Nonasymptotic coding-rate bounds for binary erasure channels with feedback

Rahul Devassy[1], Giuseppe Durisi[1], Benjamin Lindqvist[1], Wei Yang[2], Marco Dalai[3]

[1]Chalmers University of Technology, 41296 Gothenburg, Sweden;
[2]Princeton University, Princeton, NJ, 08544, USA; [3]University of Brescia, 25123 Brescia, Italy

Abstract—We present nonasymptotic achievability and converse bounds on the maximum coding rate (for a fixed average error probability and a fixed average blocklength) of variable-length full-feedback (VLF) and variable-length stop-feedback (VLSF) codes operating over a binary erasure channel (BEC). For the VLF setup, the achievability bound relies on a scheme that maps each message onto a variable-length Huffman codeword and then repeats each bit of the codeword until it is received correctly. The converse bound is inspired by the meta-converse framework by Polyanskiy, Poor, and Verdú (2010) and relies on binary sequential hypothesis testing. For the case of zero error probability, our achievability and converse bounds match. For the VLSF case, we provide achievability bounds that exploit the following feature of BEC: the decoder can assess the correctness of its estimate by verifying whether the chosen codeword is the only one that is compatible with the erasure pattern. One of these bounds is obtained by analyzing the performance of a variable-length extension of random linear fountain codes. The gap between the VLSF achievability and the VLF converse bound, when number of messages is small, is significant: 23% for 8 messages on a BEC with erasure probability 0.5. The absence of a tight VLSF converse bound does not allow us to assess whether this gap is fundamental.

I. INTRODUCTION

In a point-to-point communication system with full feedback, the transmitter has noiseless access to all the previously received symbols. For discrete memoryless channels (DMCs), it turns out that this additional information does not increase capacity when codes of fixed blocklengths are used. Specifically, Shannon [1] proved that the capacity with full-feedback fixed-blocklength codes is no larger than the one achievable in the no-feedback case. Dobrushin [2] established a similar result for the reliability function of symmetric DMCs (the general case is, however, open). However, if the use of variable-length codes is permitted, the availability of full feedback turns out to be beneficial. Burnashev [3] derived the reliability function for the case when full feedback is available and variable-length feedback (VLF) codes are used, for all rates between zero and capacity. He showed that this reliability function is

$$E(R) = C_1 \left(1 - \frac{R}{C} \right), \qquad R \in (0, C). \qquad (1)$$

Here, C is the channel capacity and C_1 denotes the maximum relative entropy between two arbitrary conditional output distributions. Note that (1) is strictly larger than the reliability

function for the no-feedback case. Burnashev's proof relies on the asymptotic analysis of an achievability and a converse bound on the maximum rate obtainable with VLF codes, for a given average blocklength and a fixed average error probability. Yamamoto and Itoh [4] gave an alternative proof of Burnashev's achievability bound, which relies on a two-phase scheme: a standard transmission phase where feedback is not used at the transmitter is followed by a confirmation phase where the transmitter uses feedback to confirm/contradict the decision of the receiver. Berlin *et al.* [5] provided a stronger version of Burnashev's converse bound, whose proof parallels the two-phase scheme in [4]. A one-phase scheme that achieves (1) was proposed in [6].

Polyanskiy *et al.* [7] obtained a nonasymptotic converse bound that improves on Burnashev's one [3]. In the same work, an achievability bound is provided, which is used to show that with VLF codes one can approach capacity faster than in the fixed-blocklength case. Specifically, the *channel dispersion* [8, Eq. (221)] turns out to be zero. The achievability bound used in [7] to prove this result is actually based on variable-length stop-feedback (VLSF) codes. In the VLSF setup, the feedback link is used by the receiver only to send a single bit indicating to stop the transmission of the current message. This setup is of interest from a practical point of view, because it encompasses hybrid automatic repetition request (ARQ) schemes. Note that VLSF codes are a special case of VLF codes.

