Variable-length Feedback Codes with Several Decoding Times for the Gaussian Channel

Recep Can Yavas, Victoria Kostina, and Michelle Effros

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9517993

***Abstract*—We investigate variable-length feedback (VLF) codes for the Gaussian point-to-point channel under maximal power, average error probability, and average decoding time constraints. Our proposed strategy chooses $K < \infty$ decoding times $n_1, n_2, \dots, n_K$ rather than allowing decoding at any time $n = 0, 1, 2, \dots$. We consider stop-feedback, which is one-bit feedback transmitted from the receiver to the transmitter at times $n_1, n_2, \dots$ only to inform her whether to stop. We prove an achievability bound for VLF codes with the asymptotic approximation $\ln M \approx \frac{NC(P)}{1-\epsilon} - \sqrt{N \ln_{(K-1)}(N) \frac{V(P)}{1-\epsilon}}$, where $\ln_{(K)}(\cdot)$ denotes the K-fold nested logarithm function, N is the average decoding time, and $C(P)$ and $V(P)$ are the capacity and dispersion of the Gaussian channel, respectively. Our achievability bound evaluates a non-asymptotic bound and optimizes the decoding times $n_1, \dots, n_K$ within our code architecture.**

***Index Terms*—Variable-length stop-feedback codes, Gaussian channel, second-order achievability bound.**

For a full version of this paper, see arXiv:2103.09373.

I. Introduction

Although Shannon's work [1] shows that feedback does not increase the capacity of memoryless, point-to-point channels, it is known that feedback has several important benefits in channel coding such as simplifying coding schemes and improving higher-order achievable rates. Several results demonstrate this effect in the fixed-length regime. Feedback simplifies coding in Horstein's scheme for the binary symmetric channel [2] and Schalkwijk and Kailath's scheme [3] for the Gaussian channel. Wagner *et al.* [4] show that feedback improves the second-order achievable rate for any discrete memoryless channel (DMC) with multiple capacity-achieving input distributions giving distinct dispersions.

The benefits of feedback increase for codes with arbitrary decoding times (called variable-length or rateless codes). In [5], Burnashev shows that feedback significantly improves the optimal error exponent of variable-length codes for DMCs. In [6], Polyanskiy *et al.* extend the work of Burnashev to the finite-length regime with non-vanishing error probabilities, introducing variable-length feedback (VLF) and variable-length feedback with termination (VLFT) codes and deriving achievability and converse bounds for their performance. In VLF codes, the receiver decides when to stop transmissions and decode; in VLFT codes, the transmitter decides when to stop transmissions using its knowledge of the source message and the feedback that it receives from the receiver. As a special case of VLF codes, Polyanskiy *et al.* define variable-length stop-feedback (VLSF) codes that use feedback from the receiver to the transmitter at each potential decoding time to inform the transmitter whether to stop transmitting; in this regime, the codewords are fixed when the code is designed, and feedback affects how much of a codeword is sent but not that codeword's value. The result in [6, Th. 2] shows that variable-length decoding improves the first-order term in the asymptotic expansion of the maximum achievable message size from NC to $\frac{NC}{1-\epsilon}$, where C is the capacity of the DMC, N is the average decoding time, and ϵ is the average error probability. The second-order achievable term within the class of VLF and VLFT codes is $O(\ln N)$, which means that VLF and VLFT codes have zero dispersion, and the convergence to the capacity is much faster than that achieved by the fixed-length codes [7], [8].

Variations of VLSF and VLFT codes are studied in [9]–[16]. In [9], Kim *et al.* consider VLSF codes where decoding must occur at a decoding time less than or equal to some constant ℓ and the decoding times satisfy $n_i = id$ for some $d \in \mathbb{Z}^+$. In [10], Altuğ *et al.* modify the VLSF coding paradigm by replacing the average decoding time constraint with a constraint on the probability that the decoding time exceeds a target value; the benefit in the first-order term does not appear under this probabilistic delay constraint [10]. Truong and Tan [11], [12] extend the results in [6] to the Gaussian point-to-point and multiple access channels under an average power constraint. Trillingsgaard *et al.* [13] study the VLSF scenario where a common message is transmitted across a K-user discrete memoryless broadcast channel.

For the Gaussian channel, the effect of feedback depends both on whether the power constraint limits maximum or average power[1] and on whether the code is fixed-length or variable-length. For example, in the fixed-length regime, feedback does not improve the code performance in its first-, second-, or third-order terms under the maximal power constraint [8], but does improve the achievable second-order term under the average power constraint [17]. In the variable-length regime where the decoder can decode at any time and an average power constraint is employed, feedback improves performance in the first-order term [12].

