# Finite-Blocklength Performance of Sequential Transmission over BSC with Noiseless Feedback

Hengjie Yang and Richard D. Wesel
Department of Electrical and Computer Engineering
University of California, Los Angeles, Los Angeles, CA 90095, USA
Email: {hengjie.yang, wesel}@ucla.edu

*Abstract*—**In this paper, we consider the problem of sequential transmission over the binary symmetric channel (BSC) with full, noiseless feedback. Naghshvar *et al.* proposed a one-phase encoding scheme, for which we refer to as the small-enough difference (SED) encoder, which can achieve capacity and Burnashev's optimal error exponent for symmetric binary-input channels. They also provided a non-asymptotic upper bound on the average blocklength, which implies an achievability bound on rates. However, their achievability bound is loose compared to the simulated performance of SED encoder, and even lies beneath Polyanskiy's achievability bound of a system limited to stop feedback. This paper significantly tightens the achievability bound by using a Markovian analysis that leverages both the submartingale and Markov properties of the transmitted message. Our new non-asymptotic lower bound on achievable rate lies above Polyanskiy's bound and is close to the actual performance of the SED encoder over the BSC.**

## I. Introduction

Feedback does not increase the capacity of memoryless channels [1], but it can significantly reduce the complexity of communication and the probability of error, provided that variable-length feedback (VLF) codes are allowed. In his seminal paper, Burnashev [2] first derived the optimal error exponent of VLF codes using a conceptually important two-phase transmission scheme. The first phase is called the *communication phase*, in which the transmitter seeks to increase the decoder's belief about the transmitted message by improving its posterior to above $1/2$. The second phase is called the *confirmation phase*, in which the transmitter seeks to increase the posterior of the most likely message identified from the communication phase to above a target threshold, at which it can be reliably decoded.

For the binary symmetric channel (BSC) with noiseless feedback, Horstein [3] first proposed a simple, elegant transmission scheme that achieves the capacity of BSC; it was then generalized by Shayevitz and Feder [4] to the concept of posterior matching. Since Horstein's work, several authors, e.g., [5]–[9], have constructed schemes to achieve the capacity or the optimal error exponent of BSC with noiseless feedback.

Recently, attention has shifted from the asymptotic regime to the finite-blocklength regime. Polyanskiy *et al.* [10], [11] first showed that variable-length coding with noiseless feedback can provide a significant advantage in achievable rate over fixed-length codes without feedback. In their analysis, a simple stop feedback scheme is enough to obtain an achievable rate larger than that of a fixed-length code without feedback.

For symmetric binary-input channels with noiseless feedback, Naghshvar *et al.* [9], [12] proposed a one-phase encoding scheme, which we refer to as the *small-enough-difference* (SED) encoder, that attains capacity and Burnashev's optimal error exponent simultaneously. They also gave a non-asymptotic upper bound on the average blocklength of the SED encoder. However, in the case of BSC, their bound corresponds to a lower bound on achievable rate that lies beneath Polyanskiy's lower bound on the achievable rate of a system limited to stop feedback. A system such as the SED encoder that exploits full noiseless feedback should provide a higher rate than a system limited to stop feedback.

In this paper, we seek a tightened upper bound of average blocklength of sequential transmission over BSC with noiseless feedback. The bounds of [9], [12] were derived by synthesizing a delicate new submartingale from two submartingales that characterize the fundamental behavior of the transmitted message. In fact, this general proof technique dates back to the work of Burnashev and Zigangirov [13]. However, this sophisticated analysis succeeds in establishing a non-asymptotic upper bound, but it does not reveal the fundamental mechanism that produces the constant term in the bound.

Following the SED encoder in [12], we present a Markovian analysis that leverages the submartingale results of Naghshvar *et al.* [9], [12] and the Markov structure of the the transmitted message during its confirmation phase. This enables us to significantly tighten the upper bound on average blocklength and to gain a deep understanding of the constant term in the bound. Specifically, we will apply a time-of-first passage analysis on the Markov chain formed by the transmitted message in the confirmation phase, which fully accounts for the times when the transmitted message *falls back* from the confirmation phase to the communication phase. Our analysis reveals that the constant term mainly results from the differential time spent in the "fallback" stage.

This paper is organized as follows. Sec. II formulates the problem of sequential transmission over the discrete memoryless channel (DMC) with noiseless feedback. Sec. III presents our main results and the proof using our Markovian analysis.