In this paper, we shall focus on the binary erasure channel (BEC) and seek nonasymptotic achievability and converse bounds, for both the VLF and the VLSF setups, which improve on the ones available in the literature. Note that the two-phase converse bounds [3, Thm. 1], [7, Thm. 6] require that all entries of the channel transition matrix of the DMC are strictly positive[1]— an assumption that does not hold for the BEC. Bounds on the maximum rate achievable over a BEC in the VLF setup are provided in [7, Thm. 7]. The achievability bound is based on a simple scheme (also suggested in [9]) where each bit is repeated until it is received correctly. The converse bound can be seen as a variable-length analogue of Fano's inequality (see [3, Lemmas 1 and 2]). This bound does not require C_1 to be finite.

The problem of constructing VLSF codes over a BEC reduces to problem of constructing rateless erasure codes. Thus, one can get achievability bounds on the maximum coding rate in the

---

[1]This is required for C_1 in (1) to be finite.

VLSF setup by analyzing the performance of family of rateless codes such as random linear fountain codes [10, Sec. 3]. The only converse bounds that are available for VLSF codes (stop feedback) hold also in the VLF setup (full feedback) to the best of the authors' knowledge. This is actually the case for both the maximum coding rate and the reliability function.

Our contributions in this paper are as follows:

- We provide nonasymptotic converse and achievability bounds on the maximum coding rate of VLF codes over BECs, which improve upon the ones provided in [7, Thm. 7]. Our converse bound relies on sequential hypothesis testing and is inspired by the meta-converse framework [8, Sec. III.E]; the achievability bound combines the simple repetition scheme used in [7, Thm. 7] with variable-length Huffman coding. For the case of zero error probability, the achievability and converse bounds match.

- For the VLSF setup, we provide nonasymptotic achievability bounds that improve on the one reported in [7, Thm. 3]. The bounds are obtained by exploiting that, for a BEC, the decoder is able to identify the correct message whenever only a single codeword is compatible with the sequence of channel outputs received up to that point (a property noted previously in e.g., [11]). The random coding argument used in one of the bounds utilizes linear codes. Hence, the resulting coding scheme can be seen as variable-length extension of random linear fountain codes.

Some proofs are omitted for space constraint; they can be found in [12].

Notation: Uppercase curly letters denote sets. The n-fold Cartesian product of a set $\mathcal{X}$ is denoted by $\mathcal{X}^n$. Uppercase boldface letters denote random quantities and lightface letters denote deterministic quantities. The distribution of a random variable $\mathbf{X}$ is denoted by $P_{\mathbf{X}}$. With $\mathbb{E}[\cdot]$ we denote expectation and with $\mathbb{E}_P[\cdot]$ we stress that the expectation is with respect to the probability law P. The indicator function is denoted by $\mathbf{1}\{\cdot\}$ and we use the symbol $\mathbb{F}_2$ to indicate the binary Galois field. With $\mathbf{X}_m^n$ we denote the random vector with entries $(\mathbf{X}_m, \mathbf{X}_{m+1}, \ldots, \mathbf{X}_n)$. Similarly, x_m^n stands for a deterministic vector with entries $(x_m, x_{m+1}, \ldots, x_n)$. We shall often use the following function:

$$\hat{\ell}(x) = \lfloor \log_2 x \rfloor + 2(1 - 2^{\lfloor \log_2 x \rfloor - \log_2 x}), \qquad x \in \mathbb{R}. \quad (2)$$

Here, $\lfloor \cdot \rfloor$ denotes the floor operator. Furthermore, we shall use $\lceil \cdot \rceil$ to denote the ceil operator. We let $\mathrm{Bern}(p)$ denote a Bernoulli-distributed random variable with parameter p and $\mathrm{Geom}(p)$ a geometrically distributed random variable with parameter p. The binary entropy function $H_{\mathrm{b}}(\cdot)$ is defined as follows:

$$H_{\mathrm{b}}(x) = -x \log_2 x - (1-x) \log_2 (1-x), \quad x \in (0,1). \quad (3)$$

II. DEFINITION

We consider a BEC with input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y} = \{0, \mathrm{e}, 1\}$, where e denotes an erasure. A VLF code for the BEC is defined as follows.

