

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

Recep Can Yavas, *Member, IEEE*, Victoria Kostina, *Senior Member, IEEE*, and Michelle Effros, *Fellow, IEEE*

Abstract—This paper investigates variable-length stop-feedback codes for memoryless channels in point-to-point, multiple access, and random access communication scenarios. The proposed codes employ L decoding times $n_1, n_2, \dots, n_L$ for the point-to-point and multiple access channels and $KL + 1$ decoding times for the random access channel with at most K active transmitters. In the point-to-point and multiple access channels, the decoder uses the observed channel outputs to decide whether to decode at each of the allowed decoding times $n_1, \dots, n_L$, at each time telling the encoder whether or not to stop transmitting using a single bit of feedback. In the random access scenario, the decoder estimates the number of active transmitters at time n_0 and then chooses among decoding times $n_{k,1}, \dots, n_{k,L}$ if it believes that there are k active transmitters. In all cases, the choice of allowed decoding times is part of the code design; given fixed value L , allowed decoding times are chosen to minimize the expected decoding time for a given codebook size and target average error probability. The number L in each scenario is assumed to be constant even when the blocklength is allowed to grow; the resulting code therefore requires only sparse feedback. The central results are asymptotic approximations of achievable rates as a function of the error probability, the expected decoding time, and the number of decoding times. A converse for variable-length stop-feedback codes with uniformly-spaced decoding times is included for the point-to-point channel.

Index Terms—Variable-length coding, multiple-access, random-access, feedback codes, sparse feedback, second-order analysis, channel dispersion, moderate deviations, sequential hypothesis testing.

I. INTRODUCTION

Although feedback does not increase the capacity of memoryless, point-to-point channels (PPCs) [3], feedback can simplify coding schemes and improve the speed of approach to capacity with blocklength. Examples that demonstrate this effect include Horstein's scheme for the binary symmetric channel (BSC) [4] and Schalkwijk and Kailath's scheme for the Gaussian channel [5], both of which leverage full channel feedback to simplify coding in the fixed-length regime. Wagner *et al.* [6] show that feedback improves the second-order term in the achievable rate as a function of blocklength for

fixed-rate coding over discrete, memoryless, point-to-point channels (DM-PPCs) that have multiple capacity-achieving input distributions giving distinct dispersions.

A. Literature Review on Variable-Length Feedback Codes

The benefits of feedback increase for codes with multiple decoding times (called variable-length or rateless codes). In [7], Burnashev shows that feedback significantly improves the optimal error exponent of variable-length codes for DM-PPCs. In [8], Polyanskiy *et al.* extend the work of Burnashev to the finite-length regime with non-vanishing error probabilities, introducing variable-length feedback (VLF) codes and deriving achievability and converse bounds on their performance. Tchamkerten and Telatar [9] show that Burnashev's optimal error exponent is achieved for a family of BSCs and Z channels, where the cross-over probability of the channel is unknown. For the BSC, Naghshvar *et al.* [10] propose a VLF coding scheme with a novel encoder called the small-enough-difference (SED) encoder and derive a non-asymptotic achievability bound. Their scheme is an alternative to Burnashev's scheme to achieve the optimal error exponent. Yang *et al.* [11] extend the SED encoder to the binary asymmetric channel, of which the BSC is a special case, and derive refined non-asymptotic achievability bounds for the binary asymmetric channel. Guo and Kostina [12] propose an instantaneous SED code for a source whose symbols progressively arrive at the encoder in real time.

The feedback in VLF codes can be limited in its amount and frequency. Here, the amount refers to how much feedback is sent from the receiver at each time feedback is available; the frequency refers to how many times feedback is available throughout the communication epoch. The extreme cases in the frequency are no feedback and feedback after every channel use. The extreme cases in the amount are full feedback and stop feedback. With full feedback, at time n_i , the receiver sends all symbols received until that time, Y^{n_i} , which can be used by the transmitter to encode the (n_{i+1}) -th symbol. With stop feedback, the receiver sends a single bit of feedback to inform the transmitter whether or not to stop transmitting. Unlike full-feedback codes, variable-length stop-feedback (VLSF) codes employ codewords that are fixed when the code is designed; that is, feedback affects how much of a codeword is sent but does not affect the codeword's value.

In [8], Polyanskiy *et al.* define VLSF codes with feedback after every channel use. The result in [8, Th. 2] shows that

Manuscript received January 31, 2023; revised October 9, 2023; accepted November 17, 2023.

When this work was completed, R. C. Yavas, V. Kostina, and M. Effros were all with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA. R. C. Yavas is currently with CNRS@CREATE, 138602, Singapore (e-mail: vkostina, effros@caltech.edu, recep.yavas@cnrsatcreate.sg). This work was supported in part by the National Science Foundation (NSF) under grant CCF-1817241 and CCF-1956386. This paper was presented in part at ISIT 2021 [1] and at ITW 2021 [2].

variable-length coding improves the first-order term in the asymptotic expansion of the maximum achievable message set size from NC to $\frac{NC}{1-\epsilon}$, where C is the capacity of the DM-PPC, N is the average decoding time (averaging is with respect to both the random message and the random noise), and ϵ is the average error probability. The second-order term achievable for VLF codes is $O(\log N)$, which means that VLF codes have zero dispersion and that the convergence to the capacity is much faster than that achieved by the fixed-length codes [13], [14]. In [15], 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, and the dispersion is no longer zero. A VLSF scenario with noisy feedback and a finite largest available decoding time is studied in [16]. For VLSF codes, Forney [17] shows an achievable error exponent that is strictly better than that of fixed-length, no-feedback codes and is strictly worse than Burnashev's error exponent for variable-length full-feedback codes. Ginzach *et al.* [18] derive the exact error exponent of VLSF codes for the BSC.

Bounds on the performance of VLSF codes that allow feedback after every channel use are derived for several network communication problems. Truong and Tan [19], [20] extend the results from [8] to the Gaussian multiple access channel (MAC) under an average power constraint. Trillingsgaard *et al.* [21] study the VLSF scenario where a common message is transmitted across a K -user discrete memoryless broadcast channel. Heidari *et al.* [22] extend Burnashev's work from the DM-PPC to the DM-MAC, deriving lower and upper bounds on the error exponents of VLF codes for the DM-MAC. Bounds on the performance of VLSF codes for the DM-MAC with an unbounded number of decoding times appear in [23]. The achievability bounds for K -transmitter MAC in [20] and [23] employ $2^K - 1$ simultaneous information density threshold rules.

While high rates of feedback are impractical for many applications — especially wireless applications on half-duplex devices — most prior work on VLSF codes (e.g., [8], [15], [19]–[23]) considers the densest feedback possible, using feedback at each of the (at most) $n_{\max}$ time steps before decoding, where $n_{\max}$ is the largest blocklength used by a given VLSF code. To consider more limited feedback scenarios, let L denote the number of potential decoding times in a VLSF code, a number that we assume to be independent of the blocklength. We further assume that feedback is available only at the L fixed decoding times $n_1, \dots, n_L$, which are fixed in the code design and known by the transmitter and receiver before the start of transmission. In [24], Kim *et al.* choose the decoding time for each message from the set $\{d, 2d, \dots, Ld\}$ for some positive integer d and $L < \infty$. In [25], Williamson *et al.* numerically optimize the values of L decoding times and employ punctured convolutional codes and a Viterbi algorithm. In [26], Vakili *et al.* introduce a sequential differential optimization (SDO) algorithm to optimize the choices of the L potential decoding times $n_1, \dots, n_L$, approximating the random decoding time τ by a Gaussian random variable.

Vakili *et al.* apply the SDO algorithm to non-binary low-density parity-check codes over binary-input, additive white Gaussian channels; the mean and variance of τ are determined through simulation. Heidarzadeh *et al.* [27] extend [26] to account for the feedback rate and apply the SDO algorithm to random linear codes over the binary erasure channel. In [28], we develop a communication strategy for a random access scenario with a total of K transmitters; in this scenario, neither the transmitters nor the receiver knows the set of active transmitters, which can vary from one epoch to the next. The code in [28] is a VLSF code with decoding times $n_0 < n_1 < \dots < n_K$. The decoder decodes messages only if it decides at that time that k out of total K transmitters are active at time n_k . It informs the transmitters about its decision by sending a one-bit signal at each time n_i until the time at which it decodes. We show that our random access channel (RAC) code with sparse stop feedback achieves performance identical in the capacity and dispersion terms to that of the best-known code without feedback for a MAC in which the set of active transmitters is known a priori. An extension of [28] to low-density parity-check codes appears in [29]. Building upon an earlier version of the present paper [1], Yang *et al.* [30] construct an integer program to minimize the upper bound on the average blocklength subject to constraints on the average error probability and the minimum gap between consecutive decoding times. By employing a combination of the Edgeworth expansion [31, Sec. XVI.4] and the Petrov expansion (Lemma 2), that paper develops an approximation to the cumulative distribution function of the information density random variable $\iota(X^n; Y^n)$; the numerical comparison of their approximation and the empirical cumulant distribution function shows that the approximation is tight even for small values of n . Their analysis uses this tight approximation to numerically evaluate the non-asymptotic achievability bound (Theorem 1, below) for the BSC, binary erasure channel, and binary-input Gaussian PPC for all $L \leq 32$. The resulting numerical results show performance that closely approaches Polyanskiy's VLSF achievability bound [8] with a relatively small L . For the binary erasure channel, [30] also proposes a new zero-error code that employs systematic transmission followed by random linear fountain coding; the proposed code outperforms Polyanskiy's achievability bound.

Sparse feedback is known to achieve the optimal error exponent for VLF codes. Yamamoto and Itoh [32] construct a two-phase scheme that achieves Burnashev's optimal error exponent [7]. Although their scheme allows an unlimited number of feedback instances and decoding times, it is sparse in the sense that feedback is available only at times $\alpha n, n, (1 + \alpha)n, 2n, \dots$ for some $\alpha \in (0, 1)$ and integer n . Lalitha and Javidi [33] show that Burnashev's optimal error exponent can be achieved by only $L = 3$ decoding times by truncating the Yamamoto–Itoh scheme.

Decoding for VLSF codes can be accomplished by running a sequential hypothesis test (SHT) on each possible message. At each increasingly larger stopping times, the SHT compares a hypothesis H_0 corresponding to a particular transmitted message to the hypothesis H_1 corresponding to the marginal distribution of the channel output. In [34], Berlin *et al.* derive

a bound on the average stopping time of an SHT. They then use this bound to derive a non-asymptotic converse bound for VLF codes. This result is an alternative proof for the converse of Burnashev's error exponent [7].

B. Contributions of This Work

Like [25]–[27], this paper studies VLSF codes under a finite constraint L on the number of decoding times. While [25]–[27] focus on practical coding and performance, our goal is to derive new achievability bounds on the asymptotic rate achievable by VLSF codes between $L = 1$ (the fixed-length regime analyzed in [13], [35]) and $L = n_{\max}$ (the classical variable-length regime defined in [8, Def. 1] where all decoding times $1, 2, \dots, n_{\max}$ are available).

Our contributions are summarized as follows.

- 1) We derive second-order achievability bounds for VLSF codes over DM-PPCs, DM-MACs, DM-RACs, and the Gaussian PPC with maximal power constraints. These bounds are presented in Theorems 2, 5, 6, and 7, respectively. In our analysis for each problem, we consider the asymptotic regime where the number of decoding times L is fixed while the average decoding time N grows without bound, i.e., $L = O(1)$ with respect to N . Each of our asymptotic bounds follows from the corresponding non-asymptotic bound that employs an information-density threshold rule with a stop-at-time-zero procedure. Asymptotically optimizing the values of the L decoding times yields the given results. By viewing the proposed decoder as a special case of SHT-based decoders, we show a more general non-asymptotic achievability bound; Theorem 8 employs an arbitrary SHT to decide whether a message is transmitted.
- 2) Linking the error probability of any given VLSF code to that of an SHT, in Theorem 9, we prove a converse bound in the spirit of the meta-converse bound from [13, Th. 27]. Analyzing the new bound with infinitely many uniformly-spaced decoding times over Cover–Thomas symmetric channels, in Theorem 3, we prove a converse bound for VLSF codes; the resulting bound is tight up to its second-order term. Unfortunately, since analyzing our meta-converse bound is challenging in the general case of an arbitrary DM-PPC and an arbitrary number L of decoding times (see [36, Th. 3.2.3] for the structure of optimal SHTs with finitely many decoding times), whether or not the second-order term is tight in the general case remains an open question.

Below, we detail these contributions. Our main result shows that for VLSF codes with $L = O(1) \geq 2$ decoding times over a DM-PPC, message set size M satisfying

$$\log M \approx \frac{NC}{1-\epsilon} - \sqrt{N \log_{(L-1)}(N)} \frac{V}{1-\epsilon} \quad (1)$$

is achievable. Here $\log_{(L)}(\cdot)$ denotes the L -fold nested logarithm, N is the average decoding time, ϵ is the average error probability, and C and V are the capacity and dispersion of the DM-PPC, respectively. Similar formulas arise for the DM-MAC and DM-RAC, where C and V are replaced by

the sum-rate mutual information and the sum-rate dispersion. The speed of convergence to $\frac{C}{1-\epsilon}$ depends on L . It is slower than the convergence to C in the fixed-length scenario, which has second-order term $O(\sqrt{N})$ [13]. The $L = 2$ case in (1) recovers the rate of convergence for the variable-length scenario without feedback, which has second-order term $O(\sqrt{N \log N})$ [8, Proof of Th. 1]; that rate is achieved with $n_1 = 0$. The nested logarithm term in (1) arises because after writing the average decoding time as $\mathbb{E}[\tau] = n_1 + \sum_{i=1}^{L-1} (n_{i+1} - n_i) \mathbb{P}[\tau > n_i]$, the decoding time choices in (22), below, satisfy $(n_{i+1} - n_i) \mathbb{P}[\tau > n_i] = o(\sqrt{n_1})$ for $i \in [L-1]$, making the effect of each decoding time on the average decoding time asymptotically similar. We then use the SDO algorithm introduced in [26] to show that our particular choice of $n_1, \dots, n_L$ is second-order optimal (see Appendix B.II). Despite the order-wise dependence of the rate of convergence on L , (1) grows so slowly with L that it suggests little benefit to choosing a large L . For example, when $L = 4$, $\sqrt{N \log_{(L-1)}(N)}$ 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 L ; bounding the benefit from above requires a converse result. We note, however, that the numerical results in [30] support our conclusion from the asymptotic achievability bound (1) that indicates that the improvement over achievable $\log M$ from L to $L+1$ decoding times diminishes as L increases.

For the PPC and MAC, the feedback rate of our code is $\frac{\ell}{n_{k,\ell}}$ if the decoding time is $n_{k,\ell}$; for the RAC, that rate becomes $\frac{\ell}{(k-1)L + \ell + 1}$ if the decoding time is $n_{k,\ell}$. In both cases, our feedback rate approaches 0 as the decoding time grows. In contrast, VLSF codes like in [8], [17] use feedback rate 1 bit per channel use. In VLSF codes for the RAC, the decoder decodes at one of the available times $n_{k,1}, n_{k,2}, \dots, n_{k,L}$ if it estimates that the number of active transmitters is $k \neq 0$; we reserve a single decoding time n_0 for decoding the possibility that no active transmitters are active. Theorem 6 extends the RAC code in [28] from $L = 1$ to any $L \geq 2$.

The converse result in Theorem 3 shows that in order to achieve (1) with evenly spaced decoding times, one needs at least $L = \Omega\left(\sqrt{\frac{N}{\log_{(L-1)}(N)}}\right)$ decoding times. In contrast, our optimized codes achieve (1) with a finite L that does not grow with the average decoding time N , which highlights the importance of optimizing the values of decoding times in a VLSF code.

Table I summarizes the literature on VLSF codes and the new results from this work, showing how they vary with the number of decoding times and the channel type.

In what follows, Section II gives notation and definitions. Sections III–VI introduce variable-length sparse stop-feedback codes for the DM-PPC, DM-MAC, DM-RAC, and the Gaussian PPC, respectively, and present our main theorems for those channel models; Section VII concludes the paper. The proofs appear in the Appendix.

TABLE I
THE PERFORMANCE OF VLSF CODES ACCORDING TO THE NUMBER OF DECODING TIMES L AND THE CHANNEL TYPE

Number of decoding times	Channel type	First-order term	Second-order term	
			Lower bound	Upper bound
Fixed-length, no-feedback ($L = 1$)	DM-PPC	NC	$-\sqrt{NV}Q^{-1}(\epsilon)$ [13]	$-\sqrt{NV}Q^{-1}(\epsilon)$ [13]
Variable-length ($1 < L < \infty$)	DM-PPC	$\frac{NC}{1-\epsilon}$	$-\sqrt{N \log_{(L-1)}(N) \frac{V}{1-\epsilon}}$ (Theorem 2)	$+O(1)$ [8]
Variable-length ($L = \infty$)	DM-PPC	$\frac{NC}{1-\epsilon}$	$-\log N + O(1)$ [8]	$+O(1)$ [8]
Fixed-length, no-feedback ($L = 1$)	DM-MAC	NI_K	$-\sqrt{NV_K}Q^{-1}(\epsilon)$ [37]	$+O(\sqrt{N})$ [38]
Variable-length ($1 < L < \infty$)	DM-MAC	$\frac{NI_K}{1-\epsilon}$	$-\sqrt{N \log_{(L-1)}(N) \frac{V_K}{1-\epsilon}}$ (Theorem 5)	$+O(1)$ [20]
Variable-length ($L = \infty$)	DM-MAC	$\frac{NI_K}{1-\epsilon}$	$-\log N + O(1)$ eq. (44)	$+O(1)$ [20]
Variable-length ($L = \infty$)	Gaussian MAC (average power)	$\frac{NC(KP)}{1-\epsilon}$	$-O(\sqrt{N})$ [20]	$+O(1)$ [20]
($L = 1$)	DM-RAC	$N_k I_k$	$-\sqrt{N_k V_k}Q^{-1}(\epsilon_k)$ [28]	$+O(\sqrt{N_k})$ [38]
($L = 1$)	Gaussian RAC (maximal power)	$N_k C(kP)$	$-\sqrt{N_k V_k(P)}Q^{-1}(\epsilon_k)$ [39]	$+O(\sqrt{N_k})$ [38]
($1 < L < \infty$)	DM-RAC	$\frac{N_k I_k}{1-\epsilon_k}$	$-\sqrt{N_k \log_{(L-1)}(N_k) \frac{V_k}{1-\epsilon_k}}$ (Theorem 6)	$+O(1)$ [20]
Variable-length ($1 < L < \infty$)	Gaussian PPC (maximal power)	$\frac{NC(P)}{1-\epsilon}$	$-\sqrt{N \log_{(L-1)}(N) \frac{V(P)}{1-\epsilon}}$ (Theorem 7)	$+O(1)$ [19]
Variable-length ($L = \infty$)	Gaussian PPC (average power)	$\frac{NC(P)}{1-\epsilon}$	$-\log N + O(1)$ [19]	$+O(1)$ [19]

II. PRELIMINARIES

A. Notation

For any positive integers k and n , $[k] \triangleq \{1, \dots, k\}$, $x^n \triangleq (x_1, \dots, x_n)$, and $x^{a:b} \triangleq (x_a, x_{a+1}, \dots, x_b)$. The collection of length- n vectors from the transmitter index set $\mathcal{A}$ is denoted by $x_{\mathcal{A}}^n \triangleq (x_a^n: a \in \mathcal{A})$; we drop the superscript n if $n = 1$, i.e., $x_{\mathcal{A}}^1 = x_{\mathcal{A}}$. The collection of non-empty strict subsets of a set $\mathcal{A}$ is denoted by $\mathcal{P}(\mathcal{A}) \triangleq \{\mathcal{B}: \mathcal{B} \subseteq \mathcal{A}, 0 < |\mathcal{B}| < |\mathcal{A}|\}$. All-zero and all-one vectors are denoted by $\mathbf{0}$ and $\mathbf{1}$, respectively; dimension is determined from the context. The sets of positive integers and non-negative integers are denoted by $\mathbb{Z}_+$ and $\mathbb{Z}_{\geq}$, respectively. We write $x^n \stackrel{\pi}{=} y^n$ if there exists a permutation π of x^n such that $\pi(x^n) = y^n$, and $x^n \not\stackrel{\pi}{=} y^n$ if such a permutation does not exist. The identity matrix of dimension n is denoted by I_n . The Euclidean norm of vector x^n is denoted by $\|x^n\| \triangleq \sqrt{\sum_{i=1}^n x_i^2}$. Unless specified otherwise, all logarithms and exponents have base e . Information is measured in nats. The standard $O(\cdot)$, $o(\cdot)$, and $\Omega(\cdot)$ notations are defined as $f(n) = O(g(n))$ if $\limsup_{n \rightarrow \infty} |f(n)/g(n)| < \infty$, $f(n) = o(g(n))$ if $\lim_{n \rightarrow \infty} |f(n)/g(n)| = 0$, and $f(n) = \Omega(g(n))$

if $\lim_{n \rightarrow \infty} |f(n)/g(n)| > 0$. The distribution of a random variable X is denoted by P_X ; $\mathcal{N}(\mu, V)$ denotes the Gaussian distribution with mean μ and covariance matrix V , $Q(\cdot)$ represents the complementary standard Gaussian cumulative distribution function $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left\{-\frac{t^2}{2}\right\} dt$, and $Q^{-1}(\cdot)$ is its functional inverse. We define the nested logarithm function

$$\log_{(L)}(x) \triangleq \begin{cases} \log(x) & \text{if } L = 1, x > 0 \\ \log(\log_{(L-1)}(x)) & \text{if } L \geq 2, \log_{(L-1)}(x) > 0; \end{cases} \quad (2)$$

$\log_{(L)}(x)$ is undefined for all other (L, x) pairs.

We denote the Radon-Nikodym derivative of distribution P with respect to distribution Q by $\frac{dP}{dQ}$. We denote the relative entropy and relative entropy variance between P and Q by $D(P\|Q) = \mathbb{E}\left[\log \frac{dP}{dQ}(X)\right]$ and $V(P\|Q) = \text{Var}\left[\log \frac{dP}{dQ}(X)\right]$, respectively, where $X \sim P$. The σ -algebra generated by random variable X is denoted by $\mathcal{F}(X)$. A random variable X is called arithmetic if there exists some

$d > 0$ such that $\mathbb{P}[X \in d\mathbb{Z}] = 1$. The largest d that satisfies this condition is called the span. If such a d does not exist, then the random variable is non-arithmetic. Denote $X^+ \triangleq \max\{0, X\}$ and $X^- \triangleq -\min\{0, X\}$ for any random variable X .

B. Discrete Memoryless Channel and Information Density

A DM-PPC is defined by the triple $(\mathcal{X}, P_{Y|X}, \mathcal{Y})$, where $\mathcal{X}$ is the finite input alphabet, $P_{Y|X}$ is the channel transition kernel, and $\mathcal{Y}$ is the finite output alphabet. The n -letter input-output relationship of the channel is given by $P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n P_{Y|X}(y_i|x_i)$ for all n , x^n , and y^n .

