Variable-Length Stop-Feedback Codes With Finite Optimal Decoding Times for BI-AWGN Channels

Hengjie Yang, Recep Can Yavas, Victoria Kostina, and Richard D. Wesel

***Abstract*—In this paper, we are interested in the performance of a variable-length stop-feedback (VLSF) code with m optimal decoding times for the binary-input additive white Gaussian noise channel. We first develop tight approximations on the tail probability of length-n cumulative information density. Building on the work of Yavas *et al.*, for a given information density threshold, we formulate the integer program of minimizing the upper bound on average blocklength over all decoding times subject to the average error probability, minimum gap and integer constraints. Eventually, minimization of locally minimum upper bounds over all thresholds will yield the globally minimum upper bound and this is called the two-step minimization. For the integer program, we present a greedy algorithm that yields possibly suboptimal integer decoding times. By allowing a positive real-valued decoding time, we develop the gap-constrained sequential differential optimization (SDO) procedure that sequentially produces the optimal, real-valued decoding times. We identify the error regime in which Polyanskiy's scheme of stopping at zero does not improve the achievability bound. In this error regime, the two-step minimization with the gap-constrained SDO shows that a finite m suffices to attain Polyanskiy's bound for VLSF codes with $m = \infty$.**

I. Introduction

Feedback has been shown to be useful both in the variable-length and fixed-length regimes, even though it does not improve the capacity of a memoryless, point-to-point channel [1]. In the variable-length regime, feedback has been shown to simplify the construction of coding schemes [2]–[4], to significantly improve the optimal error exponent [5], and to achieve universality [6]–[8]. In the fixed-length regime, feedback is shown to improve the second-order coding rate for the compound-dispersion discrete memoryless channels [9].

In [10], Polyanskiy *et al.* introduced variable-length feedback (VLF) codes, variable-length feedback with termination (VLFT) codes, and a special VLF code called a *variable-length stop-feedback (VLSF)* code. The infinite-length VLSF codewords are fixed before the start of transmission and feedback only affects the portion of a codeword being transmitted rather than the value of that codeword. During transmission, a feedback symbol "0" indicates that the decoder is not ready to decode and the transmission should continue, whereas a "1" signifies that the decoder is ready to decode and the transmitter must stop. Using VLSF codes, Polyanskiy *et al.* demonstrated that $\frac{C}{1-\epsilon}$ is achievable by stopping the code at $\tau = 0$ with a small probability, where C denotes channel capacity, and ϵ denotes the target error probability [10].

The VLSF code defined in [10] can be thought of as a VLF code with infinitely many decoding times, i.e., the number of decoding times $m = \infty$. However, in practical systems, the feedback opportunities are limited, i.e., $m < \infty$, and the decoder is only allowed to decode at time instants $n_1, n_2, \ldots, n_m$. In [11], Kim *et al.* investigated VLSF codes with m periodic decoding times and derived a lower bound on throughput. In order to minimize the average blocklength, Vakilinia *et al.* [12] developed the *sequential differential optimization* (SDO) algorithm that produces decoding time n_{k+1} based on the knowledge of n_k, n_{k-1}, and their successful decoding probabilities approximated by a differentiable function. The SDO in [12] uses the Gaussian tail probability to approximate the probability of successful decoding. Later, variations of SDO were developed to improve the Gaussian model accuracy [13], [14]. The SDO algorithm is used to optimize systems that employ incremental redundancy and hybrid automatic repeat request (ARQ) [15], and to code for the binary erasure channel [16], [17]. However, in this paper, we show that the Gaussian model is still imprecise for small values of n. Additionally, the existing SDO procedure fails to consider the inherent gap constraint that two decoding times must be separated by at least one.

In [18], Yavas *et al.* developed an achievability bound for VLSF codes with m decoding times for the additive white Gaussian noise channel with capacity C, dispersion V, and maximal power constraint P. The asymptotic expansion of the maximum message size M is given by $\ln M \approx \frac{lC}{1-\epsilon} - \sqrt{l \ln_{(m-1)}(l) \frac{V}{1-\epsilon}}$ where $\ln_{(k)}(\cdot)$ denotes the k-fold nested logarithm, l and ϵ are the upper bounds on average blocklength and error probability of the VLSF code, respectively. They showed that a slight increase in m can dramatically improve the achievable rate of VLSF codes. Unfortunately, due to the nested logarithm term, Yavas *et al.* were only able to show achievability bounds for $m \leq 4$ for average blocklength less than 2000. They also demonstrated that within their code construction, the decoding times chosen by the SDO will yield the same second- and third-order coding rates as attained by their construction of decoding times.

In this paper, we are interested in the performance of a VLSF code with m optimal decoding times for the binary-input additive white Gaussian noise (BI-AWGN) channel. We first develop tight approximations on the tail probability of length-n cumulative information density. Building on the result of Yavas *et al.* [18], for a fixed information density

H. Yang and R. D. Wesel are with the Department of Electrical and Computer Engineering, University of California, Los Angeles (e-mail: {hengjie.yang, wesel}@ucla.edu). R. C. Yavas and V. Kostina are with the Department of Electrical Engineering, California Institute of Technology (e-mail: {ryavas, vkostina}@caltech.edu). This research is supported by National Science Foundation (NSF) grant CCF-1955660.

threshold γ, we formulate an integer program of minimizing the upper bound on average blocklength over all decoding times $n_1, n_2, \ldots, n_m$ subject to average error probability, minimum gap and integer constraints. Finally, minimization of locally minimum upper bounds over information density threshold γ yields the globally minimum upper bound, and this method is called the *two-step minimization*. For the integer program, we present a greedy algorithm that yields possibly suboptimal integer decoding times. By allowing positive real-valued decoding times, we develop the *gap-constrained SDO* algorithm that captures the minimum gap constraint for the relaxed program. In [10], Polyanskiy *et al.* demonstrated that the rate $\frac{C}{1-\epsilon}$ is achievable by allowing the VLSF code to stop at zero with a small probability. In this paper, we identify the error regime where Polyanskiy's scheme of stopping at zero does not improve the achievability bound. In this error regime, the two-step minimization with the gap-constrained SDO shows that a finite m suffices to attain Polyanskiy's bound for VLSF codes with $m = \infty$.

This paper is organized as follows. Section II introduces the notation, the BI-AWGN channel model, and the VLSF code with m decoding times. Section III develops tight approximations on the tail probability of length-n cumulative information density. Section IV introduces the integer program, the two-step minimization, and a greedy algorithm, develops the gap-constrained SDO procedure for the relaxed program, identifies the error regime where stopping at zero does not help, and shows numerical comparisons. Section V concludes the paper.

