Gaussian Multiple and Random Access Channels: Finite-Blocklength Analysis

Recep Can Yavas, Victoria Kostina, and Michelle Effros

***Abstract*—This paper presents finite-blocklength achievability bounds for the Gaussian multiple access channel (MAC) and random access channel (RAC) under average-error and maximal-power constraints. Using random codewords uniformly distributed on a sphere and a maximum likelihood decoder, the derived MAC bound on each transmitter's rate matches the MolavianJazi-Laneman bound (2015) in its first- and second-order terms, improving the remaining terms to $\frac{1}{2}\frac{\log n}{n} + O\left(\frac{1}{n}\right)$ bits per channel use. The result then extends to a RAC model in which neither the encoders nor the decoder knows which of K possible transmitters are active. In the proposed rateless coding strategy, decoding occurs at a time n_t that depends on the decoder's estimate t of the number of active transmitters k. Single-bit feedback from the decoder to all encoders at each potential decoding time n_i, $i \leq t$, informs the encoders when to stop transmitting. For this RAC model, the proposed code achieves the same first-, second-, and third-order performance as the best known result for the Gaussian MAC in operation.**

***Index Terms*—Gaussian multiple access channel, Gaussian random access channel, third-order asymptotics, finite blocklength, maximum likelihood decoder, dispersion.**

I. Introduction

Emerging communication systems such as the Internet of Things, wireless cellular networks, and machine-to-machine communication systems impose two significant requirements on the code design: low latency constraints and random activity in a large number of communicating devices. These constraints lead us to study random access channels in the finite blocklength regime, where an unknown number of transmitters is active, and communication delay is finite. Current random access strategies mostly use either orthogonalization (TDMA, FDMA, and CDMA) or collision avoidance (e.g., slotted ALOHA). Orthogonalization methods divide up resources (e.g., time, frequency, or signal space) among the transmitters. In slotted ALOHA, each transmitter randomly chooses a time slot to transmit its message, and the decoder declares an error if two or more transmitters are active in a time slot. Performance of these methods is inferior to the information-theoretic bounds achieved through simultaneous resource use. For example, slotted ALOHA achieves only 37% of the single-transmitter capacity [2].

We consider a communication scenario where K transmitters are communicating with a single receiver through a Gaussian channel. We study two problems in this network: multiple access and random access communication. In the multiple access problem, the identity of active transmitters is known to all transmitters and to the receiver. In the random access problem, the set of active transmitters is unknown to the transmitters and to the receiver.

For $K = 1$, Shannon's 1948 paper [3] derives the capacity

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

using codewords with symbols drawn independently and identically distributed (i.i.d.) according to the Gaussian distribution with variance $P - \delta$ for a very small value δ; here P is the maximal (per-codeword) power constraint, and the noise variance is 1. In [4], Shannon shows the performance improvement in the achievable reliability function using codewords drawn uniformly at random on an n-dimensional sphere of radius $\sqrt{nP}$ and a maximum likelihood decoder. Tan and Tomamichel [5] use the same distribution and decoder to prove the achievability of a maximal rate of

$$C(P) - \sqrt{\frac{V(P)}{n}}Q^{-1}(\epsilon) + \frac{1}{2}\frac{\log n}{n} + O\left(\frac{1}{n}\right) \tag{2}$$

at blocklength n, average error probability ϵ, and maximal power P, where

$$V(P) = \frac{P(P+2)}{2(1+P)^2} \tag{3}$$

is the *dispersion* of the point-to-point Gaussian channel; Polyanskiy *et al.* prove a matching converse in [6]. Under a maximal-error constraint, the first- and second-order terms in (2) remain the same under both maximal- and average-power constraints across codewords; under an average-error constraint, average- and maximal-power constraints yield different first- and second-order performance limits [7, Ch. 4]. In this paper, we only consider average-error and maximal-power constraints.

MolavianJazi and Laneman [8] and Scarlett *et al.* [9] generalize the asymptotic expansion in (2) to the two-transmitter Gaussian MAC, bounding the achievable rate as a function of the 3×3 dispersion matrix $\mathsf{V}(P_1, P_2)$, an analogue of $V(P)$ assuming transmitters with per-codeword power constraints P_1 and P_2. The bound in [8] uses codewords uniformly distributed on the power sphere and threshold decoding based on the *information density*; the bound in [9] uses constant composition codes and a quantization argument for the Gaussian channel. This paper improves those bounds using codewords uniformly distributed on the power sphere and a maximum likelihood decoding rule.

R. C. Yavas, V. Kostina, and M. Effros are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA (e-mail: ryavas, vkostina, effros@caltech.edu). This work was supported in part by the National Science Foundation (NSF) under grant CCF-1817241. A part of this work was presented at the 2020 International Symposium on Information Theory (ISIT'20) [1].

The literature on RAC communications includes works like [10]–[12], where the number of active transmitters is known, and [13], where neither the transmitters nor the receiver knows the number of active transmitters. In [13], Ordentlich and Polyanskiy propose a concatenated code with a linear inner code that detects the number of active users and an outer code that decodes their messages. A two-layer code for joint erasure correction and collision resolution appears in [14].

Recently, RACs with massive numbers of users have attracted significant attention. The Gaussian "many access" channel, with a total number of users, K, that grows with the blocklength, n, as $K = O(n)$, is considered in [13], [15], [16]. Chen and Guo [15] find the capacity of the Gaussian many access channel, and Chen *et al.* [16] derive the capacity of the Gaussian many access channel in a random access scenario where the number of users, K, is unknown. For the criterion of average per-user error probability, Polyanskiy [17] and Zadik *et al.* [18] derive non-asymptotic random coding achievability bounds when K transmitters are active. Extensions of these ideas to quasi-static fading MACs and RACs appear in [19] and [20], respectively. Unlike [13], [15]–[20], we here assume that K does not grow with n.

In [21], we develop a communication strategy for a general RAC where neither the transmitters nor the receiver knows the set of active transmitters. A central result of that work shows that for permutation-invariant RACs, under mild conditions it is possible to achieve performance identical in the first- and second-order terms to the best performance known to be achievable for the underlying MAC. These results are obtained using a rateless coding scheme where decoding occurs at one of a fixed collection of possible decoding times $n_0, \ldots, n_K$, and K is the maximal number of transmitters. The chosen decoding time n_t depends on the receiver's estimate t of the number of active transmitters. At each decoding time, the receiver makes an attempt to decode by applying a single threshold rule; the receiver sends a single bit of feedback to all transmitters in order to specify when communication is completed. In [22], Liu and Effros achieve improved third-order bounds using a maximum likelihood decoder under both i.i.d. random code and random LDPC code constructions. Although the coding strategies proposed in [21], [22] apply to the Gaussian RAC, the random encoder design in [21] uses an i.i.d. input distribution. As shown in [23], this codeword distribution guarantees performance strictly inferior to that obtained when blocklength-n codewords are uniformly distributed on the n-dimensional sphere of radius $\sqrt{nP}$.

Motivated by the desire to build superior RAC codes for Gaussian channels, we here propose a new code design for the Gaussian RAC. In the proposed code design, random codewords are designed by concatenating K partial codewords of blocklengths $n_1, n_2 - n_1, \ldots, n_K - n_{K-1}$, each drawn from a uniform distribution on a sphere of radius $\sqrt{(n_i - n_{i-1})P}$. When k transmitters are active, the resulting codewords are uniformly distributed on a restricted subset of the sphere of radius $\sqrt{n_k P}$. The receiver uses output typicality to determine the number of transmitters and then applies a maximum likelihood decoding rule. Despite the restricted subset of codewords that result from our design, we achieve the same first-, second-, and third-order performance as the MAC code. While this paper focuses on Gaussian channels with maximal-power and average-error constraints, we note that the ideas developed here may be useful beyond this example channel and communication scenario.

The use of codes with multiple possible decoding times arises in a variety of communication scenarios. Codes that allow arbitrary decoding times are called rateless or variable-length codes. In [24], Burnashev finds the optimal error exponent of variable-length feedback (VLF) codes for discrete memoryless channels. Extending [24] to the finite blocklength regime, Polyanskiy *et al.* [25] formalize VLF and variable-length feedback with termination (VLFT) codes for discrete memoryless channels. They show that the second-order term in the asymptotic expansion of the maximum achievable message size for VLF and VLFT codes is $O(\log n)$, which means that the dispersion is zero. Truong and Tan [26] extend the zero-dispersion result for VLF and VLFT codes to the Gaussian point-to-point channel under an average-power constraint. In [27], Truong and Tan study VLF and VLFT codes for the Gaussian MAC. Building upon the results in the present paper, in [28] we derive an achievability bound for the Gaussian point-to-point channel with a maximal-power constraint, an average decoding delay constraint, and at most K decoding times. The coding schemes in [25]–[28] utilize so-called 'stop-feedback' [25], i.e., feedback that is only used to direct the encoder to stop transmitting. While VLF and VLFT codes in [24]–[28] employ multiple decoding times for a fixed number of transmitters, the Gaussian RAC codes studied in the present paper allow a single decoding time n_k to decode messages from k active transmitters.

The proposed coding problem is also related to the sparse recovery problem, which is a RAC problem in which each transmitter has only a single message to transmit; this message conveys the encoder's identity, and the decoder is tasked only with determining the identities of active transmitters. The group testing problem is a special case of the sparse recovery problem where the MAC output applies the logical OR operation to all inputs. For example, [29]–[32] study a group testing problem where an unknown subset of k out of K items is defective, and the decoder uses item signatures to identify which items are defective with an average probability of error approaching to zero. In the scenario where $k = O(1)$, Atia and Saligrama [30] show that the number of measurements (i.e., the blocklength n) required to identify k out of K items behaves as $O(k \log \frac{K}{k})$. In [31], Scarlett and Cevher extend this result to the scenario where $k = O(K^\theta)$ and $\theta \in (0, 1)$; in [32], they study a general channel model, which also covers the Gaussian MAC. In [31], [32], $2^k - 1$ information density threshold tests are used at the decoder, and a fixed number of defective items is considered. In the present paper, we combine a maximum likelihood decoding rule with a single threshold test based on the received power to decode messages from an unknown number of active transmitters.

The organization of the paper is as follows. In Section II, we define notation. The system model, main result, and discussions for the Gaussian MAC and Gaussian RAC appear

in Sections III and IV. The proofs of the achievability bounds for the two-transmitter Gaussian MAC, the K-transmitter Gaussian MAC, and the Gaussian RAC appear in Sections V, VI, and VII–VIII. Section IX concludes the paper.

II. Notation

We use bold symbols to denote vectors (e.g., $\mathbf{x}$). For any integer $k \geq 1$, we define $[k] \triangleq \{1, \ldots, k\}$. For any set $\mathcal{A}$, we denote by $\mathcal{P}(\mathcal{A}) \triangleq \{\mathcal{S} \subseteq \mathcal{A}, \mathcal{S} \neq \emptyset\}$ the set of non-empty subsets of $\mathcal{A}$. For any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\mathcal{N} \subseteq [n]$, $\mathbf{x}^{\mathcal{N}} = (x_i \colon i \in \mathcal{N})$ denotes the sub-vector of $\mathbf{x}$ with components in $\mathcal{N}$. For vectors $\mathbf{x}_1, \ldots, \mathbf{x}_K$ of the same dimension and index set $\mathcal{S} \in \mathcal{P}([K])$, $\mathbf{x}_{\mathcal{S}} = (\mathbf{x}_s \colon s \in \mathcal{S})$. and $\mathbf{x}_{\langle \mathcal{S} \rangle} \triangleq \sum_{s \in \mathcal{S}} \mathbf{x}_s$. Our notation for vectors and their collections is summarized in Table I, below.

For vectors $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, we write $\mathbf{x} \overset{\pi}{=} \mathbf{y}$ if there exists a permutation π of the elements of $\mathbf{y}$ such that $\mathbf{x} = \pi(\mathbf{y})$, and $\mathbf{x} \overset{\pi}{\neq} \mathbf{y}$ if $\mathbf{x} \neq \pi(\mathbf{y})$ for all permutations π of the elements of $\mathbf{y}$. We denote the inner product of $\mathbf{x}$ and $\mathbf{y}$ by $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i$ and the Euclidean norm of $\mathbf{x}$ by $\|\mathbf{x}\| \triangleq \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Vector inequalities are understood element-wise, i.e., $\mathbf{x} \leq \mathbf{y}$ if and only if $x_i \leq y_i$ for all $i \in [n]$. All-zero and all-one vectors are denoted by $\mathbf{0}$ and $\mathbf{1}$, respectively.

Matrices are denoted by sans serif font (e.g., A). The $n \times n$ identity matrix is denoted by I_n. Logarithms and exponents are base e. The indicator function is denoted by $1\{\cdot\}$. For any scalar function $f(\cdot)$ and any vector $\mathbf{x} \in \mathbb{R}^n$, we form the vector of function values $f(\mathbf{x}) = (f(x_i) \colon i \in [n])$. For a set $\mathcal{D} \subseteq \mathbb{R}^n$, a vector $\mathbf{c} \in \mathbb{R}^n$, and a scalar a, $a\mathcal{D} + c \triangleq \{a\mathbf{x} + \mathbf{c} : \mathbf{x} \in \mathcal{D}\}$. The sphere with dimension n, radius r, and center at the origin is denoted by $\mathbb{S}^n(r) \triangleq \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = r\}$.

The distribution of a random variable X is denoted by P_X. We write $P_X \to P_{Y|X} \to P_Y$ to indicate that P_Y is the marginal distribution of $P_X P_{Y|X}$. To indicate that the random variables (or vectors) X and Y are identically distributed, we write $X \sim Y$. The multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$. We employ the complementary Gaussian cumulative distribution function $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left\{-\frac{t^2}{2}\right\} dt$. The functional inverse of $Q(\cdot)$ is denoted by $Q^{-1}(\cdot)$.

We use big-O notation $f(n) = O(g(n))$ if and only if there exist constants c and n_0 such that $|f(n)| \leq c|g(n)|$ for all $n > n_0$; we use little-o notation $f(n) = o(g(n))$ if and only if for every $\epsilon > 0$, there exists a constant n_0 such that $|f(n)| \leq \epsilon |g(n)|$ for all $n > n_0$.

TABLE I
Vector Notation Example

Notation	Description
$\mathbf{x}_s = (x_{s,1}, \ldots, x_{s,n})$	The length-n vector that is a member of a collection indexed by $s \in \mathcal{S}$
$\mathbf{x}_{\mathcal{S}} = (\mathbf{x}_s \colon s \in \mathcal{S})$	The size-$\lvert\mathcal{S}\rvert$ ordered collection of length-n vectors
$\mathbf{x}_{\mathcal{S}}^{\mathcal{N}} = ((x_{s,t} \colon t \in \mathcal{N}) \colon s \in \mathcal{S})$	The size-$\lvert\mathcal{S}\rvert$ ordered collection of length-$\lvert\mathcal{N}\rvert$ vectors with time indices in $\mathcal{N} \subseteq [n]$
$\mathbf{x}_{\langle \mathcal{S} \rangle} = \sum_{s \in \mathcal{S}} \mathbf{x}_s$	Summation of length-n vectors from the collection $\mathcal{S}$

III. An RCU Bound and its Analysis for the Gaussian MAC

A. An RCU Bound for General MACs

We begin by defining a two-transmitter MAC code.

Definition 1: An (M_1, M_2, ϵ)-MAC code for the channel with transition law $P_{Y_2|X_1X_2}$ consists of two encoding functions $\mathsf{f}_1 \colon [M_1] \to \mathcal{X}_1$ and $\mathsf{f}_2 \colon [M_2] \to \mathcal{X}_2$ and a decoding function $\mathsf{g} \colon \mathcal{Y}_2 \to [M_1] \times [M_2]$ such that

$$\frac{1}{M_1 M_2} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \mathbb{P}\big[\mathsf{g}(Y_2) \neq (m_1, m_2) \mid (X_1, X_2) = (\mathsf{f}_1(m_1), \mathsf{f}_2(m_2))\big] \leq \epsilon, \quad (4)$$

where Y_2 is the channel output under inputs X_1 and X_2, and ϵ is the average-error constraint.

We define the information densities for a MAC with channel transition law $P_{Y_2|X_1X_2}$ as

$$\imath_1(x_1; y|x_2) \triangleq \log \frac{P_{Y_2|X_1X_2}(y|x_1, x_2)}{P_{Y_2|X_2}(y|x_2)} \quad (5a)$$

$$\imath_2(x_2; y|x_1) \triangleq \log \frac{P_{Y_2|X_1X_2}(y|x_1, x_2)}{P_{Y_2|X_1}(y|x_1)} \quad (5b)$$

$$\imath_{1,2}(x_1, x_2; y) \triangleq \log \frac{P_{Y_2|X_1X_2}(y|x_1, x_2)}{P_{Y_2}(y)}, \quad (5c)$$

where P_{X_1} and P_{X_2} are the channel input distributions, and $P_{X_1}P_{X_2} \to P_{Y_2|X_1X_2} \to P_{Y_2}$. The information density random vector is defined as

$$\boldsymbol{\imath}_2 \triangleq \begin{bmatrix} \imath_1(X_1; Y_2|X_2) \\ \imath_2(X_2; Y_2|X_1) \\ \imath_{1,2}(X_1, X_2; Y_2) \end{bmatrix}, \quad (6)$$

where (X_1, X_2, Y_2) is distributed according to $P_{X_1}P_{X_2}P_{Y_2|X_1X_2}$.

Theorem 1, below, generalizes Polyanskiy *et al.*'s random-coding union (RCU) achievability bound [6, Th. 16] to the MAC. Theorem 1 is derived by Liu and Effros [22] in their work on LDPC codes and is inspired by a new RCU bound for the Slepian-Wolf setting [33, Th. 2]. Its proof combines random code design and a maximum likelihood decoder, which decodes to the message pair with the maximum information density $\imath_{1,2}(X_1, X_2; Y_2)$. Our main result on the Gaussian MAC, Theorem 2, below, analyzes the RCU bound with P_{X_1} and P_{X_2} uniform on the power spheres.

Theorem 1 (RCU bound for the MAC [22, Th. 6]): Fix input distributions P_{X_1} and P_{X_2}. Let

$$P_{X_1,\bar{X}_1,X_2,\bar{X}_2,Y_2}(x_1, \bar{x}_1, x_2, \bar{x}_2, y) = P_{X_1}(x_1)P_{X_1}(\bar{x}_1)P_{X_2}(x_2)P_{X_2}(\bar{x}_2)P_{Y_2|X_1X_2}(y|x_1, x_2). \quad (7)$$

There exists an (M_1, M_2, ϵ)-MAC code for $P_{Y_2|X_1X_2}$ such that

$$\begin{aligned} \epsilon \leq \mathbb{E}\Big[\min\Big\{1, \; & (M_1 - 1)\mathbb{P}\big[\imath_1(\bar{X}_1; Y_2|X_2) \geq \imath_1(X_1; Y_2|X_2) \mid X_1, X_2, Y_2\big] \\ & + (M_2 - 1)\mathbb{P}\big[\imath_2(\bar{X}_2; Y_2|X_1) \geq \imath_2(X_2; Y_2|X_1) \mid X_1, X_2, Y_2\big] \\ & + (M_1 - 1)(M_2 - 1)\mathbb{P}\big[\imath_{1,2}(\bar{X}_1, \bar{X}_2; Y_2) \geq \imath_{1,2}(X_1, X_2; Y_2) \\ & \mid X_1, X_2, Y_2\big]\Big\}\Big]. \end{aligned} \quad (8)$$

Remark 1: As noted in [34], Theorem 1 generalizes to the K-transmitter MAC. Define the conditional information densities for the K-transmitter MAC as

$$\imath_{\mathcal{S}}(x_{\mathcal{S}}; y|x_{\mathcal{S}^c}) \triangleq \log \frac{P_{Y_K|X_{[K]}}(y|x_{[K]})}{P_{Y_K|X_{\mathcal{S}^c}}(y|x_{\mathcal{S}^c})}, \tag{9}$$

where $\mathcal{S} \subset [K]$, $\mathcal{S} \neq \emptyset$, and $\mathcal{S}^c = [K] \setminus \mathcal{S}$, and the unconditional information density as

$$\imath_{[K]}(x_{[K]}; y) \triangleq \log \frac{P_{Y_K|X_{[K]}}(y|x_{[K]})}{P_{Y_K}(y)}. \tag{10}$$

Following arguments identical to those in the proof of Theorem 1, inequality (8) extends to the K-transmitter MAC as

$$\epsilon \leq \mathbb{E}\Bigg[\min\Bigg\{1, \sum_{\mathcal{S}\in\mathcal{P}([K])} \Bigg(\prod_{s\in\mathcal{S}}(M_s-1)\Bigg)\mathbb{P}\big[\imath_{\mathcal{S}}(\bar{X}_{\mathcal{S}}; Y_K|X_{\mathcal{S}^c}) \geq \imath_{\mathcal{S}}(X_{\mathcal{S}}; Y_K|X_{\mathcal{S}^c}) \mid X_{[K]}, Y_K\big]\Bigg\}\Bigg]. \tag{11}$$

B. A Third-order Achievability Bound for the Gaussian MAC

We begin by modifying our code definition to incorporate maximal-power constraints (P_1, P_2) on the channel inputs. Let $(\mathbf{X}_1, \mathbf{X}_2)$ and $\mathbf{Y}_2$ be the MAC inputs and output, respectively.

Definition 2: An $(n, M_1, M_2, \epsilon, P_1, P_2)$-MAC code for a two-transmitter MAC comprises encoding functions $\mathsf{f}_1\colon [M_1] \to \mathbb{R}^n$ and $\mathsf{f}_2\colon [M_2] \to \mathbb{R}^n$, and a decoding function $\mathsf{g}\colon \mathbb{R}^n \to [M_1] \times [M_2]$ such that

$$\|\mathsf{f}_i(m_i)\|^2 \leq nP_i \quad \forall i \in \{1,2\},\ m_i \in [M_i]$$

$$\frac{1}{M_1M_2}\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2} \mathbb{P}\left[\mathsf{g}(\mathbf{Y}_2) \neq (m_1, m_2) \mid (\mathbf{X}_1, \mathbf{X}_2) = (\mathsf{f}_1(m_1), \mathsf{f}_2(m_2))\right] \leq \epsilon.$$

The following notation is used in presenting our achievability result for the Gaussian MAC with $k \geq 1$ transmitters. Over n channel uses, the channel has inputs $\mathbf{X}_1, \ldots, \mathbf{X}_k \in \mathbb{R}^n$, additive noise $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{I}_n)$, and output

$$\mathbf{Y}_k = \mathbf{X}_{\langle [k] \rangle} + \mathbf{Z}. \tag{12}$$

The channel transition law induced by (12) can be written as

$$P_{\mathbf{Y}_k|\mathbf{X}_{[k]}}(\mathbf{y}|\mathbf{x}_{[k]}) = \prod_{i=1}^{n} P_{Y_k|X_{[k]}}(y_i|x_{1i}, \ldots, x_{ki}), \tag{13}$$

where

$$P_{Y_k|X_{[k]}}(y|x_{[k]}) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\big(y - x_{\langle [k] \rangle}\big)^2}{2}\right\}. \tag{14}$$

When $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{V})$, and V is a $d \times d$ positive semi-definite matrix, the multidimensional analogue of the inverse $Q^{-1}(\cdot)$ of the complementary Gaussian cumulative distribution is

$$\mathcal{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) \triangleq \left\{\mathbf{z} \in \mathbb{R}^d : \mathbb{P}\left[\mathbf{Z} \leq \mathbf{z}\right] \geq 1 - \epsilon\right\}. \tag{15}$$

For $d = 1$, we have $Q^{-1}(\epsilon) = \min\{z\colon z \in Q_{\mathrm{inv}}(1, \epsilon)\}$.

Recall that $C(P)$ is the capacity function (1). The capacity vector for the two-transmitter Gaussian MAC is defined as

$$\mathbf{C}(P_1, P_2) \triangleq \begin{bmatrix} C(P_1) \\ C(P_2) \\ C(P_1 + P_2) \end{bmatrix}. \tag{16}$$

The dispersion matrix [8, eq. (25)] for the two-transmitter Gaussian MAC is defined as

$$\mathsf{V}(P_1, P_2) \triangleq \begin{bmatrix} V(P_1) & V_{1,2}(P_1, P_2) & V_{1,12}(P_1, P_2) \\ V_{1,2}(P_1, P_2) & V(P_2) & V_{2,12}(P_1, P_2) \\ V_{1,12}(P_1, P_2) & V_{2,12}(P_1, P_2) & V_{12}(P_1, P_2) \end{bmatrix}, \tag{17}$$

where $V(P)$ is the dispersion function (3), and

$$V_{1,2}(P_1, P_2) = \frac{1}{2}\frac{P_1P_2}{(1+P_1)(1+P_2)} \tag{18}$$

$$V_{i,12}(P_1, P_2) = \frac{1}{2}\frac{P_i(2+P_1+P_2)}{(1+P_i)(1+P_1+P_2)}, \quad i \in \{1,2\} \tag{19}$$

$$V_{12}(P_1, P_2) = V(P_1+P_2) + \frac{P_1P_2}{(1+P_1+P_2)^2}. \tag{20}$$

The following theorem is the main result of this section.

Theorem 2: For any $\epsilon \in (0,1)$ and any $P_1, P_2 > 0$, an $(n, M_1, M_2, \epsilon, P_1, P_2)$-MAC code for the two-transmitter Gaussian MAC exists provided that

$$\begin{bmatrix} \log M_1 \\ \log M_2 \\ \log M_1M_2 \end{bmatrix} \in n\mathbf{C}(P_1, P_2) - \sqrt{n}Q_{\mathrm{inv}}(\mathsf{V}(P_1, P_2), \epsilon) + \frac{1}{2}\log n\mathbf{1} + O(1)\mathbf{1}. \tag{21}$$

Proof: See Section V. ∎

Theorem 2 extends to the general K-transmitter Gaussian MAC. The definition of an $(n, M_{[K]}, \epsilon, P_{[K]})$-MAC code for the K-transmitter Gaussian MAC with message set sizes $M_1, \ldots, M_K$ and power constraints $P_1, \ldots, P_K$ is a natural extension of Definition 2, which defines the two-transmitter MAC code. The following theorem bounds the achievable region for the K-transmitter Gaussian MAC.