This research is supported by National Science Foundation (NSF) grant CCF-1955660. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect views of the NSF.

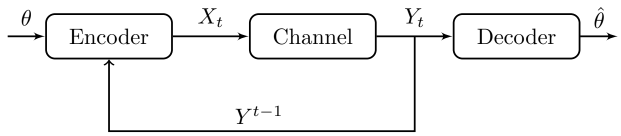


Fig. 1. System diagram of a DMC with full, noiseless feedback.

Sec. IV demonstrates the simulated performance of the SED encoder and compares our results with previous achievability bounds by Polyanskiy and bounds from the lemma of Naghshvar *et al*. Due to space constraints, we only present the proof sketch. Complete proof details can be found in [14].

## II. Problem Setup

Consider the problem of sequential transmission (or variable-length coding) over a DMC with full, noiseless feedback as depicted in Fig. 1. The DMC is described by the finite input set $\mathcal{X}$, finite output set $\mathcal{Y}$, and a collection of conditional probabilities $P(Y|X)$. The Shannon capacity of the DMC is given by

$$C = \max_{P_X} I(X;Y), \tag{1}$$

where $P_X$ denotes the probability distribution over the finite set $\mathcal{X}$. Let $C_1$ be the maximal Kullback-Leibler (KL) divergence between the conditional output distributions,

$$C_1 = \max_{x,x'\in\mathcal{X}} D\big(P(Y|X=x)\|P(Y|X=x')\big). \tag{2}$$

We also denote

$$C_2 = \max_{y\in\mathcal{Y}} \log \frac{\max_{x\in\mathcal{X}} P(Y=y|X=x)}{\min_{x\in\mathcal{X}} P(Y=y|X=x)}. \tag{3}$$

All logarithms in this paper are base 2. We assume $C, C_1, C_2$ are positive and finite. It can be easily shown that $0 < C \leq C_1 \leq C_2 < \infty$. For the BSC($p$) with crossover probability $0 < p < 1/2$, letting $q = 1-p$, we have

$$C = 1 - H(p) \tag{4}$$

$$C_1 = p\log\frac{p}{q} + q\log\frac{q}{p} \tag{5}$$

$$C_2 = \log\frac{q}{p}. \tag{6}$$

Let $\theta$ be the transmitted message uniformly drawn from the message set $\Omega = \{1, 2, \ldots, M\}$. The total transmission time (or the number of channel uses, or blocklength) $\tau$ is a random variable that is governed by a stopping rule that is a function of the observed channel outputs. Thanks to the noiseless, feedback channel, the transmitter is also informed of the channel outputs and thus the stopping time.

The transmitter wishes to communicate $\theta$ to the receiver. To this end, it produces channel inputs $X_t$ for $t = 1, 2, \ldots, \tau$ as a function of $\theta$ and past channel outputs $Y^{t-1} = (Y_1, Y_2, \ldots, Y_{t-1})$, available to the transmitter through the full, noiseless feedback channel. Namely,

$$X_t = e_t(\theta, Y^{t-1}), \quad t = 1, \ldots, \tau, \tag{7}$$

for some encoding function $e_t : \Omega \times \mathcal{Y}^{t-1} \to \mathcal{X}$.

After observing $\tau$ channel outputs $Y_1, Y_2, \ldots, Y_\tau$, the receiver makes a final estimate $\hat{\theta}$ of the transmitted message $\theta$ as a function of $Y^\tau$, i.e.,

$$\hat{\theta} = d(Y^\tau), \tag{8}$$

for some decoding function $d : \mathcal{Y}^\tau \to \Omega$.

The probability of error of the scheme is given by

$$P_e \triangleq \Pr\{\hat{\theta} \neq \theta\}. \tag{9}$$

For a fixed DMC and for a given $\epsilon > 0$, the goal is to find encoding and decoding rules described in (7), (8), and a stopping rule that defines the stopping time $\tau$ such that $P_e \leq \epsilon$ and the average blocklength $\mathbb{E}[\tau]$ is minimized.

