Random Access Channel Coding in the Finite Blocklength Regime

Recep Can Yavas, Victoria Kostina, and Michelle Effros

***Abstract*—Consider a random access communication scenario over a channel whose operation is defined for any number of possible transmitters. As in the model recently introduced by Polyanskiy for the Multiple Access Channel (MAC) with a fixed, known number of transmitters, the channel is assumed to be invariant to permutations on its inputs, and all active transmitters employ identical encoders. Unlike the Polyanskiy model, in the proposed scenario, neither the transmitters nor the receiver knows which transmitters are active. We refer to this agnostic communication setup as the Random Access Channel (RAC). Scheduled feedback of a finite number of bits is used to synchronize the transmitters. The decoder is tasked with determining from the channel output the number of active transmitters, k, and their messages but not which transmitter sent which message. The decoding procedure occurs at a time n_t depending on the decoder's estimate, t, of the number of active transmitters, k, thereby achieving a rate that varies with the number of active transmitters. Single-bit feedback at each time $n_i, i \leq t$, enables all transmitters to determine the end of one coding epoch and the start of the next. The central result of this work demonstrates the achievability on a RAC of performance that is first-order optimal for the MAC in operation during each coding epoch. While prior multiple access schemes for a fixed number of transmitters require $2^k - 1$ simultaneous threshold rules, the proposed scheme uses a single threshold rule and achieves the same dispersion.**

***Index Terms*—Channel coding, random access channel, finite blocklength regime, achievability, second-order asymptotics, rateless codes.**

I. Introduction

Access points like WiFi hot spots and cellular base stations are, for wireless devices, the gateway to the network. Unfortunately, access points are also the network's most critical bottleneck. As more kinds of devices become network-reliant, both the number of communicating devices and the diversity of their communication needs grow. Little is known about how to code under high variation in the number and variety of communicators.

Multiple-transmitter single-receiver channels are well understood in information theory when the number and identities of transmitters are fixed and known. Unfortunately, even in this known-transmitter regime, information-theoretic solutions are too complex to implement. As a result, orthogonalization methods, such as TDMA, FDMA, and orthogonal CDMA, are used instead. Orthogonalization strategies simplify coding by allocating resources (e.g., time slots) among the transmitters, but applying such methods to discrete memoryless MACs can at best attain a sum-rate equal to the single-transmitter capacity of the channel, which is often significantly smaller than the maximal multi-transmitter sum-rate.

Most random access protocols currently in use rely on collision avoidance, which cannot surpass the single-transmitter capacity of the channel and may be significantly worse since the unknown transmitter set makes it difficult to schedule or coordinate among transmitters. Collision avoidance is achieved through variations of the legacy (slotted) ALOHA and carrier sense multiple access (CSMA) algorithms. ALOHA, which uses random transmission times and back-off schedules, achieves only about 37% of the single-transmitter capacity of the channel [2]. In CSMA, each transmitter tries to avoid collisions by verifying the absence of other traffic before starting a transmission over the shared channel; when collisions do occur, all transmissions are aborted, and a jamming signal is sent to ensure that all transmitters are aware of the collision. The procedure starts again at a random time, which again introduces inefficiencies. The state of the art in random access coding is "treating interference as noise," which is part of newer CDMA-based standards. While this strategy can deal with random access better than ALOHA, it is still far inferior to the theoretical limits.

Even from a purely theoretical perspective, a satisfactory solution to random access remains to be found. The MAC model in which a fixed number, k, out of the total available K transmitters are active was studied by D'yachkov and Rykov [3] and Mathys [4] for zero-error coding on a noiseless adder MAC, and by Bassalygo and Pinsker [5] for an asynchronous model in which the information is considered erased if more than one transmitter is active at a time. See [6] for a more detailed history. Two-layer MAC decoders, with outer layer codes that work to remove channel noise and inner layer codes that work to resolve conflicts, are proposed in [7], [8]. Like the codes in [3]–[5], the codes in [6], [7] are designed for a predetermined number of transmitters, k; it is not clear how robust they are to randomness in the transmitters' arrivals and departures. In [9], Minero et al. study a random access model in which the receiver knows the transmitter activity pattern, and the transmitters opportunistically send data at the highest possible rate. The receiver recovers only a portion of the messages sent, depending on the current level of activity in the channel.

R. C. Yavas, V. Kostina, and M. Effros are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA. E-mails: {*ryavas, vkostina, effros*}*@caltech.edu*. This work was supported in part by the National Science Foundation (NSF) under grant CCF-1817241. A part of this work was presented at ISIT'18 [1].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2020.3047630, IEEE Transactions on Information Theory

A. Our Contributions and Related Works

This paper poses the question of whether it is possible, in a scenario where no one knows how many transmitters are active, for the receiver to almost always recover the messages sent by all active transmitters. Surprisingly, we find that not only is reliable decoding possible in this regime, but, for the class of permutation-invariant channels considered in [6], our proposed RAC code performs as well in its capacity and dispersion terms as the best-known code for a MAC with the transmitter activity known a priori [10]–[13]. Since the capacity region of a MAC varies with the number of transmitters, it is tempting to believe that the transmitters of a random access system must somehow vary their codebook size in order to match their transmission rate to the capacity region of the MAC in operation. Instead, we here allow the decoder to vary its decoding time depending on the observed channel output—thereby adjusting the rate at which each transmitter communicates by changing not the size but the blocklength of each transmitter's codebook.

Codes that can accommodate variable decoding times are called *rateless codes*. Rateless codes originate with the work of Burnashev [14], who computed the error exponent of variable-length coding over a known point-to-point channel. Polyanskiy et al. [15] provide a dispersion-style analysis of the same scenario. A practical implementation of rateless codes for an erasure channel with an unknown erasure probability appears in [16]. An analysis of rateless coding over an unknown binary symmetric channel appears in [17] and is extended to an arbitrary discrete memoryless channel in [18], [19] using a decoder that tracks Goppa's empirical mutual information and decodes once that quantity passes a threshold. In [20], Jeffrey's prior is used to weight unknown channels. A rateless code for noiseless random access communication is described in [21]; each user transmits replicas of its message in multiple time slots, possibly colliding with the messages of other transmitters. At the end of each time slot, the decoder attempts to apply successive interference cancellation starting with the messages received without collision and subsequently removing the associated interference from the time slots in which replicas are transmitted. The decoder then decides whether to terminate an epoch or to ask the transmitters to send more replicas.

Unlike the codes described in [14]–[21], which allow truly arbitrary decoding times, in this paper we allow decoding only at a predetermined list of possible times $n_0, n_1, n_2, \ldots$. This strategy both eases practical implementation and reduces feedback. In particular, the schemes in [14]–[21] transmit a single-bit acknowledgment message from the decoder to the encoder(s) once the decoder completes its decoding process. Because the decoding time is random, this so-called "single-bit" feedback forces the transmitter(s) to listen to the channel constantly, at every time step trying to discern whether or not a transmission was received. This either requires full-duplex devices or doubles the effective blocklength and can be quite expensive. Thus while the receiver technically sends only "one bit" of feedback, the transmitters receive one bit of feedback (with the alphabet {"transmission","no transmission"}) in every time step, giving a feedback rate of 1 bit per channel use rather than a total of 1 bit. In our framework, acknowledgment bits are sent only at times $n_0, n_1, n_2, \ldots, n_t$, where each n_i is the pre-determined decoding time used if the receiver believes that i transmitters are active. Thus the transmitters must listen only at a finite collection of time steps. For example, when $n_0 < n_1 < \cdots < n_t$, as is assumed here for simplicity, the total number of feedback bits equals one plus the receiver's estimate of the number of transmitters, a feedback rate approaching 0 bits per channel use as the blocklength grows.

In the central portion of this paper, we view the random access channel as a collection of all possible MACs that might arise as a result of the transmitter activity pattern. Barring the intricacies of multiuser decoding, the model that views an unknown channel as a collection of possible channels without assigning an a priori probability to each is known as the *compound channel* model [22]. In the context of single-transmitter compound channels, it is known that if the decoding time is fixed, the transmission rate cannot exceed the capacity of the weakest channel from the collection [22], though the dispersion may be better (smaller) [23]. With feedback and a variable decoding time, one can do much better [17]–[20].

In [6], Polyanskiy argues for removing the transmitter identification task from the physical layer encoding and decoding procedures of a MAC. As he points out, such a scenario was previously discussed by Berger [24] in the context of conflict resolution. Polyanskiy further suggests studying MACs whose conditional channel output distributions are insensitive to input permutations. For such channels, if all transmitters use the same codebook, then the receiver can at best hope to recover the messages sent without recovering who transmitted which message (the transmitter identity). In some networks the transmitter identification task can be insignificant. For example, in some sensor networks, we might be interested in the collected measurements but indifferent to the identities of the collecting sensors. In scenarios where transmitter identity is required, it can be included in the payload.

In Section IV, we propose a code for a random access communication channel model built from a family of permutation-invariant MACs. Our code employs identical encoders at all transmitters and identity-blind decoding at the receiver. Although not critical for the feasibility of our approach, these assumptions lead to a number of pleasing simplifications of both our scheme and its analysis. For example, using identical encoders at all transmitters simplifies design and implementation. Further, the collection of MACs comprising our compound RAC model can be parameterized by the number of active transmitters rather than by the full transmitter activity pattern.

We provide a second-order analysis of the rate universally achieved by our multiuser scheme over all transmitter activity patterns, taking into account the possibility that the decoder may misdetect the current activity pattern and decode for the wrong channel. Leveraging our observation that for a symmetric MAC, the fair rate point is not a corner point of the capacity region, we are able to show that a single-threshold decoding rule attains the fair rate point. This differs significantly from traditional MAC analyses, which use $2^k - 1$ simultaneous threshold rules. In the context of a MAC with a known number of transmitters, second-order analyses of multiple-threshold decoding rules are given in [10]–[13] (finite alphabet MAC) and in [25] (Gaussian MAC). A non-asymptotic analysis of variable-length coding with "single-bit" feedback over a (known) Gaussian MAC appears in [26].

Other relevant recent works on the MAC include the following. To account for massive numbers of transmitters, in [27], [28], Chen and Guo introduce a notion of capacity for the multiple access scenario in which the maximal number of transmitters grows with the blocklength and an unknown subset of transmitters is active at a given time. They show that time sharing, which achieves the conventional MAC capacity, is inadequate to achieve capacity in that regime. In [29], Sarwate and Gastpar show that rate-0 feedback, such as the feedback in our approach, does not increase the capacity of the discrete memoryless MAC. In compound MACs, limited feedback can increase capacity. For example, one strategy uses a simple training phase to estimate the channel state and employs feedback to send the state estimate to the transmitter. Such schemes cannot increase the capacity beyond the rate achievable when the state is known to the encoders and the decoder [29].

The sparse recovery problem is identical to a special case of the RAC problem in which each transmitter sends only its "signature" to the receiver. Here, the decoder's only task is to determine who is active. Active transmitters in this variant of the RAC problem may correspond to defective items or positive test outcomes in the sparse recovery problem, and successful decoding is identified with successfully detecting the set of defective or confirmed-positive elements. A group testing problem in which an unknown subset of k defective items out of K items total is observed through an OR MAC, is studied in [30]–[34]; this problem is a special case of the sparse recovery problem. In these works, the decoder reaches a conclusion about tested items at a fixed blocklength n. Atia and Saligrama [32] consider a noiseless group testing scenario in which the number of transmitted elements, k, does not grow with the total number of elements, K, showing that the smallest possible number of measurements needed to detect the defective items is $O(k \log \frac{K}{k})$. In in [33], Scarlett and Cevher extend this result to the scenario where k scales as $O(K^\theta)$ for $\theta \in (0, 1)$. In [34], Scarlett and Cevher derive the information-theoretic limits of the exact and partial support recovery problems for general probabilistic models, where exact recovery refers to detecting all k defective items, and partial recovery refers to detecting at least s out of k defective items. While we consider a nonvanishing average error probability and operate in the central limit theorem regime, [30]–[34] assume vanishing average error probability and operate in the large deviations regime. The main difference between the decoder designs in [30]–[34] and our decoder design is that [30]–[34] use $2^k - 1$ simultaneous information density threshold tests at a single blocklength n, while our decoder uses a single information density threshold test at multiple decoding times, allowing successful detection with a computationally less complex decoder even when the number of active transmitters to be detected is unknown.

B. Paper Organization

Our system model and proposed communication strategy are laid out in Section II. The main result, showing that for a nontrivial class of channels our proposed RAC code performs as well in terms of capacity and dispersion as the best-known code for a MAC with the transmitter activity known a priori, is presented in Section III. The proofs are presented in Section IV. Section V includes discussions of the effect of using maximum likelihood decoding, the choice of an input distribution in the random code design, the difficulties in proving a converse, an extension of our strategy that enables transmitter identity decoding, and performance bounds under the per-user error probability criterion. Interestingly, the problem of decoding for $k \geq 1$ unknown transmitters is substantially different from the problem of detecting whether there are any active transmitters at all. In Section VI, we employ universal hypothesis testing to solve the latter problem. Section VII concludes the paper with a discussion of our results and their implications.

II. Problem Setup

For any positive integers i, j, let $[i] = \{1, \ldots, i\}$ and $[i:j] = \{i, \ldots, j\}$, where $[i:j] = \emptyset$ when $i > j$. We denote an n dimensional vector by $x^n = (x_1, \ldots, x_n)$. When the dimension of a vector x^n is clear from the context, we denote x^n by $\mathbf{x}$. All-zero and all-one vectors are denoted by $\mathbf{0}$ and $\mathbf{1}$, respectively. For a collection of length-n vectors $x_1^n, \ldots, x_K^n$ and any subset $\mathcal{C} \subseteq [K]$, we denote the corresponding sub-collection of vectors by $x_{\mathcal{C}}^n = (x_c^n : c \in \mathcal{C})$. For collection of vectors $x_{\mathcal{C}}^n$ and index $i \in [n]$, $x_{\mathcal{C},i}$ denotes the collection of scalars obtained by taking i-th coordinate from each vector in $x_{\mathcal{C}}^n$. For any vectors $x_{\mathcal{C}}$ and $y_{\mathcal{C}}$, we write $x_{\mathcal{C}} \leq y_{\mathcal{C}}$ if $x_c \leq y_c$ for all $c \in \mathcal{C}$, $x_{\mathcal{C}} \overset{\pi}{=} y_{\mathcal{C}}$ if there exists a permutation π of $y_{\mathcal{C}}$ such that $x_{\mathcal{C}} = \pi(y_{\mathcal{C}})$, and $x_{\mathcal{C}} \overset{\pi}{\neq} y_{\mathcal{C}}$ if $x_{\mathcal{C}} \neq \pi(y_{\mathcal{C}})$ for all permutations π of $y_{\mathcal{C}}$. For any set $\mathcal{A}$ and integer $k \leq |\mathcal{A}|$, $\binom{\mathcal{A}}{k} = \{\mathcal{B} : \mathcal{B} \subseteq \mathcal{A}, |\mathcal{B}| = k\}$. For a random variable X, we write $X \sim P_X$ to specify that X is distributed according to distribution P_X. We use $Q(\cdot)$ to denote the Gaussian complementary cumulative distribution function, giving $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left\{\frac{-u^2}{2}\right\} du$. We employ the standard $o(\cdot)$ and $O(\cdot)$ notations, giving $f(n) = o(g(n))$ if $\lim_{n\to\infty} \left|\frac{f(n)}{g(n)}\right| = 0$ and $f(n) = O(g(n))$ if $\limsup_{n\to\infty} \left|\frac{f(n)}{g(n)}\right| < \infty$.

A *stationary, memoryless, symmetric, random access channel* (henceforth called simply a RAC) is a memoryless channel with one receiver and an unknown number of transmitters. It is described by a family of stationary, memoryless MACs

$$\left\{\left(\mathcal{X}^k, P_{Y_k|X_{[k]}}(y_k|x_{[k]}), \mathcal{Y}_k\right)\right\}_{k=0}^{K}, \tag{1}$$

each indexed by a number of transmitters, k; the maximal number of transmitters is $K \leq \infty$. When $k = 0$, no transmitters are active; we discuss this case separately below. For $k \geq 1$, the k-transmitter MAC has input alphabet $\mathcal{X}^k$, output

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2020.3047630, IEEE Transactions on Information Theory

alphabet $\mathcal{Y}_k$, and conditional distribution $P_{Y_k|X_{[k]}}$. When k transmitters are active, the RAC output is $Y = Y_k$. The input and output alphabets $\mathcal{X}$ and $\mathcal{Y}_k$ can be abstract.

A. Assumptions on the Channel

We assume that the impact of a channel input on the channel output is independent of the transmitter from which it comes; therefore, each channel in (1) is assumed to be *permutation-invariant* [6], giving

$$P_{Y_k|X_{[k]}}(y_k|x_{[k]}) = P_{Y_k|X_{[k]}}(y_k|\hat{x}_{[k]}) \tag{2}$$

for all $\hat{x}_{[k]} \overset{\pi}{=} x_{[k]}$ and $y_k \in \mathcal{Y}_k$, $k \in [K]$. We further assume that for any $s < k$, an s-transmitter MAC is physically identical to a k-transmitter MAC operated with s active and $k - s$ silent transmitters. At each time step of the communication period, each silent transmitter transmits a silence symbol, here denoted by $0 \in \mathcal{X}$. This *reducibility* constraint gives

$$P_{Y_s|X_{[s]}}(y|x_{[s]}) = P_{Y_k|X_{[k]}}(y|x_{[s]}, 0^{k-s}) \tag{3}$$

for all $s < k$, $x_{[s]} \in \mathcal{X}_{[s]}$, and $y \in \mathcal{Y}_s$. An immediate consequence of reducibility is that $\mathcal{Y}_s \subseteq \mathcal{Y}_k$ for any $s < k$. Another consequence is that when there are no active transmitters, the MAC $\left(\mathcal{X}^0, P_{Y_0|X_{[0]}}(y|x_{[0]}), \mathcal{Y}_0\right)$ satisfies $\mathcal{X}^0 = \{0\}$ and $P_{Y_0|X_{[0]}}(y|x_{[0]}) = P_{Y_k|X_{[k]}}(y|0^k)$ for all k.

B. RAC Communication Strategy

We here propose a new RAC communication strategy. In the proposed strategy, communication occurs in epochs, with each epoch beginning in the time step following the previous epoch's end. Each epoch ends when the receiver's scheduled broadcast to all transmitters indicates a decoding event, signaling that the prior transmission can stop and a new transmission can begin. At this point, each transmitter decides whether to be active or silent in the new epoch; the decision is binding for the length of the epoch, meaning that a transmitter must either actively transmit for all time steps in the epoch or remain silent for the same period. Thus, while the total number of transmitters, K, is potentially unlimited and can change arbitrarily from one epoch to the next, the number of active transmitters, k, remains constant throughout each epoch.

Each active transmitter uses the epoch to describe a message W from the alphabet $[M]$. When the active transmitters are $[k]$, the messages are $W_{[k]} \in [M]^k$, where $W_{[k]}$ are independent and uniformly distributed. The receiver makes a decision at each time $n_0, n_1, \ldots, n_K$, choosing to end the epoch (without decoding) at time n_0 if it believes at time n_0 that no transmitters are active, and choosing to decode at time n_t if it believes at time n_t that the number of active transmitters is t. The transmitters are informed of the decoder's decision through a single-bit feedback Z_s at each time n_s with $s \in \{0, 1, \ldots, t\}$; here $Z_s = 0$ for all $s < t$ and $Z_t = 1$, with "1" signaling the end of one epoch and the beginning of the next.

It is important to stress that in this domain each transmitter knows nothing about the set of active transmitters $\mathcal{A} \subset \mathbb{N}$ beyond its own membership and what it learns from the receiver's feedback, and the receiver knows nothing about $\mathcal{A}$ beyond what it learns from the channel output Y; we call this *agnostic* random access. In addition, since designing a different encoder for each transmitter is expensive from the perspective of both code design and code operation, as in [6], we assume through most of this paper that every transmitter employs the same encoder; we call this *identical encoding*. Under the assumptions of permutation-invariance and identical encoding, what the transmitters and receiver can learn about $\mathcal{A}$ is quite limited. Together, these properties imply that the decoder can at best distinguish *which messages were transmitted* rather than *by whom they were sent*. In practice, transmitter identity could be included in the header of each $\log M$-bit message or at some other layer of the stack; transmitter identity is not, however, handled by the RAC code. Instead, since the channel output statistics depend on the dimension of the channel input but not the identity of the active transmitters, the receiver's task is to decode the messages transmitted but not the identities of their senders. We therefore assume without loss of generality that $|\mathcal{A}| = k$ implies $\mathcal{A} = [k]$. Thus the family of k-transmitter MACs in (2) fully describes the behavior of a RAC.[1]

The single-bit feedback strategy described above uses *rateless coding* to deal with the agnostic nature of random access. Specifically, the code design fixes the blocklengths $(n_0, n_1, \ldots, n_K)$, where n_t is the decoding blocklength when the decoder believes that the number of active transmitters k is equal to t. As we show in Section IV below, with an appropriately designed decoding rule, correct decoding is performed at time n_k with high probability. Naturally, the greater the number of active transmitters, the longer it takes to decode (i.e., $n_0 < n_1 < \cdots < n_K$).[2] Since the argument employed to bound the performance of our proposed code relies on a random design algorithm, we index the family of possible codes by the elements of some set $\mathcal{U}$ and include $u \in \mathcal{U}$ as an argument for both the RAC encoder and the RAC decoder. We then represent random encoding as the application of a code indexed by some random variable $U \in \mathcal{U}$ chosen independently for each new epoch. Deterministic codes are represented under this code definition by setting the distribution on U as $\mathbb{P}[U = u_0] = 1$ for some $u_0 \in \mathcal{U}$. The following definition formalizes such rateless codes for agnostic random access.

C. Code Definition

Definition 1. *An $(M, \{(n_k, \epsilon_k)\}_{k=0}^K)$ RAC code comprises a (rateless) encoding function*

$$\mathsf{f}\colon \mathcal{U} \times [M] \to \mathcal{X}^{n_K} \tag{4}$$

[1] Section V-D treats a variant of our RAC communication strategy that enables decoding of transmitter identity. Mathematically, the variants are quite similar.