Theorem 3: For any $\epsilon \in (0,1)$, and $P_i > 0$, $i \in [K]$, an $(n, M_{[K]}, \epsilon, P_{[K]})$-MAC code for the K-transmitter Gaussian MAC exists provided that

$$\left(\sum_{s\in\mathcal{S}} \log M_s \colon \mathcal{S} \in \mathcal{P}([K])\right) \in n\mathbf{C}(P_{[K]}) - \sqrt{n}Q_{\mathrm{inv}}(\mathsf{V}(P_{[K]}), \epsilon) + \frac{1}{2}\log n\mathbf{1} + O\left(1\right)\mathbf{1}, \tag{22}$$

where $\mathbf{C}(P_{[K]})$ is the capacity vector

$$\mathbf{C}(P_{[K]}) \triangleq \big(C(P_{\langle \mathcal{S} \rangle}) : \mathcal{S} \in \mathcal{P}([K])\big) \in \mathbb{R}^{2^K-1}, \tag{23}$$

and $\mathsf{V}(P_{[K]})$ is the $\left(2^K-1\right) \times \left(2^K-1\right)$ dispersion matrix with the elements $\mathsf{V}_{\mathcal{S}_1,\mathcal{S}_2}(P_{[K]})$, $\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{P}([K])$, given by

$$\mathsf{V}_{\mathcal{S}_1,\mathcal{S}_2}(P_{[K]}) \triangleq \frac{P_{\langle \mathcal{S}_1 \rangle}P_{\langle \mathcal{S}_2 \rangle} + 2P_{\langle \mathcal{S}_1 \cap \mathcal{S}_2 \rangle} + \left(P_{\langle \mathcal{S}_1 \cap \mathcal{S}_2 \rangle}\right)^2 - P^2_{\langle \mathcal{S}_1 \cap \mathcal{S}_2 \rangle}}{2(1+P_{\langle \mathcal{S}_1 \rangle})(1+P_{\langle \mathcal{S}_2 \rangle})}. \tag{24}$$

Proof: See Section VI. ∎

Before concluding this section, we make several remarks about Theorems 2 and 3 above.

1) Theorems 2 and 3 apply the RCU bound (Theorem 1) with independent inputs uniformly distributed on the n-dimensional origin-centered spheres with radii $\sqrt{nP_i}$, $i \in$

$[K]$. Theorem 2 matches the first- and second-order terms of MolavianJazi and Laneman [8] and Scarlett *et al.* [9], and improves the third-order term from $O\left(n^{1/4}\right)\mathbf{1}$ in [8] and $O\left(n^{1/4}\log n\right)\mathbf{1}$ in [9] to $\frac{1}{2}\log n\mathbf{1}+O(1)\mathbf{1}$.

2) Our proof technique in Theorem 2 differs from the technique in [8] in two key ways. First, we use a maximum likelihood decoder in place of the set of simultaneous threshold rules based on unconditional and conditional information densities from [8]; the change of the decoding rule is essential for obtaining the third-order term $\frac{1}{2}\log n\mathbf{1}+O(1)\mathbf{1}$ in Theorem 2. Second, we refine the analysis bounding the probability that the information density random vector $\boldsymbol{\imath}_2$ belongs to a set $\mathcal{D}\subseteq\mathbb{R}^3$. Our non-i.i.d. input distribution prevents direct application of the Berry-Esseen theorem. However, given that the inner product of the inputs $\langle\mathbf{X}_1,\mathbf{X}_2\rangle$ equals a constant, the information density random vector $\boldsymbol{\imath}_2$ can be written as a sum of independent random vectors. Therefore, we apply the Berry-Esseen theorem after conditioning on the inner product $\langle\mathbf{X}_1,\mathbf{X}_2\rangle$ and then integrate the resulting probabilities over the range of the inner product. In order to approximate the resulting probability by the probability that a Gaussian vector belongs to the same set, we use a result (Lemma 5 in Section V-A, below) that approximates the normalized inner product $\frac{1}{\sqrt{nP_1P_2}}\langle\mathbf{X}_1,\mathbf{X}_2\rangle$ by a standard Gaussian random variable. We then derive a bound (Lemma 4 in Section V-A, below) on the total variation distance between two Gaussian vectors. This analysis appears in Section V-F.

   This approach contrasts with [8], which bounds the probability that the information density random vector $\boldsymbol{\imath}_2$ belongs to a set $\mathcal{D}$. Writing $\boldsymbol{\imath}_2$ as a vector-valued function of an average of i.i.d. Gaussian vectors, [8, Prop. 1] applies a central limit theorem for functions of sums to prove $O\left(\frac{1}{n^{1/4}}\right)$ convergence to normality. Our technique, described above, improves the rate of convergence to normality to $O\left(\frac{1}{\sqrt{n}}\right)$, which is the rate of convergence for i.i.d. sums. This improvement implies that the threshold-based decoding rule in [8] achieves a third-order term $O(1)\mathbf{1}$.

3) Our technique for proving Theorems 2 and 3 parallels those used for non-singular discrete memoryless channels [7, Th. 53] and for the point-to-point Gaussian channel [5]. In [7, Th. 53], Polyanskiy applies the RCU bound using a refined large deviations result [6, Lemma 47]; the use of a non-i.i.d. input distribution for the Gaussian channel prevents the direct application of [6, Lemma 47]. In [5, eq. (52)], Tan and Tomamichel derive an alternative to [6, Lemma 47] for the point-to-point Gaussian channel in order to accommodate codewords drawn uniformly on an n-dimensional sphere. While evaluating the RCU bound in this paper, we extend the bound in [5, eq. (52)] to the Gaussian MAC.

4) For the symmetric setting, where $P_i=P$ and $M_i=M$ for all $i\in[K]$, Theorem 3 reduces to the scalar inequality, below. This result refines the result in [8, Th. 2] to the third-order term and generalizes it to the K-transmitter MAC.

   Corollary 1: For any $\epsilon\in(0,1)$ and $P>0$, an $(n,M\mathbf{1},\epsilon,P\mathbf{1})$-MAC code for the K-transmitter Gaussian MAC exists provided that

$$\begin{aligned}K\log M &\le nC(KP)\\ &-\sqrt{n(V(KP)+V_{\text{cr}}(K,P))}Q^{-1}(\epsilon)+\frac{1}{2}\log n+O(1).\end{aligned}\tag{25}$$

   Again, $C(\cdot)$ and $V(\cdot)$ are the capacity (1) and dispersion (3) functions, respectively, and $V_{\text{cr}}(K,P)$ is the cross dispersion term

$$V_{\text{cr}}(K,P)\triangleq\frac{K(K-1)P^2}{2(1+KP)^2}.\tag{26}$$

   Proof: See Appendix A. ∎

5) In [35], Fong and Tan derive a converse for the Gaussian MAC with a second-order term $O(\sqrt{n\log n})\mathbf{1}$. Kosut [36] improves the second-order term in the converse to $O\left(\sqrt{n}\right)\mathbf{1}$. The coefficients of the second-order term in [36] do not match the second-order term in the achievability bounds proven in Theorem 2. As discussed in [21], closing the gap between the second-order terms of the MAC achievability and converse results is a challenging open problem.

IV. A Nonasymptotic Bound and its Analysis for the Gaussian Random Access Channel

A. System Model

Channel model: Given an unknown set of active transmitters $\mathcal{A}$, the Gaussian channel (14) depends on $\mathcal{A}$ only through the number of active transmitters, $|\mathcal{A}|=k$, that is, $P_{Y_k|X_{\mathcal{A}}}=P_{Y_k|X_{[k]}}$. Therefore, in order to capture the scenario of a memoryless Gaussian channel with K possible transmitters, a single receiver, and an unknown activity pattern $\mathcal{A}\subseteq[K]$, we describe the Gaussian RAC by a family of Gaussian MACs $\{P_{Y_k|X_{[k]}}\}_{k=0}^{K}$ (14), each indexed by the number of active transmitters $k\in\{0,\ldots,K\}$. This RAC model is introduced in [21] for general channels satisfying permutation-invariance and reducibility assumptions; the Gaussian RAC satisfies these assumptions. As in [21], we choose a *compound channel* model in order to avoid the need to assign probabilities to each activity pattern $\mathcal{A}$.

Communication strategy: We apply the epoch-based *rateless* communication strategy that we proposed in [21]. Each transmitter is either active or silent during a whole epoch. At each of times $n_0,n_1,\ldots$, the decoder broadcasts to all transmitters a single bit — sending value 1 if it can decode and 0 otherwise. The transmission of 1 at time n_t ends the current epoch and starts the next, indicating that the decoder's estimate of the number of transmitters is t. As in [17], [21], we employ identical encoding, with each active transmitter i using the same encoding function to describe its message $W_i\in[M]$. Identical encoding here requires $P_i=P$ and $M_i=M$ for all i. The task of the decoder is to decode a list of messages sent by the active transmitters $\mathcal{A}$ but not the identities of those transmitters. The messages in $W_{\mathcal{A}}$ are independent and uniformly distributed on alphabet $[M]$.

Since encoding is identical and the channel is invariant to permutation of its inputs, we assume without loss of generality that $|\mathcal{A}| = k$ implies $\mathcal{A} = [k]$. Intuitively, given identical encoding and our Gaussian channel, one would expect that interference increases with the number of active transmitters k, and therefore that the decoding time n_k increases with k. Since the capacity per transmitter for the k-transmitter Gaussian MAC, $\frac{1}{k}C(kP)$, decreases with k, we can choose $n_0 < \cdots < n_K$ for M large enough. (See [21, Lemma 1] for more general sufficient conditions under which $n_0 < \cdots < n_K$ is optimal.) As a notational convenience, we use n_K to represent the largest decoding time. Unless it stops the encoders' transmissions earlier, at time n_K, the decoder sees

$$\mathbf{Y}_k = \mathbf{X}_{\langle [k] \rangle} + \mathbf{Z} \in \mathbb{R}^{n_K} \quad \text{for } k \in [K], \tag{27}$$

where $\mathbf{X}_1, \ldots, \mathbf{X}_k$ are n_K-dimensional channel inputs, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{I}_{n_K})$ is the Gaussian noise, and $\mathbf{Y}_k$ is the n_K-dimensional output when k transmitters are active. When no transmitters are active, $\mathbf{X}_{\langle [0] \rangle} = \mathbf{0}$ and $\mathbf{Y}_0 = \mathbf{Z}$. At each time $n_t < n_K$, the decoder has access to the first n_t components of vector $\mathbf{Y}_k$, which is denoted by $\mathbf{Y}_k^{[n_t]}$.

As in [21], we assume an *agnostic* random access model, where the transmitters know nothing about the set $\mathcal{A}$ of active transmitters except their own membership and the feedback from the receiver. The receiver knows nothing about $\mathcal{A}$ except what it can learn from the channel output $\mathbf{Y}_k$.

Code definition: The following definition formalizes the rateless Gaussian RAC code described above.

Definition 3: An $\big(\{n_j, \epsilon_j\}_{j=0}^K, M, P\big)$-RAC code for the Gaussian RAC with K transmitters consists of a single encoding function $\mathsf{f} \colon \mathcal{U} \times [M] \to \mathbb{R}^{n_K}$ and decoding functions $\mathsf{g}_k \colon \mathcal{U} \times \mathbb{R}^{n_k} \to [M]^k \cup \{\mathsf{e}\}$ for $k = 0, \ldots, K$, where the input $u \in \mathcal{U}$ to the encoder and decoders is common randomness shared by all transmitters and the receiver.[1] If it cannot decode at time n_k, the decoder outputs the erasure symbol "e" and broadcasts value 0 to the transmitters, informing them that they should keep transmitting. If it can decode at time n_k, the decoder broadcasts value 1 to the transmitters, informing them that they should stop transmitting. The codewords satisfy the maximal-power constraints

$$\Big\| \mathsf{f}(u, m)^{[n_j]} \Big\|^2 \le n_j P \text{ for } m \in [M], u \in \mathcal{U}, j \in [K]. \tag{28}$$

If k transmitters are active, then the average error probability in decoding k messages at time n_k is bounded as

$$\frac{1}{M^k} \sum_{m_{[k]} \in [M]^k} \mathbb{P}\Bigg[\Big\{ \bigcup_{t\colon n_t \le n_k, t \ne k} \Big\{ \mathsf{g}_t(U, \mathbf{Y}_k^{[n_t]}) \ne \mathsf{e} \Big\} \Big\} \bigcup \Big\{ \mathsf{g}_k(U, \mathbf{Y}_k^{[n_k]}) \overset{\pi}{\ne} m_{[k]} \Big\} \Bigg| \mathbf{X}_{[k]}^{[n_k]} = \mathsf{f}(U, m_{[k]})^{[n_k]} \Bigg] \le \epsilon_k, \tag{29}$$

where $\mathsf{f}(U, m_i)$ is the codeword for the message $m_i \in [M]$, U is the common randomness random variable, and the output $\mathbf{Y}_k$ is generated according to (27). If no transmitters are active, then the decoder decodes to the unique message $[M]^0 \triangleq \{0\}$ with probability of error bounded as

$$\mathbb{P}\Big[\mathsf{g}_0(U, \mathbf{Y}_0^{[n_0]}) \ne 0 \Big] \le \epsilon_0. \tag{30}$$

[1] The realization u of the common randomness random variable U initializes the encoders and the decoder. At the start of each communication epoch, u is shared by all transmitters and the receiver. We show in [21, Appendix D] that the alphabet size of U need never exceed $K + 1$.

B. A Third-order Achievability Result for the Gaussian RAC

The following theorem is the main result of this section.

Theorem 4: Fix $K < \infty$, $\epsilon_k \in (0, 1)$ for $k \in \{0\} \cup [K]$, and M. An $\big(\{n_j, \epsilon_j\}_{j=0}^K, M, P\big)$-RAC code exists for the Gaussian RAC with K possible transmitters provided that

$$k \log M \le n_k C(kP) - \sqrt{n_k (V(kP) + V_{\text{cr}}(k, P))} Q^{-1}(\epsilon_k) + \frac{1}{2} \log n_k + O(1) \tag{31}$$

for $k \in [K]$, and

$$n_0 \ge \frac{4(1 + P^2)}{P^2} \log n_1 + o(\log n_1), \tag{32}$$

where $C(\cdot)$, $V(\cdot)$, and $V_{\text{cr}}(\cdot, \cdot)$ are the capacity (1), dispersion (3), and cross dispersion functions (26), respectively. All uses of $O(\cdot)$ and $o(\cdot)$ are taken with respect to n_1.

Remark 2: From (31), $n_1 \to \infty$ implies that $n_2, \ldots, n_K$ also grow without bound. Since all target error values ϵ_k are assumed to be constants with respect to n_1, choosing decoding times $n_0, \ldots, n_K$ so that (31) and (32) hold with equality results in $n_k = O(n_1)$ for $k \ge 2$, and $n_0 = O(\log n_1)$ (see (35), below).

Proof: Theorem 4 follows from the non-asymptotic achievability bound in Theorem 5, below, which bounds the average error probability of the proposed Gaussian RAC code. See Section VIII for details. ∎

Theorem 5: Fix constants $\lambda_k > 0$ for $k \in \{0\} \cup [K]$ and distribution $P_{\mathbf{X}}$ on $\mathbb{R}^{n_K}$. Then, there exists an $\big(\{n_j, \epsilon_j\}_{j=0}^K, M, P\big)$-RAC code with

$$\epsilon_0 \le \mathbb{P}\Big[\Big| \Big\| \mathbf{Y}_0^{[n_0]} \Big\|^2 - n_0 \Big| > n_0 \lambda_0 \Big] \tag{33}$$

$$\epsilon_k \le \frac{k(k-1)}{2M} + \mathbb{P}\left[\bigcup_{i=1}^{k} \bigcup_{\substack{j: n_j \le n_k \\ j \ge 1}} \Big\{ \Big\| \mathbf{X}_i^{[n_j]} \Big\|^2 > n_j P \Big\} \right] \tag{34a}$$

$$+ \mathbb{P}\Bigg[\bigcup_{\substack{t: n_t \le n_k \\ t \ne k}} \Big\{ \Big| \Big\| \mathbf{Y}_k^{[n_t]} \Big\|^2 - n_t(1 + tP) \Big| \le n_t \lambda_t \Big\} \bigcup \Big\{ \Big| \Big\| \mathbf{Y}_k^{[n_k]} \Big\|^2 - n_k(1 + kP) \Big| > n_k \lambda_k \Big\} \Bigg] \tag{34b}$$

$$+ \mathbb{E}\Big[\min\Big\{ 1, \sum_{s=1}^{k} \binom{k}{s} \binom{M-k}{s} \mathbb{P}\big[\imath_{[s]}(\bar{\mathbf{X}}_{[s]}^{[n_k]}; \mathbf{Y}_k^{[n_k]} | \mathbf{X}_{[s+1:k]}^{[n_k]}) \ge \imath_{[s]}(\mathbf{X}_{[s]}^{[n_k]}; \mathbf{Y}_k^{[n_k]} | \mathbf{X}_{[s+1:k]}^{[n_k]}) \mid \mathbf{X}_{[k]}^{[n_k]}, \mathbf{Y}_k^{[n_k]} \big] \Big\} \Big] \tag{34c}$$

for all $k \in [K]$, where $\mathbf{X}_{[K]}, \bar{\mathbf{X}}_{[K]}, \mathbf{Y}_k \in \mathbb{R}^{n_K}$ are distributed according to $P_{\mathbf{X}_{[K]}, \bar{\mathbf{X}}_{[K]}, \mathbf{Y}_k}(\mathbf{x}_{[K]}, \bar{\mathbf{x}}_{[K]}, \mathbf{y}_k) = \Big(\prod_{j \in [K]} P_{\mathbf{X}}(\mathbf{x}_j) P_{\mathbf{X}}(\bar{\mathbf{x}}_j) \Big) P_{\mathbf{Y}_k | \mathbf{X}_{[k]}}(\mathbf{y}_k | \mathbf{x}_{[k]})$, and $P_{\mathbf{Y}_k | \mathbf{X}_{[k]}}$ is given in (27).

Proof: The terms in (34a) capture the probability that at least two transmitters send the same message and the probability of a power constraint violation, respectively. We treat the event that at least two transmitters send the same message as an error because the analysis relies on the codeword independence across transmitters. The probability in (34b) captures the probability that the decoder decodes at a wrong decoding time, and the expectation in (34c) captures the probability that the decoder decodes an incorrect message list at the correct decoding time n_k for k active transmitters. See Section VII for details. ■

We conclude this section with some remarks concerning Theorems 4 and 5.

1) Theorem 4 shows that for the Gaussian RAC, our proposed rateless code performs as well in the first-, second-, and third-order terms as the best known MAC communication scheme without feedback (Corollary 1). In other words, the first three terms on the right-hand side of (31) for k active transmitters match the first three terms of the largest achievable sum-rate in our achievability bound in (25) for the k-transmitter MAC.
2) To prove Theorem 4, we particularize the distribution of the random codewords, $P_{\mathbf{X}}$, in Theorem 5 as follows. The first n_1 symbols are drawn uniformly from $\mathbb{S}^{n_1}(\sqrt{n_1 P})$. The sub-vector of symbols indexed from $n_{j-1}+1$ to n_j is drawn uniformly from $\mathbb{S}^{n_j-n_{j-1}}(\sqrt{(n_j-n_{j-1})P})$ for $j = 2, \ldots, K$. These K sub-codewords, each uniformly distributed on an incremental power sphere, are independent. Under this $P_{\mathbf{X}}$, the maximal-power constraint in (28) is satisfied with equality for each number of active transmitters. Rather than using an encoding function that depends on the feedback from the receiver to the transmitters, we use an encoding function that is suitable for all possible transmitter activity patterns and does not depend on the receiver's feedback. Given that a decision is made at time n_k, the active transmitters have transmitted only the first n_k symbols of the codewords representing their messages during that epoch, and the remaining $n_K - n_k$ symbols of the codewords are not transmitted.
3) As noted in [13], our achievability proofs leverage the fact that the number of active transmitters can be reliably estimated from the total received power. This is possible because when k active transmitters send k distinct messages, the average received power $\frac{1}{n_k}\mathbb{E}\left[\left\|\mathbf{Y}_k^{[n_k]}\right\|^2\right]$ at time n_k, concentrates around its mean value, $1+kP$, and this mean is distinct for each $k \in \{0\} \cup [K]$. The decoding function used at time n_k combines the maximum likelihood decoding rule for the k-transmitter MAC with a typicality rule based on the power of the output. The typicality rule decides to decode at time n_k if the average received power at time n_k lies in the interval $\left[1+kP-\frac{P}{2}, 1+kP+\frac{P}{2}\right]$ for $k \geq 1$ and $\left[1-O(n_0{}^{-\frac{1}{2}}), 1+O(n_0{}^{-\frac{1}{2}})\right]$ for $k=0$. In this case, the decoder decodes k messages at time n_k by using the maximum likelihood decoding rule. When at least two transmitters send the same message (e.g., $\mathbf{X}_1^{[n_k]} = \mathbf{X}_2^{[n_k]}$), $\frac{1}{n_k}\mathbb{E}\left[\left\|\mathbf{Y}_k^{[n_k]}\right\|^2\right] \geq 1+(k+2)P$. In our decoder design, we choose not to handle this scenario because the probability that at least two transmitters send the same message is negligible as shown in (177) in Section VII-B, below.
4) Theorem 5 applies without change to non-Gaussian RACs with power constraints satisfying the conditions in [21, Th. 1]; the tightness of the bound depends on how well k can be estimated from the received power.
5) The proof of Theorem 4 indicates that the constant term $O(1)$ in (31) depends on the number of active transmitters k, but not on the total number of transmitters K. Not requiring the decoder to determine transmitter identity is crucial for this $O(1)$ bound to hold.
6) By choosing $n_1, \ldots, n_K$ such that the inequalities in (31) are satisfied with equality for each k, we can express each n_k as a function of $n_1, \epsilon_1, \epsilon_k, k$, and P, given by

$$n_k = n_1\frac{kC_1}{C_k} + \sqrt{n_1}\left(\frac{1}{C_k}\sqrt{\frac{kC_1V_k}{C_k}}Q^{-1}(\epsilon_k) - \frac{k}{C_k}\sqrt{V_1}Q^{-1}(\epsilon_1)\right) + \frac{k-1}{2C_k}\log n_1 + O(1), \quad (35)$$

where $C_k = C(kP)$ and $V_k = V(kP) + V_{\text{cr}}(k,P)$. We derive (35) by computing the Taylor series expansion of the equation for n_k (31) in terms of k, P, ϵ_k, and $\log M$; we then replace $\log M$ by (31) for $k=1$. Fig. 1 shows the approximate decoding times $\{n_k\}_{k=1}^6$ (neglecting the $O(1)$ term in (35)), where $P=1$, $\epsilon_k = 10^{-3}$ for all k, and the smallest decoding time $n_1 \in [100, 1000]$.
7) Theorem 4 implies that the input distribution used for the Gaussian RAC also achieves the performance in Theorem 3 for the K-transmitter Gaussian MAC. As long as $n_j - n_{j-1} \geq cn_K$ holds for some constant $c > 0$ for all $j \in [K]$, requiring separate power constraints on each

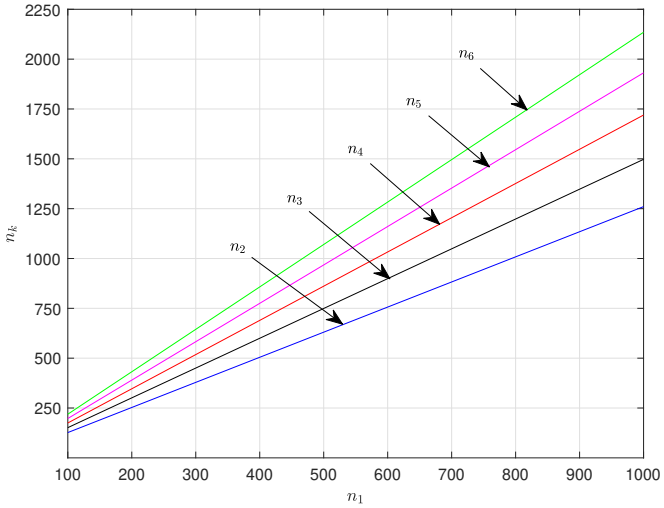

Fig. 1. Let $P=1$, $\epsilon_k = 10^{-3}$ for all k. The decoding times $\{n_k\}_{k=1}^6$ are given according to (35), where the $O(1)$ term in (35) is ignored, and $n_1 \in [100, 1000]$.

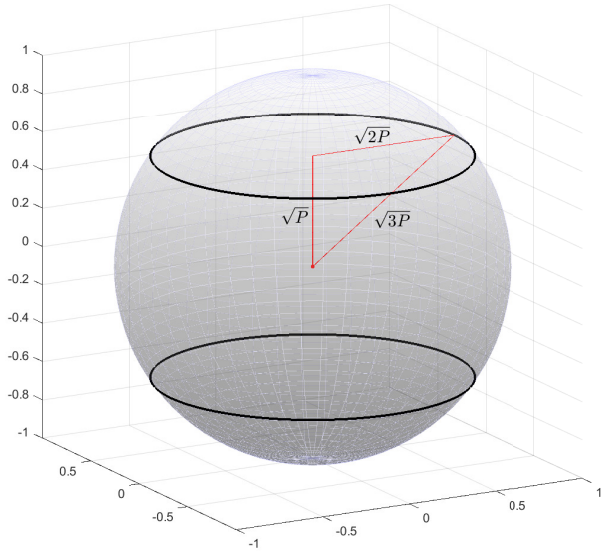

Fig. 2. Let $K = 2$, $n_1 = 2, n_2 = 3$, and $P_1 = P_2 = P = \frac{1}{3}$. The support of the input distribution for the Gaussian RAC is the Cartesian product of $\mathbb{S}^{n_1}(\sqrt{n_1 P})$ (here a circle with radius $\sqrt{2P}$) and $\mathbb{S}^{n_2-n_1}(\sqrt{(n_2-n_1)P})$ (here the set $\{-\sqrt{P}, \sqrt{P}\}$.) This set, shown above as a pair of circles, is a subset of $\mathbb{S}^{n_2}(\sqrt{n_2 P})$; the set $\mathbb{S}^{n_2}(\sqrt{n_2 P})$ is the support of the input distribution used in Theorem 3 for the Gaussian MAC.

sub-block of the codewords as

$$\left\| \mathsf{f}_i(m_i)^{[n_j]} \right\|^2 \leq n_j P_i \text{ for } m_i \in [M_i], i \in [K], j \in [K] \tag{36}$$

does not degrade our performance bound, which matches the first three terms in the expansion in Theorem 3. The support of the distribution from which the codewords are drawn for the Gaussian MAC and RAC is illustrated in Fig. 2.