While high rates of feedback are impractical for many applications — especially wireless applications on half-duplex

R. C. Yavas, V. Kostina, and M. Effros are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA (e-mail: ryavas, vkostina, effros@caltech.edu). This work was supported in part by the National Science Foundation (NSF) under grant CCF-1817241 and CCF-1956386.

[1]The maximal power constraint is also called the short-term, per-codeword, or peak power constraint; the average power constraint is also known as the long-term or expected power constraint.

devices — most prior work on VLSF codes [9]–[13] allows an unbounded number of potential decoding times, i.e., $K = \infty$. A notable exception is [14], which considers VLF and VLFT codes with $K < \infty$ decoding times. In [14], Vakilinia *et al.* introduce a sequential differential optimization (SDO) algorithm to optimize the choices of K potential decoding times $n_1, \ldots, n_K$, approximating the distribution of the random decoding time τ by a Gaussian random variable. They apply the SDO algorithm to non-binary low-density parity-check codes over binary-input additive white Gaussian noise channels, determining the mean and variance of τ through simulation. Extensions of [14] include [15], which uses a new reliability-output Viterbi algorithm at the decoder, and [16], which extends [14] to account for the feedback rate and applies the SDO algorithm to random linear codes over the binary erasure channel. As noted in [14]–[16], VLSF codes can be viewed as incremental redundancy hybrid automatic repeat request (IR-HARQ) schemes.

This paper presents the first asymptotic expansion for an achievable rate bound on VLSF codes with K falling between the extremes: $K = 1$ (the fixed-length regime analyzed in [7], [18]), and $K = \infty$ (the classical variable-length regime defined in [6, Def. 1]). We consider VLSF codes over the Gaussian point-to-point channel and limit the number of decoding times to some finite integer K that does not grow with the average decoding time N. We impose a new maximal power constraint, bounding the power of codewords at every potential decoding time. The feedback rate of our code is $\frac{k}{n_k}$ when the decoding time is n_k. Thus our feedback rate approaches 0 as n_k grows while most other VLSF codes use feedback rate 1 bit per channel use. Throughout the paper, we employ the average error and average decoding time criteria. Our main result shows that for VLSF codes with $2 \leq K < \infty$ decoding times, message set size M satisfying

$$\ln M \approx \frac{NC(P)}{1-\epsilon} - \sqrt{N \ln_{(K-1)}(N) \frac{V(P)}{1-\epsilon}} \tag{1}$$

is achievable. Here $\ln_{(K)}(\cdot)$ denotes the K-fold nested logarithm function, N is the average decoding time, and $C(P)$ and $V(P)$ are the capacity and dispersion of the Gaussian channel, respectively. The order of the second-order term in (1) depends on K. The convergence to $\frac{C(P)}{1-\epsilon}$ in (1) is slower than the convergence to $C(P)$ in the fixed-length scenario, which has second-order term $O(\sqrt{N})$ [7]. The $K = 2$ case in (1) recovers the variable-length scenario without feedback, which has second-order term $O(\sqrt{N \ln N})$ [6, Th. 1], achieved with $n_1 = 0$; our bound in [19, Th. 4] shows that when $K = \infty$, the asymptotic approximation in (1) is achievable with the second-order term replaced by $-O(\sqrt{N})$. Our result in (1) demonstrates how the performance of VLSF codes interpolates between these two extremes. To show that the decoding times $n_1, \ldots, n_K$ that achieve (1) are chosen optimally within our code structure, we use the SDO algorithm introduced in [14] (see [19, Sections IV-C, IV-E]). Despite the order-wise dependence on K, (1) grows so slowly with K that it suggests little benefit to choosing a large K. For example, when $K = 4$, $\sqrt{N \ln_{(K-1)}(N)}$ behaves very similarly to $O(\sqrt{N})$ for practical values of N (e.g., $N \in [10^3, 10^5]$). Notice, however, that the given achievability result provides a *lower* bound on the benefit of increasing K; bounding the benefit from above would require a converse, a topic left to future work.

In what follows, Section II introduces VLSF codes with K decoding times, and Section III presents our main results and discusses their implications. Proof sketches appear in Section IV, with details available in [19].