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

$$\imath(x^n; y^n) \triangleq \log \frac{P_{Y^n|X^n}(y^n|x^n)}{P_{Y^n}(y^n)}, \quad (3)$$

where P_{Y^n} is the Y^n marginal of $P_{X^n}P_{Y^n|X^n}$. If the inputs $X_1, X_2, \dots, X_n$ are independently and identically distributed (i.i.d.) according to P_X , then

$$\imath(x^n; y^n) = \sum_{i=1}^n \imath(x_i; y_i), \quad (4)$$

where the single-letter information density is given by

$$\imath(x; y) \triangleq \log \frac{P_{Y|X}(y|x)}{P_Y(y)}, \quad x \in \mathcal{X}, y \in \mathcal{Y}. \quad (5)$$

The mutual information and dispersion are defined as

$$I(X; Y) \triangleq \mathbb{E}[\imath(X; Y)] \quad (6)$$

$$V(X; Y) \triangleq \text{Var}[\imath(X; Y)], \quad (7)$$

respectively, where $(X, Y) \sim P_X P_{Y|X}$.

Let $\mathcal{P}$ denote all distributions on the alphabet $\mathcal{X}$. The capacity of the DM-PPC is

$$C = \max_{P_X \in \mathcal{P}} I(X; Y), \quad (8)$$

and the dispersion of the DM-PPC is

$$V = \min_{P_X \in \mathcal{P}: I(X; Y) = C} V(X; Y). \quad (9)$$

III. VLSF CODES FOR THE DM-PPC

A. VLSF Codes with L Decoding Times

We consider VLSF codes with a finite number of potential decoding times $n_1 < n_2 < \dots < n_L$ over a DM-PPC. The receiver chooses to end the transmission at the first time $n_\ell \in \{n_1, \dots, n_L\}$ that it is ready to decode. The transmitter learns of the receiver's decision via a single bit of feedback at each of times $n_1, \dots, n_\ell$. Feedback bit "0" at time n_i means that the receiver is not yet ready to decode, and transmission should continue; feedback bit "1" means that the receiver can decode at time n_i , which signals the transmitter to stop. Using this feedback, the transmitter and the receiver are synchronized and aware of the current state of the transmission at all times. Since n_L is the last decoding time available, the receiver always makes a final decision if time n_L is reached. Unlike [7], [25], [32], we do not allow re-transmission of the message after

time n_L . Since the transmitter and the receiver both know the values of decoding times, the receiver does not need to send feedback at the last available time n_L . We assume that the transmitter and the receiver know the channel transition kernel $P_{Y|X}$. We employ average decoding time and average error probability constraints. Definition 1, below, formalizes our code description.

Definition 1: Fix $\epsilon \in (0, 1)$, positive integers L and M , and a positive scalar N . An (N, L, M, ϵ) -VLSF code for the DM-PPC comprises

- 1) non-negative integer-valued decoding times $n_1 < \dots < n_L$,
- 2) 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,¹
- 3) an encoding function $\mathbf{f}_n: \mathcal{U} \times [M] \rightarrow \mathcal{X}$, for each $n = 1, \dots, n_L$, that assigns a codeword

$$\mathbf{f}(u, m)^{n_L} \triangleq (\mathbf{f}_1(u, m), \dots, \mathbf{f}_{n_L}(u, m)) \quad (10)$$

to each message $m \in [M]$ and common randomness instance $u \in \mathcal{U}$,

- 4) a non-negative integer-valued random stopping time $\tau \in \{n_1, \dots, n_L\}$ for the filtration generated by $\{U, Y^{n_i}\}_{i=1}^L$ that satisfies an average decoding time constraint

$$\mathbb{E}[\tau] \leq N, \quad (11)$$

- 5) and a decoding function $\mathbf{g}_{n_\ell}: \mathcal{U} \times \mathcal{Y}^{n_\ell} \rightarrow [M] \cup \{\mathbf{e}\}$ for each $\ell \in [L]$ (where $\mathbf{e}$ is the erasure symbol used to indicate that the receiver is not ready to decode), satisfying an average error probability constraint

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

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

Recall that Definition 1 with $L = 1$ recovers the fixed-length no-feedback codes in [13]. As in [8], [21], [28], we here need common randomness because the traditional random-coding argument does not prove the existence of a single (deterministic) code that simultaneously satisfies conditions (11) and (12) on the code. Therefore, randomized codes are necessary for our achievability argument; here, $|\mathcal{U}| \leq 2$ suffices [28, Appendix D].

We define the maximum achievable message set size $M^*(N, L, \epsilon)$ with L decoding times, average decoding time N , and average error probability ϵ as

$$M^*(N, L, \epsilon) \triangleq \max\{M: \text{an } (N, L, M, \epsilon) \text{ VLSF code exists}\}. \quad (13)$$

The maximum achievable message set size for VLSF codes with L decoding times $n_1, \dots, n_L$ that are restricted to belong to a subset $\mathcal{N} \subseteq \mathbb{Z}_{\geq}$ is denoted by $M^*(N, L, \epsilon, \mathcal{N})$.

¹The realization u of U specifies the codebook.

B. Related Work

The following discussion summarizes prior asymptotic expansions of the maximum achievable message set size for the DM-PPC.

- a) $M^*(N, 1, \epsilon)$: For $L = 1$ and $\epsilon \in (0, 1/2)$, Polyanskiy *et al.* [13, Th. 49] show that

$$\log M^*(N, 1, \epsilon) = NC - \sqrt{NV}Q^{-1}(\epsilon) + O(\log N). \quad (14)$$

For $\epsilon \in [1/2, 1)$, the dispersion V in (9) is replaced by the maximum dispersion $V_{\max} \triangleq \max_{P_X: \mathcal{I}(X;Y)=C} V(X;Y)$.

The $O(\log N)$ term is lower bounded by $O(1)$ and upper bounded by $\frac{1}{2} \log N + O(1)$. For nonsingular DM-PPCs, i.e., the channels that satisfy $\mathbb{E}[\mathcal{I}(X;Y)|Y] > 0$ for the distributions that achieve the capacity C and the dispersion V , the $O(\log N)$ term is equal to $\frac{1}{2} \log N + O(1)$ [40]. Moulin [41] derives lower and upper bounds on the $O(1)$ term in the asymptotic expansion when the channel is nonsingular with non-lattice information density.

- b) $M^*(N, \infty, \epsilon)$: For VLSF codes with $L = n_{\max} = \infty$, Polyanskiy *et al.* [8, Th. 2] show that for $\epsilon \in (0, 1)$,

$$\log M^*(N, \infty, \epsilon) \geq \frac{NC}{1-\epsilon} - \log N + O(1) \quad (15)$$

$$\log M^*(N, \infty, \epsilon) \leq \frac{NC}{1-\epsilon} + \frac{h_b(\epsilon)}{1-\epsilon}, \quad (16)$$

where $h_b(\epsilon) \triangleq -\epsilon \log \epsilon - (1-\epsilon) \log(1-\epsilon)$ is the binary entropy function (in nats). The bounds in (15)–(16) indicate that the ϵ -capacity (the first-order achievable term) is

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log M^*(N, \infty, \epsilon) = \frac{C}{1-\epsilon}. \quad (17)$$

The achievable dispersion term is zero, i.e., the second-order term in the fundamental limit in (15)–(16) is $o(\sqrt{N})$.

C. Our Achievability Bounds

Theorem 1, below, is our non-asymptotic achievability bound for VLSF codes with L decoding times.

Theorem 1: Fix a constant γ , decoding times $n_1 < \dots < n_L$, and a positive integer M . For any positive number N and $\epsilon \in (0, 1)$, there exists an (N, L, M, ϵ) -VLSF code for the DM-PPC $(\mathcal{X}, P_{Y|X}, \mathcal{Y})$ with

$$\epsilon \leq \mathbb{P}[\mathcal{I}(X^{n_L}; Y^{n_L}) < \gamma] + (M-1) \exp\{-\gamma\}, \quad (18)$$

$$N \leq n_1 + \sum_{\ell=1}^{L-1} (n_{\ell+1} - n_\ell) \mathbb{P}[\mathcal{I}(X^{n_\ell}; Y^{n_\ell}) < \gamma], \quad (19)$$

where $P_{X^{n_L}}$ is a product of distributions of L sub-vectors of lengths $n_j - n_{j-1}$, $j \in [L]$, i.e.,

$$P_{X^{n_L}}(x^{n_L}) = \prod_{j=1}^L P_{X^{n_{j-1}+1:n_j}}(x^{n_{j-1}+1:n_j}), \quad (20)$$

where $n_0 = 0$.

Proof sketch: Polyanskiy *et al.* [13] interpret the information-density threshold test for a fixed-length code as a collection of hypothesis tests aimed at determining whether the channel output is (H_0) or is not (H_1) dependent on a given

codeword. In our coding scheme, we use SHTs in a similar way. The strategy is as follows.

The VLSF decoder at each time $n_1, \dots, n_L$ runs M SHTs between a hypothesis H_0 that the channel output results from transmission of the m -th codeword, $m \in [M]$, and the hypothesis H_1 that the channel output is drawn from the unconditional channel output distribution. The former indicates that the decoder hypothesizes that message m is the sent message. The latter indicates that the decoder hypothesizes that message m has not been sent and thus can be removed from the list of possible messages to decode. Transmission stops at the first time n_i that hypothesis H_0 is accepted for some message m or the first time n_i that hypothesis H_1 is accepted for all m . If the latter happens, decoding fails and we declare an error. Transmission continues as long as one of the SHTs has not accepted either H_0 or H_1 . If H_0 is declared for multiple messages at the same decoding time, then we stop and declare an error. Since n_L is the last available decoding time, the SHTs are forced to decide between H_0 and H_1 at time n_L . Once H_0 or H_1 is decided for some message, the decision cannot be reversed at a later time.

The optimal SHT has the form of a two-sided information density threshold rule, where the thresholds depend on the individual decision times [36, Th. 3.2.3]. To simplify the analysis, we employ sub-optimal SHTs for which the upper threshold is set to a value $\gamma \in \mathbb{R}$ that is independent of the decoding times and the lower thresholds are set to $-\infty$ for $n_\ell < n_L$ and to γ for $n_\ell = n_L$. That is, we declare H_1 for a message if and only if the corresponding information density never reaches γ at any of decoding times $n_1, \dots, n_L$. Theorem 1 analyzes the error probability and the average decoding time of the sub-optimal SHT-based decoder above, and extends the achievability bound in [8, Th. 3] that considers $L = \infty$ to the scenario where only a finite number of decoding times is allowed. The bound on the average decoding time (19) is obtained by expressing the bound on the average decoding time in [8, eq. (27)] using the fact that the stopping time τ is in $\{n_1, \dots, n_L\}$. When we compare Theorem 1 with [8, Th. 3], we see that the error probability bound in (18) has an extra term $\mathbb{P}[\mathcal{I}(X^{n_L}; Y^{n_L}) < \gamma]$. This term appears since transmission always stops at or before time n_L .

Theorem 1 is related to [24, Lemma 1], which similarly treats $L < \infty$ but requires $n_{\ell+1} - n_\ell = d$ for some constant $d \geq 1$, and [25, Cor. 2], where the transmitter retransmits the message if decoding attempts at times $n_1, \dots, n_L$ are unsuccessful.

See Appendix A for the proof details. ■

Theorem 2, stated next, is our second-order achievability bound for VLSF codes with $L = O(1)$ decoding times over the DM-PPC. The proof of Theorem 2 builds upon the non-asymptotic bound in Theorem 1.

Theorem 2: Fix an integer $L = O(1) \geq 2$ and real numbers $N > 0$ and $\epsilon \in (0, 1)$. For the DM-PPC with $V > 0$, the maximum message set size (13) achievable by (N, L, M, ϵ) -VLSF codes satisfies

$$\log M^*(N, L, \epsilon) \geq \frac{NC}{1-\epsilon} - \sqrt{N \log_{(L-1)}(N) \frac{V}{1-\epsilon}}$$

$$+ O\left(\sqrt{\frac{N}{\log_{(L-1)}(N)}}\right). \quad (21)$$

The decoding times $\{n_1, \dots, n_L\}$ that achieve (21) satisfy the equations

$$\log M = n_\ell C - \sqrt{n_\ell \log_{(L-\ell+1)}(n_\ell) V} - \log n_\ell + O(1) \quad (22)$$

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

Proof sketch: Inspired by [8, Th. 2], the proof employs a time-sharing strategy between an $(N', L-1, M, \epsilon'_N)$ -VLSF code whose smallest decoding time is nonzero and a simple “stop-at-time-zero” procedure that does not involve any code and decodes an error at time 0. Specifically, we set the VLSF code as the one that achieves the bound in Theorem 1, and we use the VLSF code and the stop-at-time-zero procedure with probabilities $1-p$ and p , respectively, where p and ϵ'_N satisfy

$$\epsilon'_N = \frac{1}{\sqrt{N' \log N'}} \quad (23)$$

$$p = \frac{\epsilon - \epsilon'_N}{1 - \epsilon'_N} \quad (24)$$

The error probability of the resulting code is bounded by ϵ , and the average decoding time is

$$N = N'(1-p) = N'(1-\epsilon) + O\left(\sqrt{\frac{N'}{\log N'}}\right). \quad (25)$$

For the scenario where $L = \infty$, we again use time-sharing with the stop-at-time-zero procedure in the achievability bound in [8, Th. 2] with $\epsilon'_N = \frac{1}{\sqrt{N'}}$ instead of (23). In the asymptotic regime $L = O(1)$, the choice in (23) results in a better second-order term than that achieved by $\epsilon'_N = \frac{1}{\sqrt{N'}}$.

In the analysis of Theorem 1, we need to bound the probability $\mathbb{P}[\iota(X^{n_L}; Y^{n_L}) < \gamma] = \epsilon'_N(1 - o(1))$. Since this probability decays sub-exponentially to zero due to (23), we use a moderate deviations result from [42, Ch. 8] to bound this probability. Such a tool was not needed in the proof of [8, Th. 2] for $L = \infty$ because when $n_L = \infty$, the term $\mathbb{P}[\iota(X^{n_L}; Y^{n_L}) < \gamma]$ disappears from (18), and the average decoding time is bounded via martingale analysis instead of (19). Finally, we apply Karush-Kuhn-Tucker conditions to show that the decoding times in (22) yield a value of $\log M$ that is the maximal value achievable by the non-asymptotic bound up to terms of order $O\left(\sqrt{\frac{N}{\log_{(L-1)}(N)}}\right)$. The details of the proof appear in Appendix B. ■

The non-asymptotic achievability bounds obtained from the coding scheme described in the proof sketch of Theorem 2 are illustrated for the BSC in Fig. 1. For $L \in \{2, 3, 4\}$, the decoding times $n_1, \dots, n_L$ are chosen as described in (22) with the $O(1)$ term ignored, and ϵ'_N in the stop-at-time-zero procedure is replaced with the right-hand side of (18). For $L = 1$, Fig. 1 shows the random coding union bound in [13, Th. 16], which is a non-asymptotic achievability bound for fixed-length no-feedback codes. For $L = \infty$, Fig. 1 shows the non-asymptotic bound in [8, eq. (102)]. The curves for $L = 1$ and $L = 2$ cross because the choice of decoding times in (22) requires $\epsilon \gg \frac{1}{\sqrt{N \log N}}$ and is optimal only as $N \rightarrow \infty$. In

[30], Yang *et al.* construct a computationally intensive integer program for the numerical optimization of the decoding times for finite N . If such a precise optimization is desired, our approximate decoding times in (22) can be used as starting locations for that integer program.

Replacing the information-density-based decoding rule in the proof sketch with the optimal SHT would improve the performance achieved on the right-hand side of (21) by only $O(1)$.

Since any (N, L, M, ϵ) -VLSF code is also an (N, ∞, M, ϵ) -VLSF code, (16) provides an upper bound on $\log M^*(N, L, \epsilon)$ for an arbitrary L . The order of the second-order term, $-\sqrt{N \log_{(L-1)}(N) \frac{V}{1-\epsilon}}$, depends on the number of decoding times L . The larger L , the faster the achievable rate converges to the capacity. However, the dependence on L is weak since $\log_{(L-1)}(N)$ grows very slowly in N even if L is small. For example, for $L = 4$ and $N = 1000$, $\log_{(L-1)}(N) \approx 0.659$. For a finite L , this bound falls short of the $-\log N$ achievability bound in (16) achievable with $L = \infty$. Whether the second-order term achieved in Theorem 2 is tight remains an open problem.

The following theorem gives achievability and converse bounds for VLSF codes with decoding times uniformly spaced as $\{0, d_N, 2d_N, \dots\}$.

Theorem 3: Fix $\epsilon \in (0, 1)$. Let $d_N = o(N)$ with $d_N \rightarrow \infty$, and let $P_{Y|X}$ be any DM-PPC. Then, it holds that

$$\log M^*(N, \infty, \epsilon, d_N \mathbb{Z}_{\geq}) \geq \frac{NC}{1-\epsilon} - \frac{d_N C}{2} - \log N + o(d_N). \quad (26)$$

If the DM-PPC $P_{Y|X}$ is a Cover-Thomas symmetric DM-PPC [43, p. 190] i.e., the rows (and resp. the columns) of the transition probability matrix are permutations of each other, then

$$\log M^*(N, \infty, \epsilon, d_N \mathbb{Z}_{\geq}) \leq \frac{NC}{1-\epsilon} - \frac{d_N C}{2} + o(d_N). \quad (27)$$

Proof Sketch: The achievability bound (26) employs the sub-optimal SHT in the proof sketch of Theorem 2. To prove the converse in (27), we first derive in Theorem 9, in Appendix C below, the meta-converse bound for VLSF codes. The meta-converse bound in Theorem 9 bounds the error probability of any given VLSF code from below by the minimum achievable type-II error probability of the corresponding SHT; it is an extension and a tightening of Polyanskiy *et al.*'s converse in (16) since for $d_N = 1$, weakening it by applying a loose bound on the performance of SHTs from [36, Th. 3.2.2] recovers (16). The Cover-Thomas symmetry assumption allows us to circumvent the maximization of that minimum type-II error probability over codes since the log-likelihood ratio $\log \frac{P_{Y|X}(Y|x)}{P_Y(Y)}$ is the same regardless of the channel input x for that channel class. In both bounds in (26)–(27), we use the expansions for the average stopping time and the type-II error probability from [36, Ch. 2-3]. See Appendix C for details. ■

Theorem 3 establishes that when $\frac{d_N}{\log N} \rightarrow \infty$, the second-order term of the logarithm of maximum achievable message set size among VLSF codes with uniformly spaced decoding

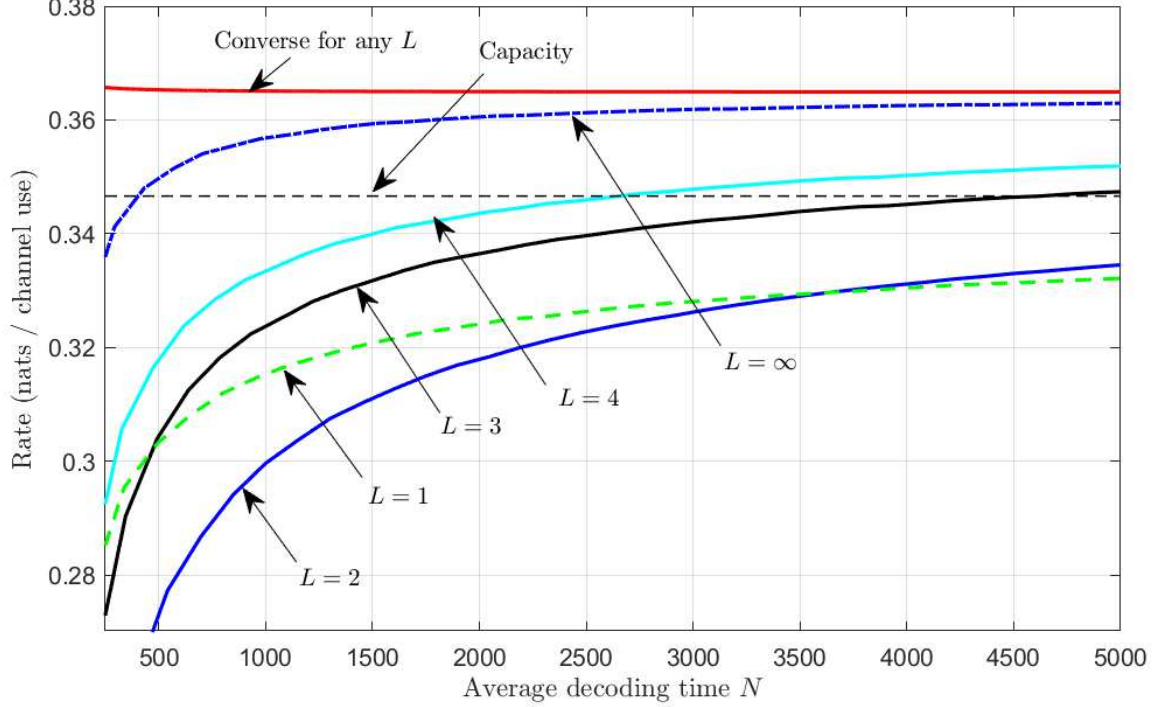


Fig. 1. The non-asymptotic achievability bounds obtained from Theorems 1 and 2 and the non-asymptotic converse bound (16) for the maximum achievable rate $\frac{\log M^*(N, L, \epsilon)}{N}$ are shown for the BSC with crossover probability 0.11, $L \in \{1, 2, 3, 4, \infty\}$, and $\epsilon = 0.05$. The curves that $L = 1$ and $L = \infty$ are Polyanskiy *et al.*'s achievability bounds from [13, Th. 16] and [8, eq. (102)], respectively.

times is $-\frac{d_N C}{2}$. Theorem 3 implies that in order to achieve the same performance as achieved in (21) with L decoding times, one needs on average $\Omega\left(\sqrt{\frac{N}{\log_{(L-1)}(N)}}\right)$ uniformly spaced stop-feedback instances, suggesting that the optimization of available decoding times considered in Theorem 2 is crucial for attaining the second-order term in (21).

The case where $d_N = \Omega(N)$ is not as interesting as the case where $d_N = o(N)$ since analyzing Theorem 9 using Chernoff bound would yield a bound on the probability that the optimal SHT makes a decision at times other than $n_1 = 0$ and $\frac{N}{1-\epsilon}(1+o(1))$. Since that probability decays exponentially with N , the scenario where L is unbounded and $d_N = \Omega(N)$ is asymptotically equivalent to $L = 2$. For example, for $d_N = \frac{1}{\ell} \frac{N}{1-\epsilon} \left(1 + O\left(\frac{1}{\sqrt{N \log N}}\right)\right)$ for some $\ell \in \mathbb{Z}_+$, the right-hand side of (21) is tight up to the second-order term.

IV. VLSF CODES FOR THE DM-MAC

We begin by introducing the definitions used for the multi-transmitter setting.