8) The coding strategy we propose in [21, Th. 1] requires an i.i.d. input distribution. One can also employ the coding strategy in [21, Th. 1] to the Gaussian MAC, drawing codewords i.i.d. from $\mathcal{N}(0, P')$ for some $P' = P - \delta_n$ and δ_n decaying to 0 sufficiently quickly as blocklength n grows without bound, provided that we discard codewords that violate the maximal-power P constraint. However, [23, eq. (5.113)] shows that the resulting achievable second-order term is inferior to that achieved by the uniform distribution on the sphere.
9) As described above, the number of active transmitters in an epoch is estimated via a sequence of decodability tests. An alternative strategy is to estimate the number of active transmitters in one shot from the received power at time n_0, and to inform the transmitters about the estimate, t, of the number of active transmitters via a $\lceil \log(K+1) \rceil$-bit feedback at time n_0. Given this knowledge, active transmitters can employ an encoding function matched to t. We show in Appendix B-A that this modified coding strategy affects our bound in (31) only in the $O(1)$ term.
10) By using distinct codebooks for each transmitter, the decoder can associate transmitter identities with the decoded messages. We show that the first three terms of the expansion in (31) are still achievable in this setting. This scenario is discussed in Appendix B-B.

V. Proof of Theorem 2

A. Supporting Lemmas

We begin by presenting the lemmas that play a key role in the proof of Theorem 2. The first two lemmas are used to bound the probability that the squared norm of the output of the channel, $\mathbf{Y}_2 = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{Z}$, does not belong to its typical interval around $1 + P_1 + P_2$.

Lemma 1 from [8, Prop. 2] uniformly bounds the Radon-Nikodym derivative of the conditional and unconditional output distributions of the Gaussian MAC (13) in response to the spherical inputs with respect to the output distributions that result under i.i.d. Gaussian inputs. The squared norm of the output in response to the i.i.d. Gaussian inputs has a chi-squared distribution.

Lemma 1 (MolavianJazi and Laneman [8, Prop. 2]):

1) 2-Transmitter MAC: Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be independent, distributed uniformly on $\mathbb{S}^n(\sqrt{nP_1})$ and $\mathbb{S}^n(\sqrt{nP_2})$, respectively. Let $\tilde{\mathbf{X}}_i \sim \mathcal{N}(\mathbf{0}, P_i \mathsf{I}_n)$, $i \in [2]$, be independent of each other. Let $P_{\mathbf{X}_1\mathbf{X}_2} \to P_{\mathbf{Y}_2|\mathbf{X}_1\mathbf{X}_2} \to P_{\mathbf{Y}_2}$, and $P_{\tilde{\mathbf{X}}_1\tilde{\mathbf{X}}_2} \to P_{\mathbf{Y}_2|\mathbf{X}_1\mathbf{X}_2} \to P_{\tilde{\mathbf{Y}}_2}$, where $P_{\mathbf{Y}_2|\mathbf{X}_1\mathbf{X}_2}$ is the Gaussian MAC (13) with $k = 2$ transmitters. Then there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $\forall (\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \in \mathbb{R}^{n \otimes 3}$, it holds that

$$\frac{P_{\mathbf{Y}_2|\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_2)}{P_{\tilde{\mathbf{Y}}_2|\tilde{\mathbf{X}}_2}(\mathbf{y}|\mathbf{x}_2)} \leq \kappa_1(P_1) = 27\sqrt{\frac{\pi}{8}}\frac{1+P_1}{\sqrt{1+2P_1}} \tag{37}$$

$$\frac{P_{\mathbf{Y}_2}(\mathbf{y})}{P_{\tilde{\mathbf{Y}}_2}(\mathbf{y})} \leq \kappa_2(P_1, P_2) = \frac{9}{2\pi\sqrt{2}}\frac{P_1+P_2}{\sqrt{P_1 P_2}}. \tag{38}$$

If there is no additive noise $\mathbf{Z}$ in (12), (38) continues to hold. Inequalities (37)–(38) are generalized to the K-transmitter Gaussian MAC as follows.

2) K-Transmitter MAC: Let $\mathbf{X}_1, \ldots, \mathbf{X}_K$ be independent, and for each $i \in [K]$, let $\mathbf{X}_i$ be distributed uniformly on $\mathbb{S}^n(\sqrt{nP_i})$. Let $\tilde{\mathbf{X}}_i \sim \mathcal{N}(\mathbf{0}, P_i \mathsf{I}_n)$ for $i \in [K]$, where $\mathbf{X}_i$ are independent of each other. Let $P_{\mathbf{X}_{[K]}} \to P_{\mathbf{Y}_K|\mathbf{X}_{[K]}} \to P_{\mathbf{Y}_K}$, and $P_{\tilde{\mathbf{X}}_{[K]}} \to P_{\mathbf{Y}_K|\mathbf{X}_{[K]}} \to P_{\tilde{\mathbf{Y}}_K}$, where $P_{\mathbf{Y}_K|\mathbf{X}_{[K]}}$ is the Gaussian MAC in (13) with K transmitters. Then there exists $n_K \in \mathbb{N}$ such that for all $n \geq n_K$, for any $\mathbf{x}_{[K]} \in \mathbb{R}^{n \otimes K}$, $\mathbf{y} \in \mathbb{R}^n$, and non-empty $\mathcal{S} \in \mathcal{P}([K])$, it holds that

$$\frac{P_{\mathbf{Y}_K|\mathbf{X}_{\mathcal{S}^c}}(\mathbf{y}|\mathbf{x}_{\mathcal{S}^c})}{P_{\tilde{\mathbf{Y}}_K|\tilde{\mathbf{X}}_{\mathcal{S}^c}}(\mathbf{y}|\mathbf{x}_{\mathcal{S}^c})} \leq \kappa_{|\mathcal{S}|}(P_s : s \in \mathcal{S}), \tag{39}$$

where $\kappa_{|\mathcal{S}|}(P_s : s \in \mathcal{S})$ is a constant depending only on the power values $(P_s : s \in \mathcal{S})$.

The proof of (39), which is given in [23, eq. (5.138)], relies on a recursive formula for the distribution of $\mathbf{Y}_K$.

Lemma 2, stated next, bounds the tail probabilities of the chi-squared distribution from above.

Lemma 2 (Laurent and Massart [37, Lemma 1]): Let χ^2_n be a random variable with a chi-squared distribution and n degrees of freedom. Then for $t > 0$,

$$\mathbb{P}\left[\chi^2_n - n \geq 2\sqrt{nt} + 2t\right] \leq \exp\{-t\} \tag{40}$$

$$\mathbb{P}\left[\chi_n^2 - n \le -2\sqrt{nt}\right] \le \exp\{-t\}. \tag{41}$$

Lemma 3, stated next, is used as the main tool to obtain large deviation bounds on the information density random variables that arise when we apply the RCU bound.

Lemma 3 (Tan and Tomamichel [5, eq. (52)]): Let $\mathbf{Z} = (Z_1, \ldots, Z_n) \sim \mathcal{N}(\mathbf{0}, \mathsf{I}_n)$, $\mathbf{x} = (\sqrt{nP}, 0, \ldots, 0)$, and let $s > 0$ and $P > 0$ be constants. Then for any $a \in \mathbb{R}$, $\mu > 0$, and n large enough,

$$\mathbb{P}\left[Z_1 \in \left[\frac{a}{\sqrt{nP}}, \frac{a+\mu}{\sqrt{nP}}\right] \middle| \|\mathbf{x}+\mathbf{Z}\|^2 = ns\right] \le \frac{L(P,s)\mu}{\sqrt{n}}, \tag{42}$$

where

$$L(P,s) \triangleq \frac{8(Ps)^{3/2}}{\sqrt{2\pi}} \sqrt{\frac{1+4Ps-\sqrt{1+4Ps}}{(\sqrt{1+4Ps}-1)^5}}. \tag{43}$$

We state the multidimensional Berry-Esseen theorem for sums of independent but not necessarily identical random vectors. The theorem is used as the main tool to bound the probability that the information density random vector belongs to a given set.

Theorem 6 (Bentkus [38]): Let $\mathbf{U}_1, \ldots, \mathbf{U}_n$ be zero mean, independent random vectors in $\mathbb{R}^d$, and let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{I}_d)$. Denote $\mathbf{S} = \sum_{i=1}^n \mathbf{U}_i$, and $T = \sum_{i=1}^n \mathbb{E}\left[\|\mathbf{U}_i\|^3\right]$. Assume that $\mathrm{Cov}\left[\mathbf{S}\right] = \mathsf{I}_d$. Then, there exists a constant $c > 0$ such that

$$\sup_{\mathcal{A} \in \mathfrak{C}_d} |\mathbb{P}\left[\mathbf{S} \in \mathcal{A}\right] - \mathbb{P}[\mathbf{Z} \in \mathcal{A}]| \le cd^{1/4}T, \tag{44}$$

where $\mathfrak{C}_d$ is the set of all convex, Borel measurable subsets of $\mathbb{R}^d$.

Raič [39, Th. 1.1] establishes that the constant $cd^{1/4}$ in (44) can be replaced by $42d^{1/4} + 16$. Tan and Kosut [40] provide the following corollary to Theorem 6 for the case of a general nonsingular $\mathrm{Cov}\left[\mathbf{S}\right]$.

Corollary 2 (Tan and Kosut [40, Corollary 8]): For the setup in Theorem 6, assume that $\mathrm{Cov}\left[\mathbf{S}\right] = n\mathsf{V}$, where $\lambda_{\min}(\mathsf{V}) > 0$ denotes the minimum eigenvalue of V, and $T = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\|\mathbf{U}_i\|^3\right]$. Let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{V})$. Then, there exists a constant $c > 0$ such that

$$\sup_{\mathcal{A} \in \mathfrak{C}_d} \left|\mathbb{P}\left[\frac{1}{\sqrt{n}}\mathbf{S} \in \mathcal{A}\right] - \mathbb{P}[\mathbf{Z} \in \mathcal{A}]\right| \le \frac{cd^{1/4}T}{\sqrt{n}\lambda_{\min}(\mathsf{V})^{3/2}}. \tag{45}$$

Lemmas 4 and 5, below, are used to bound the probability that the information density random vector belongs to a set. The total variation distance between the measures P_X and P_Y on $\mathbb{R}^d$ is defined as

$$\begin{aligned}\mathrm{TV}(P_X, P_Y) &\triangleq \sup_{\mathcal{D} \in \mathbb{R}^d} |\mathbb{P}\left[X \in \mathcal{D}\right] - \mathbb{P}\left[Y \in \mathcal{D}\right]| \\ &= \frac{1}{2}\int_{x \in \mathbb{R}^d} |dP_X(x) - dP_Y(x)|. \end{aligned} \tag{46}$$

Lemma 4, stated next, bounds the total variation distance between two Gaussian vectors.

Lemma 4: Let Σ_1 and Σ_2 be two positive definite $d \times d$ matrices, and let $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^d$ be two constant vectors. Then,

$$\begin{aligned}&\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}_1, \Sigma_1), \mathcal{N}(\boldsymbol{\mu}_2, \Sigma_2)) \\ &\le \frac{2+\sqrt{6}}{4}\left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathsf{I}_d\right\|_F \\ &\quad + \frac{1}{2}\sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\Sigma_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}, \end{aligned} \tag{47}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Proof: See Appendix C. ∎

A weaker version of the bound in Lemma 4 by Devroye *et al.* appears in [41, Th. 1.1]. Like our proof, the proof of [41, Th. 1.1] relies on Pinsker's inequality. We improve the factor in front of the Frobenius norm from 1.5 in [41, Th 1.1] to $\frac{2+\sqrt{6}}{4} \approx 1.113$ by using the result in [42, Th. 1.1] to lower bound the logdeterminant of the matrix $\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathsf{I}_d$ in (47).

Lemma 5, stated next, gives an upper bound on the total variation distance between the marginal distribution of the first k dimensions of a random variable distributed uniformly on $\mathbb{S}^n(\sqrt{n})$ and the k-dimensional standard Gaussian random vector.

Lemma 5 (Stam [43, Th. 2]): Let $\mathbf{X} = (X_1, \ldots, X_n)$ be distributed uniformly on $\mathbb{S}^n(\sqrt{n})$. Let $\mathbf{X}^{[k]} = (X_1, \ldots, X_k)$ contain the first k coordinates of $\mathbf{X}$. Then,

$$\mathrm{TV}(P_{\mathbf{X}^{[k]}}, \mathcal{N}(\mathbf{0}, \mathsf{I}_k)) \le n^{\frac{1}{2}k}(n-k-2)^{-\frac{1}{2}k} - 1 \tag{48}$$

for $n > k + 2$.

We use Lemma 5 with $k = 1$ to approximate the inner product $\langle \mathbf{X}_1, \mathbf{X}_2 \rangle$ by a Gaussian random variable, which facilitates an application of the Berry-Esseen theorem in Section V-F.

The proof of Theorem 2 relies on a random coding argument and Theorem 1. The asymptotic analysis of the RCU bound (Theorem 1) borrows some techniques from the point-to-point case [5].

B. Encoding and Decoding for the MAC

We select the distributions of the independent inputs $\mathbf{X}_1$ and $\mathbf{X}_2$ as the uniform distributions on $\mathbb{S}^n(\sqrt{nP_1})$ and $\mathbb{S}^n(\sqrt{nP_2})$, which are the n-dimensional spheres centered at the origin with radii $\sqrt{nP_1}$ and $\sqrt{nP_2}$, respectively. The resulting distribution is

$$P_{\mathbf{X}_1}(\mathbf{x}_1)P_{\mathbf{X}_2}(\mathbf{x}_2) = \frac{\delta(\|\mathbf{x}_1\|^2 - nP_1)}{S_n(\sqrt{nP_1})}\frac{\delta(\|\mathbf{x}_2\|^2 - nP_2)}{S_n(\sqrt{nP_2})}, \tag{49}$$

where $\delta(\cdot)$ is the Dirac delta function, and

$$S_n(r) \triangleq \frac{2\pi^{n/2}}{\Gamma(n/2)}r^{n-1} \tag{50}$$

is the surface area of an n-dimensional sphere $\mathbb{S}^n(r)$ with radius r. We draw M_1 codewords i.i.d. from $P_{\mathbf{X}_1}$ and M_2 codewords i.i.d. from $P_{\mathbf{X}_2}$, respectively. We denote these by $\mathsf{f}_i(m_i)$ for $m_i \in [M_i]$, $i \in \{1, 2\}$.

In order to use Theorem 1, the channel $P_{Y_2|X_1X_2}$ is particularized to the two-transmitter Gaussian MAC in (13).

Upon receiving the output sequence $\mathbf{y}$, the decoder employs a maximum likelihood decoding rule, given by

$$\mathsf{g}(\mathbf{y}) = \begin{cases} (m_1, m_2) & \text{if } \imath_{1,2}(\mathsf{f}_1(m_1), \mathsf{f}_2(m_2); \mathbf{y}) \\ & > \imath_{1,2}(\mathsf{f}_1(m_1'), \mathsf{f}_2(m_2'); \mathbf{y}) \\ & \text{for all } (m_1', m_2') \neq (m_1, m_2), \\ & (m_1', m_2') \in [M_1] \times [M_2] \\ \text{error} & \text{otherwise.} \end{cases} \tag{51}$$

We treat all ties in (51) as errors because the probability that two codewords result in exactly the same information density is negligible due to the continuity of the noise. Substituting the transition law of the Gaussian MAC (13) and the uniform input distributions on the power spheres (49) into (5a)–(5c), we compute for any $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \in \mathbb{R}^{n \otimes 3}$

$$\begin{aligned} \imath_1(\mathbf{x}_1; \mathbf{y}|\mathbf{x}_2) &= \frac{n}{2} \log \frac{1}{2\pi} + \langle \mathbf{y} - \mathbf{x}_2, \mathbf{x}_1 \rangle - \frac{\|\mathbf{y} - \mathbf{x}_2\|^2}{2} \\ &\quad - \frac{nP_1}{2} - \log P_{\mathbf{Y}_2|\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_2) \end{aligned} \tag{52}$$

$$\begin{aligned} \imath_2(\mathbf{x}_2; \mathbf{y}|\mathbf{x}_1) &= \frac{n}{2} \log \frac{1}{2\pi} + \langle \mathbf{y} - \mathbf{x}_1, \mathbf{x}_2 \rangle - \frac{\|\mathbf{y} - \mathbf{x}_1\|^2}{2} \\ &\quad - \frac{nP_2}{2} - \log P_{\mathbf{Y}_2|\mathbf{X}_1}(\mathbf{y}|\mathbf{x}_1) \end{aligned} \tag{53}$$

$$\begin{aligned} \imath_{1,2}(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) &= \frac{n}{2} \log \frac{1}{2\pi} + \langle \mathbf{y}, \mathbf{x}_1 + \mathbf{x}_2 \rangle - \frac{\|\mathbf{y}\|^2}{2} \\ &\quad - \frac{\|\mathbf{x}_1 + \mathbf{x}_2\|^2}{2} - \log P_{\mathbf{Y}_2}(\mathbf{y}). \end{aligned} \tag{54}$$

Observe that for each $\mathbf{x}_2$ and $\mathbf{y}$, $\imath_1(\mathbf{x}_1; \mathbf{y}|\mathbf{x}_2)$ depends on $\mathbf{x}_1$ only through the inner product $\langle \mathbf{y} - \mathbf{x}_2, \mathbf{x}_1 \rangle$, and for each $\mathbf{y}$, $\imath_{1,2}(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})$ depends on $(\mathbf{x}_1, \mathbf{x}_2)$ only through $\langle \mathbf{y}, \mathbf{x}_1 + \mathbf{x}_2 \rangle - \langle \mathbf{x}_1, \mathbf{x}_2 \rangle$. By the input-output relation in (12), the conditional information density for two transmitters, $\imath_1(\mathbf{x}_1; \mathbf{y}|\mathbf{x}_2)$, can be reduced to the unconditional information density for a single transmitter as

$$\imath_1(\mathbf{x}_1; \mathbf{y}|\mathbf{x}_2) = \imath_1(\mathbf{x}_1; \mathbf{y} - \mathbf{x}_2) \triangleq \log \frac{P_{\mathbf{Y}_1|\mathbf{X}_1}(\mathbf{y} - \mathbf{x}_2|\mathbf{x}_1)}{P_{\mathbf{Y}_1}(\mathbf{y} - \mathbf{x}_2)}, \tag{55}$$

where $\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{Z}$ is the output of the channel with a single transmitter.

C. Typical Set for the MAC

For the rest of the proof, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{I}_n)$ denotes the Gaussian noise, which is independent of the channel inputs $\mathbf{X}_1$ and $\mathbf{X}_2$. Note that the expectations of the squared norms of $\mathbf{X}_1 + \mathbf{Z}, \mathbf{X}_2 + \mathbf{Z}$ and $\mathbf{Y}_2$ are $n(1+P_1), n(1+P_2)$, and $n(1+P_1+P_2)$, respectively. We define a typical set for vector $(\mathbf{X}_1 + \mathbf{Z}, \mathbf{X}_2 + \mathbf{Z}, \mathbf{Y}_2)$ by

$$\mathcal{F} \triangleq \bigtimes_{\mathcal{S} \in \mathcal{P}([2])} \mathcal{F}(\mathcal{S}) \subseteq \mathbb{R}^{n \otimes 3}, \tag{56}$$

where

$$\mathcal{F}(\mathcal{S}) \triangleq \left\{ \mathbf{x}_{\langle \mathcal{S} \rangle} + \mathbf{z} \in \mathbb{R}^n \colon \frac{1}{n} \left\| \mathbf{x}_{\langle \mathcal{S} \rangle} + \mathbf{z} \right\|^2 \in \mathcal{I}(\mathcal{S}) \right\} \tag{57}$$

$$\mathcal{I}(\mathcal{S}) \triangleq [1 + P_{\langle \mathcal{S} \rangle} - n^{-1/3}, 1 + P_{\langle \mathcal{S} \rangle} + n^{-1/3}]. \tag{58}$$

We next show that for n large enough,

$$\mathbb{P}\left[(\mathbf{X}_1 + \mathbf{Z}, \mathbf{X}_2 + \mathbf{Z}, \mathbf{Y}_2) \notin \mathcal{F}\right] \leq \exp\{-c_2 n^{1/3}\}, \tag{59}$$

where $c_2 > 0$ is a constant.

To bound the probability that the triplet $(\mathbf{X}_1 + \mathbf{Z}, \mathbf{X}_2 + \mathbf{Z}, \mathbf{Y}_2)$ does not belong to the typical set $\mathcal{F}$, we use Lemma 1 to approximate the squared norms $\|\mathbf{X}_1 + \mathbf{Z}\|^2$, $\|\mathbf{X}_2 + \mathbf{Z}\|^2$, and $\|\mathbf{Y}_2\|^2$ by multiples of chi-squared distributed random variables with n degrees of freedom. We then use Lemma 2 to bound the two-sided tail probability of these chi-squared distributed random variables. Weakening the upper bound (40) in Lemma 2 using $2\sqrt{2nt} \geq 2\sqrt{nt} + 2t$ for $0 < t \leq \frac{n}{8} \leq (3 - 2\sqrt{2})n$, we get the following concentration inequalities for the squared norms of the random vectors $\mathbf{X}_1 + \mathbf{Z}$ and $\mathbf{Y}_2$

$$\begin{aligned} &\mathbb{P}\left[\left| \|\mathbf{X}_1 + \mathbf{Z}\|^2 - n(1 + P_1) \right| > nt_1 \right] \\ &\qquad \leq 2\kappa_1(P_1) \exp\left\{ -\frac{nt_1^2}{8(1+P_1)^2} \right\} \end{aligned} \tag{60}$$

$$\begin{aligned} &\mathbb{P}\left[\left| \|\mathbf{Y}_2\|^2 - n(1 + P_1 + P_2) \right| > nt_2 \right] \\ &\qquad \leq 2\kappa_2(P_1, P_2) \exp\left\{ -\frac{nt_2^2}{8(1+P_1+P_2)^2} \right\} \end{aligned} \tag{61}$$

for $t_1 \in (0, 1 + P_1)$, and $t_2 \in (0, 1 + P_1 + P_2)$, where $\kappa_1(P_1)$ and $\kappa_2(P_1, P_2)$ are constants defined in Lemma 1. We deduce (59) by the union bound and setting $t_1 = t_2 = n^{-1/3}$ in (60)–(61).

D. A Large Deviation Bound on the Mutual Information Random Variables

We introduce the following functions that are analogous to the one used in the point-to-point channel in [5, eq. (27)]

$$g_1(t; \mathbf{y}, \mathbf{x}_2) \triangleq \mathbb{P}\left[\imath_1(\bar{\mathbf{X}}_1; \mathbf{Y}_2|\mathbf{X}_2) \geq t \mid \mathbf{X}_2 = \mathbf{x}_2, \mathbf{Y}_2 = \mathbf{y}\right] \tag{62}$$

$$g_2(t; \mathbf{y}, \mathbf{x}_1) \triangleq \mathbb{P}\left[\imath_2(\bar{\mathbf{X}}_2; \mathbf{Y}_2|\mathbf{X}_1) \geq t \mid \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y}_2 = \mathbf{y}\right] \tag{63}$$

$$g_{1,2}(t; \mathbf{y}) \triangleq \mathbb{P}\left[\imath_{1,2}(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2; \mathbf{Y}_2) \geq t \mid \mathbf{Y}_2 = \mathbf{y}\right], \tag{64}$$

where

$$\begin{aligned} &P_{\mathbf{X}_1 \mathbf{X}_2 \bar{\mathbf{X}}_1 \bar{\mathbf{X}}_2 \mathbf{Y}_2}(\mathbf{x}_1, \mathbf{x}_2, \bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{y}) \\ &= P_{\mathbf{X}_1}(\mathbf{x}_1) P_{\mathbf{X}_2}(\mathbf{x}_2) P_{\mathbf{X}_1}(\bar{\mathbf{x}}_1) P_{\mathbf{X}_2}(\bar{\mathbf{x}}_2) P_{\mathbf{Y}_2|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2). \end{aligned}$$

The following lemma, which generalizes [5, eq. (53)] to the Gaussian MAC, gives upper bounds on these functions. We use Lemma 6 in the evaluation of the RCU bound.

Lemma 6: Let $(\mathbf{y} - \mathbf{x}_2, \mathbf{y} - \mathbf{x}_1, \mathbf{y}) \in \mathcal{F}$, where the set $\mathcal{F}$ is defined in (56). Then, for n large enough,

$$g_1(t; \mathbf{y}, \mathbf{x}_2) \leq \frac{G_1 \exp\{-t\}}{\sqrt{n}} \tag{65a}$$

$$g_2(t; \mathbf{y}, \mathbf{x}_1) \leq \frac{G_2 \exp\{-t\}}{\sqrt{n}} \tag{65b}$$

$$g_{1,2}(t; \mathbf{y}) \leq \frac{G_{1,2} \exp\{-t\}}{\sqrt{n}}, \tag{65c}$$

where G_1, G_2, and $G_{1,2}$ are positive constants depending only on P_1, P_2, and (P_1, P_2), respectively.

Proof: The bounds in (65a) and (65b) follow from the equivalence (stated in (55)) between the conditional information density for two transmitters and the unconditional information density for a single transmitter combined with the analysis in [5, Sec. IV-E]. The constants in (65a) and (65b) are

$$G_i = (3\log 2)L(P_i, 1+P_i), \quad i \in \{1,2\}, \tag{66}$$

where $L(\cdot,\cdot)$ is the function defined in (43).

