Asymptotic Expansion and Error Exponent of Two-Phase Variable-Length Coding with Feedback for Discrete Memoryless Channels

Tsung-Yi Chen, Adam R. Williamson and Richard D. Wesel
tsungyi.chen@northwestern.edu, adamroyce@ucla.edu, wesel@ee.ucla.edu

***Abstract*—This paper studies variable-length coding with noiseless feedback for discrete memoryless channels. Yamamoto and Itoh's two-phase scheme achieves the optimal error-exponent, but due to the block-coding nature it is not optimal in the expansion of the message size $\log M$. Polyanskiy et al. showed that with feedback, the back-off from capacity is logarithmic in the expected latency ℓ. The $O(\log \ell)$ back-off is achieved by using an incremental redundancy (IR) scheme that only utilizes feedback to determine the stopping time. However, the achievable error-exponent of the IR scheme is not optimal. This paper shows that a two-phase coding scheme where each phase uses an IR scheme achieves the optimal error-exponent while maintaining an expansion on the message size that yields the $O(\log \ell)$ back-off.**

I. Introduction

Shannon showed in 1956 that even in the presence of full feedback (i.e., instantaneous feedback of the received symbol) the capacity for a single-user memoryless channel remains the same as without feedback [1]. The next phase in the information-theoretic analysis of feedback showed that feedback can greatly improve the error-exponent. Numerous papers including [2]–[6] explicitly showed this improvement. Burnashev's seminal work [7] showed an elegant expression of the optimal error-exponent for a discrete memoryless channel (DMC) with noiseless feedback. Burnashev employed a technique that can be considered as a form of active hypothesis testing [8], which uses feedback to adapt future transmitted symbols based on the current state of the receiver.

An important advantage of using feedback is the ability to decide when to stop transmitting additional symbols about the intended message. A mutual agreement, or a confirmation, must take place to enable reliable communication. This can happen in two ways [9]: Receiver confirmation (RC) occurs when the receiver decides whether it has decoded with sufficient reliability (e.g. passing a checksum) to terminate communication and feeds this decision back to the transmitter. The alternative to RC is transmitter confirmation (TC), in which the transmitter decides (based on feedback from the receiver) whether the receiver has decoded with sufficient reliability (or even if it has decoded correctly, since the transmitter knows the true message).

This material is based upon work supported by the National Science Foundation (NSF) under Grant Numbers 1162501. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

TC schemes often use distinct transmissions for a message phase and a confirmation phase. Practical TC and RC systems can usually be assigned to one of two categories based on when confirmation is possible. Single-codeword repetition (SCR) only allows confirmation at the end of a complete codeword and repeats the same codeword until confirmation. In contrast, incremental redundancy (IR) systems [10] transmit a sequence of distinct coded symbols with numerous opportunities for confirmation within the sequence before the codeword is repeated. In some cases of IR, the sequence of distinct coded symbols is infinite and is therefore never repeated. If the sequence of symbols is finite and thus repeated, we call this a repeated IR system.

Forney's analysis [11] provided an early connection between these practical system designs and theoretical analysis by deriving error-exponent bounds for a DMC using an SCR-RC scheme. Following Forney's work, Yamamoto and Itoh [12] replaced Forney's SCR-RC scheme with a SCR-TC scheme in which the receiver feeds back its decoding result (based only on the codeword sent during the current message phase). The transmitter confirms or rejects the decoded message during a confirmation phase, continuing with additional message and confirmation phases if needed. This relatively simple SCR-TC scheme allows block codes to achieve the optimal error-exponent of Burnashev for DMCs.

The main contribution of this paper is to tie the results of error-exponent and non-asymptotic analysis together for communications with noiseless feedback. Specifically, this paper studies a two-phase coding scheme where both phases use variable-length codes instead of fixed-length codes. For this two-phase scheme, we obtain the same expansion of the message size as in [13] up to logarithmic terms, while achieving the optimal error-exponent.

The rest of the paper is organized as follows: Sec. II introduces the notation and definition used throughout the paper, and Sec. III briefly reviews relevant previous works. Sec. IV states the main results and provides the proofs of the results. Finally Sec. V concludes the paper.