As noted in [12], a sufficient statistic of $Y^t$ for $\theta$ is the *belief state* of the receiver,

$$\boldsymbol{\rho}(t) \triangleq [\rho_1(t), \rho_2(t), \ldots, \rho_M(t)], \quad t = 0, 1, 2, \ldots, \tau, \tag{10}$$

where for each $i \in \Omega$, $\rho_i(t) \triangleq \Pr\{\theta = i|Y^t\}$ for $t \geq 1$, and $Y^0 = \emptyset$. The receiver's initial belief of $\theta = i$ is $\rho_i(0) = \Pr\{\theta = i\} = 1/M$. According to Bayes' rule, upon receiving $y_t$, $\rho_i(t)$ can be updated by

$$\rho_i(t) = \frac{\rho_i(t-1)P(Y=y_t|X=e_t(i,Y^{t-1}))}{\sum_{j\in\Omega}\rho_j(t-1)P(Y=y_t|X=e_t(j,Y^{t-1}))}. \tag{11}$$

Thanks to the noiseless feedback, the transmitter will be informed of $y_t$ at $t+1$ and thus can calculate the same $\boldsymbol{\rho}(t)$. The stopping time $\tau$ and decoding rule considered in [12] are given by

$$\tau = \min\{t : \max_{i\in\Omega} \rho_i(t) \geq 1 - \epsilon\} \tag{12}$$

$$\hat{\theta} = \arg\max_{i\in\Omega} \rho_i(\tau). \tag{13}$$

Clearly, with the above scheme, the probability of error meets the desired constraint, i.e.,

$$P_e = \mathbb{E}[1 - \max_{i\in\Omega} \rho_i(\tau)] \leq \epsilon. \tag{14}$$

For any DMC, Naghshvar *et al.* [9], [12] proposed an encoder, which we refer to as the *small-enough-difference* (SED) encoder, for symmetric binary-input channels (thus also for the BSC). This encoder is implemented using a partitioning algorithm, which, after calculating $\boldsymbol{\rho}(t-1)$, partitions $\Omega$ into two subsets $S_0(t-1)$ and $S_1(t-1)$ such that

$$0 \leq \sum_{i\in S_0(t-1)} \rho_i(t-1) - \sum_{i\in S_1(t-1)} \rho_i(t-1) \leq \min_{i\in S_0(t-1)} \rho(t-1). \tag{15}$$

Then, $X_t = 0$ if $\theta \in S_0(t-1)$ and $X_t = 1$ otherwise.

With the stopping time in (12) and the SED encoder in (15), Naghshvar *et al.* proved the following non-asymptotic upper bound on $\mathbb{E}[\tau]$ via a delicate submartingale synthesis.

**Theorem 1** (Remark 7, [12])**.** *The proposed scheme described in* (12), (13), *and* (15), *for symmetric binary-input channels satisfies,*

$$\mathbb{E}[\tau] \leq \frac{\log M + \log\log\frac{M}{\epsilon}}{C} + \frac{\log\frac{1}{\epsilon} + 1}{C_1} + \frac{96 \cdot 2^{2C_2}}{CC_1}. \tag{16}$$

## III. The Markovian Analysis on Average Blocklengths

In this section, we consider the problem of sequential transmission (or variable-length coding) over BSC with full, noiseless feedback. Specifically, we follow Naghshvar *et al.*'s framework described in Sec. II, i.e., the stopping time in (12), the decoding rule in (13), and the SED encoder in (15). Our analysis focuses on the BSC($p$) with crossover probability $0 < p < 1/2$.

Our main result is stated in the following theorem.

**Theorem 2.** *The proposed scheme described in* (12), (13)*, and* (15) *for the BSC*($p$)*,* $0 < p < 1/2$*, satisfies*

$$\mathbb{E}[\tau] \leq \frac{\log M}{C} + \frac{\left\lceil \frac{\log \frac{1-\epsilon}{\epsilon}}{C_2} \right\rceil C_2}{C_1} + \frac{pC_2}{C_1}\left(\frac{C_1 + C_2}{C} - \frac{C_2}{C_1}\right) + \frac{C_1}{C}. \tag{17}$$

The proof of Theorem 2 is given by our *Markovian analysis*. Unlike the proof technique of Theorem 1, first, we decompose the process into a communication phase and a confirmation phase that also accounts for the fallback of the transmitted message, i.e., the time when the transmitted message falls back from the confirmation phase to the communication phase and then returns to the confirmation phase. Next, we utilize submartingale results for the communication phase, but exploit the Markov structure of the confirmation phase to perform a time-of-first passage analysis. The constant term yielded from the time-of-first passage analysis explicitly captures the penalty of falling back, which is given by the third term of the RHS of (17). Over the course of the Markovian analysis, we will convince the reader that this constant term is in fact the differential time of the fallback between the actual process and the fictitious process of a random walk.