Bounding the function $g_{1,2}(t;\mathbf{y})$ is more challenging. While $\|\mathbf{X}_1\|^2$ is a constant under the uniform distribution on a power sphere, $\|\mathbf{X}_1+\mathbf{X}_2\|^2$ is not. The proof of (65c) follows steps similar to [5, Sec. IV-E]. First, we change the measure from $P_{\mathbf{X}_1}P_{\mathbf{X}_2}P_{\mathbf{Y}_2}$ to $P_{\mathbf{X}_1}P_{\mathbf{X}_2}P_{\mathbf{Y}_2|\mathbf{X}_1\mathbf{X}_2}$ to get

$$g_{1,2}(t;\mathbf{y}) = \mathbb{E}[\exp\{-\imath_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2)\} \\ 1\{\imath_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2)\geq t\} \mid \mathbf{Y}_2=\mathbf{y}]. \tag{67}$$

To bound (67), we define function $h_{1,2}(\mathbf{y};a,\mu)$ for constants $a\in\mathbb{R}$ and $\mu>0$ as

$$h_{1,2}(\mathbf{y};a,\mu) \\ \triangleq \mathbb{P}\Big[\imath_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2)\in[a,a+\mu]\Big|\mathbf{Y}_2=\mathbf{y}\Big] \tag{68}$$

$$= \mathbb{P}\Bigg[\langle \mathbf{X}_1+\mathbf{X}_2,\mathbf{Y}_2\rangle - \frac{\|\mathbf{X}_1+\mathbf{X}_2\|^2}{2} \\ \in[a',a'+\mu]\,\Bigg|\,\mathbf{Y}_2=\mathbf{y}\Bigg], \tag{69}$$

where a' is shifted from a by some amount depending on $\mathbf{y}$, and (69) follows from (54). By spherical symmetry of the distribution of $\mathbf{Y}_2$, (69) depends on $\mathbf{y}$ only through its norm $\|\mathbf{y}\|$. We have

$$h_{1,2}(s;a,\mu) \triangleq h_{1,2}(\mathbf{y};a,\mu) \\ = \mathbb{P}\Bigg[\langle \mathbf{X}_1+\mathbf{X}_2,\mathbf{Y}_2\rangle - \frac{\|\mathbf{X}_1+\mathbf{X}_2\|^2}{2} \\ \in[a',a'+\mu]\,\Bigg|\,\|\mathbf{Y}_2\|^2=ns\Bigg], \tag{70}$$

where $\|\mathbf{y}\|^2 = ns$, and $s\in\mathcal{I}([2])$, and $\mathcal{I}(\mathcal{S})$ is defined in (58). Recall that the support of the norm $\|\mathbf{X}_1+\mathbf{X}_2\|^2$ is $[n(\sqrt{P_1}-\sqrt{P_2})^2, n(\sqrt{P_1}+\sqrt{P_2})^2]$. To avoid the cases where $\|\mathbf{X}_1+\mathbf{X}_2\|^2$ is too small, we separate the probability term (70) according to whether or not the event

$$\mathcal{B} = \Big\{\|\mathbf{X}_1+\mathbf{X}_2\|^2 < n(P_1+P_2-\sqrt{P_1P_2})\Big\} \tag{71}$$

occurs under the condition that $\|\mathbf{Y}_2\|^2=ns$. Here, the choice $\sqrt{P_1P_2}$ is arbitrary and can be replaced by any constant in $(0, 2\sqrt{P_1P_2})$.

In (70), conditioning on the event $\mathcal{B}$ and then bounding the corresponding probability terms by 1 gives

$$h_{1,2}(s;a,\mu) \leq \mathbb{P}\Big[\mathcal{B}\,\Big|\,\|\mathbf{Y}_2\|^2=ns\Big] + \mathbb{P}\Bigg[\langle \mathbf{X}_1+\mathbf{X}_2,\mathbf{Y}_2\rangle \\ - \frac{\|\mathbf{X}_1+\mathbf{X}_2\|^2}{2}\in[a',a'+\mu]\,\Bigg|\,\|\mathbf{Y}_2\|^2=ns,\mathcal{B}^c\Bigg]. \tag{72}$$

For n large enough, we bound the first term on the right-hand side of (72) by

$$\mathbb{P}\Big[\mathcal{B}\,\Big|\,\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2=ns\Big] \leq \exp\{-nC\}, \tag{73}$$

where $C>0$ is a constant. The proof of (73) appears in Appendix D.

By spherical symmetry, the distribution of $\langle \mathbf{X}_1+\mathbf{X}_2, \mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\rangle$ depends on $\mathbf{X}_1+\mathbf{X}_2$ only through the norm $\|\mathbf{X}_1+\mathbf{X}_2\|$. Therefore, fixing $\mathbf{X}_1+\mathbf{X}_2$ to $\mathbf{x} = (\sqrt{nu},0,\ldots,0)$, we find that for any $u\in[P_1+P_2-\sqrt{P_1P_2},(\sqrt{P_1}+\sqrt{P_2})^2]$, $s\in\mathcal{I}([2])$, and n large enough,

$$\mathbb{P}\Big[\langle \mathbf{X}_1+\mathbf{X}_2,\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\rangle - \frac{nu}{2}\in[a',a'+\mu] \\ \Big|\,\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2=ns, \|\mathbf{X}_1+\mathbf{X}_2\|^2=nu\Big] \\ = \mathbb{P}\left[Z_1+\frac{\sqrt{nu}}{2}\in\left[\frac{a'}{\sqrt{nu}},\frac{a'+\mu}{\sqrt{nu}}\right]\,\middle|\,\|\mathbf{x}+\mathbf{Z}\|^2=ns\right] \tag{74}$$

$$\leq \frac{L(u,s)\mu}{\sqrt{n}} \tag{75}$$

$$\leq \frac{3}{2}\frac{L(u,1+P_1+P_2)\mu}{\sqrt{n}}, \tag{76}$$

where (75) follows by Lemma 3, and (76) holds for n large enough by the continuity of the map $s\mapsto L(u,s)$ since $s\in\mathcal{I}([2])$. Using (76), we bound the second term in (72) as

$$\mathbb{P}\Big[\langle \mathbf{X}_1+\mathbf{X}_2,\mathbf{Y}_2\rangle - \frac{\|\mathbf{X}_1+\mathbf{X}_2\|^2}{2}\in[a',a'+\mu] \\ \Big|\,\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2=ns,\mathcal{B}^c\Big] \\ \leq \max_{u\in[P_1+P_2-\sqrt{P_1P_2},(\sqrt{P_1}+\sqrt{P_2})^2]}\frac{3\mu L(u,1+P_1+P_2)}{2\sqrt{n}}. \tag{77}$$

By (72), (73), (77), and because $L(u,1+P_1+P_2)$ is bounded above for $u\in[P_1+P_2-\sqrt{P_1P_2},(\sqrt{P_1}+\sqrt{P_2})^2]$, there exists a constant $K_2(P_1,P_2)>0$ such that

$$h_{1,2}(s;a,\mu) \leq K_2(P_1,P_2)\frac{\mu}{\sqrt{n}} \tag{78}$$

for n large enough. By following the same steps as [5, eq. (55)-(57)], we conclude that

$$g_{1,2}(t;\mathbf{y}) \leq \frac{G_{1,2}\exp\{-t\}}{\sqrt{n}}, \tag{79}$$

where $G_{1,2} = (2\log 2)K_2(P_1,P_2)$. ∎

E. Evaluating the RCU Bound for the MAC

We here bound the right-hand side of (8) in Theorem 1. The information density random vector is defined as

$$\boldsymbol{\imath}_2 \triangleq \begin{bmatrix}\imath_1(\mathbf{X}_1;\mathbf{Y}_2|\mathbf{X}_2)\\ \imath_2(\mathbf{X}_2;\mathbf{Y}_2|\mathbf{X}_1)\\ \imath_{1,2}(\mathbf{X}_1,X_2;\mathbf{Y}_2)\end{bmatrix}, \tag{80}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are distributed according to (49), and $P_{\mathbf{X}_1}P_{\mathbf{X}_2}\to P_{\mathbf{Y}_2|\mathbf{X}_1+\mathbf{X}_2}\to P_{\mathbf{Y}_2}$.

Define the typical events

$$\mathcal{E}(\mathcal{S}) \triangleq \big\{\mathbf{X}_{\langle\mathcal{S}\rangle}+\mathbf{Z}\in\mathcal{F}(\mathcal{S})\big\} \tag{81}$$

$$\mathcal{E} \triangleq \bigcap_{\mathcal{S}\in\mathcal{P}([2])}\mathcal{E}(\mathcal{S}) \tag{82}$$

$$\mathcal{A} \triangleq \left\{ \boldsymbol{\imath}_2 \geq \log \begin{bmatrix} M_1(G_1)^2\alpha_1 \\ M_2(G_2)^2\alpha_1 \\ M_1M_2(G_{1,2})^2\alpha_2 \end{bmatrix} - \frac{1}{2}\log n\mathbf{1} \right\}, \quad (83)$$

where G_1, G_2 and $G_{1,2}$ are the constants given in (65), $\mathcal{F}(\mathcal{S})$ is defined in (57), and

$$\alpha_s \triangleq 2\binom{2}{s}, \quad s = 1, 2. \quad (84)$$

Denote for brevity

$$g_1 \triangleq g_1(\imath_1(\mathbf{X}_1; \mathbf{Y}_2|\mathbf{X}_2); \mathbf{Y}_2, \mathbf{X}_2) \quad (85a)$$
$$g_2 \triangleq g_2(\imath_2(\mathbf{X}_2; \mathbf{Y}_2|\mathbf{X}_1); \mathbf{Y}_2, \mathbf{X}_1) \quad (85b)$$
$$g_{1,2} \triangleq g_{1,2}(\imath_{1,2}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_2); \mathbf{Y}_2), \quad (85c)$$

where $g_1(\cdot), g_2(\cdot)$, and $g_{1,2}(\cdot)$, are defined in (62)–(64).

The right-hand side of (8) is bounded in (86)–(90) at the top of the next page. Here, c_2 is the positive constant defined in (59). Equality (86) follows from the definitions of the functions $g_1(t; \mathbf{y}, \mathbf{x}_2)$ and $g_{1,2}(t; \mathbf{y})$ and splitting the expectation into two cases according to whether the event $\{\mathcal{A}^c \cup \mathcal{E}^c\}$ occurs or not. Inequality (87) follows by bounding the minimum inside the first expectation in (86) by 1; bounding the minimum inside the second expectation in (86) by its second argument; writing the indicator function $1\{\mathcal{A} \cap \mathcal{E}\}$ as a multiplication of 3 indicator functions using the definitions in (82) and (83) and distributing that multiplication over the summation. Inequality (88) follows from Lemma 6 and by bounding the probability terms by 1. Inequality (89) is obtained by applying the union bound to $\mathbb{P}[\mathcal{A}^c \cup \mathcal{E}^c]$ and by using Lemma 6 with $t = \log \frac{M_1(G_1)^2\alpha_1}{\sqrt{n}}$, $t = \log \frac{M_2(G_2)^2\alpha_1}{\sqrt{n}}$, and $t = \log \frac{M_1M_2(G_{1,2})^2\alpha_2}{\sqrt{n}}$ to bound the three remaining terms, respectively. Inequality (90) follows from (59).

F. A Multidimensional Berry-Esseen Type Inequality

To complete the proof of Theorem 2, it remains only to evaluate the probability $\mathbb{P}[\mathcal{A}^c]$ in (90). If the operational rate pair $\left(\frac{\log M_1}{n}, \frac{\log M_2}{n}\right)$ does not lie at a corner point of the achievable capacity region, applying the union bound to $\mathbb{P}[\mathcal{A}^c]$ gives a tight achievability bound since two of the three probability terms that appear after applying the union bound to $\mathbb{P}[\mathcal{A}^c]$ are $O\left(\frac{1}{\sqrt{n}}\right)$. However, for the corner points, $\mathbb{P}[\mathcal{A}^c]$ needs to be bounded without using the union bound in order to obtain a tighter achievability bound (see [23, Sec. 5.1.1]). In this section, we bound $\mathbb{P}[\mathcal{A}^c]$ jointly by deriving a multidimensional Berry-Esseen type inequality.

Due to the non-i.i.d. input distribution, the random vector $\boldsymbol{\imath}_2$ cannot be separated into a sum of n random vectors. Therefore, to approximate $\boldsymbol{\imath}_2$, we define the modified conditional and unconditional information densities whose denominators have Gaussian distributions corresponding to

$$\tilde{\imath}_1(\mathbf{x}_1; \mathbf{y}|\mathbf{x}_2) \triangleq \sum_{i=1}^{n} \log \frac{P_{Y_2|X_1X_2}(y_i|x_{1i}, x_{2i})}{P_{\tilde{Y}_2|\tilde{X}_2}(y_i|x_{2i})} \quad (91a)$$

$$\tilde{\imath}_2(\mathbf{x}_2; \mathbf{y}|\mathbf{x}_1) \triangleq \sum_{i=1}^{n} \log \frac{P_{Y_2|X_1X_2}(y_i|x_{1i}, x_{2i})}{P_{\tilde{Y}_2|\tilde{X}_1}(y_i|x_{1i})} \quad (91b)$$

$$\tilde{\imath}_{1,2}(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) \triangleq \sum_{i=1}^{n} \log \frac{P_{Y_2|X_1X_2}(y_i|x_{1i}, x_{2i})}{P_{\tilde{Y}_2}(y_i)}, \quad (91c)$$

where $\tilde{X}_i \sim \mathcal{N}(0, P_i)$, $i \in [2]$, and $P_{\tilde{X}_1}P_{\tilde{X}_2} \to P_{Y_2|X_1X_2} \to P_{\tilde{Y}_2} = \mathcal{N}(0, 1 + P_1 + P_2)$. Denote the modified and centered information density random vector by

$$\tilde{\boldsymbol{\imath}}_2 \triangleq \frac{1}{\sqrt{n}} \left(\begin{bmatrix} \tilde{\imath}_1(\mathbf{X}_1; \mathbf{Y}_2|\mathbf{X}_2) \\ \tilde{\imath}_2(\mathbf{X}_2; \mathbf{Y}_2|\mathbf{X}_1) \\ \tilde{\imath}_{1,2}(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_2) \end{bmatrix} - n\mathbf{C}(P_1, P_2) \right), \quad (92)$$

where $\mathbf{C}(P_1, P_2) = \frac{1}{n}\mathbb{E}[\boldsymbol{\imath}_2]$ is the capacity vector defined in (16). Define the threshold vector

$$\boldsymbol{\tau} \triangleq \log \begin{bmatrix} M_1(G_1)^2\kappa_1(P_1)\alpha_1 \\ M_2(G_2)^2\kappa_1(P_2)\alpha_1 \\ M_1M_2(G_{1,2})^2\kappa_2(P_1, P_2)\alpha_2 \end{bmatrix} - \frac{1}{2}\log n\mathbf{1} - n\mathbf{C}(P_1, P_2). \quad (93)$$

Our method to bound the probability $\mathbb{P}[\mathcal{A}^c]$ involves 5 steps.

Step 1: We first replace $\boldsymbol{\imath}_2$ by $\tilde{\boldsymbol{\imath}}_2$. Unlike $\boldsymbol{\imath}_2$, $\tilde{\boldsymbol{\imath}}_2$ can be written as a sum of n dependent random vectors. Prior uses of this approach include [5, eq. (65)] for the point-to-point channel and [8, eq. (2)] for the MAC. We then bound $\mathbb{P}[\mathcal{A}^c]$ in terms of the modified information density random vector $\tilde{\boldsymbol{\imath}}_2$. By (83) and Lemma 1,

$$\mathbb{P}[\mathcal{A}^c] = 1 - \mathbb{P}\left[\boldsymbol{\imath}_2 - \mathbb{E}[\boldsymbol{\imath}_2] \geq \left(\boldsymbol{\tau} - \log \begin{bmatrix} \kappa_1(P_1) \\ \kappa_1(P_2) \\ \kappa_2(P_1, P_2) \end{bmatrix} \right) \right] \quad (94)$$

$$\leq 1 - \mathbb{P}\left[\tilde{\boldsymbol{\imath}}_2 \geq \frac{1}{\sqrt{n}}\boldsymbol{\tau} \right]. \quad (95)$$

From (91a)–(91c), we see that

$$\tilde{\boldsymbol{\imath}}_2 \sim \frac{1}{\sqrt{n}} \begin{bmatrix} \frac{(n - \|\mathbf{Z}\|^2)P_1 + 2\langle \mathbf{X}_1, \mathbf{Z} \rangle}{2(1+P_1)} \\ \frac{(n - \|\mathbf{Z}\|^2)P_2 + 2\langle \mathbf{X}_2, \mathbf{Z} \rangle}{2(1+P_2)} \\ \frac{(n - \|\mathbf{Z}\|^2)(P_1+P_2) + 2\langle \mathbf{X}_1, \mathbf{X}_2 \rangle + 2\langle \mathbf{Z}, \mathbf{X}_1 + \mathbf{X}_2 \rangle}{2(1+P_1+P_2)} \end{bmatrix}. \quad (96)$$

Although the right-hand side of (96) is not a sum of n independent random vectors, the conditional distribution of $\tilde{\boldsymbol{\imath}}_2$ given $(\mathbf{X}_1, \mathbf{X}_2)$ is such a sum. Therefore, the multidimensional Berry-Esseen theorem is applicable to the corresponding conditional probability. In the remainder of Step 1, we detail the distribution of $\tilde{\boldsymbol{\imath}}_2$.

By spherical symmetry, the conditional distribution of $\tilde{\boldsymbol{\imath}}_2$ given $(\mathbf{X}_1, \mathbf{X}_2) = (\mathbf{x}_1, \mathbf{x}_2)$ depends on $(\mathbf{x}_1, \mathbf{x}_2)$ only through the inner product $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle$ given that each squared norm satisfies $\|\mathbf{x}_i\|^2 = nP_i$, $i \in [2]$. Define the normalized inner product random variable

$$H \triangleq \frac{\langle \mathbf{X}_1, \mathbf{X}_2 \rangle}{\sqrt{nP_1P_2}}, \quad (97)$$

and set

$$\mathbf{x}_1 = (\sqrt{nP_1}, 0, \ldots, 0) \quad (98)$$
$$\mathbf{x}_2 = (h\sqrt{P_2}, \sqrt{(n - h^2)P_2}, 0, \ldots, 0) \quad (99)$$

for some $h \in [-\sqrt{n}, \sqrt{n}]$, which satisfy

$$\frac{\langle \mathbf{x}_1, \mathbf{x}_2 \rangle}{\sqrt{nP_1P_2}} = h. \quad (100)$$

$$
\begin{aligned}
&\mathbb{E}\Big[\min\Big\{1,(M_1-1)\mathbb{P}\left[\imath_1(\bar{\mathbf{X}}_1;\mathbf{Y}_2|\mathbf{X}_2)\geq\imath_1(\mathbf{X}_1;\mathbf{Y}_2|\mathbf{X}_2)\mid\mathbf{X}_1,\mathbf{X}_2,\mathbf{Y}_2\right]\\
&\quad+(M_2-1)\mathbb{P}\left[\imath_2(\bar{\mathbf{X}}_2;\mathbf{Y}_2|\mathbf{X}_1)\geq\imath_2(\mathbf{X}_2;\mathbf{Y}_2|\mathbf{X}_1)\mid\mathbf{X}_1,\mathbf{X}_2,\mathbf{Y}_2\right]\\
&\quad+(M_1-1)(M_2-1)\mathbb{P}\left[\imath_{1,2}(\bar{\mathbf{X}}_1,\bar{\mathbf{X}}_2;\mathbf{Y}_2)\geq\imath_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2)\mid\mathbf{X}_1,\mathbf{X}_2,\mathbf{Y}_2\right]\Big\}\Big]\\
&=\mathbb{E}\Big[\min\Big\{1,(M_1-1)g_1+(M_2-1)g_2+(M_1-1)(M_2-1)g_{1,2}\Big\}1\left\{\mathcal{A}^c\cup\mathcal{E}^c\right\}\Big]\\
&\quad+\mathbb{E}\Big[\min\Big\{1,(M_1-1)g_1+(M_2-1)g_2+(M_1-1)(M_2-1)g_{1,2}\Big\}1\left\{\mathcal{A}\cap\mathcal{E}\right\}\Big] &(86)\\
&\leq\mathbb{P}\left[\mathcal{A}^c\cup\mathcal{E}^c\right]+\mathbb{P}\left[\mathcal{E}(\{1\})\right]\,M_1\,\mathbb{E}\left[g_1 1\left\{\imath_1(\mathbf{X}_1;\mathbf{Y}_2|\mathbf{X}_2)\geq\log\frac{M_1(G_1)^2\alpha_1}{\sqrt{n}}\right\}\,\middle|\,\mathcal{E}(\{1\})\right]\\
&\quad+\mathbb{P}\left[\mathcal{E}(\{2\})\right]\,M_2\,\mathbb{E}\left[g_2 1\left\{\imath_2(\mathbf{X}_2;\mathbf{Y}_2|\mathbf{X}_1)\geq\log\frac{M_2(G_2)^2\alpha_1}{\sqrt{n}}\right\}\,\middle|\,\mathcal{E}(\{2\})\right]\\
&\quad+\mathbb{P}\left[\mathcal{E}(\{1,2\})\right]\,M_1M_2\,\mathbb{E}\left[g_{1,2}1\left\{\imath_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2)\geq\log\frac{M_1M_2(G_{1,2})^2\alpha_2}{\sqrt{n}}\right\}\,\middle|\,\mathcal{E}(\{1,2\})\right] &(87)\\
&\leq\mathbb{P}\left[\mathcal{A}^c\cup\mathcal{E}^c\right]+\frac{M_1G_1}{\sqrt{n}}\mathbb{E}\left[\exp\{-\imath_1(\mathbf{X}_1;\mathbf{Y}_2|\mathbf{X}_2)\}1\left\{\imath_1(\mathbf{X}_1;\mathbf{Y}_2|\mathbf{X}_2)\geq\log\frac{M_1(G_1)^2\alpha_1}{\sqrt{n}}\right\}\,\middle|\,\mathcal{E}(\{1\})\right]\\
&\quad+\frac{M_2G_2}{\sqrt{n}}\mathbb{E}\left[\exp\{-\imath_2(\mathbf{X}_2;\mathbf{Y}_2|\mathbf{X}_1)\}1\left\{\imath_2(\mathbf{X}_2;\mathbf{Y}_2|\mathbf{X}_1)\geq\log\frac{M_2(G_2)^2\alpha_1}{\sqrt{n}}\right\}\,\middle|\,\mathcal{E}(\{2\})\right]\\
&\quad+\frac{M_1M_2G_{1,2}}{\sqrt{n}}\mathbb{E}\left[\exp\{-\imath_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2)\}1\left\{\imath_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2)\geq\log\frac{M_1M_2(G_{1,2})^2\alpha_2}{\sqrt{n}}\right\}\,\middle|\,\mathcal{E}(\{1,2\})\right] &(88)\\
&\leq\mathbb{P}\left[\mathcal{A}^c\right]+\mathbb{P}\left[\mathcal{E}^c\right]+\frac{\frac{2}{\alpha_1}+\frac{1}{\alpha_2}}{\sqrt{n}} &(89)\\
&\leq\mathbb{P}\left[\mathcal{A}^c\right]+\exp\left\{-c_2n^{1/3}\right\}+\frac{1}{\sqrt{n}} &(90)
\end{aligned}
$$

Putting (98)–(99) into (96) gives that the conditional distribution of $\tilde{\boldsymbol{\imath}}_2$ given $H=h$ equals the conditional distribution of $\tilde{\boldsymbol{\imath}}_2$ given $(\mathbf{X}_1,\mathbf{X}_2)=(\mathbf{x}_1,\mathbf{x}_2)$, which equals the conditional distribution of the random variable

$$\boldsymbol{\mu}(h)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{J}_i(h), \tag{101}$$

where

$$\boldsymbol{\mu}(h)\triangleq\mathbb{E}\left[\tilde{\boldsymbol{\imath}}_2|H=h\right]=h\begin{bmatrix}0\\0\\\frac{\sqrt{P_1P_2}}{1+P_1+P_2}\end{bmatrix} \tag{102}$$

$$\mathbf{J}_i(h)\triangleq\begin{bmatrix}\frac{(1-Z_i^2)P_1+2x_{1i}Z_i}{2(1+P_1)}\\\frac{(1-Z_i^2)P_2+2x_{2i}Z_i}{2(1+P_2)}\\\frac{(1-Z_i^2)(P_1+P_2)+2(x_{1i}+x_{2i})Z_i}{2(1+P_1+P_2)}\end{bmatrix},\quad i\in[n]. \tag{103}$$

Here, $\mathbf{J}_i(h)$ depends on h through the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ given in (98)–(99). Conditioned on the event that $H=h$, the modified information density random vector $\tilde{\boldsymbol{\imath}}_2$ behaves as a sum of conditionally independent but not identical random vectors $\frac{1}{n}\boldsymbol{\mu}(h)+\frac{1}{\sqrt{n}}\mathbf{J}_i(h)$ in (101).

We next find the distribution of H. By spherical symmetry, the distribution of H does not depend on $\mathbf{X}_1$. Therefore, we can set $\mathbf{X}_1=\mathbf{x}_1$ and get

$$H\sim\frac{X_{21}}{\sqrt{P_2}}, \tag{104}$$

where X_{21} denotes the first coordinate of $\mathbf{X}_2$. Therefore, H is distributed according to the marginal distribution of the first coordinate of a random vector distributed uniformly on $\mathbb{S}^n(\sqrt{n})$. The distribution of H is computed as (e.g., [43, Th. 1])

$$P_H(h)=\frac{\Gamma(\frac{n}{2})}{\sqrt{\pi n}\Gamma(\frac{n-1}{2})}\left(1-\frac{h^2}{n}\right)_+^{\frac{n-3}{2}}, \tag{105}$$

where $\Gamma(\cdot)$ denotes the Gamma function, and $x_+\triangleq\max\{0,x\}$ for all $x\in\mathbb{R}$. The support of H is $[-\sqrt{n},\sqrt{n}]$. From (105), we compute

$$\mathbb{E}\left[H\right]=0,\quad\mathrm{Var}\left[H\right]=1. \tag{106}$$

By Stirling's approximation, $H\to\mathcal{N}(0,1)$ in distribution as $n\to\infty$ (e.g., [43, Th. 1]). Recall that an upper bound on the total variation distance between P_H and $\mathcal{N}(0,1)$ is given in Lemma 5.