II. Notations and Definitions

Denote the input alphabet as $\mathcal{X}$ and the output alphabet as $\mathcal{Y}$. Capital letters represent random variables (r.v.'s) and the associated instances are denoted by small letters. For example, X^n denotes an n-dimensional r.v. with instances x^n taking values in $\mathcal{X}^n$. Denote the jth to kth elements of x^n as x_j^k.

A discrete memoryless channel (DMC) $P_{Y|X} : \mathcal{X} \mapsto \mathcal{Y}$ without feedback has finite input and output alphabets and follows

$$P_{Y^n|X^n}(y^n|x^n) = \prod_{j=1}^{n} P_{Y|X}(y_j|x_j) \,.$$

For DMCs with feedback we consider an input process $\{X_j\}_{j=1}^{\infty}$ and a sequence of channels $\{P_{Y_j|X_j}\}_{j=1}^{\infty}$. Let the finite dimensional distribution of $(X^n, \bar{X}^n, Y^n)$ be:

$$\begin{aligned} &P_{X^n,Y^n,\bar{X}^n}(x^n, y^n, \bar{x}^n) \\ &= P_{X^n}(x^n)P_{X^n}(\bar{x}^n)\prod_{j=1}^{n} P_{Y_j|X^jY^{j-1}}(y_j|x^j, y^{j-1}), \end{aligned} \tag{1}$$

i.e., the distribution of $\bar{X}^n$ is identical to X^n but independent of Y^n. The information density $i(x^n; y^n)$ is defined as

$$i(x^n; y^n) = \log \frac{dP_{Y^n|X^n}(y^n|x^n)}{dP_{Y^n}(y^n)}. \tag{2}$$

In this paper we only consider channels with essentially bounded information density $i(X; Y)$.

Let $D(P||Q)$ be the divergence between two distributions P and Q. For a given DMC $P_{Y|X}$, denote C as the capacity of the channel and let C_1 be defined as:

$$C_1 = \max_{x,x' \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{Y|X=x}(y) \log \frac{P_{Y|X=x}(y)}{P_{Y|X=x'}(y)} \tag{3}$$

$$= \max_{x,x' \in \mathcal{X}} D(P_{Y|X=x}||P_{Y|X=x'}) \,. \tag{4}$$

In order to be self-contained, we state the definition of variable-length feedback (VLF) codes in [13]:

Definition 1: An (ℓ, M, ϵ) variable-length feedback code (VLF code) is defined as:

1) A common r.v. $U \in \mathcal{U}$ with a probability distribution P_U revealed to both transmitter and receiver before the start of transmission.
2) A sequence of encoders $f_n : \mathcal{U} \times \mathcal{W} \times \mathcal{Y}^{n-1} \to \mathcal{X}$ that defines the channel inputs $X_n = f_n(U, W, Y^{n-1})$. Here W is the message r.v. uniform in $\mathcal{W} = \{1, \ldots, M\}$.
3) A sequence of decoders $g_n : \mathcal{U} \times \mathcal{Y}^n \to \mathcal{W}$ providing the estimate of W at time n.
4) A stopping time $\tau \in \mathbb{N}$ w.r.t. the filtration $\mathcal{F}_n = \sigma\{U, Y^n\}$ such that $\mathbb{E}[\tau] \leq \ell$.
5) The decision $\hat{W} = g_\tau(U, Y^\tau)$ satisfies $P[\hat{W} \neq W] \leq \epsilon$.

The fundamental limit of variable-length feedback codes is given by the following quantity:

$$M_f^*(\ell, \epsilon) = \max\{M : \exists \text{ an } (\ell, M, \epsilon) \text{ VLF code }\}. \tag{5}$$

A stop-feedback code is a VLF code that satisfies the following property:

$$f_n(U, W, Y^{n-1}) = f_n(U, W) \,. \tag{6}$$

In other words, the encoder only uses feedback to determine whether it is time to stop transmitting for the current message. VLF codes generally include TC and RC feedback systems, and stop-feedback codes are common RC systems in practice; the receiver computes the reliability of the decoded symbols as in [14] or uses a cyclic redundancy check (CRC) to inform the transmitter to stop or continue, which requires only 1 bit of information.

