2021 IEEE Information Theory Workshop (ITW)

Nested Sparse Feedback Codes for Point-to-Point, Multiple Access, and Random Access Channels

Recep Can Yavas, Victoria Kostina, and Michelle Effros

2021 IEEE Information Theory Workshop (ITW) | 978-1-6654-0312-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ITW48936.2021.9611385

***Abstract*—This paper investigates variable-length feedback codes for discrete memoryless channels in point-to-point, multiple access, and random access communication. The proposed nested code employs L decoding times $n_1, n_2, \ldots, n_L$ for the point-to-point and multiple access channels and KL decoding times $\{n_{k,\ell}\colon 1 \leq k \leq K, 1 \leq \ell \leq L\}$ for the random access channel with at most K active transmitters; in the latter case, decoding times $n_{k,\ell}$, $1 \leq \ell \leq L$ are reserved for decoding in the scenario where the decoder believes that the number of active transmitters is k. The code has a nested structure, i.e., codewords used to decode messages from k active transmitters are prefix of codewords used to decode messages from $k+1$ active transmitters. The code employs single-bit, scheduled feedback from the receiver to the transmitters at each potential decoding time to inform the transmitters whether or not it is able to decode. Transmitters cease transmission, thereby truncating their codewords, when no further transmissions are required by the decoder. The choice of decoding times is optimized to minimize the expected decoding time subject to an error probability constraint, and second order achievability bounds are derived.**

I. Introduction

Noiseless feedback does not increase the capacity of discrete memoryless (DM) point-to-point channels (PPCs) [1]. Neither does it improve the error exponent of symmetric DM-PPCs in the fixed-length regime [2]. However, feedback has several benefits, including simplified coding schemes [3], [4] and improved second-order achievable rates [5].

Feedback becomes even more beneficial to code performance when employed in variable-length codes that allow decoding at arbitrary time instants. In [6], Burnashev shows that variable-length codes with feedback achieve significant improvement in the achievable error exponent. Polyanskiy *et al.* [7] introduce variable-length feedback (VLF) codes for DM-PPCs and extend Burnashev's result to the finite blocklength regime with non-vanishing error probabilities. Polyanskiy *et al.* [7] prove that the achievable rates of VLF codes converge more quickly to the channel capacity than those of fixed-length codes without feedback, giving VLF codes a particular advantage at short blocklengths. As a special case of variable-length feedback codes, Polyanskiy *et al.* define variable-length stop-feedback (VLSF) codes that use one-bit feedback at each of the potential decoding times. In VLSF codes, feedback is used only to inform the transmitter whether it should end the transmission or continue to transmit. Hence, codewords are not a function of the received feedback signal.

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.

Several works that study VLSF codes in different scenarios include [8]–[13]. In all of these works, the decoder is allowed to decode at any of the time instants $0, 1, 2, \ldots$, i.e., the number of potential decoding times $L = \infty$. By large deviations theorem, using more than cN decoding times does not yield a greater second-order term than of achieved by an unbounded L. Therefore, in this paper, we evaluate the scenario with $L = \infty$ as $L = \Omega(N)$. Some examples that consider VLSF codes with $L = O(1)$ decoding times are [14]–[17]. The setting with $L = O(1)$ is also studied in the literature under the name of incremental redundancy hybrid automatic repeat request codes. Our work in [17] derives the first asymptotic expansion for the achievable rates of VLSF codes with L decoding times for the Gaussian PPC with maximal power constraints. Using the code construction in [17], the performance improvement due to increasing L diminishes rapidly beyond $L > 4$. VLSF codes with sparse feedback achieve fairly high coding rates. For example, 95.4% of the ϵ-capacity can be achieved by VLSF codes with $L = 4$, $N = 1000$, and 0 dB signal-to-noise ratio.

This paper extends VLSF codes with L decoding times to the DM-PPC, discrete memoryless multiple access channel (DM-MAC), and discrete memoryless random access channel (DM-RAC). Heidari *et al.* [18] extend Burnashev's work to the DM-MAC and derive lower and upper bounds on the error exponent of VLF codes for the DM-MAC. Bounds on the performance of VLSF codes appear in [9] for the Gaussian MAC with expected power constraints, and in [12] for the DM-MAC. In both [9] and [12], $2^k - 1$ simultaneous information density threshold rules are employed for the k-transmitter MAC. The central result of [19] is that for permutation-invariant RACs under some mild symmetry conditions, it is possible to attain the first- and second-order terms of the best-known code for the MAC in operation. In [19], the code employs K decoding times $n_1, \ldots, n_K$, where the decoder decodes messages at n_k if it believes that the number of active transmitters is k, and at each potential decoding time, one-bit feedback is sent from the receiver to the transmitters.