For brevity, throughout Sec. III, denote by $\theta = i \in \Omega$ the transmitted message unless otherwise specified.

### A. Previous Results of Naghshvar et al. and Polyanskiy

We first review several key results Naghshvar *et al.* demonstrated in [9] and [12] and Polyanskiy's VLF upper bound derived by Williamson *et al.* [15].

Let $\rho_i(t)$ denote the posterior of the transmitted message $\theta = i \in \Omega$. The log-likelihood ratio of $\theta = i$ is denoted

$$U_i(t) = \log \frac{\rho_i(t)}{1 - \rho_i(t)}. \tag{18}$$

For a given $\epsilon > 0$, define the genie-aided stopping time $\tau_i(\epsilon)$ of $\theta = i$ as

$$\tau_i(\epsilon) = \min\{t : \rho_i(t) \geq 1 - \epsilon\}. \tag{19}$$

With the SED encoder in (15), Naghshvar *et al.* proved that $\{U_i(t)\}_{t=0}^{\infty}$ forms a submartingale.

**Lemma 1** (Naghshvar *et al.*, [9])**.** *With the SED encoder in* (15)*,* $\{U_i(t)\}_{t=0}^{\infty}$ *forms a submartingale with respect to the filtration* $\mathcal{F}_t = \sigma\{Y^t\}$*, satisfying*

$$\mathbb{E}[U_i(t+1)|\mathcal{F}_t] \geq U_i(t) + C, \quad \text{if } U_i(t) < 0 \tag{20}$$
$$\mathbb{E}[U_i(t+1)|\mathcal{F}_t] = U_i(t) + C_1, \quad \text{if } U_i(t) \geq 0 \tag{21}$$
$$|U_i(t+1) - U_i(t)| \leq C_2. \tag{22}$$

*Proof:* See Appendix in [9] or [14]. ∎

**Remark 1.** *Lemma 1 characterizes the fundamental behavior of the transmitted message* $\theta$ *due to the SED encoder and the BSC*($p$)*. In [14], we also show that for any* $t \geq 1$*,*

$$C \leq \mathbb{E}[U_i(t+1) - U_i(t)] \leq C_1. \tag{23}$$

**Lemma 2** (Naghshvar *et al.*, [12])**.** *Assume that the sequence* $\{\xi_t\}_{t=0}^{\infty}$ *forms a submartingale with respect to a filtration* $\{\mathcal{F}_t\}$*. Furthermore, assume there exist positive constants* $K_1, K_2$*, and* $K_3$ *such that*

$$\mathbb{E}[\xi_{t+1}|\mathcal{F}_t] \geq \xi_t + K_1, \quad \text{if } \xi_t < 0$$
$$\mathbb{E}[\xi_{t+1}|\mathcal{F}_t] \geq \xi_t + K_2, \quad \text{if } \xi_t \geq 0$$
$$|\xi_{t+1} - \xi_t| \leq K_3, \quad \text{if } \max\{\xi_{t+1}, \xi_t\} \geq 0.$$

*Consider the stopping time* $v = \min\{t : \xi_t \geq B\}$*,* $B > 0$[2]*. Then we have*

$$\mathbb{E}[v] \leq \frac{B - \xi_0}{K_2} + \xi_0 \mathbb{1}_{\{\xi_0 < 0\}}\left(\frac{1}{K_2} - \frac{1}{K_1}\right) + \frac{3K_3^2}{K_1 K_2}. \tag{24}$$

Clearly, the submartingale in Lemma 1 can be incorporated into Lemma 2 by setting $\xi_t = U_i(t), K_1 = C, K_2 = C_1, K_3 = C_2$ and $B = \log \frac{1-\epsilon}{\epsilon}$. Thus, appealing to Lemma 2, we obtain the following tightened upper bound of $\mathbb{E}[\tau]$ over Theorem 1.

**Corollary 1.** *The proposed scheme described in* (12)*,* (13)*, and* (15) *for BSC*($p$)*,* $0 < p < 1/2$*, satisfies*

$$\mathbb{E}[\tau] \leq \frac{\log M}{C} + \frac{\log \frac{1-\epsilon}{\epsilon}}{C_1} + \frac{3C_2^2}{CC_1}. \tag{25}$$

Following Polyanskiy [11], Williamson *et al.* [15] derived the VLF upper bound on average blocklength for the BSC.