A. Definitions

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

In what follows, the subscript and superscript indicate the corresponding transmitter indices and the codeword lengths,

respectively. 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 for each non-empty $\mathcal{A} \subseteq [K]$ as

$$\iota_K(x_{\mathcal{A}}; y) \triangleq \log \frac{P_{Y_K|X_{\mathcal{A}}}(y|x_{\mathcal{A}})}{P_{Y_K}(y)} \quad (28)$$

$$\iota_K(x_{\mathcal{A}}; y|x_{\mathcal{A}^c}) \triangleq \log \frac{P_{Y_K|X_{[K]}}(y|x_{[K]})}{P_{Y_K|X_{\mathcal{A}^c}}(y|x_{\mathcal{A}^c})}, \quad (29)$$

where $\mathcal{A}^c = [K] \setminus \mathcal{A}$. Note that in (28)–(29), the information density functions depend on the transmitter set $\mathcal{A}$ unless further symmetry conditions are assumed (e.g., in some cases we assume that the components of $P_{X_{[K]}}$ are i.i.d., and $P_{Y_K|X_{[K]}}$ is invariant to permutations of the inputs $X_{[K]}$).

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

$$I_K(X_{\mathcal{A}}; Y_K) \triangleq \mathbb{E}[\iota_K(X_{\mathcal{A}}; Y_K)] \quad (30)$$

$$I_K(X_{\mathcal{A}}; Y_K|X_{\mathcal{A}^c}) \triangleq \mathbb{E}[\iota_K(X_{\mathcal{A}}; Y_K|X_{\mathcal{A}^c})]. \quad (31)$$

The dispersions are defined as

$$V_K(X_{\mathcal{A}}; Y_K) \triangleq \text{Var}[\iota_K(X_{\mathcal{A}}; Y_K)] \quad (32)$$

$$V_K(X_{\mathcal{A}}; Y_K|X_{\mathcal{A}^c}) \triangleq \text{Var}[\iota_K(X_{\mathcal{A}}; Y_K|X_{\mathcal{A}^c})]. \quad (33)$$

For brevity, we define

$$I_K \triangleq I_K(X_{[K]}; Y_K) \quad (34)$$

$$V_K \triangleq \text{Var}[\iota_K(X_{[K]}; Y_K)]. \quad (35)$$

A VLSF code for the MAC with K transmitters is defined similarly to the VLSF code for the PPC.

Definition 2: Fix $\epsilon \in (0, 1)$, $N \in (0, \infty)$, and positive integers $M_k, k \in [K]$. An $(N, L, M_{[K]}, \epsilon)$ -VLSF code for the MAC comprises

- 1) non-negative integer-valued decoding times $n_1 < \dots < n_L$,
- 2) K finite alphabets $\mathcal{U}_k, k \in [K]$, defining common randomness random variables $U_1, \dots, U_K$,
- 3) K sequences of encoding functions $f_n^{(k)}: \mathcal{U}_k \times [M_k] \rightarrow \mathcal{X}_k, k \in [K]$,
- 4) a stopping time $\tau \in \{n_1, \dots, n_L\}$ for the filtration generated by $\{U_1, \dots, U_K, Y_K^{n_\ell}\}_{\ell=1}^L$, satisfying an average decoding time constraint (11), and
- 5) L decoding functions $g_{n_\ell}: \mathcal{U}_{[K]} \times \mathcal{Y}_K^{n_\ell} \rightarrow \prod_{k=1}^K [M_k] \cup \{e\}$ for $\ell \in [L]$, satisfying an average error probability constraint

$$\mathbb{P}[g_\tau(U_{[K]}, Y_K^\tau) \neq W_{[K]}] \leq \epsilon, \quad (36)$$

where the independent messages $W_1, \dots, W_K$ are uniformly distributed on the sets $[M_1], \dots, [M_K]$, respectively.

B. Our Achievability Bounds

Our main results are second-order achievability bounds for the rates approaching a point on the sum-rate boundary of the MAC achievable region expanded by a factor of $\frac{1}{1-\epsilon}$.

Theorem 4, below, is a non-asymptotic achievability bound for any DM-MAC with K transmitters and L decoding times.

Theorem 4: Fix constants $\epsilon \in (0, 1)$, $\gamma \in \mathbb{R}$, $\lambda^{(A)} > 0$ for $\mathcal{A} \in \mathcal{P}([K])$, integers $0 \leq n_1 < \dots < n_L$, and distributions $P_{X_k}, k \in [K]$. For any DM-MAC with K transmitters $(\prod_{k=1}^K \mathcal{X}_k, P_{Y_K|X_{[K]}}, \mathcal{Y}_K)$, there exists an $(N, L, M_{[K]}, \epsilon)$ -VLSF code with

$$\epsilon \leq \mathbb{P}[\iota_K(X_{[K]}^{n_L}; Y_K^{n_L}) < \gamma] \quad (37)$$

$$+ \prod_{k=1}^K (M_k - 1) \exp\{-\gamma\} \quad (38)$$

$$+ \sum_{\ell=1}^L \sum_{\mathcal{A} \in \mathcal{P}([K])} \mathbb{P}[\iota_K(X_{\mathcal{A}}^{n_\ell}; Y_K^{n_\ell}) > N(I_K(X_{\mathcal{A}}; Y) + \lambda^{(A)})] \quad (39)$$

$$+ \sum_{\mathcal{A} \in \mathcal{P}([K])} \left(\prod_{k \in \mathcal{A}^c} (M_k - 1) \right) \exp\{-\gamma + NI_K(X_{\mathcal{A}}; Y_K) + N\lambda^{(A)}\} \quad (40)$$

$$N \leq n_1 + \sum_{\ell=1}^{L-1} (n_{\ell+1} - n_\ell) \mathbb{P}[\iota_K(X_{[K]}^{n_\ell}; Y_K^{n_\ell}) < \gamma]. \quad (41)$$

Proof sketch: The proof of Theorem 4 uses a random coding argument that employs K independent codebook ensembles, each with distribution $P_{X_k}^{n_L}, k \in [K]$. The receiver employs L decoders that operate by comparing an information density $\iota_K(x_{[K]}^{n_\ell}; y^{n_\ell})$ for each possible transmitted codeword set to a threshold. At time n_ℓ , decoder g_{n_ℓ} computes the information

densities $\iota_K(X_{[K]}^{n_\ell}(m_{[K]}); Y_K^{n_\ell})$; if there exists a unique message vector $\hat{m}_{[K]}$ satisfying $\iota_K(X_{[K]}^{n_\ell}(\hat{m}_{[K]}); Y_K^{n_\ell}) > \gamma$, then the receiver decodes to the message vector $\hat{m}_{[K]}$; if there exists multiple such message vectors, then the receiver stops the transmission and decodes an error. If no such message vectors exist at time n_ℓ , then the receiver emits output e and passes the decoding time n_ℓ without decoding if $n_\ell < n_L$ and decodes an error if $n_\ell = n_L$. The term (37) bounds the probability that the information density corresponding to the true messages is below the threshold for all decoding times; (38) bounds the probability that all messages are decoded incorrectly; and (39)-(40) bound the probability that the messages from the transmitter index set $\mathcal{A} \subseteq [K]$ are decoded incorrectly, and the messages from the index set $\mathcal{A}^c$ are decoded correctly. The proof of Theorem 4 appears in Appendix D. ■

Theorem 5, below, is a second-order achievability bound in the asymptotic regime $L = O(1)$ for any DM-MAC. It follows from an application of Theorem 4.

Theorem 5: Fix $\epsilon \in (0, 1)$, an integer $L = O(1) \geq 2$, and distributions $P_{X_k}, k \in [K]$. For any K -transmitter DM-MAC $(\prod_{k=1}^K \mathcal{X}_k, P_{Y_K|X_{[K]}}, \mathcal{Y}_K)$, there exists a K -tuple $M_{[K]}$ and an $(N, L, M_{[K]}, \epsilon)$ -VLSF code satisfying

$$\sum_{k \in [K]} \log M_k = \frac{NI_K}{1-\epsilon} - \sqrt{N \log_{(L-1)}(N) \frac{V_K}{1-\epsilon}} + O\left(\sqrt{\frac{N}{\log_{(L-1)}(N)}}\right). \quad (42)$$

Proof: See Appendix D. ■

In the application of Theorem 4 to prove Theorem 5, we choose the parameters $\lambda^{(A)}$ and γ so that the terms in (39)-(40) decay exponentially with N , which become negligible compared to (37) and (38). Between (37) and (38), the term (37) is dominant when L does not grow with N , and (38) is dominant when L grows linearly with N .

Like the single-threshold rule from [28] for the RAC, the single-threshold rule employed in the proof of Theorem 4 differs from the decoding rules employed in [20] for VLSF codes over the Gaussian MAC with expected power constraints and in [23] for the DM-MAC. In both [20] and [23], $L = n_{\max} = \infty$, and the decoder employs $2^K - 1$ simultaneous threshold rules for each of the boundaries that define the achievable region of the MAC with K transmitters. Those rules fix thresholds $\gamma^{(A)}, \mathcal{A} \in \mathcal{P}([K])$, and decode messages $m_{[K]}$ if for all $\mathcal{A} \in \mathcal{P}([K])$, the codeword for $m_{[K]}$ satisfies

$$\iota_K(X_{\mathcal{A}}^{n_\ell}(m_{\mathcal{A}}); Y_K^{n_\ell}|X_{\mathcal{A}^c}^{n_\ell}(m_{\mathcal{A}^c})) > \gamma^{(A)}, \quad (43)$$

for some $\gamma^{(A)}, \mathcal{A} \in \mathcal{P}([K])$. Our decoder can be viewed as a special case of (43) obtained by setting $\gamma^{(A)} = -\infty$ for $\mathcal{A} \neq [K]$.

Analyzing Theorem 4 in the asymptotic regime $L = \Omega(N)$, we determine that there exists a K -tuple $M_{[K]}$ and an $(N, \infty, M_{[K]}, \epsilon)$ -VLSF code satisfying

$$\sum_{k \in [K]} \log M_k = \frac{NI_K}{1-\epsilon} - \log N + O(1). \quad (44)$$

Both (42) and (44) are achieved at rate points that approach a point on the sum-rate boundary of the K -MAC achievable region expanded by a factor of $\frac{1}{1-\epsilon}$.

For any VLSF code, $L = \infty$ case can be treated as $L = \Omega(N)$ regardless of the number of transmitters since if we truncate an infinite-length code at time $n_{\max} = 2N$, by Chernoff bound, the resulting penalty term added to the error probability decays exponentially with N , whose effect in (44) is $o(1)$. See Appendix D.III for the proof of (44).

For $L = n_{\max} = \infty$, Trillinggaard and Popovski [23] numerically evaluate their non-asymptotic achievability bound for a DM-MAC while Truong and Tan [20] provide an achievability bound with second-order term $-O(\sqrt{N})$ for the Gaussian MAC with average power constraints. Applying our single-threshold rule and analysis to the Gaussian MAC with average power constraints improves the second-order term in [20] from $-O(\sqrt{N})$ to $-\log N + O(1)$ for all non-corner points in the achievable region. The main challenge in [20] is to derive a tight bound on the expected value of the maximum over $\mathcal{A} \subseteq [K]$ of stopping times $\tau^{(\mathcal{A})}$ for the corresponding threshold rules in (43). In our analysis, we avoid that challenge by employing a single-threshold decoder whose average decoding time is bounded by $\mathbb{E}[\tau^{([K])}]$.

Under the same model and assumptions on L , to achieve non-corner rate points that do not lie on the sum-rate boundary, we modify our single-threshold rule to (43), where $\mathcal{A}$ is the transmitter index set corresponding to the capacity region's active sum-rate bound at the (non-corner) point of interest. Following steps similar to the proof of (44) gives second-order term $-\log N + O(1)$ for those points as well. For corner points, more than one boundary is active²; therefore, more than one threshold rule in (43) is needed at the decoder. In this case, again for $L = \infty$, [20] proves an achievability bound with a second-order term $-O(\sqrt{N})$. Whether this bound can be improved to $-\log N + O(1)$ as in (44) remains an open problem.

V. VLSF CODES FOR THE DM-RAC WITH AT MOST K TRANSMITTERS

Definition 3 (Yavas et al. [28, eq. (1)]): A permutation-invariant, reducible DM-RAC for the maximal number of transmitters $K < \infty$ is defined by a family of DM-MACs $\left\{ \left(\mathcal{X}^k, P_{Y_k|X_{[k]}}, \mathcal{Y}_k \right) \right\}_{k=0}^K$, where the k -th DM-MAC defines the channel for k active transmitters.

By assumption, each of the DM-MACs satisfies the *permutation-invariance* condition

$$P_{Y_k|X_{[k]}}(y|x_{[k]}) = P_{Y_k|X_{[k]}}(y|x_{\pi[k]}) \quad (45)$$

for all permutations $\pi[k]$ of $[k]$, and $y \in \mathcal{Y}_k$, and the *reducibility* condition

$$P_{Y_s|X_{[s]}}(y|x_{[s]}) = P_{Y_k|X_{[k]}}(y|x_{[s]}, 0^{k-s}) \quad (46)$$

²The capacity region of a K -transmitter MAC is characterized by the region bounded by $2^K - 1$ planes. By definition of a corner point, at least two inequalities corresponding to these planes are active at a corner point.

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.

The permutation-invariance (45) and reducibility (46) conditions simplify the presentation and enable us to show, using a single-threshold rule at the decoder [28], that the symmetrical rate point $(R, R, \dots, R)$ at which the code operates lies on the sum-rate boundary of each of the underlying DM-MACs,

The VLSF RAC code defined here combines our rateless communication strategy from [28] with the sparse feedback VLSF PPC and MAC codes with optimized average decoding times described above. Specifically, the decoder estimates the value of k at time n_0 . If the estimate $\hat{k}$ is not zero, it decodes at one of the L decoding times $n_{\hat{k},1} < n_{\hat{k},2} < \dots < n_{\hat{k},L}$ (rather than just the single time $n_{\hat{k}}$ used in [28], [39]). For every $k \in [K]$, the locations of the L decoding times are optimized to attain the minimum average decoding delay. As in [28], we do not assume any probability distribution on the user activity pattern. We seek instead to optimize the rate-reliability trade-off simultaneously for all possible activity patterns. (By the permutation-invariance assumption, there are only K distinguishable activity patterns to consider here indexed by the number of active transmitters.) If the decoder concludes that no transmitters are active, then it ends the transmission at time n_0 decoding no messages. At each time $n_{i,\ell}$, $i < \hat{k}$, the receiver broadcasts “0” to the transmitters, signaling that they should continue to transmit. At time $n_{\hat{k},\ell}$, the receiver broadcasts feedback bit “1” to the transmitters if it is able to decode $\hat{k}$ messages; otherwise, it outputs an erasure symbol “e” and sends feedback bit “0”, again signaling that decoding has not occurred and transmission should continue.

As in [28], [39], we assume that 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 [44], 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 $i \in [K]$. As in [28], [44], the decoder is required to decode the list of messages transmitted by the active transmitters but not the identities of these transmitters.

To deal with the scenario where the number of transmitters in the RAC grows linearly with the blocklength, i.e., $K = \Omega(N)$, [44] employs the per-user error probability (PUPE) constraint rather than the joint error probability used here and in the analysis of the MAC (e.g., [20], [28], [39]). The PUPE is a weaker error probability constraint since, under PUPE, an error at one decoder does not count as an error at all other decoders. In [28], it is shown that when $K = O(1)$, PUPE and joint error probability constraints have the same second-order performance for random access coding. As a result, there is no advantage to using PUPE rather than the more stringent joint error criterion when $K = O(1)$. Therefore, we employ the joint error probability constraint throughout.

We formally define VLSF codes for the RAC as follows.

Definition 4: Fix $\epsilon_0, \dots, \epsilon_K \in (0, 1)$ and $N_0, \dots, N_K \in (0, \infty)$. An $(\{N_k\}_{k=0}^K, L, M, \{\epsilon_k\}_{k=0}^K)$ -VLSF code with identical encoders comprises

- 1) a set of integers $\mathcal{N} \triangleq \{n_0\} \cup \{n_{k,\ell} : k \in [K], \ell \in [L]\}$

(without loss of generality, we assume that $n_{K,L}$ is the largest available decoding time),

- 2) a common randomness random variable U on an alphabet $\mathcal{U}$,
- 3) a sequence of encoding functions $f_n: \mathcal{U} \times [M] \rightarrow \mathcal{X}$, $n = 1, 2, \dots, n_{K,L}$, defining M length- $n_{K,L}$ codewords,
- 4) 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

$$\mathbb{E}[\tau_k] \leq N_k \quad (47)$$

if $k \in \{0\} \cup [K]$ transmitters are active, and

- 5) $KL + 1$ decoding functions $g_{n_0}: \mathcal{U} \times \mathcal{Y}_0^{n_0} \rightarrow \{\emptyset\} \cup \{\mathbf{e}\}$ and $g_{n_{k,\ell}}: \mathcal{U} \times \mathcal{Y}_k^{n_{k,\ell}} \rightarrow [M]^k \cup \{\mathbf{e}\}$, $k \in [K]$ and $\ell \in [L]$, satisfying an average error probability constraint

$$\mathbb{P}\left[g_{\tau_k}(U, Y_k^{\tau_k}) \neq W_{[k]}\right] \leq \epsilon_k \quad (48)$$

when $k \in [K]$ messages $W_{[k]} = (W_1, \dots, W_k)$ are transmitted, where $W_1, \dots, W_k$ are independent and equiprobable on the set $[M]$, and

$$\mathbb{P}[g_{\tau_0}(U, Y_0^{\tau_0}) \neq \emptyset] \leq \epsilon_0 \quad (49)$$

when no transmitters are active.

To guarantee that the symmetrical rate point arising from identical encoding lies on the sum-rate boundary for all $k \in [K]$, following [28], we assume that there exists an input distribution P_X that satisfies the interference assumptions

$$P_{X_{[t]}|Y_k} \neq P_{X_{[s]}|Y_k} P_{X_{[s+1:t]}|Y_k} \quad \forall s < t \leq k \leq K. \quad (50)$$

Permutation-invariance (45), reducibility (46), and interference (50) together imply that the mutual information per transmitter, $\frac{I_k}{k}$, strictly decreases with increasing k (see [28, Lemma 1]). This property guarantees the existence of decoding times satisfying $n_{k_1, \ell_1} < n_{k_2, \ell_2}$ for any $k_1 < k_2$ and $\ell_1, \ell_2 \in [L]$.

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

$$P_{Y_{k_1}} \neq P_{Y_{k_2}} \quad \forall k_1 \neq k_2 \in \{0\} \cup [K], \quad (51)$$

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

An example of a permutation-invariant and reducible DM-RAC that satisfies interference (50) and distinguishability (51) assumptions is the adder-erasure RAC in [28], [45]

$$Y_k = \begin{cases} \sum_{i=1}^k X_i, & \text{w.p. } 1 - \delta \\ \mathbf{e} & \text{w.p. } \delta, \end{cases} \quad (52)$$

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

Theorem 6: Fix $\epsilon \in (0, 1)$, finite integers $K \geq 1$ and $L \geq 2$, and a distribution P_X satisfying (50)–(51). For any permutation-invariant (45) and reducible (46)

DM-RAC $\left\{(\mathcal{X}^k, P_{Y_k|X_{[k]}}, \mathcal{Y}_k)\right\}_{k=0}^K$, there exists an $(\{N_k\}_{k=0}^K, L, M, \{\epsilon_k\}_{k=0}^K)$ -VLSF code satisfying

$$k \log M = \frac{N_k I_k}{1 - \epsilon_k} - \sqrt{N_k \log_{(L-1)}(N_k) \frac{V_k}{1 - \epsilon_k}}$$

$$+ O\left(\sqrt{\frac{N_k}{\log_{(L)}(N_k)}}\right) \quad (53)$$

for $k \in [K]$, and

$$N_0 = c \log N_1 + o(\log N_1) \quad (54)$$

for some $c > 0$.

Proof sketch: The coding strategy to prove Theorem 6 is as follows. The decoder applies a $(K+1)$ -ary hypothesis test using the output sequence Y^{n_0} and decides an estimate $\hat{k}$ of the number of active transmitters $k \in \{0, 1, \dots, K\}$. If the hypothesis test declares that $\hat{k} = 0$, then the receiver stops the transmission at time n_0 , decoding no messages. If $\hat{k} \neq 0$, then the receiver decodes $\hat{k}$ messages at one of the times $n_{\hat{k},1}, \dots, n_{\hat{k},L}$ using the VLSF code in Theorem 5 for the $\hat{k}$ -transmitter DM-MAC with L decoding times. If the receiver decodes at time $n_{\hat{k},\ell}$, then it sends feedback bit ‘0’ at all previous decoding times $\{n \in \mathcal{N}: n < n_{\hat{k},\ell}\}$ and feedback bit ‘1’ at time $n_{\hat{k},\ell}$. Note that alternatively, the receiver can send its estimate $\hat{k}$ using $\lceil \log_2(K+1) \rceil + L$ bits at time n_0 , informing the transmitters that it will decode at some time $\{n_{\hat{k},1}, \dots, n_{\hat{k},L}\}$; in this case, the number of feedback bits decreases from the worst-case $KL + 1$ that results from the strategy described above. The details of the proof appear in Appendix E. ■

VI. VLSF CODES FOR THE GAUSSIAN PPC WITH MAXIMAL POWER CONSTRAINTS

A. Gaussian PPC

The output of a memoryless, Gaussian PPC of blocklength n in response to the input $X^n \in \mathbb{R}^n$ is

$$Y^n = X^n + Z^n, \quad (55)$$

where $Z_1, \dots, Z_n$ are drawn i.i.d. from $\mathcal{N}(0, 1)$, independent of X^n .

The channel’s capacity $C(P)$ and dispersion $V(P)$ are

$$C(P) = \frac{1}{2} \log(1 + P) \quad (56)$$

$$V(P) = \frac{P(P+2)}{2(1+P)^2}. \quad (57)$$

B. Related Work on the Gaussian PPC

We first introduce the maximal and average power constraints on VLSF codes for the PPC. Given a VLSF code with L decoding times $n_1, \dots, n_L$, the maximal power constraint requires that the length- n prefixes, $n \in \{n_1, \dots, n_L\}$, of each codeword all satisfy a power constraint P , i.e.,

$$\|f(u, m)^{n_\ell}\|^2 \leq n_\ell P \quad \text{for all } m \in [M], u \in \mathcal{U}, \ell \in [L]. \quad (58)$$

The average power constraint on the length- n_L codewords, as defined by [20, Def. 1], is

$$\mathbb{E}[\|f(U, W)^{n_L}\|^2] \leq NP. \quad (59)$$

The definitions of $(N, L, M, \epsilon, P)_{\max}$ and $(N, L, M, \epsilon, P)_{\text{ave}}$ -VLSF codes for the Gaussian PPC are similar to Definition 1 with the addition of maximal (58) and average (59) power

constraints, respectively. Similar to (13), $M^*(N, L, \epsilon, P)_{\max}$ (resp. $M^*(N, L, \epsilon, P)_{\text{ave}}$) denotes the maximum achievable message set size with L decoding times, average decoding time N , average error probability ϵ , and maximal (resp. average) power constraint P .

In the following, we discuss prior asymptotic expansions of $M^*(N, L, \epsilon, P)_{\max}$ and $M^*(N, L, \epsilon, P)_{\text{ave}}$ for the Gaussian PPC, where $L \in \{1, \infty\}$.

