VLC with probability of error P_e, and stopping time T. Suppose the message at Encoder 2, W_2, is made available to all terminals. For the new setup, as W_2 is available at the Decoder, the average probability of error is $P_e^1 \triangleq P\{\hat{W}_1 \neq W_1\}$. Note that $P_e \geq P_e^1$. We refer to such setup as W_2-assisted MAC. For a maximum a *posteriori* decoder, after n uses of the channel and assuming the realization $Y^n = y^n$ and $W_2 = w_2$, define

$$T_1^\delta \triangleq \inf \left\{ n : \max_{1 \leq i \leq M_1} P(W_1 = i | y^n, w_2) \geq 1 - \delta \right\},$$

where $\delta > 0$ is a fixed real number. Also, let $\tau_1 \triangleq \min\{T, T_1^\delta\}$. Note that τ_1 is a stopping time w.r.t the filtration $\{\mathcal{F}_{W_2} \times \mathcal{F}_t\}_{t>0}$. The following lemma provides a lower-bound on the probability of error for such setup.

Lemma 1. *The probability of error, P_e, for a hypothesis testing over a W_2-assisted MAC and variable length codes satisfies the following inequality*

$$P_e \geq \frac{\min\{P(H), P(H^c)\}}{4} e^{-D_1 \mathbb{E}[T]},$$

where $\{H, H^c\}$ are the two hypothesizes and T is the stopping time of the variable length code.

Lemma 2. *For a given MAC with finite D_3 the following holds*

$$\zeta p(w_1, w_2 | y^{n-1}) \leq p(w_1, w_2 | y^n) \leq \frac{p(w_1, w_2 | y^{n-1})}{\zeta},$$

where $\zeta \triangleq \min_{x_1, x_2, y} Q(y | x_1, x_2)$.

The above lemmas are extensions of Lemma 1 and Proposition 2 in [6] for MAC. The proofs follow from similar arguments and are omitted.

Lemma 3. *Given a MAC with $D_3 < \infty$, and for any (N, M_1, M_2) VLC with probability of error P_e the following holds*

$$P_e \geq \frac{\zeta \delta}{4} e^{-D_1 \mathbb{E}[T - \tau_1]}, \tag{13}$$

where $\zeta \triangleq \min_{x_1, x_2, y} Q(y | x_1, x_2)$.

Proof: Suppose the VLC is used for a W_2-assisted MAC. As discussed before, $P_e \geq P_e^1$. We modify the encoding and the decoding functions of the VLC used for the MAC. Let $\mathcal{H}_1 \subseteq \mathcal{M}_1$ be a subset of the message set

$\mathcal{M}_1$. The subset $\mathcal{H}_1$ is to be determined at time τ_1. The new decoding function, at time T, decides whether the message belongs to $\mathcal{H}_1$. The new encoding functions are the same as the original one until the time τ_1. Then, after τ_1, the transmitters perform a VLC to resolve the binary hypothesis $\{W_1 \in \mathcal{H}_1\}$ and $\{W_1 \notin \mathcal{H}_1\}$. This hypothesis problem is performed from τ_1 to T. With these modifications, the error probability of this binary hypothesis problem is a lower-bound on P_e. In what follows, we present a construction for $\mathcal{H}_1$. Then, we apply Lemma 1 to complete the proof.

Let $P_e^1(y^n, w_2) \triangleq 1 - \max_{1 \le i \le M_1} P(W_1 = i | y^n, w_2)$. The quantity $P_e^1(y^{\tau_1}, w_2)$ can be calculated at all terminals. By definition, at time $\tau_1 - 1$, the inequality $P(W_1 = i | Y^{\tau_1 - 1}, W_2) < 1 - \delta$ holds almost surely for all $i \in [1 : M_1]$. This implies that $P_e^1(Y^{\tau_1 - 1}, W_2) > \delta$. Hence, by Lemma 2 at time τ_1 the inequality $P_e^1(Y^{\tau_1}, W_2) \ge \zeta\delta$ holds almost surely. We consider two cases $P_e^1(y^{\tau_1}, w_2) \le \delta$ and $P_e^1(y^{\tau_1}, w_2) > \delta$, where δ is the constant used in the definition of T_1^δ. For the first case, $\mathcal{H}_1$ is the set consisting of the message with the highest *a posteriori* probability. Since $P_e^1(y^{\tau_1}, w_2) \le \delta$, then $P(\mathcal{H}_1) \ge 1 - \delta$. In addition, as $P_e^1(y^{\tau_1}, w_2) \ge \zeta\delta$, then $P(\mathcal{H}_1^c) > \zeta\delta$. For the second case, set $\mathcal{H}_1$ to be a set of messages such that $P(\mathcal{H}_1) > \delta/2$ and $P(\mathcal{H}_1) < 1 - \delta$. Such set exists, since $P(W_1 = i | Y^{\tau - 1}, W_2) < 1 - \delta$ holds for all messages $i \in [1 : M_1]$.

Note that by the above construction, for each case, $P(\mathcal{H}_1) \in [\zeta\delta, 1 - \zeta\delta]$. Thus, from Lemma 1 and the argument above, the inequality

$$P\{\hat{W}_1 \ne W_1 | Y^\tau, W_2\} \ge \frac{\zeta\delta}{4} e^{-D_1 \mathbb{E}[T - \tau | Y^\tau, W_2]}$$

holds almost surely. Next, we take the expectation of the above expression. The lemma follows by the convexity of e^{-x} and Jensen's inequality.

∎

Next, we apply the same argument for the case where W_1 is available at all the terminals. For that define

$$T_2^\delta \triangleq \inf\left\{n : \max_{1 \le j \le M_2} P(W_2 = j | y^n, w_1) \ge 1 - \delta\right\},$$

and let $\tau_2 \triangleq \min\{T, T_2^\delta\}$. By symmetry, Lemma 3 holds for this case and we obtain

$$P_e \ge \frac{\zeta\delta}{4} e^{-D_2 \mathbb{E}[T - \tau_2]}. \tag{14}$$

Next, define the following stopping times:

$$T_3^\delta \triangleq \inf\left\{n : \max_{i,j} P(W_1 = i, W_2 = j | y^n) \ge 1 - \delta\right\}.$$

Also, let $\tau_3 = \min\{T, T_3^\delta\}$. using a similar argument as in the above, we can show that

$$P_e \ge \frac{\zeta\delta}{4} e^{-D_3 \mathbb{E}[T - \tau_3]}. \tag{15}$$

For that, after time τ_3, we formulate a binary hypothesis problem in which the transmitters determine whether $(W_1, W_2) \in \mathcal{H}_3$ or not. Here, $\mathcal{H}_3$ is a subset which is constructed using a similar method as for $\mathcal{H}_1$ in the proof of Lemma 3. We further allow the transmitters to communicate with each other after τ_3. The maximum of the right-hand sides of (13), (14) and (15) gives a lower-bound on P_e. The lower-bound depends on the expectation of the stopping times $\tau_i, i = 1, 2, 3$. In what follows, we provide a lower-bound on $\mathbb{E}[\tau_i]$. Define the following random processes.