II. Preliminaries

A. Notation

For $k \in \mathbb{Z}_+$, $[k] \triangleq \{1, 2, \ldots, k\}$. We use x_i^j to denote a sequence $(x_i, x_{i+1}, \ldots, x_j)$, $1 \le i \le j$. When the context is clear, x_1^n is abbreviated as x^n. All logarithms are taken to the base 2. We use $\phi(x), \Phi(x), Q(\cdot)$ to respectively denote the probability density function (PDF), cumulative distribution function (CDF), and the tail probability of a standard normal $\mathcal{N}(0,1)$.

B. Channel Model and VLSF Codes with m Decoding Times

Let X^n be a sequence of independent and identically distributed (i.i.d.) random variables, with each X_i uniformly distributed over $\{-1, 1\}$. The output Y^n of a memoryless, point-to-point BI-AWGN channel in response to X^n is given by

$$Y^n = \sqrt{P}X^n + Z^n, \tag{1}$$

where $\sqrt{P}$ denotes the amplitude of binary-phase shift keying (BPSK), and $Z_1, Z_2, \ldots, Z_n$ are i.i.d. standard normal random variables. The SNR of the BI-AWGN channel is given by P.

For a BI-AWGN channel with a uniformly distributed input symbol, the information density $\imath(x;y) \triangleq \log \frac{P(y|x)}{P(y)}$ is given by

$$\imath(x;y) = 1 - \log\left(1 + \exp\left(-2xy\sqrt{P}\right)\right). \tag{2}$$

Since the channel is memoryless, the cumulative information density for x^n and y^n is given by

$$\imath(x^n;y^n) \triangleq \log \frac{P(y^n|x^n)}{P(y^n)} = \sum_{i=1}^{n} \imath(x_i;y_i). \tag{3}$$

For a BI-AWGN channel, the channel capacity $C = \mathbb{E}[\imath(X;Y)]$ and dispersion $V = \mathrm{var}(\imath(X;Y))$.

Next, we follow [18] in describing a VLSF code with m decoding times for the BI-AWGN channel. Due to BPSK, we omit the power constraint from the definition.

An (l, n_1^m, M, ϵ) VLSF code, where l is a positive real, n_1^m and M are non-negative integers satisfying $n_1 < n_2 < \cdots < n_m$, $\epsilon \in (0,1)$, is defined by

1) A finite alphabet $\mathcal{U}$ and a probability distribution P_U on $\mathcal{U}$ defining the common randomness random variable U that is revealed to both the transmitter and the receiver before the start of transmission.
2) A sequence of encoders $f_n : \mathcal{U} \times [M] \to \mathcal{X}$, $n = 1, 2, \ldots, n_m$, defining channel inputs
$$X_n = f_n(U, W), \tag{4}$$
where $W \in [M]$ is the equiprobable message.
3) A non-negative integer-valued random stopping time $\tau \in \{n_1, n_2, \ldots, n_m\}$ of the filtration generated by $\mathcal{G}_m = \sigma\{U, Y^{n_m}\}$ that satisfies an average decoding time constraint
$$\mathbb{E}[\tau] \le l. \tag{5}$$
4) m decoding functions $g_{n_i} : \mathcal{U} \times \mathcal{Y}^{n_i} \to [M]$, providing the best estimate of W at time n_i, $i = 1, 2, \ldots, m$. The final decision $\hat{W}$ is computed at time instant τ, i.e., $\hat{W} = g_\tau(U, Y^\tau)$ and must satisfy
$$P_e \triangleq \mathbb{P}[\hat{W} \ne W] \le \epsilon. \tag{6}$$

The rate of a VLSF code is given by $R \triangleq \log M / \mathbb{E}[\tau]$. In the above definition, the cardinality $\mathcal{U}$ specifies the number of deterministic codes under consideration to construct the random code. In [8, Appendix D], Yavas *et al.* showed that $|\mathcal{U}| \le 2$ suffices.

III. Tight Approximations on $\mathbb{P}[\imath(X^n;Y^n) \ge \gamma]$

In the analysis of (l, n_1^m, M, ϵ) VLSF codes, a key step is to develop a differentiable function $F_\gamma(n)$ to approximate or to bound the tail probability $\mathbb{P}[\imath(X^n;Y^n) \ge \gamma]$ with a fixed γ. In [12]–[15], $\mathbb{P}[\imath(X^n;Y^n) \ge \gamma]$ is approximated as a Gaussian tail probability, e.g., $Q\left(\frac{\gamma - nC}{\sqrt{nV}}\right)$ used in [13]. However, we will show that for short blocklength n, the Gaussian model is imprecise and a better approximation is desired.

In probability theory, the Edgeworth expansion [19] has been known as a powerful tool to approximate the distribution of the sum of n i.i.d. random variables. In this paper, we apply the order-s Edgeworth expansion to approximate $\mathbb{P}[\imath(X^n;Y^n) \ge \gamma]$ for moderate and large values of n. We refer the reader to [20, Chapter 2] for a detailed introduction.

Theorem 1 (Equation (2.18), [20]). *Let $W_1, W_2, \ldots, W_n$ be a sequence of i.i.d. random variables with zero mean*

and a finite variance σ^2. Define $G_n(x) \triangleq \mathbb{P}[\sum_{i=1}^n W_i \leq x\sigma\sqrt{n}]$. Let $\chi_W(t) \triangleq \mathbb{E}[e^{itW}]$ be the characteristic function of W. If $\mathbb{E}[|W|^{s+2}] < \infty$ for some $s \in \mathbb{Z}^+$ and $\limsup_{|t|\to\infty} |\chi_W(t)| < 1$ (known as Cramér's condition), then,

$$G_n(x) = \Phi(x) + \phi(x)\sum_{j=1}^{s} n^{-\frac{j}{2}} p_j(x) + o(n^{-s/2}), \quad (7)$$

where

$$p_j(x) = -\sum_{\{k_m\}} He_{j+2r-1}(x) \prod_{m=1}^{j} \frac{1}{k_m!}\left(\frac{\kappa_{m+2}}{(m+2)!}\right)^{k_m}, \quad (8)$$

$$He_j(x) = j! \sum_{k=0}^{\lfloor j/2 \rfloor} \frac{(-1)^k x^{j-2k}}{k!(j-2k)!2^k}, \quad (9)$$

$$\kappa_m = m! \sum_{\{k_l\}} (-1)^{r-1}(r-1)! \prod_{l=1}^{m} \frac{1}{k_l!}\left(\frac{\mathbb{E}[W^l]}{\sigma^l l!}\right)^{k_l}, \quad (10)$$

where in (8), the set $\{k_m\}$ consists of all non-negative solutions to $\sum_{m=1}^{j} m k_m = j$, $r \triangleq \sum_{m=1}^{j} k_m$. The set $\{k_l\}$ and r in (10) are defined analogously.