[2] For small blocklengths, the ordering may depend on the desired error probabilities $(\epsilon_0, \epsilon_1, \ldots, \epsilon_K)$. The proposed strategy works for any ordering of $n_0, n_1, \ldots, n_K$, though the error probability analysis requires mild modification to accommodate a given ordering.

and a collection of decoding functions

$$\mathsf{g}_k\colon \mathcal{U}\times\mathcal{Y}_k^{n_k}\to[M]^k\cup\{\mathsf{e}\},\quad k=0,1,\dots,K, \tag{5}$$

where e *denotes the erasure symbol, which is the decoder's output when it is not ready to decode. At the start of each epoch, a common randomness random variable* $U\in\mathcal{U}$*, with* $U\sim P_U$*, is generated independently of the transmitter activity and revealed to the transmitters and the receiver, thereby initializing the encoders and the decoder. If* k *transmitters are active, then with probability at least* $1-\epsilon_k$*, the* k *messages are correctly decoded at time* n_k*. That is,*[3]

$$\frac{1}{M^k}\sum_{w_{[k]}\in[M]^k}\mathbb{P}\left[\left\{\mathsf{g}_k(U,Y_k^{n_k})\overset{\pi}{\neq}w_{[k]}\right\}\bigcup\left\{\bigcup_{t=0}^{k-1}\{\mathsf{g}_t(U,Y_k^{n_t})\neq\mathsf{e}\}\right\}\middle|\, W_{[k]}=w_{[k]}\right]\le\epsilon_k, \tag{6}$$

where $W_{[k]}$ *are the independent and equiprobable messages of transmitters* $[k]$*, and the given probability is calculated using the conditional distribution* $P_{Y_k^{n_k}|X_{[k]}^{n_k}}=P^{n_k}_{Y_k|X_{[k]}}$*; here* $X_i^{n_k}=\mathsf{f}(U,W_i)^{n_k}$*,* $i=1,\dots,k$*. At time* n_s*, the decoder outputs the erasure symbol "*e*" if it decides that the number of active transmitters is not* s*. If* $k=0$ *transmitters are active, the unique message "0", denoted* $[M]^0\triangleq\{0\}$ *to simplify the notation, is decoded at time* n_0 *with probability at least* $1-\epsilon_0$*. That is,*

$$\mathbb{P}\left[\mathsf{g}_0(U,Y_0^{n_0})\neq 0|W_{[0]}=0\right]\le\epsilon_0. \tag{7}$$

In practice, we can implement a RAC code with random code choice U using common randomness. Common randomness available to the transmitters and the receiver allows all nodes to choose the same random variable U to specify a new codebook in each epoch. Operationally, this common randomness can be implemented by allowing the receiver to choose random instance U at the start of each epoch and to broadcast that value to the transmitters just after the feedback bit that ends the previous epoch. Alternatively, all communicators can use synchronized pseudo-random number generators. Broadcasting the value of U increases the epoch-ending feedback from 1 bit to $\lceil\log|\mathcal{U}|\rceil+1$ bits; Theorem 8 shows that $|\mathcal{U}|\le K+1$ suffices to achieve the optimal performance. In Section IV, we employ a general random coding argument to show that a given error vector $(\epsilon_0,\dots,\epsilon_K)$ is achievable when averaged over the ensemble of codes. Unfortunately, this traditional approach does not show the existence of a deterministic RAC code (i.e., a code with $|\mathcal{U}|=1$) that achieves the given error vector $(\epsilon_0,\dots,\epsilon_K)$. The challenge here is that our proof showing that the random code's expected error probability meets each of the $K+1$ error constraints does not suffice to show that any of the codes in the ensemble meets all of our error constraints simultaneously. A similar issue arises in [15], [35]. For example, in [15], a variable-length feedback code is designed with the aim of achieving average error probability no greater than ϵ and expected decoding time no greater than ℓ. To design a single code satisfying both constraints, [15] relies on common randomness. Similarly, [35] describes a variable-length feedback code designed to satisfy an error exponent criterion for every channel in a continuum of binary symmetric or Z channels. Their proof that a single, deterministic code can simultaneously satisfy this continuum of constraints exploits the ordering among the channels in the given family. While channel symmetry can sometimes be leveraged to show the existence of a deterministic code [15, eq. (29)], the symmetries in a RAC are quite different from those in point-to-point channels. We leave the question of whether a single-code solution exists for the RAC to future work.

The code model introduced in Definition 1 employs identical encoding in addition to common randomness. Under identical encoding, each transmitter uses the same encoder, f, to form a codeword of length n_K. That codeword is fed into the channel symbol by symbol. According to Definition 1, if k transmitters are active, then with probability at least $1-\epsilon_k$, the decoder recovers the transmitted messages correctly after observing the first n_k channel outputs. As noted previously, the decoder g_k does not attempt to recover transmitter identity; successful decoding means that the list of messages in the decoder output coincides with the list of messages sent. The error event defined in Definition 1 differs from the one in [6]. Our definition (6) requires that all transmitted messages are decoded correctly. In contrast, [6] bounds a per-user probability of error (PUPE), which measures the fraction of transmitted messages that are missing from the list of decoded messages. In Section V-E, we discuss the error probability for our code under the PUPE criterion.

D. Information Density Definitions

The following definitions are useful for the discussion that follows. When k transmitters are active, the input distribution is $P_{X_{[k]}}$, and the marginal output distribution is P_{Y_k}. The information density and conditional information density are defined[4] as

$$\imath_k(x_\mathcal{A};y_k)\triangleq\log\frac{P_{Y_k|X_\mathcal{A}}(y_k|x_\mathcal{A})}{P_{Y_k}(y_k)} \tag{8}$$

$$\imath_k(x_\mathcal{A};y_k|x_\mathcal{B})\triangleq\log\frac{P_{Y_k|X_\mathcal{A},X_\mathcal{B}}(y_k|x_\mathcal{A},x_\mathcal{B})}{P_{Y_k|X_\mathcal{B}}(y_k|x_\mathcal{B})} \tag{9}$$

for any $\mathcal{A},\mathcal{B}\subseteq[k]$, $x_\mathcal{A}\in\mathcal{X}_\mathcal{A}, x_\mathcal{B}\in\mathcal{X}_\mathcal{B}$, and $y_k\in\mathcal{Y}_k$; here $\imath_k(x_\mathcal{A};y_k|x_\mathcal{B})\triangleq\imath_k(x_\mathcal{A};y_k)$ when $\mathcal{B}=\emptyset$ and $\imath_k(x_\mathcal{A};y_k|x_\mathcal{B})\triangleq 0$ when $y_k\notin\mathcal{Y}_k$ or $\mathcal{A}=\emptyset$. The corresponding mutual informations are

$$I_k(X_\mathcal{A};Y_k)\triangleq\mathbb{E}[\imath_k(X_\mathcal{A};Y_k)] \tag{10}$$

$$I_k(X_\mathcal{A};Y_k|X_\mathcal{B})\triangleq\mathbb{E}[\imath_k(X_\mathcal{A};Y_k|X_\mathcal{B})]. \tag{11}$$

Throughout the paper, we also denote for brevity

$$I_k\triangleq I_k(X_{[k]};Y_k) \tag{12}$$

$$V_k\triangleq\mathrm{Var}\left[\imath_k(X_{[k]};Y_k)\right]. \tag{13}$$

[3] Recall that $\overset{\pi}{=}$ and $\overset{\pi}{\neq}$ denote equality and inequality up to a permutation.

[4] We here employ notation for discrete alphabets. In the general case, it can be replaced by the logarithm of the Radon-Nikodym derivative, giving $\imath_k(x_\mathcal{A};y_k)=\log\frac{dP_{Y_k|X_\mathcal{A}=x_\mathcal{A}}}{dP_{Y_k}}(y_k)$.

The multi-letter information density and conditional information densities are defined as

$$\imath_k(x^n_{\mathcal{A}}; y^n_k) \triangleq \log \frac{P_{Y^n_k|X^n_{\mathcal{A}}}(y^n_k|x^n_{\mathcal{A}})}{P_{Y^n_k}(y^n_k)} \tag{14}$$

$$\imath_k(x^n_{\mathcal{A}}; y^n_k|x^n_{\mathcal{B}}) \triangleq \log \frac{P_{Y^n_k|X^n_{\mathcal{A}},X^n_{\mathcal{B}}}(y^n_k|x^n_{\mathcal{A}}, x^n_{\mathcal{B}})}{P_{Y^n_k|X^n_{\mathcal{B}}}(y^n_k|x_{\mathcal{B}})}. \tag{15}$$

E. Assumptions on the Input Distribution

To ensure the existence of codes satisfying the error constraints in Definition 1, we assume that there exists a P_X such that when $X_1, X_2, \ldots, X_K$ are distributed i.i.d. P_X, then the conditions in (16)–(21) below are satisfied.

The *friendliness* assumption states that for all $s \le k \le K$,

$$I_k(X_{[s]}; Y_k|X_{[s+1:k]} = 0^{k-s}) \ge I_k(X_{[s]}; Y_k|X_{[s+1:k]}). \tag{16}$$

Friendliness implies that by remaining silent, inactive transmitters enable communication by the active transmitters at rates at least as large as those achievable if the inactive transmitters had actively participated and their codewords were known to the receiver.

The *interference* assumption states that for any s and t, $X_{[s]}$ and $X_{[s+1:t]}$ are conditionally dependent given Y_k, giving

$$P_{X_{[t]}|Y_k} \neq P_{X_{[s]}|Y_k} P_{X_{[s+1:t]}|Y_k} \quad \forall 1 \le s < t \le k, \forall k. \tag{17}$$

Assumption (17) eliminates trivial RACs in which transmitters do not interfere.

In order for the decoder to be able to distinguish the time-n_0 output $Y_0^{n_0}$ that results when no transmitters are active from the time-n_0 output $Y_k^{n_0}$ that results when $k \ge 1$ transmitters are active, we assume that there exists a $\delta_0 > 0$ such that the output distributions satisfy

$$\sup_{y \in \mathcal{Y}_K} |F_k(y) - F_0(y)| \ge \delta_0 \text{ for all } k \in [K], \tag{18}$$

where $F_k(y)$ denotes the cumulative distribution function (CDF) of P_{Y_k} for $k \in \{0, \ldots, K\}$.[5] The measure of discrepancy between distributions on the left-hand side of (18) is known as the Kolmogorov-Smirnov distance. The assumption in (18) is only needed to detect the scenario when no transmitters are active; the remainder of the code functions proceed unhampered when (18) fails. When K is finite, (18) is equivalent to $P_{Y_0} \neq P_{Y_k}$ for all $k \in [K]$.

Finally, the *moment* assumptions

$$\mathrm{Var}\left[\imath_k(X_{[k]}; Y_k)\right] > 0 \tag{19}$$

$$\mathbb{E}[|\imath_k(X_{[k]}; Y_k) - I_k(X_{[k]}; Y_k)|^3] < \infty \tag{20}$$

enable the second-order analysis presented in Theorem 1, below. In the case when $\imath_t(X_{[s]}; Y_k) > -\infty$ almost surely, we also require

$$\mathrm{Var}\left[\imath_t(X_{[s]}; Y_k)\right] < \infty \quad \forall s \le t \le k. \tag{21}$$

Moment assumptions like (19)–(21) are common in the finite-blocklength literature, e.g., [12], [36].

In the discussion that follows, we say that a channel satisfies our channel assumptions ((2), (3), (16)–(21)) if there exists an input distribution P_X under which those conditions are satisfied. All discrete memoryless channels (DMCs) satisfy finite second- and third-moment assumptions (20)–(21) [36, Lemma 46], as do Gaussian noise channels. Common channel models from the literature typically satisfy a non-zero second-moment assumption (19) as well. Example channels that meet our channel assumptions ((2), (3), and (16)–(21)) include the Gaussian RAC,

$$Y_k = \sum_{i=1}^{k} X_i + Z, \tag{22}$$

where each $X_i \in \mathbb{R}$ operates under power constraint P and $Z \sim \mathcal{N}(0, N)$ for some $N > 0$, and the adder-erasure RAC [8],

$$Y_k = \begin{cases} \sum_{i=1}^{k} X_i, & \text{w.p. } 1 - \delta \\ \mathsf{e} & \text{w.p. } \delta, \end{cases} \tag{23}$$

where $X_i \in \{0, 1\}$ and $Y_k \in \{0, \ldots, k\} \cup \{\mathsf{e}\}$. In [8], the adder-erasure RAC (23) is used to model a scenario where a digital encoder and decoder communicate over an analog channel using a modulator and demodulator. The modulator converts the bits into analog signals; the channel output equals the sum of the transmitted signals plus random noise; the demodulator quantizes that output, declaring an erasure, e, if reliable quantization is not possible due to high noise. Thus, one can view the adder-erasure RAC as a discretization of the Gaussian RAC.

For the Gaussian RAC, $\imath_t(X_{[s]}; Y_k) > -\infty$ almost surely, and (21) is satisfied. For the adder-erasure RAC, $\imath_t(X_{[s]}; Y_k) = -\infty$ for some channel realizations and user activity patterns, and (21) is not required.

We conclude this section with a series of lemmas that describe the natural orderings possessed by RACs that satisfy our permutation-invariance, reducibility, friendliness, and interference constraints ((2), (3), (16), and (17)). These properties are key to the feasibility of the approach proposed in our achievability argument in Section III. Proofs are relegated to Appendix A.

The first lemma shows that the quality of the channel for each active transmitter deteriorates as the number of active transmitters grows (even though the sum capacity may increase).

Lemma 1. *Let $X_1, X_2, \ldots, X_k \sim$ i.i.d. P_X. Under permutation-invariance (2), reducibility (3), friendliness (16), and interference (17),*

$$\frac{I_k}{k} < \frac{I_s}{s} \quad \textit{for } k > s \ge 1. \tag{24}$$

The second lemma shows that a similar relationship holds even when the number of transmitters is fixed.

Lemma 2. *Let $X_1, X_2, \ldots, X_k \sim$ i.i.d. P_X. Under permutation-invariance (2), reducibility (3) and interference (17),*

$$\frac{1}{k} I_k(X_{[k]}; Y_k) < \frac{1}{s} I_k(X_{[s]}; Y_k|X_{[s+1:k]}) \quad \textit{for } k > s \ge 1. \tag{25}$$

[5] Although the CDF is defined for real-valued random variables, i.e., $\mathcal{Y}_k \subseteq \mathcal{Y}_K \subseteq \mathbb{R}$ is required, it can be generalized to abstract alphabets by introducing a partial order $\le$ on the set $\mathcal{Y}_K$. Then $F_k(y) \triangleq \mathbb{P}[Y_k \le y]$.

Lemma 2 ensures that the equal-rate point of the k-MAC lies on the sum-rate boundary and away from all the corner points of the rate region achieved with P_X. In their work on the group testing problem [31, Th. 3], Malyutov and Mateev prove a non-strict version of (25) for permutation-invariant channels (2). They use this non-strict version of (25) to conclude that their achievability and converse results in [31, Th. 1 and 2] coincide for permutation-invariant channels. Adding the reducibility (3) and interference (17) assumptions to the permutation-invariance assumption (2) enables us to prove the strict inequality in Lemma 2, which in turn enables the use of a single threshold rule at the decoder, as discussed in Section IV.

Lemma 3 compares the expected values of the information densities for different channels.

Lemma 3. *Let $X_1, X_2, \ldots, X_k \sim$ i.i.d. P_X. If a RAC is permutation-invariant (2), reducible (3), friendly (16), and exhibits interference (17), then for any $1 \leq s \leq t < k$,*

$$\mathbb{E}[\imath_t(X_{[s]}; Y_k)] \leq I_k(X_{[s]}; Y_k) < I_t(X_{[s]}; Y_t). \tag{26}$$

The orderings in Lemma 1–3 are used in bounding the performance of our agnostic random access code.

III. Main Result

A. An Asymptotic Achievability Result

Our main result is the following bound on achievable rates for the RAC.

Theorem 1. *(Achievability) For any RAC*

$$\left\{\left(\mathcal{X}^k, P_{Y_k|X_{[k]}}(y_k|x_{[k]}), \mathcal{Y}_k\right)\right\}_{k=0}^{K}$$

satisfying (2) and (3), any $K < \infty$, and any fixed P_X satisfying (16)–(21), there exists an $(M, \{(n_k, \epsilon_k)\}_{k=0}^K)$ code provided that

$$k \log M \leq n_k I_k - \sqrt{n_k V_k} Q^{-1}(\epsilon_k) - \frac{1}{2}\log n_k + O(1) \tag{27}$$

for all $k \in [K]$, and

$$n_0 \geq c_0 \log n_1 + o(\log n_1), \tag{28}$$

where c_0 is a known positive constant. The $O(1)$ term in (27) is constant with respect to n_1; it depends on the number of active transmitters, k, but not on the total number of transmitters, K.

The code in Theorem 1 assigns equal rates $R_{[k]} = (R, \ldots, R)$, $R = \frac{\log M}{n_k}$, to all active transmitters. The sum-rate kR converges as $O\left(\frac{1}{\sqrt{n_k}}\right)$ to $I_k(X_{[k]}; Y_k)$ for some input distribution $P_{X_{[k]}}(x_{[k]}) = \prod_{i=1}^k P_X(x_i)$ for all k. Note that P_X is independent of the number of active transmitters, k. If the RAC is discrete and memoryless and a single P_X maximizes $I_k(X_{[k]}; Y_k)$ for every k, then the achievable rate in (27) not only converges to the symmetrical rate point on the capacity region of the MAC in operation but also achieves the best-known second-order term [10]–[13][6] (see Section III-B for details.)

To better understand Theorem 1, consider a channel satisfying (16)–(21) for which the same distribution P_X maximizes I_k for all k. For example, for the adder-erasure RAC in (23), setting P_X to be Bernoulli(1/2) maximizes I_k for all k. By Lemma 1, for M large enough and any $\epsilon_1, \epsilon_2, \ldots, \epsilon_K$, one can pick[7] $n_1 < n_2 < \ldots < n_K$ so that the gap between the right and left sides of (27) is $O(1)$ for all k. Therefore, Theorem 1 certifies that for some channels, rateless codes with encoders that are, until feedback, agnostic to the transmitter activity pattern perform as well in both first- and second-order terms as the best-known scheme [10]–[13] designed with complete knowledge of transmitter activity. Moreover for any fixed $0 < \epsilon_0 < 1$, the probability that at time $n_0 \geq c_0 \log n_1 + o(\log n_1)$ the decoder correctly detects the scenario where no transmitters are active is no smaller than $1 - \epsilon_0$. Thus, a new epoch can begin very quickly when no transmitters are active in the current epoch.

The constant c_0 in (28) depends on the output distributions P_{Y_k}, $k = 0, \ldots, K$, and on the hypothesis test chosen in Section VI but not on the target probability of error ϵ_0. In contrast, the $o(\log n_1)$ term in (28) depends on ϵ_0. See Section VI (eq. (151)) for an example where we bound the dependence of the $o(\log n_1)$ term on ϵ_0 under the log-likelihood ratio test.

Our achievability result in Theorem 1 assumes that the total number of transmitters, K, is constant. The asymptotic regime in which K grows with the decoding times, $n_1, n_2, \ldots, n_K$, seeks to characterize scenarios with massive numbers of communicators [6], [28], [33]. Understanding the fundamental limits of random access communications in that regime presents an interesting challenge for future work.

B. Comparison With the Existing Achievability Results

1) Discrete Memoryless RACs: Our achievable region (Theorem 1) is consistent with the achievability results for the 2-transmitter MACs given in [10]–[13]. The proofs in [10]–[12] use i.i.d. random code design, an approach that we follow in Theorem 1. In [13], Scarlett et al. use constant-composition codes. In [10]–[12], the achievable rate region of a discrete memoryless MAC is expressed as a three-dimensional vector inequality that relies on a 3×3 dispersion matrix V_2 defined in [12, eq. (48)]; the entry of V_2 at location $(3,3)$ is V_2 (13) for some input distribution (P_{X_1}, P_{X_2}). For rate pairs approaching interior (i.e., non-corner) points on the sum-rate boundary for $(P_{X_1^*}, P_{X_2^*})$, i.e., rate pairs satisfying

$$(R_1, R_2) \in \{(r_1 + o(1), r_2 + o(1)):$$

[6] Note that we are comparing the RAC achievable rate with rate-0 feedback to the MAC capacity without feedback. Wagner et al. [37] show that if a discrete, memoryless, point-to-point channel has at least two capacity-achieving input distributions and their dispersions V_1 (13) are distinct, then using one-bit feedback improves the achievable second-order term. Although rate-0 feedback does not change the capacity region of a discrete memoryless MAC [29], in light of [37] it is plausible that even one-bit feedback can improve the achievable second-order term for some MACs.

[7] As noted previously, focusing on scenarios with decoding times ordered as $n_1 < n_2 < \cdots < n_K$ simplifies the exposition but is not critical to the approach.

$$\begin{aligned} r_1 &< I_2(X_1^*; Y_2^* | X_2^*) \\ r_2 &< I_2(X_2^*; Y_2^* | X_1^*) \\ r_1 + r_2 &= I_2(X_1^*, X_2^*; Y_2^*)\}, \end{aligned} \tag{29}$$

the achievable region in [10]–[12] reduces to the scalar inequality

$$R_1 + R_2 \le I_2^* - \sqrt{\frac{V_2^*}{n}} Q^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right), \tag{30}$$

where

$$I_2^* \triangleq I_2(X_1^*, X_2^*; Y_2^*) \tag{31}$$

is the sum-rate capacity and V_2^* is the dispersion V_2 (13) evaluated using $(P_{X_1^*}, P_{X_2^*})$. The bound in (30) implies that the only component of V_2 employed in the second-order characterization of the region (29) is V_2^*. The result in (30) is proved in [38, Prop. 4 case ii)].

In [13, Th. 1], Scarlett et al. use constant-composition codes to show that the dispersion matrix V_2 in the second-order achievable region can be improved to $\tilde{\mathsf{V}}_2$, defined in [13, eq. (13)]. Further, they show that $\tilde{\mathsf{V}}_2 \preceq \mathsf{V}_2$, where $\preceq$ designates positive semidefinite order. Therefore, the second-order rate region that is obtained using constant-composition codes includes that achieved with i.i.d. random coding when the target error probability satisfies $\epsilon < \frac{1}{2}$. Scarlett et al. [13] present two examples for which $\tilde{\mathsf{V}}_2 \prec \mathsf{V}_2$, demonstrating that the inclusion can be strict. The $(3, 3)$ component of $\tilde{\mathsf{V}}_2$ is

$$\begin{aligned} \tilde{V}_2^* = V_2^* &- \mathrm{Var}\left[\mathbb{E}\left[\imath_2(X_1^*, X_2^*; Y_2^*)|X_1^*\right]\right] \\ &- \mathrm{Var}\left[\mathbb{E}\left[\imath_2(X_1^*, X_2^*; Y_2^*)|X_2^*\right]\right], \end{aligned} \tag{32}$$

where $P_{X_1^*} P_{X_2^*} P_{Y_2^*|X_1^*, X_2^*} = P_{X_1^*} P_{X_2^*} P_{Y_2|X_1,X_2}$. The right side of (30) is achievable with V_2^* replaced by $\tilde{V}_2^*$. In Lemma 4, below, we derive a saddle point condition for general MACs without cost constraints. Lemma 4 implies that

$$\tilde{V}_2^* = V_2^*. \tag{33}$$

This means that while constant-composition code design can yield achievability results with second-order terms superior to those derived through i.i.d. code design, on the sum-rate boundary that superior performance is observed only at corner points. For any rate point approaching an interior point on the sum-rate boundary, the i.i.d. random code design employed in this paper achieves first- and second-order performance identical to that achieved by constant-composition code design.