Definition 1: ([7, Def. 1]) An (ℓ, M, ϵ)–VLF code, where ℓ is a positive real, M is a positive integer, and $\epsilon \in [0, 1]$, consists of:

1) A random variable $\mathbf{U}$, defined on a set $\mathcal{U}$ with[2] $|\mathcal{U}| \leq 2$, whose realization is revealed to the encoder and the decoder before the start of transmission. The random variable $\mathbf{U}$ acts as common randomness and enables the use of randomized encoding and decoding strategies.

2) A sequence of encoders $f_n : \mathcal{U} \times \mathcal{W} \times \mathcal{Y}^{n-1} \to \mathcal{X}, n \geq 1$ that generate the channel inputs

$$\mathbf{X}_n = f_n\big(\mathbf{U}, \mathbf{W}, \mathbf{Y}_1^{n-1}\big). \quad (4)$$

Here, $\mathbf{W}$ denotes the message, which is uniformly distributed on $\mathcal{W} = \{1, 2, \ldots, M\}$. Note that the channel input at time n depends on all previous channel outputs (full feedback).

3) A sequence of decoders $g_n : \mathcal{U} \times \mathcal{Y}^n \to \mathcal{W}$ that provide the estimate of $\mathbf{W}$ at time n.

4) A nonnegative integer-valued random variable τ, which is a stopping time of the filtration

$$\mathcal{G}_n = \sigma\{\mathbf{U}, \mathbf{Y}_1^n\} \quad (5)$$

and satisfies

$$\mathbb{E}[\tau] \leq \ell. \quad (6)$$

5) The final estimate $\widehat{\mathbf{W}} = g_\tau(\mathbf{U}, \mathbf{Y}_1^\tau)$ of $\mathbf{W}$, which satisfies the error-probability constraint

$$\Pr\big\{\widehat{\mathbf{W}} \neq \mathbf{W}\big\} \leq \epsilon. \quad (7)$$

The rate R of an (ℓ, M, ϵ)–VLF code is defined as

$$R = \frac{\log_2 M}{\mathbb{E}[\tau]}. \quad (8)$$

Furthermore, we define the minimum average blocklength of VLF codes with M codewords and error probability not exceeding ϵ as follows:

$$\ell_f^*(M, \epsilon) = \min\{\ell \ : \ \exists (\ell, M, \epsilon)\text{–VLF code}\}. \quad (9)$$

VLSF codes are a special case of VLF codes. The peculiarity of VLSF codes is that the sequence of encoders is not allowed to depend on the past channel outputs, i.e.,

$$f_n : \mathcal{U} \times \mathcal{W} \to \mathcal{X}, \ n \geq 1. \quad (10)$$

In the VLSF case, the feedback link is used by the receiver only to inform the transmitter that the message has been decoded (stop/decision feedback).

III. EXISTING RESULTS FOR BEC

In this section, we review the results available in literature on the minimum average blocklength $\ell_f^*(M, \epsilon)$ for the BEC. The following achievability bound is obtained by time-sharing between a scheme that drops the message to be transmitted without using the channel at all, and a scheme that repeats the channel input until it is received correctly.

Theorem 1: ([7, Thm. 7]) For a BEC with erasure probability δ, there exists an (ℓ, M, ϵ)–VLF code with

$$\ell \leq \frac{(1-\epsilon)\lceil \log_2 M \rceil}{1-\delta}. \quad (11)$$

---

[2]The bound on the cardinality of $\mathcal{U}$ given in [7] (i.e., $|\mathcal{U}| \leq 3$) can be improved by using the Fenchel-Eggleston theorem [13, p. 35] in place of Caratheodory's theorem.

Next, we provide a converse bound.