Remark 1. *In Theorem 1, the Cramér's condition holds if the random variable W has a proper density function. The polynomial $He_j(x)$ is known as the Hermite polynomial of degree j. κ_m denotes the order-m cumulant of random variable W/σ. (10) indicates that κ_m is a homogeneous polynomial in moments of degree m. In [21], the authors presented a proof of (9) and (10) and provided an efficient algorithm to compute the set $\{k_m\}$ in (8).*

As an application of Theorem 1, let $W = 1 - \log\left(1 + e^{-2P-2Z\sqrt{P}}\right) - C$, where $Z \sim \mathcal{N}(0,1)$. Clearly, W has a proper density function and $\mathbb{E}[|W|^{s+2}] < \infty$ holds for any $s \in \mathbb{Z}_+$. Hence, for moderate and large values of n, the differentiable function $F_\gamma(n)$ we use to approximate $\mathbb{P}[\iota(X^n;Y^n) \geq \gamma]$ is given by the order-s Edgeworth expansion, i.e.,

$$F_\gamma(n) = Q\left(\frac{\gamma - nC}{\sqrt{nV}}\right) - \phi\left(\frac{\gamma - nC}{\sqrt{nV}}\right)\sum_{j=1}^{s} n^{-\frac{j}{2}} p_j\left(\frac{\gamma - nC}{\sqrt{nV}}\right). \quad (11)$$

A caveat of using the order-s Edgeworth expansion is that for small values of n, the order-s Edgeworth expansion oscillates around 0 due to truncation of an infinite series, making it no longer a suitable approximation function to the tail probability. To remedy the situation, we resort to the Petrov expansion [22] for small n.

Theorem 2 (Theorem 1, [22])**.** *Let $W_1, W_2, \ldots, W_n$ be a sequence of i.i.d. random variables with zero mean and a finite variance σ^2. Define $G_n(x) \triangleq \mathbb{P}\left[\sum_{i=1}^n W_i \leq x\sigma\sqrt{n}\right]$. If $x \geq 0$, $x = o(\sqrt{n})$, and the moment generating function $\mathbb{E}[e^{tW}] < \infty$ for $|t| < H$ for some $H > 0$, then*

$$G_n(x) = 1 - Q(x)\exp\left\{\frac{x^3}{\sqrt{n}}\Lambda\left(\frac{x}{\sqrt{n}}\right)\right\}\left[1 + O\left(\frac{x+1}{\sqrt{n}}\right)\right], \quad (12)$$

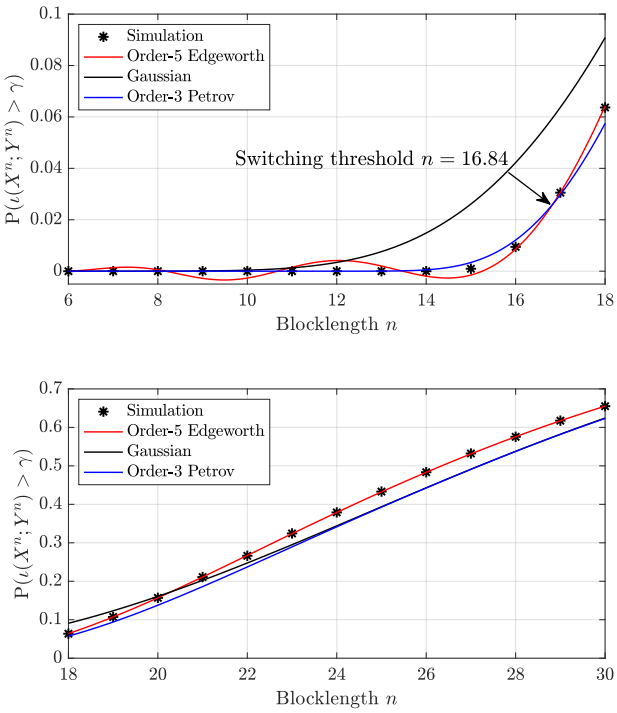

Fig. 1. Comparison of various approximation models for $\mathbb{P}[\iota(X^n;Y^n) \geq \gamma]$ with a fixed $\gamma > 0$. In this example, $k = 6$, $\epsilon = 10^{-2}$, $\gamma = \log\frac{2^k-1}{\epsilon/2} = 13.62$ for BI-AWGN channel at 0.2 dB.

$$G_n(-x) = Q(x)\exp\left\{\frac{-x^3}{\sqrt{n}}\Lambda\left(\frac{-x}{\sqrt{n}}\right)\right\}\left[1 + O\left(\frac{x+1}{\sqrt{n}}\right)\right], \quad (13)$$

where $\Lambda(t) = \sum_{k=0}^{\infty} a_k t^k$ is called the Cramér series[1] .

In [22], Petrov provided the order-2 Cramér series $\Lambda^{[2]}(t)$,

$$\Lambda^{[2]}(t) = \frac{\kappa_3}{6\kappa_2^{3/2}} + \frac{\kappa_4\kappa_2 - 3\kappa_3^2}{24\kappa_2^3}t + \frac{\kappa_5\kappa_2^2 - 10\kappa_4\kappa_3\kappa_2 + 15\kappa_3^3}{120\kappa_2^{9/2}}t^2. \quad (14)$$

For small n satisfying $n < \gamma/C$, the function $F_\gamma(n)$ we use to approximate $\mathbb{P}[\iota(X^n;Y^n) \geq \gamma]$ is given by the order-3 Petrov expansion, where the order of 3 is determined by κ_5 in (14),

$$F_\gamma(n) = Q\left(\frac{\gamma - nC}{\sqrt{nV}}\right)\exp\left\{\frac{(\gamma - nC)^3}{n^2V^{3/2}}\Lambda^{[2]}\left(\frac{\gamma - nC}{n\sqrt{V}}\right)\right\}. \quad (15)$$

In our implementation, we found that the order-5 Edgeworth expansion meets our desired approximation accuracy at large n. The switch from the order-5 Edgeworth expansion to the order-3 Petrov expansion occurs at the largest value for which two expansions are equal with a common value less than $1/2$. Fig. 1 shows the comparison of different approximation models for $\mathbb{P}[\iota(X^n;Y^n) \geq \gamma]$ with $\gamma = 13.62$ for BI-AWGN channel at 0.2 dB. The Gaussian model in Fig. 1 is given by $Q\left(\frac{\gamma - nC}{\sqrt{nV}}\right)$. Fig. 1 shows that the Gaussian model fails to capture the true tail probability at small n. The order-5

[1]Details on Cramér series can be found in the proof of [22, Theorem 2].