Lemma 4. *Let $P_{Y_2|X_1,X_2}$ be a 2-transmitter MAC with finite sum-rate capacity. Assume that the σ-algebra on the abstract input alphabets $\mathcal{X}_i$ includes all singletons on $\mathcal{X}_i$, $i = 1, 2$. Let $(X_1^*, X_2^*, Y_2^*) \sim P_{X_1^*} P_{X_2^*} P_{Y_2|X_1,X_2}$, where $(P_{X_1^*}, P_{X_2^*})$ is a sum-rate capacity achieving input distribution, i.e.,*

$$I_2^* \triangleq I_2(X_1^*, X_2^*; Y_2^*) = \sup_{P_{X_1} P_{X_2}} I_2(X_1, X_2; Y_2) < \infty. \tag{34}$$

Then, for $i = 1, 2$,

$$\mathbb{E}\left[\imath_2(X_1^*, X_2^*; Y_2^*)|X_i^*\right] = I_2^*, \tag{35}$$

where (35) holds $P_{X_i^*}$-almost surely.

Proof: See Appendix B. ■

A version of Lemma 4 for discrete memoryless MACs appears in [39, Prop. 1]. The result is proved by verifying that (35) satisfies the Karush-Kuhn-Tucker (KKT) conditions for the maximization problem in (34) (Although the maximization problem in (34) is not convex, it satisfies a regularity condition ensuring the necessity of the KKT conditions for optimality [39].) We extend [39, Prop. 1] to general MACs by demonstrating a saddle point condition for MACs. The saddle point condition is more general in the sense that it applies to abstract alphabets.

From (35), we deduce that

$$\mathrm{Var}\left[\mathbb{E}\left[\imath_2(X_1^*, X_2^*; Y_2^*)|X_i^*\right]\right] = 0, \quad i = 1, 2. \tag{36}$$

Substituting (36) into (32), we obtain (33).

The result in (34)–(35) extends the following well-known properties of point-to-point DMCs to MACs. In [40, Th. 4.5.1], the KKT conditions in (34)–(35) for point-to-point DMCs are

$$I_1^* \triangleq \max_{P_{X_1}} I_1(X_1; Y_1) \tag{37}$$

$$\mathbb{E}\left[\imath_1(X_1^*; Y_1^*)|X_1^*\right] = I_1^* \quad \text{if } P_{X_1^*}(x_1) > 0 \tag{38}$$

$$\mathbb{E}\left[\imath_1(X_1^*; Y_1^*)|X_1^* = x_1\right] \le I_1^* \quad \text{if } P_{X_1^*}(x_1) = 0; \tag{39}$$

these conditions are necessary and sufficient for optimality. As noted in [36, Lemma 62], (38)–(39) indicate that for a capacity-achieving input distribution $P_{X_1^*}$,

$$\mathrm{Var}\left[\mathbb{E}\left[\imath_1(X_1^*; Y_1^*)|X_1^*\right]\right] = 0. \tag{40}$$

From (40) and the law of total variance, it follows that the unconditional and conditional variances of $\imath_1(X_1^*; Y_1^*)$ given X_1^* are equal, i.e.,

$$V_1 = \mathbb{E}\left[\mathrm{Var}\left[\imath_1(X_1^*; Y_1^*)|X_1^*\right]\right]. \tag{41}$$

For point-to-point DMCs, Moulin [41] shows that the second-order term $\tilde{V}_1$ achievable using constant-composition coding equals the right-hand side of (41), meaning that i.i.d. random code design and constant-composition random code design achieve the same fundamental limits for point-to-point DMCs.

2) The Gaussian RAC: While the RAC code definition (Definition 1) does not impose cost constraints on the codewords, cost constraints can be added where needed. In the case of the Gaussian RAC defined in (22), the maximal power constraint P on the codewords requires that

$$\left\|\mathsf{f}(u, w)^{n_k}\right\|^2 \le n_k P \tag{42}$$

for all $u \in \mathcal{U}$, $w \in [M]$, and $k \in [K]$, where $\|\cdot\|$ denotes the Euclidean norm. If any encoder attempts to transmit a codeword that does not satisfy (42), we count that event as an error. Hence, the maximal power constraints add the term

$$\mathbb{P}\left[\bigcup_{j=1}^{k} \bigcup_{i=1}^{k} \left\{\left\|X_i^{n_j}\right\|^2 > n_j P\right\}\right] \tag{43}$$

to the error terms in (6).

For the Gaussian k-MAC under maximal power constraints, drawing codewords i.i.d. according to distribution $P_X \sim \mathcal{N}(0, P - \delta_{n_k})$ for any $\delta_{n_k} \to 0$ as $n_k \to \infty$ yields a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2020.3047630, IEEE Transactions on Information Theory

worse second-order performance bound than the one achieved by drawing codewords uniformly at random from the n_k-dimensional power sphere [25], [42]. MolavianJazi and Laneman [25] and Scarlett et al. [13] derive the improved second-order term for the Gaussian MAC by drawing codewords uniformly at random over an n_k-dimensional power sphere and by combining constant-composition code design with a quantization argument, respectively. In [43], for the Gaussian MAC and RAC, we prove the achievability of the same second-order term as [13], [25] with an improved third-order term $\frac{1}{2}\log n_k$. The proof employs codewords designed by concatenating spherically distributed sub-blocks and a maximum likelihood decoding rule combined with a threshold rule based on the output power.

C. An Example RAC

The following example investigates rates achievable for the adder-erasure RAC in (23).

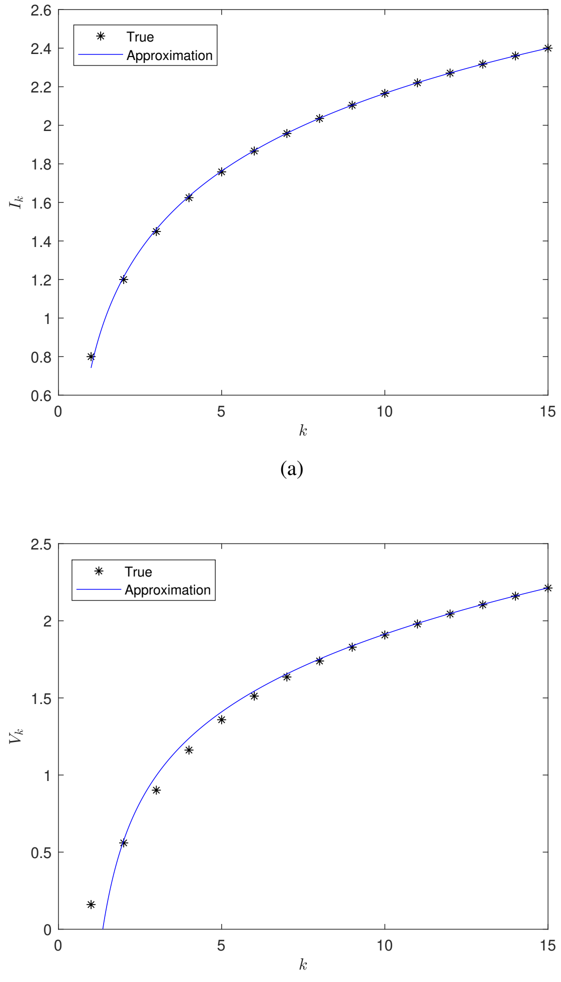
Example 1. For the adder-erasure RAC, the capacity achieving distribution is the equiprobable (Bernoulli(1/2)) distribution for all k. (See the proof of Theorem 7 in Appendix C.) For this channel, one can exactly calculate I_k and V_k for this channel for every k (labelled "True" in Fig. 1). The approximating characterizations

$$I_k = (1-\delta)\left(\frac{1}{2}\log\frac{\pi e k}{2} - \frac{\log e}{12k^2}\right) + O(k^{-3}) \tag{44}$$

$$\begin{aligned} V_k = (1-\delta)\Bigg[&\frac{\delta}{4}\log^2\frac{\pi e k}{2} + \frac{\log^2 e}{2} - \frac{\log^2 e}{2k} \\ &- \left(\frac{\log e}{2} + \frac{\delta\log\frac{\pi e k}{2}}{12}\right)\frac{\log e}{k^2}\Bigg] + O\left(\frac{\log k}{k^3}\right), \end{aligned} \tag{45}$$

which capture the first- and second-order behavior of I_k and V_k for each k, are, nonetheless, useful since they highlight how each depends on k and δ. These values, without the $O(\cdot)$ terms in (44)–(45), are labelled "Approximation" in Fig. 1. The approximations are quite tight even for small k. Both I_k and $\sqrt{V_k}$ are of order $O(\log k)$, indicating that as k grows, the sum-rate capacity grows, albeit slowly, while the per-user rate vanishes as $O\left(\frac{\log k}{k}\right)$. The dispersion V_k also grows, and the speed of approach to the sum-rate capacity is slower. Interestingly, the dispersion behavior is different for the pure adder RAC ($\delta = 0$), in which case $V_k = \frac{1}{2} + O\left(\frac{1}{k}\right)$ is almost constant as a function of k. The derivation of (44) and (45) relies on an approximation for the probability mass function of the $(k, 1/2)$ Binomial distribution using a higher order Stirling's approximation (Appendix C).

Fig. 2 shows the approximate rate per transmitter, $R_k = \frac{\log M}{n_k}$ (neglecting the $O(1)$ term in (27)), achieved by the proposed scheme as a function of the number of active transmitters, k, and the choice of blocklength n_1 for a fixed error probability $\epsilon_k = 10^{-6}$ for all k. Fixing n_1 and ϵ_k fixes the maximum achievable message size, M, according to (27). The remaining n_k for $k \geq 2$ are found by choosing the smallest n_k that satisfies (27) using the given M and ϵ_k. Each curve illustrates how the rate per transmitter (R_k) decreases



(a)

(b)

Fig. 1: (a) Sum-rate capacity I_k (in bits) and (b) dispersion V_k (in bits^2) for the adder-erasure RAC with $\delta = 0.2$.

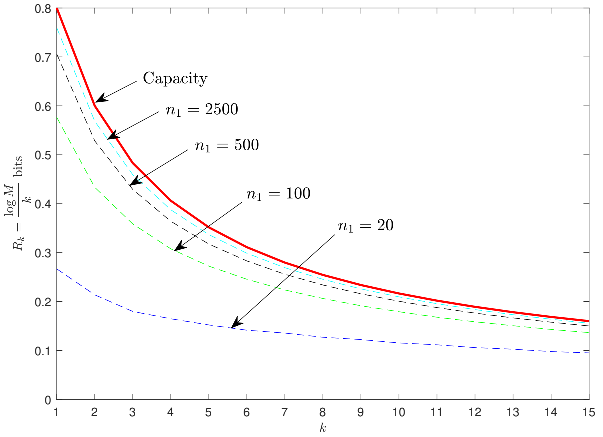


Fig. 2: Capacity and approximate achievable rates (in bits per user) for the adder-erasure RAC with erasure probability $\delta = 0.2$ are given for the target error probability $\epsilon_k = 10^{-6}$ for all k. For each curve, the message size M is fixed so that the rates $\{R_k\}$ are achievable with n_1 set to $20, 100, 500$, and 2500, respectively.

as the number of active users k increases. The curves differ in their choice of blocklength n_1 and the resulting changes in M and $n_0, n_2, \ldots, n_K$. Here n_1 is fixed to $20, 100, 500$ and 2500. For a fixed k, the points on the same vertical line demonstrate how the gap between the per-user capacity and the finite-blocklength achievable rate decreases as blocklength increases.

D. A Non-asymptotic Achievability Result

Theorem 1 follows from Theorem 2, stated next, which bounds the error probability of the RAC code defined in Section IV. When k transmitters are active, the error probability ϵ_k captures both errors in the estimate t of k and errors in the reproduction $\hat{W}_{[t]}$ of $W_{[k]}$ when $t = k$. Theorem 2 is formulated for an arbitrary choice of a statistic $h\colon \mathcal{Y}^{n_0} \mapsto \mathbb{R}$ used to decide whether any transmitters are active. Possible choices for $h(\cdot)$ appear in (126) and (133) in Section VI below.

Theorem 2. *Fix constants γ_0, $\lambda_{s,t}^k \geq 0$, and $\gamma_t > 0$ for all $1 \leq s \leq t \leq k$. For any RAC $\left\{\left(\mathcal{X}^k, P_{Y_k|X_{[k]}}(y_k|x_{[k]}), \mathcal{Y}_k\right)\right\}_{k=0}^K$ satisfying (2) and (3), any $K \leq \infty$[8], and any fixed input distribution P_X, there exists an $(M, \{(n_k, \epsilon_k)\}_{k=0}^K)$ code such that*

$$\epsilon_0 \leq \mathbb{P}\left[h(Y_0^{n_0}) > \gamma_0\right], \tag{46}$$

and for all $k \geq 1$,

$$\begin{aligned}
\epsilon_k \leq\ & \mathbb{P}[\imath_k(X_{[k]}^{n_k}; Y_k^{n_k}) \leq \log \gamma_k] && \text{(47a)}\\
& + \mathbb{P}\left[h(Y_k^{n_0}) \leq \gamma_0\right] && \text{(47b)}\\
& + \frac{k(k-1)}{2M} && \text{(47c)}\\
& + \sum_{t=1}^{k-1} \binom{k}{t} \mathbb{P}[\imath_t(X_{[t]}^{n_t}; Y_k^{n_t}) > \log \gamma_t] && \text{(47d)}\\
& + \sum_{t=1}^{k} \sum_{s=1}^{t-1} \binom{k}{t-s} \mathbb{P}\Big[\imath_t(X_{[s+1:t]}^{n_t}; Y_k^{n_t}) \\
& \qquad > n_t \mathbb{E}[\imath_t(X_{[s+1:t]}; Y_k)] + \lambda_{s,t}^k\Big] && \text{(47e)}\\
& + \sum_{t=1}^{k} \sum_{s=1}^{t} \binom{k}{t-s} \binom{M-k}{s} \mathbb{P}\Big[\imath_t(\bar{X}_{[s]}^{n_t}; Y_k^{n_t} | X_{[s+1:t]}^{n_t}) \\
& \qquad > \log \gamma_t - n_t \mathbb{E}[\imath_t(X_{[s+1:t]}; Y_k)] - \lambda_{s,t}^k\Big], && \text{(47f)}
\end{aligned}$$

where for any n, $(X_{[k]}^n, \bar{X}_{[k]}^n, Y_k^n)$ is a random sequence drawn i.i.d. $\sim P_{X_{[k]}\bar{X}_{[k]}Y_k}(x_{[k]}, \bar{x}_{[k]}, y_k) = \left(\prod_{i=1}^k P_X(x_i) P_X(\bar{x}_i)\right) P_{Y_k|X_{[k]}}(y_k|x_{[k]})$.

The operational regime of interest is when $\epsilon_0, \ldots, \epsilon_k$ are constant; that is, ϵ_k does not vanish as n_k grows. For $k = 0$, the error term in (46) is the probability that the decoder does not correctly determine that the number of active transmitters is 0 at time n_0. For $k > 0$, (47a) is the probability that the true codeword set produces a low information density. This is the dominating term in the regime of interest. All remaining terms are negligible, as shown in the refined asymptotic analysis of the bound in Theorem 2 (see Section IV-C, below.) The remaining terms bound the probability that the decoder incorrectly estimates the number of active transmitters as 0 (47b), the probability that two or more transmitters send the same message (47c),[9] the probability that the decoder estimates the number of active transmitters as t for some $1 \leq t < k$ and decodes those t messages correctly (47d), and the probability that the decoder estimates the number of active transmitters as t for some $1 \leq t \leq k$ and decodes the messages from s of those t transmitters incorrectly and the messages from the remaining $t - s$ of those transmitters correctly (47e)–(47f).

For $k = 1, 2$, the expression in (47) particularizes to

$$\begin{aligned}
\epsilon_1 \leq\ & \mathbb{P}[\imath_1(X_1^{n_1}; Y_1^{n_1}) \leq \log \gamma_1] + \mathbb{P}\left[h(Y_1^{n_0}) \leq \gamma_0\right]\\
& + (M-1)\mathbb{P}[\imath_1(\bar{X}_1^{n_1}; Y_1^{n_1}) > \log \gamma_1 - \lambda_{1,1}^1] \qquad (48)\\
\epsilon_2 \leq\ & \mathbb{P}[\imath_2(X_{[2]}^{n_2}; Y_2^{n_2}) \leq \log \gamma_2] + \mathbb{P}\left[h(Y_2^{n_0}) \leq \gamma_0\right]\\
& + \frac{1}{M} + 2\mathbb{P}[\imath_1(X_1^{n_1}; Y_2^{n_1}) > \log \gamma_1]\\
& + 2\mathbb{P}[\imath_2(X_2^{n_2}; Y_2^{n_2}) \geq n_2 I_2(X_2; Y_2) + \lambda_{1,2}^2]\\
& + (M-1)\mathbb{P}[\imath_1(\bar{X}_1^{n_1}; Y_2^{n_1}) > \log \gamma_1 - \lambda_{1,1}^2]\\
& + 2(M-2)\mathbb{P}[\imath_2(\bar{X}_1^{n_2}; Y_2^{n_2} | X_2^{n_2})\\
& \qquad > \log \gamma_2 - n_2 I_2(X_2; Y_2) - \lambda_{1,2}^2]\\
& + \frac{(M-2)(M-3)}{2} \mathbb{P}[\imath_2(\bar{X}_{[2]}^{n_2}; Y_2^{n_2}) > \log \gamma_2 - \lambda_{2,2}^2].
\end{aligned}$$
$$(49)$$

For the MAC with K transmitters, i.e., the scenario where K transmitters are always active, the only decoding time is n_K. The error terms associated with incorrect decoding times are no longer needed in this case, and the error probability bound in (47) becomes

$$\begin{aligned}
\epsilon_K \leq\ & \mathbb{P}[\imath_K(X_{[K]}^{n_K}; Y_K^{n_K}) \leq \log \gamma_K] + \frac{K(K-1)}{2M} && \text{(50a)}\\
& + \sum_{s=1}^{K-1} \binom{K}{K-s} \mathbb{P}\Big[\imath_K(X_{[s+1:K]}^{n_K}; Y_K^{n_K}) \\
& \qquad > n_K \mathbb{E}[\imath_K(X_{[s+1:K]}; Y_K)] + \lambda_{s,K}^K\Big] && \text{(50b)}\\
& + \sum_{s=1}^{K} \binom{K}{K-s} \binom{M-K}{s} \mathbb{P}\Big[\imath_K(\bar{X}_{[s]}^{n_K}; Y_K^{n_K} | X_{[s+1:K]}^{n_K}) \\
& \qquad > \log \gamma_K - n_K \mathbb{E}[\imath_K(X_{[s+1:K]}; Y_K)] - \lambda_{s,K}^K\Big]. && \text{(50c)}
\end{aligned}$$

A description of the proposed RAC code and the proofs of Theorems 1 and 2 appear in Section IV.

IV. The RAC Code and Its Performance

A. Code Design

We construct the RAC code used in the proofs of Theorems 1 and 2 as follows.

[8] Note that while Theorem 1 requires $K < \infty$, Theorem 2 allows $K = \infty$. For $K = \infty$, (47) holds for every finite k, since the bound on ϵ_k depends only on the RAC with at most k active transmitters. The $K = \infty$ case is the only point in this work where the assumption $n_0 < n_1 < n_2 < \ldots$ is not merely convenient but, in fact, critical.

[9] Given the use of identical encoders, multiple encoders sending the same message can be beneficial or harmful, depending on the channel. To simplify the analysis, we treat this (exponentially rare) event as an error.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2020.3047630, IEEE Transactions on Information Theory

Encoder Design: The common randomness random variable $U = (U(1), \ldots, U(M))$ has distribution

$$P_U \triangleq P_{U(1)} \times \cdots \times P_{U(M)}, \tag{51}$$

where $P_{U(w)} = P_X^{n_K}$, $w = 1, \ldots, M$, and P_X is a fixed distribution on alphabet $\mathcal{X}$. Each realization of U defines a codebook with M i.i.d. vectors $U(1), \ldots, U(M)$ of dimension n_K (the codewords). Note that the cardinality of the alphabet U is $|\mathcal{X}|^{Mn_K}$. In [15, Th. 19], Polyanskiy et al. use Carathéodory's Theorem to show that the common randomness U can be replaced with common randomness U' with cardinality at most $K+2$. We reduce this alphabet size to $K+1$ in Appendix D. As described in Definition 1, an $(M, \{(n_k, \epsilon_k)\}_{k=0}^{K})$ RAC code with identical encoders employs the same encoder $\mathsf{f}(\cdot)$ at every transmitter. The encoder $\mathsf{f}(U, \cdot)$ depends on U as

$$\mathsf{f}(U, w) = U(w) \qquad \text{for } w = 1, \ldots, M. \tag{52}$$

For brevity, we omit U in the encoding and decoding functions and write $\mathsf{f}(U, w) = \mathsf{f}(w)$ for $w = 1, \ldots, M$, and $\mathsf{g}_k(U, y^{n_k}) = \mathsf{g}_k(y^{n_k})$ for $y^{n_k} \in \mathcal{Y}_K^{n_k}, k \in \{0, \ldots, K\}$. Recall that $\mathsf{f}(w)$ is a n_K-dimensional vector. We use $\mathsf{f}(w)^{n_k}$ to denote the first n_k coordinates of vector $\mathsf{f}(w)$. For any collection of messages $w_{[k]} \in [M]^k$, we use $\mathsf{f}\left(w_{[k]}\right) \triangleq (\mathsf{f}(w_1), \ldots, \mathsf{f}(w_k))$ to denote the collection of encoded descriptions produced by the encoders.

Decoder Design: Upon receiving the first n_0 samples of the channel output Y, the decoder runs the following composite hypothesis test

$$\mathsf{g}_0(y^{n_0}) = \begin{cases} 0 & \text{if } \ h(y^{n_0}) \le \gamma_0 \\ \mathsf{e} & \text{otherwise} \end{cases} \tag{53}$$

to decide whether there are any active transmitters. Decoder output 0 signifies that the decoder decides that all transmitters are silent, sending a feedback bit '1' to all transmitters to start a new coding epoch. Decoder output e indicates that the receiver believes that there are active transmitters; the decoder transmits feedback bit '0' to the transmitters, telling them that it is not ready to decode, and therefore that transmissions must continue. Statistic $h\colon \mathcal{Y}^{n_0} \mapsto \mathbb{R}$ is used to decide whether any transmitters are active.