- a) $M^*(N, 1, \epsilon, P)_{\max}$: For $L = 1$, $P > 0$, and $\epsilon \in (0, 1)$, Tan and Tomamichel [35, Th. 1] and Polyanskiy *et al.* [13, Th. 54] show that

$$\begin{aligned} \log M^*(N, 1, \epsilon, P)_{\max} \\ = NC(P) - \sqrt{NV(P)}Q^{-1}(\epsilon) + \frac{1}{2} \log N + O(1). \end{aligned} \quad (60)$$

The converse for (60) is derived in [13, Th. 54] and the achievability for (60) in [35, Th. 1]. The achievability scheme in [35, Th. 1] 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 (60) is $O(1)$.

- b) $M^*(N, 1, \epsilon, P)_{\text{ave}}$: For $L = 1$ with an average-power-constraint, Yang *et al.* show in [46] that

$$\begin{aligned} \log M^*(N, 1, \epsilon, P)_{\text{ave}} = NC\left(\frac{P}{1-\epsilon}\right) - \\ - \sqrt{N \log N V\left(\frac{P}{1-\epsilon}\right)} + O(\sqrt{N}). \end{aligned} \quad (61)$$

Yang *et al.* use a power control argument to show the achievability of (61). 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 $\left(N, 1, |\mathcal{A}|, \frac{2}{\sqrt{N \log N}}, \frac{P}{1-\epsilon}(1 - o(1))\right)$ -VLSF code with a single decoding time N . The codewords are generated i.i.d. uniformly on the sphere with center at 0 and 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 (61) follows from an application of the meta-converse [13, Th. 26].

- c) $M^*(N, \infty, \epsilon, P)_{\text{ave}}$: For VLSF codes with $L = n_{\max} = \infty$ and average power constraint (59), Truong and Tan show in [19, Th. 1] that for $\epsilon \in (0, 1)$ and $P > 0$,

$$\log M^*(N, \infty, \epsilon, P)_{\text{ave}} \geq \frac{NC(P)}{1-\epsilon} - \log N + O(1) \quad (62)$$

$$\log M^*(N, \infty, \epsilon, P)_{\text{ave}} \leq \frac{NC(P)}{1-\epsilon} + \frac{h_b(\epsilon)}{1-\epsilon}, \quad (63)$$

where h_b is the binary entropy function. The results in (62)–(63) are analogous to the fundamental limits for DM-PPCs (15)–(16) and follow from arguments similar to those in [8]. Since the information density $\imath(X; Y)$ for the Gaussian channel is unbounded, bounding the expected value of the decoding time in the proof of [19, Th. 1] requires different techniques from those applicable to DM-PPCs [8].

Table II combines the $L = 1$ summary from [47, Table I] with the corresponding results for $L > 1$ to summarize the performance of VLSF codes for the Gaussian channel in different communication scenarios.

C. Main Result

The theorem below is our main result for the Gaussian PPC under the maximal power constraint (58).

Theorem 7: Fix an integer $L = O(1) \geq 2$ and real numbers $P > 0$ and $\epsilon \in (0, 1)$. For the Gaussian channel with maximal power constraint (58), the maximum message set size achievable by (N, L, M, ϵ, P) -VLSF codes satisfies

$$\begin{aligned} \log M^*(N, L, \epsilon, P)_{\max} \geq \frac{NC(P)}{1-\epsilon} - \sqrt{N \log_{(L-1)}(N) \frac{V(P)}{1-\epsilon}} \\ + O\left(\sqrt{\frac{N}{\log_{(L-1)}(N)}}\right). \end{aligned} \quad (64)$$

The decoding times that achieve (64) satisfy the equations

$$\begin{aligned} \log M^*(N, L, \epsilon, P) \\ = n_\ell C(P) - \sqrt{n_\ell \log_{(L-\ell+1)}(n_\ell) V(P)} - \log n_\ell + O(1) \end{aligned} \quad (65)$$

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

Proof: See Appendix F. ■

Note that the achievability bound in Theorem 7 has the same form as the one in Theorem 2 with C and V replaced with the Gaussian capacity $C(P)$ and the Gaussian dispersion $V(P)$, respectively. The bound in (64) holds for the average power constraint as well since any code that satisfies the maximal power constraint also satisfies the average power constraint.

From Shannon's work in [48], it is 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. As a result, 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 the maximal/average power constraint and feedback/no feedback [14], [35], [46], [47] in Table I) employ random codewords drawn uniformly at random on the power sphere. A notable exception is Truong and Tan's result in (62) [19, Th. 1], which considers VLSF codes with an average power constraint; that result employs i.i.d. Gaussian inputs. The Gaussian distribution works in this scenario because when $L = \infty$, the usually dominant term $\mathbb{P}[\imath(X^{n_L}; Y^{n_L}) < \gamma]$ in (18) disappears. The second term $(M - 1) \exp\{-\gamma\}$ in (18) is not affected by the input distribution. Unfortunately, the approach from [19, Th. 1] does not work here since drawing codewords i.i.d. $\mathcal{N}(0, P)$ satisfies the average power constraint (59) but not the maximal power constraint (58). When $L = O(1)$ and the probability $\mathbb{P}[\imath(X^{n_L}; Y^{n_L}) < \gamma]$ dominates, using i.i.d. $\mathcal{N}(0, P)$ inputs achieves a worse second-order term in the asymptotic expansion (64) of the maximum achievable message set size. For the case $L = O(1)$, we draw codewords according to the rule that the sub-codewords indexed from $n_{j-1} + 1$ to n_j are drawn uniformly on the $(n_j - n_{j-1})$ -dimensional sphere of radius

TABLE II
THE PERFORMANCE OF VLSF CODES FOR THE GAUSSIAN CHANNEL IN SCENARIOS DISTINGUISHED BY THE NUMBER OF AVAILABLE DECODING TIMES L , THE TYPE OF THE POWER CONSTRAINT, AND THE PRESENCE OF FEEDBACK.

			First-order term	Second-order term	
				Lower Bound	Upper Bound
Fixed-length ($L = 1$)	No Feedback	Max. power	$NC(P)$	$-\sqrt{NV(P)}Q^{-1}(\epsilon)$ ([13], [35])	$-\sqrt{NV(P)}Q^{-1}(\epsilon)$ ([13])
		Ave. power	$NC\left(\frac{P}{1-\epsilon}\right)$	$-\sqrt{N \log NV\left(\frac{P}{1-\epsilon}\right)}$ ([46])	$-\sqrt{N \log NV\left(\frac{P}{1-\epsilon}\right)}$ ([46])
	Feedback	Max. power	$NC(P)$	$-\sqrt{NV(P)}Q^{-1}(\epsilon)$ ([13], [35])	$-\sqrt{NV(P)}Q^{-1}(\epsilon)$ ([14])
		Ave. power	$NC\left(\frac{P}{1-\epsilon}\right)$	$-O(\log_{(L)}(N))$ ([47])	$+\sqrt{N \log NV\left(\frac{P}{1-\epsilon}\right)}$ ([47])
Variable-length ($L < \infty$)	Max. power		$\frac{NC(P)}{1-\epsilon}$	$-\sqrt{N \log_{(L-1)}(N) \frac{V(P)}{1-\epsilon}}$ (Theorem 7)	$+O(1)$ ([19])
	Ave. power		$\frac{NC(P)}{1-\epsilon}$	$-\sqrt{N \log_{(L-1)}(N) \frac{V(P)}{1-\epsilon}}$ (Theorem 7)	$+O(1)$ ([19])
Variable-length ($L = n_{\max} = \infty$)	Max. power		$\frac{NC(P)}{1-\epsilon}$	$-O(\sqrt{N})$ ([1])	$+O(1)$ ([19])
	Ave. power		$\frac{NC(P)}{1-\epsilon}$	$-\log N$ ([19])	$+O(1)$ ([19])

$\sqrt{(n_j - n_{j-1})P}$ for $j \in [L]$, independently of each other. Note that this input distribution is dispersion-achieving for the fixed-length no-feedback case, i.e., $L = 1$ [13] and is superior to choosing codewords i.i.d. $\mathcal{N}(0, P)$, even under the average power constraint. In particular, i.i.d. $\mathcal{N}(0, P)$ inputs achieve (21), where the dispersion $V(P)$ is replaced by the variance $\tilde{V}(P) = \frac{P}{1+P}$ 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 [49, eq. (2.56)]). Whether or not our input distribution is optimal in the second-order term remains an open question.

VII. CONCLUSIONS

This paper investigates the maximum achievable message set size for sparse VLSF codes over the DM-PPC (Theorem 2), DM-MAC (Theorem 5), DM-RAC (Theorem 6), and Gaussian PPC (Theorem 7) in the asymptotic regime where the number of decoding times L is constant as the average decoding time N grows without bound. Under our second-order achievability bounds, the performance improvement due to adding more decoding time opportunities to our code quickly diminishes as L increases. For example, for the BSC with crossover probability 0.11, at average decoding time $N = 2000$, our VLSF coding bound with only $L = 4$ decoding times achieves 96% of the rate of Polyanskiy *et al.*'s VLSF coding bound for $L = \infty$. Incremental redundancy automatic repeat request codes, which are some of the most common feedback codes, employ only a small number of decoding times and stop feedback. Our analysis shows that such a code design is not only practical but also has performance competitive with the best known dense feedback codes.

In all channel types considered, the first-order term in our achievability bounds is $\frac{NC}{1-\epsilon}$, where N is the average decoding time, ϵ is the error probability, and C is the capacity (or the sum-rate capacity in the multi-transmitter case), and the second-order term is $O\left(\sqrt{N \log_{(L-1)}(N)}\right)$. For DM-

PPCs, there is a mismatch between the second-order term of our achievability bound for VLSF codes with $L = O(1)$ decoding times (Theorem 2) and the second-order term of the best known converse bound (16); the latter applies to $L = \infty$, and therefore to any L . Towards closing the gap between the achievability and converse bounds, in Theorem 9 in Appendix C, below, we derive a non-asymptotic converse bound that links the error probability of a VLSF code with the minimum achievable type-II error probability of an SHT. However, since the threshold values of the optimal SHT with L decoding times do not have a closed-form expression [36, pp. 153-154], analyzing the non-asymptotic converse bound in Theorem 9 is a difficult task. Whether the second-order term in Theorem 2 is optimal is a question left to future work.

In sparse VLSF codes, optimizing the values of L available decoding times is important since to achieve the same performance as $L = O(1)$ optimized decoding times (Theorem 2), one needs $\Omega\left(\sqrt{\frac{N}{\log_{(L-1)}(N)}}\right)$ uniformly spaced decoding times (Theorem 3).

APPENDIX A PROOF OF THEOREM 1

In this section, we derive an achievability bound based on a general SHT, which we use to prove Theorem 1.

A.1. A General SHT-based Achievability Bound

1) *SHT definitions:* We begin by formally defining an SHT. We extend the definition in [36, Ch. 3] to allow non-i.i.d. distributions and finitely many testing times. Let $\{Z_i\}_{i=1}^{n_L}$ be the observed sequence. Consider two hypotheses for the distribution of Z^{n_L}

$$H_0: Z^{n_L} \sim P_0 \quad (66)$$

$$H_1: Z^{n_L} \sim P_1, \quad (67)$$

where P_0 and P_1 are distributions on a common alphabet $\mathcal{Z}^{n_L}$. Let $\mathcal{N} \subseteq \{0, 1, 2, \dots, n_L\}$ be the set of times that the hypothesis is tested. Let $P_i^{(n_\ell)}$ denote the marginal distribution of the first n_ℓ symbols in P_i , $i \in \{0, 1\}$. At time $n_\ell \in \mathcal{N}$, we either decide $H_0: Z^{n_\ell} \sim P_0^{(n_\ell)}$, $H_1: Z^{n_\ell} \sim P_1^{(n_\ell)}$, or we wait until the next available time $n_{\ell+1}$ in $\mathcal{N}$. Let τ be a stopping time adapted to the filtration $\{\mathcal{F}(X^n)\}_{n \in \mathcal{N}}$. Let δ be a $\{0, 1\}$ -valued, $\mathcal{F}(\tau)$ -measurable function. An SHT is a triple $(\delta, \tau, \mathcal{N})$, where δ is called the decision rule, τ is called the stopping time, and $\mathcal{N}$ is the set of available decision times. Type-I and type-II error probabilities are defined as

$$\alpha \triangleq \mathbb{P}[\delta = 1 | H_0] \quad (68)$$

$$\beta \triangleq \mathbb{P}[\delta = 0 | H_1]. \quad (69)$$

Below, we derive an achievability using a general SHT.

2) *Achievability Bound:* Given some input distribution $P_{X^{n_L}}$, define the common randomness random variable U on $\mathbb{R}^{M n_L}$ with the distribution

$$P_U = \underbrace{P_{X^{n_L}} \times P_{X^{n_L}} \times \dots \times P_{X^{n_L}}}_{M \text{ times}}. \quad (70)$$

The realization of U defines M length- n_L codewords $X^{n_L}(1), X^{n_L}(2), \dots, X^{n_L}(M)$. Denote the set of available decoding times by

$$\mathcal{N} \triangleq \{n_1, \dots, n_L\}. \quad (71)$$

Let $\{(\delta_m, \tilde{\tau}_m, \mathcal{N})\}_{m=1}^M$ be M copies of an SHT that distinguishes between the hypotheses

$$H_0: (X^{n_L}, Y^{n_L}) \sim P_{X^{n_L}} \times P_{Y|X}^{n_L} \quad (72)$$

$$H_1: (X^{n_L}, Y^{n_L}) \sim P_{X^{n_L}} \times P_{Y^{n_L}} \quad (73)$$

for each message $m \in [M]$, where the type-I and type-II error probabilities are α and β , respectively. Define for $m \in [M]$ and $j \in \{0, 1\}$,

$$\tau_m^j \triangleq \begin{cases} \tilde{\tau}_m & \text{if } \delta_m = j \\ \infty & \text{otherwise.} \end{cases} \quad (74)$$

Theorem 8, below, is an achievability bound that employs an arbitrary SHT with L decoding times.

Theorem 8: Fix $L \leq \infty$, integers $M > 0$ and $0 \leq n_1 < n_2 < \dots < n_L \leq \infty$, a distribution $P_{X^{n_L}}$ as in (70), and M copies of an SHT $\{(\delta_m, \tilde{\tau}_m, \{n_1, \dots, n_L\})\}_{m=1}^M$ as in (72)–(74). There exists an (N, L, M, ϵ) -VLSF code for the DM-PPC $(\mathcal{X}, P_{Y|X}, \mathcal{Y})$ with

$$\epsilon \leq \alpha + (M - 1)\beta \quad (75)$$

$$N \leq \mathbb{E} \left[\min \left\{ \min_{m \in [M]} \{\tau_m^0\}, \max_{m \in [M]} \{\tau_m^1\} \right\} \right]. \quad (76)$$

Proof: We generate M i.i.d. codewords according to (70). For each of M messages, we run the hypothesis test given in (72)–(73). We decode at the earliest time that one of the following events happens

- H_0 is declared for some message $m \in [M]$,
- H_1 is declared for all $m \in [M]$.

The decoding output is m if H_0 is declared for m ; if there exist more than one such m or if there exists no such m , the decoder declares an error.

Mathematically, the random decoding time of this code is expressed as

$$\tau^* = \min \left\{ \min_{m \in [M]} \{\tau_m^0\}, \max_{m \in [M]} \{\tau_m^1\} \right\}. \quad (77)$$

Note that τ^* is bounded by n_L by construction. The average decoding time bound in (76) immediately follows from (77). The decoder output is

$$\hat{W} \triangleq \begin{cases} m & \text{if } \exists! m \in [M] \text{ s. t. } \tau^* = \tau_m^0 \\ \text{error} & \text{otherwise.} \end{cases} \quad (78)$$

Since the messages are equiprobable, without loss of generality, assume that message $m = 1$ is transmitted. An error occurs if and only if H_1 is decided for $m = 1$ or if H_0 is decided for some $m \neq 1$, giving

$$\epsilon = \mathbb{P} \left[\{\delta_1 = 1\} \cup \left\{ \bigcup_{m=2}^M \{\delta_m = 0\} \right\} \right]. \quad (79)$$

Applying the union bound to (79) shows (75). ■

A.II. Proof of Theorem 1

Theorem 1 particularizes the SHT in Theorem 8 as an information density threshold rule.

In addition to the random code design in (70), let $P_{X^{n_L}}$ satisfy (20). We here specify the stopping rule τ_m and the decision rule δ_m for the SHT in (72)–(73).

Define the information density for message m and decoding time n_ℓ as

$$S_{m, n_\ell} \triangleq \iota(X^{n_\ell}(m); Y^{n_\ell}) \text{ for } m \in [M], \ell \in [L]. \quad (80)$$

Note that S_{m, n_ℓ} is the log-likelihood ratio between the distributions in hypotheses H_0 and H_1 . We fix a threshold $\gamma \in \mathbb{R}$ and construct the SHTs

$$\tau_m = \inf \{n_\ell \in \mathcal{N} : S_{m, n_\ell} \geq \gamma\} \quad (81)$$

$$\tilde{\tau}_m = \min \{\tau_m, n_L\} \quad (82)$$

$$\delta_m = \begin{cases} 0 & \text{if } S_{m, \tilde{\tau}_m} \geq \gamma \\ 1 & \text{if } S_{m, \tilde{\tau}_m} < \gamma \end{cases} \quad (83)$$

for all $m \in [M]$, that is, we decide H_0 for message m at the first time n_ℓ that S_{m, n_ℓ} passes γ ; if this never happens for $n_\ell \in \{n_1, \dots, n_L\}$, then we decide H_1 for m . Without loss of generality, assume that message 1 is transmitted.

Bounding (76) from above, we get

$$N \leq \mathbb{E}[\min\{\tau_1, n_L\}] \quad (84)$$

$$= \sum_{n=0}^{\infty} \mathbb{P}[\min\{\tau_1, n_L\} > n] \quad (85)$$

$$= n_1 + \sum_{\ell=1}^{L-1} (n_{\ell+1} - n_\ell) \mathbb{P}[\tau_1 > n_\ell]. \quad (86)$$

The probability $\mathbb{P}[\tau_1 > n_\ell]$ is further bounded as

$$\mathbb{P}[\tau_1 > n_\ell] = \mathbb{P}\left[\bigcap_{j=1}^{\ell} \{i(X^{n_j}(1); Y^{n_j}) < \gamma\}\right] \quad (87)$$

$$\leq \mathbb{P}[i(X^{n_\ell}(1); Y^{n_\ell}) < \gamma]. \quad (88)$$

Combining (86) and (88) proves (19).

We bound the type-I error probability of the given SHT as

$$\alpha \triangleq \mathbb{P}[\delta_1 = 1] \quad (89)$$

$$= \mathbb{P}[\tau_1 = \infty] \quad (90)$$

$$= \mathbb{P}\left[\bigcap_{j=1}^L \{i(X^{n_j}(1); Y^{n_j}) < \gamma\}\right] \quad (91)$$

$$\leq \mathbb{P}[i(X^{n_L}(1); Y^{n_L}) < \gamma], \quad (92)$$

where (91) uses the definition of the decision rule (83). The type-II error probability is bounded as

$$\beta \triangleq \mathbb{P}[\delta_2 = 0] \quad (93)$$

$$\leq \mathbb{P}[\tau_2 < \infty] \quad (94)$$

$$= \mathbb{E}[\exp\{-i(X^{n_L}(1); Y^{n_L})\} 1\{\tau_1 < \infty\}] \quad (95)$$

$$= \mathbb{E}[\exp\{-i(X^\tau(1); Y^\tau)\} 1\{\tau_1 < \infty\}] \quad (96)$$

$$\leq \exp\{-\gamma\}, \quad (97)$$

where (95) follows from changing measure from $P_{X^{n_L}(2)Y^{n_L}} = P_{X^{n_L}}P_{Y^{n_L}}$ to $P_{X^{n_L}(1), Y^{n_L}} = P_{X^{n_L}}P_{Y^{n_L}|X}$. Equality (96) uses the same arguments as in [8, eq. (111)-(118)] and the fact that $\{\exp\{-i(X^{n_\ell}(1); Y^{n_\ell})\} : n_\ell \in \mathcal{N}\}$ is a martingale due to the product distribution in (20). Inequality (97) follows from the definition of τ_1 in (81). Applying (75) together with (92) and (97) proves (18). ■

In his analysis of the error exponent regime, Forney [17] uses a slightly different threshold rule than the one in (81). Specifically, he uses a maximum a posteriori threshold rule, which can also be written as

$$\log \frac{P_{Y^{n_\ell}|X^{n_\ell}}(Y^{n_\ell}|X^{n_\ell}(m))}{\frac{1}{M} \sum_{j=1}^M P_{Y^{n_\ell}|X^{n_\ell}}(Y^{n_\ell}|X^{n_\ell}(j))} \geq \gamma, \quad (98)$$

whose denominator is the output distribution induced by the code rather than by the random codeword distribution $P_X^{n_\ell}$.

APPENDIX B PROOF OF THEOREM 2

The proof uses an idea that is similar to that in [13, Th. 2], which combines the achievability bound of a VLSF code with a sub-exponentially decaying error probability with the stop-at-time-zero procedure. The difference is that we set the sub-exponentially decaying error probability as $\epsilon'_N = \frac{1}{\sqrt{N' \log N'}}$ while [13, Th. 2] sets it to $\frac{1}{N'}$. This modification yields a better second-order term for finite L .

Inverting (25), we get

$$N' = \frac{N}{1 - \epsilon} \left(1 + O\left(\frac{1}{\sqrt{N \log N}}\right)\right). \quad (99)$$

Next, we particularize the decision rules in the SHT at times $n_2, \dots, n_L$ to the information density threshold

rule. Lemma 1, below, is an achievability bound for an $\left(N, L, M, \frac{1}{\sqrt{N \log N}}\right)$ -VLSF code that employs the information density threshold rule with the optimized decoding times and the threshold value.

Lemma 1: Fix an integer $L = O(1) \geq 1$. For the DM-PPC with $V > 0$, the maximum message set size (13) achievable by $\left(N, L, M, \frac{1}{\sqrt{N \log N}}\right)$ -VLSF codes satisfies

$$\log M^* \left(N, L, \frac{1}{\sqrt{N \log N}}\right) \geq NC - \sqrt{N \log_{(L)}(N) V} + O\left(\sqrt{\frac{N}{\log_{(L)}(N)}}\right). \quad (100)$$

The decoding times $n_1, \dots, n_L$ that achieve (100) satisfy the equations

$$\log M = n_\ell C - \sqrt{n_\ell \log_{(L-\ell+1)}(n_\ell) V} - \log n_\ell + O(1) \quad (101)$$

for $\ell \in [L]$.