II. Problem Statement

A. Notation

For any positive integers k and n, $[k] \triangleq \{1, \ldots, k\}$ and $x^n \triangleq (x_1, \ldots, x_n)$. For any x^n and integers $n_1 \leq n_2 \leq n$, $x_{n_1}^{n_2} \triangleq (x_{n_1}, x_{n_1+1}, \ldots, x_{n_2})$. The Euclidean norm of vector x^n is denoted by $\|x^n\| \triangleq \sqrt{\sum_{i=1}^n x_i^2}$. We use $\ln(\cdot)$ to denote the natural logarithm. We measure information in nats. We use the standard $O(\cdot)$ and $o(\cdot)$ notations, i.e., $f(n) = O(g(n))$ if $\limsup_{n\to\infty} |f(n)/g(n)| < \infty$ and $f(n) = o(g(n))$ if $\lim_{n\to\infty} |f(n)/g(n)| = 0$. We denote the distribution of a random variable X by P_X, and we write $\mathcal{N}(\mu, \sigma^2)$ to denote the univariate Gaussian distribution with mean μ and variance σ^2. We use $Q(\cdot)$ to represent the complementary Gaussian cumulative distribution function $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left\{-\frac{t^2}{2}\right\} dt$ and $Q^{-1}(\cdot)$ to represent its functional inverse.

The k-fold nested logarithm function is defined as

$$\ln_{(k)}(x) \triangleq \begin{cases} \ln(x) & \text{if } k = 1,\ x > 0 \\ \ln(\ln_{(k-1)}(x)) & \text{if } k > 1,\ \ln_{(k-1)}(x) > 0, \end{cases} \tag{2}$$

and undefined elsewhere.

B. Channel Model

The output of a memoryless, point-to-point Gaussian channel in response to input $X^n \in \mathbb{R}^n$ is

$$Y^n = X^n + Z^n, \tag{3}$$

where $Z_1, \ldots, Z_n$ are $\mathcal{N}(0, 1)$ random variables independent of X^n and of each other.

The channel's capacity and dispersion are

$$C(P) \triangleq \frac{1}{2} \ln(1 + P) \tag{4}$$

$$V(P) \triangleq \frac{P(P+2)}{2(1+P)^2}, \tag{5}$$

respectively. The information density of a channel $P_{Y^n|X^n}$ under input distribution P_{X^n} is defined as

$$\imath(x^n; y^n) \triangleq \ln \frac{P_{Y^n|X^n}(y^n|x^n)}{P_{Y^n}(y^n)}, \tag{6}$$

where P_{Y^n} is the marginal of $P_{X^n} P_{Y^n|X^n}$.

C. VLSF Codes with K Decoding Times

We consider VLSF codes with a finite number of potential decoding times $n_1 < n_2 < \cdots < n_K$. The receiver chooses to end the transmission at the first time $n_k \in \{n_1, \ldots, n_K\}$ that it is ready to decode. The transmitter learns of the receiver's decision via a single bit of feedback at each of times $n_1, \ldots, n_k$. Feedback bit "0" at time n_i means that the receiver is not yet ready to decode and the transmitter should continue; feedback bit "1" means that the receiver can decode at time n_i and the transmitter must stop. We impose a maximal power constraint at each possible decoding time and employ average decoding time and average error probability constraints. The definition, below, formalizes our code.

Definition 1: Fix $\epsilon \in (0, 1)$, positive scalars N and P, and non-negative integers $n_1 < \ldots < n_K$ and M. An $(N, \{n_i\}_{i=1}^K, M, \epsilon, P)$ VLSF code comprises

1) a finite alphabet $\mathcal{U}$ and a probability distribution P_U on $\mathcal{U}$ defining a common randomness random variable U that is revealed to both the transmitter and the receiver before the start of the transmission,[2]
2) a sequence of encoding functions $\mathsf{f}_n : \mathcal{U} \times [M] \to \mathbb{R}$, $n = 1, \ldots, n_K$ that assign a codeword
$$\mathsf{f}(u, m)^{n_K} \triangleq (\mathsf{f}_1(u, m), \ldots, \mathsf{f}_{n_K}(u, m)) \tag{7}$$
to each message $m \in [M]$ and common randomness instance $u \in \mathcal{U}$. Each codeword satisfies the nested maximal power constraint P on all sub-codewords, giving
$$\|\mathsf{f}(u, m)^{n_k}\|^2 \leq n_k P \quad \forall\, m \in [M], u \in \mathcal{U}, k \in [K], \tag{8}$$
3) a non-negative integer-valued random stopping time $\tau \in \{n_1, \ldots, n_K\}$ for the filtration generated by $\{U, Y^{n_i}\}_{i=1}^K$ that satisfies an average decoding time constraint
$$\mathbb{E}\left[\tau\right] \leq N, \tag{9}$$
4) K decoding functions $\mathsf{g}_{n_k} : \mathcal{U} \times \mathbb{R}^{n_k} \to [M]$ for $k \in [K]$, satisfying an average error probability constraint
$$\mathbb{P}\left[\mathsf{g}_\tau(U, Y^\tau) \neq W\right] \leq \epsilon, \tag{10}$$
where the message W is equiprobable on the set $[M]$, and $X^\tau = \mathsf{f}(U, W)^\tau$.