For each $k \ge 1$, decoder g_k observes output y^{n_k} and employs a single threshold rule

$$\mathsf{g}_k(y^{n_k}) = \begin{cases} w_{[k]} & \text{if } \ \imath_k(\mathsf{f}\left(w_{[k]}\right)^{n_k}; y^{n_k}) > \log \gamma_k \\ & \quad \text{and } w_i < w_j \ \forall\, i < j \\ \mathsf{e} & \text{otherwise} \end{cases} \tag{54}$$

for some constant γ_k chosen before the transmission starts. Under permutation-invariance (2) and identical encoding (4), all permutations of the message vector $w_{[k]}$ give the same information density. We use the ordered permutation specified in (54) as a representative of the equivalence class with respect to the binary relation $\overset{\pi}{=}$. The choice of a representative is immaterial since decoding is identity-blind. When there is more than one ordered $w_{[k]}$ that satisfies the threshold condition, decoder g_k chooses among these options arbitrarily. All such events are counted as errors in the analysis in Section IV-B, below. If the decoder output is a message vector $w_{[k]}$, then the decoder sends feedback bit '1', telling them to stop transmission. Otherwise, the decoder sends feedback bit '0', and the epoch continues. For $k \ge 1$, the decoder $\mathsf{g}_k(y^{n_k})$ depends on U through its dependence on the encoding function $\mathsf{f}\left(w_{[k]}\right)$; for $k = 0$, $\mathsf{g}_0(y^{n_0})$ does not depend on U.

The proof of Theorem 2, below, bounds the error probability for the proposed code.

B. Proof of Theorem 2

In the discussion that follows, we bound the error probability of the code $(\mathsf{f}, \{\mathsf{g}_k\}_{k=0}^{K})$ defined above. For $k = 0$, the only error event is that the received vector at time n_0, $Y_0^{n_0}$, fails to pass the test

$$\epsilon_0 \le \mathbb{P}\left[\mathsf{g}_0(Y_0^{n_0}) \neq 0 | W_0 = 0\right] \tag{55}$$

given in (53). For $k > 0$, the analysis relies on the independence of codewords $\mathsf{f}(W_i)$ and $\mathsf{f}(W_j)$ from distinct transmitters i and j. Given identical encoders and i.i.d. codeword design, this assumption is valid provided that $W_i \neq W_j$; we therefore count events of the form $W_i = W_j$ as errors. Let $\mathbb{P}_{\text{rep}}$ denote the probability of such a repetition; the union bound gives

$$\mathbb{P}_{\text{rep}} \le \frac{k(k-1)}{2M}. \tag{56}$$

The discussion that follows uses $w_{[k]}^* = (1, 2, \ldots, k)$ as an example instance of a message vector $w_{[k]}$ in which $w_i \neq w_j$ for all $i \neq j$. The set $\overline{\mathcal{W}}_{[s]}$ describes all ordered message vectors that do not share any messages in common with $w_{[k]}^*$, i.e.,

$$\overline{\mathcal{W}}_{[s]} \triangleq \{\overline{w}_{[s]} \in [M]^s \colon \overline{w}_1 > k, \overline{w}_i < \overline{w}_j \ \ \forall i < j\}. \tag{57}$$

Let the components of the vectors $(X_{[k]}^{n_k}, \bar{X}_{[k]}^{n_k}, Y_k^{n_k})$ be i.i.d. with joint distribution

$$\begin{aligned} &P_{X_{[k]}\bar{X}_{[k]}Y_k}(x_{[k]}, \bar{x}_{[k]}, y_k) \\ &\quad = P_{X_{[k]}}(x_{[k]}) P_{X_{[k]}}(\bar{x}_{[k]}) P_{Y_k|X_{[k]}}(y_k | x_{[k]}). \end{aligned} \tag{58}$$

Recall that the information density $\imath_t(x_{[t]}^{n_t}; y_t^{n_t})$ in (14) is defined with respect to $(X_{[t]}^{n_t}, Y_t^{n_t})$, not with respect to $(\bar{X}_{[t]}^{n_t}, Y_t^{n_t})$. The resulting error bound proceeds as shown in (59)–(64); here $X_{[k]}$ is the vector of transmitted codewords, and $\bar{X}_{[s]}(\overline{w}_{[s]})$ is an i.i.d. copy of $\bar{X}_{[s]}$, which represents the codeword for a collection of messages $\overline{w}_{[s]} \in \overline{\mathcal{W}}_{[s]}$ that was not transmitted. Line (60) separates the case where at least one message is repeated from the case where there are no repetitions. Lines (61)–(62) enumerate the error events in the no-repetition case; these include all cases where the transmitted codeword passes the binary hypothesis test (53) for "no active transmitters" (61), all cases where a subset of the transmitted codewords meets the threshold for some $t < k$ (61), all cases where a codeword that is incorrect in s dimensions and correct in $t-s$ dimensions meets the threshold for $t \le k$ (62), and all cases where the transmitted codeword fails to meet the threshold (62). We apply the union bound and the symmetry of the code design to represent the probability of each case by the probability of an example instance times the number of instances. Equations (63)-(64) apply the bound

$$\epsilon_k = \frac{1}{M^k}\sum_{w_{[k]}\in[M]^k}\mathbb{P}[\{\mathsf{g}_0(Y_k^{n_0})\neq \mathsf{e}\}\cup\{\cup_{t=1}^{k-1}\mathsf{g}_t(Y_k^{n_t})\neq\mathsf{e}\}\cup\{\mathsf{g}_k(Y_k^{n_k})\overset{\pi}{\neq} w_{[k]}\}|W_{[k]}=w_{[k]}] \tag{59}$$

$$\leq \mathbb{P}_{\text{rep}} + (1-\mathbb{P}_{\text{rep}})\mathbb{P}[\{\mathsf{g}_0(Y_k^{n_0})\neq \mathsf{e}\}\cup\{\cup_{t=1}^{k-1}\mathsf{g}_t(Y_k^{n_t})\neq\mathsf{e}\}\cup\{\mathsf{g}_k(Y_k^{n_k})\overset{\pi}{\neq} w^*_{[k]}\}|W_{[k]}=w^*_{[k]}] \tag{60}$$

$$\leq \mathbb{P}_{\text{rep}} + \mathbb{P}[\mathsf{g}_0(Y_k^{n_0})\neq\mathsf{e}|W_{[k]}=w^*_{[k]}] + \sum_{t=1}^{k-1}\binom{k}{t}\mathbb{P}[\mathsf{g}_t(Y_k^{n_t})\overset{\pi}{=} w^*_{[t]}|W_{[k]}=w^*_{[k]}] \tag{61}$$

$$+\sum_{t=1}^{k}\sum_{s=1}^{t}\binom{k}{t-s}\mathbb{P}[\cup_{\overline{w}_{[s]}\in\overline{\mathcal{W}}_{[s]}}\{\mathsf{g}_t(Y_k^{n_t})\overset{\pi}{=}(\overline{w}_{[s]}, w^*_{[s+1:t]})\}|W_{[k]}=w^*_{[k]}] + \mathbb{P}[\mathsf{g}_k(Y_k^{n_k})=\mathsf{e}|W_{[k]}=w^*_{[k]}] \tag{62}$$

$$\leq \frac{k(k-1)}{2M} + \mathbb{P}\left[h(Y_k^{n_0})\leq\gamma_0\right] + \sum_{t=1}^{k-1}\binom{k}{t}\mathbb{P}[\imath_t(X_{[t]}^{n_t};Y_k^{n_t})>\log\gamma_t] \tag{63}$$

$$+\sum_{t=1}^{k}\sum_{s=1}^{t}\binom{k}{t-s}\mathbb{P}[\cup_{\overline{w}_{[s]}\in\overline{\mathcal{W}}_{[s]}}\{\imath_t(\bar{X}_{[s]}^{n_t}(\overline{w}_{[s]}), X_{[s+1:t]}^{n_t};Y_k^{n_t})>\log\gamma_t\}] + \mathbb{P}[\imath_k(X_{[k]}^{n_k};Y_k^{n_k})\leq\log\gamma_k] \tag{64}$$

in (56) and replace decoders by the threshold rules in their definitions.

The delay in applying the union bound to the first probability in (64) is deliberate. It allows us to exploit the symmetry assumptions on the channel and to use a single threshold rule instead of 2^k-1 threshold rules as in [10]–[13]. Applying the bound

$$\mathbb{P}\left[\bigcup_{\overline{w}_{[s]}\in\overline{\mathcal{W}}_{[s]}}\{\imath_t(\bar{X}_{[s]}^{n_t}(\overline{w}_{[s]}), X_{[s+1:t]}^{n_t};Y_k^{n_t})>\log\gamma_t\}\right] \tag{65}$$

$$= \mathbb{P}\Bigg[\bigcup_{\overline{w}_{[s]}\in\overline{\mathcal{W}}_{[s]}}\{\imath_t(\bar{X}_{[s]}^{n_t}(\overline{w}_{[s]}), X_{[s+1:t]}^{n_t};Y_k^{n_t})>\log\gamma_t\}$$
$$\bigcap\left\{\imath_t(X_{[s+1:t]}^{n_t};Y_k^{n_t}) > n_t\mathbb{E}[\imath_t(X_{[s+1:t]};Y_k)]+\lambda_{s,t}^k\right\}\Bigg]$$
$$+\mathbb{P}\Bigg[\bigcup_{\overline{w}_{[s]}\in\overline{\mathcal{W}}_{[s]}}\{\imath_t(\bar{X}_{[s]}^{n_t}(\overline{w}_{[s]}), X_{[s+1:t]}^{n_t};Y_k^{n_t})>\log\gamma_t\}$$
$$\bigcap\left\{\imath_t(X_{[s+1:t]}^{n_t};Y_k^{n_t}) \leq n_t\mathbb{E}[\imath_t(X_{[s+1:t]};Y_k)]+\lambda_{s,t}^k\right\}\Bigg]$$
$$\leq \mathbb{P}\left[\imath_t(X_{[s+1:t]}^{n_t};Y_k^{n_t}) > n_t\mathbb{E}[\imath_t(X_{[s+1:t]};Y_k)]+\lambda_{s,t}^k\right]$$
$$+\mathbb{P}\Bigg[\bigcup_{\overline{w}_{[s]}\in\overline{\mathcal{W}}_{[s]}}\{\imath_t(\bar{X}_{[s]}^{n_t}(\overline{w}_{[s]});Y_k^{n_t}|X_{[s+1:t]}^{n_t}) >$$
$$\log\gamma_t - n_t\mathbb{E}[\imath_t(X_{[s+1:t]};Y_k)]-\lambda_{s,t}^k\}\Bigg] \tag{66}$$

before applying the union bound to the first probability in (64) yields a tighter bound. Combining (64) and (66) and applying the union bound to the second probability in (66) completes the proof. □

C. Proof of Theorem 1

We begin by enumerating our choice of parameters. Let

$$\log\gamma_k = n_kI_k - \tau_k\sqrt{n_kV_k} \tag{67}$$

$$\lambda_{s,t}^k = \frac{n_t}{2}\left(I_t(X_{[s]};Y_t|X_{[s+1:t]}) - \frac{s}{t}I_t\right) \tag{68}$$

$$n_k = \gamma_k^2\left(\frac{e}{k}(M-k)\right)^{-2k} \tag{69}$$

for every $1\leq s\leq t\leq k$, where

$$\tau_k \triangleq Q^{-1}\left(\epsilon_k - \frac{B_k+C_k}{\sqrt{n_k}}\right), \tag{70}$$

C_k is a constant to be chosen in (102),

$$B_k \triangleq \frac{6T_k}{V_k^{3/2}} \tag{71}$$

is the Berry-Esseen constant [44, Chapter XVI.5 Th. 2] (which is finite by the moment assumptions (19) and (20)), and

$$T_k \triangleq \mathbb{E}\left[|\imath_k(X_{[k]};Y_k)-I_k|^3\right]. \tag{72}$$

The choice of the threshold γ_k (67) follows the approach established for the point-to-point channel in [36]. The constants $\left\{\lambda_{s,t}^k\right\}$ used in the error probability bound (47e)–(47f) are set in (68) to ensure that $\lambda_{s,t}^k>0$ when $s<t$ (see Lemma 2) and that $\lambda_{s,t}^k=0$ when $s=t$. The blocklengths n_k in (69) are chosen to ensure that for a large enough M, $n_1<\ldots<n_K$ (see Lemma 1).

Applying the choices in (67) and (69) and the Taylor series expansion of $Q^{-1}(\cdot)$, the size of the codebook admits the following expansion

$$k\log M = n_kI_k - \sqrt{n_kV_k}Q^{-1}(\epsilon_k) - \frac{1}{2}\log n_k + O(1). \tag{73}$$

Therefore, to prove Theorem 1, we need to show that the probability of decoding error at time n_k is bounded by ϵ_k. Towards that end, we sequentially bound the terms in Theorem 2 using the parameters chosen in (67)–(69).

- (47a): As noted previously, this is the dominant term. Since $\imath_k(X_{[k]}^{n_k};Y_k^{n_k})$ is a sum of n_k independent random variables, by the Berry-Esseen theorem [44, Chapter XVI.5 Th. 2], (67), (70), and (71),

$$\mathbb{P}\left[\imath_k(X_{[k]}^{n_k};Y_k^{n_k})\leq\log\gamma_k\right]\leq\epsilon_k-\frac{C_k}{\sqrt{n_k}}. \tag{74}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2020.3047630, IEEE Transactions on Information Theory

- (47b): The test statistic $h(\cdot)$ and the threshold γ_0 given in (53) are chosen in Section VI to satisfy

$$\mathbb{P}\left[h(Y_k^{n_0}) \leq \gamma_0\right] \leq \frac{E_k}{\sqrt{n_k}} \tag{75}$$

$$\mathbb{P}\left[h(Y_0^{n_0}) > \gamma_0\right] \leq \epsilon_0 \tag{76}$$

for some constant $E_k > 0$. Lemma 5, below, bounds the type-II error in (75) in terms of n_0 when the type-I error in (76) is bounded by ϵ_0.

Lemma 5. *Fix $\epsilon_0 \in (0,1)$. Assume that (18) holds. Then there exists a test function $h(\cdot)$ such that (76) is satisfied and*

$$\mathbb{P}\left[h(Y_k^{n_0}) \leq \gamma_0\right] \leq \exp\{-n_0 C' + o(n_0)\} \tag{77}$$

for some $C' > 0$ depending on the output distributions P_{Y_i} for $i = 0, \ldots, K$.

Proof: See Section VI. ■

From (73), $n_k = O(n_1)$ for $k \geq 1$. To make (77) behave as $O\left(\frac{1}{\sqrt{n_k}}\right)$ in Lemma 5, we pick n_0 as in (28) with $c_0 = \frac{1}{2C'}$.

- (47c): According to (69), the upper bound $\frac{k(k-1)}{2M}$ on $\mathbb{P}_{\text{rep}}$ in (56) decays exponentially with n_k.
- (47d): Define p as

$$p \triangleq \mathbb{P}[\imath_t(X_{[t]}; Y_k) > -\infty]. \tag{78}$$

We next analyze (47d) for the cases $p = 1$ and $p < 1$.

Case 1: $p = 1$. By Lemma 3 and moment assumption (21),

$$I_t - \mathbb{E}\left[\imath_t(X_{[t]}; Y_k)\right] - \tau_t \sqrt{\frac{V_t}{n_t}} > 0 \tag{79}$$

for sufficiently large n_t. Chebyshev's inequality gives

$$\begin{aligned}&\mathbb{P}[\imath_t(X_{[t]}^{n_t}; Y_k^{n_t}) > \log\gamma_t] \\ &\leq \frac{\mathrm{Var}[\imath_t(X_{[t]}; Y_k)]}{n_t\left(I_t - \mathbb{E}\left[\imath_t(X_{[t]}; Y_k)\right] - \tau_t\sqrt{\frac{V_t}{n_t}}\right)^2}.\end{aligned} \tag{80}$$

The right side of (80) behaves as $O\left(\frac{1}{n_t}\right)$.

Case 2: $p < 1$. Here

$$\begin{aligned}&\mathbb{P}[\imath_t(X_{[t]}^{n_t}; Y_k^{n_t}) > \log\gamma_t] \\ &\leq \mathbb{P}[\imath_t(X_{[t]}^{n_t}; Y_k^{n_t}) > -\infty] && (81)\\ &= p^{n_t}, && (82)\end{aligned}$$

where (82) holds because $\imath_t(X_{[t]}^{n_t}; Y_k^{n_t})$ is the sum of n_t i.i.d. random variables, and that sum is greater than $-\infty$ if and only if all the summands satisfy the same inequality. From (80) and (82), (47d) contributes $O\left(\frac{1}{n_k}\right)$ to our error bound.

- (47e): As in the analysis of (47d), we define

$$q \triangleq \mathbb{P}[\imath_t(X_{[s+1:t]}; Y_k) > -\infty], \tag{83}$$

and treat the cases $q = 1$ and $q < 1$ separately. Observe that for $q = 1$, Chebyshev's inequality implies

$$\begin{aligned}&\mathbb{P}\left[\imath_t(X_{[s+1:t]}^{n_t}; Y_k^{n_t}) > n_t\mathbb{E}[\imath_t(X_{[s+1:t]}; Y_k)] + \lambda_{t,s}^k\right] \\ &\leq \frac{\mathrm{Var}\left[\imath_t(X_{[s+1:t]}; Y_k)\right]}{n_t\left(\frac{1}{2}(I_t(X_{[s]}; Y_t | X_{[s+1:t]}) - \frac{s}{t}I_t)\right)^2},\end{aligned} \tag{84}$$

which is of order $O\left(\frac{1}{n_t}\right)$ by the moment assumption (21) and Lemma 2.

For $q < 1$,

$$\mathbb{P}\left[\imath_t(X_{[s+1:t]}^{n_t}; Y_k^{n_t}) > n_t\mathbb{E}[\imath_t(X_{[s+1:t]}; Y_k)] + \lambda_{t,s}^k\right] \leq q^{n_t}. \tag{85}$$

Therefore (47e) contributes $O\left(\frac{1}{n_k}\right)$ to our error bound.

- (47f): First, consider the case where $s < t \leq k$. By Lemma 3 and Chernoff's bound,

$$\begin{aligned}&\mathbb{P}[\imath_t(\bar{X}_{[s]}^{n_t}; Y_k^{n_t} | X_{[s+1:t]}^{n_t}) \\ &\quad > \log\gamma_t - n_t\mathbb{E}[\imath_t(X_{[s+1:t]}; Y_k)] - \lambda_{s,t}^k] && (86)\\ &\leq \mathbb{P}[\imath_t(\bar{X}_{[s]}^{n_t}; Y_k^{n_t} | X_{[s+1:t]}^{n_t}) \\ &\quad > \log\gamma_t - n_t I_t(X_{[s+1:t]}; Y_t) - \lambda_{s,t}^k] && (87)\\ &\leq \mathbb{E}\left[\exp\left\{\imath_t\left(\bar{X}_{[s]}^{n_t}; Y_k^{n_t} | X_{[s+1:t]}^{n_t}\right)\right\}\right] \\ &\quad \cdot \exp\{-(\log\gamma_t - n_t I_t(X_{[s+1:t]}; Y_t) - \lambda_{s,t}^k)\} && (88)\\ &= \exp\{-(\log\gamma_t - n_t I_t(X_{[s+1:t]}; Y_t) - \lambda_{s,t}^k)\}. && (89)\end{aligned}$$

Using Stirling's bound

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k, \tag{90}$$

we find that for all $s \leq t \leq k$

$$\begin{aligned}\log\binom{M-k}{s} &\leq s\log\left(\frac{e(M-k)}{s}\right) && (91)\\ &\leq s\log\left(\frac{e(M-t)}{t}\right) + s\log\left(\frac{t}{s}\right) && (92)\\ &= \frac{s}{t}\left(\log\gamma_t - \frac{1}{2}\log n_t\right) + s\log\left(\frac{t}{s}\right), && (93)\end{aligned}$$

where (93) follows from (69). From (67), (68), (89), and (93), we have

$$\begin{aligned}&\binom{M-k}{s}\mathbb{P}[\imath_t(\bar{X}_{[s]}^{n_t}; Y_k^{n_t} | X_{[s+1:t]}^{n_t}) \\ &\quad > \log\gamma_t - n_t I_t(X_{[s+1:t]}; Y_t) - \lambda_{s,t}^k] && (94)\\ &\leq \exp\Bigg\{ - n_t\frac{1}{2}\left(I_t(X_{[s]}; Y_t | X_{[s+1:t]}) - \frac{s}{t}I_t\right) \\ &\quad + \left(1 - \frac{s}{t}\right)\tau_t\sqrt{n_t V_t} - \frac{s}{2t}\log n_t + s\log\left(\frac{t}{s}\right)\Bigg\}. && (95)\end{aligned}$$

Lemma 2 ensures that the exponent in (95) is negative for n_t large enough.

For $s = t < k$, from (89) and (93) with $s = t$, we get

$$\binom{M-k}{t}\mathbb{P}[\imath_t(\bar{X}_{[t]}^{n_t}; Y_k^{n_t}) > \log\gamma_t] \leq \frac{\binom{M-k}{t}}{\gamma_t} \leq \frac{1}{\sqrt{n_t}}. \tag{96}$$

For $s = t = k$, following the change of measure technique (e.g., [45, Prop. 17.1]), one can rewrite an expectation with respect to measure Q as an expectation with respect to measure P, giving

$$Q\left[Z \in \mathcal{A}\right] = \mathbb{E}_P\left[\left(\frac{P[Z]}{Q[Z]}\right)^{-1} 1\left\{Z \in \mathcal{A}\right\}\right]. \tag{97}$$

Switching to the measure $P_{X_{[k]}} P_{Y_k|X_{[k]}}$ in this way, by (90) and the parameter choice (69), we write

$$\begin{aligned} &\binom{M-k}{k} \mathbb{P}[\imath_k(\bar{X}_{[k]}^{n_k}; Y_k^{n_k}) > \log \gamma_k] \\ &\le \left(\frac{e}{k}(M-k)\right)^k \mathbb{E}\left[\exp\{-\imath_k(X_{[k]}^{n_k}; Y_k^{n_k})\}\right. && (98) \\ &\quad \left.\cdot 1\{\imath_k(X_{[k]}^{n_k}; Y_k^{n_k}) > \log \gamma_k\}\right] \\ &\le \frac{D_k}{n_k}, && (99) \end{aligned}$$

where

$$D_k \triangleq 2\left(\frac{\log 2}{\sqrt{2\pi V_k}} + 2B_k\right) \tag{100}$$

and B_k is defined in (71). To justify (99), notice that $\imath_k(X_{[k]}^{n_k}; Y_k^{n_k})$ is a sum of i.i.d. random variables; in [36, Lemma 47], Polyanskiy et al. derive a sharp bound on the expectation

$$\mathbb{E}\left[\exp\left(-\sum_{i=1}^{n} Z_i\right) 1\left\{\sum_{i=1}^{n} Z_i > \gamma\right\}\right] \tag{101}$$

when the Z_i's are independent. Applying that bound with $Z_i = \imath_k(X_{[k],i}; Y_{k,i})$ yields (99). Note that D_k is finite by the moment assumptions (19) and (20). Combining the bounds for the three cases in (95), (96), and (99), we conclude that (47f) contributes $O\left(\frac{1}{\sqrt{n_k}}\right)$ to the total error. Finally, we set the constant C_k in (70) to ensure

$$(47b) + (47c) + (47d) + (47e) + (47f) \le \frac{C_k}{\sqrt{n_k}}. \tag{102}$$

The existence of such a constant is guaranteed by our analysis above demonstrating that the terms (47b)–(47f) do not contribute more than $O\left(\frac{1}{\sqrt{n_k}}\right)$ to the total.[10]

Due to (74) and (102), the total probability of making an error at time n_k is bounded by ϵ_k, and the proof of Theorem 1 is complete. □

V. Discussion of the Main Result

A. Refining the Third-Order Term Using a Maximum Likelihood Decoder

For a RAC that satisfies the conditions in Theorem 1 and the conditional variance condition

$$\mathbb{E}\left[\mathrm{Var}\left[\imath_k(X_{[k]}; Y_k)|Y_k\right]\right] > 0 \quad \forall s \in [k], \tag{103}$$

we can improve the achievable third-order performance in (27) from $-\frac{1}{2}\log n_k$ to $+\frac{1}{2}\log n_k$. Prior work showing the achievability of the $+\frac{1}{2}\log n$ third-order term includes [46, Th. 53] for point-to-point channels satisfying (103) with $k = 1$, [47, Th. 1] for the Gaussian point-to-point channel, [48, Th. 7], [49, Th. 14] for discrete memoryless MACs satisfying (103), and [43, Th. 2 and 4], [50, Th. 2 and 4] for the Gaussian MAC and RAC. We can achieve the result here by replacing the threshold rule in (54) with a combination of a hypothesis test and a maximum likelihood decoder, giving

$$\mathsf{g}_k(U, y^{n_k}) = \begin{cases} \arg\max\limits_{w_{[k]}} \imath_k(\mathsf{f}(w_{[k]})^{n_k}; y^{n_k}) & \text{if } h_k(y^{n_k}) \le \gamma_k \\ \mathsf{e} & \text{otherwise,} \end{cases} \tag{104}$$

where the maximum is over the ordered message vectors $w_{[k]}$, and $h_k(\cdot)$ is a suitable test function that allows us to distinguish P_{Y_k} from any P_{Y_t} with $t \neq k$. As in prior work, the analysis applies the random coding union bound from [36, Th. 16]. As discussed in Section VI, suitable test functions $h_k(\cdot)$ can be found provided that $P_{Y_k} \neq P_{Y_t}$ for all $t \neq k$. For instance, in [43], we use $h_k(y^{n_k}) = \left|\frac{1}{n_k}\left\|y^{n_k}\right\|^2 - (1 + kP)\right|$ for the Gaussian RAC, where P is the maximal power constraint. The result does not apply to channels such as the adder-erasure RAC (23), which does not satisfy the condition in (103).