Theorem 2: ([7, Thm. 7],[3, Lemmas 1 and 2]) For every (ℓ, M, ϵ)–VLF code with $0 \leq \epsilon \leq 1 - 1/M$ operating over a BEC with erasure probability δ, we have

$$\ell \geq \frac{(1-\epsilon)\log_2 M - H_{\mathsf{b}}(\epsilon)}{1-\delta}. \tag{12}$$

This converse bound can be obtained by constructing an appropriate martingale using the conditional entropy of the *a posteriori* distribution of the message given the channel output. Note that the bounds (11) and (12) coincide for $\epsilon = 0$ whenever the number of messages M is a power of 2.

IV. NOVEL BOUNDS FOR VLF CODES

In this section, we present an achievability and a converse bound on $\ell_f^*(M, \epsilon)$ that improve upon the ones given in Theorems 1 and 2. The idea behind the achievability bound is to combine the scheme in Theorem 1 with a Huffman code, whose purpose is to reduce the average blocklength when the number of messages is not a power of two. The converse bound relies on sequential hypothesis testing and is inspired by the meta-converse framework [8, Sec. III.E]. As we shall see, achievability and converse bounds are tight when $\epsilon = 0$, *for every* integer M. Our achievability bound is given in Theorem 3 below.

Theorem 3: For a BEC with erasure probability δ, there exists an (ℓ, M, ϵ)–VLF code with

$$\ell \leq \frac{(1-\epsilon)\hat{\ell}(M)}{1-\delta} \tag{13}$$

where $\hat{\ell}(\cdot)$ is defined in (2).

Proof: See [12, App. A]. ∎

The converse bound is based on binary sequential hypothesis testing [14]. Let[3] $(\mathbf{D}, \tau)$ denote a generic binary sequential hypothesis test between two stochastic processes P and Q. Here, τ is a stopping time and $\mathbf{D}$ is a decision rule (0 indicates that P is chosen and 1 that Q is chosen). Let $\mathcal{R}(P, Q)$ denote the set of all possible binary sequential hypothesis tests. We are interested in the minimum average number of samples $\ell_{\alpha_P, \alpha_Q}(P, Q)$ required by a binary sequential hypothesis test to identify P and Q correctly with probability at least α_P and α_Q, respectively. Formally,

$$\ell_{\alpha_P, \alpha_Q}(P, Q) = \min_{\substack{(\mathbf{D}, \tau) \in \mathcal{R}(P, Q), \\ P[\mathbf{D}=0] \geq \alpha_P, \\ Q[\mathbf{D}=1] \geq \alpha_Q}} \mathbb{E}_P[\tau]. \tag{14}$$

In Lemma 4 below we establish a connection between $\ell_{\alpha_P, \alpha_Q}(P, Q)$ and the parameters of a given VLF code. For the sake of generality, the lemma is formulated for the case of arbitrary DMCs (this requires a suitable generalization of the definition of VLF codes provided in Definition 1 to arbitrary DMCs).

Lemma 4: Consider an (ℓ, M, ϵ)–VLF code for the DMC $P_{\mathbf{Y}|\mathbf{X}}$. Let ϵ_Q denote the probability of error when this code is used over the DMC $Q_{\mathbf{Y}|\mathbf{X}}$. Let $\left\{ P_{\mathbf{U}, \mathbf{X}_1^n, \mathbf{Y}_1^n} \right\}_{n=0}^{\infty}$ and $\left\{ Q_{\mathbf{U}, \mathbf{X}_1^n, \mathbf{Y}_1^n} \right\}_{n=0}^{\infty}$ be the probability distribution of

---

[3]We use the same notation as in [15, Ch. 3].