In this work, we revisit the impact of feedback in variable-rate codes for the DM-PPC, DM-MAC, and DM-RAC. For non-corner points in the achievable MAC region and $L = \Omega(N)$, we employ a single threshold rule to improve the second-order term achieved in [9] from $-O(\sqrt{N})$ to $-\ln N$. For the DM-RAC, we employ the channel model from [19], which comprises a single receiver and an unknown number of active transmitters. We here extend our approach from [19] for the RAC from one possible decoding time n_k for each possible

978-1-6654-0312-2/21/$31.00 ©2021 IEEE

estimate k of the number of active transmitters to $L \geq 1$ decoding times $n_{k,1}, n_{k,2}, \ldots, n_{k,L}$. This extension increases the expected achievable rate at the expense of increasing the number of feedback bits in the communication epoch. Note that the feedback employed for the DM-RAC is still sparse as the total number of feedback bits is at most KL.

The paper is organized as follows. We define notation and channel models in Section II. Section III introduces VLSF codes. Section IV presents and discusses the main results. Proofs are relegated to the extended version [20].

II. System Model

A. Notation and Definitions

For any positive integers k and n, $[k] \triangleq \{1, \ldots, k\}$ and $x^n \triangleq (x_1, \ldots, x_n)$. For a collection of length-n vectors $x_1^n, \ldots, x_K^n$ and any subset $\mathcal{A} \subseteq [K]$, we denote the sub-vectors indexed with the elements in $\mathcal{A}$ by $x_{\mathcal{A}}^n \triangleq \{x_a^n \colon a \in \mathcal{A}\}$. We write $x_{\mathcal{A}} \stackrel{\pi}{=} y_{\mathcal{A}}$ to indicate that there exists a permutation π of $y_{\mathcal{A}}$ such that $x_{\mathcal{A}} = \pi(y_{\mathcal{A}})$, and we write $x_{\mathcal{A}} \stackrel{\pi}{\neq} y_{\mathcal{A}}$ to indicate that there exists no such permutation. We use $\ln(\cdot)$ to denote the natural logarithm. We measure information in nats. We use the standard $O(\cdot)$, $\Omega(\cdot)$, and $o(\cdot)$ notations, i.e., $f(n) = O(g(n))$ if $\limsup_{n\to\infty} |f(n)|/g(n) < \infty$, $f(n) = \Omega(g(n))$ if $\liminf_{n\to\infty} f(n)/g(n) > 0$, 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. 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

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

and undefined for all other (k, x) pairs.

For a channel transition probability $P_{Y_k|X_{[k]}}$ with k transmitters, let P_{Y_k} denote the marginal output distribution induced by the input distribution $P_{X_{[k]}}$. The unconditional and conditional information densities are defined as

$$\imath_k^{\mathcal{A}}(x_{\mathcal{A}}; y) \triangleq \ln \frac{P_{Y_k|X_{\mathcal{A}}}(y|x_{\mathcal{A}})}{P_{Y_k}(y)} \tag{2}$$

$$\imath_k^{\mathcal{A}}(x_{\mathcal{A}}; y|x_{\mathcal{A}^c}) \triangleq \ln \frac{P_{Y_k|X_{[k]}}(y|x_{[k]})}{P_{Y_k|X_{\mathcal{A}^c}}(y|x_{\mathcal{A}^c})} \tag{3}$$

for any $\mathcal{A} \subseteq [k]$, where $\mathcal{A} \neq \emptyset$ and $\mathcal{A}^c = [k] \setminus \mathcal{A}$.

The corresponding mutual informations under the input distribution $P_{X_{[k]}}$ and the channel transition probability $P_{Y_k|X_{[k]}}$ are defined as

$$I_k(X_{\mathcal{A}}; Y_k) \triangleq \mathbb{E}\left[\imath_k^{\mathcal{A}}(X_{\mathcal{A}}; Y_k)\right] \tag{4}$$

$$I_k(X_{\mathcal{A}}; Y_k|X_{\mathcal{A}^c}) \triangleq \mathbb{E}\left[\imath_k^{\mathcal{A}}(X_{\mathcal{A}}; Y_k|X_{\mathcal{A}^c})\right]. \tag{5}$$

For brevity, we define

$$I_k \triangleq I_k(X_{[k]}; Y_k) \tag{6}$$

$$V_k \triangleq \mathrm{Var}\left[\imath_k^{[k]}(X_{[k]}; Y_k)\right]. \tag{7}$$

B. Channel Models: DM-PPC, DM-MAC, and DM-RAC

Definition 1: A DM-PPC is described by $(\mathcal{X}, P_{Y|X}, \mathcal{Y})$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite input and output alphabets of the channel, and the conditional probability $P_{Y|X}$ describes the channel transition probabilities.