Random variable U is common randomness shared by the transmitter and receiver. As in [6], [13], [20], the traditional random-coding argument does not prove the existence of a single (deterministic) code that simultaneously satisfies two conditions on the code (e.g., (9) and (10)). Therefore, randomized codes are necessary for our achievability argument; here, $|\mathcal{U}| \leq 2$ suffices [20, Appendix D].

The average power constraint on length-n_K codewords in Truong and Tan's VLSF code [12, Def. 1] is given by
$$\mathbb{E}\left[\|\mathsf{f}(U, W)^{n_K}\|^2\right] = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{n_K} \mathbb{E}\left[(\mathsf{f}_i(U, m))^2\right] \tag{11}$$
$$\leq NP. \tag{12}$$

[2] The realization u of U specifies the codebook.

As noted in [12, eq. (100)-(103)], for any code with stopping time τ, the expected value on the left-hand side of (11) can be replaced by the expected value $\mathbb{E}\left[\|\mathsf{f}(U, W)^\tau\|^2\right]$ because setting the symbols after time-slot τ to 0 does not affect the average error probability. Our maximal power constraint (8) implies the average power constraint (12), which follows by taking the expected value of (8) with respect to the stopping time $\tau \in \{n_1, \ldots, n_K\}$ and the equiprobable message W.

Given K decoding times and nested maximal power constraint P, we define the maximum achievable message size $M^*(N, K, \epsilon, P)$ as
$$M^*(N, K, \epsilon, P) \triangleq \max\{M : \exists \text{ an } (N, \{n_i\}_{i=1}^K, M, \epsilon, P) \text{ VLSF code}\}. \tag{13}$$
We define an $(N, \{n_i\}_{i=1}^K, M, \epsilon, P)_{\text{ave}}$ VLSF code under the average power constraint and the corresponding maximum achievable message size $M^*(N, K, \epsilon, P)_{\text{ave}}$ similarly, replacing the nested maximal power constraint (8) by the average power constraint (12).

D. Related Work

The following discussion summarizes prior asymptotic expansions of the maximum achievable message size for the Gaussian channel.

1) $M^*(N, 1, \epsilon, P)$: For $K = 1$, $P > 0$, and $\epsilon \in (0, 1)$, Tan and Tomamichel [18, Th. 1] and Polyanskiy *et al.* [7, Th. 54] show that
$$\ln M^*(N, 1, \epsilon, P) = NC(P) - \sqrt{NV(P)}Q^{-1}(\epsilon) + \frac{1}{2}\ln N + O(1). \tag{14}$$
The converse for (14) follows from [7, Th. 54]; the achievability for (14) follows from [18, Th. 1], which generates i.i.d. codewords uniformly distributed on the n-dimensional sphere with radius $\sqrt{nP}$ and applies maximum likelihood (ML) decoding. These results imply that random codewords uniformly distributed on a sphere and ML decoding are, together, third-order optimal, meaning that the gap between the achievability and converse bounds in (14) is $O(1)$.

2) $M^*(N, 1, \epsilon, P)_{\text{ave}}$: For $K = 1$ with an average-power-constraint, Yang *et al.* show in [21] that
$$\ln M^*(N, 1, \epsilon, P)_{\text{ave}} = N\,C\left(\frac{P}{1-\epsilon}\right) - \sqrt{N \ln N\, V\left(\frac{P}{1-\epsilon}\right)} + O(\sqrt{N}). \tag{15}$$
Yang *et al.* use a power control argument to show the achievability of (15). They divide the messages into disjoint sets $\mathcal{A}$ and $[M] \setminus \mathcal{A}$, where $|\mathcal{A}| = M(1-\epsilon)(1-o(1))$. For the messages in $\mathcal{A}$, they use an $\left(N, \{N\}, |\mathcal{A}|, \frac{2}{\sqrt{N \ln N}}, \frac{P}{1-\epsilon}(1-o(1))\right)$ VLSF code with a single decoding time N. The codewords are generated i.i.d. uniformly on the sphere with radius $\sqrt{N\frac{P}{1-\epsilon}(1-o(1))}$. The messages in $[M] \setminus \mathcal{A}$ are assigned the all-zero codeword. The converse for (15) follows from an application of the meta-converse [7, Th. 26].