Edgeworth expansion oscillates around 0 when $n < 16$ and is extremely accurate when $n \geq 16$. In contrast, the order-3 Petrov expansion is loose yet close to the Gaussian model when $n \geq 24$ and becomes tight when $n \leq 14$. Therefore, the combination of the order-2 Petrov expansion and the order-5 Edgeworth expansion at switching threshold $n = 16.84$ provides a remarkably precise estimate of the tail probability $\mathbb{P}[\iota(X^n;Y^n) \geq \gamma]$.

IV. VLSF Codes With m Optimal Decoding Times

In this section, we develop numerical tools to evaluate the achievable rate of a VLSF code with m optimal decoding times. We mainly consider the error regime where Polyanskiy's scheme of stopping at zero does not improve the achievability bound [10].

A. An Integer Program and a Greedy Algorithm

In [18], Yavas *et al.* proved an achievability bound for an (l, n_1^m, M, ϵ) VLSF code for the AWGN channel. With a slight modification, this result is directly applicable to the BI-AWGN channel.

Theorem 3 (Theorem 3, [18])**.** *Fix a constant $\gamma > 0$ and decoding times $n_1 < \cdots < n_m$. For any positive numbers l and $\epsilon \in (0,1)$, there exists an (l, n_1^m, M, ϵ) VLSF code for the BI-AWGN channel* (1) *with*

$$\mathbb{E}[\tau] \leq n_1 + \sum_{i=1}^{m-1}(n_{i+1}-n_i)\mathbb{P}\left[\bigcap_{j=1}^{i}\{\iota(X^{n_j};Y^{n_j}) < \gamma\}\right], \tag{16}$$

$$P_e \leq \mathbb{P}[\iota(X^{n_m};Y^{n_m}) < \gamma] + (M-1)2^{-\gamma}, \tag{17}$$

where $P_{X^{n_m}}$ is the product of distribution of m subvectors of length $n_j - n_{j-1}$, $j \in [m]$, with the convention $n_0 = 0$. Namely,

$$P_{X^{n_m}}(x^{n_m}) = \prod_{j=1}^{m} P_{X_{n_{j-1}+1}^{n_j}}(x_{n_{j-1}+1}^{n_j}). \tag{18}$$

Remark 2. *In [18] (and its full version [23]), Yavas et al. obtained Theorem 3 by constructing a random VLSF code according to distribution* (18) *and applying an information density threshold decoder that favors the largest message index whose cumulative information density exceeds γ for the first time among any other message indices at decoding times $\{n_1, n_2, \ldots, n_m\}$.*

In (17)*, the first term upper bounds the probability that the true message never crosses γ and the second term upper bounds the probability that any other message crosses γ sooner than the true message.*

Interested readers can refer to the full version [23] of [18] for the proof of Theorem 3. For our purposes, Theorem 3 motivates the following integer program. Define

$$N(\gamma, n_1^m) \triangleq n_1 + \sum_{i=1}^{m-1}(n_{i+1}-n_i)\mathbb{P}[\iota(X^{n_i};Y^{n_i}) < \gamma], \tag{19}$$

$$\mathcal{F}_m(\gamma, M, \epsilon) \triangleq \{n_1^m : n_{i+1} - n_i \geq 1, i \in [m-1], \text{ and } \mathbb{P}[\iota(X^{n_m};Y^{n_m}) < \gamma] + (M-1)2^{-\gamma} \leq \epsilon\}. \tag{20}$$

For a given $m \in \mathbb{Z}_+, M \in \mathbb{Z}_+$, $\epsilon \in (0,1)$, and $\gamma \geq \log\frac{M-1}{\epsilon}$,

$$\begin{aligned} \min \quad & N_\gamma(n_1^m) \\ \text{s.t.} \quad & n_1^m \in \mathcal{F}_m(\gamma, M, \epsilon) \\ & n_1^m \in \mathbb{Z}_+^m. \end{aligned} \tag{21}$$

In the integer program (21), we consider the minimum gap and average error probability constraints as in (20), and the constraint that all decoding times must be integers.

Let $\tilde{N}(\gamma)$ denote the locally minimum upper bound $N(\gamma, n_1^m)$ on $\mathbb{E}[\tau]$ for a given γ in program (21). Then, $\min_\gamma \tilde{N}(\gamma)$ yields the globally minimum upper bound $N(\gamma, n_1^m)$. In this paper, we solve the globally minimum upper bound $N(\gamma, n_1^m)$ using this two-step minimization approach.

In general, an integer program is NP-complete. For the specific integer program (21), additional challenge is caused by the fact that there is no closed-form expression for $\mathbb{P}[\iota(X^{n_k};Y^{n_k}) < \gamma]$ and $n_1, n_2, \ldots, n_m$ are required to be monotonically increasing integers. While a complete solution to the integer program (21) remains open, we establish the following results.

Lemma 1. *Fix $m \in \mathbb{Z}_+, M \in \mathbb{Z}_+, \epsilon \in (0,1)$ and $\gamma \geq \frac{M-1}{\epsilon}$, and let $n_0 \triangleq 0$. Let $n_1^m \in \mathcal{F}_m(\gamma, M, \epsilon)$ be m integer-valued decoding times that achieve $N(\gamma, n_1^m)$. Suppose there exists an integer $\tilde{n}$ such that $n_{k-1} < \tilde{n} < n_k$ for some $k \in [m]$. Then,*

$$N(\gamma, n_1^{k-1}, \tilde{n}, n_k^m) < N(\gamma, n_1^m). \tag{22}$$

Proof: For brevity, let $S_n \triangleq \iota(X^n;Y^n)$. Let $n_1^m \in \mathcal{F}_m(\gamma, M, \epsilon)$ be the m integer-valued decoding times that achieve $N(\gamma, n_1^m)$. If $k = 1$, i.e., $\tilde{n} < n_1$, then (22) holds trivially. Assume that $2 \leq k \leq m$ and there exists an integer $\tilde{n}$ such that $n_{k-1} < \tilde{n} < n_k$. It is straightforward to show that

$$\begin{aligned} &N(\gamma, n_1^m) - N(\gamma, n_1^{k-1}, \tilde{n}, n_k^m) \\ &= (n_k - n_{k-1})\mathbb{P}[S_{n_{k-1}} < \gamma] - (\tilde{n} - n_{k-1})\mathbb{P}[S_{n_{k-1}} < \gamma] \\ &\quad - (n_k - \tilde{n})\mathbb{P}[S_{\tilde{n}} < \gamma] \geq 0, \end{aligned}$$

where the last step follows from $\mathbb{P}[S_{\tilde{n}} < \gamma] < \mathbb{P}[S_{n_{k-1}} < \gamma]$. ∎

Define

$$N_m^*(\gamma, M, \epsilon) \triangleq \min_{n_1^m \in \mathcal{F}_m(\gamma, M, \epsilon)} N(\gamma, n_1^m). \tag{23}$$