$$H_t^1 \triangleq H(W_1 | \mathcal{F}_{W_2} \times \mathcal{F}_t),$$
$$H_t^2 \triangleq H(W_2 | \mathcal{F}_{W_1} \times \mathcal{F}_t),$$
$$H_t^3 \triangleq H(W_1, W_2 | \mathcal{F}_t),$$

Lemma 4. *Given a (M_1, M_2, N)-VLC, for any $\epsilon > 0$ there exist L and a probability distribution $P_{X_1^L X_2^L Y^L}$ that factors as in* (5) *such that the following inequalities hold almost surely for $1 \le t \le N$*

$$\mathbb{E}[H_{t+1}^1 - H_t^1 | \mathcal{F}_{W_2} \times \mathcal{F}_t] \ge -(I_L^1 + \epsilon),$$
$$\mathbb{E}[H_{t+1}^2 - H_t^2 | \mathcal{F}_{W_1} \times \mathcal{F}_t] \ge -(I_L^2 + \epsilon),$$
$$\mathbb{E}[H_{t+1}^3 - H_t^3 | \mathcal{F}_t] \ge -(I_L^3 + \epsilon).$$

where $i = 1, 2, 3$, and I_L^i is defined as in (6)-(8).

Proof: The proof is provided in Appendix B. ∎

We need the following lemma to proceed. The lemma is a result of Lemma 4 in [2], and we omit its proof.

Lemma 5. *For any $t \geq 1$ and $i = 1, 2, 3$, the following inequality holds almost surely w.r.t $\mathcal{F}_{W_1} \times \mathcal{F}_{W_2} \times \mathcal{F}_t$*

$$\log H_t^i - \log H_{t+1}^i \leq \max_{\substack{j,l \in [1:M_1] \\ k,m \in [1:M_2]}} \max_{y \in \mathcal{Y}} \frac{\hat{Q}_{j,k}(y)}{\hat{Q}_{l,m}(y)}.$$

From Lemma 4 and the fact that $H_t^i \leq \log_2 M_i < \infty$, the processes $\{H_t^i + (I_L^1 + \epsilon)t\}_{t>0}$ are submartingales for $i = 1, 2, 3$. In addition, from Lemma 5 and the inequalities $\mathbb{E}[\tau_i] \leq \mathbb{E}[T] \leq N < \infty$, we can apply Doob's Optional Stopping Theorem for each submartingale $\{H_t^i + (I_L^1 + \epsilon)t\}_{t>0}$. Then, we get:

$$\log M_i \leq \mathbb{E}[H_{\tau_i}^i] + \mathbb{E}[\tau_i](I_L^i + \epsilon) \tag{16}$$

where $M_3 = M_1 M_2$.

Lemma 6. *The following inequality holds for each $i = 1, 2, 3$*

$$\mathbb{E}[H_{\tau_i}^i] \leq h_b(\delta) + (\delta + \frac{P_e}{\delta}) \log_2 M_i.$$

Proof: We prove the lemma for the case $i = 1$. The proof for $i = 2, 3$ follows from a similar argument. For $i = 1$, we obtain

$$\mathbb{E}[H_{\tau_1}^1] = P\{P_e(Y^{\tau_1}, W_2) > \delta\}\mathbb{E}[H_{\tau_1}^i | P_e(Y^{\tau_1}, W_2) > \delta] + P\{P_e(Y^{\tau_1}, W_2) \leq \delta\}\mathbb{E}[H_{\tau_1}^1 | P_e(Y^{\tau_1}, W_2) \leq \delta]$$
$$\leq P\{P_e(Y^{\tau_1}, W_2) > \delta\} \log_2 M_1 + P\{P_e(Y^{\tau_1}, W_2) \leq \delta\}\mathbb{E}[H_{\tau_1}^i | P_e(Y^{\tau_1}, W_2) \leq \delta]. \tag{17}$$

Note that the event $\{P_e(Y^{\tau_1}, W_2) > \delta\}$ implies that $\tau_1 = T$, and $P_e(y^{\tau_1}, W_2) > \delta$ for all $0 \leq n \leq T$. Hence, this event is included in the event $\{P_e(Y^T, W_2) > \delta\}$. Thus, applying Markov inequality gives

$$P\{P_e(Y^{\tau_1}, W_2) > \delta\} \leq P\{P_e(Y^T, W_2) > \delta\} \leq \frac{P_e}{\delta}.$$

As a result of the above argument, the right-hand side of (17) does not exceed the following

$$\frac{P_e}{\delta} \log_2 M_1 + \mathbb{E}[H_{\tau_1}^1 | P_e(Y^{\tau_1}, W_2) \leq \delta].$$

From Fano's inequality we obtain

$$\mathbb{E}[H_{\tau_1}^1 | P_e(Y^{\tau_1}, W_2) \leq \delta] \leq h_b(\delta) + \delta \log_2 M_1.$$

The proof is complete from the above inequality. ∎

As a result of the above lemma and (16), the inequality $\mathbb{E}[\tau_i] \geq \frac{\log M_i}{I_L^i + \epsilon} - \frac{h_b(\delta)}{I_L^i + \epsilon}$ holds. Finally, combining this inequality with (13)-(15) completes the proof of the theorem.

B. An Alternative Proof for the Upper-Bound

In this part of the paper, we provide a series of Lemmas that are used to prove the Theorem. Define the following random processes.

Lemma 7. *For an (M_1, M_2, N)-VLC with probability of error P_e the following inequality holds*

$$\mathbb{E}[H_T^i] \leq h_b(P_e) + P_e \log_2(M_1 M_2 - 1), \quad for \quad i = 1, 2, 3.$$

Proof: The proof follows from Fano's Lemma as in [2]. ∎

Lemma 8. *There exists $\epsilon > 0$ such that, if $H_t^i \leq \epsilon$, then*

$$\mathbb{E}[\log H_{t+1}^1 - \log H_t^1 | \mathcal{F}_{W_2} \times \mathcal{F}_t] \geq -(D_1 + \epsilon),$$
$$\mathbb{E}[\log H_{t+1}^2 - \log H_t^2 | \mathcal{F}_{W_1} \times \mathcal{F}_t] \geq -(D_2 + \epsilon),$$
$$\mathbb{E}[\log H_{t+1}^3 - \log H_t^3 | \mathcal{F}_t] \geq -(D_3 + \epsilon)$$

holds almost surely, where $D_i, i = 1, 2, 3$ are defined in (11).

Proof: The proof is given in Appendix C. ∎

Lemma 9. *For $i = 1, 2, 3$, define random process $\{Z_t^{(i)}\}_{t \geq 1}$ as*

$$Z_t^{(i)} = \left(\frac{\log H_t^i - \log \epsilon}{D_i} + t + f_i(\log \frac{H_t^i}{\epsilon}) \right) \mathbb{1}\{H_t^i \leq \epsilon\}$$
$$+ \left(\frac{H_t^i - \epsilon}{I_L^i} + t \right) \mathbb{1}\{H_t^i \geq \epsilon\} \tag{18}$$

where the function f_i is defined as $f_i(y) = \frac{1 - e^{-\mu_i y}}{D_i \mu_i}$. Then, there exists $\mu_i > 0$ such that $Z_t^{(i)}$ is a submartingale w.r.t $\mathcal{F}_{W_1} \times \mathcal{F}_{W_2} \times \mathcal{F}_t$.