B. Choosing the Input Distribution P_X

Although there are RACs for which a single input distribution P_X achieves the capacity for all k-MACs, $k \in [K]$, (e.g., the adder-erasure channel), the permutation-invariance (2) and reducibility (3) assumptions do not imply that such a distribution exists for all RACs. In the following, we discuss how to choose the input distribution when the optimal input distribution varies with k.

Given a permutation-invariant (2) and reducible (3) RAC, M, $\boldsymbol{\epsilon} = (\epsilon_0, \ldots, \epsilon_K)$, and any P_X such that (16)–(21) are satisfied for the given RAC under input distribution P_X, let

$$\mathcal{R}(M, \boldsymbol{\epsilon}, P_X) = \{(R_0, \ldots, R_K) \colon (27) \text{ and } (28) \text{ hold}\} \tag{105}$$

denote the achievable rate region under input distribution P_X. Here

$$R_k = \frac{\log M}{n_k} \text{ for all } k \in \{0, \ldots, K\}. \tag{106}$$

Let

$$\mathcal{R}(M, \boldsymbol{\epsilon}) = \bigcup_{P_X \colon (16)\text{–}(21) \text{ hold}} \mathcal{R}(M, \boldsymbol{\epsilon}, P_X) \tag{107}$$

denote the achievable rate region over all i.i.d. input distributions. A point in this set is called *dominant* if no other points in the set are element-wise greater than or equal to that point. To optimize the achievable rate vector over the allowed input distributions, we must choose a distribution P_{X^*} that achieves a dominant point for the set $\mathcal{R}(M, \boldsymbol{\epsilon})$. Note that for the dominant points of $\mathcal{R}(M, \boldsymbol{\epsilon})$ corresponding to different values of P_{X^*}, there is an $O(1)$ difference between the left and right sides of the inequalities in (27). If the achievable rate region $\mathcal{R}(M, \boldsymbol{\epsilon})$ is not convex, it can be improved to its convex hull using time sharing. For the modifications to the coding strategy that enable us to incorporate time sharing, see [10], [12], [13].

To illustrate what happens when different P_{X^*} values achieve different dominant points of $\mathcal{R}(M, \boldsymbol{\epsilon})$, we consider the following example.

[10] Our bounds on (47b)–(47f) technically depend on γ_k and therefore on C_k. However, it is easy to see that their dependence on C_k is weak, and for large enough n_k, it can be eliminated entirely. Thus the choice of C_k satisfying (102) is possible.

Example 2. Consider a RAC with $K = 2$, $\mathcal{X} = \mathcal{Y}_2 = \{0, 1\}$, and transition probability matrix $P_{Y_2|X_1,X_2}$

$Y_2 \backslash X_1X_2$	00	01	10	11
0	$1-b$	b	b	$1-a$
1	b	$1-b$	$1-b$	a

(108)

where $a, b \in [0, 1]$. This RAC is permutation-invariant since the "01" and the "10" columns are identical. When $k = 1$, the channel reduces to the binary symmetric channel with crossover probability b. Fig. 3 illustrates the set of achievable rate vectors $\mathcal{R}(M, \boldsymbol{\epsilon})$ (neglecting the $O(1)$ term in (27)) with $\log M = 1000$ and $\boldsymbol{\epsilon} = 10^{-3}\mathbf{1}$ for two choices of parameters in the channel in (108). In Fig. 3a, $a = 0.7, b = 0.11$, and in Fig. 3b, $a = b = 0.11$; for each, the finite blocklength and capacity boundaries are demonstrated. In Fig. 3a, the dominant points are highlighted. The input distribution $P_{X^*} = (0.65, 0.35)$ (i.e., the Bernoulli(0.35) distribution) achieves the dominant point $(R_1, R_2) = (0.400, 0.204)$; the corresponding region $\mathcal{R}(M, \boldsymbol{\epsilon}, P_{X^*})$ is shown as the region bounded by the dashed lines. In Fig. 3b, the only dominant point $(0.437, 0.227)$ is achieved by the input distribution $P_{X^*} = (0.5, 0.5)$ (i.e., the Bernoulli(0.5) distribution.) Therefore, for the channel in Fig. 3b, the achievable rate region $\mathcal{R}(M, \boldsymbol{\epsilon})$ coincides with $\mathcal{R}(M, \boldsymbol{\epsilon}, P_{X^*})$, and we must choose P_{X^*} as our input distribution. For this channel, $P_{X^*} = (0.5, 0.5)$ simultaneously maximizes the mutual informations I_1 and I_2, and the maxima are $I_1 = I_2 = 0.5$.

C. Discussion of the Converse

Even for MACs with only 2 transmitters, the capacity region for the MAC remains incompletely understood. A brief summary of related results follows. For any blocklength n and average error probability $\epsilon \in (0, 1)$, let

$$\mathcal{R}(n, \epsilon) = \left\{ \left(\frac{\log M_1}{n}, \frac{\log M_2}{n} \right) : \exists \text{ an } (n, M_1, M_2, \epsilon) \text{ code} \right\} \tag{109}$$

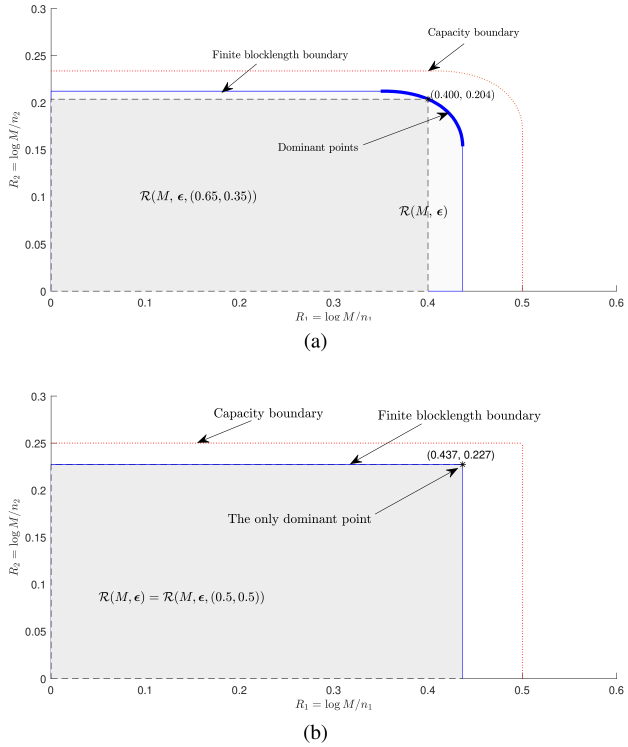
denote the set of achievable rate pairs, where M_i is the message size for transmitter $i \in \{1, 2\}$. The capacity region of the MAC [51], [52] is

$$\begin{aligned} \mathcal{C} = \bigcup_{P_Q P_{X_1|Q} P_{X_2|Q}} \{(R_1, R_2)\colon & \\ R_1 &\leq I_2(X_1; Y_2 | X_2, Q) \\ R_2 &\leq I_2(X_2; Y_2 | X_1, Q) \\ R_1 + R_2 &\leq I_2(X_1, X_2; Y_2 | Q)\}, \end{aligned} \tag{110}$$

where Q is the time sharing random variable. In [53], Dueck uses the blowing-up lemma to derive the first strong converse for discrete memoryless MACs. In [54], for discrete memoryless MACs, Ahlswede uses a wringing technique to show

$$\mathcal{R}(n, \epsilon) \subseteq \mathcal{C} + O\left(\frac{\log n}{\sqrt{n}} \right) \mathbf{1}, \tag{111}$$

which improves Dueck's result. The coefficients of the term $O\left(\frac{\log n}{\sqrt{n}}\right)\mathbf{1}$ in (111) are bounded by a multiple of the product of input and output alphabet sizes $|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}_2|$. In [55, Th. 1], Fong and Tan improve Ahlswede's second-order term $O\left(\frac{\log n}{\sqrt{n}}\right)\mathbf{1}$ to $O\left(\sqrt{\frac{\log n}{n}}\right)\mathbf{1}$ for the Gaussian MAC. They derive this result by applying Ahlswede's wringing technique [54] to quantized channel inputs. In [56], Kosut further improves the second-order term to $O\left(\frac{1}{\sqrt{n}}\right)\mathbf{1}$. The second-order term in [56, Th. 7] has the same order and, for some channels, the same sign as the best-known second-order achievable term in [13]. Kosut's result applies to all discrete memoryless MACs and to the Gaussian MAC. To prove this converse, Kosut introduces a new measure of dependence between two random variables called "wringing dependence." A key aspect of the approach is to restrict the channel inputs so that the wringing dependence between them is small.



(a)

(b)

Fig. 3: The achievable rate region from Theorem 1 (excluding the $O(1)$ term) applied to the channel in (108) with $\log M = 1000$ and $\epsilon_k = 10^{-3}$ for $k \in [2]$. The results are shown for (a) $a = 0.7$ and $b = 0.11$ and blocklengths $(n_1, n_2) = (2501, 4904)$, and (b) for $a = b = 0.11$ and blocklengths $(n_1, n_2) = (2290, 4399)$.

In [57], Moulin proposes a new converse technique for maximum-error capacity. His approach relies on strong large deviations for binary hypothesis tests and leads to a second-order term as in (27) when no time sharing is needed. Since the capacity regions for the maximum and average error probability can differ [58], Moulin's result does not give a converse for the average-error capacity. Whether it is possible to derive a converse for the average-error capacity with a second-order term matching the ones in [10]–[13], [25] remains an open problem.

In the sparse recovery literature, where achievability proofs typically consider the expected error probability evaluated

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2020.3047630, IEEE Transactions on Information Theory

under i.i.d. codebook design (see, e.g., [30]–[34]), converses derive lower bounds on the expected error probability *assuming i.i.d. code design*. Although a lower bound on the expected error probability for our problem could be derived using tools from [33], such a bound would yield a bound for the best i.i.d. random code rather than a bound for all possible codes.

D. A RAC Code That Decodes Transmitter Identity

While the use of identical encoding at all transmitters has a number of practical advantages, the techniques employed in this work are not limited to that case.

We next briefly explore the use of distinct encoders at each transmitter of a RAC. Under permutation-invariance (2) and identical encoding, the decoder cannot distinguish which transmitter sent each of the decoded messages. Maintaining permutation-invariance but replacing identical encoders with a different instance of the same random codebook for each encoder, we get a code that achieves the same first- and second-order terms as in Theorem 1, with a decoder that can also associate the corresponding transmitter identity to each decoded message. The following definition formalizes the resulting RAC codes.

Definition 2. *An* $(M, \{(n_k, \epsilon_k)\}_{k=0}^K)$ *identity-preserving code comprises a collection of encoding functions*

$$\mathsf{f}_k\colon \mathcal{U}\times[M]\to\mathcal{X}^{n_K},\quad k=1,\dots,K, \tag{112}$$

and a collection of decoding functions

$$\mathsf{g}_k\colon \mathcal{U}\times\mathcal{Y}_k^{n_k}\to\left\{[M]^k\times\binom{[K]}{k}\right\}\cup\{\mathsf{e}\},\ k=0,1,\dots,K, \tag{113}$$

where erasure symbol e *is the decoder's output when the decoder is not ready to decode. At the start of each epoch, a random variable* $U\in\mathcal{U}$*, with* $U\sim P_U$*, is generated independently of the transmitter activity, and revealed to the transmitters and the receiver for use in initializing the encoders and the decoder. If the set of active transmitters* $\mathcal{A}\subseteq[K]$ *satisfies* $|\mathcal{A}|=k>0$*, i.e.,* k *transmitters are active, then the messages of* $\mathcal{A}$ *and their corresponding transmitter identities are decoded correctly at time* n_k*, with probability at least* $1-\epsilon_k$*, i.e.,*

$$\frac{1}{M^k}\sum_{w_{\mathcal{A}}\in[M]^k}\mathbb{P}\Bigg[\{\mathsf{g}_k(U,Y_k^{n_k})\neq(w_{\mathcal{A}},\mathcal{A})\}\bigcup\left\{\bigcup_{t=0}^{k-1}\{\mathsf{g}_t(U,Y_k^{n_t})\neq\mathsf{e}\}\right\}\Bigg|\,W_{\mathcal{A}}=w_{\mathcal{A}}\Bigg]\leq\epsilon_k, \tag{114}$$

where $W_{\mathcal{A}}$ *are the independent and equiprobable messages of the transmitters in* $\mathcal{A}$*, and the given probability is calculated using the conditional distribution* $P_{Y_k^{n_k}|X_{\mathcal{A}}^{n_k}}=P_{Y_k|X_{\mathcal{A}}}^{n_k}$ *where* $X_i^{n_k}=\mathsf{f}_i(U,W_i)^{n_k}$*,* $i\in\mathcal{A}$*. If* $\mathcal{A}=\emptyset$*, then the probability that at time* n_0 *the receiver decodes to the unique message in set* $[M]^0=\{0\}$ *is no smaller than* $1-\epsilon_0$*. That is,*

$$\mathbb{P}\left[\mathsf{g}_0(U,Y_0^{n_0})\neq 0|W_{[0]}=0\right]\leq\epsilon_0. \tag{115}$$

If we continue to assume permutation-invariance (2) and to employ the same input distribution P_X at all encoders, then the channel output statistics again depend on the dimension of the channel input but not on the identity of the active transmitters. In this case, we can apply the proof from the identical-encoding single-threshold-decoding argument in Section IV-A to derive an achievability result for the general case.[11] In particular, consider a code with KM (rather than M) messages. Replacing M by KM in Theorem 1 implies that our RAC code with identical encoders gives a penalty of $-k\log K$ on the right-hand side of the rate bound (27). Suppose that we use this identical-encoding code to design a general code in which codewords indexed from $(t-1)M+1$ to tM are used exclusively by transmitter t for $t=1,\dots,K$. Since each message belongs to a single transmitter, the list of decoded messages reveals the identities of the active transmitters. Under this allocation of codewords, the repetition error $\mathbb{P}_{\text{rep}}$ in (56) disappears since transmitters send messages from distinct sets. The error probability from decoding the wrong codeword values decreases since there are fewer legitimate codeword combinations to consider. Therefore, in the case where K is a finite constant and the receiver decodes both messages and transmitter identities, the first three terms in (27) are preserved, and the penalty $-k\log K$ only affects the constant term $O(1)$ in (27).

When applied to a scenario with $M=1$ and identity decoding, the bound in Theorem 2, modified as described in the preceding paragraph, extends the non-asymptotic achievability bound in the group testing problem [33, Th. 4] to the scenario where an unknown number k out of a total of K items are defective. In the scenario considered in [33], the number of defective items k is known, and our MAC bound (50) with K replaced by k, M replaced by $KM=K$, and the term $\frac{K(K-1)}{2M}$ removed applies. The resulting bound is similar to [33, Th. 4]. The difference is that the bound in (50) uses a single information density threshold rule, while [33, Th. 4] uses 2^k-1 simultaneous information density threshold rules.

E. Per-user Probability of Error

We extend the definition of the PUPE from [6, Def. 1] to the RAC with $k\in[K]$ active transmitters as

$$e_k\triangleq\frac{1}{M^k}\sum_{w_{[k]}\in[M]^k}\sum_{i=1}^{k}\frac{1}{k}\mathbb{P}\left[w_i\notin\mathsf{g}_T(U,Y_k^{n_T})|W_{[k]}=w_{[k]}\right], \tag{116}$$

where $Y_k^{n_T}$ is the received output at time n_T, and

$$T\triangleq\min\{t\in\{0\}\cup[K]\colon\mathsf{g}_t(U,Y_k^{n_t})\neq\mathsf{e}\} \tag{117}$$

is the random variable describing the decoder's estimate of the number of active transmitters.[12] We set $T=K$ if $\mathsf{g}_t(U,Y_k^{n_t})=\mathsf{e}$ for all $t\in\{0\}\cup[K]$. For $k=0$, we define $e_0\triangleq\mathbb{P}\left[\mathsf{g}_0(U,Y_0^{n_0})\neq 0|W_{[0]}=0\right]$ as in (7).

[11] This simple argument was suggested by Dr. Jonathan Scarlett.

[12] Note that the joint error probability in (6) can likewise be written as

$$\frac{1}{M^k}\sum_{w_{[k]}\in[M]^k}\mathbb{P}\left[\mathsf{g}_T(U,Y_k^{n_T})\overset{\pi}{\neq}w_{[k]}\middle|W_{[k]}=w_{[k]}\right].$$

For a RAC with a total of K transmitters and a MAC with K transmitters, the following corollary to Theorem 2 gives non-asymptotic achievability bounds under the PUPE criterion (116).

Corollary 1. *Fix constants γ_0, $\lambda^k_{s,t} \geq 0$, and $\gamma_t > 0$ for all $1 \leq s \leq t \leq k$. For any k and n, let $(X^n_{[k]}, \bar{X}^n_{[k]}, Y^n_k)$ be a random sequence drawn i.i.d. $\sim P_{X_{[k]}\bar{X}_{[k]}Y_k}(x_{[k]}, \bar{x}_{[k]}, y_k) = \left(\prod_{i=1}^k P_X(x_i)P_X(\bar{x}_i)\right) P_{Y_k|X_{[k]}}(y_k|x_{[k]})$.*

A) For any RAC $\left\{\left(\mathcal{X}^k, P_{Y_k|X_{[k]}}(y_k|x_{[k]}), \mathcal{Y}_k\right)\right\}_{k=0}^K$ satisfying (2) and (3), any $K \leq \infty$, and any fixed input distribution P_X, there exists an $(M, \{(n_k, e_k)\}_{k=0}^K)$ RAC code under the PUPE criterion (116) such that

$$e_0 \leq \mathbb{P}\left[h(Y_0^{n_0}) > \gamma_0\right], \tag{118}$$

and for all $k \geq 1$,

$$\begin{aligned}
e_k \leq\ & \mathbb{P}[\imath_k(X^{n_k}_{[k]}; Y^{n_k}_k) \leq \log \gamma_k] && (119a)\\
& + \mathbb{P}\left[h(Y_k^{n_0}) \leq \gamma_0\right] + \frac{k(k-1)}{2M} && (119b)\\
& + \sum_{t=1}^{k-1} \binom{k-1}{t} \mathbb{P}[\imath_t(X^{n_t}_{[t]}; Y^{n_t}_k) > \log \gamma_t] && (119c)\\
& + \sum_{t=1}^{k} \sum_{s=1}^{t-1} \binom{k-1}{t-s} \mathbb{P}\Big[\imath_t(X^{n_t}_{[s+1:t]}; Y^{n_t}_k) \\
& \qquad > n_t \mathbb{E}[\imath_t(X_{[s+1:t]}; Y_k)] + \lambda^k_{s,t}\Big] && (119d)\\
& + \sum_{t=1}^{k} \sum_{s=1}^{t} \binom{k-1}{t-s} \binom{M-k}{s} \\
& \qquad \mathbb{P}\Big[\imath_t(\bar{X}^{n_t}_{[s]}; Y^{n_t}_k | X^{n_t}_{[s+1:t]}) \\
& \qquad > \log \gamma_t - n_t \mathbb{E}[\imath_t(X_{[s+1:t]}; Y_k)] - \lambda^k_{s,t}\Big]. && (119e)
\end{aligned}$$

B) For a MAC with K transmitters satisfying (2), there exists a MAC code for M messages and decoding blocklength n_K such that

$$\begin{aligned}
e_K \leq\ & \mathbb{P}[\imath_K(X^{n_K}_{[K]}; Y^{n_K}_K) \leq \log \gamma_K] + \frac{K(K-1)}{2M} \\
& + \sum_{s=1}^{K-1} \binom{K-1}{K-s} \mathbb{P}\Big[\imath_K(X^{n_K}_{[s+1:K]}; Y^{n_K}_K) \\
& \qquad > n_K \mathbb{E}[\imath_K(X_{[s+1:K]}; Y_K)] + \lambda^K_{s,K}\Big] \\
& + \sum_{s=1}^{K} \binom{K-1}{K-s} \binom{M-K}{s} \\
& \qquad \mathbb{P}\Big[\imath_K(\bar{X}^{n_K}_{[s]}; Y^{n_K}_K | X^{n_K}_{[s+1:K]}) > \log \gamma_K \\
& \qquad - n_K \mathbb{E}[\imath_K(X_{[s+1:K]}; Y_K)] - \lambda^K_{s,K}\Big].
\end{aligned} \tag{120}$$

Proof: Notice that in (119), the only modification from Theorem 2 is the replacement of the coefficients $\binom{k}{t}$ in (47d) and $\binom{k}{t-s}$ in (47e)–(47f) by the coefficients $\binom{k-1}{t}$ and $\binom{k-1}{t-s}$, respectively. To see how Corollary 1 is derived from Theorem 2, observe that the PUPE (116) measures the fraction of transmitted messages missing from the list of decoded messages. Therefore, to bound the PUPE for the RAC, we can multiply the error probability bounds in (47) that correspond to the case where t out of k messages are decoded by $\frac{k-(t-s)}{k}$, where s is the number of messages decoded incorrectly.