3) $M^*(N, \infty, \epsilon, P)_{\text{ave}}$: : For VLSF codes with $K = \infty$, $n_i = i - 1$ for all i, and average power constraint (12), Truong and Tan show in [11, Th. 1] that for any $\epsilon \in (0, 1)$ and $P > 0$,

$$\frac{NC(P)}{1-\epsilon} - \ln N + O(1) \leq \ln M^*(N, \infty, \epsilon, P)_{\text{ave}} \tag{16}$$

$$\leq \frac{NC(P) + h_b(\epsilon)}{1-\epsilon}, \tag{17}$$

where $h_b(\epsilon) \triangleq -\epsilon \ln \epsilon - (1-\epsilon)\ln(1-\epsilon)$ is the binary entropy function (in nats). The bounds in (16)–(17) indicate that the ϵ-capacity (the first-order achievable term) is

$$\liminf_{N \to \infty} \frac{1}{N} \ln M^*(N, K, \epsilon, P) = \frac{C(P)}{1-\epsilon}. \tag{18}$$

The achievable dispersion term is zero, i.e., the second-order term in the fundamental limit in (16)–(17) is $o(\sqrt{N})$. The results in (16)–(17) are analogous to the fundamental limits for DMCs [6, Th. 2] and follow from arguments similar to those in [6]. Since the information density $\imath(X; Y)$ for the Gaussian channel is unbounded, bounding the expectation of the decoding time in the proof of [11, Th. 1] requires different techniques from those applicable to DMCs [6].

III. Main Result

Our main result is an asymptotic achievability bound for the scenario where $K \geq 2$ decoding times are available.

Theorem 1: Fix a finite integer $K \geq 2$ and real numbers $P > 0$ and $\epsilon \in (0, 1)$. For the Gaussian channel with noise variance 1 (3), the maximum message size (13) achievable by $(N, \{n_i\}_{i=1}^K, M, \epsilon, P)$ VLSF codes satisfies

$$\ln M^*(N, K, \epsilon, P) \geq \frac{NC(P)}{1-\epsilon} - \sqrt{N \ln_{(K-1)}(N) \frac{V(P)}{1-\epsilon}} + O\left(\sqrt{\frac{N}{\ln_{(K-1)}(N)}}\right). \tag{19}$$

As noted previously, any $(N, \{n_i\}_{i=1}^K, M, \epsilon, P)$ VLSF code is also an $(N, \{n_i\}_{i=1}^\infty, M, \epsilon, P)_{\text{ave}}$ VLSF code. Therefore, the converse in (17) provides an upper bound on $\ln M^*(N, K, \epsilon, P)$. Theorem 1 and (17) together imply that the ϵ-capacity (18) achievable within the class of VLSF codes in Def. 1 is $\frac{C(P)}{1-\epsilon}$. Neither switching from the average to maximal power constraint nor limiting the number of decoding times to finite K changes the first-order term. In fact, the same ϵ-capacity is achievable by a variable-length code without feedback that decodes at time 0 with probability $\epsilon(1 - o(1))$ (see the achievability proof of [6, Th. 1]). However, comparing Theorem 1 and (16), we see that the second-order term of our new achievability bound is significantly worse than the earlier results. Whether this is the consequence of our tighter power constraint or finite K or a weakness of our code construction is a topic for future research.