---

the process $\mathbf{U}, \{(\mathbf{X}_n, \mathbf{Y}_n)\}_{n=1}^{\infty}$ under $P_{\mathbf{Y}|\mathbf{X}}$ and $Q_{\mathbf{Y}|\mathbf{X}}$, respectively. The distributions of the stochastic processes depend on the chosen (ℓ, M, ϵ)–VLF code through its encoder according to (4). We consider binary sequential hypothesis testing between the two processes, under the assumption that the realization of $\mathbf{U}$ is known to the test before processing $(\mathbf{X}_1, \mathbf{Y}_1)$. We have

$$\ell \geq \ell_{1-\epsilon, \epsilon_Q} \left(\left\{ P_{\mathbf{U}, \mathbf{X}_1^n, \mathbf{Y}_1^n} \right\}_{n=0}^{\infty}, \left\{ Q_{\mathbf{U}, \mathbf{X}_1^n, \mathbf{Y}_1^n} \right\}_{n=0}^{\infty} \right). \tag{15}$$

Proof: See Appendix A. ∎

The bound (15) can be viewed as the variable-length analogue of the meta-converse theorem [8, Thm. 26]. The meta-converse theorem links the average error probabilities resulting by using the same fixed-blocklength code over two different channels by means of binary hypothesis testing. Similarly, Lemma 4 relates the average error probabilities and the average blocklengths resulting by using a given VLF code over two different channels by means of binary sequential hypothesis testing.

To obtain a converse bound from (15), we take $Q_{\mathbf{Y}|\mathbf{X}} = Q_{\mathbf{Y}}$, with $Q_{\mathbf{Y}}$ being the capacity-achieving output distribution of the BEC. Then, we solve the minimization in (14) by using the sequential probability ratio rest (SPRT) [14] (see [12, App. C] for a short review). This yields the following bound.

Theorem 5: Every (ℓ, M, ϵ)–VLF code operating over a BEC with erasure probability δ satisfies

$$\ell \geq \frac{(1-\epsilon)\hat{\ell}(M(1-\epsilon))}{1-\delta}. \tag{16}$$

Proof: See [12, App. D]. ∎

We would like to emphasize that although (16) is tighter than the converse bound reported in [7, Thm. 6], a generalization of Theorem 5 to DMCs with finite C_1 yields a converse bound that is in general looser than the ones reported in [3, Thm. 1] and [7, Thm. 6]. The peculiarity of the BEC is that the decoder is able to determine if its estimate is correct or not by assessing whether the estimated codeword is the only codeword compatible with the erasure pattern. This implies that a two-phase scheme with a confirmation from the encoder is not required.

Note that the right-hand sides of (13) and (16) coincide when $\epsilon = 0$. This fact is collected in the following corollary.

Corollary 6: The minimum average blocklength $\ell_f^*(M, 0)$ of an $(\ell, M, 0)$–VLF code over a BEC with erasure probability δ is given by

$$\ell_f^*(M, 0) = \frac{\hat{\ell}(M)}{1-\delta} \tag{17}$$

where $\hat{\ell}(\cdot)$ is defined in (2).

V. NOVEL BOUNDS FOR VLSF CODES

We now focus on the VLSF setup and provide achievability bounds for the case $\epsilon = 0$. Achievability bounds for arbitrary ϵ can be obtained by allowing the receiver to send a stop signal at time zero with probability ϵ (see [7, Sec. III.D]). The corresponding achievability bounds can be readily obtained from the ones presented in this section by multiplying them by $(1 - \epsilon)$.

As already mentioned, in the BEC case the decoder can assess the correctness of its message estimate by verifying whether the codeword corresponding to the chosen message is the only one that is compatible with the received sequence. It is therefore natural to consider a decoder whose stopping time is given by

$$\tau = \inf\left\{n \geq 1 : \Pr\left\{\mathbf{W} = \widehat{W}_n \mid \mathbf{Y}_1^n = y_1^n\right\} = 1\right\} \quad (18)$$

where $\widehat{W}_n$ denotes the message estimate at the decoder after n channel uses:

$$\widehat{W}_n = \arg\max_{w \in \mathcal{W}} \Pr\{\mathbf{W} = w \mid \mathbf{Y}_1^n = y_1^n\}. \quad (19)$$

The decoding rule (18)–(19) combined with random coding (independent and identically distributed (i.i.d.) $\mathrm{Bern}(0.5)$ ensemble) yields the following achievability bound.