Moving toward a TC system, we introduce the following definition of sequential probability ratio test (SPRT) for simple hypothesis testing based on [15]:

Definition 2: Let P_θ be the distribution associated with the hypothesis H_θ. Let Λ_n be the log-likelihood ratio of the observed samples $Y_j, j = 1, 2, \ldots$:

$$\Lambda_n = \log \frac{\prod_{j=1}^{n} P_1(Y_j)}{\prod_{j=1}^{n} P_0(Y_j)} \tag{7}$$

$$= \sum_{j=1}^{n} \Lambda(Y_j) \,. \tag{8}$$

Let $t_0 < 0 < t_1$ be the decision thresholds to be optimized. After the nth observation, the sequential test uses the following decision rule:

$$\text{Stop sampling and accept } \begin{cases} H_1 \text{ if } \Lambda_n \geq t_1 \\ H_0 \text{ if } \Lambda_n \leq t_0 \end{cases} ,$$

Otherwise take another sample.

The stopping time for the SPRT is therefore given as:

$$\tau = \inf\{n : \Lambda_n \geq t_1 \text{ or } \Lambda_n \leq t_0\}. \tag{9}$$

Let E_0 and E_1 be the events where the test selects hypotheses H_0 and H_1, respectively. Let α be the error of the first kind, $P_0[E_1]$, and β be the error of the second kind, $P_1[E_0]$.

Many interesting results follow from SPRT [15] [16]. Some relevant results are the following:

$$\frac{\alpha}{1-\beta} \leq \exp\{-t_1\} \,, \tag{10}$$

$$\frac{\beta}{1-\alpha} \leq \exp\{t_0\} \,. \tag{11}$$

Further simplifying the inequalities we have

$$\alpha \leq \exp\{-t_1\} \,, \tag{12}$$

$$\beta \leq \exp\{t_0\} \,, \tag{13}$$

which are useful when both α and β are very small. We first define a class of VLF codes called two-phase feedback codes:

Definition 3: Let $f_n^{(1)}$ and $f_n^{(2)}$ be two sequences of stop-feedback encoders. An (ℓ, M, ϵ) two-phase feedback code is an (ℓ, M, ϵ) VLF code with encoders satisfying the following properties:

$$\{f_n\} = \{f_1^{(1)}, f_2^{(1)}, \ldots, f_{\tau_1}^{(1)}, f_1^{(2)}, f_2^{(2)}, \ldots\} \,, \tag{14}$$

where τ_1 is a stopping time w.r.t. the filtration $\mathcal{F}_n = \sigma(U, Y^n)$.

Note that due to the transition from $f^{(1)}$ to $f^{(2)}$, a two-phase feedback code is no longer a stop-feedback code, although it consists of two stop-feedback codes. Def. 3 is one step toward "more active" feedback codes compared to stop-feedback codes, and the definition includes stop-feedback codes as a special case by choosing $\tau_1 = \infty$. The level of "activeness" can generally increase as we increase the number of phases m with several stopping times between the phases $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_m$. We focus on two-phase results in this paper (m=2). For notational convenience we denote the fundamental limit of a two-phase feedback code as $M^*_{\text{2-ph}}(\ell,\epsilon)$, the maximum M such that there exists an (ℓ, M, ϵ) two-phase feedback code. Note that by definition $M^*_{\text{2-ph}}(\ell,\epsilon) \le M^*_f(\ell,\epsilon)$.

III. Previous Work

In the error-exponent literature, Burnashev showed that the optimal error-exponent is $E(R) = C_1(1 - R/C)$, where C_1 is given in (3). As observed by Naghshvar and Javidi [8], reformulating the problem of channel coding with noiseless feedback as an active hypothesis testing problem yields an explicit policy that has error probability achieving Burnashev's error-exponent. In this light, Burnashev's original achievability approach can be seen as an active hypothesis testing policy.