In the achievability bound in Theorem 1, the order of the second-order term, $-\sqrt{N \ln_{(K-1)}(N) \frac{V(P)}{1-\epsilon}}$, depends on the number of decoding times K. The rate convergence to the capacity grows with K. However, this dependence on

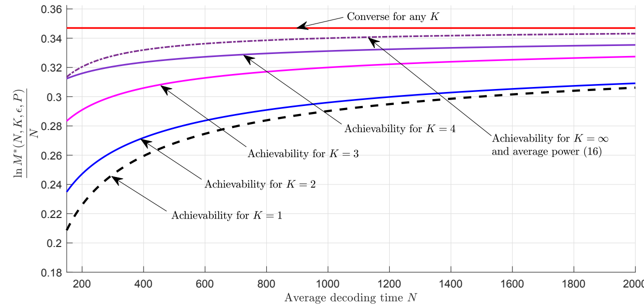

Fig. 1. The achievability bounds for the maximum achievable rate $\frac{\ln M^*(N,K,\epsilon,P)}{N}$ for $K = 1$ (14) and $K \in \{2, 3, 4\}$ (Theorem 1), and the achievability (16) and converse (17) bounds for $\frac{\ln M^*(N,\infty,\epsilon,P)_{\text{ave}}}{N}$ are shown. Here, $P = 1$ and $\epsilon = 10^{-3}$. The achievability bounds use the asymptotic approximation, i.e., we ignore the $O(\cdot)$ term in (16) and (19).

K is weak since $\ln_{(K-1)}(N)$ grows very slowly in N even when K is large. For example, for $K = 4$ and $N = 1000$, $\ln_{(K-1)}(N) \approx 0.659$. Furthermore, for $\epsilon = 10^{-3}$, $P = 1$, and $N = 1000$, 83.6% of the ϵ-capacity is achieved with $K = 1$, 85.3% with $K = 2$, 92.2% with $K = 3$, and 95.4% with $K = 4$. The achievability bounds for $K \in [4]$ and the maximal power constraint, and the achievability (16) and converse (17) bounds for $K = \infty$ and the average power constraint are illustrated in Fig. 1.

Theorem 1 builds on the following achievability bound for an $\left(N, \{n_i\}_{i=1}^K, M, \frac{1}{\sqrt{N \ln N}}, P\right)$ VLSF code.

Theorem 2: Fix an integer $K \geq 1$ and a real number $P > 0$. For the Gaussian channel with noise variance 1 (3), the maximum message size (13) achievable by $\left(N, \{n_i\}_{i=1}^K, M, \frac{1}{\sqrt{N \ln N}}, P\right)$ VLSF codes satisfies

$$\ln M^*\left(N, K, \frac{1}{\sqrt{N \ln N}}, P\right) \geq NC(P) - \sqrt{N \ln_{(K)}(N) V(P)} + O\left(\sqrt{\frac{N}{\ln_{(K)}(N)}}\right). \tag{20}$$

The $K = 1$ case in Theorem 2 is recovered by [22, Th. 7], which investigates moderate deviations bounds in channel coding.

Theorems 1 and 2 follow from an application of the following non-asymptotic achievability bound.

Theorem 3: Fix a constant γ and decoding times $n_1 < \cdots < n_K$. For any positive numbers N, P, and $\epsilon \in (0, 1)$, there exists an $(N, \{n_i\}_{i=1}^K, M, \epsilon, P)$ VLSF code for the Gaussian channel (3) with

$$\epsilon \leq \mathbb{P}\left[\imath(X^{n_K}; Y^{n_K}) < \gamma\right] + (M-1)\exp\{-\gamma\} + \mathbb{P}\left[\bigcup_{i=1}^K \left\{\|X^{n_i}\|^2 > n_i P\right\}\right] \tag{21}$$

$$N \leq n_1 + \sum_{i=1}^{K-1} (n_{i+1} - n_i) \mathbb{P}\left[\bigcap_{j=1}^{i} \{\imath(X^{n_j}; Y^{n_j}) < \gamma\}\right], \tag{22}$$

where $P_{X^{n_K}}$ is a product of K distributions on subvectors of dimensions $n_j - n_{j-1}$, $j \in [K]$, i.e.,

$$P_{X^{n_K}}(x^{n_K}) = \prod_{j=1}^{K} P_{X_{n_{j-1}+1}^{n_j}}(x_{n_{j-1}+1}^{n_j}). \tag{23}$$

The proof sketches for Theorems 1–3 appear in Section IV.