Definition 2: A K-transmitter DM-MAC is defined by $\left(\prod_{i=1}^K \mathcal{X}_i, P_{Y|X_{[K]}}, \mathcal{Y}\right)$, where $\mathcal{X}_i$ is the finite input alphabet for transmitter $i \in [K]$, $\mathcal{Y}$ is the finite output alphabet of the channel, and $P_{Y_K|X_{[K]}}$ is the channel transition probability.

We define a DM-RAC that consists of an unknown number of active transmitters and a single receiver as in [19].

Definition 3: A permutation-invariant, reducible DM-RAC for the maximal number of transmitters $K < \infty$ is defined by a family of DM-MACs $\left\{\left(\mathcal{X}^k, P_{Y_k|X_{[k]}}, \mathcal{Y}_k\right)\right\}_{k=1}^K$, where the k-th DM-MAC defines the channel for k active transmitters.

Each of the DM-MACs satisfies the *permutation-invariance* assumption

$$P_{Y_k|X_{[k]}}(y|x_{[k]}) = P_{Y_k|X_{[k]}}(y|\hat{x}_{[k]}) \tag{8}$$

for all $\hat{x}_{[k]} \stackrel{\pi}{=} x_{[k]}$ and $y \in \mathcal{Y}_k$, $k \in [K]$, and different DM-MACs are connected by *reducibility* assumption

$$P_{Y_s|X_{[s]}}(y|x_{[s]}) = P_{Y_k|X_{[k]}}(y|x_{[s]}, 0^{k-s}) \tag{9}$$

for all $s < k$, $x_{[s]} \in \mathcal{X}_{[s]}$, and $y \in \mathcal{Y}_s$, where $0 \in \mathcal{X}$ specifies a unique "silence" symbol that is transmitted when a transmitter is silent.

In addition to simplifying the presentation, the permutation-invariance (8) and reducibility (9) assumptions allow us to show that the symmetrical rate point $(R, R, \ldots, R)$ at which the code operates lies on the sum-rate boundary of the underlying DM-MAC, enabling the use of a single-threshold rule at the decoder.

III. VLSF Codes

A. VLSF Code Definitions for the PPC and MAC

For both the PPC and MAC, we consider the VLSF code with L decoding times introduced in [17] for the Gaussian PPC with maximal power constraints. The code employs L predetermined decoding times $n_1 < n_2 < \cdots < n_L$. The receiver stops the transmission at the first potential decoding time $n_\ell \in \{n_1, \ldots, n_L\}$ at which it is able to decode. The transmitters are informed of the receiver's decision by a one-bit feedback signal at times $n_1, \ldots, n_\ell$. Feedback bit "0" at time n_i means that the decoder is not ready to decode, the decoder output at time n_i is an erasure symbol "e", and the transmitters should continue to transmit symbols. Feedback bit "1" means that the decoder is ready to declare a message outcome, and transmission should stop. We require the average decoding time of a VLSF code to be bounded by N, and the average error probability to be bounded by ϵ. Below, we formally define VLSF codes for the DM-PPC.

Definition 4: Fix $\epsilon \in (0, 1)$, $N \in (0, \infty)$, integers $0 \leq n_1 < \cdots < n_L$, and $M > 0$. An $(N, \{n_\ell\}_{\ell=1}^L, M, \epsilon)$ VLSF code for the PPC 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) a sequence of encoding functions $\mathsf{f}_n\colon \mathcal{U} \times [M] \to \mathcal{X}$, $n = 1, \ldots, n_L$ that assign a codeword

$$\mathsf{f}(u,m)^{n_L} \triangleq (\mathsf{f}_1(u,m), \ldots, \mathsf{f}_{n_L}(u,m)) \tag{10}$$

to each message $m \in [M]$ and common randomness instance $u \in \mathcal{U}$,
3) a non-negative integer-valued random stopping time $\tau \in \{n_1, \ldots, n_L\}$ for the filtration generated by $\{U, Y^{n_\ell}\}_{\ell=1}^L$, satisfying an average decoding time constraint

$$\mathbb{E}[\tau] \leq N, \tag{11}$$

4) L decoding functions $\mathsf{g}_{n_\ell}\colon \mathcal{U} \times \mathcal{Y}^{n_\ell} \to [M] \cup \{\mathsf{e}\}$ for $\ell \in [L]$, satisfying an average error probability constraint

$$\mathbb{P}[\mathsf{g}_\tau(U, Y^\tau) \neq W] \leq \epsilon, \tag{12}$$

where the message W is uniformly distributed on the set $[M]$, and $X^\tau = \mathsf{f}(U, W)^\tau$.