Theorem 7: For a BEC with erasure probability δ, there exists an $(\ell, M, 0)$–VLSF code with

$$\ell \leq \frac{1}{1-\delta}\left(1 - \sum_{i=1}^{M-1}\binom{M-1}{i}\frac{(-1)^i}{2^i - 1}\right). \quad (20)$$

Proof: See Appendix B. ∎

The achievability bound (20) suffers from two pitfalls: (i) the bound is loose when M is small because the random coding ensemble contains few codebooks with abnormally large average blocklengths; (ii) since the bound requires the computation of differences of binomial coefficients, it becomes difficult to compute when M is larger than 10^4. Next, we present a different achievability bound that addresses these two shortcomings. To tighten (20) for small M we use an expurgation technique similar to the ones utilized by Shannon, Gallager, and Berlekamp [16, p. 529]. Specifically, we view each codebook as a random matrix with M rows and an infinite number of columns and we assign the following probability distribution on the VLSF code ensemble: each column is drawn uniformly and independently from the set of binary vectors with $\lceil M/2 \rceil$ zeros. Furthermore, to obtain an expression that is computable for arbitrary values of M, we upper-bound the average blocklength of the expurgated ensemble using the union bound. The achievability bound thus obtained is given in the following theorem.

Theorem 8: For a BEC with erasure probability δ, there exists an $(\ell, M, 0)$–VLSF code with

$$\ell \leq \frac{1}{1-\delta}\left(\lfloor m \rfloor + 1 + \frac{\mu^{m - \lfloor m \rfloor}}{\mu - 1}\right) \quad (21)$$

where $m = \log_\mu(M-1)$ and μ is related to the number of messages M as follows:

$$\mu = 2 + \frac{1}{\lceil M/2 \rceil - 1}. \quad (22)$$

Proof: See [12, App. G]. ∎

The parameter μ in (22) is the reciprocal of the probability that the first two bits in a random vector that is uniformly distributed over the set of vectors in $\mathbb{F}_2^M$ with $\lceil M/2 \rceil$ zeros, are equal.

For the case when the number of messages M is a power of 2, one can obtain an achievability bound that is tighter than (21), that
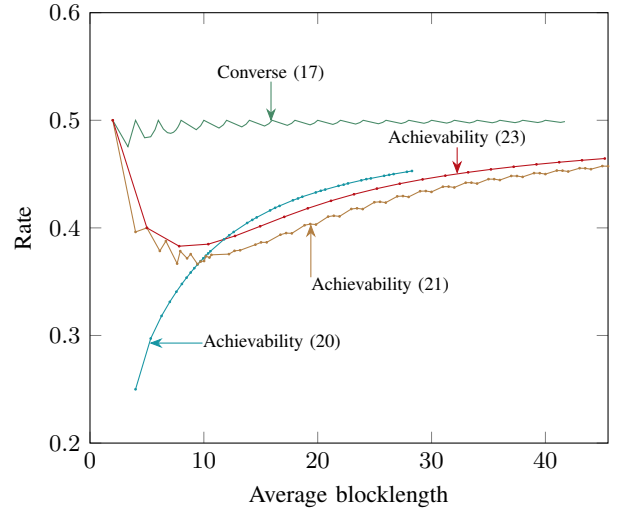

Fig. 1: Achievability and converse bounds for zero error VLSF codes over BEC with $\delta = 0.5$. The converse bound is the one given in (17) for VLF codes.

does not require the union bound, and that is easily computable. The bound relies on a linear codebook ensemble in which the columns of the random generator matrix are distributed uniformly over the set of all nonzero vectors from $\mathbb{F}_2^{\log_2 M}$. Specifically, consider a received vector of length n and let $\mathcal{I}_n$ be the set containing the indices of unerased symbols in the received vector. We are interested in the dimension of the subspace spanned by the columns of the generator matrix with index in $\mathcal{I}_n$. This dimension evolves as a Markov chain with a single absorbing state (the state that corresponds to maximum dimension $\log_2 M$). it follows that the average blocklength (averaged over the ensemble) coincides with the expected absorption time of the Markov chain, which follows a discrete phase-type distribution [17, Ch. 2]. This achievability scheme can be seen as a variable-length analogue of random linear fountain codes [10, Sec. 3]. The performance of the achievability scheme just described is characterized in the following theorem.