Yamamoto and Itoh [12] showed that a two-phase block-coding scheme achieves the same error-exponent. The result is surprising because despite being a simple block-coding scheme with ARQ, a relatively "inactive" scheme, the error-exponent is the same as complicated coding schemes that are extremely active. This is because the cost of using block codes vanishes as the blocklength goes to infinity.

The drawback of a block-coding scheme is evident when the average rate is analyzed in the non-asymptotic regime. Polyanskiy et al. [13] showed that stop-feedback codes achieve the following expansion :

$$\log M^*_f(\ell,\epsilon) = \frac{\ell C}{1-\epsilon} - O(\log \ell)\,. \tag{15}$$

On the other hand, the achievability result based on active hypothesis testing in [8] can be stated as

$$\log M^*_f(\ell,\epsilon) \ge \ell C - \frac{C\log(1/\epsilon)}{C_1} - O(1)\,, \tag{16}$$

which shows that the constant terms can be significant when ϵ is very small while ℓ is only moderately large.

Although not shown in [13], it is evident that feedback cannot improve the $\sqrt{\ell}$ penalty term for a block-coding scheme due to the channel dispersion V [17]. To see this, let the fundamental limit of a code without feedback $M^*(n,\epsilon)$ be as defined in [17]. Refining the analysis in [18]–[20], [17] showed that

$$\log M^*(n,\epsilon) = nC - \sqrt{nV}Q^{-1}(\epsilon) + O(\log n)\,. \tag{17}$$

Let $\eta \in (0,1)$ be a parameter to be chosen later. The first phase of the Yamamoto-Itoh scheme uses a length $\eta n = n_1$ block code with M codewords while the second phase uses a length $(1-\eta)n = n_2$ block code with 2 codewords, representing acknowledge (ACK) and not acknowledge (NACK), respectively. Since the error probabilities in the second phase (both $P[\text{ACK} \to \text{NACK}]$ and $P[\text{NACK} \to \text{ACK}]$) can be made exponentially small, the expected latency ℓ is given as

$$\ell = n(1-\epsilon)^{-1} + o(1),$$

where the overall error probability ϵ' is given as

$$\epsilon' = \epsilon \exp\{-n_2 C_1 + o(n_2)\}. \tag{18}$$

The two-phase block-coding scheme therefore gives the following expansion of M:

$$\log M(\ell,\epsilon') = n_1 C - \sqrt{n_1 V}Q^{-1}(\epsilon) + O(\log n_1) \tag{19}$$
$$= \ell_1(1-\epsilon)C - \sqrt{\ell_1(1-\epsilon)V}Q^{-1}(\epsilon) + O(\log \ell_1), \tag{20}$$

where $\ell_1 = \eta\ell$. The $\sqrt{\ell}$ penalty term will not disappear except for the trivial case where $\epsilon = 1$.

While stop-feedback codes achieve the expansion in (15), the error-exponent is given as $C - R$ instead of $C_1\left(1 - \frac{R}{C}\right)$. Observing the results of [12], the key to achieving the optimal error-exponent relies on the second phase. Motivated by the observation, this paper studies the error-exponent and the expansion for two-phase feedback codes.

IV. Main Results

This section summarizes the main results of this paper. All of the results focus on DMCs with bounded information density.

Theorem 1: For a DMC with capacity C and a finite C_1 we have

$$\log M^*_{\text{2-ph}}(\ell,\epsilon_\ell) \ge \ell C - (1+\delta)\log \ell - O(1), \tag{21}$$

for some $\delta > 0$. In addition, ϵ_ℓ satisfies

$$-\log \epsilon_\ell = C_1\left(\ell - \frac{\log M^*_{\text{2-ph}}}{C}\right) + O(1)\,. \tag{22}$$

Furthermore, for some $\delta > 0$ and a fixed $\epsilon \in (0,1)$ we have

$$\log M^*_{\text{2-ph}}(\ell,\epsilon) \ge \frac{\ell C}{1-\epsilon} - (1+\delta)\log \ell - O(1)\,. \tag{23}$$

With the converse in [7, Lem. 1 and 2] and the above theorem we can conclude that $M^*_{\text{2-ph}}$ has the same expansion as in (15) while also achieving the optimal error-exponent.