Outline of the proof: Suppose $W_2 = m$ for some $m \in [1 : M_2]$. Given this event and using the same argument as in the proof of Theorem 1 in [2] we can show that $Z_t^{(i)}|W_2 = m$ is a submartingale for all m. More precisely, the inequality

$$\mathbb{E}\{Z_t^{(i)} - Z_{t+1}^{(i)}|\mathcal{F}_{W_1} \times \mathcal{F}_{W_2}\} \leq 0,$$

holds almost surely w.r.t $\mathcal{F}_{W_1} \times \mathcal{F}_{W_2}$. Taking the expectation of the both sides in the above inequality gives

$$\mathbb{E}\{Z_t^{(i)} - Z_{t+1}^{(i)}\} \leq 0, \quad \forall t \geq 0, \ i = 1, 2, 3.$$

Thus, $Z_t^{(i)}$ is a submartingale for $i = 1, 2, 3$ and w.r.t $\mathcal{F}_{W_1} \times \mathcal{F}_{W_2} \times \mathcal{F}_t$. ∎

Corollary 2. *Suppose $\alpha_1, \alpha_2, \alpha_3$ are non-negative numbers such that $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Define $Z_t = \alpha_1 Z_t^{(1)} + \alpha_2 Z_t^{(2)} + \alpha_3 Z_t^{(3)}$. Then, Z_t is a submartingale w.r.t $\mathcal{F}_{W_1} \times \mathcal{F}_{W_2} \times \mathcal{F}_t$.*

The Theorem follows from the above lemma, and the proof is given in Appendix D.

V. THE SHAPE OF THE LOWER AND UPPER BOUNDS

In this Section, we point out a few remarks on $E_u(R_1, R_2)$ and the lower-bound $E_l(R_1, R_2)$ defined in Theorem 1. Furthermore, we provide an alternative representation for the bounds and show that the lower and upper-bounds match for a class of MACs.

We first compare the lower bound in (9) and the upper-bound in Corollary 1. For a given arbitrary rate pair (R_1, R_2) inside the feedback-capacity of a given MAC, consider a sequence of VLCs with rates (R_1, R_2) and with average probability of error approaching zero. Then, the following holds:

$$\lim_{\epsilon \to 0} \lim_{P_e \to 0} \lim_{M_1 M_2 \to \infty} \frac{E_u(R_1, R_2)}{E_l(R_1, R_2)} = \frac{D_u}{D_l}$$

As a result of the above remark, it is concluded that for small enough probability of error, the bounds are different only in the constants D_u and D_l.

Next, provide an alternative representation for the lower/upper-bound. For that, suppose (R_1, R_2) is a point inside the capacity region $\mathcal{C}$. By $(\|\underline{R}\|, \theta_R)$ denote the polar coordinate of (R_1, R_2) in $\mathbb{R}^2$. It is shown in the following Remark that the optimum $\underline{\lambda}$ in E_u and E_l is independent of the Euclidean norm of (R_1, R_2), i.e., $\|\underline{R}\|$.

Remark 1. Given an arbitrary $\alpha > 0$ and a rate pair (R_1, R_2) in the capacity region, the optimum $\underline{\lambda}$ for $E_l(R_1, R_2)$ is the same as the one for $E_l(\alpha R_1, \alpha R_2)$.

Proof: Note that one can write $E_l(R_1, R_2)$ as

$$E_l(R_1, R_2) = D_l \left(1 - \max_{\substack{\lambda_1, \lambda_2, \lambda_3 \geq 0 \\ \lambda_1 + \lambda_2 + \lambda_3 = 1}} \frac{\sum_{i=1}^{3} \lambda_i R_i}{C_\lambda} \right),$$
$$= D_l \left(1 - \frac{\sum_{i=1}^{3} \lambda_i^* R_i}{C_{\lambda^*}} \right),$$

where $\underline{\lambda}^*$ is the optimum $\underline{\lambda}$ for E_l. Next, replace (R_1, R_2) with $(\alpha R_1, \alpha R_2)$ for some constant $\alpha > 0$. Then, we obtain

$$E_l(\alpha R_1, \alpha R_2) = D_l \left(1 - \alpha \max_{\substack{\lambda_1, \lambda_2, \lambda_3 \geq 0 \\ \lambda_1 + \lambda_2 + \lambda_3 = 1}} \frac{\sum_{i=1}^3 \lambda_i R_i}{C_\lambda} \right),$$

$$\overset{(a)}{=} D_l \left(1 - \alpha \frac{\sum_{i=1}^3 \lambda_i^* R_i}{C_{\lambda^*}} \right),$$

where (a) follows as the objective function for the maximization is the same as the one in $E_l(R_1, R_2)$. This implies that there is an identical $\underline{\lambda}^*$ which optimizes the expression in $E_l(R_1, R_2)$ and $E_l(\alpha R_1, \alpha R_2)$. ∎
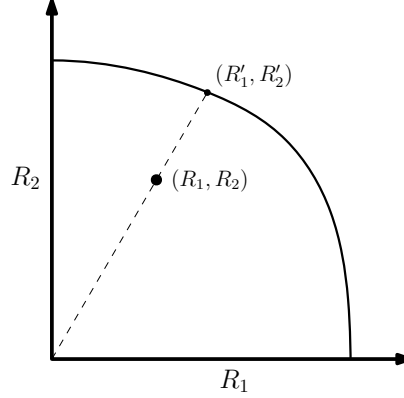

Fig. 1. Given a rate pair (R_1, R_2) which is inside the capacity region, consider the line passing (R_1, R_2) and the origin. Then, (R_1', R_2') is the point of intersection of this line with the boundary of the capacity region.

Now, consider the line passing (R_1, R_2) and the origin. Let (R_1', R_2') denote the point of intersection of this line with the boundary of the capacity region. Fig. 1 shows how (R_1', R_2') is determined. Since, $R_i' = \alpha R_i, i = 1, 2$ for some $\alpha > 0$, then the optimum $\underline{\lambda}$ in $E_l(R_1', R_2')$ is the same as the one in $E_l(R_1, R_2)$. Therefore, from this argument and the fact that $R_i = \frac{R_i'}{\alpha}, i = 1, 2$, we can rewrite $E_l(R_1, R_2)$ as

$$E_l(R_1, R_2) = \min_{\substack{\lambda_1, \lambda_2, \lambda_3 \geq 0 \\ \lambda_1 + \lambda_2 + \lambda_3 = 1}} D_l \left(1 - \frac{1}{\alpha} \frac{\sum_{i=1}^3 \lambda_i R_i'}{C_\lambda} \right),$$

$$\overset{(a)}{=} D_l \left(1 - \frac{1}{\alpha} \right),$$

where (a) follows, since (R_1', R_2') is on the capacity boundary. Note that $\alpha = \frac{\|R\|}{\|R'\|}$. Therefore, $E_l(R_1, R_2) = D_l \left(1 - \frac{\|R\|}{\|R'\|} \right)$. Moreover, note that $\|R'\|$ depends on (R_1, R_2) only through θ_R; in particular, it equals to $C(\theta_R)$ which is a function of θ_R. With this notation, we can rewrite E_l as