**Theorem 3** (Polyanskiy's VLF bound, [15])**.** *For a given* $\epsilon > 0$ *and a positive integer* $M$*, there exists a stop-feedback VLF code for BSC*($p$)*, with average blocklength satisfying*

$$\mathbb{E}[\tau] \leq \frac{\log(M-1)}{C} + \frac{\log \frac{1}{\epsilon}}{C} + \frac{\log 2(1-p)}{C}. \tag{26}$$

### B. The Markovian Analysis: Proof of Theorem 2

Consider a genie-aided decoder with the genie-aided stopping rule described in (19). Clearly, $\tau \leq \tau_i(\epsilon)$ for any $\theta = i \in \Omega$, by definition. Thus,

$$\mathbb{E}[\tau] = \mathbb{E}\big[\mathbb{E}[\tau|\theta = i]\big] \leq \mathbb{E}\big[\mathbb{E}[\tau_\theta(\epsilon)|\theta = i]\big] = \mathbb{E}[\tau_\theta(\epsilon)]. \tag{27}$$

[2]The stopping time $v$ can be shown to be a.s. finite in general for any positive threshold $B$. See [14] for more details.

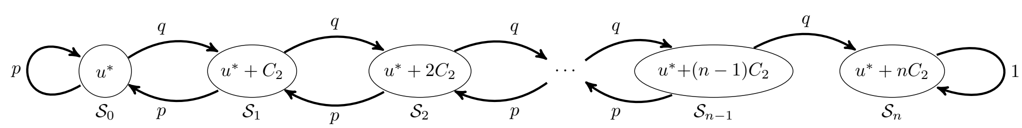


Fig. 2. An instance of the generalized Markov chain initiated at $U^*(t)=u^*$, where $u^*\in\mathcal{S}_0$ is some constant. The value inside the $i$-th circle is an element of state $\mathcal{S}_i$, $0\le i\le n$.

Note that the SED encoder does not rely on the location of $\theta$, we also have $\mathbb{E}[\tau_\theta(\epsilon)]=\mathbb{E}[\tau_i(\epsilon)|\theta=i]=\mathbb{E}[\tau_i(\epsilon)]$. Thus, it suffices to analyze $\mathbb{E}[\tau_i(\epsilon)]$ henceforth.

Now, we decompose $\mathbb{E}[\tau_i(\epsilon)]$ into

$$\begin{aligned}\mathbb{E}[\tau_i(\epsilon)]&=\mathbb{E}[\tau_i(1/2)+\tau_i(\epsilon)-\tau_i(1/2)]\\&=\mathbb{E}[\tau_i(1/2)]+\mathbb{E}_u\left[\mathbb{E}[\tau_i(\epsilon)-\tau_i(1/2)\mid U_i(\tau_i(1/2))=u]\right],\end{aligned}\tag{28}$$

where $\tau_i(1/2)=\min\{t:\rho_i(t)\ge 1/2\}$ following (19) and $u$ represents the log-likelihood ratio of the transmitted message when $\rho_i(t)$ crosses $1/2$ for the first time. By definition and Lemma 1, $0\le u<C_2$.

The decomposition in (28) provides a key insight on the average blocklength of the sequential transmission. It indicates that the overall average blocklength may be obtained as the sum of the expected time of first crossing of $1/2$ by $\rho_i(t)$ and the expected time after the first crossing of $1/2$ until $\rho_i(t)$ exceeds $1-\epsilon$. Our next step is to upper bound each term in the RHS of (28), which is given by Lemma 3 and 4, respectively. Finally, summing the two upper bounds yields Theorem 2.

**Lemma 3.** *With the SED encoder in* (15)*, the stopping time $\tau_i(1/2)$ of the transmitted message $\theta=i\in\Omega$ satisfies*

$$\mathbb{E}[\tau_i(1/2)]<\frac{\log M}{C}+\frac{C_1}{C}.\tag{29}$$

*Proof:* Let $\mathcal{F}_t=\sigma\{Y^t\}$ denote the history of receiver's knowledge up to time $t$. For brevity, let $T\triangleq\tau_i(1/2)$ be the shorthand notation for random variable $\tau_i(1/2)$. Consider $\eta_t=\frac{U_i(t)}{C}-t$. We can easily show that the new sequence $\{\eta_t\}_{t=0}^\infty$ is also a submartingale. Since $T$ is a.s. finite, by Doob's optional stopping theorem [16],

$$\begin{aligned}\frac{-\log(M-1)}{C}&=\mathbb{E}[\eta_0]\le\mathbb{E}[\eta_T]\\&=\frac{\mathbb{E}[U_i(T)-U_i(T-1)]+\mathbb{E}[U_i(T-1)]}{C}-\mathbb{E}[T]\\&\le\frac{C_1+0}{C}-\mathbb{E}[T].\end{aligned}$$

Therefore, rewriting the above inequality in terms of $\mathbb{E}[T]$ establishes the lemma.