The given code definition differs slightly from those in [7], [17]. The code in [17] requires the codewords $\mathsf{f}(u,m)^{n_L}$ to satisfy a list of maximal power constraints, while the code in Definition 4 does not. The code in Definition 4 differs from that in [7, Def. 1] only in that Definition 4 limits the number of potential decoding times by L, while [7, Def. 1] imposes no limit.

A VLSF code for the MAC is defined similarly to the VLSF code for the PPC. For simplicity, we define VLSF codes for the MAC only for the two-transmitter case. This definition extends naturally to the K-transmitter MAC.

Definition 5: Fix $\epsilon \in (0,1)$, $N \in (0, \infty)$, integers $0 \leq n_1 < \cdots < n_L$, and $M_1, M_2 > 0$. An $(N, \{n_\ell\}_{\ell=1}^L, M_1, M_2, \epsilon)$ VLSF code for the MAC comprises

1) two finite alphabets $\mathcal{U}_1$ and $\mathcal{U}_2$ defining two common randomness random variables U_1 and U_2,
2) two sequences of encoding functions $\mathsf{f}_n^{(i)}\colon \mathcal{U}_i \times [M_i] \to \mathcal{X}_i$, $i = 1, 2$,
3) a stopping time τ for the filtration generated by $\{U_1, U_2, Y^{n_\ell}\}_{\ell=1}^L$, satisfying an average decoding time constraint (11), and
4) L decoding functions $\mathsf{g}_{n_\ell}\colon \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{Y}^{n_\ell} \to \{[M_1] \times [M_2]\} \cup \{\mathsf{e}\}$ for $\ell \in [L]$, satisfying an average error probability constraint

$$\mathbb{P}[\mathsf{g}_\tau(U_1, U_2, Y^\tau) \neq (W_1, W_2)] \leq \epsilon, \tag{13}$$

where the independent messages W_1 and W_2 are uniformly distributed on the sets $[M_1]$ and $[M_2]$, respectively.

B. VLSF Code Definition for the RAC

The VLSF RAC code defined here combines the rateless communication strategy that we introduce in [19] with the VLSF PPC and MAC codes described above. In this case, when the decoder concludes that k transmitters are active, it can decode at any of the L decoding times $n_{k,1} < n_{k,2} < \cdots < n_{k,L}$ as opposed to just at n_k as in [19], [21]. At time $n_{k,\ell}$, the receiver broadcasts feedback bit "1" to the transmitters if it is able to decode k messages; otherwise, it outputs an erasure symbol "e" and sends feedback bit "0" signaling that decoding has not occurred and transmission should continue.

As in [19], [21], we assume a compound RAC model, that is, we do not assign probabilities to the potential set of transmitters $\mathcal{A} \subseteq [K]$, and an agnostic RAC model, that is, the transmitters know nothing about the set $\mathcal{A}$ except their own membership and the receiver's feedback at potential decoding times. We employ identical encoding [22], that is, all transmitters use the same codebook. This implies that the RAC code operates at the symmetrical rate point, i.e., $M_i = M$ for $i \in [K]$. Due to the identical encoding, the decoder is required to decode the list of the messages transmitted by the active transmitters but not the identities of the transmitters.

We formally define VLSF codes for the RAC as follows.

Definition 6: Fix $\epsilon \in (0,1)$, $N_1, \ldots, N_K \in (0, \infty)$, and a set of integers $\mathcal{N} \triangleq \{n_{k,\ell} \geq 0\colon k \in [K], \ell \in [L]\}$. An $(\{N_k\}_{k=1}^K, \mathcal{N}, M, \epsilon)$ identical-encoder VLSF code comprises