Theorem 4. *Fix $M \in \mathbb{Z}_+$, $\epsilon \in (0,1)$ and $\gamma \geq \log\frac{M-1}{\epsilon}$. Let $n^* = \min\{n \in \mathbb{Z}_+ : \mathbb{P}[\iota(X^n;Y^n) < \gamma] \leq \epsilon - (M-1)2^{-\gamma}\}$. For $m < n^*$, it holds that*

$$N_{m+1}^*(\gamma, M, \epsilon) < N_m^*(\gamma, M, \epsilon). \tag{24}$$

Proof: Let $n_0 \triangleq 0$. Let $n_1^m \in \mathcal{F}_m(\gamma, M, \epsilon)$ be the sequence of decoding times that achieves $N_m^*(\gamma, M, \epsilon)$. It is straightforward to show that the optimal $n_m = n^*$. If $m < n^*$, this implies that there exists some $k \in [m]$ such that $n_k - n_{k-1} \geq 2$. For this k, choose an integer $\tilde{n}$ such that $n_{k-1} < \tilde{n} < n_k$. Clearly, $(n_1^{k-1}, \tilde{n}, n_k^m) \in \mathcal{F}_{m+1}(\gamma, M, \epsilon)$. Therefore,

$$N_{m+1}^*(\gamma, M, \epsilon) \leq N(\gamma, n_1^{k-1}, \tilde{n}, n_k^m) < N_m^*(\gamma, M, \epsilon), \tag{25}$$

where the last inequality in (25) follows from Lemma 1. ∎

Theorem 4 motivates the following greedy algorithm for a fixed γ: Start from $m = n^*$ where $n^* \triangleq \min\{n \in \mathbb{Z}_+ : \mathbb{P}[\iota(X^n;Y^n) < \gamma] \le \epsilon - (M-1)2^{-\gamma}\}$. Suppose that n_1^m is the solution for m. Then, the solution $\tilde{n}_1^{m-1}$ for $m-1$ is identified by removing the decoding time n_i in n_1^{m-1} that minimizes $N_\gamma(n_1^{i-1}, n_{i+1}^m)$. Note that the decoding time n_m is always retained to ensure that the target error probability is met via (17).

B. The Relaxed Program and the Gap-Constrained SDO

To facilitate a program that is computationally tractable, we consider the relaxed program that allows $n_1^m \in \mathbb{R}_+^m$: For a given $m \in \mathbb{Z}_+, M \in \mathbb{Z}_+$, $\epsilon \in (0,1)$ and $\gamma \ge \log \frac{M-1}{\epsilon}$,

$$\begin{aligned} \min \quad & N(\gamma, n_1^m) \\ \text{s.t.} \quad & n_1^m \in \mathcal{F}_m(\gamma, M, \epsilon), \end{aligned} \tag{26}$$

where the tail probability $\mathbb{P}[\iota(X^n;Y^n) \ge \gamma]$ is approximated by a monotonically increasing and differentiable function $F_\gamma(n)$ with $F_\gamma(0) = 0$ and $F_\gamma(\infty) = 1$, for instance, the piecewise function[2] introduced in Section III. Let

$$f_\gamma(n) \triangleq \frac{dF_\gamma(n)}{dn}. \tag{27}$$

For the relaxed program (26) with a fixed γ, the optimal, real-valued decoding times $n_1^*, n_2^*, \ldots, n_m^*$ are given by the following theorem.

Theorem 5. *For a given $m \in \mathbb{Z}_+, M \in \mathbb{Z}_+$, $\epsilon \in (0,1)$ and $\gamma \ge \log \frac{M-1}{\epsilon}$, the optimal real-valued decoding times $n_1^*, n_2^*, \ldots, n_m^*$ in program* (26) *satisfy*

$$n_m^* = F_\gamma^{-1}\left(1 - \epsilon + (M-1)2^{-\gamma}\right), \tag{28}$$

$$n_{k+1}^* = n_k^* + \max\left\{1, \frac{F_\gamma(n_k^*) - F_\gamma(n_{k-1}^*) - \lambda_{k-1}}{f_\gamma(n_k^*)}\right\}, \tag{29}$$

$$\lambda_k = \max\{\lambda_{k-1} + f_\gamma(n_k^*) - F_\gamma(n_k^*) + F_\gamma(n_{k-1}^*), 0\}, \tag{30}$$

where $k \in [m-1]$, $\lambda_0 \triangleq 0$ and $n_0^* \triangleq 0$.

Proof: For brevity, define $\boldsymbol{n} \triangleq (n_1, n_2, \ldots, n_m)$. By introducing the Lagrangian multipliers ν, λ_1^{m-1}, the Lagrangian of program (26) is given by

$$\begin{aligned} \mathcal{L}(\boldsymbol{n}, \nu, \lambda_1^{m-1}) &= n_1 + \nu(1 - F_\gamma(n_m) + (M-1)2^{-\gamma} - \epsilon) \\ &+ \sum_{i=1}^{m-1}(n_{i+1} - n_i)(1 - F_\gamma(n_i)) + \sum_{i=1}^{m-1}\lambda_i(n_i - n_{i+1} + 1). \end{aligned}$$

[2]The first derivative of $F_\gamma(n)$ at the switching threshold does not exist. Nonetheless, one can assign the right (or left) derivative as the derivative for the switching threshold so that the solution is not affected significantly.

By the Karush-Kuhn-Tucker (KKT) conditions, the optimal decoding times $\boldsymbol{n}^* = (n_1^*, n_2^*, \ldots, n_m^*)$ must satisfy

$$\frac{\partial \mathcal{L}}{\partial n_k}\Big|_{\boldsymbol{n}=\boldsymbol{n}^*} = F_\gamma(n_k^*) - F_\gamma(n_{k-1}^*) - (n_{k+1}^* - n_k^*)f_\gamma(n_k^*) + \lambda_k - \lambda_{k-1} = 0, \quad k \in [m-1], \tag{31}$$

$$\frac{\partial \mathcal{L}}{\partial n_m}\Big|_{\boldsymbol{n}=\boldsymbol{n}^*} = 1 - F_\gamma(n_{m-1}^*) - \nu f_\gamma(n_m^*) = 0, \tag{32}$$

$$\nu(1 - F_\gamma(n_m^*) + (M-1)2^{-\gamma} - \epsilon) = 0, \tag{33}$$

$$\lambda_k(n_k^* - n_{k+1}^* + 1) = 0, \quad k \in [m-1]. \tag{34}$$

Since $F_\gamma(n) \in (0,1)$ and $f_\gamma(n) > 0$ for $n > 0$, (32) indicates that $\nu > 0$. Hence, we obtain $n_m^* = F_\gamma^{-1}(1 - \epsilon + (M-1)2^{-\gamma})$ from (33).