IV. Proof Sketches

A. Proof Sketch for Theorem 1

Our achievability proof is inspired by Polyanskiy *et al.*'s coding scheme in [6] for DMCs, where the number of decoding times is unlimited. In [6], to prove an achievability result with an average error probability ϵ and an average decoding time N, the decoder decodes to an arbitrary message at time $n_1 = 0$ with probability $p = \frac{\epsilon - \epsilon'_N}{1-\epsilon'_N}$. With probability $1-p$, the code uses another VLSF code with average decoding time N' and average error probability $\epsilon'_N = \frac{1}{N'}$. Truong and Tan [11], [12] adapt this coding scheme with the same choice of ϵ'_N to the Gaussian point-to-point and multiple access channels with an unlimited number of decoding times and an average power constraint. Like [6], [11], [12], our coding strategy fixes the smallest decoding time n_1 to 0. To achieve the best second-order term within our code structure, we set $\epsilon'_N \triangleq \frac{1}{\sqrt{N' \ln N'}}$. Before transmission starts, the decoder generates a random variable $D \sim \mathrm{Bernoulli}(p)$ with

$$p \triangleq \frac{\epsilon - \epsilon'_N}{1 - \epsilon'_N}. \tag{24}$$

If $D = 1$, then the decoder decodes to an arbitrary message at time $n_1 = 0$; otherwise the encoder and decoder use an $(N', \{n_k\}_{k=2}^K, M, \epsilon'_N, P)$ VLSF code. The average error probability for this coding scheme is bounded by

$$1 \cdot p + \epsilon'_N \cdot (1-p) = \epsilon, \tag{25}$$

and the average decoding time is bounded by

$$N = \mathbb{E}[\tau] = p \cdot 0 + (1-p) \cdot N' = (1-p)N'. \tag{26}$$

From (24) and (26), we get the asymptotic expansion

$$N' = \frac{N}{1-\epsilon}\left(1 + O\left(\frac{1}{\sqrt{N \ln N}}\right)\right). \tag{27}$$

By Theorem 2, there exists an $(N', \{n_k\}_{k=2}^K, M, \epsilon'_N, P)$ VLSF code with

$$\ln M = N'C(P) - \sqrt{N' \ln_{(K-1)}(N')\, V(P)} + O\left(\sqrt{\frac{N'}{\ln_{(K-1)}(N')}}\right). \tag{28}$$

Plugging (27) into (28) and using the Taylor series expansion of the function $\sqrt{N' \ln_{(K-1)}(N')}$ completes the proof. ∎

B. Proof Sketch and Discussion for Theorem 2

1) Random encoder design: To prove Theorem 2, we choose the distribution of the random codewords, $P_{X^{n_K}}$, in Theorem 3 as in our prior work [23, Th. 4]. Fixing decoding times $n_1, n_2, \ldots, n_K$, we generate M i.i.d. codewords as follows. Set $n_0 = 0$. The sub-codewords are drawn independently, with $X_{n_{j-1}+1}^{n_j}$, $j \in [K]$, drawn from the uniform distribution on the $(n_j - n_{j-1})$-dimensional sphere with radius $\sqrt{(n_j - n_{j-1})P}$. The resulting codewords are uniformly distributed on the subset of the n_K-dimensional power sphere that satisfies our K maximal power constraints (8) with equality.

2) Probability analysis: The VLSF code in Theorem 2 considers a vanishing error probability, $\frac{1}{\sqrt{N \ln N}}$, that does not decay exponentially with N. Therefore, with an appropriate choice of γ, bounding the first probability term in (21) and the probability term in (22) requires moderate deviations techniques. We use [24, Ch. 8, Th. 4] and [25, Prop. 2] to bound these probabilities. The decoding times $n_1, \ldots, n_K$ are chosen to minimize the right-hand side of (22). ∎

Applying our analysis of multiple access code performance in [23] to the Gaussian point-to-point channel, we see that for any finite K and sufficiently large increments $n_i - n_{i-1}$ for all $i \in [K]$, using the restricted subset in our random codebook design instead of the entire n_K-dimensional power sphere results in no change in (14) up to the third-order term.