Proof: Lemma 1 analyzes Theorem 1. See Appendix B.1, below. For $L = O(1)$, the proof is significantly different than the corresponding result in [8, eq. (102)] for $L = \infty$ because the dominant error probability term $\mathbb{P}[i(X^{n_L}; Y^{n_L}) < \gamma]$ in (18) disappears when $L = \infty$. ■

We use the average decoding time N and average error probability ϵ of a VLSF code in Lemma 1 in the places of N' and ϵ'_N in (23). By Lemma 1, there exists an $(N', L-1, M, \epsilon'_N)$ -VLSF code with

$$\log M = N'C - \sqrt{N' \log_{(L-1)}(N') V} + O\left(\sqrt{\frac{N'}{\log_{(L-1)}(N')}}\right). \quad (102)$$

Plugging (99) into (102) and applying the necessary Taylor series expansions complete the proof. ■

Lemma 1 is an achievability bound in the moderate deviations regime since the error probability $\frac{1}{\sqrt{N \log N}}$ decays sub-exponentially to zero. The fixed-length scenario in Lemma 1, i.e., $L = 1$, is recovered by [50], which investigates the moderate deviations regime in channel coding. A comparison between the right-hand side of (100) and [50, Th. 2] highlights the benefit of using VLSF codes in the moderate deviations regime. The second-order rate achieved by a VLSF code with $L \geq 2$ decoding times, average decoding time N , and error probability $\frac{1}{\sqrt{N \log N}}$ is achieved by a fixed-length code with blocklength N and error probability $\frac{1}{\sqrt{\log_{(L-1)}(N) \log_{(L)}(N)}}$.

In the remainder of the appendix, we prove Lemma 1 and show the second-order optimality of the parameters used in the proof of Theorem 2 including the decoding times set in (22).

B.1. Proof of Lemma 1

We first present two lemmas used in the proof of Lemma 1 (step 1). We then choose the distribution $P_X^{n_L}$ of the random codewords (step 2) and the parameters $n_1, \dots, n_L, \gamma$ in Theorem 1 (step 3). Finally, we analyze the bounds in Theorem 1 using the supporting lemmas (step 4).

1) *Supporting lemmas:* Lemma 2, below, is the moderate deviations result that bounds the probability that a sum of n zero-mean i.i.d. random variables is above a function of n that grows at most as quickly as $n^{2/3}$.

Lemma 2 (Petrov [42, Ch. 8, Th. 4]): Let $Z_1, \dots, Z_n$ be i.i.d. random variables. Let $\mathbb{E}[Z_1] = 0$, $\sigma^2 = \text{Var}[Z_1]$, and $\mu_3 = \mathbb{E}[Z_1^3]$. Suppose that the moment generating function $\mathbb{E}[\exp\{tZ\}]$ is finite in the neighborhood of $t = 0$. (This condition is known as Cramér's condition.) Let $0 \leq z_n = O(n^{1/6})$. As $n \rightarrow \infty$, it holds that

$$\begin{aligned} \mathbb{P}\left[\sum_{i=1}^n Z_i \geq z_n \sigma \sqrt{n}\right] \\ = Q(z_n) \exp\left\{\frac{z_n^3 \mu_3}{6\sqrt{n}\sigma^3}\right\} + O\left(\frac{1}{\sqrt{n}} \exp\left\{-\frac{z_n^2}{2}\right\}\right) \end{aligned} \quad (103)$$

Lemma 3, below, gives the asymptotic expansion of the root of an equation. We use Lemma 3 to find the asymptotic expansion for the gap between two consecutive decoding times n_ℓ and $n_{\ell+1}$.

Lemma 3: Let $f(x)$ be a differentiable increasing function that satisfies $f'(x) \rightarrow 0$ as $x \rightarrow \infty$. Suppose that

$$x + f(x) = y. \quad (104)$$

Then, as $x \rightarrow \infty$ it holds that

$$x = y - f(y) (1 - o(1)). \quad (105)$$

Proof of Lemma 3: Define the function $F(x) \triangleq x + f(x) - y$. Applying Newton's method with the starting point $x_0 = y$ yields

$$x_1 = x_0 - \frac{F(x_0)}{F'(x_0)} \quad (106)$$

$$= y - \frac{f(y)}{1 + f'(y)} \quad (107)$$

$$= y - f(y)(1 - f'(y) + O(f'(y)^2)). \quad (108)$$

Recall that $f'(y) = o(1)$ by assumption. Equality (108) follows from the Taylor series expansion of the function $\frac{1}{1+x}$ around $x = 0$. Let

$$x^* = y - f(y)(1 - o(1)). \quad (109)$$

From Taylor's theorem, it follows that

$$f(x^*) = f(y) - f'(y_0)f(y)(1 - o(1)), \quad (110)$$

for some $y_0 \in [y - f(y)(1 - o(1)), y]$. Therefore, $f'(y_0) = o(1)$, and $f(x^*) = f(y)(1 - o(1))$. Putting (109)–(110) in (104), we see that x^* is a solution to the equality in (104). ■

2) *Random encoder design:* We set the distribution of the random codewords $P_{X^{n_L}}$ as the product of P_X^* 's, where P_X^* is the capacity-achieving distribution with minimum dispersion, i.e.,

$$P_{X^{n_L}} = (P_X^*)^{n_L} \quad (111)$$

$$P_X^* = \arg \min_{P_X} \{V(X; Y) : I(X; Y) = C\}. \quad (112)$$

3) *Choosing the threshold γ and decoding times $n_1, \dots, n_L$:* We choose $\gamma, n_1, \dots, n_L$ so that the equalities

$$\gamma = n_\ell C - \sqrt{n_\ell \log_{(L-\ell+1)}(n_\ell) V} \quad (113)$$

hold for all $\ell \in [L]$. This choice minimizes our upper bound (19) on the average decoding time up to the second-order term in the asymptotic expansion. See Appendix B.II for the proof. Applying Lemma 3 with

$$x = n_{\ell+1} \quad (114)$$

$$y = n_\ell - \frac{1}{C} \sqrt{n_\ell \log_{(L-\ell+1)}(n_\ell) V} \quad (115)$$

$$f(x) = -\frac{1}{C} \sqrt{n_{\ell+1} \log_{(L-\ell)}(n_{\ell+1}) V} \quad (116)$$

for $\ell \in \{1, \dots, L-1\}$, gives the following gaps between consecutive decoding times

$$\begin{aligned} n_{\ell+1} - n_\ell = \frac{1}{C} \left(\sqrt{n_\ell \log_{(L-\ell)}(n_\ell) V} \right. \\ \left. - \sqrt{n_{\ell+1} \log_{(L-\ell+1)}(n_{\ell+1}) V} \right) (1 + o(1)). \end{aligned} \quad (117)$$

4) *Analyzing the bounds in Theorem 1:* The information density $\imath(X; Y)$ of a DM-PPC is a bounded random variable. Therefore, $\imath(X; Y)$ satisfies Cramér's condition (see Lemma 2).³ For each $\ell \in [L]$, applying Lemma 2 with $\gamma, n_1, \dots, n_L$ satisfying (113) gives

$$\begin{aligned} \mathbb{P}[\imath(X^{n_\ell}; Y^{n_\ell}) < \gamma] \\ \leq Q\left(\sqrt{\log_{(L-\ell+1)}(n_\ell)}\right) \exp\left\{\frac{-(\log_{(L-\ell+1)}(n_\ell))^{3/2} \mu_3}{6\sqrt{n} V^{3/2}}\right\} \\ + O\left(\frac{1}{\sqrt{n}} \exp\left\{-\frac{\log_{(L-\ell+1)}(n_\ell)}{2}\right\}\right) \\ \leq \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\log_{(L-\ell)}(n_\ell)}} \frac{1}{\sqrt{\log_{(L-\ell+1)}(n_\ell)}} \\ \cdot \left(1 + O\left(\frac{(\log_{(L-\ell+1)}(n_\ell))^{(3/2)}}{\sqrt{n_\ell}}\right)\right) \end{aligned} \quad (118)$$

for $\ell < L$, where

$$\mu_3 \triangleq \mathbb{E}[(\imath(X; Y) - C)^3] < \infty, \quad (120)$$

and (119) follows from the Taylor series expansion $\exp\{x\} = 1 + x + O(x^2)$ as $x \rightarrow 0$, and the well-known bound (e.g., [42, Ch. 8, eq. (2.46)])

$$Q(x) \leq \frac{1}{\sqrt{2\pi}} \frac{1}{x} \exp\left\{-\frac{x^2}{2}\right\} \quad \text{for } x > 0. \quad (121)$$

For $\ell = L$, Lemma 2 gives

$$\begin{aligned} \mathbb{P}[\imath(X^{n_L}; Y^{n_L}) < \gamma] \\ \leq \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n_L}} \frac{1}{\sqrt{\log n_L}} \left(1 + O\left(\frac{(\log n_L)^{(3/2)}}{\sqrt{n_L}}\right)\right). \end{aligned} \quad (122)$$

³Here, Cramér's condition is the bottleneck that determines whether our proof technique applies to a specific DM-PPC. For DM-PPCs with infinite input or output alphabets, Cramér's condition may or may not be satisfied. Our proof technique applies to any DM-PPC for which the information density satisfies Cramér's condition.

By Theorem 1, there exists a VLSF code with L decoding times $n_1 < n_2 < \dots < n_L$ such that the expected decoding time is bounded as

$$N \leq n_1 + \sum_{\ell=1}^{L-1} (n_{\ell+1} - n_\ell) \mathbb{P}[\iota(X^{n_\ell}; Y^{n_\ell}) < \gamma]. \quad (123)$$

By (117), we have $n_{\ell+1} = n_\ell(1 + o(1))$ for $\ell \in [L-1]$. Multiplying these asymptotic equations, we get

$$n_\ell = n_1(1 + o(1)) \quad (124)$$

for $\ell \in [L]$. Plugging (117), (119), and (124) into (123), we get

$$N \leq n_1 + \frac{\sqrt{V}}{\sqrt{2\pi}C} \frac{\sqrt{n_1}}{\sqrt{\log_{(L)}(n_1)}} (1 + o(1)). \quad (125)$$

Applying Lemma 3 to (125), we get

$$n_1 \geq N - \frac{\sqrt{V}}{\sqrt{2\pi}C} \frac{\sqrt{N}}{\sqrt{\log_{(L)}(N)}} (1 + o(1)). \quad (126)$$

Comparing (126) and (117), we observe that for n_1 large enough,

$$n_1 < N < n_2 < \dots < n_L. \quad (127)$$

Further, from (113) and (125), we have

$$n_L = N \left(1 + O \left(\sqrt{\frac{\log N}{N}} \right) \right). \quad (128)$$

Finally, we set message set size M such that

$$\log M = \gamma - \log N. \quad (129)$$

Plugging (122) and (129) into (18), we bound the error probability as

$$\mathbb{P}[\mathbf{g}_{\tau^*}(U, Y^{\tau^*}) \neq W] \leq \mathbb{P}[\iota(X^{n_L}; Y^{n_L}) < \gamma] + (M-1) \exp\{-\gamma\} \quad (130)$$

$$\leq \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n_L}} \frac{1}{\sqrt{\log n_L}} (1 + o(1)) + \frac{1}{N} \quad (131)$$

$$\leq \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{N}} \frac{1}{\sqrt{\log N}} (1 + o(1)) + \frac{1}{N}, \quad (132)$$

where (132) follows from (127). Inequality (132) implies that the error probability is bounded by $\frac{1}{\sqrt{N \log N}}$ for N large enough. Plugging (126) and (129) into (113) with $\ell = 1$, we conclude that there exists an $(N, L, M, \frac{1}{\sqrt{N \log N}})$ -VLSF code with

$$\begin{aligned} \log M &\geq NC - \sqrt{N \log_{(L)}(N)V} \\ &\quad - \frac{1}{\sqrt{2\pi}} \sqrt{\frac{NV}{\log_{(L)}(N)}} (1 + o(1)) - \log N, \end{aligned} \quad (133)$$

which completes the proof. ■

B.II. Second-order optimality of the decoding times in Theorem 2

From the code construction described in Theorems 1–2, the average decoding time is

$$N(n_2, \dots, n_L, \gamma) = N'(1 - \epsilon) \frac{1}{1 - \epsilon'_N}, \quad (134)$$

where

$$N' = n_2 + \sum_{i=2}^{L-1} (n_{i+1} - n_i) \mathbb{P}[\iota(X^{n_i}; Y^{n_i}) < \gamma] \quad (135)$$

$$\epsilon'_N = \mathbb{P}[\iota(X^{n_L}; Y^{n_L}) < \gamma] + M \exp\{-\gamma\}. \quad (136)$$

We here show that given a fixed M , the parameters $n_2, n_3, \dots, n_L, \gamma$ chosen according to (113) and (129) (and also the error value ϵ'_N chosen in (23) since ϵ'_N is a function of (n_L, γ)) minimize the average decoding time in (134) in the sense that the second-order expansion of $\log M$ in terms of N is maximized. That is, our parameter choice optimizes our bound on our code construction.

1) *Optimality of $n_2, \dots, n_{L-1}$.* We first set n_L and γ to satisfy the equations

$$\frac{1}{\sqrt{n_L \log n_L}} = \mathbb{P}[\iota(X^{n_L}; Y^{n_L}) < \gamma] + (M-1) \exp\{-\gamma\} \quad (137)$$

$$\log M = \gamma - \log n_L, \quad (138)$$

and optimize the values of $n_2, \dots, n_{L-1}$ under (137)–(138). Section B.II.2 proves the optimality of the choices in (137)–(138).

Under (137)–(138), the optimization problem in (134)–(136) reduces to

$$\begin{aligned} \min \quad & N'(n_2, \dots, n_{L-1}) \\ &= n_2 + \sum_{i=2}^{L-1} (n_{i+1} - n_i) \mathbb{P}[\iota(X^{n_i}; Y^{n_i}) < \gamma] \\ \text{s.t.} \quad & \frac{1}{\sqrt{n_L \log n_L}} = \mathbb{P}[\iota(X^{n_L}; Y^{n_L}) < \gamma] \\ & \quad + (M-1) \exp\{-\gamma\}. \end{aligned} \quad (139)$$

Next, we define the functions

$$g(n) \triangleq \frac{nC - \gamma}{\sqrt{nV}} \quad (140)$$

$$F(n) \triangleq Q(-g(n)) = 1 - Q(g(n)) \quad (141)$$

$$f(n) \triangleq F'(n) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{g(n)^2}{2}\right\} g'(n). \quad (142)$$

Assume that $\gamma = \gamma_n$ is such that $g(n) = O(n^{1/6})$, and $\lim_{n \rightarrow \infty} g(n) = \infty$. Then by Lemma 2, $\mathbb{P}[\iota(X^n; Y^n) < \gamma]$, which is a step-wise constant function of n , is approximated by differentiable function $1 - F(n)$ as

$$\mathbb{P}[\iota(X^n; Y^n) < \gamma] = (1 - F(n))(1 + o(1)). \quad (143)$$

Taylor series expansions give

$$1 - F(n) = \frac{1}{g(n)} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{g(n)^2}{2}\right\} (1 + o(1)) \quad (144)$$

$$f(n) = (1 - F(n))g(n)g'(n)(1 + o(1)) \quad (145)$$

$$g'(n) = \frac{C}{\sqrt{nV}}(1 + o(1)). \quad (146)$$

Let $\mathbf{n}^* = (n_2^*, \dots, n_{L-1}^*)$ denote the solution to the optimization problem in (139) with $\mathbb{P}[i(X^n; Y^n) < \gamma]$ replaced by $(1 - F(n))$. Then $\mathbf{n}^*$ must satisfy the Karush-Kuhn-Tucker conditions $\nabla N'(\mathbf{n}^*) = \mathbf{0}$, giving

$$\left. \frac{\partial N'}{\partial n_2} \right|_{\mathbf{n}=\mathbf{n}^*} = F(n_2^*) - (n_3^* - n_2^*)f(n_2^*) = 0 \quad (147)$$

$$\left. \frac{\partial N'}{\partial n_\ell} \right|_{\mathbf{n}=\mathbf{n}^*} = F(n_\ell^*) - F(n_{L-1}^*) - (n_{\ell+1}^* - n_\ell^*)f(n_\ell^*) = 0 \quad (148)$$

for $\ell = 3, \dots, L-1$.

Let $\tilde{\mathbf{n}} = (\tilde{n}_2, \dots, \tilde{n}_{L-1})$ be the decoding times chosen in (113). We evaluate

$$g(\tilde{n}_\ell) = \sqrt{\log_{(L-\ell+1)}(\tilde{n}_\ell)}(1 + o(1)) \quad (149)$$

$$1 - F(\tilde{n}_\ell) = \frac{1}{\sqrt{2\pi}} \frac{1}{g(\tilde{n}_{\ell+1})g(\tilde{n}_\ell)}(1 + o(1)) \quad (150)$$

$$f(\tilde{n}_\ell) = \frac{1}{\sqrt{2\pi}} \frac{g'(\tilde{n}_\ell)}{g(\tilde{n}_{\ell+1})} \quad (151)$$

$$\tilde{n}_{\ell+1} - \tilde{n}_\ell = \frac{g(\tilde{n}_{\ell+1})}{g'(\tilde{n}_\ell)}(1 + o(1)) \quad (152)$$

for $\ell = 2, \dots, L-1$. Plugging (149)–(152) into (147)–(148) for $\left. \frac{\partial N'}{\partial n_\ell} \right|_{\mathbf{n}=\tilde{\mathbf{n}}}$, we get

$$\nabla N'(\tilde{\mathbf{n}}) = \left(1 - \frac{1}{\sqrt{2\pi}}, -\frac{1}{\sqrt{2\pi}}, -\frac{1}{\sqrt{2\pi}}, \dots, -\frac{1}{\sqrt{2\pi}} \right) (1 + o(1)). \quad (153)$$

Our goal is to find a vector $\Delta \mathbf{n} = (\Delta n_2, \dots, \Delta n_{L-1})$ such that

$$\nabla N'(\tilde{\mathbf{n}} + \Delta \mathbf{n}) = \mathbf{0}, \quad (154)$$

Assume that $\Delta n = O(\sqrt{n})$. By plugging $n + \Delta n$ into (144)–(146) and using the Taylor series expansion of $g(n + \Delta n)$, we get

$$1 - F(n + \Delta n) = (1 - F(n)) \cdot \exp\{-\Delta n g(n) g'(n)\}(1 + o(1)) \quad (155)$$

$$f(n + \Delta n) = f(n) \exp\{-\Delta n g(n) g'(n)\}(1 + o(1)). \quad (156)$$

Using (155)–(156), and putting $\tilde{\mathbf{n}} + \Delta \mathbf{n}$ in (147)–(148), we solve (154) as

$$\Delta n_2 = -\frac{\log \sqrt{2\pi}}{g(\tilde{n}_2)g'(\tilde{n}_2)}(1 + o(1)) \quad (157)$$

$$= -\frac{\sqrt{V} \log \sqrt{2\pi}}{C} \frac{\sqrt{\tilde{n}_2}}{\sqrt{\log_{(L-1)}(\tilde{n}_2)}}(1 + o(1)) \quad (158)$$

$$\Delta n_i = \frac{1}{2} \frac{g(\tilde{n}_{i-1})^2}{g(\tilde{n}_i)g'(\tilde{n}_i)} = o(\Delta n_2)(1 + o(1)) \quad (159)$$

for $i = 3, \dots, L-1$. Hence, $\tilde{\mathbf{n}} + \Delta \mathbf{n}$ satisfies the optimality criterion, and $\mathbf{n}^* = \tilde{\mathbf{n}} + \Delta \mathbf{n}$.

It remains only to evaluate the gap $N'(\mathbf{n}^*) - N'(\tilde{\mathbf{n}})$. We have

$$\begin{aligned} N'(\mathbf{n}^*) - N'(\tilde{\mathbf{n}}) &= \left(\Delta n_2 + \sum_{i=2}^{L-1} (\tilde{n}_{i+1} - \tilde{n}_i) Q(g(\tilde{n}_i)) \right. \\ &\quad \cdot (\exp\{-\Delta n_i g(\tilde{n}_i) g'(\tilde{n}_i)\} - 1) \left. \right) (1 + o(1)) \end{aligned} \quad (160)$$

$$= \left(\Delta n_2 + \left(1 - \frac{1}{\sqrt{2\pi}} \right) \frac{1}{g(\tilde{n}_1)g'(\tilde{n}_1)} - \sum_{i=3}^{L-1} \Delta n_i \right) \cdot (1 + o(1)) \quad (161)$$

$$= -B \frac{\sqrt{\tilde{n}_2}}{\sqrt{\log_{(L-1)}(\tilde{n}_2)}}(1 + o(1)), \quad (162)$$

where $B = \left(\log \sqrt{2\pi} + \frac{1}{\sqrt{2\pi}} - 1 \right) \frac{\sqrt{V}}{C}$ is a positive constant. From the relationship between n_ℓ and n_2 in (124) and the equality (162), we get

$$N'(\tilde{\mathbf{n}}) = N'(\mathbf{n}^*) + B \frac{\sqrt{N'(\mathbf{n}^*)}}{\sqrt{\log_{(L-1)}(N'(\mathbf{n}^*))}}(1 + o(1)). \quad (163)$$

Plugging (163) into our VLSF achievability bound (133) gives

$$\begin{aligned} \log M &\geq N'(\mathbf{n}^*)C - \sqrt{N'(\mathbf{n}^*) \log_{(L-1)}(N'(\mathbf{n}^*))}V \\ &\quad - O\left(\sqrt{\frac{N'(\mathbf{n}^*)}{\log_{(L-1)}(N'(\mathbf{n}^*))}} \right). \end{aligned} \quad (164)$$

Comparing (164) and (133), note that the decoding times chosen in (113) have the optimal second-order term in the asymptotic expansion of the maximum achievable message set size within our code construction. Moreover, the order of the third-order term in (164) is the same as the order of the third-order term in (133). ■

The method of approximating the probability $\mathbb{P}[i(X^n; Y^n) \geq \gamma]$, which is an upper bound for $\mathbb{P}[\tau \leq n]$ (see (86)), by a differentiable function $F(n)$ is introduced in [26, Sec. III] for the optimization problem in (139). In [26], Vakilinia *et al.* approximate the distribution of the random stopping time τ by the Gaussian distribution, where $\mathbb{E}[\tau]$ and $\text{Var}[\tau]$ are found empirically. They derive the Karush-Kuhn-Tucker conditions in (147)–(148), which is known as the SDO algorithm. A similar analysis appears in [27] for the binary erasure channel. The analyses in [26], [27] use the SDO algorithm to numerically solve the equations (147)–(148) for a fixed L , M , and ϵ . Unlike [26], [27], we find the analytic solution to (147)–(148) as decoding times $n_2, \dots, n_L$ approach infinity, and we derive the achievable rate in Theorem 2 as a function of L .

2) *Optimality of n_L and γ* : Let $(\mathbf{n}^*, \gamma^*) = (n_2^*, \dots, n_L^*, \gamma^*)$ be the solution to $\nabla N(\mathbf{n}^*, \gamma^*) = \mathbf{0}$ in (134). Section A finds the values of $n_2^*, n_3^*, \dots, n_{L-1}^*$ that minimize N' . Minimizing N' also minimizes N in (134) since ϵ'_N depends only on n_L and γ , and ϵ is a constant.