Similarly, under the PUPE, the coefficient $\binom{K}{K-s}$ in the K-transmitter MAC bound (50) is replaced by $\binom{K-1}{K-s}$ in (120) since we can multiply the error probability bounds in (50b)–(50c), corresponding to the case where s out of K messages are decoded incorrectly, by $\frac{s}{K}$. ■

From the proof of Theorem 1, the error probability bounds in (119c)–(119e) behave as $O\left(\frac{1}{\sqrt{n_k}}\right)$. This implies that under the PUPE criterion (116), our encoding and decoding scheme described in Section IV-A achieves the same first three order terms as Theorem 1. Only the constant $O(1)$ term in (27) is affected by the change from the joint error probability to the PUPE.

The PUPE criterion becomes critical in applications of the Gaussian RAC with $K \to \infty$, where the energy per bit $(\frac{nP}{2\log_2 M})$ and the number of bits sent by each transmitter $(\log_2 M)$ are fixed as the blocklength n grows, and all K transmitters are active. In [6], Polyanskiy shows that in this regime, the joint error probability goes to 1 as $K \to \infty$. As we saw in (120), the PUPE introduces scaling factors $\frac{s}{K}$ in front of the error terms corresponding to s out of K messages decoded incorrectly, for $s = 1, \dots, K$. In the regime $K \to \infty$, the number of these terms is infinite, and the PUPE can be strictly less than 1 even as the joint error probability approaches 1. In [6], Polyanskiy shows that the PUPE behaves nontrivially in this regime.

VI. Tests for No Active Transmitters

In this section, we give an analysis of the error probabilities of the composite binary hypothesis test that we use to decide between H_0: "no active transmitters," and H_1: "$k \in [K]$ active transmitters;" that is

$$\begin{aligned}
& H_0 : Y^{n_0} \sim P^{n_0}_{Y_0} \\
& H_1 : Y^{n_0} \sim P^{n_0}_{Y_k} \text{ for some } 1 \leq k \leq K.
\end{aligned} \tag{121}$$

In the context of Theorem 2, the maximal number of transmitters, K, can be infinite. In that case, enumerating all alternative possibilities as in (121) becomes infeasible, and a universal (goodness-of-fit) test

$$\begin{aligned}
& H_0 \colon Y^n \sim P^n_{Y_0} \\
& H_1 \colon Y^n \nsim P^n_{Y_0}
\end{aligned} \tag{122}$$

is appropriate.

Following [59], a *test statistic* $h_n \colon \mathcal{Y}^n \mapsto \mathbb{R}$ is a function that maps the observed sequence y^n to a real number used to measure the correspondence between that sequence and the null hypothesis. A (randomized) test corresponding to the test statistic h_n is a binary random variable that depends only on $h_n(Y^n)$. The test is deterministic if it outputs H_0 if $h_n(y^n) \leq \gamma_0$ for some constant γ_0, and H_1 otherwise.

Type-I and type-II errors corresponding to a deterministic test with the statistic h_n are defined as

$$\alpha(h_n) \triangleq P_{Y_0}[h_n(Y^n) > \gamma_0] \tag{123}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2020.3047630, IEEE Transactions on Information Theory

$$\beta(h_n) \triangleq Q[h_n(Y^n) \leq \gamma_0], \tag{124}$$

where Q is the unknown alternative distribution of Y, and γ_0 is a constant determined by the desired error criterion. Throughout the following discussion and in our application of these results in Lemma 5, we employ deterministic tests. For these deterministic tests, we choose γ_0 to ensure that we meet the zero-transmitter error bound $\alpha(h_n) \leq \epsilon_0$, and then we show that $\beta(h_n)$ decays exponentially with n for each Q in $\{P_{Y_1}, \ldots, P_{Y_K}\}$ to ensure (28) in Theorem 1.

In Sections A and B, below, we consider Hoeffding's test and the Kolmogorov-Smirnov test as possible hypothesis tests for recognizing the zero-transmitter scenario. Both tests are universal in the sense that the test statistic does not vary with the alternative output distributions $P_{Y_1}, \ldots, P_{Y_K}$. They both give an exponentially decaying type-II error for a fixed type-I error $\epsilon_0 \in (0, 1)$. The disadvantage of Hoeffding's test is that its traditional form requires the channel output alphabet to be finite for every k (as in the adder-erasure RAC in (23)); the advantage of Hoeffding's test is that it achieves the same exponent as the Neyman-Pearson Lemma, which is optimal for a given collection of output distributions $P_{Y_1}, \ldots, P_{Y_K}$, but is not universal, meaning that a different test statistic is necessary for each collection $\{P_{Y_k} : k \in [K]\}$. In contrast to Hoeffding's test, the Kolmogorov-Smirnov test does not require $\mathcal{Y}$ to be finite; however, when applied to a setting with finite $\mathcal{Y}$, it achieves a type-II error exponent that is inferior to that achieved by Hoeffding's test. In Section VI-C, we compare the performances of these universal test statistics to that of the log-likelihood ratio (LLR) threshold test, which is third-order optimal in terms of the type-II error exponent for composite hypothesis testing [60] and relies explicitly on alternative output distributions $P_{Y_1}, \ldots, P_{Y_K}$.

A. *Hoeffding's Test*

Denote the empirical distribution of an observed sequence $y_1, \ldots, y_n$ by

$$\hat{P}_{y^n}(a) \triangleq \frac{1}{n}\sum_{i=1}^{n} 1\{y_i = a\} \quad \forall\, a \in \mathcal{Y}. \tag{125}$$

Hoeffding's test is based on the relative entropy, denoted by $D(\cdot\|\cdot)$, between $\hat{P}_{y^n}$ and P_{Y_0}, giving the test statistic

$$h_n^H(y^n) = D(\hat{P}_{y^n} \| P_{Y_0}). \tag{126}$$

Note that if P_{Y_0} is a continuous distribution, $h_n^H(y^n) = +\infty$.

Theorem 3 (Hoeffding's test [61])**.** *Let $\mathcal{Y}$ be a finite set, and let Q be an unknown alternative distribution for Y_0. If P_{Y_0} is absolutely continuous with respect to Q, and $P_{Y_0} \neq Q$, then the type-I and type-II errors of Hoeffding's test satisfy*

$$\alpha(h_n^H) \leq \exp\{-n\gamma_0 + O(\log n)\} \tag{127}$$

$$\beta(h_n^H) \leq \exp\left\{-n \inf_{P: D(P\|P_{Y_0}) < \gamma_0} D(P\|Q) + O(\log n)\right\}. \tag{128}$$

In [61], a more restrictive assumption ($P_{Y_0}(y) > 0$ and $Q(y) > 0$ for all $y \in \mathcal{Y}$) is used. Absolute continuity is sufficient according to the proofs given in [59] and [62, Th. 2.3], which both rely on Sanov's theorem. The error exponents of Hoeffding's test coincide with the exponents of the optimal (Neyman-Pearson Lemma) binary hypothesis test. Therefore, Hoeffding's test is asymptotically universally most powerful.

Setting $\gamma_0 = \frac{|\mathcal{Y}|\log n}{n}$ achieves type-I error $\epsilon_0 \to 0$ as $n \to \infty$; therefore, the type-I error condition is satisfied for any $\epsilon_0 > 0$ and sufficiently large n. Under this choice, type-II error $\exp\{-nD(P_{Y_0}\|Q) + o(n)\}$ is achieved (see [62, Th. 2.3]). Therefore, in (77), the maximum type-II error decays with exponent

$$C' = \inf_{k \in [K]} D(P_{Y_0} \| P_{Y_k}) \tag{129}$$

$$\geq 2 \inf_{k \in [K]} \left\{ \left(\sup_{x \in \mathbb{R}} |F_k(x) - F_0(x)| \right)^2 + \frac{4}{9}\left(\sup_{x \in \mathbb{R}} |F_k(x) - F_0(x)| \right)^4 \right\} \tag{130}$$

$$\geq 2\delta_0^2 + \frac{4}{9}\delta_0^4. \tag{131}$$

The inequality in (130) is due to [63, eq. (5)-(6)] and Pinsker's inequality [64]. The inequality in (131) follows from (18).

In [59], Zeitouni and Gutman extend Hoeffding's test to continuous distributions. Their test, which also uses the empirical distribution, employs "δ-smoothing" of the decision regions obtained by a relative entropy comparison. The Zeitouni-Gutman test is optimal under a slightly weaker optimality criterion than the standard first-order type-II error exponent criterion. Using [59, Th. 2], it can be shown that the Zeitouni-Gutman test also yields the desired exponentially decaying maximum type-II error.

B. *Kolmogorov-Smirnov Test*

The Kolmogorov-Smirnov test [65], [66] relies on the empirical CDF

$$\hat{F}^{(n)}(x|y^n) \triangleq \frac{1}{n}\sum_{i=1}^{n} 1\{y_i \leq x\} \quad \forall\, x \in \mathbb{R} \tag{132}$$

of the observed sequence $y_1, \ldots, y_n \in \mathbb{R}$. The Kolmogorov-Smirnov test uses a deterministic test

$$h_n^{KS}(y^n) = \sup_{x \in \mathbb{R}} |\hat{F}^{(n)}(x|y^n) - F_0(x)| \tag{133}$$

to test whether the observed sequence y^n is well-explained by P_{Y_0} with the CDF F_0.

The following theorem bounds the probability that the Kolmogorov-Smirnov statistic exceeds a threshold γ_0.

Theorem 4 (Dvoretzky-Kiefer-Wolfowitz [67], [68])**.** *Let $Y_1, \ldots, Y_n$ be drawn i.i.d. according to an arbitrary distribution P_{Y_0} with the CDF F_0 on $\mathbb{R}$. For any $n \in \mathbb{N}$ and $\gamma_0 > 0$, it holds that*

$$\alpha(h_n^{KS}) \leq 2\exp\{-2n\gamma_0^2\}. \tag{134}$$

In [67], Dvoretzky et al. prove Theorem 4 with an unspecified multiplicative constant C in front of the exponential on the right side of (134). In [68], Massart establishes that $C = 2$.

In our operational regime of interest, we set the type-I error to a given constant ϵ_0, which by Theorem 4 corresponds to setting the threshold γ_0 to

$$\gamma_0 = \sqrt{\frac{\log\frac{2}{\epsilon_0}}{2n}} = O\left(\frac{1}{\sqrt{n}}\right). \tag{135}$$

We next bound the type-II errors for every $k \in [K]$. For each $k \in \{0, \ldots, K\}$, let F_k denote the CDF of P_{Y_k}. The type-II error when $k \geq 1$ transmitters are active is bounded as

$$\begin{align}
\beta_k(h_n^{KS}) &= \mathbb{P}\left[\sup_{x\in\mathbb{R}}|\hat{F}^{(n)}(x|Y_k^n) - F_0(x)| \leq \gamma_0\right] \tag{136}\\
&\leq \mathbb{P}\Bigg[\sup_{x\in\mathbb{R}}\Big(|F_k(x) - F_0(x)| \\
&\qquad - |\hat{F}^{(n)}(x|Y_k^n) - F_k(x)|\Big) \leq \gamma_0\Bigg] \tag{137}\\
&\leq \mathbb{P}\Bigg[\sup_{x\in\mathbb{R}}|\hat{F}^{(n)}(x|Y_k^n) - F_k(x)| \\
&\qquad \geq \sup_{x\in\mathbb{R}}|F_k(x) - F_0(x)| - \gamma_0\Bigg] \tag{138}\\
&\leq 2\exp\Bigg\{-2n\left(\sup_{x\in\mathbb{R}}|F_k(x) - F_0(x)|\right)^2 \\
&\qquad + O(\sqrt{n})\Bigg\}, \tag{139}
\end{align}$$

where (137) follows from triangle inequality $|x+y| \geq |x| - |y|$, and (139) follows from Theorem 4 and (135). Applying (18) to (139), we conclude that the maximum type-II error in (77) decays exponentially with n, with exponent

$$\begin{align}
C' &= 2\inf_{k\in[K]}\left(\sup_{x\in\mathbb{R}}|F_k(x) - F_0(x)|\right)^2 \tag{140}\\
&\geq 2\delta_0^2. \tag{141}
\end{align}$$

Comparing (140) and (130), from (18), we see that the type-II error exponent achieved by the Kolmogorov-Smirnov test is always inferior to that achieved by Hoeffding's test.

C. The Optimal Composite Hypothesis Test

From (131) and (141), we know that there exists a positive constant c_0 such that

$$n_0 \geq c_0 \log n_1 + o(\log n_1) \tag{142}$$

suffices to meet the error requirements of the composite hypothesis test given in (75) and (76). Since the proposed tests are universal, Theorem 2 allows us to decode any message set of $k \leq K$ active transmitters without knowing the total number of transmitters, K. In this section, we find the smallest first three terms on the right side of (142) that we can achieve when K is finite and we allow the composite hypothesis test to depend on the distributions $P_{Y_1}, \ldots, P_{Y_K}$.

Let $\beta_{\epsilon_0}(P_{Y_0}, \{P_{Y_k}\}_{k=1}^K)$ denote the minimax type-II error among the alternative distributions $P_{Y_1}, \ldots, P_{Y_K}$ such that type-I error (under P_{Y_0}) does not exceed ϵ_0; that is,

$$\beta_{\epsilon_0}(P_{Y_0}, \{P_{Y_k}\}_{k=1}^K) \triangleq \min_{h_n : \alpha(h_n) \leq \epsilon_0}\ \max_{k\in[K]} \beta_k(h_n), \tag{143}$$

where the minimum is over all tests including deterministic and randomized tests.

The LLR test statistic $h_n^{\mathrm{LLR}} \colon \mathcal{Y}^n \mapsto \mathbb{R}^K$ is given by

$$h_n^{\mathrm{LLR}}(y^n) = \sum_{i=1}^n h_1^{\mathrm{LLR}}(y_i), \tag{144}$$

where

$$h_1^{\mathrm{LLR}}(y) \triangleq \begin{bmatrix} \log\frac{P_{Y_0}(y)}{P_{Y_1}(y)} \\ \log\frac{P_{Y_0}(y)}{P_{Y_2}(y)} \\ \vdots \\ \log\frac{P_{Y_0}(y)}{P_{Y_K}(y)} \end{bmatrix}. \tag{145}$$

Given a threshold vector $\boldsymbol{\tau} \in \mathbb{R}^K$, the corresponding LLR test outputs H_0 if $h_n^{\mathrm{LLR}}(y^n) \geq \boldsymbol{\tau}$, and H_1 otherwise.

The gap in the type-II error exponent (C' in (77)) between the general optimal tests and the LLR tests with the optimal threshold vector $\boldsymbol{\tau}$ is $O\left(\frac{1}{n}\right)$ [60]; therefore, we only consider minimizing over the LLR tests in (143) for asymptotic optimality.

Denote by $\mathbf{D}$ and V the mean and covariance matrix of the random vector $h_1^{\mathrm{LLR}}(Y_0)$, respectively. Define

$$\begin{align}
D_{\min} &\triangleq \min_{k\in[K]} D(P_{Y_0}\|P_{Y_k}) \tag{146}\\
\mathcal{I}_{\min} &\triangleq \{k \in [K] \colon D(P_{Y_0}\|P_{Y_k}) = D_{\min}\} \tag{147}\\
\mathsf{V}_{\min} &\triangleq \mathsf{Cov}\left[\left(h_1^{\mathrm{LLR}}(Y_0)\right)_{\mathcal{I}_{\min}}\right] \in \mathbb{R}^{|\mathcal{I}_{\min}|\times|\mathcal{I}_{\min}|}. \tag{148}
\end{align}$$

The following theorem gives the asymptotics of the minimax type-II error defined in (143).

Theorem 5. *Assume that P_{Y_0} is absolutely continuous with respect to P_{Y_k}, $0 < D(P_{Y_0}\|P_{Y_k}) < \infty$ for $k = 1, \ldots, K$, V is positive definite, and $T = \mathbb{E}[\|h_1^{\mathrm{LLR}}(Y_0) - \mathbf{D}\|_2^3] < \infty$. Then for any $\epsilon_0 \in (0, 1)$, the asymptotic minimax type-II error satisfies*

$$\begin{align}
\beta_{\epsilon_0}(P_{Y_0}, \{P_{Y_k}\}_{k=1}^K) = \exp\Bigg\{&-nD_{\min} + \sqrt{n}b \\
&- \frac{1}{2}\log n + O(1)\Bigg\}, \tag{149}
\end{align}$$

where b is the solution to

$$\mathbb{P}\left[\mathbf{Z} \leq b\mathbf{1}\right] = 1 - \epsilon_0, \tag{150}$$

for $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{V}_{\min}) \in \mathbb{R}^{|\mathcal{I}_{\min}|}$. Moreover, the minimax error in (149) is achieved by a LLR test with some threshold vector $\boldsymbol{\tau}$.

Proof: See Appendix E. ∎

Rewriting (149), defining b as given in (150), and using the condition in (75) with any fixed E_k, we see that a decision about whether any of the transmitters are active can be made at time

$$\begin{align}
n_0 = \frac{1}{2D_{\min}}\log n_1 &+ \frac{b}{\sqrt{2D_{\min}^3}}\sqrt{\log n_1} \\
&- \frac{1}{2D_{\min}}\log\log n_1 + O(1) \tag{151}
\end{align}$$

while guaranteeing both that the probability that we do not decode at time n_0 when no transmitters are active does not exceed ϵ_0 and that the probability that we decode at time n_0 when $k > 0$ transmitters are active does not exceed $\frac{E_k}{\sqrt{n_k}}$. Note that E_k only affects the constant term $O(1)$ in (151). Theorem 5 implies that the coefficients in front of $\log n_1$, $\sqrt{\log n_1}$, and $\log \log n_1$ in (151) are optimal. Juxtaposing (129) and (151), we see that Hoeffding's test achieves the optimal first-order error exponent (that is, the optimal coefficient in front of $\log n_1$).

VII. Summary and Conclusions

We study the agnostic random access model, in which each transmitter knows nothing about the set of active transmitters beyond what it learns from limited scheduled feedback from the receiver, and the receiver knows nothing about the set of active transmitters beyond what it learns from the channel output. In our proposed rateless coding strategy, the decoder attempts to decode only at a fixed, finite collection of decoding times. At each decoding time n_t, it sends a single bit of feedback to all transmitters indicating whether or not its estimate for the number of active transmitters is t. We prove non-asymptotic and second-order achievability results for the equal rate point $(R, \ldots, R)$ under our assumptions on the channel (permutation-invariance (2), reducibility (3), friendliness (16), and interference (17)). For a nontrivial class of discrete, memoryless RACs, our proposed RAC code performs as well in its capacity and dispersion terms as the best-known code for the discrete memoryless MAC in operation; that is, it performs as well as if the transmitter set were known *a priori*. The assumptions of permutation-invariance (2), reducibility (3), and interference (17) together with our use of identical encoding guarantee (by Lemma 2) that the equal rate point always lies on the sum-rate boundary rather than on one of the corner points. For example, for two users, the capacity region is a symmetric pentagon. This ensures that our simplified, single-threshold decoding rule results in no loss in the first- or second-order achievable rate terms, making the codes far more practical than prior schemes [10]–[13] in which decoders employ $2^k - 1$ simultaneous threshold-rules. In Section V-D, we show that as long as $K < \infty$, there is no loss in the first two terms even if the decoder is tasked with decoding transmitter identity.

We also provide a tight approximation for the capacity and dispersion of the adder-erasure RAC (23), which is an example channel satisfying our symmetry conditions.

In order to decide whether there are any active transmitters without enumerating all K alternative hypotheses, we analyze universal hypothesis tests. Results are given both for the case where the channel output alphabet is finite and the case where the channel output alphabet is countably or uncountably infinite. Using existing literature, it is possible in both cases to obtain exponentially decaying maximum type-II error under the condition that $\sup_{x \in \mathbb{R}} |F_k(x) - F_0(x)| \geq \delta_0 > 0$ for all $k \in [K]$. We also derive the best third-order asymptotics of the minimax type-II error (Theorem 5).

Acknowledgment

We are grateful to the reviewers—Professor Jonathan Scarlett and two anonymous reviewers—for their thorough, careful, and insightful feedback, which is reflected in the paper.

References

[1] M. Effros, V. Kostina, and R. C. Yavas, "Random access channel coding in the finite blocklength regime," in *Proc. 2018 IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, CO, USA, Jun. 2018, pp. 1261–1265.

[2] L. G. Roberts, "ALOHA packet system with and without slots and capture," *SIGCOMM Comput. Commun. Rev.*, vol. 5, no. 2, pp. 28–42, Apr. 1975.

[3] A. G. D'yachkov and V. V. Rykov, "On a coding model for a multiple-access adder channel," *Problemy Peredachi Informatsii*, vol. 17, no. 2, pp. 26–38, 1981.

[4] P. Mathys, "A class of codes for a t active users out of n multiple-access communication system," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1206–1219, Nov. 1990.

[5] L. A. Bassalygo and M. S. Pinsker, "Restricted asynchronous multiple access," *Problemy Peredachi Informatsii*, vol. 19, no. 4, pp. 92–96, 1983.

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

[7] O. Ordentlich and Y. Polyanskiy, "Low complexity schemes for the random access Gaussian channel," in *Proc. 2017 IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2528–2532.

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

[9] P. Minero, M. Franceschetti, and D. N. C. Tse, "Random access: An information-theoretic perspective," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 909–930, Feb. 2012.

[10] Y.-W. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in *Proc. 2012 IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, USA, Jul. 2012, pp. 831–835.

[11] E. MolavianJazi and J. N. Laneman, "Simpler achievable rate regions for multiaccess with finite blocklength," in *Proc. 2012 IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, USA, Jul. 2012, pp. 36–40.

[12] 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.

[13] J. Scarlett, A. Martinez, and A. G. i Fàbregas, "Second-order rate region of constant-composition codes for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 157–172, Jan. 2015.

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

[15] 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.

[16] M. Luby, "LT codes," in *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, Vancouver, BC, Canada, Nov. 2002, pp. 271–280.

[17] A. Tchamkerten and E. Telatar, "A feedback strategy for binary symmetric channels," in *Proc. 2002 IEEE Int. Symp. Inf. Theory (ISIT)*, Lausanne, Switzerland, Jun. 2002, p. 362.

[18] S. C. Draper, B. J. Frey, and F. R. Kschischang, "Efficient variable length channel coding for unknown DMCs," in *Proc. 2004 Int. Symp. Inf. Theory*, Chicago, IL, USA, Jun. 2004, pp. 379–379.

[19] N. Shulman, "Communication over an unknown channel via common broadcasting," Ph.D. dissertation, Tel Aviv University, Jul. 2003.