1) a common randomness random variable U on an alphabet $\mathcal{U}$,
2) a sequence of encoding functions $\mathsf{f}_n\colon \mathcal{U} \times [M] \to \mathcal{X}$, $n = 1, 2, \ldots, n_{K,L}$, defining M length-$n_{K,L}$ codewords, where $n_{K,L}$ is assumed to be the largest decoding time in $\mathcal{N}$,
3) K non-negative integer-valued random stopping times $\tau_k \in \mathcal{N}$ for the filtration generated by $\{U, Y_k^n\}_{n \in \mathcal{N}}$, satisfying that

$$\mathbb{E}[\tau_k] \leq N_k \tag{14}$$

when $k \in [K]$ messages $W_{[k]}$ are transmitted,
4) KL decoding functions $\mathsf{g}_{n_{k,\ell}}\colon \mathcal{U} \times \mathcal{Y}_k^{n_{k,\ell}} \to [M]^k \cup \{\mathsf{e}\}$ for $k \in [K]$ and $\ell \in [L]$, satisfying an average error probability constraint

$$\mathbb{P}\left[\mathsf{g}_{\tau_k}(U, Y_k^{\tau_k}) \overset{\pi}{\neq} W_{[k]}\right] \leq \epsilon \tag{15}$$

when $k \in [K]$ messages $W_{[k]}$ are transmitted, where $W_{[k]}$ are independent and equiprobable on the set $[M]$.

IV. Main Results

Our main results are second-order achievability bounds for the VLSF codes with sparse feedback over the DM-PPC, DM-MAC, and DM-RAC.

A. DM-PPC

Theorem 1: Fix $\epsilon \in (0,1)$, an integer $L = O(1) \geq 2$, and a distribution P_X. For any DM-PPC $(\mathcal{X}, P_{Y|X}, \mathcal{Y})$, there exists an $(N, \{n_\ell\}_{\ell=1}^L, M, \epsilon)$ VLSF code provided that

$$\ln M \leq \frac{N I_1}{1-\epsilon} - \sqrt{N \ln_{(L-1)}(N) \frac{V_1}{1-\epsilon}}$$

$$+ O\left(\sqrt{\frac{N}{\ln_{(L-1)}(N)}}\right). \quad (16)$$

The decoding times that achieve (16) satisfy the equations

$$\ln M = n_\ell I_1 - \sqrt{n_\ell \ln_{(L-\ell+1)}(n_\ell) V_1} - \ln n_\ell + O(1) \quad (17)$$

for $\ell \in \{2, \ldots, L\}$, and $n_1 = 0$.

Theorem 1 is proved by analyzing the bounds in [23, Th. 3].

In [7, Th. 2], Polyanskiy *et al.* show that the second-order term $-\ln N$ is achievable for the DM-PPC in a scenario where all time instants are available for decoding, i.e., $L = \Omega(N)$. They also show that the second-order term in the converse is $+O(1)$ for $L = \Omega(N)$; the converse applies to the scenario with $L = O(1)$ as well. How to close the gap between the achievability bound in Theorem 1 and the converse bound in [7, Th. 4] remains an open problem.

The coding strategy and the technique to prove Theorem 1 are closely related to those of [17, Th. 1], where VLSF codes with L decoding times for the Gaussian PPC are considered. The main difference is that the sub-vectors of the random codewords employed in [17] are distributed uniformly over their corresponding power spheres, while for the DM-PPC, we employ an i.i.d. code ensemble according to some P_X. Adapting the analysis of the optimization problem in [23, Sec. IV-C, E], applicable to the Gaussian PPC with maximal power constraints, to the i.i.d. code ensembles, we deduce that the choice of decoding times in (17) minimizes the average decoding time N within our code structure.

As discussed in [17], Theorem 1 suggests that under our code structure there is little benefit to having more than $L = 4$ decoding times for practical values of N (i.e., $N \in [10^3, 10^5]$).

Using the maximization lemmas in [24, Appendix J], we see that the optimal P_X^* that maximizes the right-hand side of (16) up to the second-order term is the capacity-achieving input distribution with the minimum dispersion V_1.

B. DM-MAC With Two Transmitters

Theorem 2: Fix $\epsilon \in (0, 1)$, an integer $L = O(1) \geq 2$, and distributions P_{X_1} and P_{X_2}. For any DM-MAC with two transmitters $(\mathcal{X}_1 \times \mathcal{X}_2, P_{Y|X_{[2]}}, \mathcal{Y})$, there exists an $(N, \{n_\ell\}_{\ell=1}^L, M_1, M_2, \epsilon)$ VLSF code with

$$\ln M_1 + \ln M_2 \leq \frac{NI_2}{1-\epsilon} - \sqrt{N \ln_{(L-1)}(N) \frac{V_2}{1-\epsilon}} + O\left(\sqrt{\frac{N}{\ln_{(L-1)}(N)}}\right), \quad (18)$$

provided that the rate pair $\left(\frac{\ln M_1}{N}, \frac{\ln M_2}{N}\right)$ approaches a point on the sum-rate boundary, i.e., the (M_1, M_2) pair satisfies

$$\left(\frac{\ln M_1}{N}, \frac{\ln M_2}{N}\right) \in \Big\{(r_1 + o(1), r_2 + o(1)) \colon r_1 + r_2 = \frac{I_2}{1-\epsilon}, r_1 < \frac{I_2(X_1; Y|X_2)}{1-\epsilon}, r_2 < \frac{I_2(X_2; Y|X_1)}{1-\epsilon}\Big\}. \quad (19)$$