$$E_l(R_1, R_2) = D_l \left(1 - \frac{\|R\|}{C(\theta_R)} \right)$$

Using a similar argument for E_u, we have

$$E_u(R_1, R_2) = D_u \left(1 - \frac{\|R\|}{C(\theta_R)} \right) + \delta.$$

As a conclusion of the above argument, the lower (upper) bound increases linearly with respect to a specific Euclidean distance measure defined between the transmission rate pair and the capacity boundary. Fig. 2 shows the shape of a typical upper (lower) bound as a function of the transmission rate pairs.
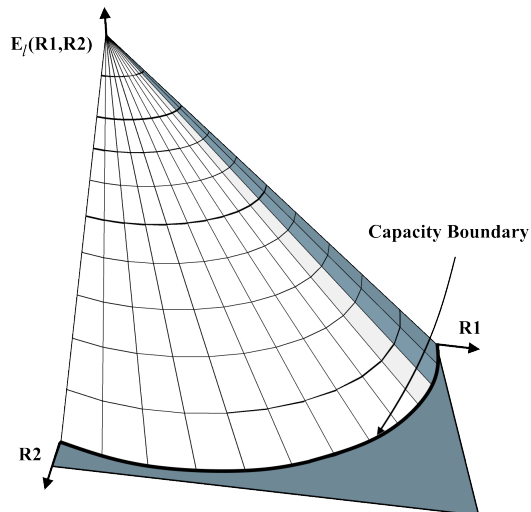
Fig. 2. The conceptual shape of the lower/upper-bound on the error exponent of a given MAC with respect to the transmission rate pair (R_1, R_2).

A. On the Tightness of the Bounds on the Error Exponent

In what follows, we provide examples of classes of channels for which the lower and upper bound coincide.

Example 1. Consider a MAC in which the output is (Y_1, Y_2) and the transition probability matrix is described by the product $Q_{Y_1|X_1} Q_{Y_2|X_2}$. This MAC consists of two parallel (independent) point-to-point channels. Suppose, C_1 and C_2 are the capacity of the first and the second parallel channel, respectively. For this MAC, one can use two parallel Yamamoto-Itoh schemes, one for each channel. Based on the results for the point-to-point case, it is not difficult to show that the error exponent for such MAC satisfies

$$E(R_1, R_2) \geq \min\{D_1(1 - \frac{R_1}{C_1}), D_2(1 - \frac{R_2}{C_2})\}, \tag{19}$$

where C_1 and C_2 are the point-to-point capacity of the channel corresponding to $Q_{Y_1|X_1}$ and $Q_{Y_2|X_2}$, respectively. Note that this lower-bound is not covered by the proposed coding strategy given in Section III. For such MAC, the upper-bound given in (12) is simplified to

$$E(R_1, R_2) \leq \min_{\lambda_1, \lambda_2 \geq 0} \min_{j \in \{1,2\}} D_j \left(1 - \frac{\lambda_j R_j}{\lambda_1 C_1 + \lambda_2 C_2}\right) + \delta.$$

The right-hand side of the above inequality is further upper-bounded by substituting $(\lambda_1, \lambda_2) = (0, 1)$ or $(\lambda_1, \lambda_2) = (1, 0)$. Therefore, we obtain

$$E(R_1, R_2) \leq \min_{j \in \{1,2\}} D_j \left(1 - \frac{R_j}{C_j}\right) + \delta$$

By letting $\delta \to 0$ as in Theorem 2, the above bound can be made arbitrary close to the lower-bound given in (19).

Example 2. Consider a MAC with input alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1, 2\}$, and output alphabet $\mathcal{Y} = \{0, 1, 2\}$. The transition probability of the channel is described by the following relation:

$$Y = X_1 \oplus_3 X_2 \oplus_3 N_p,$$

where the additions are modulo-3 addition, and N_p is a random variable with $P(N_p = 1) = P(N_p = 2) = p$, and $P(N_p = 0) = 1 - 2p$, where $0 \leq p \leq 1/2$. It can be shown that for this channel $D_l = D_u = (1 - 3p) \log \frac{1-2p}{p}$. Hence, the upper-bound in Corollary 1 can be made arbitrary close to the lower-bound in Theorem 1.

The argument in the above example can be extended to m-ary additive MACs for $m > 2$, where the transition probability of the channel is described by

$$Y = X_1 \oplus_m X_2 \oplus_m N_p,$$

where all the random variables take values from $\mathbb{Z}_m$, and N_p is a random variable with $P(N_p = i) = p$ for any $i \in \mathbb{Z}_m$, $i \neq 0$ and $P(N_p = 0) = 1 - (m-1)p$. It can be shown that for this channel

$$D_l = D_u = (1 - mp) \log \frac{1 - (m-1)p}{p}.$$

VI. CONCLUSION

We derive a lower and upper-bound on the reliability function of discrete memoryless MAC with noiseless feedback and variable-length codes. For the lower-bound, we adapt Yamamoto and Itoh's coding scheme consisting of a data and a confirmation phase. For the upper-bound, we adopt the proof techniques of Burnashev for the reliability function of the point-to-point case. The two bounds have the same shape with the difference being the constants at zero rate. We identify sequential binary hypothesis testing problems that are used to derive the upper-bound. We show that the bounds are tight for a class of MACs.

APPENDIX A
PROOF OF THEOREM 1

Proof: At each block a re-transmission occurs with probability q, an error occurs with probability P_{eb} and a correct decoding process happens with probability $1 - q - P_{eb}$. The probability of a re-transmission at each block is

$$q = P(\hat{\Theta} = \Theta_1).$$

The probability of error at each block is

$$P_{eb} = P(\Theta_1)P(\hat{\Theta} = \Theta_0|\Theta_1).$$

Therefore, with this setting the total probability of error for the transmission of a message is

$$P_e = \sum_{k=0}^{\infty} q^k P_{eb} = \frac{P_{eb}}{1 - q}. \tag{20}$$

The number of blocks required to complete the transmission of one message is a geometric random variable with probability of success $1 - q$. Thus, the expected number of blocks for transmission of a message is $\frac{1}{1-q}$.

Next, we derive an upper-bound for q and P_{eb}. For shorthand, denote $H_{12} = (H_1, H_2), \hat{H}_{12} = (\hat{H}_1, \hat{H}_2)$. Then

$$P_{eb} = P\left(\hat{H}_{12} = 00, H_{12} \neq 00\right)$$
$$= \sum_{a \in \{01,10,11\}} P(H_{12} = a)P(\hat{H}_{12} = 00|H_{12} = a).$$

Note that the effective rates of this transmission scheme are $(\frac{R_1}{1-\gamma}, \frac{R_2}{1-\gamma})$. Suppose $(\frac{R_1}{1-\gamma}, \frac{R_2}{1-\gamma})$ is inside the feedback-capacity region of the channel. Then, from the definition of the capacity region, there exist a sequence $\zeta_n, n \geq 1$ with $\zeta_n \to 0$ such that after the first stage

$$P((\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)) \leq \zeta_n.$$

Equivalently, the effective rates are inside the capacity region, if the following inequality holds for any $\lambda_i \geq 0, i = 1, 2, 3$:

$$\frac{1}{1 - \gamma}(\lambda_1 R_1 + \lambda_2 R_2 + \lambda_3(R_1 + R_2)) < C_{\underline{\lambda}}, \tag{21}$$

where $C_{\underline{\lambda}}$ is given in Definition 4. Denote $R_3 = R_1 + R_2$ and define

$$\gamma^* = \min_{\substack{\lambda_1,\lambda_2,\lambda_3 \geq 0 \\ \lambda_1+\lambda_2+\lambda_3=1}} (1 - \frac{\sum_i \lambda_i R_i}{C_{\underline{\lambda}}}). \tag{22}$$