Next, we analyze (34). There are two cases. If $\lambda_k > 0$, then $n_{k+1}^* = n_k^* + 1$. By (31), we obtain

$$\lambda_k = \lambda_{k-1} + f_\gamma(n_k^*) - F_\gamma(n_k^*) + F_\gamma(n_{k-1}^*). \tag{35}$$

If $n_{k+1}^* > n_k^* + 1$, then $\lambda_k = 0$. By (31), we obtain

$$n_{k+1}^* = n_k^* + \frac{F_\gamma(n_k^*) - F_\gamma(n_{k-1}^*) - \lambda_{k-1}}{f_\gamma(n_k^*)}. \tag{36}$$

Rewriting the above two cases in a compact form yields (29) and (30). ∎

The procedures (29) and (30) are called the *gap-constrained SDO* for the relaxed program (26). In contrast, the SDO studied in [12]–[17] is derived from the relaxed program (26) without the gap constraint[3] and admits a simple recursion

$$n_{k+1}^* = n_k^* + \frac{F_\gamma(n_k^*) - F_\gamma(n_{k-1}^*)}{f_\gamma(n_k^*)}, \quad k \in [m-1], \tag{37}$$

where $n_0^* \triangleq 0$. We will show that for small values of m, the gap-constrained SDO behaves indistinguishably as the SDO without the gap constraint in (37). However, as m becomes large, the decoding times provided by these two algorithms differ noticeably.

In practice, after solving n_m^* via (28), one would apply a bisection search between 0.5 and $\lceil n_m^* \rceil - m + 0.5$ for n_1 and the SDO to identify n_1^*. This guarantees that the nearest integer to n_1^* is at least 1.

When evaluating at small n, both $F_\gamma(n)$ and $f_\gamma(n)$ will become infinitesimally small. In this case, a direct numerical computation using (29) and (30) may cause the precision issue. Fortunately, the SDO described by (29) and (30) also admits a ratio form. Define $\lambda_k^{(r)} \triangleq \lambda_k / f_\gamma(n_k^*)$. Thus, (29) and (30) are equivalent to

$$n_{k+1}^* = n_k^* + \max\left\{1, \frac{F_\gamma(n_k^*)}{f_\gamma(n_k^*)} - \frac{F_\gamma(n_{k-1}^*)}{f_\gamma(n_k^*)} - \lambda_{k-1}^{(r)}\frac{f_\gamma(n_{k-1}^*)}{f_\gamma(n_k^*)}\right\},$$

$$\lambda_k^{(r)} = \max\left\{\lambda_{k-1}^{(r)}\frac{f_\gamma(n_{k-1}^*)}{f_\gamma(n_k)^*} + 1 - \frac{F_\gamma^*(n_k)}{f_\gamma^*(n_k)} + \frac{F_\gamma^*(n_{k-1})}{f_\gamma^*(n_k)}, 0\right\}.$$

The purpose of using $F_\gamma(\tilde{n})/f_\gamma(n)$, $f_\gamma(\tilde{n})/f_\gamma(n)$, and $\lambda_k^{(r)}$ is that they have a closed-form expression that cancels out the common infinitesimal factor in both the numerator and

[3]The error probability constraint is also different, yet it does not affect the SDO procedure.

denominator. In our implementation, we applied the ratio form of SDO.

C. Error Regime Where Stopping at Zero Does Not Help

In [10], Polyanskiy *et al.* demonstrated that the VLSF code with infinitely many stopping times can achieve $\frac{C}{1-\epsilon}$. This is accomplished by the following scheme: With probability $p = \frac{\epsilon-\epsilon'}{1-\epsilon'}$, the code immediately stops at $\tau = 0$ without any channel use, and with probability $1-p$, employs an (l', M, ϵ') VLSF code satisfying $\log M = Cl' + \log \epsilon' - a_0$, where $a_0 \triangleq \sup_{x,y} \iota(x;y)$. The overall code has an error probability

$$1 \cdot p + \epsilon'(1-p) = \epsilon, \tag{38}$$

and average blocklength

$$0 \cdot p + l'(1-p) = l'(1-p). \tag{39}$$

In this section, we identify the error regime where Polyanskiy's scheme of stopping at $\tau = 0$ does not improve the achievability bound.

Theorem 6. *For a given $a_0 \in \mathbb{R}_+$, $M \in \mathbb{Z}_+$, define*

$$\epsilon^* \triangleq \arg\min_{x\in(0,1)} \frac{\log M + a_0 - \log x}{1-x}. \tag{40}$$

If $\epsilon \in (0, \epsilon^*]$, stopping at $\tau = 0$ does not improve the achievability bound for VLSF codes.

Proof: By Polyanskiy's scheme, solving the error regime where stopping at zero does not improve the achievability bound is equivalent to identifying the error regime in which $\epsilon' = \epsilon$ is the minimizer to the following program: For a given $C, a_0 \in \mathbb{R}_+$, $M \in \mathbb{Z}_+$, and $\epsilon \in (0,1)$,

$$\begin{aligned} \min_{\epsilon'} \quad & l'(1-p) \\ \text{s.t.} \quad & \log M = Cl' + \log \epsilon' - a_0 \\ & p = \frac{\epsilon - \epsilon'}{1-\epsilon'} \\ & \epsilon' \in (0, \epsilon]. \end{aligned} \tag{41}$$

The program (41) is equivalent to the following program

$$\begin{aligned} \min_{\epsilon'} \quad & \left(\frac{1-\epsilon}{C}\right) f(\epsilon') \\ \text{s.t.} \quad & \epsilon' \in (0, \epsilon], \end{aligned} \tag{42}$$

where

$$f(x) \triangleq \frac{\log M + a_0 - \log x}{1-x}. \tag{43}$$

Since $f(x)$ is convex in $(0,1)$, there exists a unique minimizer $\epsilon^* \in (0,1)$. Therefore, if $\epsilon \le \epsilon^*$, then $\epsilon' = \epsilon$ minimizes the objective function in (42), giving $p = 0$. Namely, stopping at zero does not improve the achievability bound. ∎

We remark that there is no closed-form solution to ϵ^* in Theorem 6. Nonetheless, one can numerically solve ϵ^* for a given M and a_0.