[20] N. Blits and M. Feder, "Universal rateless coding with finite message set," in *Proc. 2012 IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, USA, Jul. 2012, pp. 1772–1776.

[21] C. Stefanovic and P. Popovski, "Aloha random access that operates as a rateless code," *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4653–4662, Nov. 2013.

[22] D. Blackwell, L. Breiman, and A. Thomasian, "The capacity of a class of channels," *The Annals of Mathematical Statistics*, pp. 1229–1241, 1959.

[23] Y. Polyanskiy, "On dispersion of compound DMCs," in *Proc. 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, USA, Oct. 2013, pp. 26–32.

[24] T. Berger, “The Poisson multiple-access conflict resolution problem,” *Multi-user communication systems*, pp. 1–27, 1981.

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

[26] L. V. Truong and V. Y. Tan, “On the Gaussian MAC with stop-feedback,” in *Proc. 2017 IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2303–2307.

[27] X. Chen and D. Guo, “Many-access channels: the Gaussian case with random user activities,” in *Proc. 2014 IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun. 2014, pp. 3127–3131.

[28] X. Chen, T. Y. Chen, and D. Guo, “Capacity of Gaussian many-access channels,” *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3516–3539, Jun. 2017.

[29] A. D. Sarwate and M. Gastpar, “Some observations on limited feedback for multiaccess channels,” in *Proc. 2009 IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, Jun. 2009, pp. 394–397.

[30] M. B. Malyutov, “The separating property of random matrices,” *Mathematical notes of the Academy of Sciences of the USSR*, vol. 23, no. 1, pp. 84–91, 1978.

[31] M. B. Malyutov and P. S. Mateev, “Planning of screening experiments for a nonsymmetric response function,” *Mathematical notes of the Academy of Sciences of the USSR*, vol. 27, no. 1, pp. 57–68, 1980.

[32] G. K. Atia and V. Saligrama, “Boolean compressed sensing and noisy group testing,” *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1880–1901, Mar. 2012.

[33] J. Scarlett and V. Cevher, “Phase transitions in group testing,” in *Proc. 27th Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA ’16. Arlington, VA, USA: Society for Industrial and Applied Mathematics, Jan. 2016, pp. 40–53.

[34] J. Scarlett and V. Cevher, “Limits on support recovery with probabilistic models: An information-theoretic framework,” *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 593–620, Jan. 2017.

[35] 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.

[36] Y. Polyanskiy, H. V. Poor, and S. Verdu, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[37] 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.

[38] E. Haim, Y. Kochman, and U. Erez, “A note on the dispersion of network problems,” in *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*, Nov. 2012, pp. 1–9.

[39] Y. Watanabe, “The total capacity of two-user multiple-access channel with binary output,” *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1453–1465, Sep. 1996.

[40] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968, vol. 2.

[41] P. Moulin, “The log-volume of optimal constant-composition codes for memoryless channels, within o(1) bits,” in *Proc. 2012 IEEE Int. Symp. Inf. Theory (ISIT)*, Cambridge, MA, USA, Jul. 2012, pp. 826–830.

[42] E. MolavianJazi and J. N. Laneman, “On the second-order cost of TDMA for Gaussian multiple access,” in *Proc. 2014 IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun. 2014, pp. 266–270.

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

[44] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed. John Wiley & Sons, 1971, vol. II.

[45] Y. Polyanskiy and Y. Wu, “Lecture notes on information theory,” Aug. 2017, [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/itlectures_v5.pdf.

[46] Y. Polyanskiy, “Channel coding: non-asymptotic fundamental limits,” Ph.D. dissertation, Princeton University, Nov. 2010.

[47] 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.

[48] Y. Liu and M. Effros, “Finite-blocklength and error-exponent analyses for LDPC codes in point-to-point and multiple access communication,” in *Proc. 2020 IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angeles, CA, USA, Jun. 2020, pp. 361–366.

[49] Y. Liu and M. Effros, “Finite-blocklength and error-exponent analyses for LDPC codes in point-to-point and multiple access communication,” *arXiv Preprints, arxiv/2005.06428*, May 2020.

[50] R. C. Yavas, V. Kostina, and M. Effros, “Gaussian multiple and random access in the finite blocklength regime,” *arXiv Preprints, arXiv/2001.03867*, Jun. 2020.

[51] H. H. J. Liao, “Multiple access channels,” Ph.D. dissertation, University of Hawaii, Honolulu, HI, USA, Sep. 1972.

[52] R. Ahlswede, “Multi-way communication channels,” in *Proc. Int. Symp. Inf. Theory (ISIT)*, Tsahkadsor, Armenia, USSR, Sep. 1971.

[53] G. Dueck, “The strong converse to the coding theorem for the multiple–access channel,” *J. Comb. Inform. Syst. Sci.*, vol. 6, no. 3, pp. 187–196, 1981.

[54] R. Ahlswede, “An elementary proof of the strong converse theorem for the multiple-access channel,” *J. Comb. Inform. Syst. Sci.*, vol. 7, no. 3, 1982.

[55] S. L. Fong and V. Y. F. Tan, “A proof of the strong converse theorem for Gaussian multiple access channels,” *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4376–4394, Aug. 2016.

[56] O. Kosut, “A second-order converse bound for the multiple-access channel via wringing dependence,” *arXiv Preprints, arxiv:2007.15664*, Jul. 2020.

[57] P. Moulin, “A new metaconverse and outer region for finite-blocklength MACs,” in *Proc. 2013 Information Theory and Applications Workshop (ITA)*, San Diego, CA, USA, Feb. 2013, pp. 1–8.

[58] G. Dueck, “Maximal error capacity regions are smaller than average error capacity regions for multi-user channels,” *Problems of Control and Information Theory*, vol. 7, pp. 409–413, 1978.

[59] O. Zeitouni and M. Gutman, “On universal hypotheses testing via large deviations,” *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 285–290, Mar. 1991.

[60] Y. Huang and P. Moulin, “Strong large deviations for composite hypothesis testing,” in *Proc. 2014 IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun. 2014, pp. 556–560.

[61] W. Hoeffding, “Asymptotically optimal tests for multinomial distributions,” *Ann. Math. Statist.*, vol. 36, no. 2, pp. 369–401, Apr. 1965.

[62] I. Csiszár and P. C. Shields, “Information theory and statistics: A tutorial,” *Commun. Inf. Theory*, vol. 1, no. 4, pp. 417–528, Dec. 2004.

[63] A. L. Gibbs and F. E. Su, “On choosing and bounding probability metrics,” *International Statistical Review / Revue Internationale de Statistique*, vol. 70, no. 3, pp. 419–435, 2002.

[64] S. Kullback, “Correction to a lower bound for discrimination information in terms of variation,” *IEEE Trans. Inf. Theory*, vol. 16, no. 5, pp. 652–652, Sep. 1970.

[65] A. N. Kolmogorov, “Sulla Determinazione Empirica di una Legge di Distribuzione,” *Giornale dell’Istituto Italiano degli Attuari*, vol. 4, pp. 83–91, 1933.

[66] N. V. Smirnov, “Approximate laws of distribution of random variables from empirical data,” *Usp. Mat. Nauk*, vol. 10, pp. 179–206, 1944.

[67] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, “Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator,” *Ann. Math. Statist.*, vol. 27, no. 3, pp. 642–669, Sep. 1956.

[68] P. Massart, “The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality,” *Ann. Probab.*, vol. 18, no. 3, pp. 1269–1283, Jul. 1990.

[69] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. Dover, 1972.

[70] L. Shepp and I. Olkin, “Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution,” in *Contributions to Probability*. Academic Press, 1981, pp. 201–206.

[71] H. G. Eggleston, *Convexity*, ser. Cambridge Tracts in Mathematics. Cambridge University Press, 1958.

[72] S. Chen, M. Effros, and V. Kostina, “Lossless source coding in the point-to-point, multiple access, and random access scenarios,” *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6688–6722, Jul. 2020.

Recep Can Yavas (S'19) is currently a Ph.D. candidate in electrical engineering at the California Institute of Technology (Caltech). He received the B.S. degree from Bilkent University in Ankara, Turkey, in 2016 and the M.S. degree from Caltech in 2017, both in electrical engineering. His research interests include information theory and probability theory.

Victoria Kostina (S'12-M'14) received the bachelor's degree from the Moscow institute of Physics and Technology in 2004, the master's degree from the University of Ottawa in 2006, and the Ph.D. degree from Princeton University in 2013. In 2014, she joined Caltech, where she is currently a Professor of electrical engineering. Her research spans information theory, coding, control, learning, and communications. She received the Natural Sciences and Engineering Research Council of Canada master's scholarship in 2009, the Princeton Electrical Engineering Best Dissertation Award in 2013, the Simons-Berkeley Research Fellowship in 2015, and the NSF CAREER Award in 2017.

Michelle Effros (S'93-M'95-SM'03-F'09) received the B.S. degree with distinction in 1989, the M.S. degree in 1990, and the Ph.D. degree in 1994, all in electrical engineering from Stanford University. She joined the faculty at the California Institute of Technology in 1994, where she is currently the George Van Osdol Professor of Electrical Engineering. Her research interests include information theory, network coding, data compression, and communications.

Prof. Effros received Stanford's Frederick Emmons Terman Engineering Scholastic Award (for excellence in engineering) in 1989, the Hughes Masters Full-Study Fellowship in 1989, the National Science Foundation Graduate Fellowship in 1990, the AT&T Ph.D. Scholarship in 1993, the NSF CAREER Award in 1995, the Charles Lee Powell Foundation Award in 1997, the Richard Feynman-Hughes Fellowship in 1997, and an Okawa Research Grant in 2000. She was cited by Technology Review as one of the world's top young innovators in 2002. She and her co-authors received the Communications Society and Information Theory Society Joint Paper Award in 2009. She became a fellow of the IEEE in 2009. She is a member of Tau Beta Pi, Phi Beta Kappa, and Sigma Xi. She 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, serving in the role of President of the Information Theory Society in 2015. 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 2006 joint special issue on Networking and Information Theory in the *IEEE Transactions on Information Theory* and the *IEEE Transactions on Networking/ACM Transactions on Networking* and as Associate Editor for Source Coding for the *IEEE Transactions on Information Theory* from 2004 to 2007. 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.

Appendix A
Proofs of Lemmas 1–3

We first state and prove Lemma 6, which we then use to prove Lemmas 2, 1, and 3 (in that order).

Lemma 6. *Let $X_1, X_2, \ldots, X_k$ be i.i.d., and let the interference (17), permutation-invariance (2), and reducibility (3) assumptions hold. Then $I_k(X_i; Y_k | X_{[i-1]})$ is strictly increasing in i, i.e., for all $i < j \leq k$,*

$$I_k(X_i; Y_k | X_{[i-1]}) < I_k(X_j; Y_k | X_{[j-1]}). \tag{A.1}$$

Proof of Lemma 6: By permutation-invariance (17) and the i.i.d. distribution of $X_1, \ldots, X_k$, we have

$$I_k(X_i; Y_k | X_{[i-1]}) = I_k(X_j; Y_k | X_{[i-1]}). \tag{A.2}$$

Let (U, V, T) be mutually independent random variables. Then $I(U; V) = I(U; T, V) = 0$. Since $I(U; T, Y) \leq I(U; T, V, Y)$, the chain rule implies that

$$I(U; Y | T) \leq I(U; Y | T, V). \tag{A.3}$$

Setting U to X_j, Y to Y_k, T to $X_{[i-1]}$, and V to $X_{[i:j-1]}$ in (A.3) and then applying (A.2) gives (A.1) with $<$ replaced by $\leq$. Equality in (A.3) is attained if and only if U and V are conditionally independent given (Y, T). As a result, equality in our modified form of (A.1) occurs if and only if X_j and $X_{[i:j-1]}$ are conditionally independent given $(Y_k, X_{[i-1]})$. We proceed to show that this is not possible using a proof by contradiction.

Assume that X_j and $X_{[i:j-1]}$ are conditionally independent given $(Y_k, X_{[i-1]})$, i.e.,

$$P_{X_{[i:j]}|Y_k, X_{[i-1]}} = P_{X_{[i:j-1]}|Y_k, X_{[i-1]}} P_{X_j|Y_k, X_{[i-1]}}. \tag{A.4}$$

Set $X_{[i-1]} = 0^{i-1}$ and use Bayes' rule to show

$$P_{X_{[i:j]}|Y_k, X_{[i-1]}=0^{i-1}} = P_{X_{[j-(i-1)]}|Y_{k-(i-1)}} \tag{A.5}$$

$$P_{X_{[i:j-1]}|Y_k, X_{[i-1]}=0^{i-1}} = P_{X_{[2:j-(i-1)]}|Y_{k-(i-1)}} \tag{A.6}$$

$$P_{X_j|Y_k, X_{[i-1]}=0^{i-1}} = P_{X_1|Y_{k-(i-1)}} \tag{A.7}$$

due to reducibility (2), permutation-invariance (3), and the i.i.d. distribution of $X_1, \ldots, X_k$. Therefore, (A.4) implies that X_1 and $X_{[2:j-(i-1)]}$ are conditionally independent given $Y_{k-(i-1)}$, which is not possible by interference assumption (17). ■

Proof of Lemma 2: We wish to show that

$$\frac{1}{k} I_k(X_{[k]}; Y_k) < \frac{1}{s} I_k(X_{[s]}; Y_k | X_{[s+1:k]}). \tag{A.8}$$

By the chain rule for mutual information, the left-hand side of (A.8) equals the average of k terms

$$\frac{1}{k} I_k(X_{[k]}; Y_k) = \frac{1}{k} \sum_{i=1}^{k} I_k(X_i; Y_k | X_{[i-1]}). \tag{A.9}$$

By permutation-invariance (2) and the chain rule, the right-hand side of (A.8) equals the average of the last s of those k terms

$$\frac{1}{s} I_k(X_{[s]}; Y_k | X_{[s+1:k]}) = \frac{1}{s} I_k(X_{[k-s+1:k]}; Y_k | X_{[k-s]}) \tag{A.10}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2020.3047630, IEEE Transactions on Information Theory

$$= \frac{1}{s} \sum_{i=k-s+1}^{k} I_k(X_i; Y_k | X_{[i-1]}). \tag{A.11}$$

Since the terms in these averages are strictly increasing in i by Lemma 6, we have the desired result. ■

Proof of Lemma 1: We wish to show that $\frac{1}{s} I_s > \frac{1}{k} I_k$. We proceed by representing I_s in terms of I_k as

$$\begin{aligned}
\frac{1}{s} I_s &= \frac{1}{s} I_k(X_{[s]}; Y_k | X_{[s+1:k]} = 0^{k-s}) && \text{(A.12)}\\
&\geq \frac{1}{s} I_k(X_{[s]}; Y_k | X_{[s+1:k]}) && \text{(A.13)}\\
&> \frac{1}{k} I_k, && \text{(A.14)}
\end{aligned}$$

where (A.12) follows from reducibility (3), (A.13) follows from friendliness (16), and (A.14) follows from Lemma 2. ■

Proof of Lemma 3: To derive the bound $\mathbb{E}[\imath_t(X_{[s]}; Y_k)] \leq I_k(X_{[s]}; Y_k) < I_t(X_{[s]}; Y_t)$, we write

$$\begin{aligned}
\mathbb{E}[\imath_t(X_{[s]}; Y_k)] &= \mathbb{E}\left[\log \frac{P_{Y_t|X_{[s]}}(Y_k|X_{[s]})}{P_{Y_t}(Y_k)}\right] && \text{(A.15)}\\
&= -D(P_{X_{[s]}} P_{Y_k|X_{[s]}} \| P_{X_{[s]}} P_{Y_t|X_{[s]}})\\
&\quad + D(P_{Y_k} \| P_{Y_t})\\
&\quad + D(P_{X_{[s]}} P_{Y_k|X_{[s]}} \| P_{X_{[s]}} P_{Y_k}) && \text{(A.16)}\\
&= -D(P_{X_{[s]}} P_{Y_k|X_{[s]}} \| P_{X_{[s]}} P_{Y_t|X_{[s]}})\\
&\quad + D(P_{Y_k} \| P_{Y_t}) + I_k(X_{[s]}; Y_k) && \text{(A.17)}\\
&\leq I_k(X_{[s]}; Y_k) && \text{(A.18)}\\
&= \sum_{i=1}^{s} I_k(X_i; Y_k | X_{[i-1]}) && \text{(A.19)}\\
&< \sum_{i=1}^{s} I_k(X_i; Y_k | X_{[i-1]}, X_{[s+1:s+k-t]}) && \text{(A.20)}\\
&= I_k(X_{[s]}; Y_k | X_{[t+1:k]}) && \text{(A.21)}\\
&\leq I_k(X_{[s]}; Y_k | X_{[t+1:k]} = 0^{k-t}) && \text{(A.22)}\\
&= I_t(X_{[s]}; Y_t), && \text{(A.23)}
\end{aligned}$$

where (A.18) follows from data processing inequality of relative entropy (e.g., [45, Th. 2.2.5]), (A.19) follows from the chain rule, (A.20) follows from permutation-invariance (2) and Lemma 6, (A.21) follows from permutation-invariance (2) and the chain rule, and (A.22) and (A.23) follow from friendliness (16) and reducibility (3), respectively. ■

Appendix B
Proof of Lemma 4

To prove Lemma 4, we first derive the saddle point condition for the MAC.

Theorem 6 (Saddle point condition for the MAC). *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be convex set of distributions on alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. Suppose that there exists a product distribution $P_{X_1^*} P_{X_2^*}$ such that*

$$\sup_{\substack{P_{X_1} P_{X_2} \\ P_{X_1} \in \mathcal{P}_1, P_{X_2} \in \mathcal{P}_2}} I_2(X_1, X_2; Y_2) = I_2(X_1^*, X_2^*; Y_2^*) = I_2^*, \tag{B.1}$$

where $P_{Y_2^*|X_1^*, X_2^*} = P_{Y_2|X_1, X_2}$. Then, for all $P_{X_1} \in \mathcal{P}_1$ and for all Q_{Y_2}, it holds that

$$\begin{aligned}
&D(P_{X_1} P_{X_2^*} P_{Y_2|X_1,X_2} \| P_{X_1} P_{X_2^*} P_{Y_2^*})\\
&\leq I_2^* && \text{(B.2)}\\
&\leq D(P_{X_1^*} P_{X_2^*} P_{Y_2|X_1,X_2} \| P_{X_1^*} P_{X_2^*} Q_{Y_2}). && \text{(B.3)}
\end{aligned}$$

Proof of Lemma 4: Lemma 4 follows by an application of Theorem 6 to the setting where $\mathcal{P}_1$ includes the set of all distributions with a singleton on $\mathcal{X}_1$ having probability 1, i.e., $\{\delta_{x_1} : x_1 \in \mathcal{X}_1\} \subseteq \mathcal{P}_1$, and $I_2^* < \infty$. Particularizing P_{X_1} in (B.2) to any $P_{X_1} = \delta_{x_1}$ with $x_1 \in \mathcal{X}_1$ yields

$$D(P_{X_2^*} P_{Y_2|X_1 = x_1, X_2} \| P_{X_2^*} P_{Y_2^*}) \leq I_2^* \tag{B.4}$$

for all $x_1 \in \mathcal{X}_1$. Since the left-hand side of (B.4) is equal to the conditional expectation of $\imath_2(X_1^*, X_2^*; Y_2^*)$ given $X_1^* = x_1$, (35) follows with less than or equal to. The equality in (35) follows since otherwise (B.4) would give the contradiction $I_2(X_1^*, X_2^*; Y_2^*) < I_2^*$. ■

Proof of Theorem 6: The proof of Theorem 6 is similar to the proof of the saddle point condition for point-to-point channels in [45, Th. 4.4] and extends [45, Th. 4.4] to the MAC. Although the optimization in (B.1) is not convex in general [39], the optimization

$$\sup_{P_{X_1} \in \mathcal{P}_1} I_2(X_1, X_2^*; Y_2), \tag{B.5}$$

where $P_{X_1 X_2^* Y_2} = P_{X_1} P_{X_2^*} P_{Y_2|X_1, X_2}$ is convex.

Inequality (B.3) follows from the golden formula (e.g., [45, Th. 3.3])

$$\begin{aligned}
I_2^* &= D(P_{X_1^*} P_{X_2^*} P_{Y_2|X_1,X_2} \| P_{X_1^*} P_{X_2^*} P_{Y_2^*}) && \text{(B.6)}\\
&= D(P_{X_1^*} P_{X_2^*} P_{Y_2|X_1,X_2} \| P_{X_1^*} P_{X_2^*} Q_{Y_2}) - D(P_{Y_2^*} \| Q_{Y_2}) && \text{(B.7)}
\end{aligned}$$

and the nonnegativity of the relative entropy. Notice that for $I_2^* = \infty$, (B.2) is trivial. Assume that $I_2^* < \infty$. Fix any $P_{X_1} \in \mathcal{P}_1$. Let $\lambda \in (0, 1)$. Set

$$P_{X_{1\lambda}} = \lambda P_{X_1} + (1 - \lambda) P_{X_1^*} \in \mathcal{P}_1. \tag{B.8}$$

Let $\theta \sim \text{Bernoulli}(\lambda)$, so that $P_{X_{1\lambda}|\theta=0} = P_{X_1^*}$ and $P_{X_{1\lambda}|\theta=1} = P_{X_1}$, and let

$$P_{X_{1\lambda} X_2^* Y_{2\lambda}} = P_{X_{1\lambda}} P_{X_2^*} P_{Y_2|X_1, X_2}. \tag{B.9}$$

Then

$$\begin{aligned}
I_2^* &\geq I_2(X_{1\lambda}, X_2^*; Y_{2\lambda}) && \text{(B.10)}\\
&= D(P_{X_{1\lambda}} P_{X_2^*} P_{Y_2|X_1,X_2} \| P_{X_{1\lambda}} P_{X_2^*} P_{Y_{2\lambda}}) && \text{(B.11)}\\
&= \lambda D(P_{X_1} P_{X_2^*} P_{Y_2|X_1,X_2} \| P_{X_1} P_{X_2^*} P_{Y_{2\lambda}})\\
&\quad + (1 - \lambda) D(P_{X_1^*} P_{X_2^*} P_{Y_2|X_1,X_2} \| P_{X_1^*} P_{X_2^*} P_{Y_{2\lambda}}) && \text{(B.12)}\\
&\geq \lambda D(P_{X_1} P_{X_2^*} P_{Y_2|X_1,X_2} \| P_{X_1} P_{X_2^*} P_{Y_{2\lambda}})\\
&\quad + (1 - \lambda) I_2^*, && \text{(B.13)}
\end{aligned}$$

where (B.13) follows from (B.3). By subtracting $(1 - \lambda) I_2^*$ from both sides of (B.13) and dividing by λ, we get

$$I_2^* \geq D(P_{X_1} P_{X_2^*} P_{Y_2|X_1,X_2} \| P_{X_1} P_{X_2^*} P_{Y_{2\lambda}}). \tag{B.14}$$

By taking $\liminf_{\lambda \to 0}$ in (B.14) and applying the lower semi-continuity of the relative entropy (e.g., [45, Th. 3.6]), (B.2) is proved. ■

Note that $(P_{X_1^*}, P_{X_2^*})$ does not have to be unique for Theorem 6 and Lemma 4 to hold.