Then, (21) implies that $\gamma < \gamma^*$. The probability of error is therefore bounded by

$$P_{eb} \leq \sum_{a \in \{01,10,11\}} P(\hat{H}_{12} = 00|H_{12} = a) \tag{23}$$

Suppose $(X_1(0), X_1(1), X_2(0), X_2(1))$ are random variables with joint distribution P_n. Then for $i, j \in \{0, 1\}$ define

$$\bar{D}_{\mathsf{P}_n}(00||ij) = \mathbb{E}_{\mathsf{P}_n}\Big[D_Q\big(X_1(0), X_2(0)||X_1(i), X_2(j)\big)\Big].$$

From the description of the transmission scheme, the codewords for the confirmation stage are selected with joint-type P_n. In addition, the decoding process is performed using ML decoding. Therefore, the following bounds hold for $a \in \{01, 10, 11\}$:

$$P(\hat{H}_{12} = 00|H_{12} = a) \leq 2^{-n\gamma\bar{D}_{\mathsf{P}_n}(00||a)}.$$

Thus, from (23), the probability of error is upper bounded by

$$P_{eb} \leq 3 \times 2^{-n\gamma D_{l,n}} \tag{24}$$

Where $D_{l,n} = \max_{\mathsf{P}_n} \min_{a \in \{01,10,11\}} \bar{D}_{\mathsf{P}_n}(00||a)$.

Next we derive an upper bound for q. We have

$$\begin{aligned}
q &= P(\hat{\Theta} = \Theta_1) \\
&= P(\Theta_0)P(\hat{\Theta} = \Theta_1|\Theta_0) + P(\Theta_1)P(\hat{\Theta} = \Theta_1|\Theta_1) \\
&\leq P(\hat{\Theta} = \Theta_1|\Theta_0) + \zeta_n,
\end{aligned}$$

where the last inequality holds because of the following inequalities 1) $P(\Theta_1) \leq \zeta_n$, and 2) $P(\Theta_0), P(\hat{\Theta} = \Theta_1|\Theta_1) \leq 1$. Note that

$$\begin{aligned}
P(\hat{\Theta} = \Theta_1|\Theta_0) &= \sum_{a \in \{01,10,11\}} P(\hat{H}_{12} = a|H_{12} = 00) \\
&\leq \sum_{a \in \{01,10,11\}} 2^{-n\gamma\bar{D}_{\mathsf{P}_n}(a||00)} \\
&\leq 3 \times 2^{-n\gamma\tilde{D}_{l,n}},
\end{aligned}$$

where $\tilde{D}_{l,n} = \min_{a \in \{01,10,11\}} \bar{D}_{\mathsf{P}_n}(a||00)$. Therefore, there exists a sequence $\{q_n\}_{n \geq 1}$ with $q_n \to 0$ such that $q < q_n + \zeta_n$. Using this inequality and the inequality at (24), we derive the following upper-bound for the total probability of error given in (20)

$$P_e \leq \frac{3}{1 - q_n - \zeta_n} 2^{-n\gamma D_{l,n}}.$$

Therefore, the error exponent is bounded from below as

$$\frac{-\log_2 P_e}{\mathbb{E}[T]} \geq \sup \frac{\gamma D_{l,n}}{(1 - q_n - \zeta_n)} + \xi_n$$

where $\xi_n = \frac{1}{n}\frac{\log_2(\frac{1-q_n-\zeta_n}{3})}{1-q_n-\zeta_n}$. Note that for any $\epsilon > 0$ there exists large enough n such that $q_n + \zeta_n < \epsilon, D_{l,n} > D_l - \epsilon, \xi_n < \epsilon$. Set $\gamma = \gamma^* - \epsilon$. Then

$$\frac{-\log_2 P_e}{\mathbb{E}[T]} \geq \gamma^* D_l - \sigma(\epsilon)$$

where σ is a function of ϵ such that $\lim_{\epsilon \to 0} \sigma(\epsilon) = 0$. Finally, the proof is complete by replacing γ^* from (22). $\blacksquare$

APPENDIX B
PROOF OF LEMMA 4

Proof: Given $Y^t = y^t, W_1 = m_1, W_2 = m_2$, we obtain

$$\mathbb{E}[H_{t+1}^1 - H_t^1 | m_2, y^t]$$
$$= -I(W_1; Y_{t+1} | m_2, y^t)$$
$$= -I(W_1; Y_{t+1} | m_2, x_2^{t+1}, y^t)$$
$$= -H(Y_{t+1} | m_2, x_2^{t+1}, y^t) + H(Y_{t+1} | m_2, x_2^{t+1}, W_1, y^t)$$
$$= -H(Y_{t+1} | m_2, x_2^{t+1}, y^t)$$
$$\quad + H(Y_{t+1} | m_2, x_2^{t+1}, W_1, X_1^{t+1}, y^t)$$
$$\overset{(a)}{=} -H(Y_{t+1} | m_2, x_2^{t+1}, y^t) + H(Y_{t+1} | x_2^{t+1}, X_1^{t+1}, y^t)$$
$$\triangleq -J_{t+1}^1(m_2, x_2^{t+1}, y^t) \tag{25}$$

where (a) follows because condition on the channel inputs $X_{1,t+1}, X_{2,t+1}$, the output Y_{t+1} is independent of W_1, W_2. We denote the right-hand side of (a) by $J_{t+1}^1(.)$ as in (25). Similarly for the case when $i = 2$ the following lower-bound holds

$$\mathbb{E}[H_{t+1}^2 - H_t^2 | m_1, y^t]$$
$$= -H(Y_{t+1} | m_1, x_1^{t+1}, y^t) + H(Y_{t+1} | X_2^{t+1}, x_1^{t+1}, y^t)$$
$$\triangleq -J_{t+1}^2(m_1, x_1^{t+1}, y^t). \tag{26}$$

Using a similar argument for the case when $i = 3$, we can show that the following inequality holds

$$\mathbb{E}[H_{t+1}^3 - H_t^3 | y^t] \geq -I(X_1^{t+1}, X_2^{t+1}; Y_{t+1} | y^t)$$
$$\triangleq -J_{t+1}^3(y^t). \tag{27}$$

Consider the quantities at the right-hand side of (25), (26) and (27), i.e., the functions $J_{t+1}^1, J_{t+1}^2, J_{t+1}^3$. We proceed by the following lemma.