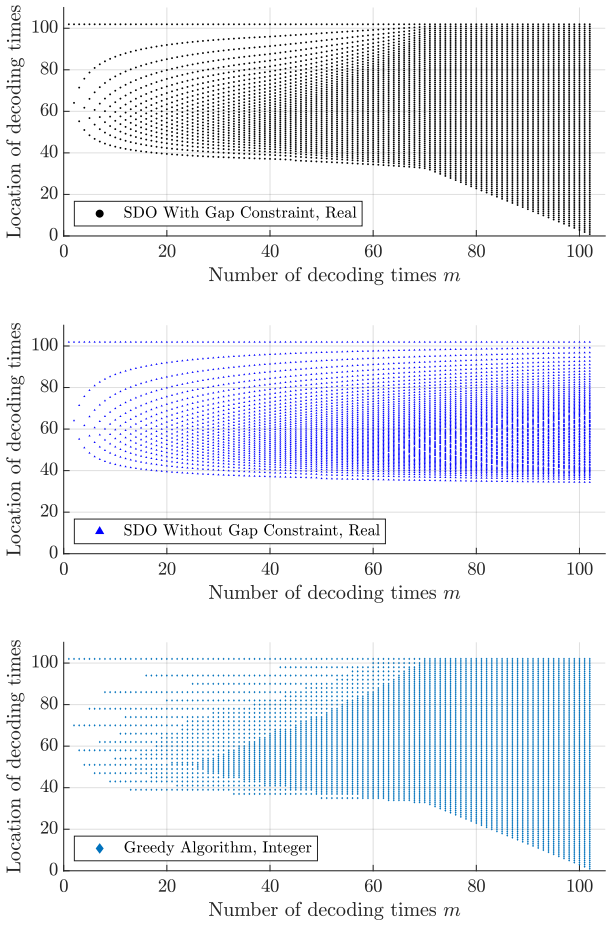

Fig. 2. Comparison of the real-valued decoding times by the SDO with gap constraint, the real-valued decoding times by the SDO without gap constraint, and the integer-valued decoding times by greedy algorithm for $k = 20$, $\epsilon = 10^{-2}$, $\delta = 1/2$, $\gamma = \log \frac{2^k-1}{\delta\epsilon}$ and BI-AWGN channel at 0.2 dB. m ranges from 1 to $\lceil n_m^* \rceil = 102$, where $n_m^* = 101.91$ is given by (28).

D. Numerical Evaluation

Let $M = 2^k$, $k \in \mathbb{Z}_+$. We consider the BI-AWGN channel at 0.2 dB with a capacity of 0.5 and the error regime in which stopping at zero does not improve the achievability bound. By Theorem 6, if $k \le 100$, $\epsilon \le 1.33 \cdot 10^{-2}$ is the error regime where stopping at zero does not help. In the following example, we consider $\epsilon = 10^{-2}$.

We consider the relaxed program and apply the two-step minimization with the gap-constrained SDO introduced in Section IV to obtain the globally minimum upper bound $N^*(\gamma, n_1^m)$. Thus, $k/N^*(\gamma, n_1^m)$ gives the achievability bound. In [10], Polyanskiy *et al.* showed that the average blocklength $\mathbb{E}[\tau]$ of a VLSF code with $m = \infty$ and no stopping at $\tau = 0$ is upper bounded by

$$\mathbb{E}[\tau] \le \frac{\log \frac{M-1}{\epsilon} + a_0}{C}, \tag{44}$$

where $a_0 \triangleq \sup_{x,y} \iota(x;y)$. This bound yields the *VLSF achievability bound* on rate. For BI-AWGN channel, $a_0 = 1$.

For a fixed γ at $k = 20$ and $\epsilon = 10^{-2}$, Fig. 2 shows how the decoding times evolve with m for the three algorithms: SDO with/without the gap constraint, and the greedy algorithm. For

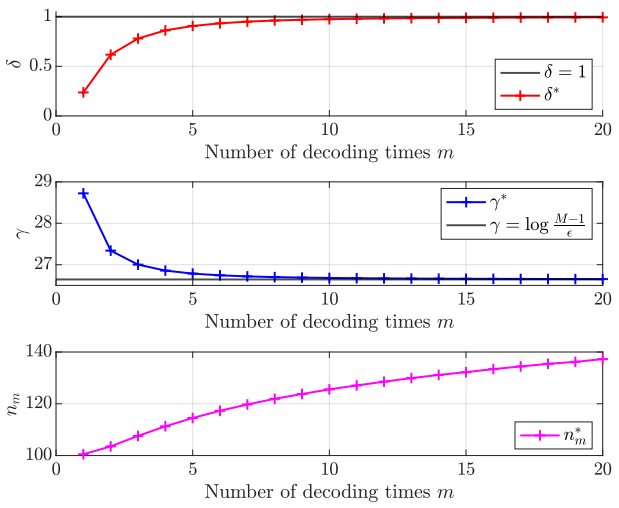

Fig. 3. Globally optimal δ^*, γ^*, and n_m^* as a function of the number of decoding times m. In the VLSF achievability bound, $\delta = 1$ and $\gamma = \log \frac{M-1}{\epsilon}$. In this example, $\epsilon = 10^{-2}$, $k = 20$ for BI-AWGN channel at 0.2 dB.

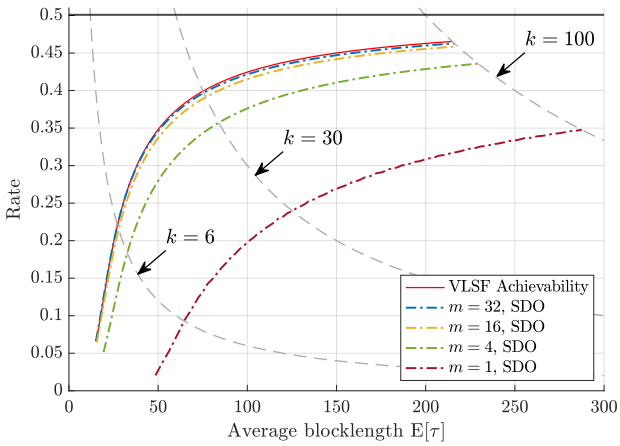

Fig. 4. Comparison of achievable rate estimation by the gap-constrained SDO and by the greedy algorithm for VLSF codes with m optimal decoding times. The gray dashed line represents the $(\mathbb{E}[\tau], R)$ pairs such that $R\mathbb{E}[\tau] = k$. In this example, $\epsilon = 10^{-2}$ and the BI-AWGN channel is at 0.2 dB.

$m \leq 20$, SDO with a gap constraint behaves indistinguishably as the SDO without a gap constraint since the SDO solution naturally has gaps larger than one. The greedy algorithm is forced to choose from the remaining decoding times, leading to a possibly suboptimal solution. For large m, SDO without the gap constraint avoids early decoding times and instead adds later decoding times so densely that their separation is less than one. In contrast, SDO with the gap constraint is forced to add early decoding times when all existing gaps become one.

The greedy algorithm lacks the optimality guarantee of the gap-constrained SDO and is computationally more intensive. Despite their distinct design perspectives, the greedy algorithm and the gap-constrained SDO arrive at essentially the same solution for large m. For $k = 20$, $\epsilon = 10^{-2}$, and $\gamma = \log \frac{(2^k-1)}{\epsilon/2}$, Fig. 2 shows that n_1 is never less than 37 when $m \leq 32$ and grows as the number of decoding times decreases.