The fact that the optimal error-exponent is achievable with a two-phase feedback code is expected due to [12]. The interesting part is that the optimal expansion up to $O(\log \ell)$ terms in [13] is also achievable. Comparing the achievability results shown in [13], we see that the pre-log factor is almost the same, but with an increase from 1 to $1+\delta$ for an arbitrary $\delta \in (0,1)$. The implicit penalty in having a small δ is that the constant term in (23) may be comparable to the leading terms when ℓ is small and must be reviewed carefully when evaluating the non-asymptotic performance. For very short ℓ and small ϵ the achievability bound must be computed numerically to obtain accurate estimates of the expected throughput.

The converse of Theorem 1 follows from Burnashev [7]. The proof of the achievability using a two-phase feedback code is separated into two parts in the following subsections.

A. Error Exponent of Two-Phase Feedback Codes

This subsection studies the error-exponent of two-phase feedback codes and some of the non-asymptotic behavior. The following lemma states that in the second phase where the encoder attempts to confirm whether the message in the first phase is decoded correctly or not, it is enough to have the average blocklength ℓ_2 scale logarithmically with ℓ_1 rather than linearly as for the achievability proof in [12]:

Lemma 1: For a DMC with capacity C and $C_1 < \infty$, there exists an (ℓ, M, ϵ_ℓ) two-phase feedback code such that

$$-\log \epsilon_\ell = C_1 \left(\ell - \frac{\log M}{C}\right) + O(1) . \quad (24)$$

In other words, the optimal error-exponent is achievable with two stop-feedback codes: an $(\ell_1, M, \epsilon_{\ell_1})$ stop-feedback code and an $(\ell_2, 2, \epsilon_{\ell_2})$ stop-feedback code where $\ell_2 = O(\log \ell_1)$.

The main idea of the achievability is to use stop-feedback codes in both phases instead of fixed-length block codes. After decoding the message W in the first phase, the message is fed back to the transmitter and both the transmitter and the receiver enter the second phase. The transmitter decides whether the message is correct and transmits an ACK or NACK in the second phase accordingly. The transmission is terminated if an ACK is decoded in the second phase, otherwise the same process repeats from the first phase.

Proof: The construction of a stop-feedback code for the first phase follows closely to the proof of [13, Thm. 2] and is largely omitted here. The relevant results of an (ℓ_1, M, ϵ_1) stop-feedback code used in the first phase are stated as follows. The error probability ϵ_1 is upper bounded as

$$\epsilon_1 \leq \exp\{\log M - \ell_1 C + a_0\} , \quad (25)$$

and by definition the stopping time of the first phase τ_1 satisfies

$$\mathbb{E}[\tau_1] \leq \ell_1 . \quad (26)$$

Now we turn our attention to the second phase and construct an $(\ell_2, 2, \epsilon_2)$ stop-feedback code that conveys the confirmation message ACK or NACK. Construct the codebook for the second phase as a repetition code. In other words, the codebook consists of two repeated sequence of the symbols x_c and x_e, which stand for "confirmation" and "error". The decoding of the second phase can be accomplished by a sequential testing of simple hypothesis H_0 against an alternative H_1 using the SPRT. We sequentially observe $Y_j, j = 1, 2, \ldots$, which are i.i.d. according to $P_0 = P_{Y|X=x_c}$ under H_0 and according to $P_1 = P_{Y|X=x_e}$ under H_1.

Denote the mean of $\Lambda(Y)$ under $P_{Y|X=x_c}$ as $-D_{c,e}$:

$$\begin{aligned}\mathbb{E}_{P_{Y|X=x_c}}[\Lambda(Y)] &= -D\left(P_{Y|X=x_c} || P_{Y|X=x_e}\right) && (27)\\ &= -D_{c,e} , && (28)\end{aligned}$$

and denote the mean of $\Lambda(Y)$ under $P_{Y|X=x_e}$ as $D_{e,c}$:

$$\begin{aligned}\mathbb{E}_{P_{Y|X=x_e}}[\Lambda(Y)] &= D\left(P_{Y|X=x_e} || P_{Y|X=x_c}\right) && (29)\\ &= D_{e,c} . && (30)\end{aligned}$$