■

**Lemma 4.** *With the SED encoder in* (15)*, the difference between stopping times $\tau_i(\epsilon)$ and $\tau_i(1/2)$ of the transmitted message $\theta=i\in\Omega$ satisfies, for any $0\le u<C_2$,*

$$\begin{aligned}&\mathbb{E}[\tau_i(\epsilon)-\tau_i(1/2)\mid\ U_i(\tau_i(1/2))=u]\\&\qquad\le\frac{\left\lceil\frac{\log\frac{1-\epsilon}{\epsilon}}{C_2}\right\rceil C_2}{C_1}+\frac{pC_2}{C_1}\left(\frac{C_1+C_2}{C}-\frac{C_2}{C_1}\right).\end{aligned}\tag{30}$$

*Proof:* The proof requires several steps. First, we show that if $\rho_i(t)\ge 1/2$ (or $U_i(t)\ge 0$), $U_i(t)$ forms a Markov chain, which is given by Lemma 5. Thus, $\mathbb{E}[\tau_i(\epsilon)-\tau_i(1/2)\mid U_i(\tau_i(1/2))=u]$ is equivalent to the expected time-of-first passage from $u$ to $\log\frac{1-\epsilon}{\epsilon}$. In order to capture the fact that $\theta$ could fall back to the communication phase and then re-enter the confirmation phase with probability 1[3], we consider the following *generalized Markov chain*.

**Definition 1.** *Let $\mathcal{S}_0=[0,C_2)$ represent the set of all possible values of log-likelihood ratio $u$ when $\rho_i(t)$ transitions from below $1/2$ to above $1/2$. Let $n\triangleq\lceil\log\frac{1-\epsilon}{\epsilon}/C_2\rceil$. Let $\mathcal{S}_j=[jC_2,jC_2+C_2)$, $1\le j\le n$. The generalized Markov chain is defined as a sequence of states $\mathcal{S}_0,\mathcal{S}_1,\ldots,\mathcal{S}_n$, satisfying*

$$\begin{aligned}P(\mathcal{S}_{j+1}|\mathcal{S}_j)&=P_{V|U}(u+C_2|u\in\mathcal{S}_{j-1})=q,\ \ 0\le j\le n-1\\P(\mathcal{S}_{j-1}|\mathcal{S}_j)&=P_{V|U}(u-C_2|u\in\mathcal{S}_j)=p,\quad 1\le j\le n,\\P(\mathcal{S}_0|\mathcal{S}_0)&=P_{V|U}(u'\in\mathcal{S}_0|u\in\mathcal{S}_0)=p,\\P(\mathcal{S}_n|\mathcal{S}_n)&=P_{V|U}(u'\in\mathcal{S}_n|u\in\mathcal{S}_n)=1.\end{aligned}$$

Fig. 2 illustrates an instance of the generalized Markov chain initiated at $U^*(t)=u^*$, where

$$U^*(t)\triangleq\begin{cases}U_i(t)-\left\lfloor\frac{U_i(t)}{C_2}\right\rfloor C_2, & \text{if}\ \ U_i(t)\ge 0\\+\infty, & \text{otherwise.}\end{cases}\tag{31}$$

Each time $U_i(t)\ge 0$, there is a Markov chain with the initial position $U^*(t)$ that can be readily determined from $U_i(t)$. Also, $U^*(t)$ remains constant as long as $U_i(t)\ge 0$.

Next, let us consider the *position-invariant stopping rule* as

$$\tau_i^*(\epsilon)=\min\left\{t:\left\lfloor\frac{U_i(t)}{C_2}\right\rfloor\ge\left\lceil\frac{\log\frac{1-\epsilon}{\epsilon}}{C_2}\right\rceil\right\}.\tag{32}$$

Thus, regardless of $U^*(t)$, the position-invariant stopping rule of (32) is achieved exactly when $U_i(t)$ enters state $\mathcal{S}_n$ of the generalized Markov chain for the first time. In contrast, the stopping rule of (19) might be achieved either at state $\mathcal{S}_n$ or $\mathcal{S}_{n-1}$, which complicates the analysis.