Theorem 9: For every integer $k \geq 1$, there exists an $(\ell, 2^k, 0)$–VLSF code for a BEC with erasure probability δ with

$$\ell \leq \frac{1}{1-\delta}\left(k + \sum_{i=1}^{k-1}\frac{2^i - 1}{2^k - 2^i}\right). \quad (23)$$

Proof: See [12, App. H]. ∎

The achievability bounds (20), (21), and (23) are plotted in Fig. 1. As expected, the bound (21) is tighter than (20) at small average blocklengths (because of expurgation) and looser at large average blocklengths (because of union bound). The achievability bound (23) is tighter than (21) for all blocklengths and looser than (20) for large average blocklengths. When $M = 2$, the achievability bounds (21) and (23) coincide with the converse bound for VLF codes given in (17). This holds because the scheme that achieves (17) (repeat each bit until it is received correctly) can be implemented with stop feedback when[4] $M = 2$.

---

[4]Recall that in the VLSF setup the decoder is allowed to send only one stop signal per message.

$$\Pr\{\mathbf{Z} = 0 \,|\, \mathbf{X}_1^n = x_1^n, \mathbf{Y}_1^n = y_1^n, \mathbf{U} = u, \tau = n\}$$
$$= \sum_{w \in \mathcal{W}} \Pr\{\mathbf{W} = w \,|\, \mathbf{X}_1^n = x_1^n, \mathbf{Y}_1^n = y_1^n, \mathbf{U} = u\} \Pr\Big\{\widehat{\mathbf{W}} = w \,|\, \mathbf{Y}_1^n = y_1^n, \mathbf{U} = u, \tau = n\Big\}. \qquad (24)$$

As M increases, the gap between the VLSF achievability bounds and the VLF converse bound increases and gets as large as 23% when $M = 8$, before vanishing asymptotically as $M \to \infty$. It remains to be seen whether this gap is fundamental.

APPENDIX A
PROOF OF LEMMA 4

Consider the random variable $\mathbf{Z} = \mathbf{1}\Big\{\mathbf{W} \neq \widehat{\mathbf{W}}\Big\}$. The conditional distribution of $\mathbf{Z}$ given $\mathbf{X}_1^n = x_1^n, \mathbf{Y}_1^n = y_1^n, \mathbf{U} = u, \tau = n$, which is given in (24), does not depend on whether the underlying channel is $P_{\mathbf{Y}|\mathbf{X}}$ or $Q_{\mathbf{Y}|\mathbf{X}}$. Indeed in (24), the first factor depends only on the encoder and second factor depends only on the decoder. Using the stopping time τ associated to the given code and the family of probability kernels defined by the conditional distribution (24) we construct a binary sequential hypothesis test $(\mathbf{D}, \tau)$. By definition, we have that under $P_{\mathbf{Y}|\mathbf{X}}$,

$$\Pr\{\mathbf{D} = 0\} = 1 - \epsilon \qquad (25)$$

and under $Q_{\mathbf{Y}|\mathbf{X}}$

$$\Pr\{\mathbf{D} = 0\} = 1 - \epsilon_Q. \qquad (26)$$

Thus,

$$\ell \geq \mathbb{E}_{\left\{P_{\mathbf{U}, \mathbf{X}_1^n, \mathbf{Y}_1^n}\right\}_{n=0}^{\infty}}[\tau] \qquad (27)$$

$$\geq \ell_{1-\epsilon, \epsilon_Q}\left(\left\{P_{\mathbf{U}, \mathbf{X}_1^n, \mathbf{Y}_1^n}\right\}_{n=0}^{\infty}, \left\{Q_{\mathbf{U}, \mathbf{X}_1^n, \mathbf{Y}_1^n}\right\}_{n=0}^{\infty}\right). \qquad (28)$$