Lemma 10. *The vector* $(J_{t+1}^1, J_{t+1}^2, J_{t+1}^3)$ *is inside the feedback-capacity region* $\mathcal{C}$ *almost surely.*

Proof: We use the alternative representation for $\mathcal{C}$ which is given in Fact 1. For any non-negative numbers $\lambda_1, \lambda_2, \lambda_3$, let

$$J_\lambda(m_1, m_2, x_1^{t+1}, x_2^{t+1}, y^t) = \lambda_1 J_{t+1}^1(m_2, x_2^{t+1}, y^t) + \lambda_2 J_{t+1}^2(m_1, x_1^{t+1}, y^t) + \lambda_3 J_{t+1}^3(y^t)$$

Note that

$$J_\lambda(m_1, m_2, x_1^{t+1}, x_2^{t+1}, y^t) \leq \sup_{P_{W_1 W_2 X_1^{t+1} X_2^{t+1} | Y^{t+1}}} \mathbb{E}\{J_\lambda(W_1, W_2, X_1^{t+1}, X_2^{t+1}, y^t)\}, \tag{28}$$

where the supremum is taken over all $P_{X_1^{t+1} X_2^{t+1} | Y^{t+1}}$ that factors as in (5). The right-hand side of the above inequality equals $\sum_i \mathbb{E}[\lambda_i J_{t+1}^i]$. Each expectation inside the summation can be bounded as follows

$$\mathbb{E}\{J_{t+1}^1(W_2, X_2^{t+1}, y^t)\} = H(Y_{t+1} | W_2, X_2^{t+1}, y^t) - H(Y_{t+1} | X_2^{t+1}, X_1^{t+1}, y^t)$$
$$\leq H(Y_{t+1} | X_2^{t+1}, y^t) - H(Y_{t+1} | X_2^{t+1}, X_1^{t+1}, y^t)$$
$$= I(X_{1,t+1}, Y_{t+1} | X_2^{t+1}, y^t)$$

Similarly,

$$\mathbb{E}\{J_{t+1}^2(W_1, X_1^{t+1}, y^t)\} \leq I(X_{2,t+1}; Y_{t+1} | X_1^{t+1} y^t)$$
$$\mathbb{E}\{J_{t+1}^3(y^t)\} \leq I(X_{1,t+1}, X_{2,t+1}; Y_{t+1} | y^t)$$

Therefore, since the channel is memoryless using the above bounds we have

$$\mathbb{E}\{J_\lambda(W_1, W_2, X_1^{t+1}, X_2^{t+1}, y^t)\}$$
$$\leq \lambda_1 I(X_{1,t+1}, Y_{t+1}|X_2^{t+1}, y^t) + \lambda_2 I(X_{2,t+1}; Y_{t+1}|X_1^{t+1}y^t) + \lambda_3 I(X_{1,t+1}, X_{2,t+1}; Y_{t+1}|y^t)$$
$$\leq C_{\underline{\lambda}}$$

∎

Since the vector $(J_{t+1}^1, J_{t+1}^2, J_{t+1}^3)$ is inside the capacity for all $1 \leq t \leq N$, then, by definition, $\forall \epsilon > 0$ there exist L and $P_{X_1^L X_2^L Y^L}$ factoring as in (5) such that

$$J_{t+1}^i \leq I_L^i + \epsilon, \quad i = 1, 2, 3$$

holds for all $1 \leq t \leq N$. This implies the statement of the lemma.

∎

APPENDIX C
PROOF OF LEMMA 8

Proof: We prove the first statement of the lemma. The second and the third statements follow by a similar argument. Given $Y^t = y^t, W_2 = m$, define the following quantities

$$f_{i|m} = P(W_1 = i|Y^t = y^t, W_2 = m)$$
$$f_{i|m}(y_{t+1}) = P(W_1 = i|Y^t = y^t, W_2 = m, Y_{t+1} = y_{t+1})$$
$$Q_{i,m}(y_{t+1}) = P(Y_{t+1} = y_{t+1}|W_1 = i, W_2 = m, Y^t = y^t),$$

where $i \in [1 : M_1], y_{t+1} \in \mathcal{Y}$. Since $H_t^1 < \epsilon$, then there exist ϵ' (as a function of ϵ) and an index $l \in [1 : M_1]$ such that $f_{l|m} \geq 1 - \epsilon'$ and $f_{i|m} \leq \frac{\epsilon'}{M_1-1}$ for all $i \in [1 : M_1], i \neq l$. Denote

$$\hat{f}_{i|m} = \frac{f_{i|m}}{1 - f_{l|m}}, \quad i \neq l.$$

Using the grouping axiom we have

$$H_t^1 = H(W_1|W_2 = m, y^t) = h_b(f_{l|m}) + (1 - f_{l|m})H(\hat{X})$$

where $\hat{X}$ is a random variable with probability distribution $P(\hat{X} = i) = \hat{f}_{i|m}, \; i \in [1 : M_1], i \neq l$. Note that

$$h_b(f_{l|m}) \approx -(1 - f_{l|m}) \log(1 - f_{l|m}).$$

Therefore,

$$H_t^1 \approx -(1 - f_{l|m})(\log(1 - f_{l|m}) - H(\hat{X}))$$
$$\approx (1 - f_{l|m}) \log(1 - f_{l|m}) \tag{29}$$

where the last approximation is due to the fact that $-\log(1 - f_{l|m}) \gg H(\hat{X})$. Next, we derive an approximation for H_{t+1}^1. Note that

$$f_{l|m}(y_{t+1}) = \frac{f_{l|m} Q_{l,m}(y_{t+1})}{\sum_j f_{j|m} Q_{j,m}(y_{t+1})}$$

The denominator can be written as

$$f_{l|m} Q_{l,m}(y_{t+1}) + (1 - f_{l|m}) \sum_{j \neq l} \hat{f}_{j|m} Q_{j,m}(y_{t+1}).$$

The above quantity is approximately equals to $Q_{l,m}(y)$. Therefore,

$$(1 - f_{l|m}(y_{t+1})) = (1 - f_{l|m}) \frac{\sum_{j \neq l} \hat{f}_{j|m} Q_{j,m}(y_{t+1})}{\sum_j f_{j|m} Q_{j,m}(y_{t+1})}$$
$$\approx (1 - f_{l|m}) \frac{\sum_{j \neq l} \hat{f}_{j|m} Q_{j,m}(y_{t+1})}{Q_{l,m}(y_{t+1})}$$

This implies that $f_{l|m}(y_{t+1}) \approx 1$. Therefore, using the same argument for H_t^1 we have

$$
\begin{aligned}
H_{t+1}^1 &\approx -(1 - f_{l|m}(y_{t+1}))(\log(1 - f_{l|m}(y_{t+1}))) \\
&= -(1 - f_{l|m}(y_{t+1}))\left[\log(1 - f_{l|m}) + \log(\frac{\sum_{j\neq l} \hat{f}_{j|m} Q_{j,m}(y_{t+1})}{Q_{l,m}(y_{t+1})})\right] \\
&\approx -(1 - f_{l|m}(y_{t+1}))\log(1 - f_{l|m}).
\end{aligned}
\tag{30}
$$

As a result of the approximations in (29) and (30), we obtain

$$
\begin{aligned}
\frac{H_{t+1}^1}{H_t^1} &\approx \frac{(1 - f_{l|m}(y))\log(1 - f_{l|m})}{(1 - f_{l|m})\log(1 - f_{l|m})} \\
&= \frac{\sum_{j\neq l} \hat{f}_{j|m} Q_{j,m}(y)}{Q_{l,m}(y)}
\end{aligned}
$$