More importantly, the position-invariant stopping rule is more stringent than the genie-aided stopping rule in that it yields an upper bound on $\tau_i(\epsilon)$, i.e.,

$$\tau\le\tau_i(\epsilon)\le\tau_i^*(\epsilon).\tag{33}$$

This can be justified by the definition of $\tau_i(\epsilon)$ in (19) and that

$$\frac{U_i(\tau_i^*(\epsilon))}{C_2}\ge\left\lfloor\frac{U_i(\tau_i^*(\epsilon))}{C_2}\right\rfloor\ge\left\lceil\frac{\log\frac{1-\epsilon}{\epsilon}}{C_2}\right\rceil\ge\frac{\log\frac{1-\epsilon}{\epsilon}}{C_2}.\tag{34}$$

[3]This can be justified by the fact that $\tau_i(1/2)$ is a.s. finite.

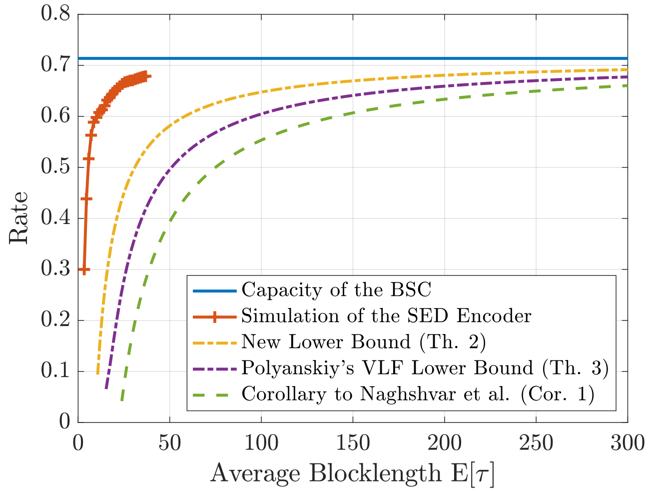


Fig. 3. The rate as a function of average blocklength over the BSC(0.05) with full, noiseless feedback. $\epsilon = 10^{-3}$.

That is, $\rho_i(\tau_i^*(\epsilon)) \geq 1 - \epsilon$, which concludes that (33) holds.

Let $V_i$ denote the expected time-of-first passage from state $\mathcal{S}_i$ to state $\mathcal{S}_n$, $i = 0, 1, \ldots, n-1$. Thus, for any $0 \leq u^* < C_2$,

$$\begin{aligned}\mathbb{E}[\tau_i(\epsilon) - &\tau_i(1/2) \mid U_i(\tau_i(1/2)) = u^*] \\ &\leq \mathbb{E}[\tau_i^*(\epsilon) - \tau_i(1/2) \mid U_i(\tau_i(1/2)) = u^*] = V_0. \quad (35)\end{aligned}$$

In the appendix of [14], the time-of-first passage analysis on the generalized Markov chain yields

$$V_0 = \frac{n}{1-2p} + \frac{p}{1-2p}\left(1 - \left(\frac{p}{1-p}\right)^n\right)(\Delta_0 - \Delta_0^*) \quad (36)$$

where $\Delta_0^*$ is the expected self-loop time from state $\mathcal{S}_0$ to state $\mathcal{S}_0$ associated with a standard random walk, $\Delta_0$ is the actual expected self-loop time from state $\mathcal{S}_0$ to state $\mathcal{S}_0$, which is also the expected time it takes to fall back to the communication phase from state $\mathcal{S}_0$ and then return to state $\mathcal{S}_0$. Here, the second term in (36) is exactly the differential time between the actual process and the fictitious random walk, which reveals the fundamental mechanism of the constant term in the upper bound. A more detailed analysis in [14] shows that

$$\Delta_0 \leq 1 + \frac{C_1 + C_2}{C}, \quad (37)$$

$$\Delta_0^* = 1 + \frac{C_2}{C_1}. \quad (38)$$

Therefore, combining (36), (37) and (38), we have

$$\begin{aligned}V_0 \leq &\frac{n}{1-2p} + \frac{p}{1-2p}\left(\frac{C_1+C_2}{C} - \frac{C_2}{C_1}\right) \\ = &\frac{nC_2}{C_1} + \frac{pC_2}{C_1}\left(\frac{C_1+C_2}{C} - \frac{C_2}{C_1}\right). \quad (39)\end{aligned}$$