Therefore, to minimize N , it only remains to find (n_L^*, γ^*) such that

$$\left. \frac{\partial N}{\partial n_L} \right|_{(\mathbf{n}, \gamma) = (\mathbf{n}^*, \gamma^*)} = 0 \quad (165)$$

$$\left. \frac{\partial N}{\partial \gamma} \right|_{(\mathbf{n}, \gamma) = (\mathbf{n}^*, \gamma^*)} = 0. \quad (166)$$

Consider the case where $L > 2$. Solving (165) and (166) using (147)–(152) gives

$$g(n_L^*) = \sqrt{\log n_L^* + \log_{(2)}(n_L^*) + \log_{(3)}(n_L^*) + O(1)} \quad (167)$$

$$0 = c_0 + N' \left(\frac{1}{\sqrt{2\pi n_L^*}} \exp \left\{ -\frac{g(n_L^*)^2}{2} \right\} (1 + o(1)) - M \exp\{-\gamma^*\} \right), \quad (168)$$

where c_0 is a positive constant. Solving (167)–(168) simultaneously, we obtain

$$\log M = \gamma^* - \log n_L^* + O(1). \quad (169)$$

Plugging (167) and (169) into (136), we get

$$\epsilon_N'^* = \frac{c_1}{\sqrt{n_L^* \log_{(2)}(n_L^*) \log n_L^*}} (1 + o(1)), \quad (170)$$

where c_1 is a constant. Let $(\tilde{\mathbf{n}}, \tilde{\gamma}) = (\tilde{n}_2, \dots, \tilde{n}_K, \tilde{\gamma})$ be the parameters chosen in (113) and (129). Note that $\epsilon_N'^*$ is order-wise different than ϵ_N' in (23). Following steps similar to (160)–(162), we compute

$$N(\mathbf{n}^*, \gamma^*) - N(\tilde{\mathbf{n}}, \tilde{\gamma}) = -O \left(\sqrt{\frac{n_L^*}{\log n_L^*}} \right). \quad (171)$$

Plugging (171) into (21) gives

$$\begin{aligned} \log M &= \frac{N(\mathbf{n}^*, \gamma^*)C}{1 - \epsilon} \\ &\quad - \sqrt{N(\mathbf{n}^*, \gamma^*) \log_{(L-1)}(N(\mathbf{n}^*, \gamma^*))} \frac{V}{1 - \epsilon} \\ &\quad + O \left(\sqrt{\frac{N(\mathbf{n}^*, \gamma^*)}{\log_{(L-1)}(N(\mathbf{n}^*, \gamma^*))}} \right). \end{aligned} \quad (172)$$

Comparing (21) and (172), we see that although (23) and (170) are different, the parameters $(\tilde{\mathbf{n}}, \tilde{\gamma})$ chosen in (113) and (129) have the same second-order term in the asymptotic expansion of the maximum achievable message set size as the parameters $(\mathbf{n}^*, \gamma^*)$ that minimize the average decoding time in the achievability bound in Theorem 1.

For $L = 2$, the summation term in (135) disappears; in this case, the solution to (165) gives

$$\epsilon_N'^* = \frac{c_2}{\sqrt{n_L^* \log n_L^*}} (1 + o(1)) \quad (173)$$

for some constant c_2 . Following the steps in (171)–(172), we conclude that the parameter choices in (113) and (129) are second-order optimal for $L = 2$ as well.

APPENDIX C PROOF OF THEOREM 3

Let P_0 and P_1 be two distributions. Let $Z \triangleq \log \frac{dP_0}{dP_1}$ be the log-likelihood ratio, and let

$$S_n = \sum_{i=1}^n Z_i, \quad (174)$$

where Z_i 's are i.i.d. and have the same distribution as Z . For $i \in \{0, 1\}$, we denote the probability measures and expectations under distribution P_i by $\mathbb{P}_i$ and $\mathbb{E}_i$, respectively. Given a threshold $a_0 \in \mathbb{R}$, define the stopping time

$$T \triangleq \inf\{n \geq 1 : S_n \geq a_0\} \quad (175)$$

and the overshoot

$$\xi_0 = S_T - a_0. \quad (176)$$

The following lemma from [36], which gives the refined asymptotics for the stopping time T , is the main tool to prove our bounds.

Lemma 4 ([36, Cor. 2.3.1, Th. 2.3.3, Th. 2.5.3, Lemma 3.1.1]): Suppose that $\mathbb{E}_0[(Z_1^+)^2] < \infty$, and Z_1 is non-arithmetic. Then, as $a \rightarrow \infty$, it holds that

$$\mathbb{E}_0[T] = \frac{1}{D(P_0 \| P_1)} (a_0 + \mathbb{E}_0[\xi_0]) \quad (177)$$

$$\begin{aligned} &= \frac{1}{D(P_0 \| P_1)} \left(a_0 + \frac{\mathbb{E}_0[Z_1^2]}{2D(P_0 \| P_1)} \right. \\ &\quad \left. - \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E}_0[S_n^-] + o(1) \right), \end{aligned} \quad (178)$$

and

$$\mathbb{P}_0[T < \infty] = 1 \quad (179)$$

$$\mathbb{P}_1[T < \infty] = e^{-a_0} \mathbb{E}_0[e^{-\xi_0}] \quad (180)$$

$$\mathbb{E}_0[e^{-\lambda \xi_0}] = \frac{1 + o(1)}{\lambda D(P_0 \| P_1)} \exp \left\{ - \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{E}_0[e^{-\lambda S_n^+}] \right\}. \quad (181)$$

C.I. Achievability Proof

Let P_X be a capacity-achieving distribution of the DM-PPC. Define the hypotheses

$$H_0: (X^{d_N}, Y^{d_N})^\infty \sim P_0^\infty = ((P_X \times P_{Y|X})^{d_N})^\infty \quad (182)$$

$$H_1: (X^{d_N}, Y^{d_N})^\infty \sim P_1^\infty = ((P_X \times P_Y)^{d_N})^\infty, \quad (183)$$

and the random variables

$$W_i \triangleq \frac{1}{d_N} \log \frac{dP_0^{d_N}}{dP_1^{d_N}} (X^{(i-1)d_N+1:id_N}, Y^{(i-1)d_N+1:id_N}) \quad (184)$$

$$= \frac{1}{d_N} \imath(X^{(i-1)d_N+1:id_N}, Y^{(i-1)d_N+1:id_N}) \quad (185)$$

for $i \in \mathbb{N}$. Note that under H_0 , $\mathbb{E}_0[W_i] = C$. Here, we use W_i in the place of Z_i in (174). Define

$$S_n \triangleq \sum_{i=1}^n W_i, \quad (186)$$

and

$$\tau \triangleq \inf\{k \geq 1: S_k \geq a_0/d_N\} \quad (187)$$

$$T \triangleq d_N \tau. \quad (188)$$

We employ the stop-at-time-zero procedure described in the proof sketch of Theorem 2 with $\epsilon'_N = \frac{1}{\mathbb{E}_0[T]}$ and the information density threshold rule (80)–(83) from the proof of Theorem 1, where the threshold γ is set to a_0 . Here, T is as in (175). We set M and a_0 so that

$$M\mathbb{P}_1[T < \infty] \leq Me^{-a_0} = \epsilon'_N = \frac{1}{\mathbb{E}_0[T]}, \quad (189)$$

where the inequality follows from (180). Following steps identical to (99), and noting that $\mathbb{P}_0[T = \infty] = 0$, we get

$$N = (1 - \epsilon)\mathbb{E}_0[T] + O(1), \quad (190)$$

and the average error probability of the code is bounded by ϵ .

To evaluate $\mathbb{E}_0[T]$, we use Lemma 4 with W_i in place of Z_i . A straightforward calculation yields

$$\mathbb{E}_0[W_1^2] = \mathbb{E}_0[W_1]^2 - \text{Var} \left[\frac{1}{d_N} \sum_{i=1}^{d_N} \iota(X_i; Y_i) \right] \quad (191)$$

$$= C^2 - \frac{1}{d_N} \text{Var} [\iota(X_1; Y_1)]. \quad (192)$$

Next, we have that

$$\mathbb{E}_0[S_n^-] = -nd_N \mathbb{E} \left[\frac{1}{nd_N} S_n 1 \left\{ \frac{1}{nd_N} S_n \leq 0 \right\} \right], \quad (193)$$

where $S_n = \sum_{j=1}^{nd_N} \iota(X_j; Y_j)$. Applying the saddlepoint approximation (e.g., [51, eq. (1.2)]) to $\frac{1}{nd_N} S_n$, we get

$$\mathbb{E}_0[S_n^-] = nd_N \int_{-\infty}^0 c(x) \sqrt{nd_N} e^{-nd_N g(x) + \log x} dx, \quad (194)$$

where $c(x)$ and $g(x)$ are bounded from below a positive constant for all $x \in (-\infty, 0]$. Applying the Laplace's integral [51, eq. (2.5)] to (194), we get

$$\mathbb{E}_0[S_n^-] = e^{-nd_N c_n + o(nd_N)} \quad (195)$$

for all $n \in \mathbb{Z}_+$, where each c_n is a positive constant depending on n . Putting (192) and (195) into (178) and (188), we get

$$\mathbb{E}_0[T] = \frac{a_0}{C} + \frac{d_N}{2} + o(d_N). \quad (196)$$

From (189)–(190), we get

$$\mathbb{E}_0[T] = \frac{N}{1 - \epsilon} + O(1) \quad (197)$$

$$\log M = a_0 - \log N. \quad (198)$$

Putting (196)–(198) together completes the proof of (26).

C.II. Converse Proof

Recall the definition of an SHT $(\delta, \tau, \mathcal{N})$ from Appendix A.11 that tests the hypotheses

$$H_0: Z^\infty \sim P_0 \quad (199)$$

$$H_1: Z^\infty \sim P_1, \quad (200)$$

where P_0 and P_1 are distributions on a common alphabet $\mathcal{Z}^\infty$. Here, $Z^\infty \triangleq (Z_1, Z_2, \dots)$ denotes a vector of infinite length whose joint distribution is either P_0 or P_1 , which need not be product distributions in general. We define the minimum achievable type-II error probability, subject to a type-I error probability bound and a maximal expected decoding time constraint, with decision times restricted to the set $\mathcal{N}$ as

$$\beta_{(\epsilon, N, \mathcal{N})}(P_0, P_1) \triangleq \min_{(\delta, \tau, \mathcal{N}): \mathbb{P}_0[\delta=1] \leq \epsilon, \max\{\mathbb{E}_0[\tau], \mathbb{E}_1[\tau]\} \leq N} \mathbb{P}_1[\delta=0], \quad (201)$$

which is the SHT version of the β_α -function defined for the fixed-length binary hypothesis test [13].

The following theorem extends the meta-converse bound [13, Th. 27], which is a fundamental theorem used to show converse results in fixed-length channel coding without feedback and many other applications.

Theorem 9: Fix any set $\mathcal{N} \subseteq \mathbb{Z}_\geq$, a real number $N > 0$, and a DM-PPC $P_{Y|X}$. Then, it holds that

$$\begin{aligned} \log M^*(N, |\mathcal{N}|, \epsilon, \mathcal{N}) \\ \leq \sup_{P_{X^\infty}} \inf_{Q_{Y^\infty}} -\log \beta_{(\epsilon, N, \mathcal{N})}(P_{X^\infty} \times P_{Y|X}^\infty, P_{X^\infty} \times Q_{Y^\infty}). \end{aligned} \quad (202)$$

Proof: The proof is similar to that in [13]. Let W denote a message equiprobably distributed on $[M]$, and let $\hat{W}$ be its reconstruction. Given any VLSF code with the set of available decoding times $\mathcal{N}$, average decoding time N , error probability ϵ , and codebook size M , let $\hat{P}_{X^\infty}$ denote the input distribution induced by the code's codebook. The code operation creates a Markov chain $W \rightarrow X^\infty \rightarrow Y^\infty \rightarrow \hat{W}$. As full feedback breaks this Markov chain, stop feedback does not since the channel inputs are conditionally independent of the channel outputs given the message W . Fix an arbitrary output distribution Q_{Y^∞} , and consider the SHT

$$H_0: (X^\infty, Y^\infty) \sim \hat{P}_{X^\infty} \times P_{Y|X}^\infty \quad (203)$$

$$H_1: (X^\infty, Y^\infty) \sim \hat{P}_{X^\infty} \times Q_{Y^\infty} \quad (204)$$

with a test $\delta = 1\{\hat{W} \neq W\}$, where $(W, \hat{W})$ are generated by the (potentially random) encoder-decoder pair of the VLSF code. The type-I and type-II error probabilities of this code-induced SHT are

$$\alpha = \mathbb{P}_0[\delta = 1] = \mathbb{P}[\hat{W} \neq W] \leq \epsilon \quad (205)$$

$$\beta = \mathbb{P}_1[\delta = 0] = \frac{1}{M}, \quad (206)$$

where (206) follows since the sequence Y^∞ is independent of X^∞ under H_1 . The stopping time of this SHT under H_0 or H_1 is bounded by N by the definition of a VLSF code. Since the error probabilities in (205)–(206) cannot be better than that of the optimal SHT, it holds that

$$\log M$$

$$\begin{aligned}
&\leq -\log \beta_{(\epsilon, N, \mathcal{N})}(\hat{P}_{X^\infty} \times P_{Y|X}^\infty, \hat{P}_{X^\infty} \times Q_{Y^\infty}) \quad (207) \\
&\leq \inf_{Q_{Y^\infty}} -\log \beta_{(\epsilon, N, \mathcal{N})}(\hat{P}_{X^\infty} \times P_{Y|X}^\infty, \hat{P}_{X^\infty} \times Q_{Y^\infty}) \quad (208) \\
&\leq \sup_{P_{X^\infty}} \inf_{Q_{Y^\infty}} -\log \beta_{(\epsilon, N, \mathcal{N})}(P_{X^\infty} \times P_{Y|X}^\infty, P_{X^\infty} \times Q_{Y^\infty}), \quad (209)
\end{aligned}$$

where (208) follows since the choice Q_{Y^∞} is arbitrary. ■

To prove (27), we apply Theorem 9 and get

$$\log M \leq -\log \beta_{(\epsilon, N, \mathcal{N})}(P_{Y|X}^\infty, P_Y^\infty), \quad (210)$$

where P_Y is the capacity-achieving output distribution, and $\mathcal{N} = \{0, d_N, 2d_N, \dots\}$. The reduction from Theorem 9 to (210) follows since $\log \frac{P_{Y|X}(Y|x)}{P_Y(Y)}$ has the same distribution for all $x \in \mathcal{X}$ for Cover–Thomas symmetric channels [43, p. 190]. In the remainder of the proof, we derive an upper bound for the right-hand side of (210).

Consider any SHT $(\delta, \tau, \mathcal{N})$ with $\mathbb{E}_0[\tau] \leq N$ and $\mathbb{E}_1[\tau] \leq N$. Our definition in (201) is slightly different than the classical SHT definition from [52] since our definition allows one to make a decision at time 0. Notice that at time 0, any test has three choices: decide H_0 , decide H_1 , or decide to start taking samples. When the test decides to start taking samples, the remainder of the procedure becomes a classical SHT. From this observation, any test satisfies

$$\epsilon \geq \alpha = \epsilon_0 + (1 - \epsilon_0 - \epsilon_1)\alpha' \geq \epsilon_0 \quad (211)$$

$$\beta \geq \epsilon_1 + (1 - \epsilon_0 - \epsilon_1)\beta' \geq (1 - \epsilon_0)\beta', \quad (212)$$

where at time 0, the test decides H_i with probability ϵ_{1-i} , and α' and β' are the type-I and type-II error probabilities conditioned on the event that the test decides to take samples at time 0, which occurs with probability $1 - \epsilon_0 - \epsilon_1$.

Let τ' denote the average stopping time of the test with error probabilities (α', β') . We have

$$\mathbb{E}_0[\tau] = (1 - \epsilon_0 - \epsilon_1)\mathbb{E}_0[\tau'] \quad (213)$$

$$= (1 - \epsilon_0)(\mathbb{E}_0[\tau'] + e^{-O(N)}) \leq N \quad (214)$$

$$\mathbb{E}_1[\tau] = (1 - \epsilon_0 - \epsilon_1)\mathbb{E}_1[\tau'] \quad (215)$$

$$= (1 - \epsilon_0)(\mathbb{E}_1[\tau'] + e^{-O(N)}) \leq N \quad (216)$$

since β decays exponentially with $\mathbb{E}_0[\tau]$.

The following argument is similar to that in [53, Sec. V-C].

Set an arbitrary $\nu > 0$ and the thresholds

$$\tilde{a}_0 = C \left(\frac{N}{1 - \epsilon_0} - \frac{d_N}{2} - o(d_N) + \nu \right) \quad (217)$$

$$\tilde{a}_1 = D(P_Y \| P_{Y|X=x}) \left(\frac{N}{1 - \epsilon_0} - \frac{d_N}{2} - o(d_N) + \nu \right), \quad (218)$$

where $x \in \mathcal{X}$ is arbitrary, and let $(\tilde{\delta}, \tilde{\tau}, \mathcal{N})$ be the SPRT associated with the thresholds $(-\tilde{a}_1, \tilde{a}_0)$, and type-I and type-II error probabilities $\tilde{\alpha}$ and $\tilde{\beta}$.

Applying [36, eq. (3.56)] to (196), we get

$$\mathbb{E}_0[\tilde{\tau}] = \frac{\tilde{a}_0}{C} + \frac{d_N}{2} + o(d_N) \quad (219)$$

$$\mathbb{E}_1[\tilde{\tau}] = \frac{\tilde{a}_1}{D(P_Y \| P_{Y|X=x})} + \frac{d_N}{2} + o(d_N). \quad (220)$$

Combining (217)–(220) gives

$$\mathbb{E}_0[\tilde{\tau}] \geq \frac{N}{1 - \epsilon_0} + \nu \quad (221)$$

$$\mathbb{E}_1[\tilde{\tau}] \geq \frac{N}{1 - \epsilon_0} + \nu. \quad (222)$$

Letting $\nu = O(\frac{1}{N})$, it follows from (213)–(216) and (221)–(222) that

$$\mathbb{E}_0[\tilde{\tau}] \geq \mathbb{E}_0[\tau'] \quad (223)$$

$$\mathbb{E}_1[\tilde{\tau}] \geq \mathbb{E}_1[\tau'] \quad (224)$$

for a large enough N . Using Wald and Wolfowitz's SPRT optimality result [54], we get

$$\alpha' \geq \tilde{\alpha} \quad (225)$$

$$\beta' \geq \tilde{\beta}. \quad (226)$$

Now it only remains to lower bound $\tilde{\beta}$. Applying [36, Th. 3.1.2, 3.1.3] and (181) gives

$$\tilde{\beta} = \tilde{\zeta} e^{-\tilde{a}_0} (1 + o(1)), \quad (227)$$

where

$$\tilde{\zeta} = \frac{1}{d_N C} \left(\exp \left\{ - \sum_{n=1}^{\infty} \frac{1}{n} \mathbb{P}_0[S_n < 0] + \mathbb{P}_1[S_n > 0] \right\} \right), \quad (228)$$

and S_n is as in (186). Since S_n is a sum of $nd_N \rightarrow \infty$ i.i.d. random variables, where the summands have a non-zero mean, the Chernoff bound implies that each of the probabilities in (228) decays exponentially with d_N . Thus,

$$\tilde{\zeta} = \frac{1}{d_N C} (1 + o(1)). \quad (229)$$

From (217) and (229), we get

$$-\log \tilde{\beta} = C \left(\frac{N}{1 - \epsilon_0} - \frac{d_N}{2} - o(d_N) + o(\log d_N) \right) \quad (230)$$

$$\leq C \left(\frac{N}{1 - \epsilon} - \frac{d_N}{2} - o(d_N) + o(\log d_N) \right), \quad (231)$$

where (231) follows from (211). Inequalities (212), (226), and (231) imply (27).

APPENDIX D PROOFS FOR THE DM-MAC

In this section, we prove our main results for the DM-MAC, beginning with Theorem 4, which is used to prove Theorem 5.

D.1. Proof of Theorem 4

For each transmitter k , $k \in [K]$, we generate M_k n_L -dimensional codewords i.i.d. from $P_{X_k}^{n_L}$. Codewords for distinct transmitters are drawn independently of each other. Denote the codeword for transmitter k and message m_k by $X_k^{n_L}(m_k)$ for $k \in [K]$ and $m_k \in [M_k]$. The proof extends the DM-PPC achievability bound in Theorem 1 that is based on a sub-optimal SHT to the DM-MAC. Below, we explain the differences.

Without loss of generality, assume that $m_{[K]} = \mathbf{1}$ is transmitted. The hypothesis test in (72)–(73) is replaced by

$$H_0: (X_{[K]}^{n_L}, Y_K^{n_L}) \sim \left(\prod_{k=1}^K P_{X_k}^{n_L} \right) \times P_{Y_K|X_{[K]}}^{n_L} \quad (232)$$

$$H_1: (X_{[K]}^{n_L}, Y_K^{n_L}) \sim \left(\prod_{k=1}^K P_{X_k}^{n_L} \right) \times P_{Y_K}^{n_L}, \quad (233)$$

which is run for every message tuple $m_{[K]} \in \prod_{k=1}^K [M_k]$. The information density (80), the stopping times (81)–(82), and the decision rule (83) are extended to the DM-MAC as

$$S_{m_{[K]}, n_\ell} \triangleq \iota_K(X_{[K]}^{n_\ell}(m_{[K]}; Y_K^{n_\ell}) \quad (234)$$

$$\tau_{m_{[K]}} \triangleq \inf\{n_\ell \in \mathcal{N} : S_{m_{[K]}, n_\ell} \geq \gamma\} \quad (235)$$

$$\tilde{\tau}_{m_{[K]}} \triangleq \min\{\tau_{m_{[K]}}, n_L\} \quad (236)$$

$$\delta_{m_{[K]}} \triangleq \begin{cases} 0 & \text{if } S_{m_{[K]}, n_\ell} \geq \gamma \\ 1 & \text{if } S_{m_{[K]}, n_\ell} < \gamma \end{cases} \quad (237)$$

for every message tuple $m_{[K]}$ and decoding time n_ℓ . For brevity, let $(X_{[K]}^{n_\ell}, Y_K^{n_\ell}, \bar{X}_{[K]}^{n_\ell})$ be drawn i.i.d. according to the joint distribution

$$\begin{aligned} P_{X_{[K]}, Y_K, \bar{X}_{[K]}}(x_{[K]}, y, \bar{x}_{[K]}) \\ = P_{Y_K|X_{[K]}}(y|x_{[K]}) \prod_{k=1}^K P_{X_k}(x_k) P_{X_k}(\bar{x}_k). \end{aligned} \quad (238)$$

Expected decoding time analysis: Following steps identical to (84)–(86), we get (41).

Error probability analysis: The following error analysis extends the PPC bounds in (79) and (89)–(97) to the DM-MAC.

In the analysis below, for brevity, we write $m_{\mathcal{A}} \neq \mathbf{1}$ to denote that $m_i \neq 1$ for $i \in \mathcal{A}$. The error probability is bounded as

$$\epsilon \leq \mathbb{P} \left[\bigcup_{m_{[K]} \neq \mathbf{1}} \{\tau_{m_{[K]}} \leq \tau_1 < \infty\} \bigcup \{\tau_1 = \infty\} \right] \quad (239)$$

$$\leq \mathbb{P}[\tau_1 = \infty] + \mathbb{P} \left[\bigcup_{m_{[K]} \neq \mathbf{1}} \{\tau_{m_{[K]}} < \infty\} \right] \quad (240)$$

$$+ \mathbb{P} \left[\bigcup_{\substack{m_{[K]} \neq \mathbf{1} : \exists i \in [K] \\ m_i = 1}} \{\tau_{m_{[K]}} < \infty\} \right], \quad (241)$$

where (240)–(241) apply the union bound to separate the probabilities of the following error events:

- 1) the information density of the true message tuple does not satisfy the threshold test for any available decoding time;
- 2) the information density of a message tuple in which all messages are incorrect satisfies the threshold test for some decoding time;
- 3) the information density of a message tuple in which the messages from some transmitters are correct and

the messages from the other transmitters are incorrect satisfies the threshold test for some decoding time.