From (101), we find the conditional covariance matrix of the modified information density random vector as

$$
\begin{aligned}
\boldsymbol{\Sigma}(h)&\triangleq\mathrm{Cov}\left[\tilde{\boldsymbol{\imath}}_2|H=h\right] &(107)\\
&=\mathrm{Cov}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{J}_i(h)\right] &(108)\\
&=\boldsymbol{\Sigma}+\frac{h}{\sqrt{n}}\mathsf{B}, &(109)
\end{aligned}
$$

where

$$\boldsymbol{\Sigma}\triangleq\begin{bmatrix}V(P_1)&V_{1,2}(P_1,P_2)&V_{1,12}(P_1,P_2)\\V_{1,2}(P_1,P_2)&V(P_2)&V_{2,12}(P_1,P_2),\\V_{1,12}(P_1,P_2)&V_{2,12}(P_1,P_2)&V(P_1+P_2)\end{bmatrix} \tag{110}$$

$$\mathsf{B} \triangleq \frac{\sqrt{P_1 P_2}}{(1+P_1)(1+P_2)(1+P_1+P_2)} \cdot \begin{bmatrix} 0 & 1+P_1+P_2 & 1+P_2 \\ 1+P_1+P_2 & 0 & 1+P_1 \\ 1+P_2 & 1+P_1 & \frac{(1+P_1)(1+P_2)}{(1+P_1+P_2)} \end{bmatrix}, \quad (111)$$

and $V(P), V_{1,2}(P_1,P_2)$, and $V_{i,12}(P_1,P_2)$, $i \in [2]$, are given in (3), (18), and (19), respectively. Note that Σ and B depend only on P_1 and P_2. Using (102), (106), (109), by the law of total expectation and variance, we compute

$$\mathbb{E}\left[\tilde{\boldsymbol{\imath}}_2\right] = 0 \quad (112)$$

$$\mathrm{Cov}\left[\tilde{\boldsymbol{\imath}}_2\right] = \mathsf{V}(P_1,P_2), \quad (113)$$

where $\mathsf{V}(P_1,P_2)$ is the dispersion matrix defined in (17).

Step 2: We next approximate the distribution of $\tilde{\boldsymbol{\imath}}_2$ by a Gaussian. Toward that end, we consider some auxiliary random variables. Based on our observation in (101), we express the probability on the right-hand side of (95) by conditioning on H and taking the expectation with respect to P_H. Define the probability measure $P_{\tilde{H}}$, and the transition probability kernels $P_{\mathbf{V}|H}$ and $P_{\mathbf{W}|H}$ as

$$P_{\tilde{H}} \triangleq \mathcal{N}(0,1) \quad (114)$$

$$P_{\mathbf{V}|H=h} \triangleq \begin{cases} \mathcal{N}\left(\boldsymbol{\mu}(h), \Sigma(h)\right) & \text{if } |h| \le \sqrt{n} \\ \mathcal{N}\left(\boldsymbol{\mu}(h), \Sigma\right) & \text{if } |h| > \sqrt{n} \end{cases} \quad (115)$$

$$P_{\mathbf{W}|H=h} \triangleq \mathcal{N}\left(\boldsymbol{\mu}(h), \Sigma\right) \qquad \text{for } h \in (-\infty,\infty). \quad (116)$$

As with $P_{\mathbf{V}|H}$, we extend the definition of the kernel $P_{\tilde{\boldsymbol{\imath}}_2|H}$ given in (101) for $|H| > \sqrt{n}$ by choosing $P_{\tilde{\boldsymbol{\imath}}_2|H=h} = \mathcal{N}(\boldsymbol{\mu}(h),\Sigma)$ for $|h| > \sqrt{n}$ in order for the joint distribution $P_{\tilde{H}} P_{\tilde{\boldsymbol{\imath}}_2|H}$ to be valid. Recall that $\tilde{H}$ is a Gaussian random variable with the same mean and variance as H, and the mean and covariance matrix according to $P_{\mathbf{V}|H=h}$ are the same as those for $P_{\tilde{\boldsymbol{\imath}}_2|H=h}$. The Gaussian kernel $P_{\mathbf{W}|H}$ is obtained from $P_{\mathbf{V}|H}$ by replacing its covariance matrix $\Sigma(H)$ by the mean value of $\Sigma(H)$, Σ.

We define the joint distributions $P_{H\,\tilde{\boldsymbol{\imath}}_2}$, $P_{\tilde{H}\,\boldsymbol{\imath}_2^*}$, $P_{\tilde{H}\,\mathbf{V}}$ and $P_{\tilde{H}\,\mathbf{W}}$ as

$$P_{H\,\tilde{\boldsymbol{\imath}}_2} = P_H P_{\tilde{\boldsymbol{\imath}}_2|H} \quad (117a)$$

$$P_{\tilde{H}\,\boldsymbol{\imath}_2^*} = P_{\tilde{H}} P_{\tilde{\boldsymbol{\imath}}_2|H} \quad (117b)$$

$$P_{\tilde{H}\,\mathbf{V}} = P_{\tilde{H}} P_{\mathbf{V}|H} \quad (117c)$$

$$P_{\tilde{H}\,\mathbf{W}} = P_{\tilde{H}} P_{\mathbf{W}|H}, \quad (117d)$$

where

$$\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathsf{V}(P_1,P_2)), \quad (118)$$

which has the desired Gaussian distribution in our Berry-Esseen type bound.

Let $\mathcal{D}$ be any convex, Borel-measurable subset of $\mathbb{R}^3$. Then,

$$\left|\mathbb{P}\left[\tilde{\boldsymbol{\imath}}_2 \in \mathcal{D}\right] - \mathbb{P}\left[\mathbf{W} \in \mathcal{D}\right]\right| \quad (119a)$$

$$\le \left|\mathbb{P}\left[\tilde{\boldsymbol{\imath}}_2 \in \mathcal{D}\right] - \mathbb{P}\left[\boldsymbol{\imath}_2^* \in \mathcal{D}\right]\right| \quad (119b)$$

$$+ \left|\mathbb{P}\left[\boldsymbol{\imath}_2^* \in \mathcal{D}\right] - \mathbb{P}\left[\mathbf{V} \in \mathcal{D}\right]\right| \quad (119c)$$

$$+ \left|\mathbb{P}\left[\mathbf{V} \in \mathcal{D}\right] - \mathbb{P}\left[\mathbf{W} \in \mathcal{D}\right]\right|, \quad (119d)$$

where the inequality in (119b) follows from the triangle inequality. The absolute differences in (119b), (119c), and (119d) reflect the change of the input measure from P_H to $P_{\tilde{H}}$, the change of the transition probability kernel from $P_{\tilde{\boldsymbol{\imath}}_2|H}$ to $P_{\mathbf{V}|H}$, and the change of the transition probability kernel from $P_{\mathbf{V}|H}$ to $P_{\mathbf{W}|H}$, respectively. We next bound (119a) by showing that the absolute difference in each of (119b)–(119d) is $O\left(\frac{1}{\sqrt{n}}\right)$. In the next three steps, we bound each of these absolute differences in turn.

Step 3: We bound the absolute difference in the right-hand side of (119b) as

$$\left|\mathbb{P}\left[\tilde{\boldsymbol{\imath}}_2 \in \mathcal{D}\right] - \mathbb{P}\left[\boldsymbol{\imath}_2^* \in \mathcal{D}\right]\right|$$

$$= \left|\int_{-\infty}^{\infty} \mathbb{P}\left[\tilde{\boldsymbol{\imath}}_2 \in \mathcal{D} | H = h\right] \left(P_H(h) - P_{\tilde{H}}(h)\right) dh\right| \quad (120)$$

$$\le \int_{-\infty}^{\infty} \left|P_H(h) - P_{\tilde{H}}(h)\right| dh \quad (121)$$

$$= 2\,\mathrm{TV}(P_H, P_{\tilde{H}}) \quad (122)$$

$$\le 2\frac{\sqrt{n}}{\sqrt{n-3}} - 2 \quad (123)$$

$$\le \frac{C_{\mathrm{H}}}{n}, \quad (124)$$

where $C_{\mathrm{H}} = 8$. Inequality (121) follows by moving the absolute value to the inside of the integral and bounding the conditional probability by 1 for all h, and (123) holds for any $n \ge 4$ by Lemma 5. Inequality (124) holds for $n \ge 4$. We conclude that (124) holds for any n since (120) is trivially bounded by 1.

Step 4: We bound the absolute difference due to changing the transition probability kernel from $P_{\tilde{\boldsymbol{\imath}}_2|H}$ to the Gaussian kernel $P_{\mathbf{V}|H}$ as

$$\left|\mathbb{P}\left[\boldsymbol{\imath}_2^* \in \mathcal{D}\right] - \mathbb{P}\left[\mathbf{V} \in \mathcal{D}\right]\right|$$

$$= \left|\mathbb{E}\left[\mathbb{P}\left[\boldsymbol{\imath}_2^* \in \mathcal{D}\middle|\tilde{H}\right] - \mathbb{P}\left[\mathbf{V} \in \mathcal{D}\middle|\tilde{H}\right]\right]\right| \quad (125)$$

$$\le \mathbb{E}\left[\left|\mathbb{P}\left[\boldsymbol{\imath}_2^* \in \mathcal{D}\middle|\tilde{H}\right] - \mathbb{P}\left[\mathbf{V} \in \mathcal{D}\middle|\tilde{H}\right]\right| 1\left\{\left|\tilde{H}\right| \le \frac{\sqrt{n}}{2}\right\}\right]$$

$$+ \mathbb{P}\left[\left|\tilde{H}\right| > \frac{\sqrt{n}}{2}\right] \quad (126)$$

$$\le \max_{h \in \left[-\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right]} \frac{C(h)}{\sqrt{n}} + \mathbb{P}\left[\left|\tilde{H}\right| > \frac{\sqrt{n}}{2}\right] \quad (127)$$

$$\le \frac{C_{\mathrm{BE}}}{\sqrt{n}} + 2\exp\left\{-\frac{n}{8}\right\} \quad (128)$$

$$\le \frac{C_{\mathrm{BE}} + C_{\mathrm{Ch}}}{\sqrt{n}}, \quad (129)$$

where

$$T(h) \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\left\|\mathbf{J}_i(h)\right\|^3\right] \quad (130)$$

$$C(h) \triangleq \frac{c\,3^{1/4} T(h)}{\lambda_{\min}(\Sigma(h))^{3/2}} \quad (131)$$

$$C_{\mathrm{BE}} \triangleq \max_{h \in \left[-\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right]} C(h) \quad (132)$$

$$C_{\mathrm{Ch}} \triangleq 4\exp\left\{-\frac{1}{2}\right\}, \quad (133)$$

each $\mathbf{J}_i(h)$ is defined in (103), and c is the Berry-Esseen constant given in Theorem 6. Here, (126) moves the absolute

value in (125) to the inside of the expectation. We then separate the expectation into two cases in order to guarantee that we apply the Berry-Esseen theorem for values of h such that $\Sigma(h)$ is positive-definite. Inequality (127) follows from Corollary 2, and (128) follows from the Chernoff bound applied to a Gaussian random variable. Inequality (129) holds for any n. For every $h \in \left[-\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right]$, $\Sigma(h)$ is a non-degenerate covariance matrix, and $T(h) < \infty$. Therefore, we conclude that $C_{\text{BE}} < \infty$.

Step 5: We next bound the probability in (119d), which is the absolute difference due to changing the covariance matrix of the Gaussian kernel from $\Sigma(h)$ to Σ, using Lemma 4, which bounds the total variation distance between two Gaussian vectors. Denote the spectral radius of a $d \times d$ symmetric matrix M by

$$\rho(\mathsf{M}) \triangleq \max_{i \in [d]} |\lambda_i(\mathsf{M})|, \tag{134}$$

where $\lambda_i(\cdot)$ is the i-th largest eigenvalue of its matrix argument. Let

$$\mathsf{A} \triangleq \Sigma^{-1/2} \mathsf{B} \Sigma^{-1/2}, \tag{135}$$

where matrices Σ and B are defined in (110)–(111). Then

$$\begin{aligned}
&|\mathbb{P}[\mathbf{V} \in \mathcal{D}] - \mathbb{P}[\mathbf{W} \in \mathcal{D}]| \\
&= \left|\mathbb{E}\left[\mathbb{P}\left[\mathbf{V} \in \mathcal{D} \middle| \tilde{H}\right] - \mathbb{P}\left[\mathbf{W} \in \mathcal{D} \middle| \tilde{H}\right]\right]\right| && (136)\\
&\le \mathbb{E}\left[\left|\mathbb{P}\left[\mathbf{V} \in \mathcal{D} \middle| \tilde{H}\right] - \mathbb{P}\left[\mathbf{W} \in \mathcal{D} \middle| \tilde{H}\right]\right|\right] && (137)\\
&\le \mathbb{E}\left[\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}(\tilde{H}), \Sigma), \mathcal{N}(\boldsymbol{\mu}(\tilde{H}), \Sigma(\tilde{H})))\right] && (138)\\
&\le \frac{2+\sqrt{6}}{4} \|\mathsf{A}\|_F \frac{\mathbb{E}\left[\left|\tilde{H}\right|\right]}{\sqrt{n}}, && (139)
\end{aligned}$$

where (137) follows by moving the absolute value inside the expectation in (136), $\mu(\cdot)$ and $\Sigma(\cdot)$ are defined in (102) and (107), respectively, and (139) follows from Lemma 4.

The matrices Σ, $\Sigma + \mathsf{B}$, and $\Sigma - \mathsf{B}$ are all positive semidefinite as they are special cases of $\mathrm{Cov}[\tilde{\boldsymbol{\imath}}_2 | H = h]$ in (107) with h equal to $0, \sqrt{n}$, and $-\sqrt{n}$, respectively. Hence $\Sigma^{-1/2}(\Sigma+\mathsf{B})\Sigma^{-1/2}$ and $\Sigma^{-1/2}(\Sigma-\mathsf{B})\Sigma^{-1/2}$ are also positive semidefinite. Since their eigenvalues are respectively given by $1+\lambda_i(\mathsf{A})$ and $1-\lambda_i(\mathsf{A})$, it follows then that $-1 \le \lambda_i(\mathsf{A}) \le 1$ for $i \in [d]$, giving $\rho(\mathsf{A}) \le 1$.[2] Using the fact that $\|\mathsf{M}\|_F \le \sqrt{d}\rho(\mathsf{M})$ for any $d \times d$ symmetric matrix M, and employing the value of the expectation in (139), we conclude that

$$|\mathbb{P}[\mathbf{V} \in \mathcal{D}] - \mathbb{P}[\mathbf{W} \in \mathcal{D}]| \le \frac{C_{\text{G}}}{\sqrt{n}}, \tag{140}$$

where $C_{\text{G}} = \frac{2\sqrt{6}+6}{4\sqrt{\pi}}$.

Combining the bounds in (124), (129), and (140), we have the following Berry-Esseen-type inequality

$$|\mathbb{P}[\tilde{\boldsymbol{\imath}}_2 \in \mathcal{D}] - \mathbb{P}[\mathbf{W} \in \mathcal{D}]| \le \frac{C_{\text{H}} + C_{\text{BE}} + C_{\text{Ch}} + C_{\text{G}}}{\sqrt{n}} \tag{141}$$

for the modified information density random vector.

[2]Actually, $\rho(\mathsf{A}) = 1$. Indeed, for $h = \sqrt{n}$, the random variables in the first and the second index of the vectors in (103) are identical. Therefore, both $\Sigma(\sqrt{n}) = \Sigma + \mathsf{B}$ and $\Sigma^{-1/2}(\Sigma + \mathsf{B})\Sigma^{-1/2}$ have an eigenvalue 0, and A has an eigenvalue -1.

G. Completion of the Proof of Theorem 2

We employ the set $\mathcal{D} = \left\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} \ge \frac{1}{\sqrt{n}}\boldsymbol{\tau}\right\}$ in (141), where $\boldsymbol{\tau}$ is given in (93). Combining (95) and (141), we conclude that the probability $\mathbb{P}[\mathcal{A}^c]$ in (90) satisfies

$$\mathbb{P}[\mathcal{A}^c] \le 1 - \mathbb{P}\left[\mathbf{W} \ge \frac{1}{\sqrt{n}}\boldsymbol{\tau}\right] + \frac{C_{\text{H}} + C_{\text{BE}} + C_{\text{Ch}} + C_{\text{G}}}{\sqrt{n}} \tag{142}$$

$$= 1 - \mathbb{P}\left[\mathbf{W} \le -\frac{1}{\sqrt{n}}\boldsymbol{\tau}\right] + \frac{C_{\text{Out}}}{\sqrt{n}}, \tag{143}$$

where $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathsf{V}(P_1, P_2))$ and

$$C_{\text{Out}} \triangleq C_{\text{H}} + C_{\text{BE}} + C_{\text{Ch}} + C_{\text{G}}. \tag{144}$$

Equality (143) follows since $\mathbf{W} \sim -\mathbf{W}$. Suppose that $\boldsymbol{\tau}$ satisfies

$$-\frac{1}{\sqrt{n}}\boldsymbol{\tau} \in Q_{\text{inv}}(\mathsf{V}(P_1, P_2), \epsilon - \gamma_n) \tag{145}$$

$$\gamma_n \triangleq \exp\left\{-c_2 n^{1/3}\right\} + \frac{1 + C_{\text{Out}}}{\sqrt{n}}, \tag{146}$$

where the constant c_2 is as in (90). Then, the right-hand side of (90) is bounded by ϵ. From the Taylor series expansion of $Q_{\text{inv}}(\mathsf{V}, \cdot)$ (e.g., [33, Lemma 13]), we conclude that (145) is equivalent to the inequality in (21), which completes the proof.

VI. Proof of Theorem 3

In this section, we sketch the proof of Theorem 3 by detailing the modifications to generalize the proof of Theorem 2 from 2 to $K \ge 2$ transmitters. Assume that $\mathcal{S} \in \mathcal{P}([K])$. Define the information densities as

$$\imath_{\mathcal{S}}(\mathbf{x}_{\mathcal{S}}; \mathbf{y} | \mathbf{x}_{\mathcal{S}^c}) \triangleq \log \frac{P_{\mathbf{Y}_K | \mathbf{X}_{[K]}}(\mathbf{y} | \mathbf{x}_{[K]})}{P_{\mathbf{Y}_K | \mathbf{X}_{\mathcal{S}^c}}(\mathbf{y} | \mathbf{x}_{\mathcal{S}^c})}, \tag{147}$$

where $\mathcal{S}^c = [K] \setminus \mathcal{S}$. The information density random vector for K transmitters is

$$\boldsymbol{\imath}_K \triangleq (\imath_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}; \mathbf{Y}_K | \mathbf{X}_{\mathcal{S}^c}) \colon \mathcal{S} \in \mathcal{P}([K])) \in \mathbb{R}^{2^K - 1}, \tag{148}$$

where $\mathbf{X}_k$ is distributed uniformly on $\mathbb{S}^n(\sqrt{nP_k})$ for $k \in [K]$, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{I}_n)$, $\mathbf{X}_1, \ldots, \mathbf{X}_K$ and $\mathbf{Z}$ are independent, and $\mathbf{Y}_K = \mathbf{X}_{\langle [K] \rangle} + \mathbf{Z}$.

Below, we use Lemma 1 and the generalization of Lemma 6 given in (152). The following lemma, which generalizes Lemma 5 to K transmitters, is the critical part of the proof of Theorem 3.

Lemma 7: Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{in})$, $i = 1, \ldots, K$, be K independent random vectors, distributed uniformly on $\mathbb{S}^n(1)$. Let $H_{ij} = \sqrt{n}\langle \mathbf{X}_i, \mathbf{X}_j \rangle$ for $1 \le i < j \le K$, and $\mathbf{H} = (H_{ij} \colon 1 \le i < j \le K)$. Then

$$\mathrm{TV}\left(P_{\mathbf{H}}, \mathcal{N}\left(\mathbf{0}, \mathsf{I}_{\frac{K(K-1)}{2}}\right)\right) \le \frac{C_K}{\sqrt{n}} \tag{149}$$

for some constant C_K depending only on K.

Proof: See Appendix E. ∎

The modifications in Section V are as follows.

1) The two-transmitter maximum likelihood decoder given in (51) is replaced by a K-transmitter maximum likelihood decoder, which chooses the message vector $m_{[K]} =$

$(m_1, \ldots, m_K)$ corresponding to the maximal information density $\imath_{[K]}(\mathsf{f}_{[K]}(m_{[K]}); \mathbf{y})$.

2) The typical set $\mathcal{F}$ defined in (56) is replaced by

$$\mathcal{F}_K \triangleq \bigtimes_{\mathcal{S}\in\mathcal{P}([K])} \mathcal{F}(\mathcal{S}) \subseteq \mathbb{R}^{n\otimes(2^K-1)}, \tag{150}$$

where $\mathcal{F}(\mathcal{S})$ is defined in (57). Inequality (59) extends to $\mathcal{F}_K$ by Lemma 1.

3) The functions given in (62)–(64) are extended as

$$g_{\mathcal{S}}(t; \mathbf{y}, \mathbf{x}_{\mathcal{S}^c}) \triangleq \mathbb{P}\big[\imath_{\mathcal{S}}(\bar{\mathbf{X}}_{\mathcal{S}}; \mathbf{Y}_K|\mathbf{X}_{\mathcal{S}^c}) \geq t \mid \mathbf{X}_{\mathcal{S}^c} = \mathbf{x}_{\mathcal{S}^c}, \mathbf{Y}_K = \mathbf{y}\big]. \tag{151}$$

In the proof of Lemma 6, we replace $P_1 + P_2$ by $P_{\langle\mathcal{S}\rangle}$, and P_1P_2 by $\sum_{\substack{i,j\in[K]\\ i<j}} P_iP_j$. Inequality (73) generalizes to the K-transmitter MAC by applying its proof from Appendix D with Lemma 1 from Section V-A. Hence, Lemma 6 generalizes as

$$g_{\mathcal{S}}(t; \mathbf{y}, \mathbf{x}_{\mathcal{S}^c}) \leq \frac{G(\mathcal{S}) \exp\{-t\}}{\sqrt{n}}, \tag{152}$$

where $G(\mathcal{S})$ is a constant depending only on the powers $(P_s : s \in \mathcal{S})$.

4) The high probability events given in (82) and (83) are replaced by

$$\mathcal{E}_K \triangleq \bigcap_{\mathcal{S}\in\mathcal{P}([K])} \mathcal{E}(\mathcal{S}), \tag{153}$$

$$\mathcal{A}_K \triangleq \bigg\{\boldsymbol{\imath}_K \geq \bigg(\log\bigg(\Big(\prod_{s\in\mathcal{S}} M_s\Big)(G(\mathcal{S})^2)\alpha_{|\mathcal{S}|,K}\bigg) : \mathcal{S} \in \mathcal{P}([K])\bigg) - \frac{1}{2}\log n\mathbf{1}\bigg\}, \tag{154}$$

where

$$\alpha_{s,K} \triangleq K\binom{K}{s}, \quad s = 1, \ldots, K. \tag{155}$$

Using the extension of the RCU bound for K transmitters given in Remark 1 and following the same steps as Section V-E, we replace the right-hand side of the inequality in (90) by

$$\mathbb{P}\left[\mathcal{A}_K^c\right] + \exp\left\{-c_K n^{1/3}\right\} + \frac{1}{\sqrt{n}}, \tag{156}$$

where c_K is a constant.

5) To understand the differences between bounding $\mathbb{P}\left[\mathcal{A}_K^c\right]$ and $\mathbb{P}\left[\mathcal{A}^c\right]$, we first extend the definition of the modified and centered information density random vector to K transmitters by defining

$$\tilde{\imath}_{\mathcal{S}}(\mathbf{x}_{\mathcal{S}}; \mathbf{y}_K|\mathbf{x}_{\mathcal{S}^c}) \triangleq \sum_{i=1}^{n} \log \frac{P_{Y_K|X_{[K]}}(y_i|x_{[K]i})}{P_{\tilde{Y}_K|\tilde{X}_{\mathcal{S}^c}}(y_i|x_{\mathcal{S}^c i})} \tag{157}$$

$$\tilde{\boldsymbol{\imath}}_K \triangleq \frac{1}{\sqrt{n}}\big[(\tilde{\imath}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}; \mathbf{Y}_K|\mathbf{X}_{\mathcal{S}^c}) : \mathcal{S} \in \mathcal{P}([K])) - n\mathbf{C}(P_{[K]})\big], \tag{158}$$

where $\mathbf{C}(P_{[K]})$ is the capacity vector defined in (23), $\tilde{X}_k \sim \mathcal{N}(0, P_k)$ for $k \in [K]$, and $\prod_{k=1}^{K} P_{\tilde{X}_k} \to P_{Y_K|X_{[K]}} \to P_{\tilde{Y}_K} = \mathcal{N}(0, 1 + P_{[K]})$.

We replace the threshold value in (93) by

$$\boldsymbol{\tau} \triangleq \log\bigg(\frac{\big(\prod_{s\in\mathcal{S}} M_s\big)\,(G(\mathcal{S}))^2\kappa_{|\mathcal{S}|}(P_{\mathcal{S}})\alpha_{|\mathcal{S}|,K}}{\sqrt{n}} : \mathcal{S} \in \mathcal{P}([K])\bigg) - n\mathbf{C}(P_{[K]}), \tag{159}$$

where $\kappa_{|\mathcal{S}|}(P_{\mathcal{S}})$ is the constant (which depends only on $P_{\mathcal{S}}$) in (39). Using the joint distribution of $(\mathbf{X}_{[K]}, \mathbf{Y}_K)$, we get

$$\tilde{\boldsymbol{\imath}}_K \sim \frac{1}{\sqrt{n}}\bigg(\frac{(n - \|\mathbf{Z}\|^2)P_{\langle\mathcal{S}\rangle}}{2(1 + P_{\langle\mathcal{S}\rangle})} + \frac{\sum_{\substack{i,j\in\mathcal{S}\\ i<j}}\langle\mathbf{X}_i, \mathbf{X}_j\rangle + \langle\mathbf{Z}, \mathbf{X}_{\langle\mathcal{S}\rangle}\rangle}{1 + P_{\langle\mathcal{S}\rangle}} : \mathcal{S} \in \mathcal{P}([K])\bigg). \tag{160}$$

Define the random vector

$$\mathbf{H} \triangleq (H_{ij} \colon 1 \leq i < j \leq K) \in \mathbb{R}^{\binom{K}{2}}, \tag{161}$$

where $H_{ij} = \frac{\langle\mathbf{X}_i, \mathbf{X}_j\rangle}{\sqrt{nP_iP_j}}$ denotes the normalized inner product of $\mathbf{X}_i$ and $\mathbf{X}_j$. The inner product random vector $\mathbf{H}$ replaces H in (104). Observe that for all different (i_1, j_1) and (i_2, j_2) pairs, $H_{i_1j_1}$ and $H_{i_2j_2}$ are independent of each other, which follows by independence of $\mathbf{X}_1, \ldots, \mathbf{X}_K$. However, $\mathbf{H}$ does not have a product distribution due to the fact that any triplets in $\mathbf{H}$ are not jointly independent.[3] While $P_{\mathbf{H}}$ is not a product distribution, Lemma 7 implies that $P_{\mathbf{H}}$ converges to the distribution of $\binom{K}{2}$ i.i.d. standard Gaussian random variables in total variation, allowing us to use the Berry-Esseen theorem just as we did for the two-transmitter MAC.