Finally, appealing to (35) and (39) concludes the proof. ■

**Lemma 5.** *The log-likelihood ratio $U_i(t)$ of the transmitted message $\theta = i \in \Omega$ satisfies*

$$P(U_i(t+1) = u + C_2 | U_i(t) = u, u \geq 0) = q, \quad (40)$$

$$P(U_i(t+1) = u - C_2 | U_i(t) = u, u \geq 0) = p. \quad (41)$$

## IV. Numerical Simulation

In this section, we consider the BSC with crossover probability $p = 0.05$ and $\epsilon = 10^{-3}$. Then, it can be calculated that

$$C = 0.7136, \; C_1 = 3.8231, \; C_2 = 4.2479. \quad (42)$$

Clearly, this setting meets the technical conditions in [12]. Thus, from (16) given by Naghshvar *et al.*,

$$\mathbb{E}[\tau] \leq \frac{\log M + \log\log M + 3.32}{0.7136} + 2.87 + 12702.89, \quad (43)$$

which turns out to be a loose bound.

The rate of a VLF code is given by $R = \frac{\log M}{\mathbb{E}[\tau]}$, suggesting that an upper bound on $\mathbb{E}[\tau]$ corresponds to an achievability bound on rate.

Fig. 3 demonstrates the simulated rate performance of the SED encoder as a function of average blocklength $\mathbb{E}[\tau]$. We also plot the achievability bounds given by Theorem 2, Theorem 3, and Corollary 1. However, due to the exponential complexity of the partitioning algorithm, we were only able to simulate up to $k = 25$. Nevertheless, one can see that our new bound exceeds the lower bound of Polyanskiy on achievable rate for a system limited to stop feedback, as would be expected for a system utilizing full, noiseless feedback. In contrast, the corollary from Naghshvar *et al.*'s lemma lies beneath Polyanskiy's VLF lower bound, indicating that it does not capture the actual performance of the SED encoder. Indeed, we show analytically in [14] that our bound in Theorem 2 is tighter than that in Corollary 1 and than that in Theorem 3 if the crossover probability $p$ is moderately large.

## Acknowledgment

The authors are grateful to Professor Tara Javidi whose talk at UCLA in 2012 inspired our initial interest in this work and whose comments about this work as it has developed have been critical to finally reaching the desired result. Adam Williamson implemented the first simulation results in his doctoral work, which demonstrated that a smaller constant term than given in [12] might be possible with further analysis. Gourav Khadge helped implement the SED encoder for a wider range of average blocklengths. Anonymous reviewers of this manuscript at various stages of its development provided invaluable constructive criticism.

## References

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

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

[3] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 9, no. 3, pp. 136–143, July 1963.

[4] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1186–1222, March 2011.

[5] J. Schalkwijk, "A class of simple and optimal strategies for block coding on the binary symmetric channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 17, no. 3, pp. 283–287, May 1971.

[6] J. Schalkwijk and K. Post, "On the error probability for a class of binary recursive feedback strategies," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 498–511, July 1973.

[7] A. Tchamkerten and E. Telatar, "A feedback strategy for binary symmetric channels," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2002, pp. 362–362.

[8] A. Tchamkerten and I. E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2126–2145, May 2006.

[9] M. Naghshvar, M. Wigger, and T. Javidi, "Optimal reliability over a class of binary-input channels with feedback," in *2012 IEEE Inf. Theory Workshop*, Sep. 2012, pp. 391–395.

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

[11] ——, "Feedback in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4903–4925, Aug 2011.

[12] M. Naghshvar, T. Javidi, and M. Wigger, "Extrinsic Jensen–Shannon divergence: Applications to variable-length coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 2148–2164, April 2015.

[13] M. V. Burnashev and K. S. Zigangirov, "On one problem of observation control," *Problemy Peredachi Inf.*, vol. 11, no. 3, pp. 44–52, 1975.

[14] H. Yang and R. D. Wesel, "Finite-blocklength performance of sequential transmission over bsc with noiseless feedback." [Online]. Available: http://arxiv.org/abs/1902.00593

[15] A. R. Williamson, T. Chen, and R. D. Wesel, "Variable-length convolutional coding for short blocklengths with decision feedback," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2389–2403, July 2015.

[16] D. Williams, *Probability with Martingales*. Cambridge, United Kingdom: Cambridge University Press, 1991.