Theorem 3: Under the setting of Theorem 2, where $L = \Omega(N)$, for any rate point approaching a point on the sum-rate boundary (19), there exists an $(N, \{n_\ell\}_{\ell=1}^L, M_1, M_2, \epsilon)$ VLSF code provided that

$$\ln M_1 + \ln M_2 \leq \frac{NI_2}{1-\epsilon} - \ln N + O(1). \quad (20)$$

Theorems 2 and 3 follow from an application of the non-asymptotic achievability bound, Theorem 4, below.

Theorem 4: Fix constants $\epsilon \in (0, 1)$, γ, $\lambda_1 > 0$, $\lambda_2 > 0$, integers $0 \leq n_1 < \cdots < n_L$, and distributions P_{X_1} and P_{X_2}. For any DM-MAC with two transmitters $(\mathcal{X}_1 \times \mathcal{X}_2, P_{Y|X_{[2]}}, \mathcal{Y})$, there exists an $(N, \{n_\ell\}_{\ell=1}^L, M_1, M_2, \epsilon)$ VLSF code with

$$\epsilon \leq \mathbb{P}\left[\imath_2^{[2]}(X_{[2]}^{n_L}; Y^{n_L}) < \gamma\right] \quad (21)$$

$$+ (M_1 - 1)(M_2 - 1)\exp\{-\gamma\} \quad (22)$$

$$+ \sum_{\ell=1}^{L} \mathbb{P}\left[\imath_2^2(X_2^{n_\ell}; Y^{n_\ell}) > NI_2(X_2; Y) + N\lambda_1\right] \quad (23)$$

$$+ (M_1 - 1)\exp\{-\gamma + NI_2(X_2; Y) + N\lambda_1\} \quad (24)$$

$$+ \sum_{\ell=1}^{L} \mathbb{P}\left[\imath_2^1(X_1^{n_\ell}; Y^{n_\ell}) > NI_2(X_1; Y) + N\lambda_2\right] \quad (25)$$

$$+ (M_2 - 1)\exp\{-\gamma + NI_2(X_1; Y) + N\lambda_2\} \quad (26)$$

$$N \leq n_1 + \sum_{i=1}^{L-1} (n_{i+1} - n_i)\mathbb{P}\left[\bigcap_{j \in [i]} \{\imath_2^{[2]}(X_{[2]}^{n_j}; Y^{n_j}) < \gamma\}\right]. \quad (27)$$

The proof of Theorem 4 uses a random coding argument that employs i.i.d. codebook ensembles with distributions P_{X_1} and P_{X_2}. The decoder uses a single threshold rule based on the information density. At time n_ℓ, it computes the information densities $\imath_2^{[2]}(X_1^{n_\ell}(m_1), X_2^{n_\ell}(m_2); Y^{n_\ell})$, and, if there exists a message pair $(\hat{m}_1, \hat{m}_2)$ satisfying $\imath_2^{[2]}(X_1^{n_\ell}(\hat{m}_1), X_2^{n_\ell}(\hat{m}_2); Y^{n_\ell}) > \gamma$, then $(\hat{m}_1, \hat{m}_2)$ is decoded. Otherwise, the decoder passes the decoding time n_ℓ without decoding. If $n_\ell < n_L$, the transmission continues until the next decoding time.

In Theorem 4, the term (21) bounds the probability that the information density corresponding to the true messages is below the threshold for all decoding times; the term (22) bounds the probability that both messages are decoded incorrectly; and the terms (23)-(26) bound the probability that one of the transmitted messages is decoded correctly and the other message is decoded incorrectly. In the application of Theorem 4 to prove Theorems 2 and 3, we choose $\lambda_1, \lambda_2, \gamma$ so that the terms in (23)-(26) decay exponentially with N, which become negligible compared to (21) and (22). Between (21) and (22), the term (21) is dominant when L does not grow with N, and (22) is dominant when L grows linearly with N.