As for the two-transmitter MAC, the distribution in (160) depends on $\mathbf{X}_{[K]}$ only through the inner product random vector $\mathbf{H}$. The conditional distribution of $\tilde{\boldsymbol{\imath}}_K$ given $\mathbf{H} = \mathbf{h}$ is the same as the conditional distribution of

$$\boldsymbol{\mu}(\mathbf{h}) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbf{J}_i(\mathbf{h}), \tag{162}$$

where

$$\boldsymbol{\mu}(\mathbf{h}) \triangleq \mathbb{E}\left[\boldsymbol{\imath}_K|\mathbf{H} = \mathbf{h}\right] = \sum_{\substack{i,j\in[K]\\ i<j}} h_{ij}\Big(\frac{\sqrt{P_iP_j}}{1 + P_{\langle\mathcal{S}\rangle}} 1\left\{i, j \in \mathcal{S}\right\} : \mathcal{S} \in \mathcal{P}([K])\Big) \tag{163}$$

$$\mathbf{J}_i(\mathbf{h}) \triangleq \Big(\frac{(1 - Z_i^2)P_{\langle\mathcal{S}\rangle} + 2\sum_{s\in\mathcal{S}} x_{si}Z_i}{2(1 + P_{\langle\mathcal{S}\rangle})} : \mathcal{S} \in \mathcal{P}([K])\Big) \tag{164}$$

for $i \in [n]$, and $\mathbf{x}_{[K]}$ are vectors on the n-dimensional power spheres, satisfying $\frac{\langle\mathbf{x}_i, \mathbf{x}_j\rangle}{\sqrt{nP_iP_j}} = h_{ij}$ for all $i < j \in$

[3] Given that $H_{12} = H_{13} = \sqrt{n}$, we have that $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_3$. Therefore, H_{23} is necessarily equal to $\sqrt{n}$ under this condition, and H_{12}, H_{13}, H_{23} are not jointly independent.

$[K]$. The conditional covariance matrix given in (109) is extended to K transmitters as

$$\Sigma(\mathbf{h}) = \mathrm{Cov}\left[\tilde{\boldsymbol{\imath}}_K|\mathbf{H}=\mathbf{h}\right] = \Sigma_K + \sum_{i,j\in[K],i<j} \frac{h_{ij}}{\sqrt{n}}\mathsf{B}_{ij}, \tag{165}$$

where the $\left(\mathbb{R}^{2^K-1}\right) \times \left(\mathbb{R}^{2^K-1}\right)$ matrices Σ_K and B_{ij} have elements

$$\Sigma_{\mathcal{S}_1\mathcal{S}_2} = \frac{P_{\mathcal{S}_1}P_{\mathcal{S}_2} + 2P_{\mathcal{S}_1\cap\mathcal{S}_2}}{2(1+P_{\mathcal{S}_1})(1+P_{\mathcal{S}_2})} \tag{166}$$

$$b_{\mathcal{S}_1\mathcal{S}_2} = \frac{\sqrt{P_iP_j}}{(1+P_{\mathcal{S}_1})(1+P_{\mathcal{S}_2})} \cdot 1\left\{\{i\in\mathcal{S}_1, j\in\mathcal{S}_2\}\cup\{i\in\mathcal{S}_2, j\in\mathcal{S}_1\}\right\} \tag{167}$$

for $\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{P}([K])$. These formulas generalize the formulas for the two-transmitter MAC given in (110) and (111). By (163), (165), and the pairwise independence of $H_{i_1j_1}$, $H_{i_2j_2}$ for all different (i_1, j_1) and (i_2, j_2) pairs, using the law of total expectation and variance, we find that

$$\mathbb{E}\left[\tilde{\boldsymbol{\imath}}_K\right] = \mathbf{0} \tag{168}$$

$$\mathrm{Cov}\left[\tilde{\boldsymbol{\imath}}_K\right] = \mathsf{V}(P_{[K]}), \tag{169}$$

where the covariance matrix $\mathsf{V}(P_{[K]})$ is defined in (24). The rest of the proof follows the proof in Section V-F, where we replace H by $\mathbf{H}$, $\tilde{H}$ by the $\binom{K}{2}$-dimensional standard Gaussian random vector $\tilde{\mathbf{H}}$, $P_{\tilde{\boldsymbol{\imath}}_2|H}$ by $P_{\tilde{\boldsymbol{\imath}}_K|\mathbf{H}}$, $P_{\mathbf{V}|H}$ by $P_{\mathbf{V}|\mathbf{H}}$, and $P_{\mathbf{W}|H}$ by $P_{\mathbf{W}|\mathbf{H}}$. For the probability transition kernels $P_{\mathbf{V}|\mathbf{H}}$ and $P_{\mathbf{W}|\mathbf{H}}$, we replace $\boldsymbol{\mu}(h)$ by $\boldsymbol{\mu}(\mathbf{h})$, Σ by Σ_K, and $\Sigma(h)$ by $\Sigma(\mathbf{h})$. We replace all conditions in the form $|h| \le t$ by $|\mathbf{h}| \le t\mathbf{1}$.

The only critical modification is that the bound on the total variation distance $\mathrm{TV}(P_H, P_{\tilde{H}})$ in (123) is replaced by the bound on the total variation distance $\mathrm{TV}(P_{\mathbf{H}}, P_{\tilde{\mathbf{H}}})$, which is $O\left(\frac{1}{\sqrt{n}}\right)$ by Lemma 7. We conclude that

$$\left|\mathbb{P}\left[\tilde{\boldsymbol{\imath}}_K \in \mathcal{D}\right] - \mathbb{P}\left[\mathbf{W}\in\mathcal{D}\right]\right| \le \frac{C_K}{\sqrt{n}} \tag{170}$$

for some constant $C_K > 0$, where $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathsf{V}(P_{[K]}))$. By combining (156) and (170) as in Section V-G, we complete the proof of Theorem 3.

VII. Proof of Theorem 5

The main difference between the coding strategies for the Gaussian MAC and RAC is that for the Gaussian RAC, an output typicality condition is added to the decoding function in order to reliably detect the number of active transmitters.

A. Encoding and Decoding

Encoding: Recall that n_K is the largest decoding time. In our encoding strategy, rather than adapting the codebook to the estimate of the number of active transmitters at the receiver, we generate codewords with length n_K. Each active transmitter transmits one symbol of its message codeword at each time step until the decoder signals at time $n_k \in \{n_0, \ldots, n_K\}$ that it is able to decode. If decoding happens at time n_k, only the initial sub-codeword of length n_k is used.

The common randomness random variable $U \in \mathbb{R}^{Mn_K}$ has the distribution

$$P_U = \underbrace{P_{U(1)} \times P_{U(2)} \times \cdots \times P_{U(M)}}_{M\text{times}}, \tag{171}$$

where $P_{U(m)} = P_{\mathbf{X}}$ for $m \in [M]$. The realization of U defines M length-n_K i.i.d. codewords. In other words, the encoding function is given by

$$\mathsf{f}(U, m) = U(m), \quad m \in [M]. \tag{172}$$

The need for using common randomness in encoding is due to the requirement that a single code must satisfy multiple constraints, i.e., the error probability constraints in (29). Showing that a random code satisfies multiple constraints does not imply the existence of a deterministic code. This issue arises in [25], where a variable-length feedback code must satisfy both an average decoding time and an average-error constraint, and in [21], where a RAC code must satisfy $K+1$ average error probability constraints. See [21, Sec. II.C] for details.

Decoding: Unlike the MAC, for the Gaussian RAC, we require the decoder to determine the time $n_k \in \{n_0, \ldots, n_K\}$ at which to decode. Therefore, we couple the maximum likelihood decoder given in (51) with a threshold rule, used to estimate the number of transmitters and a single bit of feedback at each time n_i up to and including the time n_k at which the decoder decides to decode. The maximum likelihood decoder is applied only if the threshold test is satisfied. Here, the role of the threshold rule is to reliably determine the true channel in the communication epoch. We use a threshold rule to determine the number of active transmitters because for any $P > 0$, under an input distribution $P_{\mathbf{X}}$ such that the expected input power meets the power constraint in (28) with equality (i.e., $\frac{1}{n_k}\mathbb{E}\left[\left\|\mathbf{X}^{[n_k]}\right\|^2\right] = P$), for each k, the normalized squared norm of the output $\mathbf{Y}_k^{[n_k]}$ concentrates around its mean. That mean is different for each $k \in \{0, 1, \ldots, K\}$; specifically

$$\frac{1}{n_k}\mathbb{E}\left[\left\|\mathbf{Y}_k^{[n_k]}\right\|^2\right] = 1 + kP, \quad \forall k \in \{0\}\cup[K]. \tag{173}$$

Upon receiving the first n_0 symbols of the output, $\mathbf{y}^{[n_0]}$, the decoder computes the following function

$$\mathsf{g}_0(U, \mathbf{y}^{[n_0]}) = \begin{cases} 0 & \text{if } \left|\frac{1}{n_0}\left\|\mathbf{y}^{[n_0]}\right\|^2 - 1\right| \le \lambda_0 \\ \mathsf{e} & \text{otherwise} \end{cases} \tag{174}$$

to decide whether there are any active transmitters; here λ_0 is a parameter that is determined by the error criterion ϵ_0. At time n_0, if $\mathsf{g}_0(U, \mathbf{y}^{[n_0]}) = 0$, the receiver broadcasts a bit value 1 to all transmitters, signaling that the receiver estimates "no active transmitters" and the epoch ends. Otherwise the receiver broadcasts a bit value 0 and the epoch continues.

For $k \geq 1$, the decoder applies the following function to make a decision at each subsequent time $n_k \leq n_K$

$$\mathsf{g}_k(U, \mathbf{y}^{[n_k]}) = \begin{cases} m_{[k]} & \text{if } \imath_{[k]}(\mathsf{f}(U, m_{[k]})^{[n_k]}; \mathbf{y}^{[n_k]}) \\ & \quad > \imath_{[k]}(\mathsf{f}(U, m'_{[k]})^{[n_k]}; \mathbf{y}^{[n_k]}) \\ & \quad \text{for all } m'_{[k]} \overset{\pi}{\neq} m_{[k]}, \\ & m_1 \leq \cdots \leq m_k, \\ & \left| \frac{1}{n_k} \left\| \mathbf{y}^{[n_k]} \right\|^2 - (1 + kP) \right| \leq \lambda_k \\ \mathsf{e} & \text{otherwise,} \end{cases} \tag{175}$$

where λ_k is a parameter chosen to satisfy the error criterion ϵ_k. At time n_k, if $\mathsf{g}_k(U, \mathbf{y}^{[n_k]}) \neq \mathsf{e}$ or $k = K$, then the receiver broadcasts the bit value 1 to all transmitters, signaling the end of epoch and the start of next one. Otherwise, the receiver sends feedback 0 and the epoch continues.

By the permutation-invariance of the channel in terms of the inputs $\mathbf{X}_{[k]}$ and the identical encoding in (172), all permutations of the messages $m_{[k]}$ give the same information density. Therefore, without loss of generality, the output of our decoder is always the ordered message vector in (175). The condition $\left| \frac{1}{n_k} \left\| \mathbf{y}^{[n_k]} \right\|^2 - (1 + kP) \right| \leq \lambda_k$, which does not depend on the randomly generated codebook, allows us with high probability to decode at time n_k when the number of active transmitters is k, rather than decoding earlier or failing to decode at the time n_k intended for the k-transmitter scenario.

B. Error Analysis

In this section, we bound the probability of error for the random access code in Definition 3.

No active transmitters: For $k = 0$, the only error event is that the squared norm of the output $\mathbf{Y}_0^{[n_0]}$ is away from its mean:

$$\epsilon_0 \leq \mathbb{P}\left[\left| \frac{1}{n_0} \left\| \mathbf{Y}_0^{[n_0]} \right\|^2 - 1 \right| > \lambda_0 \right]. \tag{176}$$

$k \geq 1$ *active transmitters*: When there is at least one active transmitter, the encoding function (172) and decoding rule (175) yield an error if and only if at least one of the following events occurs:

- $\mathcal{E}_{\text{codeword}}$: At least one of the k codewords associated with the sent messages $m_{[k]}$ violates the power constraint in (28) in the first n_k symbols. In this case, an error occurs since it is forbidden to transmit those codewords. We do not need to include the power constraint violation beyond the n_k-th symbol since that event is captured by the event of decoding time error, stated next.
- $\mathcal{E}_{\text{time}}$: A list of messages is decoded at a wrong decoding time $n_t \neq n_k$, or no messages is decoded during the entire epoch.
- $\mathcal{E}_{\text{message}}$: A list of messages $m'_{[k]} \neq m_{[k]}$ is decoded at time n_k.

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

Since we are employing identical encoders at all encoders, we simplify the analysis by treating the event $\mathcal{E}_{\text{rep}} = \{W_i = W_j \text{ for some } i \neq j\}$ that at least one message among transmitted messages is repeated as an error. While this case is actually advantageous to decoding, it requires special treatment since it violates the assumption of codeword independence employed in our analysis.

By the union bound,

$$\mathbb{P}\left[\mathcal{E}_{\text{rep}}\right] \leq \frac{k(k-1)}{2M}. \tag{177}$$

Applying the union bound, we bound the error probability as

$$\begin{aligned} \epsilon_k = \frac{1}{M^k} \sum_{m_{[k]} \in [M]^k} & \mathbb{P}\Bigg[\bigcup_{t:\ n_t \leq n_k, t \neq k} \left\{ \mathsf{g}_t(U, \mathbf{Y}_k^{[n_t]}) \neq \mathsf{e} \right\} \\ & \bigcup \left\{ \mathsf{g}_k(U, \mathbf{Y}_k^{[n_k]}) \overset{\pi}{\neq} m_{[k]} \right\} \Big| \ W_{[k]} = m_{[k]} \Bigg] \qquad (178) \\ \leq \mathbb{P}\left[\mathcal{E}_{\text{rep}}\right] + \mathbb{P}\left[\mathcal{E}_{\text{rep}}^c\right] & \Big(\mathbb{P}\left[\mathcal{E}_{\text{codeword}} \middle| \mathcal{E}_{\text{rep}}^c\right] \qquad (179) \\ & + \mathbb{P}\left[\mathcal{E}_{\text{time}} \middle| \mathcal{E}_{\text{rep}}^c\right] + \mathbb{P}\left[\mathcal{E}_{\text{message}} \middle| \mathcal{E}_{\text{rep}}^c\right] \Big) \qquad (180) \\ \leq \mathbb{P}\left[\mathcal{E}_{\text{rep}}\right] + \mathbb{P}\left[\mathcal{E}_{\text{codeword}} \middle| \mathcal{E}_{\text{rep}}^c\right] & \\ + \mathbb{P}\left[\mathcal{E}_{\text{time}} \middle| \mathcal{E}_{\text{rep}}^c\right] + \mathbb{P}\left[\mathcal{E}_{\text{message}} \middle| \mathcal{E}_{\text{rep}}^c\right]. & \qquad (181) \end{aligned}$$

Power constraint violation: The probability that a power constraint violation occurs in the first n_k symbols for at least one of the k distinct messages is

$$\mathbb{P}\left[\mathcal{E}_{\text{codeword}} \middle| \mathcal{E}_{\text{rep}}^c\right] = \mathbb{P}\left[\bigcup_{i=1}^{k} \bigcup_{\substack{j: n_j \leq n_k \\ j \geq 1}} \left\{ \frac{1}{n_j} \left\| \mathbf{X}_i^{[n_j]} \right\|^2 > P \right\} \right]. \tag{182}$$

Wrong decoding time: According to the decoding rule in (175), decoding occurs at time n_k if and only if the output typicality criterion is not satisfied for any t with $n_t \leq n_k$ and $t \neq k$ (that is $\left| \frac{1}{n_t} \left\| \mathbf{y}^{[n_t]} \right\|^2 - (1 + tP) \right| > \lambda_t$), and is satisfied for k (that is $\left| \frac{1}{n_k} \left\| \mathbf{y}^{[n_k]} \right\|^2 - (1 + kP) \right| \leq \lambda_k$). Note that it is possible that no message set is decoded during an entire epoch. This would happen if $\left| \frac{1}{n_t} \left\| \mathbf{y}^{[n_t]} \right\|^2 - (1 + tP) \right| > \lambda_t$ for $t \in \{0, \ldots, K\}$. The probability $\mathbb{P}\left[\mathcal{E}_{\text{time}} \middle| \mathcal{E}_{\text{rep}}^c\right]$ is computed as

$$\begin{aligned} \mathbb{P}\left[\mathcal{E}_{\text{time}} \middle| \mathcal{E}_{\text{rep}}^c\right] = \mathbb{P}\Bigg[& \bigcup_{\substack{t: n_t \leq n_k \\ t \neq k}} \left\{ \left| \frac{1}{n_t} \left\| \mathbf{Y}_k^{[n_t]} \right\|^2 - (1 + tP) \right| \leq \lambda_t \right\} \\ & \bigcup \left\{ \left| \frac{1}{n_k} \left\| \mathbf{Y}_k^{[n_k]} \right\|^2 - (1 + kP) \right| > \lambda_k \right\} \Bigg]. \end{aligned} \tag{183}$$

Wrong message: By using the RCU bound in Remark 1 and the permutation-invariance of the information density, we bound $\mathbb{P}\left[\mathcal{E}_{\text{message}} \middle| \mathcal{E}_{\text{rep}}^c\right]$ as

$$\begin{aligned} \mathbb{P}\left[\mathcal{E}_{\text{message}} \middle| \mathcal{E}_{\text{rep}}^c\right] \leq \mathbb{E}\Bigg[& \min\left\{ 1, \sum_{s=1}^{k} \binom{k}{s} \binom{M-k}{s} \right. \\ & \mathbb{P}\Big[\imath_{[s]}(\bar{\mathbf{X}}_{[s]}^{[n_k]}; \mathbf{Y}_k^{[n_k]} | \mathbf{X}_{[s+1:k]}^{[n_k]}) \\ & \left. \geq \imath_{[s]}(\mathbf{X}_{[s]}^{[n_k]}; \mathbf{Y}_k^{[n_k]} | \mathbf{X}_{[s+1:k]}^{[n_k]}) \ \Big| \ \mathbf{X}_{[k]}^{[n_k]}, \mathbf{Y}_k^{[n_k]} \Big] \right\} \Bigg]. \end{aligned} \tag{184}$$

Combining (176), (177) and (181)–(184) completes the proof. Note that compared to the achievability proof of the Gaussian MAC in (11), the multiplicative constant in (184) is $\binom{M-k}{s}$ instead of $(M-1)^s$ since we are given that the transmitted messages are distinct.

VIII. Proof of Theorem 4

In this section, we analyze the achievability bound in Theorem 5 by particularizing the input distribution, $P_{\mathbf{X}}$ in Theorem 5, choosing the free parameters λ_k, decoding times $n_0, n_1, \ldots, n_K$, and bounding the probability and expectation terms in (34). In the rest of the proof, we assume that the decoding times satisfy $n_0 < n_1 < \cdots < n_K$, which we make explicit in (211).

A. Particularizing $P_{\mathbf{X}}$

We modify the input distribution used in Theorem 2 for the Gaussian MAC so that the randomly generated codewords meet the power constraints with probability 1.

A random codeword distributed according to $P_{\mathbf{X}}$ has length n_K and consists of K independent sub-codewords. The j-th sub-codeword has length $|\mathcal{N}(j)|$, where

$$\mathcal{N}(j) \triangleq \begin{cases} [n_1] & \text{if } j = 1 \\ \{n_{j-1}+1, n_{j-1}+2, \ldots, n_j\} & \text{if } 2 \le j \le K \end{cases} \tag{185}$$

for $j \in [K]$ is the index set for the j-th block in our code design. Thus, the input distribution $P_{\mathbf{X}}$ in Theorem 5 is

$$P_{\mathbf{X}}(\mathbf{x}) = \prod_{j=1}^{K} P_{\mathbf{X}^{\mathcal{N}(j)}}\left(\mathbf{x}^{\mathcal{N}(j)}\right), \tag{186}$$

where

$$P_{\mathbf{X}^{\mathcal{N}(j)}}\left(\mathbf{x}^{\mathcal{N}(j)}\right) = \frac{\delta\left(\left\|\mathbf{x}^{\mathcal{N}(j)}\right\|^2 - |\mathcal{N}(j)|P\right)}{S_{|\mathcal{N}(j)|}(\sqrt{|\mathcal{N}(j)|P})}, \tag{187}$$

that is, $\mathbf{X}^{\mathcal{N}(j)} \sim \text{Uniform}\left(\mathbb{S}^{|\mathcal{N}(j)|}(\sqrt{|\mathcal{N}(j)|P})\right)$, and $\mathbf{X}^{\mathcal{N}(1)}, \ldots, \mathbf{X}^{\mathcal{N}(K)}$ are independent.

Codewords chosen according to (186) satisfy the power constraints in (28) with equality, giving

$$\mathbb{P}\left[\bigcup_{i=1}^{k}\bigcup_{j=1}^{k}\left\{\frac{1}{n_j}\left\|\mathbf{X}_i^{[n_j]}\right\|^2 > P\right\}\right] = 0. \tag{188}$$

B. Error Analysis

We separate the analysis into 3 steps: deriving an output typicality bound, evaluation of the RCU bound, and evaluation of a Berry-Esseen type inequality.

Step 1: In this step, we bound the probability that the output $\mathbf{Y}_k^{[n_k]}$ does not satisfy the condition $\left|\frac{1}{n_k}\left\|\mathbf{Y}_k^{[n_k]}\right\|^2 - (1+kP)\right| \le \lambda_k$ given in the decoding rule (175). Since for $k \ge 1$, $\mathbf{Y}_k^{\mathcal{N}(1)}, \mathbf{Y}_k^{\mathcal{N}(2)}, \ldots, \mathbf{Y}_k^{\mathcal{N}(K)}$ are independent due to the input distribution in (186), Lemma 1 and Lemma 2 imply

$$\begin{aligned}&\mathbb{P}\left[\left|\left\|\mathbf{Y}_k^{[n_k]}\right\|^2 - n_k(1+kP)\right| > n_k\lambda_k\right] \\ &\le 2\left(\kappa_k(P\mathbf{1})\right)^k \exp\left\{-\frac{n_k\lambda_k^2}{8(1+kP)^2}\right\}\end{aligned} \tag{189}$$

for $\lambda_k \in (0, 1+kP)$, where $\kappa_j(P\mathbf{1})$ is the constant defined in Lemma 1. For $k = 0$, we have

$$\mathbb{P}\left[\left|\left\|\mathbf{Y}_0^{[n_0]}\right\|^2 - n_0\right| > n_0\lambda_0\right] \le 2\exp\left\{-\frac{n_0\lambda_0^2}{8}\right\} \tag{190}$$

for $\lambda_0 \in (0,1)$. We pick

$$\lambda_0 = \sqrt{\frac{-8\log\frac{\epsilon_0}{2}}{n_0}} \tag{191}$$

to ensure that the right-hand side of (190) is bounded above by ϵ_0. By setting $\lambda_t = \frac{P}{2}$ for $t \ge 1$, using (189) and (190), and applying the union bound, we bound the probability of decoding time error in (34b) by

$$\begin{aligned}B \triangleq\ & 2\kappa_1(P)\exp\left\{-\frac{n_0((k-\frac{\lambda_0}{P})P)^2}{8(1+kP)^2}\right\} \\ &+ 2\sum_{t=1}^{k}\left(\kappa_k(P\mathbf{1})\right)^t \exp\left\{-\frac{n_t((k-t-\frac{1}{2})P)^2}{8(1+kP)^2}\right\}.\end{aligned} \tag{192}$$

Step 2: To bound the expectation in (34c), we first modify the definition of the typical output set $\mathcal{F}(\mathcal{S})$ in (57) as

$$\begin{aligned}\mathcal{F}(\mathcal{S})_{\text{RAC}} \triangleq \Big\{&\mathbf{y}^{[n_k]} \in \mathbb{R}^{n_k} : \\ &\frac{1}{|\mathcal{N}(j)|}\left\|\mathbf{y}^{\mathcal{N}(j)}\right\|^2 \in \mathcal{I}(j,\mathcal{S}) \text{ for } j \in [k]\Big\}.\end{aligned} \tag{193}$$

$$\begin{aligned}\mathcal{I}(j,\mathcal{S}) \triangleq [&1 + |\mathcal{S}|P - |\mathcal{N}(j)|^{-1/3}, \\ &1 + |\mathcal{S}|P + |\mathcal{N}(j)|^{-1/3}].\end{aligned} \tag{194}$$

We then show that Lemma 6 holds under input distribution (186) with typical output set (193). That is, for every $0 < s \le k$, and $\mathbf{y}^{[n_k]}$ and $\mathbf{x}_{[k]\setminus[s]}^{[n_k]}$ such that $\mathbf{y}^{[n_k]} - \mathbf{x}_{\langle[k]\setminus[s]\rangle}^{[n_k]} \in \mathcal{F}([s])_{\text{RAC}}$, we prove that

$$\begin{aligned}&g_{[s]}(t; \mathbf{y}^{[n_k]}, \mathbf{x}_{[k]\setminus[s]}^{[n_k]}) \\ &\triangleq \mathbb{P}\Big[\imath_{[s]}(\bar{\mathbf{X}}_{[s]}^{[n_k]}; \mathbf{Y}_k^{[n_k]} | \mathbf{X}_{[k]\setminus[s]}^{[n_k]}) \ge t \\ &\quad \Big| \mathbf{X}_{[k]\setminus[s]}^{[n_k]} = \mathbf{x}_{[k]\setminus[s]}^{[n_k]}, \mathbf{Y}_k^{[n_k]} = \mathbf{y}^{[n_k]}\Big] \qquad (195) \\ &\le \frac{G'_{s,k}\exp\{-t\}}{\sqrt{n_k}}, \qquad (196)\end{aligned}$$

where $G'_{s,k}$ is a positive constant depending on s, k and P.