Choose x_e and x_c such that $D_{c,e} = C_1 \geq C$ as defined in (3), and therefore $D_{c,e} \geq D_{e,c}$. Let λ be the uniform bound of $\Lambda(Y)$ under both H_0 and H_1. Letting τ_2 be the stopping time of the second phase, we have by Wald's identity that

$$\begin{aligned}C_1 \mathbb{E}[\tau_2|H_0] &\leq P_0[E_0](-t_0 + \lambda) - P_0[E_0^c] t_1 , && (31)\\ D_{e,c} \mathbb{E}[\tau_2|H_1] &\leq P_1[E_1](t_1 + \lambda) + P_1[E_1^c] t_0 . && (32)\end{aligned}$$

Let $\xi_0 = 1 + \frac{D_{e,c}}{C_1}$ and $\xi_1 = 1 + \frac{C_1}{D_{e,c}}$. Choosing $t_0 = -\ell_2 C_1 + \lambda + \lambda'$ and $t_1 = \ell_2 D_{e,c} - \lambda$, we have under H_0

$$\begin{aligned}\mathbb{E}[\tau_2|H_0] &\leq (1-\alpha)\left(\ell_2 - \frac{\lambda'}{C_1}\right) - \alpha\left(\frac{\ell_2 D_{e,c}}{C_1} - \lambda\right) && (33)\\ &= \ell_2 - \alpha \xi_0 \ell_2 - \frac{(1-\alpha)\lambda'}{C_1} + \alpha\lambda , && (34)\end{aligned}$$

And under H_1 we have

$$\begin{aligned}\mathbb{E}[\tau_2|H_1] &\leq (1-\beta)\ell_2 + \beta\left(-\frac{\ell_2 C_1}{D_{e,c}} + \frac{\lambda' + \lambda}{D_{e,c}}\right) && (35)\\ &= \ell_2 - \beta \ell_2 \xi_2 + \frac{\beta\lambda'}{D_{e,c}} + \frac{\beta\lambda}{D_{e,c}} . && (36)\end{aligned}$$

The overall expectation is given as

$$\mathbb{E}[\tau_2] = P[H_0]\mathbb{E}[\tau_2|H_0] + P[H_1]\mathbb{E}[\tau_2|H_1] . \quad (37)$$

Recalling from (12) that both α and β decrease to zero exponentially with ℓ_2, we can choose λ' to be a constant such that $\mathbb{E}[\tau_2] \leq \ell_2$ for a large enough ℓ_2.

To analyze the overall error probability ϵ, observe that an error occurs only when both phases make an error. Let $R = \log M/\ell_1$ and let M scale with ℓ_1 as

$$\log M = \ell_1 C - \eta \log \ell_1 - a_0 , \quad (38)$$

for some $\eta > 0$ to be chosen later. Recall from (25) that the error probability in the first phase ϵ_1 is upper bounded as

$$\epsilon_1 \leq \exp\{-\eta \log \ell_1\} . \quad (39)$$

Denoting the probability that a NACK is decoded in the second phase as $P[\text{NACK}]$, we have

$$\begin{aligned}P[\text{NACK}] &= \epsilon_1(1-\beta) + (1-\epsilon_1)\alpha && (40)\\ &\leq \epsilon_1 + \alpha && (41)\\ &\leq \exp\{-a\ell_1\} && (42)\end{aligned}$$

for some $a > 0$. Recall that $\beta = P_1[E_0] \leq \exp\{t_0\}$ where $t_0 = -\ell_2 C_1 + \lambda + \lambda'$. The overall error probability ϵ is:

$$\begin{aligned}\epsilon &= \epsilon_1 P_1[E_0](1 - P[\text{NACK}])^{-1} && (43)\\ &= \exp\{-\eta \log \ell_1 + t_0 + O(1)\} && (44)\\ &= \exp\{-\eta \log \ell_1 - \ell_2 C_1 + O(1)\} . && (45)\end{aligned}$$