We remark that Fig. 2 assumes a constant γ over all number of decoding times. However, in the two-step minimization with the gap-constrained SDO, the globally optimal γ^* is a function of m. Therefore, the globally optimal n_m^* may not stay as constant as shown in Fig. 2.

Let $\delta \in [0, 1]$ and assume that the first and second terms in the right-hand side of (17) are equal to $(1-\delta)\epsilon$ and $\delta\epsilon$, respectively. Then, both γ^* and n_m^* can be thought as a function of δ^*, i.e.,

$$\gamma^*(\delta^*) = \log \frac{M-1}{\epsilon\delta^*}, \tag{45}$$

$$n_m^*(\delta^*) = F_{\gamma^*}^{-1}(1 - \epsilon + \epsilon\delta^*). \tag{46}$$

Thus, minimization over γ is equivalent to minimization over δ. For $k = 20$, $\epsilon = 10^{-2}$, and BI-AWGN channel at 0.2 dB, Fig. 3 shows how the globally optimal δ^* and the associated globally optimal γ^*, n_m^* vary with m during the two-step minimization. We see that when m is small, δ^* is far from 1, indicating a large value of γ^* and a small value of n_m^*. As m gets large, we observe that δ^* monotonically increases, which, by (45) and (46), implies that γ^* decreases and n_m^* increases. In particular, as $m \to \infty$, $\delta^* \to 1$, and consequently,

$$\lim_{\delta^* \to 1} \gamma^*(\delta^*) = \log \frac{M-1}{\epsilon}, \tag{47}$$

$$\lim_{\delta^* \to 1} n_m^*(\delta^*) = \infty. \tag{48}$$

In Polyanskiy's setting [10], the first term in (17) is zero since $n_m = \infty$ and the optimal γ can thus be computed as $\gamma = \log \frac{M-1}{\epsilon}$ from (17), implying that $\delta = 1$. Fig. 3 shows that as m increases, δ^*, γ^*, and n_m^* rapidly approach those in Polyanskiy's setting.

Fig. 4 shows the achievable rate of a VLSF code with m optimal decoding times estimated by the two-step minimization with the gap-constrained SDO algorithm. We see that a finite m suffices to achieve Polyanskiy's VLSF achievability bound derived from VLSF codes with infinitely many decoding times. For instance, for the BI-AWGN channel at 0.2 dB and $\epsilon = 10^{-2}$, the achievable rate estimated by the SDO for VLSF codes with $k \leq 6$ and $m = 32$ beats the VLSF bound. Additionally, for BI-AWGN channel at 0.2 dB, with 16 decoding times, the achievable rate by SDO is within 0.66% of the VLSF achievability bound for $k \leq 100$. With 32 decoding times, it becomes hard to distinguish the achievable rate by SDO from the VLSF achievability bound for $k \leq 30$.

V. Conclusion

This paper provides a new SDO that includes the gap constraint. Using this improved SDO, the paper demonstrates that Polyanskiy's VLSF achievability bound with infinitely many decoding times can be closely approached with a finite, and relatively small, number of decoding times.

References

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

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

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

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

[5] 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.

[6] M. Luby, "LT codes," in *The 43rd Annual IEEE Symp. Foundations Comp. Sci., 2002. Proc.*, Nov. 2002, pp. 271–280.

[7] S. Draper, B. Frey, and F. Kschischang, "Efficient variable length channel coding for unknown DMCs," in *2004 IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2004, pp. 379–379.

[8] R. C. Yavas, V. Kostina, and M. Effros, "Random access channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 67, no. 4, pp. 2115–2140, Apr. 2021.

[9] A. B. Wagner, N. V. Shende, and Y. Altuğ, "A new method for employing feedback to improve coding performance," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6660–6681, Nov. 2020.

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

[11] S. H. Kim, D. K. Sung, and T. Le-Ngoc, "Variable-length feedback codes under a strict delay constraint," *IEEE Commun. Lett.*, vol. 19, no. 4, pp. 513–516, Apr. 2015.

[12] K. Vakilinia, S. V. S. Ranganathan, D. Divsalar, and R. D. Wesel, "Optimizing transmission lengths for limited feedback with nonbinary LDPC examples," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2245–2257, Jun. 2016.

[13] H. Wang, N. Wong, A. M. Baldauf, C. K. Bachelor, S. V. S. Ranganathan, D. Divsalar, and R. D. Wesel, "An information density approach to analyzing and optimizing incremental redundancy with feedback," in *2017 IEEE Int. Sym. Inf. Theory (ISIT)*, Jun. 2017, pp. 261–265.

[14] R. D. Wesel, N. Wong, A. Baldauf, A. Belhouchat, A. Heidarzadeh, and J. F. Chamberland, "Transmission lengths that maximize throughput of variable-length coding & ACK/NACK feedback," in *2018 IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[15] N. Wong, K. Vakilinia, H. Wang, S. V. S. Ranganathan, and R. D. Wesel, "Sequential differential optimization of incremental redundancy transmission lengths: An example with tail-biting convolutional codes," in *2017 Inf. Theory and App. Workshop (ITA)*, Feb. 2017, pp. 1–5.

[16] A. Heidarzadeh, J.-F. Chamberland, P. Parag, and R. D. Wesel, "A systematic approach to incremental redundancy over erasure channels," in *2018 IEEE Int. Sym. Inf. Theory (ISIT)*, Jun. 2018, pp. 1176–1180.

[17] A. Heidarzadeh, J.-F. Chamberland, R. D. Wesel, and P. Parag, "A systematic approach to incremental redundancy with application to erasure channels," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 2620–2631, Apr. 2019.

[18] R. C. Yavas, V. Kostina, and M. Effros, "Variable-length feedback codes with several decoding times for the Gaussian channel," in *2021 IEEE Int. Sym. Inf. Theory (ISIT)*, Jul. 2021, pp. 1883–1888.

[19] F. Y. Edgeworth, "The law of error," *Cambridge Philos. Trans.*, vol. 20, pp. 36–66 and 113–141., 1905.

[20] P. Hall, *The Bootstrap and Edgeworth Expansion*. New York, NY, USA: Springer, 1992.

[21] Blinnikov, S. and Moessner, R., "Expansions for nearly Gaussian distributions," *Astron. Astrophys. Suppl. Ser.*, vol. 130, no. 1, pp. 193–205, 1998.

[22] V. V. Petrov, *Sums of independent random variables*. New York, NY, USA: Springer, Berlin, Heidelberg, 1975.

[23] R. C. Yavas, V. Kostina, and M. Effros, "Variable-length feedback codes with several decoding times for the Gaussian channel," Mar. 2021. [Online]. Available: http://arxiv.org/abs/2103.09373