The derivation of the bound in (196) follows the analysis in Section V-D. The critical goal is to verify steps (74)–(76) for the modified input distribution in (186). This requires showing that

$$\mathbb{P}\left[\langle \mathbf{X}_{\langle[s]\rangle}^{[n_k]}, \mathbf{X}_{\langle[s]\rangle}^{[n_k]} + \mathbf{Z}^{[n_k]}\rangle - \sum_{j=1}^{k}\frac{|\mathcal{N}(j)|u_j}{2} \in [a, a+\mu] \,\middle|\, \mathcal{E}\right]$$

$$\leq O\left(\frac{1}{\sqrt{n_k}}\right), \tag{197}$$

where

$$\mathcal{E} = \Big\{ \left\| \mathbf{X}^{\mathcal{N}(j)}_{\langle [s] \rangle} + \mathbf{Z}^{\mathcal{N}(j)} \right\|^2 = |\mathcal{N}(j)| s_j, \left\| \mathbf{X}^{\mathcal{N}(j)}_{\langle [s] \rangle} \right\|^2 = |\mathcal{N}(j)| u_j \text{ for } j \in [k] \Big\}, \tag{198}$$

$s_j \in \mathcal{I}(j, [s])$, and $u_j > 0$. The proof of (197) is similar to the one in [5, Appendix A] for parallel Gaussian channels since we can consider K independent sub-codewords with lengths $|\mathcal{N}(j)|$, $j \in [K]$, as K parallel channels, each having blocklength $|\mathcal{N}(j)|$, $j \in [K]$.

Taking an arbitrary $t \in [k]$, we get

$$\begin{aligned}
&\mathbb{P}\left[\langle \mathbf{X}^{[n_k]}_{\langle [s] \rangle}, \mathbf{X}^{[n_k]}_{\langle [s] \rangle} + \mathbf{Z}^{[n_k]} \rangle - \sum_{j=1}^{k} \frac{|\mathcal{N}(j)| u_j}{2} \in [a, a+\mu] \,\middle|\, \mathcal{E} \right] \\
&= \int_{\mathbb{R}^{k-1}} \mathbb{P}\left[Z_{n_{t-1}+1} + \frac{\sqrt{|\mathcal{N}(j)|}}{2} \in \left[\frac{a'}{\sqrt{|\mathcal{N}(j)|}}, \frac{a'+\mu}{\sqrt{|\mathcal{N}(j)|}} \right] \right. \\
&\qquad \left. \,\middle|\, \mathcal{E}, \{ Z_{n_{j-1}+1} = z_j, j \in [k] \setminus \{t\} \} \right] \\
&\qquad \cdot \Big(\prod_{\substack{j \in [k] \\ j \neq t}} f_{Z_{n_{j-1}+1} | \mathcal{E}}(z_j) dz_j \Big) \qquad (199) \\
&\leq \frac{L(u_t, s_t)\mu}{\sqrt{|\mathcal{N}(t)|}} \qquad (200) \\
&\leq \frac{3}{2} \frac{L(u_t, 1+sP)\mu}{\sqrt{|\mathcal{N}(t)|}} \qquad (201) \\
&\leq \frac{3}{2} \frac{\max_{j \in [k]} L(u_j, 1+sP)\mu}{\sqrt{|\mathcal{N}(t)|}}, \qquad (202)
\end{aligned}$$

where a' is related to a by a constant shift, and (199) follows by setting $\mathbf{X}^{\mathcal{N}(j)}_{\langle [s] \rangle} = (\sqrt{|\mathcal{N}(j)| u_j}, 0, \ldots, 0)$, and conditioning on the event that $\{Z_{n_{j-1}+1} = z_j \text{ for } j \neq t\}$. Since t is arbitrary in (199), we have

$$\begin{aligned}
&\mathbb{P}\left[\langle \mathbf{X}^{[n_k]}_{\langle [s] \rangle}, \mathbf{X}^{[n_k]}_{\langle [s] \rangle} + \mathbf{Z}^{[n_k]} \rangle - \sum_{j=1}^{k} \frac{|\mathcal{N}(j)| u_j}{2} \in [a, a+\mu] \,\middle|\, \mathcal{E} \right] \\
&\leq \frac{3}{2} \frac{\max_{j \in [k]} L(u_j, 1+sP)\mu}{\sqrt{\max_{t \in [k]} |\mathcal{N}(t)|}} \qquad (203) \\
&\leq \frac{3}{2} \frac{\sqrt{k} \max_{j \in [k]} L(u_j, 1+sP)\mu}{\sqrt{n_k}}, \qquad (204)
\end{aligned}$$

which implies (197), and (196) follows.

In the following discussion, we modify the analysis in Section V-E according to the input distribution in (186). Define the information density random vector $\boldsymbol{\imath}_k$ and the typical events analogous to (81)–(83) as

$$\boldsymbol{\imath}_k \triangleq (\imath_{\mathcal{S}}(\mathbf{X}^{[n_k]}_{\mathcal{S}}; \mathbf{Y}^{[n_k]}_k | \mathbf{X}^{[n_k]}_{\mathcal{S}^c}) \colon \mathcal{S} \in \mathcal{P}([k])) \tag{205}$$

$$\mathcal{E}(\mathcal{S})_{\text{RAC}} \triangleq \left\{ \mathbf{X}^{[n_k]}_{\langle \mathcal{S} \rangle} + \mathbf{Z}^{[n_k]} \in \mathcal{F}(\mathcal{S})_{\text{RAC}} \right\} \tag{206}$$

$$\mathcal{E}_{\text{RAC}} \triangleq \bigcap_{\mathcal{S} \in \mathcal{P}([k])} \mathcal{E}(\mathcal{S})_{\text{RAC}} \tag{207}$$

$$\mathcal{A}_k \triangleq \left\{ \boldsymbol{\imath}_k \geq \left(\log \left(\binom{M-k}{|\mathcal{S}|} (G'_{|\mathcal{S}|,k})^2 \alpha_{|\mathcal{S}|,k} \right) : \mathcal{S} \in \mathcal{P}([k]) \right) - \frac{1}{2} \log n_k \mathbf{1} \right\}, \tag{208}$$

where $\alpha_{s,k}$ is given in (155). By Lemma 2 and the union bound, we have

$$\mathbb{P}\left[\mathcal{E}^c_{\text{RAC}}\right] \leq \sum_{j=1}^{k} \exp\left\{ -c_k |\mathcal{N}(j)|^{1/3} \right\}, \tag{209}$$

where c_k is a positive constant. Combining (196) and (209) and following the analysis in Section V-E, we bound the expectation in (34c) by

$$\mathbb{P}\left[\mathcal{A}^c_k\right] + \sum_{j=1}^{k} \exp\left\{ -c_k |\mathcal{N}(j)|^{1/3} \right\} + \frac{1}{\sqrt{n_k}}. \tag{210}$$

Step 3: Given M and $\{\epsilon_k\}_{k=0}^K$, we set the decoding times $n_1, \ldots, n_K$ according to the equalities

$$\begin{aligned}
k \log M = {} & n_k C(kP) \\
& - \sqrt{n_k (V(kP) + V_{\text{cr}}(k,P))} Q^{-1}\left(\epsilon_k - \frac{D_k}{\sqrt{n_k}} \right) \\
& + \frac{1}{2} \log n_k + \eta_k - k \log \kappa_k(P\mathbf{1})
\end{aligned} \tag{211}$$

for all $k \in [K]$, where D_k is a positive constant to be chosen later in (225), and $\eta_k \triangleq -2 \log G'_{k,k} + (k-1) \log k - k$. Since $\frac{1}{s} C(sP) > \frac{1}{k} C(kP)$ for $s < k$ and (211), we reach a sequence of conclusions.

1) There exists a constant $c_0 > 0$ such that $\min_{j \in [k]} |\mathcal{N}(j)| \geq c_0 n_k$ for large enough M. In other words, $|\mathcal{N}(j)|$ is of the same order as n_k for all $j \in [k]$.
2) The bound on the probability of message repetition, $\frac{k(k-1)}{2M}$, decays exponentially with n_k.
3) In order to bound the expression in (192) as $B \leq O\left(\frac{1}{\sqrt{n_k}}\right)$, we choose $n_0 \geq \frac{4(1+P^2)}{P^2} \log n_1 + o(\log n_1)$.
4) By the union bound, Chebyshev's inequality, $\alpha_{k,k} = k$ in (155), and the fact that

$$\binom{M}{k} \leq \left(\frac{eM}{k} \right)^k, \tag{212}$$

we get

$$\mathbb{P}\left[\mathcal{A}^c_k\right] \leq \frac{E_k}{n_k} + \mathbb{P}\left[\imath_{[k]}(\mathbf{X}^{[n_k]}_{[k]}; \mathbf{Y}^{[n_k]}_k) < k \log M - \frac{1}{2} \log n_k - \eta_k \right] \tag{213}$$

for some positive constant E_k.

Therefore, it remains only to evaluate the probability term in (213). Define the modified and centered information density random variable

$$\tilde{\imath}_k \triangleq \frac{1}{\sqrt{n_k}} \left(\sum_{i=1}^{n_k} \log \frac{P_{Y_k | X_{[k]}}(Y_i | X_{[k],i})}{P_{\tilde{Y}_k}(Y_i)} - n_k C(kP) \right), \tag{214}$$

where $\tilde{Y}_k \sim \mathcal{N}(0, 1+kP)$. By Lemma 1 and (211), we get

$$\mathbb{P}\left[\imath_{[k]}(\mathbf{X}^{[n_k]}_{[k]}; \mathbf{Y}^{[n_k]}_k) < k \log M - \frac{1}{2} \log n_k - \eta_k \right]$$

$$\leq \mathbb{P}\left[\tilde{\imath}_k < -\sqrt{V(kP)+V_{\text{cr}}(k,P)}Q^{-1}\left(\epsilon_k - \frac{D_k}{\sqrt{n_k}}\right)\right]. \tag{215}$$

The conditional distribution of $\tilde{\imath}_k$ given $\mathbf{X}_{[k]}^{[n_k]} = \mathbf{x}_{[k]}^{[n_k]}$ is the same as the conditional distribution of $\tilde{\imath}_k$ given $\mathbf{H} = \mathbf{h}$, where

$$\mathbf{H} = (H_{ij} \colon i,j \in [k], i<j) \in \mathbb{R}^{\binom{k}{2}}, \tag{216}$$

and $H_{ij} = \frac{\langle \mathbf{X}_i^{[n_k]}, \mathbf{X}_j^{[n_k]} \rangle}{\sqrt{n_k P^2}}$. To bound the right-hand side of (215), in a manner similar to the arguments in Section VI, we only need to verify that

$$\text{TV}(P_{\mathbf{H}}, P_{\tilde{\mathbf{H}}}) \leq \frac{\psi_k}{\sqrt{n_k}} \tag{217}$$

for some constant ψ_k, where $\tilde{\mathbf{H}} \sim \mathcal{N}\left(\mathbf{0}, \mathsf{I}_{\binom{k}{2}}\right)$. To show (217), we define

$$\mathbf{H}^{(t)} \triangleq (H_{ij}^{(t)} \colon i,j \in [k], i<j) \in \mathbb{R}^{\binom{k}{2}}, \tag{218}$$

where $H_{ij}^{(t)} = \frac{\langle \mathbf{X}_i^{\mathcal{N}(t)}, \mathbf{X}_j^{\mathcal{N}(t)} \rangle}{\sqrt{|\mathcal{N}(t)|P^2}}$, then write

$$\mathbf{H} = \sum_{t=1}^{k} \frac{\sqrt{|\mathcal{N}(t)|}}{\sqrt{n_k}} \mathbf{H}^{(t)}. \tag{219}$$

By the data processing inequality of the total variation distance and the independence of $\mathbf{H}^{(t)}$, $t \in [k]$, we get

$$\text{TV}(P_{\mathbf{H}}, P_{\tilde{\mathbf{H}}}) \leq \text{TV}\left(\prod_{t=1}^{k} P_{\mathbf{H}^{(t)}}, P_{\tilde{\mathbf{H}}}^{k}\right) \tag{220}$$

$$\leq \sum_{t=1}^{k} \text{TV}(P_{\mathbf{H}^{(t)}}, P_{\tilde{\mathbf{H}}}) \tag{221}$$

$$\leq \sum_{t=1}^{k} \frac{F_k}{\sqrt{|\mathcal{N}(t)|}} \tag{222}$$

$$\leq \frac{kF_k}{\sqrt{c_0 n_k}}, \tag{223}$$

where (221) applies [44, eq. (4.5)], which bounds the total variation distance between two product measures P^k and Q^k by k times the total variation distances between P and Q. The bound in [44, eq. (4.5)] is extended to arbitrary product measures $\prod_{i=1}^{k} P_i$ and $\prod_{i=1}^{k} Q_i$ in [45, Lemma 2.1]. Inequality (222) follows from Lemma 7, F_k is the constant from Lemma 7, and (223) follows from (211), which proves (217).

By (223), and following arguments similar to those in Section VI, we conclude that

$$\mathbb{P}\left[\tilde{\imath}_k < -\sqrt{V(kP)+V_{\text{cr}}(k,P)}Q^{-1}\left(\epsilon_k - \frac{D_k}{\sqrt{n_k}}\right)\right]$$
$$\leq \epsilon_k - \frac{D_k}{\sqrt{n_k}} + \frac{C_k}{\sqrt{n_k}}, \tag{224}$$

where C_k is a Berry-Esseen constant. We choose the constant D_k such that

$$\frac{D_k}{\sqrt{n_k}} \leq \frac{k(k-1)}{2M} + B + \frac{C_k}{\sqrt{n_k}} + \frac{E_k}{n_k}$$
$$+k\exp\left\{-c_k(c_0 n_k)^{1/3}\right\} + \frac{1}{\sqrt{n_k}}, \tag{225}$$

where B is in (192). For large enough n_k, such a constant exists by the enumerated consequences of (211), above. From Theorem 5 and the inequalities (188), (210)–(213), (215), (224) and (225), we conclude that the probability of error is bounded by ϵ_k. By the Taylor series expansion of the function $Q^{-1}(\cdot)$ in (211), we complete the proof.

IX. Concluding Remarks

This paper studies the Gaussian multi-access channels in the finite-blocklength regime for two communication scenarios. In the first scenario, called the Gaussian MAC, K active transmitters are fixed and known to the transmitters and the receiver; in the second scenario, called the Gaussian RAC, an unknown subset of K transmitters is active, and neither the transmitters nor the receiver knows the set of active transmitter.

For the Gaussian MAC problem, we build on the RCU bound (Theorem 1) for general MACs to prove a third-order achievability result (Theorem 2). Our random encoder design chooses codewords distributed independently and uniformly on the n-dimensional sphere. At the receiver, we employ a maximum likelihood decoder. Compared to the result of MolavianJazi and Laneman [8], our coding scheme improves the achievable third-order term to $\frac{1}{2}\log n\mathbf{1} + O(1)\mathbf{1}$. Theorem 3 extends our result for the Gaussian MAC with two transmitters to the K-transmitter Gaussian MAC.

We generalize the rateless coding strategy in [21] for the permutation-invariant random access channels by allowing non-i.i.d. input distributions at the random encoding function. For the Gaussian RAC, our strategy uses concatenated codewords such that each sub-codeword is uniformly distributed on a power sphere and independent of the other sub-codewords. In our proposed coding strategy, the decoding occurs at finitely many time instants $n_0, \ldots, n_K$, with the choice of n_k indicating that the decoder's estimate of the number of active transmitters is k. The receiver broadcasts a single bit to all transmitters at each decoding time, indicating whether or not it is ready to decode. The decoding rule combines a threshold rule based on the total received power and a maximum likelihood decoder. Building upon our result on the Gaussian MAC, we show in Theorem 4 that our rateless Gaussian RAC code achieves the same performance up to the third-order term as the best known code for the Gaussian MAC in operation (Corollary 1). Furthermore, by forcing decoding at time n_K our feedback RAC code in Theorem 4 can be used with a K-transmitter MAC without feedback. While this can only reduce the error probability determined in Theorem 5 by eliminating the error events that result from deciding upon an incorrect number of active transmitters, that reduction is negligible in our asymptotic regime (see the proof of Theorem 4). Thus, Theorem 4 also describes the performance of the length-n_K codebook of the RAC code when used with a K-transmitter MAC without feedback. That means that although the length-n_K codebook of the RAC code is supported on only a subset of the power sphere (see Fig. 2), it achieves the same first three order terms on a K-transmitter MAC as the more traditional code in Corollary 1 that uses the entire power sphere.

APPENDIX A
PROOF OF COROLLARY 1

In order to prove Corollary 1, we show that for any M that satisfies the inequality (25), it holds that

$$(|\mathcal{S}|\log M\colon \mathcal{S})\in\mathcal{P}([K])\in n\mathbf{C}(P\mathbf{1})-\sqrt{n}Q_{\text{inv}}(\mathsf{V}(P\mathbf{1}),\epsilon)+\frac{1}{2}\log n\mathbf{1}+O(1)\,\mathbf{1}. \tag{226}$$

Let $\mathbf{Z}=(Z(\mathcal{S}):\mathcal{S}\in\mathcal{P}([K]))\sim\mathcal{N}(\mathbf{0},\mathsf{V}(P\mathbf{1}),\epsilon))$. Take M such that the asymptotic expansion in (25) holds, implying that

$$\mathbb{P}\Big[Z([K])>\sqrt{n}\left(C(KP)-\frac{K\log M}{n}\right)+\frac{1}{2}\frac{\log n}{\sqrt{n}}+O\left(\frac{1}{\sqrt{n}}\right)\Big]\le\epsilon. \tag{227}$$

Consider any $\mathcal{S}\in\mathcal{P}([K])$ with $|\mathcal{S}|<K$. Then

$$\mathbb{P}\Big[Z(\mathcal{S})>\sqrt{n}\left(C(|\mathcal{S}|P)-\frac{|\mathcal{S}|\log M}{n}\right)+\frac{1}{2}\frac{\log n}{\sqrt{n}}+O\left(\frac{1}{\sqrt{n}}\right)\Big]\le O\left(\frac{1}{n}\right), \tag{228}$$

which follows from Chebyshev's inequality since $C(sP)-\frac{s}{K}C(KP)>0$ for $s<K$.

By the union bound, (227) and (228), we get

$$\mathbb{P}\Bigg[\bigcup_{\mathcal{S}\in\mathcal{P}([K])}\Big\{Z(\mathcal{S})>\sqrt{n}\left(C(|\mathcal{S}|P)-\frac{|\mathcal{S}|\log M}{n}\right)+\frac{1}{2}\frac{\log n}{\sqrt{n}}+O\left(\frac{1}{\sqrt{n}}\right)\Big\}\Bigg]\le\epsilon+O\left(\frac{1}{n}\right), \tag{229}$$

which, by the definition (15), is equivalent to

$$(|\mathcal{S}|\log M\colon \mathcal{S}\in\mathcal{P}([K]))\in n\mathbf{C}(P\mathbf{1})-\sqrt{n}Q_{\text{inv}}\left(\mathsf{V}(P\mathbf{1}),\epsilon+O\left(\frac{1}{n}\right)\right)+\frac{1}{2}\log n\mathbf{1}+O(1)\,\mathbf{1}. \tag{230}$$

Applying the Taylor series expansion to $Q_{\text{inv}}(\mathsf{V}(P\mathbf{1}),\cdot)$ completes the proof.

APPENDIX B
CODE DESIGN VARIATIONS

A. Adopting the Codebooks Based on the Channel Estimate at Time n_0

In our encoder and decoder design, we use the fact that the received output power concentrates around its mean value. In the proof of Theorem 2, we show that $n_0=O(\log n_1)$ symbols are sufficient to ensure that the probability that the decision is made at the correct decoding time, i.e., n_k when k transmitters are active, decays with $O\left(\frac{1}{\sqrt{n_k}}\right)$. In our strategy, we make a binary decision at each decoding time $n_0,\dots,n_K$ of whether or not to decode. An alternative to this strategy would be to decide the number of active transmitters at time n_0, which is much smaller than the rest of the decoding times, and to inform the transmitters about the decoding time in the epoch at time n_0. This alternative allows for a code design that depends on the feedback from the receiver to the transmitters at time n_0. Using its knowledge of the typical interval, in which the squared norm of the output, $\frac{1}{n_0}\left\|\mathbf{Y}_k^{[n_0]}\right\|^2$, lies for each $k\le K$, the decoder estimates the number of active transmitters. We denote this value by t. The decoder could then transmit t to all transmitters, so that all parties understand that the communication epoch is going to end at time n_t. This strategy requires $\lceil\log(K+1)\rceil$ bits of feedback from the receiver to transmitters at time n_0; in contrast, the strategy in the proof of Theorem 4 requires a number of bits of feedback that varies with the decoder's estimate of the number of active transmitters with a maximum of $K+1$ bits. Let the decoder choose t as the nearest integer to $\frac{1}{P}\left(\frac{1}{n_0}\left\|\mathbf{y}^{[n_0]}\right\|^2-1\right)$. Then, the bound in (192) on the probability that the decoder errs in determining the number of active transmitters can be bounded as

$$\mathbb{P}\left[\mathcal{E}_{\text{time}}\middle|\mathcal{E}_{\text{rep}}^c\right]\le 2\left(\prod_{j=1}^{k}\kappa_j(P\mathbf{1})\right)\exp\left\{-\frac{n_0(\frac{P}{2})^2}{8(1+kP)^2}\right\} \tag{231}$$

in the case when the decision is made at time n_0. Like (192), this bound decays exponentially with n_0. Here, however, the exponential rate is smaller than (192). Hence, this modification in the strategy increases the constant c in (32), and affects the achievable $O(1)$ term in (31).

As the encoders learn the estimate of the number of active transmitters at an earlier time, an encoding function that depends on the feedback from the receiver could be employed as follows. Recall from (28) that the maximal-power constraints apply to the decoding times $n_1,\dots,n_K$, but not to n_0. Given the estimate t of the number of active transmitters k, length-n_t codewords are drawn such that the first n_1 symbols are uniformly distributed on n_1-dimensional sphere with radius $\sqrt{n_1P}$, and the symbols indexed from n_1+1 to n_t are distributed on (n_t-n_1)-dimensional sphere with radius $\sqrt{(n_t-n_1)P}$, i.e., instead of K independent spherical sub-codewords, we use two independent sub-codewords. The length of the second sub-codeword depends on the estimate t. The effect of this modification on the error analysis is that under this input distribution, the total variation bound in (223) can be improved to

$$\text{TV}(P_{\mathbf{H}},P_{\tilde{\mathbf{H}}})\le\frac{F_k}{\sqrt{n_1}}+\frac{F_k}{\sqrt{n_k-n_1}}, \tag{232}$$

which decays with the same asymptotic rate as (223). Therefore, this modification affects only the $O(1)$ term in (31), meaning that the same expansion as Theorem 4 is achieved.

B. Decoding Transmitter Identity

Another scenario of possible interest is the case, where the decoder must decode transmitter identities as well as messages. For this scenario, as we observed in the context of RACs without codeword cost constraints [21, Section V.D], one can again use the encoding and decoding rules employed in Section VII-A with a message set size of KM such that the messages indexed from $(k-1)M+1$ to kM are associated with

transmitter k. If the decoder decides to decode k messages at time n_k, a list of k out of KM messages is decoded, which automatically reveals the identities of the transmitters. By our RAC code design, it is possible that the decoder decodes multiple messages belonging to the same transmitter; this would not have been possible with the MAC decoder. Given that there are k active transmitters, in the proposed coding scheme, the random message $W_{[k]}$ is not uniformly distributed over a set of size $\binom{KM}{k}$, as it would be if any combination of k of the KM codewords could be transmitted. Instead, it is uniformly distributed over a set of size M^k since each of the k active transmitters sends exactly one of its M codewords. Nonetheless, replacing M by KM in Theorem 4 gives a valid upper bound on the resulting error probability since the codewords for each transmitter and each message are generated i.i.d. While the inclusion of all $\binom{KM}{k}$ codeword combinations in (178)–(181) may yield looser than necessary bounds on the error probability, their implication is that to decode transmitter identities the RAC code pays a penalty of $-k \log K$ on the right-hand side of (31). Since K does not grow with n_1, decoding transmitter identities affects only the $O(1)$ term in (31).

Appendix C
Proof of Lemma 4

Pinsker's inequality (e.g., [46, Th. 6.5]) states that for any distributions P and Q,

$$\mathrm{TV}(P,Q) \le \sqrt{\frac{1}{2} D(P\|Q)}. \tag{233}$$

Let $\mathrm{tr}(\cdot)$ denote trace of its matrix argument. The relative entropy between two d-dimensional Gaussian distributions with positive covariance matrices is given (e.g., [46, eq. (1.18)]) by

$$\begin{aligned} &D(\mathcal{N}(\boldsymbol{\mu}_1,\Sigma_1)\|\mathcal{N}(\boldsymbol{\mu}_2,\Sigma_2)) \\ &= \frac{1}{2}\Big(\mathrm{tr}(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathsf{I}_d) + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\Sigma_1^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) \\ &\quad - \log\det(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2})\Big). \end{aligned} \tag{234}$$

Define

$$\mathsf{G} \triangleq \Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathsf{I}_d \tag{235}$$

$$a \triangleq \frac{1}{2}\sqrt{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\Sigma_1^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)}. \tag{236}$$

Combining (233) and (234) and using the inequality $\sqrt{x+y} \le \sqrt{x}+\sqrt{y}$, we get

$$\begin{aligned} &\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}_1,\Sigma_1),\mathcal{N}(\boldsymbol{\mu}_2,\Sigma_2)) \\ &\le a + \frac{1}{2}\sqrt{\mathrm{tr}(\mathsf{G}) - \log\det(\mathsf{I}_d+\mathsf{G})}. \end{aligned} \tag{237}$$

To bound the logdeterminant term in (237) from below, we use the following result from [42, Th. 1.1]. Let $\rho(\cdot)$ denote the spectral radius, i.e., the maximum absolute eigenvalue, and let $\|\cdot\|_F$ denote the Frobenius norm. If $\rho(\mathsf{G}) < 1$, then

$$\exp\left\{\mathrm{tr}(\mathsf{G}) - \frac{\|\mathsf{G}\|_F^2}{2(1-\rho(\mathsf{G}))}\right\} \le \det(\mathsf{I}_d+\mathsf{G}). \tag{238}$$

For $\rho(\mathsf{G}) < 1$, we apply (238) to (237) and get

$$\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}_1,\Sigma_1),\mathcal{N}(\boldsymbol{\mu}_2,\Sigma_2)) \le \frac{1}{2\sqrt{2}}\frac{\|\mathsf{G}\|_F}{\sqrt{1-\rho(\mathsf{G})}} + a. \tag{239}$$

In addition, trivially, we have that

$$\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}_1,\Sigma_1),\mathcal{N}(\boldsymbol{\mu}_2,\Sigma_2)) \le 1 \tag{240}$$

$$\le \frac{\|\mathsf{G}\|_F}{\rho(\mathsf{G})} + a, \tag{241}$$

where in (241), we use the fact that Frobenius norm is an upper bound to the spectral radius. Taking the tighter bound among (239) and (241), we conclude that for $\rho(\mathsf{G}) < 1$,

$$\begin{aligned} &\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}_1,\Sigma_1),\mathcal{N}(\boldsymbol{\mu}_2,\Sigma_2)) \\ &\le \min\left\{\frac{1}{2\sqrt{2}}\frac{1}{\sqrt{1-\rho(\mathsf{G})}}, \frac{1}{\rho(\mathsf{G})}\right\}\|\mathsf{G}\|_F + a \qquad (242) \\ &\le \frac{2+\sqrt{6}}{4}\|\mathsf{G}\|_F + a. \qquad (243) \end{aligned}$$

Inequality (243) follows since the maximum of the minimum term in (242) is achieved by $\rho(\mathsf{G}) = 2\sqrt{6}-4 \approx 0.899$ and that maximum value is $\frac{2+\sqrt{6}}{4}$. Since the coefficient $\frac{2+\sqrt{6}}{4} > 1 \ge \frac{1}{\rho(\mathsf{G})}$ for $\rho(\mathsf{G}) \ge 1$, we conclude that (243) holds for any $\rho(\mathsf{G})$.