The single threshold rule employed in the proof of Theorem 4 differs from the decoding rules employed in [9] for VLSF codes over the Gaussian MAC with expected power constraints and in [12] for the DM-MAC. In both [9] and

[12], $L = \Omega(N)$, and the decoder employs three simultaneous threshold rules for each of the boundaries that define the pentagonal achievable region of the MAC. The threshold rules are

$$\imath_2^1(X_1^{n_\ell}(m_1); Y^{n_\ell} | X_2^{n_\ell}(m_2)) > \gamma_1 \tag{28}$$

$$\imath_2^2(X_2^{n_\ell}(m_2); Y^{n_\ell} | X_1^{n_\ell}(m_1)) > \gamma_2 \tag{29}$$

$$\imath_2^{[2]}(X_1^{n_\ell}(m_1), X_2^{n_\ell}(m_2); Y^{n_\ell}) > \gamma_3 \tag{30}$$

for some γ_1, γ_2, and γ_3. Our decoder is a special case of (28)–(30), obtained by setting $\gamma_1 = \gamma_2 = -\infty$.

While [12] does not provide a second-order achievability bound, [9] could only show an achievability bound with a second-order term $-O(\sqrt{N})$ for the Gaussian MAC with expected power constraints and $L = \Omega(N)$. Employing our single threshold rule and analysis to the Gaussian MAC with expected power constraints improves the second-order term $-O(\sqrt{N})$ in [9] to $-\ln N$ for the non-corner points in the achievable region. In [9], the main challenge is to derive a tight bound on the expectation of the maximum of stopping times $\tau^{(1)}$, $\tau^{(2)}$, and $\tau^{(3)}$ for the threshold rules (28)–(30), respectively. In our analysis, we avoid that challenge by employing a single threshold decoder, whose average decoding time is bounded by $\mathbb{E}\left[\tau^{(3)}\right]$.

For the achievability of non-corner rate points that do not lie on the sum-rate boundary, we can modify the single threshold rule by replacing (30) either by (28) or by (29), depending on the location of the rate point. Following similar steps to the proof of Theorem 3, the second-order term $-\ln N$ can be achieved for those points as well.

C. DM-RAC With At Most K transmitters

To guarantee the existence of decoding times $n_{k_1,\ell_1} < n_{k_2,\ell_2}$ for any $k_1 < k_2$, (ℓ_1, ℓ_2), and a large enough M, and that the symmetrical rate point arising from the identical encoding lies on the sum-rate boundary for all $k \in [K]$, we assume that there exists an input distribution P_X that satisfies the friendliness and the interference assumptions from [19]. The *friendliness* assumption is

$$I_k(X_{[s]}; Y_k | X_{[s+1:k]} = 0^{k-s}) \geq I_k(X_{[s]}; Y_k | X_{[s+1:k]}) \tag{31}$$

for all $s < k \leq K$, and the *interference* assumption is

$$P_{X_{[t]}|Y_k} \neq P_{X_{[s]}|Y_k} P_{X_{[s+1:t]}|Y_k} \quad \forall\, s < t \leq k \leq K. \tag{32}$$

See [19, Lemmas 1 and 2] for how (31) and (32) together with permutation-invariance (8) and reducibility (9) imply the desired symmetry conditions for the DM-RAC.

In order to be able to detect the number of active transmitters using the received symbols $Y^{n_{k,\ell}}$ but not the codewords themselves, we require that the input distribution P_X satisfies the *distinguishability* assumption

$$P_{Y_{k_1}} \neq P_{Y_{k_2}} \quad \forall\, k_1 \neq k_2 \in [K], \tag{33}$$

where P_{Y_k} is the marginal output distribution under the DM-MAC with k transmitters and the input distribution $P_{X_{[k]}} = (P_X)^k$.

An example of permutation-invariant and reducible DM-RACs that satisfy the friendliness (31), interference (32), and distinguishability (33) assumptions is the adder-erasure RAC in [19], [25]

$$Y_k = \begin{cases} \sum_{i=1}^k X_i, & \text{w.p. } 1-\delta \\ \mathsf{e} & \text{w.p. } \delta, \end{cases} \tag{34}$$

where $X_i \in \{0,1\}$, $Y_k \in \{0, \ldots, k\} \cup \{\mathsf{e}\}$, and $\delta \in (0,1)$.