Note that

$$
P(Y_{t+1} = y | W_2 = m, y^t) \approx Q_{l,m}(y)
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}\{\log \frac{H_{t+1}^1}{H_t^1} | y^t\} &\approx \mathbb{E}\{\log \frac{\sum_{j\neq l} \hat{f}_{j|m} Q_{j,m}(Y_{t+1})}{Q_{l,m}(Y_{t+1})}\} \\
&= \sum_y Q_{l,m}(y) \log \frac{\sum_{j\neq l} \hat{f}_{j|m} Q_{j,m}(y)}{Q_{l,m}(y)} \\
&\stackrel{(a)}{=} -D(Q_{l,m} || \sum_{j\neq l} \hat{f}_{j|m} Q_{j,m}) \\
&\stackrel{(b)}{\geq} -\sum_{j\neq l} \hat{f}_{j|m} D(Q_{l,m} || Q_{j,m}) \\
&\geq -\max_{j\neq l} D(Q_{l,m} || Q_{j,m}) \\
&\stackrel{(c)}{\geq} -(D_1 + \epsilon)
\end{aligned}
$$

where (a) is due to the definition of Kullback–Leibler divergence, (b) is due to the convexity of Kullback–Leibler divergence, and (c) is due to the definition of D_1. ∎

APPENDIX D
PROOF OF THEOREM 2

Proof. Since $\{Z_t\}$ is a submartingale, then $Z_0 \leq \mathbb{E}[Z_T]$. By the definition of $\{Z_t\}$ we have $\mathbb{E}[Z_T] = \sum_{i=1}^3 \alpha_i \mathbb{E}[Z_T^i]$. For any of processes $\{Z_t^i\}$, the following hold:

$$
\begin{aligned}
\mathbb{E}[Z_T^i] &= \mathbb{E}\left[\frac{H_T^i - \epsilon}{I_L^i + \epsilon} 1_{\{H_T^i \geq \epsilon\}}\right] + \mathbb{E}\left[\left(\frac{\log H_T^i - \log \epsilon}{D_i + \epsilon} + f_i(\log \frac{H_T^i}{\epsilon})\right) 1_{\{H_T^i \leq \epsilon\}}\right] + \mathbb{E}[T] \\
&\leq \mathbb{E}\left[\frac{H_T^i - \epsilon}{I_L^i + \epsilon}\right] + \mathbb{E}\left[\frac{\log H_T^i - \log \epsilon}{D_i + \epsilon} + f_i(\log \frac{H_T^i}{\epsilon})\right] + \mathbb{E}[T] \\
&\stackrel{(a)}{\leq} \mathbb{E}\left[\frac{H_T^i - \epsilon}{I_L^i + \epsilon}\right] + \mathbb{E}\left[\frac{\log H_T^i - \log \epsilon}{D_i + \epsilon}\right] + \frac{1}{\mu_i D_i} + \mathbb{E}[T] \\
&= \frac{\mathbb{E}[H_T^i] - \epsilon}{I_L^i + \epsilon} + \frac{\mathbb{E}[\log H_T^i] - \log \epsilon}{D_i + \epsilon} + \frac{1}{\mu_i D_i} + \mathbb{E}[T] \\
&\stackrel{(b)}{\leq} \frac{\mathbb{E}[H_T^i] - \epsilon}{I_L^i + \epsilon} + \frac{\log \mathbb{E}[H_T^i] - \log \epsilon}{D_i + \epsilon} + \frac{1}{\mu_i D_i} + \mathbb{E}[T]
\end{aligned}
\tag{31}
$$

where (a) follows from the inequality $f_i(y) \leq \frac{1}{\mu_i D_i}$, and (b) follows by applying Jensen's inequality for the function $\log(x)$.

Define $\eta(P_e) = h_b(P_e) + P_e \log(M_1 M_2)$. Using Lemma 7, the right-hand side of (31) is upper bounded as

$$\leq \frac{\eta(P_e) - \epsilon}{(I_L^i + \epsilon)} + \frac{\log(\eta(P_e)) - \log \epsilon}{D_i + \epsilon} + \frac{1}{\mu_i D_i} + \mathbb{E}[T]$$

$$= \frac{\eta(P_e) - \epsilon}{(I_L^i + \epsilon)} + \frac{\log P_e + \log \frac{\eta(P_e)}{P_e} - \log \epsilon}{D_i + \epsilon} + \frac{1}{\mu_i D_i} + \mathbb{E}[T]$$

$$\leq \frac{\log P_e}{D_i + \epsilon} + \mathbb{E}[T](1 + \delta_i(P_e, M_1 M_2, \epsilon)), \tag{32}$$

where the function δ_i is defined as

$$\delta_i(P_e, M_1 M_2, \epsilon) = \left\| \frac{\eta(P_e) - \epsilon}{(I_L^i + \epsilon)\frac{\log M_1 M_2}{R_N^{(3)}}} + \frac{\log \frac{\eta(P_e)}{P_e} - \log \epsilon}{(D_i + \epsilon)\frac{\log M_1 M_2}{R_N^{(3)}}} + \frac{1}{\mu_i D_i \frac{\log M_1 M_2}{R_N^{(3)}}} \right\|$$

Note that we use the equation $\mathbb{E}[T] = \frac{\log M_1 M_2}{R_N^{(3)}}$ in the definition of δ_i. Observe that

$$\lim_{P_e \to 0} \lim_{M_1 M_2 \to \infty} \delta_i(P_e, M_1 M_2, \epsilon) = 0.$$

Note that $Z_0^i \leq \mathbb{E}[Z_T^i], i = 1, 2, 3$, where $Z_0^i = \frac{\log M_i - \epsilon}{I_L^i + \epsilon}$. Therefore,

$$\frac{\log M_i - \epsilon}{I_L^i + \epsilon} \leq \frac{\log P_e}{D_i + \epsilon} + \mathbb{E}[T](1 + \delta_i(P_e, M_1 M_2, \epsilon))$$

Multiplying both sides by $\frac{D_i + \epsilon}{\mathbb{E}[T]}$ and rearranging the terms give

$$-\frac{\log P_e}{\mathbb{E}[T]} \leq (D_i + \epsilon)\left(1 - \frac{R_N^{(i)}}{I_L^i + \epsilon}\right)$$

$$+ \frac{\epsilon(D_i + \epsilon)}{(I_L^i + \epsilon)\mathbb{E}[T]} + (D_i + \epsilon)\delta_i(P_e, M_1 M_2, \epsilon),$$

Define

$$\tilde{\delta}(P_e, M_1 M_2, \epsilon) = \max_i \frac{(D_i + \epsilon) \epsilon R_N^{(3)}}{(I_L^i + \epsilon') \log M_1 M_2} + (D_i + \epsilon)\delta_i(P_e, M_1 M_2, \epsilon).$$