APPENDIX B
PROOF OF THEOREM 7

We consider the VLSF codebook ensemble specified by the set of all binary matrices with M rows and infinitely many columns. Furthermore, we assign a probability distribution on this ensemble by assuming each entry in the codebook being i.i.d. $\text{Bern}(0.5)$. Using the stopping time (18) and the decoder (19), we can now create a VLSF code ensemble. By [12, Lemma 10], we can write the ensemble average blocklength as

$$\mathbb{E}[\tau] = \frac{\mathbb{E}[\tau_0]}{1 - \delta} \qquad (29)$$

where τ_0 is the stopping time when $\delta = 0$. Let $\mathcal{E}_w(\mathbf{X}_1^n)$ be the event that the bits $\mathbf{X}_1^n$, which are distributed i.i.d. $\text{Bern}(0.5)$, coincide with the first n bits of the codeword corresponding to message w. Without loss of generality, we assume that message 1 is transmitted. The ensemble average of τ_0 is given by

$$\mathbb{E}[\tau_0] = 1 + \sum_{n=1}^{\infty} \Pr\left\{\bigcup_{w=2}^{M} \mathcal{E}_w(\mathbf{X}_1^n)\right\} \qquad (30)$$

$$= \sum_{n=0}^{\infty}\left(1 - \left(1 - 2^{-n}\right)^{M-1}\right) \qquad (31)$$

$$= 1 - \sum_{i=1}^{M-1}\binom{M-1}{i}\frac{(-1)^i}{2^i - 1}. \qquad (32)$$

In (32), we used the binomial theorem and the summation formula for geometric series. Substituting (32) into (29) we conclude that

$$\mathbb{E}[\tau] = \frac{1}{1 - \delta}\left(1 - \sum_{i=1}^{M-1}\binom{M-1}{i}\frac{(-1)^i}{2^i - 1}\right). \qquad (33)$$

Since there exists at least one VLSF code in the ensemble with average blocklength lower than the ensemble average blocklength, we conclude that (20) must hold.

REFERENCES

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

[2] R. L. Dobrushin, "An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback," vol. 8, pp. 161–168, 1962.

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

[4] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, pp. 729–733, Nov. 1979.

[5] P. Berlin, B. Nakiboğlu, B. Rimoldi, and I. Telatar, "A simple converse of Burnashev's reliability function," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3074–3080, Jul. 2009.

[6] 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, Apr. 2015.

[7] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4903–4925, Aug. 2011.

[8] ——, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

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

[10] D. J. MacKay, "Fountain codes," *Inst. Electr. Eng. Proc.-Commun.*, vol. 152, no. 6, pp. 1062–1068, Dec. 2005.

[11] J. L. Massey, "Zero error," presented at the IEEE Winter School on Coding and Information Theory, Mar 2007.

[12] R. Devassy, G. Durisi, B. Lindqvist, W. Yang, and M. Dalai, "Nonasymptotic coding-rate bounds for binary erasure channels with feedback (extended version)." [Online]. Available: http://arxiv.org/pdf/1607.06837.pdf

[13] H. G. Eggleston, *Convexity*. New York: Cambridge university press, 1958.

[14] A. Wald, "Sequential tests of statistical hypotheses," *Ann. Math. Statist.*, vol. 16, no. 2, pp. 117–186, Jun. 1945.

[15] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*. Boca Raton, Florida, USA: CRC Press, 2014.

[16] C. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. {II}," *Information and Control*, vol. 10, no. 5, pp. 522 – 552, Jan. 1967.

[17] M. Neuts, *Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach*. Baltimore, MD: The Johns Hopkins Univ. Press, 1981.