From Shannon's work in [26], it is well-known that for the Gaussian channel with a maximal power constraint, drawing i.i.d. Gaussian codewords yields a performance inferior to that achieved by the uniform distribution on the power sphere. While almost all tight achievability bounds for the Gaussian channel in the fixed-length regime under a variety of settings (e.g., all four combinations of maximal or average power constraint and feedback or no feedback [8], [17], [18], [21]) employ random codewords drawn uniformly on the power sphere, Truong and Tan's result (16) for VLSF codes with an average power constraint employs i.i.d. Gaussian inputs. The Gaussian distribution works in this scenario because when $K = \infty$, the term $\mathbb{P}[\imath(X^{n_K}; Y^{n_K}) < \gamma]$ in (21), which is usually dominant, disappears. The second term $(M-1)\exp\{-\gamma\}$ in (21) is not affected by the input distribution. Unfortunately, drawing codewords i.i.d. $\mathcal{N}(0, P)$ satisfies the average power constraint (12) but not the maximal power constraint (8). When $K < \infty$ and the probability $\mathbb{P}[\imath(X^{n_K}; Y^{n_K}) < \gamma]$ dominates, using i.i.d. $\mathcal{N}(0, P)$ inputs achieves a worse second-order term in (19). This implies that when $K < \infty$, using our uniform distribution on a subset of the power sphere is again beneficial even under the average power constraint. In particular, i.i.d. $\mathcal{N}(0, P)$ inputs achieve (19), where the dispersion $V(P)$ is replaced by $\tilde{V}(P) = \frac{P}{1+P}$, which is the variance of $\imath(X; Y)$ when $X \sim \mathcal{N}(0, P)$; here $\tilde{V}(P)$ is greater than the dispersion $V(P)$ for all $P > 0$ (see [27, eq. (2.56)]).

C. Proof Sketch for Theorem 3

The proof of Theorem 3 extends the arguments of the random coding bound in [6, Th. 3] to the scenario with finite K where each codeword must satisfy a list of maximal power constraints (8). Variations of Theorem 3 are proved in [9] and [15]. ∎

Acknowledgment

Stimulating discussions with Prof. Richard Wesel and Prof. Aaron Wagner are gratefully acknowledged.

References

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

[5] M. V. Burnashev, "Data transmission over a discrete channel with feedback: Random transmission time," *Problems of Information Transmission*, vol. 12, no. 4, pp. 10–30, 1976.

[6] 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.

[7] 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.

[8] S. L. Fong and V. Y. F. Tan, "Asymptotic expansions for the AWGN channel with feedback under a peak power constraint," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, China, Jun. 2015, pp. 311–315.

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

[10] Y. Altuğ, H. V. Poor, and S. Verdú, "Variable-length channel codes with probabilistic delay guarantees," in *53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, USA, Sep. 2015, pp. 642–649.

[11] L. V. Truong and V. Y. F. Tan, "On Gaussian MACs with variable-length feedback and non-vanishing error probabilities," *arXiv:1609.00594v2*, Sep. 2016.

[12] ——, "On Gaussian MACs with variable-length feedback and non-vanishing error probabilities," *IEEE Trans. Inf. Theor.*, vol. 64, no. 4, p. 2333–2346, Apr. 2018.

[13] K. F. Trillingsgaard, W. Yang, G. Durisi, and P. Popovski, "Common-message broadcast channels with feedback in the nonasymptotic regime: Stop feedback," *IEEE Trans. Inf. Theory*, vol. 64, no. 12, pp. 7686–7718, Dec. 2018.

[14] K. Vakilinia, S. V. S. Ranganathan, D. Divsalar, and R. D. Wesel, "Optimizing transmission lengths for limited feedback with nonbinary ldpc examples," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2245–2257, 2016.

[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, Jul. 2015.

[16] A. Heidarzadeh, J. 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.

[17] L. V. Truong, S. L. Fong, and V. Y. F. Tan, "On Gaussian channels with feedback under expected power constraints and with non-vanishing error probabilities," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1746–1765, Mar. 2017.

[18] V. Y. F. Tan and M. Tomamichel, "The third-order term in the normal approximation for the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2430–2438, May 2015.

[19] R. C. Yavas, V. Kostina, and M. Effros, "Variable-length feedback codes with several decoding times for the Gaussian channel," *arXiv:2103.09373*, Mar. 2021.

[20] ——, "Random access channel coding in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 67, no. 4, pp. 2115–2140, 2021.

[21] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, "Optimum power control at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4598–4615, Sep. 2015.

[22] Y. Polyanskiy and S. Verdú, "Channel dispersion and moderate deviations limits for memoryless channels," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, USA, Sep. 2010, pp. 1334–1339.

[23] R. C. Yavas, V. Kostina, and M. Effros, "Gaussian multiple and random access in the finite blocklength regime," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angeles, CA, USA, June 2020, pp. 3013–3018.

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

[25] E. MolavianJazi and J. N. Laneman, "A second-order achievable rate region for Gaussian multi-access channels via a central limit theorem for functions," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6719–6733, Dec. 2015.

[26] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *The Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, May 1959.

[27] E. MolavianJazi, "A unified approach to Gaussian channels with finite blocklength," Ph.D. dissertation, University of Notre Dame, Jul. 2014.