Letting $\ell_2 = k \log \ell_1$, the overall error probability is

$$\begin{aligned}\epsilon &\leq \exp\{-\log \ell_1(\eta + kC_1) + \lambda' + O(1)\} && (46)\\ &= \exp\{-\ell_1(\eta + kC_1)(C - R) + \lambda' + O(1)\} . && (47)\end{aligned}$$

Choosing η and k such that $\eta + kC_1 = C_1/C$, e.g., $k = 1/C_1, \eta = C_1/C - 1$, we obtain

$$\epsilon_\ell \leq \exp\{-\ell_1 C_1 (1 - R/C) + O(1)\} . \quad (48)$$

■

B. Asymptotic Expansion of $M^*_{\text{2-ph}}$

Lemma 2: For a DMC with capacity C and a finite C_1 we have for some $\delta > 0$ that

$$\log M^*_{\text{2-ph}}(\ell, \epsilon) \geq \frac{\ell C}{1-\epsilon} - (1+\delta)\log \ell - O(1) . \quad (49)$$

Proof: It is not necessary to insist on the strict latency constraint of the two stop-feedback codes to show the expansion of the message size M. Hence we choose the thresholds of the second phase to be $t_1 = \ell_2 D_{c,e} - \lambda$ and $t_0 = -\ell_2 C_1 + \lambda$, giving the expected latency in the second phase as follows:

$$\mathbb{E}[\tau_2] \leq \ell_2 + o(1)\,. \tag{50}$$

To compute the overall expected latency, recall that we have (42). Hence the overall expected latency is given as

$$\begin{aligned}\mathbb{E}[\tau] &= \mathbb{E}[\tau_1 + \tau_2](1 - P[\text{NACK}])^{-1} && (51)\\ &\leq (\ell_1 + \ell_2 + o(1))\,(1 + b\exp\{-a\ell_1\}) && (52)\\ &= \frac{\log M}{C} + \frac{\eta \log \ell_1}{C} + k\log \ell_1 + O(1)\,. && (53)\end{aligned}$$

for some constant $b > 0$. Therefore we have:

$$\log M \geq \ell C - (\eta + kC)\log \ell_1 - O(1)\,. \tag{54}$$

To show (15) for a fixed error probability ϵ, we follow the same argument as in [13]. Consider a system that operates with an $(\ell', M, \epsilon_{\ell'})$ code with probability p and terminates immediately with probability $1 - p$. The error probability is given as $1 - p + p\epsilon_\ell$ and the expected latency is given as $\ell = p\ell'$. Setting $p = \frac{1-\epsilon}{1-\epsilon_{\ell'}}$ and observing that $1 - \epsilon_{\ell'} \geq 1 - \epsilon_\ell$ we obtain an (ℓ, M, ϵ) VLF code such that

$$\begin{aligned}\log M &\geq \ell' C - (\eta + kC)\log \ell' - O(1) && (55)\\ &\geq \frac{\ell C(1-\epsilon_\ell)}{1-\epsilon} - (\eta + kC)\log \ell - O(1) && (56)\\ &= \frac{\ell C}{1-\epsilon} - (\eta + kC)\log \ell - O(1)\,, && (57)\end{aligned}$$

where the last equality holds because ϵ_ℓ decreases exponentially in ℓ as shown in (48). Observe that the pre-log factor is increased from 1 to $\eta + kC$ by introducing the second phase to achieve the optimal error-exponent.

To solve for η and k satisfying $\eta + kC_1 = C_1/C$, note that η and k must be positive. Hence $k \in (0, 1/C)$ and $\eta \in (0, C_1/C)$. Letting $k = (1-\nu)/C$ and $\eta = \nu C_1/C$ satisfies the equation for all $\nu \in (0, 1)$. This gives $\eta + kC = 1 + \nu(C_1/C - 1)$. Properly choosing ν according to any given $\delta > 0$ yields the desired expansion:

$$\log M \geq \frac{\ell C}{1-\epsilon} - (1+\delta)\log \ell - O(1)\,, \tag{58}$$

finishing the achievability of Thm. 1. ∎

V. Concluding Remarks

This paper showed that the optimal error-exponent and the same expansion of M up to $O(\log \ell)$ terms is achievable by two-phase feedback codes. The variable-length nature of stop-feedback codes provides the gain to achieve the expansion $\ell C - O(\log \ell)$. Introducing a second phase using the SPRT, which is a stop-feedback code, achieves the optimal error-exponent and only affects the pre-log factor by a factor of $(1 + \delta)$ for any $\delta > 0$.