The terms in (240) are bounded using steps identical to (89)–(97) as

$$\mathbb{P}[\tau_1 = \infty] \leq \mathbb{P}[\iota_K(X_{[K]}^{n_L}; Y_K^{n_L}) < \gamma] \quad (242)$$

$$\mathbb{P} \left[\bigcup_{m_{[K]} \neq \mathbf{1}} \{\tau_{m_{[K]}} < \infty\} \right] \leq \prod_{k=1}^K (M_k - 1) \exp\{-\gamma\}. \quad (243)$$

For the cases where at least one message is decoded correctly, we delay the application of the union bound. Let $\mathcal{A} \in \mathcal{P}([K])$ be the set of transmitters whose messages are decoded correctly. Define

$$\begin{aligned} \mathcal{M}^{(\mathcal{A})} \triangleq \{m_{[K]} \in [M]^K : m_k = 1 \text{ for } k \in \mathcal{A}, \\ m_k \neq 1 \text{ for } k \in \mathcal{A}^c\} \end{aligned} \quad (244)$$

$$\tilde{\mathcal{M}}^{(\mathcal{A})} \triangleq \{m_{\mathcal{A}} \in [M]^{|\mathcal{A}|} : m_k \neq 1 \text{ for } k \in \mathcal{A}\}. \quad (245)$$

We first bound the probability term in (241) by applying the union bound according to which subset $\mathcal{A}$ of the transmitter set $[K]$ is decoded correctly, and get

$$\begin{aligned} \mathbb{P} \left[\bigcup_{\substack{m_{[K]} \neq \mathbf{1} : \exists i \in [K] \\ m_i = 1}} \{\tau_{m_{[K]}} < \infty\} \right] \\ \leq \sum_{\mathcal{A} \in \mathcal{P}([K])} \mathbb{P} \left[\bigcup_{m_{[K]} \in \mathcal{M}^{(\mathcal{A})}} \{\tau_{m_{[K]}} < \infty\} \right] \quad (246) \\ = \sum_{\mathcal{A} \in \mathcal{P}([K])} \mathbb{P} \left[\bigcup_{\substack{m_{\mathcal{A}^c} \in \tilde{\mathcal{M}}^{(\mathcal{A}^c)} \\ n_\ell \in \mathcal{N}}} \{\iota_K(\bar{X}_{\mathcal{A}^c}^{n_\ell}(m_{\mathcal{A}^c}), X_{\mathcal{A}}^{n_\ell}; Y_K^{n_\ell}) \geq \gamma\} \right], \end{aligned} \quad (247)$$

where $\bar{X}_{\mathcal{A}^c}^{n_\ell}(m_{\mathcal{A}^c})$ refers to the random sample from the codebooks of transmitter set $\mathcal{A}^c$, independent from the codewords $X_{\mathcal{A}^c}^{n_\ell}$ transmitted by the transmitters $\mathcal{A}^c$ and the received output Y^{n_ℓ} .

We bound the right-hand side of (247) using the same method as in [28, eq. (65)–(66)]. This step is crucial in enabling the single-threshold rule for the rate vectors approaching a point on the sum-rate boundary. Set an arbitrary $\lambda^{(\mathcal{A})} > 0$. Define two events

$$\mathcal{E}(\mathcal{A}) \triangleq \bigcup_{n_\ell \in \mathcal{N}} \left\{ \iota_K(X_{\mathcal{A}}^{n_\ell}; Y_K^{n_\ell}) > NI_K(X_{\mathcal{A}}; Y_K) + N\lambda^{(\mathcal{A})} \right\} \quad (248)$$

$$\mathcal{F}(\mathcal{A}) \triangleq \bigcup_{\substack{m_{\mathcal{A}^c} \in \tilde{\mathcal{M}}^{(\mathcal{A}^c)} \\ n_\ell \in \mathcal{N}}} \left\{ \iota_K(\bar{X}_{\mathcal{A}^c}^{n_\ell}(m_{\mathcal{A}^c}), X_{\mathcal{A}}^{n_\ell}; Y_K^{n_\ell}) \geq \gamma \right\}. \quad (249)$$

Define the threshold

$$\bar{\gamma}^{(\mathcal{A})} \triangleq \gamma - NI_K(X_{\mathcal{A}}; Y_K) - N\lambda^{(\mathcal{A})}. \quad (250)$$

We have

$$\mathbb{P}[\mathcal{F}(\mathcal{A})]$$

$$= \mathbb{P}[\mathcal{F}(\mathcal{A}) \cap \mathcal{E}(\mathcal{A})] + \mathbb{P}[\mathcal{F}(\mathcal{A}) \cap \mathcal{E}(\mathcal{A})^c] \quad (251)$$

$$\leq \mathbb{P}[\mathcal{E}(\mathcal{A})]$$

$$+ \mathbb{P} \left[\bigcup_{\substack{m_{\mathcal{A}^c} \in \tilde{\mathcal{M}}^{(\mathcal{A}^c)} \\ n_\ell \in \mathcal{N}}} \left\{ \iota_K(\bar{X}_{\mathcal{A}^c}^{n_\ell}; Y_K^{n_\ell} | X_{\mathcal{A}}^{n_\ell}) \geq \bar{\gamma}^{(\mathcal{A})} \right\} \right] \quad (252)$$

$$\leq \sum_{n_\ell \in \mathcal{N}} \mathbb{P} \left[\iota_K(X_{\mathcal{A}}^{n_\ell}; Y_K^{n_\ell}) > NI_K(X_{\mathcal{A}}; Y_K) + N\lambda^{(\mathcal{A})} \right] \\ + \prod_{k \in \mathcal{A}^c} (M_k - 1) \mathbb{P} \left[\bigcup_{n_\ell \in \mathcal{N}} \left\{ \iota_K(\bar{X}_{\mathcal{A}^c}^{n_\ell}; Y_K^{n_\ell} | X_{\mathcal{A}}^{n_\ell}) \geq \bar{\gamma}^{(\mathcal{A})} \right\} \right] \quad (253)$$

$$\leq \sum_{n_\ell \in \mathcal{N}} \mathbb{P} \left[\iota_K(X_{\mathcal{A}}^{n_\ell}; Y_K^{n_\ell}) > NI_K(X_{\mathcal{A}}; Y_K) + N\lambda^{(\mathcal{A})} \right] \\ + \prod_{k \in \mathcal{A}^c} (M_k - 1) \exp\{-\bar{\gamma}^{(\mathcal{A})}\}, \quad (254)$$

where inequality (252) uses the chain rule for mutual information, (253) applies the union bound, and (254) follows from [20, eq. (88)].

Applying the bound in (254) to each of the probabilities in (247) and plugging (242), (243), and (247) into (240)–(241), we complete the proof of Theorem 4.

D.II. Proof of Theorem 5

We employ the stop-at-time-zero procedure in the proof sketch of Theorem 2 with $\epsilon'_N = \frac{1}{\sqrt{N \log N}}$. Therefore, we first show that there exists an $(N, L, M_{[K]}, 1/\sqrt{N \log N})$ -VLSF code satisfying

$$\sum_{k=1}^K \log M_k = NI_K - \sqrt{N \log_{(L)}(N) V_K} \\ + O \left(\sqrt{\frac{NV_K}{\log_{(L)}(N)}} \right). \quad (255)$$

We set the parameters

$$\gamma = n_\ell I_K - \sqrt{n_\ell \log_{(L-\ell+1)}(n_\ell) V_K} \quad \forall \ell \in [L] \quad (256)$$

$$= \sum_{k=1}^K \log M_k + \log N \quad (257)$$

$$\lambda^{(\mathcal{A})} = \frac{NI_K(X_{\mathcal{A}^c}; Y_K | X_{\mathcal{A}}) - \sum_{k \in \mathcal{A}^c} \log M_k}{2N}, \quad \mathcal{A} \in \mathcal{P}([K]). \quad (258)$$

Note that $\lambda^{(\mathcal{A})}$'s are bounded below by a positive constant for rate points lying on the sum-rate boundary.

By Theorem 4, there exists a VLSF code with L decoding times $n_1 < n_2 < \dots < n_L$ such that the average decoding time is bounded as

$$N \leq n_1 + \sum_{\ell=1}^{L-1} (n_{\ell+1} - n_\ell) \mathbb{P} \left[\iota_K(X_{[K]}^{n_\ell}; Y_K^{n_\ell}) < \gamma \right]. \quad (259)$$

Following the analysis in (125)–(128), we conclude that

$$n_\ell = N(1 + o(1)) \quad (260)$$

for all $\ell \in [L]$. Applying the Chernoff bound to the probability terms in (39)–(40) using (256) and (260), we get that the sum of the terms in (39)–(40) is bounded by $\exp\{-NE\}$ for some $E > 0$.

Applying Lemma 2 to the probability in (37) with (256) gives

$$\mathbb{P} \left[\iota_K(X_{[K]}^{n_L}; Y_K^{n_L}) < \gamma \right] \leq \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n_L}} \frac{1}{\sqrt{\log n_L}} \\ \cdot \left(1 + O \left(\frac{(\log n_L)^{(3/2)}}{\sqrt{n_L}} \right) \right). \quad (261)$$

Applying Theorem 4 with (257), (261), and the exponential bound on the sum of the terms in (39)–(40), we bound the error probability as

$$\mathbb{P} \left[\mathbf{g}_{\tau^*}(U, Y^{\tau^*}) \neq W_{[K]} \right] \\ \leq \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{N}} \frac{1}{\sqrt{\log N}} \cdot \left(1 + O \left(\frac{(\log N)^{(3/2)}}{\sqrt{N}} \right) \right) \\ + \frac{1}{N} + \exp\{-NE\}, \quad (262)$$

which is further bounded by $\frac{1}{\sqrt{N \log N}}$ for N large enough. Following steps identical to (125)–(133), we prove the existence of a VLSF code that satisfies (255) for the DM-MAC with L decoding times and error probability $\frac{1}{\sqrt{N \log N}}$.

Finally, invoking (255) with L replaced by $L - 1$ and the stop-at-time-zero procedure in the proof sketch of Theorem 2 with $\epsilon'_N = \frac{1}{\sqrt{N \log N}}$, we complete the proof of Theorem 5.

D.III. Proof of (44)

The proof of (44) follows steps similar to those in the proof of [8, Th. 2]. Below, we detail the differences between the proofs of (44), Theorem 5, and [8, Th. 2].

- 1) In (44), we choose $cN + 1$ decoding times as $n_i = i - 1$ for $i = 1, \dots, cN + 1$ for a sufficiently large constant $c > 1$. This differs from Theorem 5 where L does not grow with N ($L = O(1)$) and the gaps between consecutive decoding times can differ. In [8, Th. 2], any integer time is available for decoding, giving $L = n_{\max} = \infty$.
- 2) We here set the parameter γ differently from how it was set in (256) and (257). The difference accounts for the error event that some of the messages are decoded incorrectly and some of the messages are decoded correctly. Specifically, we set

$$\gamma = NI_K - a \quad (263)$$

$$= \sum_{k=1}^K \log M_k + \log N + b, \quad (264)$$

where a is an upper bound on the information density $\iota_K(X_{[K]}; Y_K)$, and b is a positive constant to be determined later. Since the number of decoding times L grows linearly with N and $c > 1$, applying the Chernoff bound gives

$$(37) + (39) + (40) \leq \exp\{-NE\} \quad (265)$$

for some $E > 0$ and N large enough. Hence, the error probability ϵ in Theorem 4 is bounded by $\frac{\exp\{-b\}}{N} + \exp\{-NE\}$, which can be further bounded by $\frac{1}{N}$ by appropriately choosing the constant b .

The term (37) disappears in [8, Th. 2] because $n_L = \infty$; the terms (39) and (40) disappear in [8, Th. 2] because the channel is point-to-point. Therefore, b is set to 0 in [8, Th. 2].

- 3) We bound the average decoding time $\mathbb{E}[\tau^*]$ as

$$\mathbb{E}[\tau^*] \leq \frac{\gamma + a}{I_K} = N \quad (266)$$

using Doob's optional stopping theorem as used in [8, eq. (106)-(107)] while $\mathbb{E}[\tau^*]$ in the proof of Theorem 5 is bounded by bounding the tail probability of information density.

The steps above show the achievability of an $(N, cN, M_{[K]}, 1/N)$ code with

$$\sum_{k=1}^K \log M_k = NI_K - \log N + O(1). \quad (267)$$

- 4) Lastly, as in [8, Th. 2], we invoke the stop-at-time-zero procedure from the proof sketch of Theorem 2 with $\epsilon'_N = \frac{1}{N^\gamma}$.

APPENDIX E PROOF OF THEOREM 6

In Theorem 6, we employ a multiple hypothesis test at some early time n_0 to estimate the number of active transmitters followed by a VLSF MAC coding. Since VLSF MAC code-word design employed in Theorem 5 is unchanged (up to the coding dimension), the VLSF MAC code employs a single nested codebook, as described in the proof below. If the test decides that the number of active transmitters is $\hat{k} = 0$, then the decoder declares no active transmitters at time n_0 and stops the transmission at that time. If the estimated number of active transmitters is $\hat{k} \neq 0$, then the decoder decides to decode at one of the available times $n_{\hat{k},1}, \dots, n_{\hat{k},L}$ using the decoder for the MAC with $\hat{k}$ transmitters.

E.I. Encoding and decoding

Encoding: As in the DM-PPC and DM-MAC cases, the codewords are generated i.i.d. from the distribution $P_X^{n_{K,L}}$.

Decoding: The decoder combines a $(K+1)$ -ary hypothesis test and the threshold test that is used for the DM-MAC.

Multiple hypothesis test: Given distributions P_{Y_k} , $k \in \{0, \dots, K\}$ where $\mathcal{Y}_K$ is the common alphabet, we test the hypotheses

$$H_k: Y^{n_0} \sim P_{Y_k}^{n_0}, \quad k \in \{0, \dots, K\}. \quad (268)$$

The error probability constraints of our test are

$$\mathbb{P}[\text{Decide } H_s \text{ where } s \neq 0 | H_0] \leq \epsilon_0 \quad (269)$$

$$\mathbb{P}[\text{Decide } H_s \text{ where } s \neq k | H_k] \leq \exp\{-n_0 E + o(n_0)\} \quad (270)$$

for $k \in [K]$, where $E > 0$ is a constant.

Due to the asymmetry in (269)–(270), we employ a composite hypothesis testing to decide whether H_0 is true; that is, the test declares H_0 if

$$\log \frac{P_{Y_0}^{n_0}(y^{n_0})}{P_{Y_s}^{n_0}(y^{n_0})} \geq \eta_s \quad (271)$$

for all $s \in [K]$, where the threshold values η_s , $s \in [K]$, are chosen to satisfy (269). If the condition in (271) is not satisfied, then the test applies the maximum likelihood decoding rule, i.e., the output is H_s , where

$$s = \arg \max_{s \in [K]} P_{Y_s}^{n_0}(y^{n_0}). \quad (272)$$

From the asymptotics of the error probability bound for composite hypothesis testing in [28, Th. 5], the maximum type-II error of the composite hypothesis test is bounded as

$$\begin{aligned} \max_{k \in [K]} \mathbb{P}[\text{Decide } H_0 | H_k] \\ \leq \exp \left\{ -n_0 \min_{k \in [K]} D(P_{Y_0} \| P_{Y_k}) + O(\sqrt{n_0}) \right\}. \end{aligned} \quad (273)$$

If P_{Y_0} is not absolutely continuous with respect to P_{Y_k} , (273) remains valid when $D(P_{Y_0} \| P_{Y_k}) = \infty$ since in that case, we can achieve arbitrarily large type-II error probability exponent (see [13, Lemmas 57-58].)

From [55], the maximum likelihood test yields

$$\max_{(k,s) \in [K]^2} \mathbb{P}[\text{Decide } H_s | H_k] \leq \exp\{-nE_C + o(n)\}, \quad (274)$$

where

$$E_C = \min_{k,s} \min_{\lambda \in (0,1)} \log \sum_{y \in \mathcal{Y}_K} P_{Y_k}(y)^{1-\lambda} P_{Y_s}(y)^\lambda \quad (275)$$

is the minimum Chernoff distance between the pairs (P_{Y_k}, P_{Y_s}) , $k \neq s \in [K]$. Combining (273) and (275), the conditions in (269)–(270) are satisfied with

$$E = \min \left\{ \min_{k \in [K]} D(P_{Y_0} \| P_{Y_k}), E_C \right\} > 0. \quad (276)$$

If the hypothesis test declares the hypothesis $H_{\hat{k}}$, $\hat{k} \neq 0$, then the receiver decides to decode $\hat{k}$ messages at one of the decoding times in $\{n_{\hat{k},1}, \dots, n_{\hat{k},L}\}$ using the VLSF code in Section D for the $\hat{k}$ -MAC, where $n_{\hat{k},1}$ is set to n_0 .

E.II. Error analysis

We here bound the probability of error for the RAC code in Definition 4.

No active transmitters: For $k = 0$, the only error event is that the composite hypothesis test at time n_0 does not declare H_0 given that H_0 is true. By (269), the probability of this event is bounded by ϵ_0 .

$k \geq 1$ active transmitters: When there is at least one active transmitter, there is an error if and only if at least one of the following events occurs:

- $\mathcal{E}_{\hat{k} \neq k}$: The number of active transmitters is estimated incorrectly at time n_0 , i.e., $\hat{k} \neq k$, which results in decoding of $\hat{k}$ messages instead of k messages.

- $\mathcal{E}_{\text{mes}}$: A list of messages $m'_{[k]} \neq m_{[k]}$ is decoded at one of the times in $\{n_{k,1}, \dots, n_{k,L}\}$.

In the following discussion, we bound the probability of these events separately, and apply the union bound to combine them.

Since the encoders are identical, the event $\mathcal{E}_{\text{rep}} = \{W_i = W_j \text{ for some } i \neq j\}$, where two or more transmitters send the same yields a dependence with transmitted codewords. Since transmitted codewords are usually independent, treating $\mathcal{E}_{\text{rep}}$ as an error simplifies the analysis. By the union bound, we have

$$\mathbb{P}[\mathcal{E}_{\text{rep}}] \leq \frac{k(k-1)}{2M}. \quad (277)$$

Applying the union bound, we bound the error probability as

$$\begin{aligned} \epsilon_k &\leq \mathbb{P}[\mathcal{E}_{\text{rep}}] + \mathbb{P}[\mathcal{E}_{\text{rep}}^c] \mathbb{P}[\mathcal{E}_{\hat{k} \neq k} \cup \mathcal{E}_{\text{mes}} | \mathcal{E}_{\text{rep}}^c] \\ &\leq \mathbb{P}[\mathcal{E}_{\text{rep}}] + \mathbb{P}[\mathcal{E}_{\hat{k} \neq k} | \mathcal{E}_{\text{rep}}^c] + \mathbb{P}[\mathcal{E}_{\text{mes}} | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c]. \end{aligned} \quad (278)$$

By (270), the probability $\mathbb{P}[\mathcal{E}_{\hat{k} \neq k} | \mathcal{E}_{\text{rep}}^c]$ is bounded as

$$\mathbb{P}[\mathcal{E}_{\hat{k} \neq k} | \mathcal{E}_{\text{rep}}^c] \leq \exp\{-n_0 E + o(n_0)\}. \quad (279)$$

Since the number of active transmitters k is not available at the decoder at time 0, we here slightly modify the stop-at-time-zero procedure from the proof sketch of Theorem 2. We set the smallest decoding time $n_{j,1}$ to $n_0 \neq 0$ for all $j \in [K]$. Given the estimate $\hat{k}$ of the number of active transmitters k , we employ the stop-at-time-zero procedure with the triple $(N', \epsilon, \epsilon'_N)$ replaced by $(N'_k, \epsilon_{\hat{k}}, \epsilon_{N'_k})$.

Let $\mathcal{E}_{\text{stop}}$ denote the event that the decoder chooses to stop at time $n_{k,1} = n_0$. We bound $\mathbb{P}[\mathcal{E}_{\text{mes}} | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c]$ as

$$\begin{aligned} \mathbb{P}[\mathcal{E}_{\text{mes}} | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c] &\leq \mathbb{P}[\mathcal{E}_{\text{stop}} | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c] \\ &\quad + \mathbb{P}[\mathcal{E}_{\text{stop}}^c | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c] \mathbb{P}[\mathcal{E}_{\text{mes}} | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c \cap \mathcal{E}_{\text{stop}}^c]. \end{aligned} \quad (280)$$

By Theorem 4, when the RAC decoder decodes a list of k messages from $[M]$ at time $n_{k,\ell}$, we get

$$\mathbb{P}[\mathcal{E}_{\text{mes}} | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c \cap \mathcal{E}_{\text{stop}}^c] \quad (281)$$

$$\leq \mathbb{P}[\iota_k(X_{[k]}^{n_{k,L}}, Y_k^{n_{k,L}}) < \gamma_k] \quad (282)$$

$$+ \binom{M-k}{k} \exp\{-\gamma_k\} \quad (283)$$

$$\begin{aligned} &+ \sum_{\ell=2}^L \sum_{\mathcal{A} \in \mathcal{P}([k])} \\ &\quad \mathbb{P}[\iota_k(X_{\mathcal{A}}^{n_{k,\ell}}, Y_k^{n_{k,\ell}}) > N'_k I_k(X_{\mathcal{A}}; Y_k) + N'_k \lambda^{(k,\mathcal{A})}] \end{aligned} \quad (284)$$

$$\begin{aligned} &+ \sum_{\mathcal{A} \in \mathcal{P}([k])} \binom{M-k}{|\mathcal{A}|} \\ &\quad \exp\{-\gamma + N'_k I_k(X_{\mathcal{A}}; Y_k) + N'_k \lambda^{(k,\mathcal{A})}\}, \end{aligned} \quad (285)$$

where N'_k is the average decoding time given $\mathcal{E}_{\text{stop}}^c$, and γ_k and $\lambda^{(k,\mathcal{A})}$ are constants chosen to satisfy the equations

$$\gamma_k = n_{k,\ell} I_k - \sqrt{n_{k,\ell} \log_{(L-\ell+1)}(n_{k,\ell}) V_k} \quad (286)$$

$$= k \log M + \log N'_k + O(1) \quad (287)$$

for all $\ell \in \{2, \dots, L\}$, and

$$\lambda^{(k,\mathcal{A})} = \frac{N'_k I_k(X_{\mathcal{A}^c}; Y_k | X_{\mathcal{A}}) - |\mathcal{A}^c| \log M}{2N'_k}, \quad \mathcal{A} \in \mathcal{P}([k]). \quad (288)$$

The fact that each $\lambda^{(k,\mathcal{A})}$ is bounded below by a positive constant follows from (287), [28, Lemma 1], and the symmetry assumptions on the RAC.