Appendix D
Proof of (73)

We show a more general result. Fix any constant $u < P_1 + P_2$. We prove below that for n large enough,

$$g(y) \triangleq \mathbb{P}\left[\|\mathbf{X}_1+\mathbf{X}_2\|^2 \le nu \,\middle|\, \|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2 = y\right] \tag{244}$$

$$\le \exp\{-nC\} \tag{245}$$

for all $y \in \mathcal{I}$, where

$$\mathcal{I} \triangleq [n(1+P_1+P_2-\epsilon), n(1+P_1+P_2+\epsilon)] \tag{246}$$

$$\epsilon \triangleq n^{-1/3}, \tag{247}$$

and C is a positive constant depending on u. Taking $u = P_1+P_2-\sqrt{P_1P_2}$ in (244) then proves the desired inequality (73).

We proceed to prove (245). Since the support of $\|\mathbf{X}_1+\mathbf{X}_2\|^2$ is

$$\mathcal{S} = [n(\sqrt{P_1}-\sqrt{P_2})^2, n(\sqrt{P_1}+\sqrt{P_2})^2], \tag{248}$$

inequality (245) is trivially satisfied for $u < (\sqrt{P_1}-\sqrt{P_2})^2$. To show (245) for $(\sqrt{P_1}-\sqrt{P_2})^2 \le u < P_1+P_2$, we show two concentration results. First, we show that

$$g(y) = g(n(1+P_1+P_2))\exp\{O(n\epsilon)\} \tag{249}$$

for all $y \in \mathcal{I}$; second, we show that for n large enough,

$$p \triangleq \mathbb{P}\left[\|\mathbf{X}_1+\mathbf{X}_2\|^2 \le nu \,\middle|\, \mathcal{A}\right] \tag{250}$$

$$\le \exp\{-nC'\} \tag{251}$$

for some $C' > 0$, where the event $\mathcal{A}$ is defined as

$$\mathcal{A} \triangleq \left\{\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2 \in \mathcal{I}\right\}. \tag{252}$$

Using (249) and (251), we show (245) as follows. By conditioning the probability in (250) on each value of $\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2$, we express p as

$$p = \int_{\mathcal{I}} g(y) f_{\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2|\mathcal{A}}(y)dy \tag{253}$$
$$= g(n(1+P_1+P_2))\exp\{O(n\epsilon)\} \tag{254}$$
$$\leq \exp\{-nC'\}, \tag{255}$$

where (254) follows from (249) and

$$\min_{y\in\mathcal{I}} g(y) \leq \int_{\mathcal{I}} g(y) f_{\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2|\mathcal{A}}(y)dy \leq \max_{y\in\mathcal{I}} g(y). \tag{256}$$

Inequality (255) follows from (251). Inequalities (249) and (255) imply that since $O(n\epsilon)=o(n)$, there exists a constant $C>0$ such that for n large enough, (245) holds for all $y\in\mathcal{I}$.

We proceed to show (251). By Bayes' rule, we have

$$p = \frac{\mathbb{P}\left[\|\mathbf{X}_1+\mathbf{X}_2\|^2 \leq nu\right]\mathbb{P}\left[\mathcal{A}\,\middle|\,\|\mathbf{X}_1+\mathbf{X}_2\|^2\leq nu\right]}{\mathbb{P}[\mathcal{A}]}. \tag{257}$$

Changing measure from $P_{\mathbf{X}_1+\mathbf{X}_2}P_{\mathbf{Z}}$ to $P_{\tilde{\mathbf{U}}}P_{\mathbf{Z}}$, where $\tilde{\mathbf{U}}\sim\mathcal{N}(0,(P_1+P_2)\mathsf{I}_n)$, and then applying Lemma 1, we get

$$p \leq \frac{\kappa_2(P_1,P_2)\mathbb{P}\left[\left\|\tilde{\mathbf{U}}\right\|^2\leq nu\right]\cdot 1}{1-\kappa_2(P_1,P_2)\mathbb{P}\left[\left|\left\|\tilde{\mathbf{U}}+\mathbf{Z}\right\|^2 - n(1+P_1+P_2)\right|>n\epsilon\right]} \tag{258}$$
$$\leq \kappa_2(P_1,P_2)\frac{\exp\left\{\frac{-n(P_1+P_2-u)^2}{4(P_1+P_2)^2}\right\}}{1-2\kappa_2(P_1,P_2)\exp\{\frac{-n\epsilon^2}{8(1+P_1+P_2)^2}\}} \tag{259}$$
$$\leq 2\kappa_2(P_1,P_2)\exp\left\{\frac{-n(P_1+P_2-u)^2}{4(P_1+P_2)^2}\right\} \tag{260}$$
$$\leq \exp\{-nC'\} \tag{261}$$

for all n large enough, where $\kappa_2(P_1,P_2)$ is the constant defined in (38), and C' is a positive constant. Inequality (259) follows from the tail bounds on the chi-squared distribution in Lemma 2, and (260) follows since the denominator on the right-hand side of (259) is greater than $\frac{1}{2}$ for n large enough. Inequality (261) holds since $u<P_1+P_2$.

We proceed to prove (249). Define the events $\mathcal{B}\triangleq\{\|\mathbf{X}_1+\mathbf{X}_2\|^2\leq nu\}$ and $\mathcal{B}(\lambda)\triangleq\{\|\mathbf{X}_1+\mathbf{X}_2\|^2=\lambda\}$ for any $\lambda\in\mathcal{S}$. By Bayes' rule, we can express $g(y)$ as

$$g(y) = \frac{\mathbb{P}[\mathcal{B}]\, f_{\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2|\mathcal{B}}(y)}{f_{\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2}(y)}. \tag{262}$$

By the spherical symmetry of the distribution of $\mathbf{X}_1+\mathbf{X}_2$, the conditional distribution of $\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2$ given $\mathcal{B}(\lambda)$ does not depend on $\mathbf{u}$ when we fix $\mathbf{X}_1+\mathbf{X}_2$ to any $\mathbf{u}$ such that $\|\mathbf{u}\|^2=\lambda\in\mathcal{S}$. Therefore, the conditional distribution of $\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2$ given $\mathcal{B}(\lambda)$ equals the distribution of

$$\sum_{i=1}^{n}\left\|Z_i+\frac{\sqrt{\lambda}}{\sqrt{n}}\right\|^2, \tag{263}$$

which has non-central chi-squared distribution with n degrees of freedom and non-centrality parameter λ. That is, the probability density function is

$$f(x;n,\lambda) = \frac{1}{2}\exp\left\{-\frac{(x+\lambda)}{2}\right\}\left(\frac{x}{\lambda}\right)^{\frac{n}{4}-\frac{1}{2}} I_{\frac{n}{2}-1}(\sqrt{\lambda x}), \tag{264}$$

where $I_\nu(x)$ denotes the modified Bessel function of the first kind with order ν. Fix some $\lambda>0$, $x_1=nb$, and $x_2=n(b+\delta)$, where $0<\delta\leq\epsilon$ and $b>0$. Consider the ratio

$$\frac{f(x_1;n,\lambda)}{f(x_2;n,\lambda)} = \exp\{x_2-x_1\}\left(\frac{x_1}{x_2}\right)^{\frac{n}{4}-\frac{1}{2}}\frac{I_{\frac{n}{2}-1}(\sqrt{\lambda x_1})}{I_{\frac{n}{2}-1}(\sqrt{\lambda x_2})}. \tag{265}$$

Paris [47] bounds $I_\nu(x)/I_\nu(y)$ as

$$\exp\{x-y\}\left(\frac{x}{y}\right)^\nu < \frac{I_\nu(x)}{I_\nu(y)} < \left(\frac{x}{y}\right)^\nu \tag{266}$$

for any $0<x<y$ and $\nu>-1/2$. Using (266), we bound (265) as

$$\exp\{n\delta\}\left(1-\frac{\delta}{b+\delta}\right)^{\frac{n}{2}-1}\exp\left\{-\sqrt{n\lambda}\left(\sqrt{b+\delta}-\sqrt{b}\right)\right\}$$
$$\leq \frac{f(x_1;n,\lambda)}{f(x_2;n,\lambda)} \tag{267}$$
$$\leq \exp\{n\delta\}\left(1-\frac{\delta}{b+\delta}\right)^{\frac{n}{2}-1}. \tag{268}$$

Applying the Taylor series expansion at $\delta=0$ gives

$$\log\left(1-\frac{\delta}{b+\delta}\right) = -\frac{\delta}{b}+O(\delta^2) \tag{269}$$
$$-\sqrt{n\lambda}\left(\sqrt{b+\delta}-\sqrt{b}\right) = -\sqrt{n\lambda}\left(\frac{\delta}{2\sqrt{b}}+O(\delta^2)\right). \tag{270}$$

Substituting (269) and (270) in (267) and (268), we get

$$\frac{f(x_1;n,\lambda)}{f(x_2;n,\lambda)} = \exp\{O(n\delta)\}. \tag{271}$$

We can also verify the validity of (271) for $\lambda=0$ by using the probability density function of chi-squared distribution with n degrees of freedom instead of (264). Particularizing (271) to $b=1+P_1+P_2$, we get for all $\lambda\in\mathcal{S}$ that

$$f_{\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2|\mathcal{B}(\lambda)}(y)$$
$$= f_{\|\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}\|^2|\mathcal{B}(\lambda)}(n(1+P_1+P_2))\exp\{O(n\epsilon)\}, \tag{272}$$

which together with (262) implies (249).

Appendix E
Proof of Lemma 7

We use the induction technique from [43, Th. 4] to prove this lemma, showing that the total variation distance in (149) diminishes as n goes to infinity. We here prove that the convergence rate is $O\left(\frac{1}{\sqrt{n}}\right)$. Since the distribution of $\mathbf{H}$ is invariant to rotation, we fix

$$\mathbf{X}_1 = (1,0,0,\ldots,0). \tag{273}$$

Then $H_{1j} = \sqrt{n} X_{j1}$ for $2 \le j \le K$. Define the vectors

$$\mathbf{H}_1 \triangleq (H_{1j} : 2 \le j \le K) \tag{274}$$

$$\mathbf{H}_2 \triangleq (H_{ij} : 2 \le i < j \le K), \tag{275}$$

which consist of all the inner product random variables including $\mathbf{X}_1$, and not including $\mathbf{X}_1$, respectively. Hence $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$. Notice that $\mathbf{H}_1$ is a product distribution since X_{j1}'s are independent.

Note that we have for $2 \le i < j \le K$

$$H_{ij} = \sqrt{n} X_{i1} X_{j1} + \frac{\sqrt{n}}{\sqrt{n-1}} (1 - X_{i1}^2)^{\frac{1}{2}} (1 - X_{j1}^2)^{\frac{1}{2}} V_{ij} \tag{276}$$

$$V_{ij} = \sqrt{n-1} \langle \mathbf{Y}_i, \mathbf{Y}_j \rangle, \tag{277}$$

where $\mathbf{Y}_i = (1 - X_{i1}^2)^{-\frac{1}{2}} (X_{i2}, \ldots, X_{in}) \in \mathbb{R}^{n-1}$ for $i = 2, \ldots, K$. Denote by $p_K^{(n)}$ the distribution of the $\binom{K}{2}$-dimensional random vector $(\sqrt{n} \langle \mathbf{Z}_i, \mathbf{Z}_j \rangle : 1 \le i < j \le K)$, where the $\mathbf{Z}_i$, $i \in [K]$, are distributed independently and uniformly on $\mathbb{S}^n(1)$.

Since $\mathbf{Y}_i, i \in \{2, \ldots, K\}$, are distributed independently and uniformly on $\mathbb{S}^{n-1}(1)$, the joint distribution of $\mathbf{V} = (V_{ij} : 2 \le i < j \le K)$ is $p_{K-1}^{(n-1)}$. By (276), the conditional distribution of H_{ij} given $\mathbf{H}_1 = \mathbf{h}_1$ is the same as the distribution of

$$\frac{h_{1i} h_{1j}}{\sqrt{n}} + \frac{\sqrt{n}}{\sqrt{n-1}} \left(1 - \frac{h_{1i}^2}{n}\right)^{\frac{1}{2}} \left(1 - \frac{h_{1j}^2}{n}\right)^{\frac{1}{2}} V_{ij} \tag{278}$$

for $2 \le i < j \le K$. We define the random vector $\mathbf{H}_2^* = (H_{ij}^* : 2 \le i < j \le K)$ through $\mathbf{H}_1$ as follows. The conditional distribution of H_{ij}^* given $\mathbf{H}_1 = \mathbf{q}_1$ is the same as the distribution of

$$\frac{h_{1i} h_{1j}}{\sqrt{n}} + \frac{\sqrt{n}}{\sqrt{n-1}} \left(1 - \frac{h_{1i}^2}{n}\right)^{\frac{1}{2}} \left(1 - \frac{h_{1j}^2}{n}\right)^{\frac{1}{2}} Z_{ij} \tag{279}$$

for $2 \le i < j \le K$, where $Z_{ij} \sim \mathcal{N}(0,1)$, and H_{ij}^*, $2 \le i < j \le K$, are conditionally independent given $\mathbf{H}_1$. Now, we are ready to apply the mathematical induction.

Base case: For $K = 2$, we have

$$\mathrm{TV}(p_2^{(n)}, \mathcal{N}(0,1)) \le \frac{4}{n} \tag{280}$$

by Lemma 5 with $k = 1$.

Inductive step: For $K > 2$, assume that for any n,

$$\mathrm{TV}\left(p_{K-1}^{(n)}, \mathcal{N}\left(\mathbf{0}, \mathsf{I}_{\frac{1}{2}(K-1)(K-2)}\right)\right) \le \frac{C_{K-1}}{\sqrt{n}} \tag{281}$$

for some constant C_{K-1}. Let $P_{\tilde{\mathbf{H}}_1} = \mathcal{N}(\mathbf{0}, \mathsf{I}_{K-1})$ and $P_{\tilde{\mathbf{H}}_2} = \mathcal{N}\left(\mathbf{0}, \mathsf{I}_{\binom{K-1}{2}}\right)$. Since the total variation distance is ℓ_1 norm, applying the triangle inequality gives

$$\begin{aligned}
&\mathrm{TV}\left(p_K^{(n)}, \mathcal{N}\left(\mathbf{0}, \mathsf{I}_{\binom{K}{2}}\right)\right) \\
&= \mathrm{TV}\left(P_{\mathbf{H}_1} P_{\mathbf{H}_2|\mathbf{H}_1}, P_{\tilde{\mathbf{H}}_1} P_{\tilde{\mathbf{H}}_2}\right) && (282)\\
&\le \mathrm{TV}\left(P_{\mathbf{H}_1} P_{\mathbf{H}_2|\mathbf{H}_1}, P_{\tilde{\mathbf{H}}_1} P_{\mathbf{H}_2|\mathbf{H}_1}\right) && (283)\\
&\quad + \mathrm{TV}\left(P_{\tilde{\mathbf{H}}_1} P_{\mathbf{H}_2|\mathbf{H}_1}, P_{\tilde{\mathbf{H}}_1} P_{\mathbf{H}_2^*|\mathbf{H}_1}\right) && (284)\\
&\quad + \mathrm{TV}\left(P_{\tilde{\mathbf{H}}_1} P_{\mathbf{H}_2^*|\mathbf{H}_1}, P_{\tilde{\mathbf{H}}_1} P_{\tilde{\mathbf{H}}_2}\right). && (285)
\end{aligned}$$

Here, (283) approximates the input measure $P_{\mathbf{H}_1}$ with the corresponding i.i.d. Gaussian measure $P_{\tilde{\mathbf{H}}_1}$, (284) approximates the inner product random variables V_{ij} in the definition of the probability transition kernel given in (278) with i.i.d. standard Gaussian random variables, and (285) approximates the mean in (279) by 0 and the variance by 1. We next bound the right-hand sides of (283)–(285) in that order. We have

$$\begin{aligned}
&\mathrm{TV}\left(P_{\mathbf{H}_1} P_{\mathbf{H}_2|\mathbf{H}_1}, P_{\tilde{\mathbf{H}}_1} P_{\mathbf{H}_2|\mathbf{H}_1}\right) \\
&= \mathrm{TV}\left(P_{\mathbf{H}_1}, P_{\tilde{\mathbf{H}}_1}\right) && (286)\\
&\le (K-1) \mathrm{TV}\left(P_{H_{12}}, \mathcal{N}(0,1)\right) && (287)\\
&\le \frac{4(K-1)}{n}, && (288)
\end{aligned}$$

where (287) follows from [45, Lemma 2.1] since $P_{\mathbf{H}_1} = (P_{H_{12}})^{K-1}$ and $P_{\tilde{\mathbf{H}}_1} = (\mathcal{N}(0,1))^{K-1}$ are both product distributions, and (288) follows from Lemma 5. The total variation distance in (284) is bounded as

$$\begin{aligned}
&\mathrm{TV}\left(P_{\tilde{\mathbf{H}}_1} P_{\mathbf{H}_2|\mathbf{H}_1}, P_{\tilde{\mathbf{H}}_1} P_{\mathbf{H}_2^*|\mathbf{H}_1}\right) \\
&= \mathbb{E}\left[\mathrm{TV}\left(P_{\mathbf{H}_2|\mathbf{H}_1 = \tilde{\mathbf{H}}_1}, P_{\mathbf{H}_2^*|\mathbf{H}_1 = \tilde{\mathbf{H}}_1}\right) \middle| \tilde{\mathbf{H}}_1\right] && (289)\\
&= \mathrm{TV}\left(p_{K-1}^{(n-1)}, \mathcal{N}\left(\mathbf{0}, \mathsf{I}_{\binom{K-1}{2}}\right)\right) && (290)\\
&\le \frac{C_{K-1}}{\sqrt{n-1}}, && (291)
\end{aligned}$$

where (290) follows from the definitions (278) and (279) since the total variation distance is shift and scale invariant and (291) follows from the inductive assumption (281). The total variation distance in (285) is bounded as

$$\begin{aligned}
&\mathrm{TV}\left(P_{\tilde{\mathbf{H}}_1} P_{\mathbf{H}_2^*|\mathbf{H}_1}, P_{\tilde{\mathbf{H}}_1} P_{\tilde{\mathbf{H}}_2}\right) \\
&= \mathbb{E}\left[\mathrm{TV}\left(P_{\mathbf{H}_2^*|\mathbf{H}_1 = \tilde{\mathbf{H}}_1}, P_{\tilde{\mathbf{H}}_2}\right) \middle| \tilde{\mathbf{H}}_1\right] && (292)\\
&\le \mathbb{E}\left[\sum_{2 \le i < j \le K} \mathrm{TV}\left(P_{H_{ij}^*|\mathbf{H}_1 = \tilde{\mathbf{H}}_1}, \mathcal{N}(0,1)\right) \middle| \tilde{\mathbf{H}}_1\right] && (293)\\
&= \binom{K-1}{2} \mathbb{E}\Bigg[\mathrm{TV}\Bigg(\mathcal{N}\Bigg(\frac{\tilde{H}_{12} \tilde{H}_{13}}{\sqrt{n}}, \frac{n}{n-1}\left(1 - \frac{\tilde{H}_{12}^2}{n}\right) \\
&\qquad \left(1 - \frac{\tilde{H}_{13}^2}{n}\right)\Bigg), \mathcal{N}(0,1)\Bigg)\Bigg] && (294)\\
&\le \binom{K-1}{2} \Bigg\{ \frac{1}{2} \frac{\mathbb{E}\left[\left|\tilde{H}_{12}\right|\right]^2}{\sqrt{n}} \\
&\quad + \frac{2+\sqrt{6}}{4} \left| \frac{n}{n-1} \left(\mathbb{E}\left[1 - \frac{\tilde{H}_{12}^2}{n}\right]\right)^2 - 1 \right| \Bigg\} && (295)\\
&= \binom{K-1}{2} \left(\frac{1}{\pi\sqrt{n}} + \frac{2+\sqrt{6}}{4n}\right), && (296)
\end{aligned}$$

where (293) follows from [45, Lemma 2.1] since the conditional distribution of $\mathbf{H}_2^*$ given $\mathbf{H}_1 = \mathbf{h}_1$ is a product distribution and $P_{\tilde{\mathbf{H}}_2}$ is an i.i.d. standard Gaussian. Equality (294) follows since the conditional distribution of H_{ij}^* given $\mathbf{H}_1 = \mathbf{h}_1$ is identically distributed for $2 \le i < j \le K$. Inequality (295) follows from Lemma 4 with $d = 1$ using the i.i.d. distribution of $\tilde{H}_{12}$ and $\tilde{H}_{13}$. Combining (288), (291),

(296), and the inequality in (283) completes the proof by induction.

We note that the convergence rate of the total variation distance of interest is $O\left(\frac{1}{\sqrt{n}}\right)$ for $K > 2$, while it is faster $\left(O\left(\frac{1}{n}\right)\right)$ for $K = 2$.

ACKNOWLEDGMENT

We are grateful to Peida Tian for pointing out the paper [41] that led to an improvement in Lemma 4.

REFERENCES

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

[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] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

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

[5] 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.

[6] 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.

[7] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton University, Department of Electrical Engineering, Nov. 2010.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

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

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

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

[16] 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.

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

[18] I. Zadik, Y. Polyanskiy, and C. Thrampoulidis, "Improved bounds on Gaussian MAC and sparse regression via Gaussian inequalities," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 430–434.

[19] S. S. Kowshik and Y. Polyanskiy, "Fundamental limits of many-user mac with finite payloads and fading," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, pp. 5853–5884, 2021.

[20] S. S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy efficient random access for the quasi-static fading mac," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 2768–2772.

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

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

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

[24] 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.

[25] 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.

[26] L. V. Truong and V. Y. F. Tan, "On Gaussian MACs with variable-length feedback and non-vanishing error probabilities," *arXiv:1609.00594v2*, Sep. 2016, [Online]. Available: https://arxiv.org/abs/1609.00594v2.

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

[28] R. C. Yavas, V. Kostina, and M. Effros, "Variable-length feedback codes with several decoding times for the gaussian channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Melbourne, Victoria, Australia, 2021, pp. 1883–1888.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] Y. Liu and M. Effros, "Finite-blocklength and error-exponent analyses for LDPC codes in point-to-point and multiple access communication," *arXiv:2005.06428*, May 2020, [Online]. Available: https://arxiv.org/abs/2005.06428.

[35] 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.

[36] O. Kosut, "A second-order converse bound for the multiple-access channel via wringing dependence," *arxiv:2007.15664*, Jul. 2020, [Online]. Available: https://arxiv.org/abs/2007.15664.

[37] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by modelselection," *The Annals of Statistics*, vol. 28, no. 5, pp. 1302 – 1338, Oct. 2000.

[38] V. Bentkus, "A Lyapunov-type bound in $\mathbb{R}^d$," *Theory of Probability & Its Applications*, vol. 49, no. 2, pp. 311–323, 2005.

[39] M. Raič, "A multivariate Berry–Esseen theorem with explicit constants," *Bernoulli*, vol. 25, no. 4A, pp. 2824–2853, 2019.

[40] 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.

[41] L. Devroye, A. Mehrabian, and T. Reddad, "The total variation distance between high-dimensional Gaussians," *arXiv:1810.08693*, Oct. 2018, [Online]. Available: https://arxiv.org/abs/1810.08693.

[42] S. M. Rump, "Estimates of the determinant of a perturbed identity matrix," *Linear Algebra and its Applications*, vol. 558, pp. 101–107, Dec. 2018.

[43] A. J. Stam, "Limit theorems for uniform distributions on spheres in high-dimensional Euclidean spaces," *Journal of Applied Probability*, vol. 19, no. 1, pp. 221–228, Mar. 1982.

[44] W. Hoeffding and J. Wolfowitz, "Distinguishability of sets of distributions," *Ann. Math. Statist.*, vol. 29, no. 3, pp. 700–718, Sep. 1958.

[45] W. Sendler, "A note on the proof of the zero-one law of Blum and Pathak," *The Annals of Probability*, vol. 3, no. 6, pp. 1055–1058, 1975.

[46] 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.
[47] R. Paris, "An inequality for the Bessel function $J_\nu(\nu x)$," *SIAM journal on mathematical analysis*, vol. 15, no. 1, pp. 203–205, 1984.

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) is a Professor of Electrical Engineering and of Computing and Mathematical Sciences at Caltech. She received a bachelor's degree from Moscow Institute of Physics and Technology (2004), where she was affiliated with the Institute for Information Transmission Problems of the Russian Academy of Sciences, a master's degree from University of Ottawa (2006), and a PhD from Princeton University (2013). She received the Natural Sciences and Engineering Research Council of Canada postgraduate scholarship (2009–2012), the Princeton Electrical Engineering Best Dissertation Award (2013), the Simons-Berkeley research fellowship (2015) and the NSF CAREER award (2017). Kostina's research spans information theory, coding, control, learning, and communications.

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.