The ability to achieve the expansion shown in [13] can be seen as the "sequentiality gain" [21] of feedback while the ability to achieve a larger error-exponent can be seen as the "adaptivity gain". The result of this paper showed that by introducing a little amount of "activeness", i.e., the second phase stop-feedback code, the optimal error-exponent is achievable while maintaining almost the same expansion on the message size M.

As discuss in [21], the "adaptivity" gain will become crucial as the error probability ϵ tends to zero. Since we are free to choose ℓ to be large in the expansion results, the pre-log factor can be as close to the optimal one as possible. The implicit penalty caused by the $O(1)$ terms, however, must be reviewed carefully when ℓ is small. Obtaining numerical results of actual performance often requires a system simulation to evaluate the tightness of the achievability bounds.

References

[1] C. Shannon, "The zero error capacity of a noisy channel," *Information Theory, IRE Transactions on*, vol. 2, no. 3, pp. 8–19, 1956.

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

[3] A. Kramer, "Improving communication reliability by use of an intermittent feedback channel," *IEEE Trans. Inf. Theory*, vol. 15, no. 1, pp. 52–60, 1969.

[4] K. S. Zigangirov, "Upper bounds for the error probability for channels with feedback," *Probl. Pered. Inform.*, vol. 6, no.1, pp. 87–92, 1970.

[5] B. Nakiboglu and R. Gallager, "Error exponents for variable-length block codes with feedback and cost constraints," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 945 –963, Mar. 2008.

[6] R. G. Gallager, "Variations on a theme by Schalkwijk and Kailath," *IEEE Trans. Inf. Theory*, vol. 56, no.1, pp. 6–17, Jan. 2010.

[7] M. V. Burnashev, "Data transmission over a discrete channel with feedback. random transmission time," *Probl. Inf. Transm*, vol. 12, no. 4, pp. 10–30, 1976.

[8] M. Naghshvar and T. Javidi, "Active sequential hypothesis testing," *arXiv:1203.4626v3 [cs.IT]*, Oct. 2012.

[9] T.-Y. Chen, A. R. Williamson, N. Seshadri, and R. D. Wesel, "Feedback communication systems with limitations on incremental redundancy," submitted for publication. Available: http://arxiv.org/abs/1309.0707.

[10] D. Mandelbaum, "An adaptive-feedback coding scheme using incremental redundancy (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 388–389, 1974.

[11] J. Forney, G.D., "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, 1968.

[12] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 25, pp. 729–733, Nov. 1979.

[13] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, vol. 57(8), pp. 4903–4925, Aug. 2011.

[14] A. Williamson, T.-Y. Chen, and R. Wesel, "Reliability-based error detection for feedback communication with low latency," in *ISIT 2013*, 2013, pp. 2552–2556.

[15] A. Wald, *Sequential Analysis*. New York, NY, USA: John Wiley & Sons Inc, 1947.

[16] H. Chernoff, *Sequential Analysis and Optimal Design*. Bristol, England: Society for Industrial and Applied Mathematics, 1972.

[17] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307 –2359, may 2010.

[18] L. Weiss, "On the strong converse of the coding theorem for symmetric channels without memory," *Quart. Appl. Math.*, 1960.

[19] R. L. Dobrushin, "Mathematical problems in the Shannon theory of optimal coding of information," in *Proc. 4th Berkeley Symp. Mathematics, Statistics, and Probability*, vol. 1, 1961, pp. 211 – 252.

[20] V. Strassen, "Asymptotische abschätzungen in Shannon's informationstheorie," in *Trans. 3d Prague Conf. Inf. Theory*, Prague, 1962, pp. 689 – 723.

[21] M. Naghshvar and T. Javidi, "Sequentiality and adaptivity gains in active hypothesis testing," *IEEE J. Sel. Areas Signal Process.*, vol. 7, no. 5, Oct. 2013.