Appendix C
Adder-erasure RAC

Here, we approximate the sum-capacity and dispersion of the adder-erasure RAC for a large number of transmitters (k).

Theorem 7. *The optimal input distribution for the adder-erasure RAC defined in (23) is the Bernoulli(1/2) distribution at all encoders. That input distribution achieves the sum-rate capacity, and*

$$I_k = (1-\delta)\left(\frac{1}{2}\log\frac{\pi e k}{2} - \frac{\log e}{12k^2}\right) + O(k^{-3}) \tag{C.1}$$

$$V_k = (1-\delta)\Bigg[\frac{\delta}{4}\log^2\frac{\pi e k}{2} + \frac{\log^2 e}{2} - \frac{\log^2 e}{2k} - \left(\frac{\log e}{2} + \frac{\delta \log\frac{\pi e k}{2}}{12}\right)\frac{\log e}{k^2}\Bigg] + O\left(\frac{\log k}{k^3}\right). \tag{C.2}$$

The calculation leading to Theorem 7 is presented in Lemmas 7–8, which rely on Stirling's approximation and the Taylor series expansion.

Consider a binomial random variable $X \sim \mathrm{Binom}(n, 1/2)$. Lemma 7, below, shows that the probability mass that this Binomial distribution puts at k is well approximated by

$$\tilde{P}_X(k) \triangleq \frac{1}{\sqrt{\frac{\pi n}{2}}} e^{-\frac{(k-\frac{n}{2})^2}{\frac{n}{2}}}\left(1 + \frac{f(k)}{n} + \frac{g(k)}{n^2}\right), \tag{C.3}$$

where

$$f(x) \triangleq -\frac{1}{12}\frac{(2x-n)^4}{n^2} + \frac{1}{2}\frac{(2x-n)^2}{n} - \frac{1}{4} \tag{C.4}$$

$$g(x) \triangleq \frac{1}{288}\frac{(2x-n)^8}{n^4} - \frac{3}{40}\frac{(2x-n)^6}{n^3} + \frac{19}{48}\frac{(2x-n)^4}{n^2} - \frac{11}{24}\frac{(2x-n)^2}{n} + \frac{1}{32}. \tag{C.5}$$

Define the interval

$$\mathcal{K} \triangleq \left[\frac{n}{2} - \frac{A}{2}\sqrt{n\log n},\ \frac{n}{2} + \frac{A}{2}\sqrt{n \log n}\right] \tag{C.6}$$

for some constant $A > 0$.

Lemma 7. *Let* $X \sim \mathrm{Binom}(n, 1/2)$*. Then for any* $k \in \mathcal{K}$*,*

$$P_X(k) = \binom{n}{k}2^{-n} = \tilde{P}_X(k)\left(1 + O\left(\frac{\log^6 n}{n^3}\right)\right). \tag{C.7}$$

Proof of Lemma 7: We apply Stirling's approximation [69, eq. (6.1.37)]

$$n! = \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}\left(1 + \frac{1}{12n} + \frac{1}{288n^2} + O(n^{-3})\right), \tag{C.8}$$

and a Taylor series expansion of $\binom{n}{k}$ around $x = 0$, where

$$k = \frac{n}{2} + \frac{x}{2}\sqrt{n\log n}, \tag{C.9}$$

to $P_X(k) = \binom{n}{k}2^{-n}$, to derive (C.7). ■

Let $V(X)$

$$V(X) = \mathrm{Var}\left[\log\frac{1}{P_X(X)}\right]. \tag{C.10}$$

denote the varentropy of X.

Lemma 8 (Entropy and varentropy of $\mathrm{Binom}(n, 1/2)$)**.** *For* $X \sim \mathrm{Binom}(n, 1/2)$,

$$H(X) = \frac{1}{2}\log\frac{\pi e n}{2} - \frac{\log e}{12n^2} + O(n^{-3}) \tag{C.11}$$

$$V(X) = \frac{\log^2 e}{2} - \frac{\log^2 e}{2n} - \frac{\log^2 e}{2n^2} + O(n^{-3}). \tag{C.12}$$

Proof of Lemma 8: Let $\tilde{T}(k)$ denote the first 3 terms of the Taylor series expansion of $\log\frac{1}{\tilde{P}_X(k)}$ around $\frac{n}{2}$ evaluated at k, giving

$$\tilde{T}(k) \triangleq \frac{1}{2}\log\frac{\pi n}{2} + \log e\Bigg(\frac{(k-\frac{n}{2})^2}{\frac{n}{2}} - \frac{f(k)}{n} + \frac{-g(k) + \frac{f^2(k)}{2}}{n^2}\Bigg). \tag{C.13}$$

Recall the definition of interval $\mathcal{K}$ from (C.6). Then we can write the entropy $H(X)$ as

$$H(X) = \sum_{k=0}^{n}\frac{\binom{n}{k}}{2^n}\log\left(\frac{2^n}{\binom{n}{k}}\right) \tag{C.14}$$

$$\begin{aligned} &= \mathbb{E}\left[\tilde{T}(X)\right] \\ &\quad + \mathbb{E}\left[\left(\log\frac{1}{P_X(X)} - \tilde{T}(X)\right)1\{X \in \mathcal{K}\}\right] \\ &\quad + \mathbb{E}\left[\left(\log\frac{1}{P_X(X)} - \tilde{T}(X)\right)1\{X \notin \mathcal{K}\}\right]. \end{aligned} \tag{C.15}$$

Using the moments of $\mathrm{Binom}(n, 1/2)$ (e.g., [69, eq. (26.1.20)]), the first term in (C.15) is

$$\mathbb{E}\left[\tilde{T}(X)\right] = \frac{1}{2}\log\frac{\pi e n}{2} - \frac{\log e}{12n^2}. \tag{C.16}$$

By Lemma 7, the second term in (C.15) is

$$\mathbb{E}\left[\left(\log\frac{1}{P_X(X)} - \tilde{T}(X)\right)1\{X \in \mathcal{K}\}\right] = O\left(\frac{\log^6 n}{n^3}\right). \tag{C.17}$$

By Hoeffding's inequality,

$$\mathbb{P}\left[X \notin \mathcal{K}\right] \leq 2n^{-\frac{A^2 \log e}{2}}, \tag{C.18}$$

where A is the constant in (C.6). Since the minimum of $P_X(k)$ over k is achieved at $k = n$, using (C.18), we get

$$\mathbb{E}\left[\log\frac{1}{P_X(X)}1\{X \notin \mathcal{K}\}\right] = O\left(\frac{\log^6 n}{n^3}\right) \tag{C.19}$$

for $A \geq \frac{3}{\sqrt{\log e}}$. Similarly, by taking the derivative of $\tilde{T}(k)$, one can show that $\tilde{T}(k) \leq \tilde{T}(n) \leq n$ for all $k \in [0, n]$, which gives

$$\mathbb{E}\left[\tilde{T}(X)1\{X \notin \mathcal{K}\}\right] = O\left(\frac{\log^6 n}{n^3}\right). \tag{C.20}$$

Combining (C.15)–(C.17), (C.19)–(C.20) gives

$$H(X) = \frac{1}{2}\log\frac{\pi e n}{2} - \frac{\log e}{12n^2} + O\left(\frac{\log^6 n}{n^3}\right). \tag{C.21}$$

Via an argument similar to (C.19) and (C.20), we can show that for $A \geq \frac{4}{\sqrt{\log e}}$, the contribution of $k \notin \mathcal{K}$ to the varentropy is $O\left(\frac{\log^6 n}{n^3}\right)$. Therefore, using the moments of $\mathrm{Binom}(n, 1/2)$ and Lemma 7, we can approximate the varentropy $V(X)$ as

$$\begin{aligned} V(X) &= \mathbb{E}\left[\log^2\frac{1}{P_X(X)}\right] - (H(X))^2 && \text{(C.22)}\\ &= \mathbb{E}\left[(\tilde{T}(X))^2\right] - (H(X))^2 + O\left(\frac{\log^6 n}{n^3}\right) && \text{(C.23)}\\ &= \log^2 e\left(\frac{1}{2} - \frac{1}{2n} - \frac{1}{2n^2}\right) + O\left(\frac{\log^6 n}{n^3}\right). && \text{(C.24)}\end{aligned}$$

The above analyses use the first 3 terms of the Stirling series (C.8) to obtain the remainder $O\left(\frac{\log^6 n}{n^3}\right)$. Applying the same analyses with 4 terms of the Stirling series improves the remainder to $O(n^{-3})$, as claimed in (C.11) and (C.12) in the statement of Lemma 8. ■

We are now equipped to prove Theorem 7.

Proof of Theorem 7: Define

$$E \triangleq 1\{Y = \mathsf{e}\}. \tag{C.25}$$

By the chain rule for entropy, we have for the adder-erasure RAC

$$\begin{aligned} I_k(X_{[k]}; Y_k) &= H(Y_k) - H(Y_k|X_{[k]}) && \text{(C.26)}\\ &= H(Y_k, E) - H(E) && \text{(C.27)}\\ &= H(Y_k|E) && \text{(C.28)}\\ &= (1-\delta)H(Y_k|E=0). && \text{(C.29)}\end{aligned}$$

Given the independent inputs $X_i \sim \mathrm{Bernoulli}(p_i)$ for $i \in [k]$, $H(Y_k|E=0)$ is equal to the entropy of the sum of k independent Bernoulli random variables with parameters $(p_1, \ldots, p_k)$, which is maximized when $p_i = 1/2$ for all i [70]. Therefore, for any $\delta \in [0,1]$, the equiprobable input distribution at all encoders, $X_i^* \sim \mathrm{Bernoulli}(1/2)$, maximizes the mutual information $I_k(X_{[k]}; Y_k)$ for all k. Let $(X_{[k]}^* Y_k^*) \sim P_{X_{[k]}^*} P_{Y_k|X_{[k]}}$. Then

$$I_k(X_{[k]}^*; Y_k^*) = (1-\delta)H(Z), \tag{C.30}$$

where $Z \sim \mathrm{Binom}(k, 1/2)$, and (C.1) follows from Lemma 8. Furthermore,

$$\imath_k(X_{[k]}^*; Y_k^*) = \begin{cases} 0 & \text{w.p. } \delta \\ \log\frac{2^k}{\binom{k}{i}} & \text{w.p. } (1-\delta)\frac{\binom{k}{i}}{2^k}, \quad 0 \leq i \leq k, \end{cases} \tag{C.31}$$

which gives

$$V_k = \mathrm{Var}\left[\imath_k(X_{[k]}^*; Y_k^*)\right] = (1-\delta)\left[V(Z) + \delta(H(Z))^2\right], \tag{C.32}$$

and (C.2) follows from Lemma 8. ■

Appendix D
Bound on the Cardinality $|\mathcal{U}|$

While the analysis in Section IV-B employs common randomness U with $|\mathcal{U}| = |\mathcal{X}|^{Mn_K}$, [15, Th. 19] shows that $|\mathcal{U}| \leq K+2$ suffices to achieve the optimal performance. Theorem 8, stated next, improves the cardinality bound on $|\mathcal{U}|$ from $K+2$ [15, Th. 19] to $K+1$ by using the connectedness of the set of achievable error vectors defined in (D.1).

Theorem 8. *If an $(M, \{(n_k, \epsilon_k)\}_{k=0}^K)$ RAC code exists, then there exists an $(M, \{(n_k, \epsilon_k)\}_{k=0}^K)$ RAC code with $|\mathcal{U}| \leq K+1$.*

Proof of Theorem 8: For fixed $M, n_0, \ldots, n_K$, let G_u denote the set of achievable error vectors compatible with message size M, blocklengths $n_0, \ldots, n_K$, and cardinality $|\mathcal{U}| \leq u$; that is,

$$\begin{aligned} G_u = \{(\epsilon_0', \ldots, \epsilon_K') : \exists (M, \{(n_k, \epsilon_k')\}_{k=0}^K) \text{ code with} \\ |\mathcal{U}| \leq u\}. \end{aligned} \tag{D.1}$$

Let G denote the set of achievable error vectors compatible with message size M and blocklengths $n_0, \ldots, n_K$; that is,

$$G = \{(\epsilon_0', \ldots, \epsilon_K') : \exists (M, \{(n_k, \epsilon_k')\}_{k=0}^K) \text{ code}\}. \tag{D.2}$$

As observed in [15, Proof of Th. 19], $G = G_{|\mathcal{X}|^{Mn_K}}$ is the convex hull of G_1. Indeed, every vector $(\epsilon_0', \ldots, \epsilon_K')$ in G is a convex combination of vectors in G_1, and the coefficients of the convex combination are determined by the distribution of the common randomness random variable U.

Furthermore, G_1 is a connected set. To see this, take any $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2 \in G_1$. For any $\boldsymbol{\epsilon}' \geq \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \in G_1$, the line segments $L_i = \{\lambda\boldsymbol{\epsilon}_i + (1-\lambda)\mathbf{1} : \lambda \in [0,1]\}$, $i = 1, 2$, also belong to G_1, and the path $L_1 \cup L_2$ connects $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$. Therefore, G_1 is a connected set.

Since $G = \mathrm{conv}(G_1) \subset \mathbb{R}^{K+1}$, and G_1 is a connected set, by Fenchel-Eggleston-Carathéodory's theorem [71, Th. 18 (ii)], $G = G_{K+1}$ holds. Therefore, $(\epsilon_0, \ldots, \epsilon_K) \in G$ implies that $(\epsilon_0, \ldots, \epsilon_K) \in G_{K+1}$. ■

Appendix E
Composite Hypothesis Testing

We begin with a lemma that is used in the proof of Theorem 5. See Fig. 4 for an illustration of Lemma 9.

Lemma 9. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuous function that satisfies coordinate-wise partial ordering, i.e., $f(\mathbf{x}) \leq f(\mathbf{y})$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ with $\mathbf{x} \leq \mathbf{y}$. Then for any a in the image of f (denoted $a \in \mathrm{Im} f$), it holds that*

$$b^\star = \min_{\mathbf{b} \in \mathbb{R}^d : f(\mathbf{b}) \geq a} \max_{1 \leq j \leq d} b_j = \min_{x \in \mathbb{R} : f(x\mathbf{1}) \geq a} x. \tag{E.1}$$

Proof: Since $a \in \mathrm{Im} f$, there exists some $\mathbf{b} \in \mathbb{R}^d$ such that $f(\mathbf{b}) = a$. Denote by $b_{\min}$ and $b_{\max}$ the minimum and maximum components of $\mathbf{b}$, respectively. Since f is nondecreasing,

$$f(b_{\min}\mathbf{1}) \leq a = f(\mathbf{b}) \leq f(b_{\max}\mathbf{1}). \tag{E.2}$$

Therefore, since the function mapping b to $f(b\mathbf{1})$ is continuous and nondecreasing, by the intermediate value theorem there

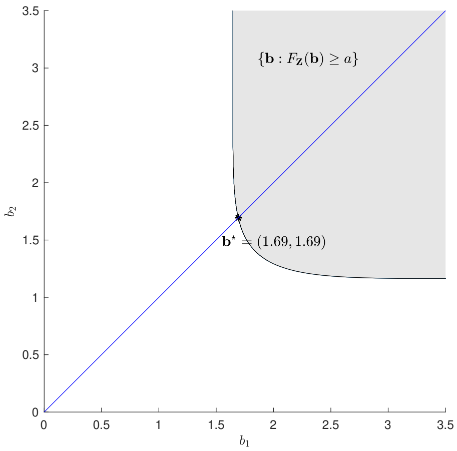


Fig. 4: An example to illustrate Lemma 9. Here $f(\mathbf{b}) = F_{\mathbf{Z}}(\mathbf{b})$ is the CDF of $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{V})$, where $\mathsf{V} = \left[\begin{smallmatrix} 1 & 0.4 \\ 0.4 & 0.5 \end{smallmatrix}\right]$. The shaded region illustrates the set $\{\mathbf{b} \in \mathbb{R}^2 : f(\mathbf{b}) \geq a = 0.95\}$. Lemma 9 shows that the minimax on this set is achieved at a point described by a scalar multiple of $\mathbf{1}$. For this example, the optimizer is $\mathbf{b}^\star = (1.69, 1.69)$.

exists some $b \leq b_{\max}$ such that $f(b\mathbf{1}) = a$. Equation (E.1) follows. ■

Let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{V})$. Define the multidimensional counterpart of the function $Q^{-1}(\cdot)$ as

$$\mathcal{Q}_{\text{inv}}(\mathsf{V}, \epsilon) \triangleq \left\{\mathbf{z} \in \mathbb{R}^K : \mathbb{P}\left[\mathbf{Z} \leq \mathbf{z}\right] \geq 1 - \epsilon\right\}. \tag{E.3}$$

Proof of Theorem 5: For any $\epsilon_0 \in (0, 1)$, consider all composite hypothesis tests in the form given in (121) that achieve type-I error no greater than ϵ_0. Let

$$\mathcal{E}_{\epsilon_0}(P_{Y_0}, \{P_{Y_k}\}_{k=1}^K) \triangleq \Big\{(e_1, \ldots, e_K) : \exists \text{ a (randomized) test}$$

such that

$$\mathbb{P}\left[\text{Decide } H_1 | H_0\right] \leq \epsilon_0,$$
$$\mathbb{P}\left[\text{Decide } H_0 | H_1\right] = e_k, 1 \leq k \leq K\Big\} \tag{E.4}$$

denote the set of type-II errors achievable by these tests. Huang and Moulin [60, Th. 1][13] show that the asymptotic form of the error region defined in (E.4) is given by

$$\mathcal{E}_{\epsilon_0}(P_{Y_0}, \{P_{Y_k}\}_{k=1}^K)$$
$$= \exp\left\{-n\mathbf{D} + \sqrt{n}\mathcal{Q}_{\text{inv}}(\mathsf{V}, \epsilon_0) - \frac{1}{2}\log n\mathbf{1} + O(1)\mathbf{1}\right\}. \tag{E.5}$$

By the definition of the minimax error (143) and the characterization of the achievable error region asymptotics in (E.5), we have

$$\beta_{\epsilon_0}(P_{Y_0}, \{P_{Y_k}\}_{k=1}^K)$$
$$= \min_{\mathbf{z} \in \exp\{-n\mathbf{D} + \sqrt{n}\mathcal{Q}_{\text{inv}}(\mathsf{V}, \epsilon_0) - \frac{1}{2}\log n\mathbf{1} + O(1)\mathbf{1}\}} \max_{1 \leq k \leq K} z_k. \tag{E.6}$$

Applying Lemma 9 with $f(\mathbf{z}) = \mathbb{P}\left[-n\mathbf{D} + \sqrt{n}\mathbf{Z} \leq \mathbf{z}\right]$ and $a = 1 - \epsilon_0$, where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{V})$, we obtain

$$\beta_{\epsilon_0}(P_{Y_0}, \{P_{Y_k}\}_{k=1}^K)$$
$$= \min_{z \in \mathbb{R} : f(z\mathbf{1}) \geq 1 - \epsilon_0} \exp\left\{z - \frac{1}{2}\log n + O(1)\right\}. \tag{E.7}$$

Since $f(z\mathbf{1})$ is nondecreasing and continuous in z,

$$f(z^\star \mathbf{1}) = 1 - \epsilon_0 \tag{E.8}$$

holds, where $z^\star$ is the argument that achieves the minimum on the right-hand side of (E.7). Recall the definitions of $D_{\min}$ and $\mathcal{I}_{\min}$ from (146)–(147). By Chernoff's bound on $f(\mathbf{z})$, for any $z = nE + o(n)$ with $E > -D_{\min}$, we have $f(z\mathbf{1}) = 1 - o(1)$. Similarly, for $E < -D_{\min}$, we have $f(z\mathbf{1}) = o(1)$, giving

$$z^\star = -nD_{\min} + o(n). \tag{E.9}$$

We proceed to show that the minimum on the right-hand side of (E.7) is achieved at

$$z^\star = -nD_{\min} + \sqrt{n}b + O\left(1\right), \tag{E.10}$$

where b is defined in (150). Here

$$\mathbb{P}\left[-nD_{\min}\mathbf{1} + \sqrt{n}\mathbf{Z}_{\mathcal{I}_{\min}} \leq z^\star \mathbf{1}\right]$$
$$= \mathbb{P}\left[-n\mathbf{D} + \sqrt{n}\mathbf{Z} \leq z^\star \mathbf{1}\right]$$
$$+\mathbb{P}\Big[\{-nD_{\min}\mathbf{1} + \sqrt{n}\mathbf{Z}_{\mathcal{I}_{\min}} \leq z^\star \mathbf{1}\}$$
$$\bigcap \{-n\mathbf{D}_{\mathcal{I}^c_{\min}} + \sqrt{n}\mathbf{Z}_{\mathcal{I}^c_{\min}} \not\leq z^\star \mathbf{1}\}\Big] \tag{E.11}$$
$$= 1 - \epsilon_0 + O\left(\frac{1}{n}\right), \tag{E.12}$$

where (E.12) follows from (E.8), (E.9), and the union bound and Chebyshev's inequality on $\mathbb{P}\left[-n\mathbf{D}_{\mathcal{I}^c_{\min}} + \mathbf{Z}_{\mathcal{I}^c_{\min}} \not\leq z^\star \mathbf{1}\right]$. By the Taylor series expansion of $\mathcal{Q}_{\text{inv}}(\mathsf{V}, \cdot)$, we conclude that

$$\mathbb{P}\left[\mathbf{Z}_{\mathcal{I}_{\min}} \leq \frac{1}{\sqrt{n}}(z^\star + nD_{\min})\mathbf{1} + O\left(\frac{1}{n}\right)\right] = 1 - \epsilon_0, \tag{E.13}$$

which implies (E.10). Combining (E.7) and (E.10) completes the proof. ■

[13] In the converse part of the proof of [60, Th. 1], Huang and Moulin show that for any LLR test (144) with threshold vector $\boldsymbol{\tau}$ such that the type-I error is bounded by ϵ_0, it holds that $\boldsymbol{\tau} = n\mathbf{D} - \sqrt{n}\mathbf{b} + O(1)\mathbf{1}$ for some $\mathbf{b} \in \mathcal{Q}_{\text{inv}}(\mathsf{V}, \epsilon_0)$. Then, it is assumed that $\mathbf{b} = O(1)\mathbf{1}$, and [60, Lemma 2] is applied. However, according to the definition of $\mathcal{Q}_{\text{inv}}(\mathsf{V}, \epsilon_0)$ in (E.3), $\mathbf{b}$ can have coordinates growing with n, which violates this assumption. In [72, Th. 11], Chen et al. confirm that the asymptotic expansion in (E.5) holds. They prove the converse part of the expansion (E.5) by evaluating a converse bound that they derive in [72, Lemma 9] for the composite hypothesis testing.