For any non-negative numbers $\lambda_i, i = 1, 2, 3$ the following inequality holds:

$$-\frac{\log P_e}{\mathbb{E}[T]} \leq (D_i + \epsilon)\left(1 - \frac{R_N^{(i)}}{I_L^i + \epsilon}\right) + \tilde{\delta}, \tag{33}$$

$$\leq (D_i + \epsilon)\left(1 - \frac{\lambda_i R_N^{(i)}}{\lambda_i I_L^i + \epsilon}\right) + \tilde{\delta},$$

$$\leq (D_i + \epsilon)\left(1 - \frac{\lambda_i R_N^{(i)}}{\sum_j \lambda_j I_L^j + \epsilon'}\right) + \tilde{\delta},$$

$$\leq (D_i + \epsilon)\left(1 - \frac{\lambda_i R_N^{(i)}}{\sup \sum_j \lambda_j I_L^j + \epsilon}\right) + \tilde{\delta},$$

$$= (D_i + \epsilon)\left(1 - \frac{\lambda_i R_N^{(i)}}{C_\lambda + \epsilon}\right) + \tilde{\delta},$$

Since the transmission rates are inside the capacity region, $\lambda_i R_N^{(i)} \leq C_\lambda$ and we obtain

$$
\frac{\log P_e}{\mathbb{E}[T]} \leq D_i \left(1 - \frac{\lambda_i R_N^{(i)}}{C_\lambda + \epsilon}\right) + \epsilon + \tilde{\delta}(P_e, M_1 M_2, \epsilon),
$$

$$
\overset{(a)}{=} D_i \left(1 - \frac{\lambda_i R_N^{(i)}}{C_\lambda}\right) + D_i \frac{\lambda_i R_N^{(i)} \epsilon}{C_\lambda(C_\lambda + \epsilon)} + \epsilon + \tilde{\delta}(P_e, M_1 M_2, \epsilon),
$$

$$
\overset{(b)}{\leq} D_i \left(1 - \frac{\lambda_i R_N^{(i)}}{C_\lambda}\right) + D_{\max} \frac{\epsilon}{C_\lambda} + \epsilon + \tilde{\delta}(P_e, M_1 M_2, \epsilon),
$$

where $D_{\max} = \max\{D_1, D_2, D_3\}$, (a) follows by adding and subtracting the term $D_i(\frac{\lambda_i R_N^{(i)}}{C_\lambda})$, and (b) follows as $\frac{\lambda_i R_N^{(i)}}{C_\lambda + \epsilon} \leq 1$. Define $\delta(P_e, M_1 M_2, \epsilon) = \epsilon(1 + \frac{D_{\max}}{C_\lambda}) + \tilde{\delta}(P_e, M_1 M_2, \epsilon)$. The theorem follows by taking the minimum over $\lambda_i, i = 1, 2, 3$ and the fact that the following condition is satisfied:

$$
\lim_{\epsilon \to 0} \lim_{P_e \to 0} \lim_{M_1 M_2 \to \infty} \delta(P_e, M_1 M_2, \epsilon) = 0.
$$

Note that in the above proof it is assumed that the capacity region is nonempty. This assumption implies that $C_\lambda > 0$ for all $\underline{\lambda} \neq \underline{0}$ with non-negative components. $\qquad \square$

APPENDIX E
PROOF OF COROLLARY 1

From (33) in the proof of Theorem 2, we obtain:

$$
-\frac{\log P_e}{\mathbb{E}[T]} \leq \min_{i \in \{1,2,3\}} (D_i + \epsilon) \left(1 - \frac{R_N^{(i)}}{I_L^i + \epsilon}\right) + \tilde{\delta}
$$

$$
\leq D_{\max} \min_{i \in \{1,2,3\}} \left(1 - \frac{R_N^{(i)}}{I_L^i}\right) + \delta
$$

$$
= D_{\max} \min_{\substack{\alpha_1, \alpha_2, \alpha_3 \geq 0 \\ \alpha_1 + \alpha_2 + \alpha_3 = 1}} \left(1 - \sum_{i=1}^{3} \alpha_i \frac{R_N^{(i)}}{I_L^i}\right) + \delta, \tag{34}
$$

where $D_{\max} = \max\{D_1, D_2, D_3\}$, and $\delta = \tilde{\delta} + \epsilon \sup(1 + \frac{D_{\max}}{I_L^i})$. For non-negative $\lambda_i, i = 1, 2, 3$, set $\alpha_i = \frac{\lambda_i I_L^i}{\sum_j \lambda_j I_L^j}$. Next, replace $\alpha_i, i = 1, 2, 3$ in (34) with the above term. Therefore, (34) does not exceed the following

$$
D_{\max} \min_{\substack{\lambda_1, \lambda_2, \lambda_3 \geq 0 \\ \lambda_1 + \lambda_2 + \lambda_3 = 1}} \left(1 - \frac{\sum_i \lambda_i R_N^{(i)}}{\sum_j \lambda_j I_L^j}\right) + \delta.
$$

The proof is completed by noting that $\sum_j \lambda_j I_L^j \leq C_\lambda$.

REFERENCES

[1] C. E. Shannon. The zero error capacity of a noisy channel. IRE Trans. on Inform. Th., 2(3):8–19, 1956

[2] M. V. Burnashev, "Data transmission over a discrete channel with feedback. Random transmission time," Problemy Peredachi Informatsii, vol. 12, no. 4, pp. 10–30, Oct.-Dec. 1976.

[3] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," IEEE Trans. Info. Theory, vol. 12, no. 2, pp. 172–182, Apr 1966.

[4] R. L. Dobrushin, "An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback," Problemy Peredachi Informatsii, vol. 8, pp. 161–160, 1962.

[5] E. A. Haroutunian, "Lower bound for error probability in channels with feedback," Problemy Peredachi Informatsii, vol. 13, pp. 36–44, 1977.

[6] P. Berlin, B. Nakiboglu, B. Rimoldi, and E. Telatar, "A simple converse of burnashev's reliability function," IEEE Trans. Info. Theory, vol. 55, no. 7, pp. 3074–3080, July 2009.

[7] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified schalkwijk-barron scheme for channels with noiseless feedback (corresp.)," IEEE Trans. Info. Theory, vol. 25, no. 6, pp. 729–733, Nov 1979.

[8] N. T. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," IEEE Trans. Inf. Theory, vol. 21, no. 1, pp. 100–102, 1975.

[9] F. M. J. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels," IEEE Trans. Inf. Theory, vol. 28, no. 1, pp. 93–95, 1982.

[10] G. Kramer, Directed Information for channels with Feedback. PhD thesis, Swiss Federal Institute of Technology, Zurich, 1998.

[11] L. V. Truong and V. Y. F. Tan, "Error exponent of the common-message broadcast channel with variable-length feedback," IEEE Int. Symp. on Info. Theory (ISIT), Aachen, 2017, pp. 186-190.

[12] M. Heidari, F. Shirani and S. S. Pradhan, "On the necessity of structured codes for communications over MAC with feedback," IEEE Int. Symp. on Info. Theory (ISIT), Aachen, 2017, pp. 2298-2302.

[13] A. Nazari, S. S. Pradhan and A. Anastasopoulos, "Error Exponent for Multiple Access Channels: Upper Bounds," in IEEE Trans. Info. Theory, vol. 61, no. 7, pp. 3605-3621, July 2015.

[14] R. Gallager, "A perspective on multi-access channels," IEEE Trans. Information Theory, vol. 31, pp. 124–142, Mar. 1985.