Theorem 5: Fix $\epsilon \in (0,1)$, finite integers $K \geq 1$ and $L \geq 2$, and a distribution P_X satisfying (31)–(33). For any permutation-invariant (8) and reducible (9) DM-RAC $\left\{(\mathcal{X}^k, P_{Y_k|X_{[k]}}, \mathcal{Y}_k)\right\}_{k=1}^K$, there exists an $(\{N_k\}_{k=1}^K, \{n_{k,\ell} \colon k \in [K], \ell \in [L]\}, M, \epsilon)$ VLSF code provided that

$$k \ln M \leq \frac{N_k I_k}{1-\epsilon} - \sqrt{N_k \ln_{(L-1)}(N_k) \frac{V_k}{1-\epsilon}} + O\left(\sqrt{\frac{N_k}{\ln_{(L-1)}(N_k)}}\right). \tag{35}$$

for $k \in [K]$.

The coding strategy to prove Theorem 5 is as follows. The decoder at time $n_{k,1}$ applies a binary composite hypothesis test using the output sequences $Y^{n_{k,1}}$ to decide whether the null hypothesis H_0 that $Y^{n_{k,1}}$ is drawn i.i.d. from P_{Y_k}, or the alternative hypothesis H_1 that $Y^{n_{k,1}}$ is drawn i.i.d. from one of the alternative distributions P_{Y_s}, $s \in [K] \setminus \{k\}$, is true. From (33), using the log-likelihood ratio test in [19, Sec. VI-C], we show that the probabilities of both type-I and type-II errors decay exponentially with $n_{k,1}$

$$P_{Y_k}[H_1] \leq \exp\{-n_{k,1} E_{k,k}\} \tag{36}$$

$$P_{Y_s}[H_0] \leq \exp\{-n_{k,1} E_{k,s}\} \quad \text{for } s \neq k, \tag{37}$$

where $E_{k,i} > 0$ for $k, i \in [K]$. If the hypothesis test outputs H_0, then the decoder decodes k messages at one of the times $n_{k,1}, \ldots, n_{k,L}$ using the VLSF code for the DM-MAC with k transmitters and L decoding times. If the hypothesis test outputs H_1, the decoder skips the decoding times $n_{k,1}, \ldots, n_{k,L}$ without decoding, and broadcasts feedback bit "0" at each of them to inform the transmitters about its decision. The proof of Theorem 5 combines the log-likelihood ratio test and the threshold rule used for VLSF codes with $k \in [K]$ transmitters.

An alternative to determining the number of active transmitters using a composite hypothesis testing applied at each of times $n_{1,1}, n_{2,1}, \ldots, n_{K,1}$ is to estimate the number of active transmitters using a K-ary hypothesis test at time $n_{1,1}$. In this case, the decoder feeds back the estimate, $\hat{k}$, to the transmitters using $\lceil \log_2 K \rceil$ bits. In this scenario, transmitters listen to the feedback signal only at times $n_{1,1}$ and $n_{\hat{k}1}, \ldots, n_{\hat{k}L}$; therefore, the number of feedback bits used is bounded by $\lceil \log_2 K \rceil + L$ rather than KL. Both the composite hypothesis test in [19] and the K-ary hypothesis test in [26] yield a probability of incorrect decision that decays exponentially with the underlying decoding time. This implies that both tests achieve the asymptotic expansion in (35).

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] R. Dobrushin, "An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback," *Problemy Kibernetiki*, vol. 8, no. 161-168, p. 5, 1962.

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

[5] 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.

[6] 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.

[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] 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.

[9] ——, "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.

[10] 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.

[11] 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.

[12] K. F. Trillingsgaard and P. Popovski, "Variable-length coding for short packets over a multiple access channel with feedback," in *11th International Symposium on Wireless Communications Systems (ISWCS)*, 2014, pp. 796–800.

[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 Trans. Communications*, vol. 64, no. 6, pp. 2245–2257, June 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] R. C. Yavas, V. Kostina, and M. Effros, "Variable-length feedback codes with several decoding times for the Gaussian channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Melbourne, VIC, Australia, July 2021, pp. 1883–1888.

[18] M. Heidari, A. Anastasopoulos, and S. S. Pradhan, "On the reliability function of discrete memoryless multiple-access channel with feedback," in *IEEE Information Theory Workshop (ITW)*, Guangzhou, China, Nov. 2018, pp. 1–5.

[19] 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.

[20] ——, "Nested sparse feedback codes for point-to-point, multiple access, and random access channels," *arXiv preprint*, Sep. 2021.

[21] ——, "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.

[22] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, June 2017, pp. 2523–2527.

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

[24] 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.

[25] M. Ebrahimi, F. Lahouti, and V. Kostina, "Coded random access design for constrained outage," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, June 2017, pp. 2732–2736.

[26] P. Moulin, "Asymptotically achievable error probabilities for multiple hypothesis testing," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Barcelona, Spain, July 2016, pp. 1541–1545.