Following the analysis in Appendix D.II, we conclude that

$$\begin{aligned} k \log M &= N'_k I_k - \sqrt{N'_k \log_{(L-1)}(N'_k) V_k} \\ &\quad + O\left(\sqrt{\frac{N'_k V_k}{\log_{(L-1)}(N'_k)}}\right) \end{aligned} \quad (289)$$

$$\mathbb{P}[\mathcal{E}_{\text{mes}} | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c \cap \mathcal{E}_{\text{stop}}^c] \leq \frac{1}{\sqrt{N'_k \log N'_k}}. \quad (290)$$

Note that by (277) and (289), the bound on $\mathbb{P}[\mathcal{E}_{\text{rep}}]$ decays exponentially with N_k . A consequence of (286) and (289) is that

$$N'_k = n_{k,\ell}(1 + o(1)) \quad (291)$$

for all $\ell \geq 2$ and $k \in [K]$.

Note that from (289), the right-hand side of (277) is bounded by $\frac{1}{N'_k}$ for N'_k large enough. We set the time n_0 so that the right-hand side of (279) is bounded by $\frac{1}{4\sqrt{N'_k \log N'_k}}$ for all $k \in [K]$. This condition is satisfied if

$$n_0 \geq \frac{1}{2E} \log N'_k + o(\log N'_k). \quad (292)$$

The above arguments imply that

$$\mathbb{P}[\mathcal{E}_{\text{rep}}] + \mathbb{P}[\mathcal{E}_{\hat{k} \neq k} | \mathcal{E}_{\text{rep}}^c] \leq \frac{1}{2\sqrt{N'_k \log N'_k}} \quad (293)$$

for N'_k large enough. As in the DM-MAC case, we set

$$p \triangleq \mathbb{P}[\mathcal{E}_{\text{stop}} | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c] = \frac{\epsilon'_k - \frac{1}{\sqrt{N'_k \log N'_k}}}{1 - \frac{1}{\sqrt{N'_k \log N'_k}}} \quad (294)$$

where

$$\epsilon'_k = \epsilon_k - \frac{1}{2\sqrt{N'_k \log N'_k}}. \quad (295)$$

Combining (278), (280), (290), and (293)–(294), the error probability of the RAC code is bounded by

$$\begin{aligned} &\mathbb{P}[\mathcal{E}_{\text{rep}}] + \mathbb{P}[\mathcal{E}_{\hat{k} \neq k} | \mathcal{E}_{\text{rep}}^c] + \mathbb{P}[\mathcal{E}_{\text{stop}} | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c] \\ &\quad + \mathbb{P}[\mathcal{E}_{\text{stop}}^c | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c] \mathbb{P}[\mathcal{E}_{\text{mes}} | \mathcal{E}_{\text{rep}}^c \cap \mathcal{E}_{\hat{k} \neq k}^c \cap \mathcal{E}_{\text{stop}}^c] \end{aligned} \quad (296)$$

$$\leq \frac{1}{2\sqrt{N'_k \log N'_k}} + p + (1-p) \frac{1}{\sqrt{N'_k \log N'_k}} \quad (297)$$

$$= \epsilon_k. \quad (298)$$

The average decoding time of the code is bounded as

$$N_k \leq \mathbb{E} \left[\tau_k^* | \mathcal{E}_{k \neq k}^* \cup \mathcal{E}_{\text{rep}} \right] \mathbb{P} \left[\mathcal{E}_{k \neq k}^* \cup \mathcal{E}_{\text{rep}} \right] + \mathbb{E} \left[\tau_k^* | \mathcal{E}_{k \neq k}^c \cap \mathcal{E}_{\text{rep}}^c \right] \mathbb{P} \left[\mathcal{E}_{k \neq k}^c \cap \mathcal{E}_{\text{rep}}^c \right] \quad (299)$$

$$\leq \frac{N_{K,L}}{2\sqrt{N'_k \log N'_k}} + n_0 p + N'_k (1-p). \quad (300)$$

From (291) and (294)–(295), we get

$$N'_k = \frac{N_k}{1 - \epsilon'_k} \left(1 + O \left(\frac{1}{\sqrt{N_k \log N_k}} \right) \right). \quad (301)$$

Plugging (301) into (289) completes the proof.

APPENDIX F PROOF OF THEOREM 7

The non-asymptotic achievability bound in Theorem 1 applies to the Gaussian PPC with maximal power constraint P (58) with the modification that the error probability (18) has an additional term for power constraint violations

$$\mathbb{P} \left[\bigcup_{\ell=1}^L \left\{ \|X^{n_\ell}\|^2 > n_\ell P \right\} \right]. \quad (302)$$

The proof follows similarly to the proof of Theorem 2 as we employ the stop-at-time-zero procedure in the proof sketch of Theorem 2. We extend Lemma 1 to the Gaussian PPC, showing

$$\begin{aligned} & \log M^* \left(N, L, \frac{1}{\sqrt{N \log N}}, P \right) \\ & \geq NC(P) - \sqrt{N \log_{(L)}(N) V(P)} + O \left(\sqrt{\frac{N}{\log_{(L)}(N)}} \right). \end{aligned} \quad (303)$$

The input distribution $P_{X^{n_L}}$ used in the proof of (303) differs from the one used in the proof of Lemma 1, causing changes in the analysis on the probability $\mathbb{P}[\iota(X^{n_L}; Y^{n_L}) < \gamma]$ and the threshold γ in (113). Below, we detail these differences.

1) *The input distribution $P_{X^{n_L}}$:* We choose the distribution of the random codewords, $P_{X^{n_L}}$, in Theorem 1 as follows. Set $n_0 = 0$. For each codeword, independently draw sub-codewords $X^{n_{j-1}+1:n_j}$, $j \in [L]$ from the uniform distribution on the $(n_j - n_{j-1})$ -dimensional sphere of radius $\sqrt{(n_j - n_{j-1})P}$. Let $P_{X^{n_L}}$ denote the distribution of the length- n_L random codewords described above. Since codewords chosen under $P_{X^{n_L}}$ never violate the power constraint (58), the power violation probability in (302) is 0. Furthermore, the power constraint is satisfied with equality at each of the dimensions $n_1, \dots, n_L$; our analysis in [39] shows that for any finite L , and sufficiently large increments $n_\ell - n_{\ell-1}$ for all $\ell \in [L]$, using this restricted subset instead of the entire n_L -dimensional power sphere results in no change in the asymptotic expansion (60) for the fixed-length no-feedback codes up to the third-order term.

2) *Bounding the probability of the information density random variable:* We begin by bounding the probability

$$\mathbb{P}[\iota(X^{n_\ell}; Y^{n_\ell}) < \gamma], \quad \ell \in [L], \quad (304)$$

that appears in Theorem 1 under the input distribution described above. Under this choice of $P_{X^{n_L}}$, the random variable $\iota(X^{n_\ell}; Y^{n_\ell})$ is not a sum of n_ℓ i.i.d. random variables. We wish to apply the moderate deviations result in Lemma 2. To do this, we first introduce the following lemma from [56], which uniformly bounds the Radon-Nikodym derivative of the channel output distribution in response to the uniform distribution on a sphere as compared to the channel output distribution in response to i.i.d. Gaussian inputs.

Lemma 5 (MolavianJazi and Laneman [56, Prop. 2]): Let X^n be distributed uniformly over the n -dimensional sphere of radius $\sqrt{nP}$. Let $\tilde{X}^n \sim \mathcal{N}(\mathbf{0}, P\mathbf{I}_n)$. Let P_{Y^n} and $P_{\tilde{Y}^n}$ denote the channel output distributions in response to P_{X^n} and $P_{\tilde{X}^n}$, respectively, where $P_{Y^n|X^n}$ is the point-to-point Gaussian channel (55). Then there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ and $y^n \in \mathbb{R}^n$, it holds that

$$\frac{dP_{Y^n}(y^n)}{dP_{\tilde{Y}^n}(y^n)} \leq J(P) \triangleq 27\sqrt{\frac{\pi}{8}} \frac{1+P}{\sqrt{1+2P}}. \quad (305)$$

Let $P_{\tilde{Y}^n}^{n_\ell}$ be $\mathcal{N}(\mathbf{0}, (1+P)\mathbf{I}_{n_\ell})$. By Lemma 5, we bound (304) as

$$\begin{aligned} & \mathbb{P}[\iota(X^{n_\ell}; Y^{n_\ell}) < \gamma] \\ & = \mathbb{P} \left[\log \frac{dP_{Y^{n_\ell}|X^{n_\ell}}(Y^{n_\ell}|X^{n_\ell})}{dP_{\tilde{Y}^{n_\ell}}(Y^{n_\ell})} < \gamma + \log \frac{dP_{Y^{n_\ell}}(Y^{n_\ell})}{dP_{\tilde{Y}^{n_\ell}}(Y^{n_\ell})} \right] \end{aligned} \quad (306)$$

$$\leq \mathbb{P} \left[\log \frac{dP_{Y^{n_\ell}|X^{n_\ell}}(Y^{n_\ell}|X^{n_\ell})}{dP_{\tilde{Y}^{n_\ell}}(Y^{n_\ell})} < \gamma + \ell \log J(P) \right], \quad (307)$$

where $J(P)$ is the constant given in (305), and (307) follows from the fact that $P_{Y^{n_\ell}}$ is the product of ℓ output distributions of dimensions $n_j - n_{j-1}$, $j \in [\ell]$, each induced by a uniform distribution over a sphere of the corresponding radius. As argued in [13], [35], [39], [56], by spherical symmetry, the distribution of the random variable

$$\log \frac{dP_{Y^{n_\ell}|X^{n_\ell}}(Y^{n_\ell}|X^{n_\ell})}{dP_{\tilde{Y}^{n_\ell}}(Y^{n_\ell})} \quad (308)$$

depends on X^{n_ℓ} only through its norm $\|X^{n_\ell}\|$. Since $\|X^{n_\ell}\|^2 = n_\ell P$ with probability 1, any choice of x^{n_ℓ} such that $\|x^{n_i}\|^2 = n_i P$ for $i \in [\ell]$ gives

$$\begin{aligned} & \mathbb{P} \left[\log \frac{dP_{Y^{n_\ell}|X^{n_\ell}}(Y^{n_\ell}|X^{n_\ell})}{dP_{\tilde{Y}^{n_\ell}}(Y^{n_\ell})} < \gamma + \ell \log J(P) \right] = \\ & \mathbb{P} \left[\log \frac{dP_{Y^{n_\ell}|X^{n_\ell}}(Y^{n_\ell}|X^{n_\ell})}{dP_{\tilde{Y}^{n_\ell}}(Y^{n_\ell})} < \gamma + \ell \log J(P) \mid X^{n_\ell} = x^{n_\ell} \right]. \end{aligned} \quad (309)$$

We set $x^{n_\ell} = (\sqrt{P}, \sqrt{P}, \dots, \sqrt{P}) = \sqrt{P}\mathbf{1}$ to obtain an i.i.d. sum in (309). Given $X^{n_\ell} = \sqrt{P}\mathbf{1}$, the distribution of (308) is the same as the distribution of the sum

$$\sum_{i=1}^{n_\ell} A_i \quad (310)$$

of n_ℓ i.i.d. random variables

$$A_i = C(P) + \frac{P}{2(1+P)} \left(1 - Z_i^2 + \frac{2}{\sqrt{P}} Z_i \right), \quad i \in [n_\ell], \quad (311)$$

where $Z_1, \dots, Z_{n_\ell}$ are drawn independently from $\mathcal{N}(0, 1)$ (see e.g., [13, eq. (205)]). The mean and variance of A_1 are

$$\mathbb{E}[A_1] = C(P) \quad (312)$$

$$\text{Var}[A_1] = V(P). \quad (313)$$

From (307)–(310), we get

$$\mathbb{P}[\iota(X^{n_\ell}; Y^{n_\ell}) < \gamma] \leq \mathbb{P} \left[\sum_{i=1}^{n_\ell} A_i < \gamma + \ell \log J(P) \right]. \quad (314)$$

To verify that Lemma 2 is applicable to the right-hand side of (314), it only remains to show that $\mathbb{E}[(A_1 - C(P))^3]$ is finite, and $A_1 - C(P)$ satisfies Cramér's condition, that is, there exists some $t_0 > 0$ such that $\mathbb{E}[\exp\{t(A_1 - C(P))\}] < \infty$ for all $|t| < t_0$. From (311), $(A_1 - C(P))^3$ has the same distribution as a 6-degree polynomial of the Gaussian random variable $Z \sim \mathcal{N}(0, 1)$. This polynomial has a finite mean since all moments of Z are finite. Let $c \triangleq \frac{P}{2(1+P)}$, $f \triangleq \frac{2}{\sqrt{P}}$, and $t' \triangleq tc$. To show that Cramér's condition holds, we compute

$$\begin{aligned} & \mathbb{E}[\exp\{t(A_1 - C(P))\}] \\ &= \mathbb{E}[\exp\{t'(1 - Z^2 + fZ)\}] \end{aligned} \quad (315)$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2} + t'(1 - x^2 + fx) \right\} dx \quad (316)$$

$$= \frac{1}{\sqrt{1+2t'}} \exp \left\{ t' + \frac{t'f}{2(1+2t')} \right\}. \quad (317)$$

Thus, $\mathbb{E}[\exp\{t(A_1 - C(P))\}] < \infty$ for $t' > -\frac{1}{2}$, and $t_0 = \frac{1}{2c} > 0$ satisfies Cramér's condition.

3) *The threshold γ* : We set $\gamma, n_1, \dots, n_L$ so that the equalities

$$\gamma = n_\ell C(P) - \sqrt{n_\ell \log_{(L-\ell+1)}(n_\ell) V(P)} - \ell \log J(P) \quad (318)$$

hold for all $\ell \in [L]$.

The rest of the proof follows identically to (118)–(133) with C and V replaced by $C(P)$ and $V(P)$, respectively, giving

$$\begin{aligned} \log M &\geq NC(P) - \sqrt{N \log_{(L)}(N) V(P)} \\ &\quad - \frac{1}{\sqrt{2\pi}} \sqrt{\frac{NV(P)}{\log_{(L)}(N)}} (1 + o(1)) - \log N - L \log J(P), \end{aligned} \quad (319)$$

which completes the proof.

REFERENCES

- [1] 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)*, July 2021, pp. 1883–1888.
- [2] —, "Nested sparse feedback codes for point-to-point, multiple access, and random access channels," in *IEEE Information Theory Workshop (ITW)*, Kanazawa, Japan, Oct. 2021, pp. 1–6.
- [3] C. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.
- [4] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 9, no. 3, pp. 136–143, July 1963.
- [5] 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.
- [6] 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.
- [7] 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, Oct. 1976.
- [8] 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.
- [9] A. Tchamkerten and I. E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2126–2145, May 2006.
- [10] M. Naghshvar, T. Javidi, and M. Wigger, "Extrinsic Jensen–Shannon divergence: Applications to variable-length coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 2148–2164, Apr. 2015.
- [11] H. Yang, M. Pan, A. Antonini, and R. D. Wesel, "Sequential transmission over binary asymmetric channels with feedback," *IEEE Trans. Inf. Theory*, vol. 68, no. 11, pp. 7023–7042, Nov. 2022.
- [12] N. Guo and V. Kostina, "Reliability function for streaming over a DMC with feedback," *IEEE Trans. Inf. Theory*, vol. 69, no. 4, pp. 2165–2192, Apr. 2023.
- [13] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [14] 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, June 2015, pp. 311–315.
- [15] 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.
- [16] J. Östman, R. Devassy, G. Durisi, and E. G. Ström, "Short-packet transmission via variable-length codes in the presence of noisy stop feedback," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 214–227, Jan. 2021.
- [17] G. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, Mar. 1968.
- [18] S. Ginzach, N. Merhav, and I. Sason, "Random-coding error exponent of variable-length codes with a single-bit noiseless feedback," in *IEEE Inf. Theory Workshop (ITW)*, Kaohsiung, Taiwan, Nov. 2017, pp. 584–588.
- [19] 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.
- [20] —, "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.
- [21] 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.
- [22] M. Heidari, A. Anastasopoulos, and S. S. Pradhan, "On the reliability function of discrete memoryless multiple-access channel with feedback," in *2018 IEEE Information Theory Workshop (ITW)*, Guangzhou, China, Nov. 2018, pp. 1–5.
- [23] K. F. Trillingsgaard and P. Popovski, "Variable-length coding for short packets over a multiple access channel with feedback," in *2014 11th International Symposium on Wireless Communications Systems (ISWCS)*, Barcelona, Spain, Aug. 2014, pp. 796–800.
- [24] 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.
- [25] A. R. Williamson, T. Chen, and R. D. Wesel, "Variable-length convolutional coding for short blocklengths with decision feedback," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2389–2403, July 2015.
- [26] 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, June 2016.

- [27] 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.
- [28] 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.
- [29] Y. Liu and M. Effros, "Finite-blocklength and error-exponent analyses for LDPC codes in point-to-point and multiple access communication," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angeles, California, USA, June 2020, pp. 361–366.
- [30] H. Yang, R. C. Yavas, V. Kostina, and R. D. Wesel, "Variable-length stop-feedback codes with finite optimal decoding times for BI-AWGN channels," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Espoo, Finland, July 2022, pp. 2327–2332.
- [31] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed. John Wiley & Sons, 1971, vol. II.
- [32] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback (corresp.)," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, pp. 729–733, Nov. 1979.
- [33] A. Lalitha and T. Javidi, "On error exponents of almost-fixed-length channel codes and hypothesis tests," *arXiv:2012.00077*, Nov. 2020.
- [34] P. Berlin, B. Nakiboğlu, B. Rimoldi, and E. Telatar, "A simple converse of Burnashev's reliability function," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3074–3080, July 2009.
- [35] 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.
- [36] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, 1st ed. Chapman and Hall CRC, 2014.
- [37] V. Y. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 881–903, Feb. 2014.
- [38] O. Kosut, "A second-order converse bound for the multiple-access channel via wringing dependence," *IEEE Tran. Inf. Theory*, vol. 68, no. 6, pp. 3552–3584, June 2022.
- [39] R. C. Yavas, V. Kostina, and M. Effros, "Gaussian multiple and random access channels: Finite-blocklength analysis," *IEEE Trans. Inf. Theory*, vol. 67, no. 11, pp. 6983–7009, Nov. 2021.
- [40] M. Tomamichel and V. Y. F. Tan, "A tight upper bound for the third-order asymptotics of discrete memoryless channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, July 2013, pp. 1536–1540.
- [41] P. Moulin, "The log-volume of optimal codes for memoryless channels, asymptotically within a few nats," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2278–2313, Apr. 2017.
- [42] V. V. Petrov, *Sums of independent random variables*. New York, USA: Springer, Berlin, Heidelberg, 1975.
- [43] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. NJ, USA: Wiley, 2006.
- [44] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, June 2017, pp. 2523–2527.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] 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.
- [49] E. MolavianJazi, "A unified approach to Gaussian channels with finite blocklength," Ph.D. dissertation, University of Notre Dame, July 2014.
- [50] 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, USA, Sep. 2010, pp. 1334–1339.
- [51] R. W. Butler, *Saddlepoint Approximations with Applications*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2007.
- [52] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, vol. 16, no. 2, pp. 117–186, June 1945.
- [53] Y. Li and V. Y. F. Tan, "Second-order asymptotics of sequential hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 7222–7230, Nov. 2020.
- [54] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *The Annals of Mathematical Statistics*, vol. 19, no. 3, pp. 326–339, Sep. 1948.
- [55] C. Leang and D. Johnson, "On the asymptotics of m-hypothesis Bayesian detection," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 280–282, 1997.
- [56] 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.

Recep Can Yavas (S'18–M'22) received the B.S. degree (Hons.) in electrical engineering from Bilkent University, Ankara, Turkey, in 2016. He received the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology (Caltech) in 2017 and 2023, respectively. He is currently a research fellow at CNRS at CREATE, Singapore. His research interests include information theory, probability theory, and multi-armed bandits.

Victoria Kostina (S'12–M'14–SM'22) is a professor of electrical engineering and of computing and mathematical sciences at Caltech. She received the bachelor's degree from Moscow Institute of Physics and Technology (MIPT) in 2004, the master's degree from University of Ottawa in 2006, and the Ph.D. degree from Princeton University in 2013. During her studies at MIPT, she was affiliated with the Institute for Information Transmission Problems of the Russian Academy of Sciences.

Her research interests lie in information theory, coding, communications, learning, and control. She has served as an Associate Editor for IEEE Transactions of Information Theory, and as a Guest Editor for the IEEE Journal on Selected Areas in Information Theory. She received the Natural Sciences and Engineering Research Council of Canada postgraduate scholarship during 2009–2012, Princeton Electrical Engineering Best Dissertation Award in 2013, Simons-Berkeley research fellowship in 2015 and the NSF CAREER award in 2017.

Michelle Effros (S'93–M'95–SM'03–F'09) is the George Van Osdel Professor of Electrical Engineering and Vice Provost at the California Institute of Technology. She was a co-founder of Code On Technologies, a technology licensing firm, which was sold in 2016. Dr. Effros is a fellow of the IEEE and has received a number of awards including Stanford's Frederick Emmons Terman Engineering Scholastic Award (for excellence in engineering), the Hughes Masters Full-Study Fellowship, the National Science Foundation Graduate Fellowship, the AT&T Ph.D. Scholarship, the NSF CAREER Award, the Charles Lee Powell Foundation Award, the Richard Feynman-Hughes Fellowship, an Okawa Research Grant, and the Communications Society and Information Theory Society Joint Paper Award. She was cited by Technology Review as one of the world's top 100 young innovators in 2002, became a fellow of the IEEE in 2009, and is a member of Tau Beta Pi, Phi Beta Kappa, and Sigma Xi. She received the B.S. (with distinction), M.S., and Ph.D. degrees in electrical engineering from Stanford University. Her research interests include information theory (with a focus on source, channel, and network coding for multi-node networks) and theoretical neuroscience (with a focus on neurostability and memory).

Dr. Effros served as the Editor of the IEEE Information Theory Society Newsletter from 1995 to 1998 and as a Member of the Board of Governors of the IEEE Information Theory Society from 1998 to 2003 and from 2008 to 2017. She served as President of the IEEE Information Theory Society in 2015 and as Executive Director for the film “The Bit Player,” a movie about Claude Shannon, which came out in 2018. She was a member of the Advisory Committee and the Committee of Visitors for the Computer and Information Science and Engineering (CISE) Directorate at the National Science Foundation from 2009 to 2012 and in 2014, respectively. She served on the IEEE Signal Processing Society Image and Multi-Dimensional Signal Processing (IMDSP) Technical Committee from 2001 to 2007 and on ISAT from 2006 to 2009. She served as Associate Editor for the joint special issue on Networking and Information Theory in the IEEE Transactions on Information Theory and the IEEE/ACM Transactions on Networking, as Associate Editor for the special issue honoring the scientific legacy of Ralf Koetter in the IEEE Transactions on Information Theory and, from 2004 to 2007 served as Associate Editor for Source Coding for the IEEE Transactions on Information Theory. She has served on numerous technical program committees and review boards, including serving as general co-chair for the 2009 Network Coding Workshop and technical program committee co-chair for the 2012 IEEE International Symposium on Information Theory and the 2023 IEEE Information Theory Workshop.