Gaussian Multiple and Random Access in the Finite Blocklength Regime

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, spherical distribution.**

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 are 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 signals) among the transmitters; while in slotted ALOHA, each transmitter randomly chooses a time slot to transmit their message and 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 [1].

In this work, 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. In the multiple access problem, the identity of active transmitters is known to all transmitters and the receiver. In the random access problem, the set of active transmitters is unknown to the transmitters and the receiver.

For $K = 1$, Shannon's 1948 paper [2] 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 very small δ; here P is the maximal (per-codeword) power constraint and the noise variance is 1. In [3], Shannon demonstrates the performance improvement in the achievable realibility function using codewords drawn uniformly on an n-dimensional sphere of radius $\sqrt{nP}$ and a maximum likelihood decoder. Tan and Tomamichel [4] use the same distribution and decoder to prove the achievability of a maximal rate of

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

under blocklength n and average error probability ϵ, 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 [5]. The first- and second-order terms in (2) remain the same under maximal-error and both maximal- and average-power constraints across codewords; they differ under average-error and average-power constraints [6, Ch. 4]. In this paper, we only consider average-error and maximal-power constraints.

Extending the asymptotic expansion in (2) to a Gaussian MAC, in which multiple transmitters communicate independent messages to a single receiver over a Gaussian channel with blocklength n, is not trivial. MolavianJazi and Laneman [7] and Scarlett et al. [8] generalize the result in (2) to the two-transmitter 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 [7] uses codewords uniformly distributed on the power sphere and threshold decoding based on the *mutual information random variable*; the bound in [8] 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 maximum likelihood decoding.

The literature on RAC communications includes works like [9], [10], [11], where the number of active transmittters

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.

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

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 [12], [14], [15]. Chen and Guo [14] find the capacity of the Gaussian many access channel, and Chen et al. [15] 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 [16] and Zadik et al. [17] 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 [18] and [19], respectively. In this work, K does not grow with n.

In [20], 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 the decoding time n_t depends on the receiver's estimate t of the number of active transmitters. Decoding occurs at one of a fixed collection of possible decoding times $n_0, \dots, n_K$, where K is the maximal number of transmitters. At each decoding time, the receiver makes an attempt to decode by applying a single threshold rule; the receiver sends a single-bit ACK/NACK feedback to all transmitters in order to specify when communication is completed. In [21], Liu and Effros achieve improved third-order bounds using a maximum-likelihood decoder. Although the coding strategies proposed in [20], [21] apply to the Gaussian RAC, the random encoder design in [20] uses an i.i.d. input distribution. As shown in [22], 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 coding scheme for the Gaussian RAC. In the proposed code design, random codewords are designed by concatenating K codewords of blocklengths $n_1, n_2 - n_1, \dots, n_K - n_{K-1}$, 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 decoder. 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 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, respectively. 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, respectively. 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 $[k] \triangleq \{1, \dots, k\}$. 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 $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ 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 $\mathbf{x}_1, \dots, \mathbf{x}_K$ 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 $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$, we write $\mathbf{x} \stackrel{\pi}{=} \mathbf{y}$ if there exists a permutation π of elements of $\mathbf{y}$ such that $\mathbf{x} = \pi(\mathbf{y})$, and $\mathbf{x} \stackrel{\pi}{\neq} \mathbf{y}$ if $\mathbf{x} \neq \pi(\mathbf{y})$ for all permutations π of 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 $1\{\cdot\}$. Unless specified otherwise, 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 $\mathbb{S}^{n-1}(r) \triangleq \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = r\}$.

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 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 $n > n_0$.

III. An RCU Bound and its Analysis for the Gaussian Multiple Access Channel

A. An RCU Bound for General MACs

We begin by defining a two-transmitter MAC channel code.

TABLE I
VECTOR NOTATIONS

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

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

$$\frac{1}{M_1M_2}\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}\mathbb{P}\big[\mathsf{g}(Y_2)\neq(m_1,m_2) \mid (X_1,X_2)=(\mathsf{f}_1(m_1),\mathsf{f}_2(m_2))\big]\leq\epsilon, \quad (4)$$

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

We define the mutual 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 mutual information 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 the random-coding union (RCU) achievability bound of Polyanskiy et al. [5, Th. 16] to the MAC. The proof, derived earlier by Liu and Effros [21] in their work on the analysis of LDPC codes and inspired by a new RCU bound for the Slepian-Wolf setting [23, Th. 2], combines random code design and maximum likelihood decoding. 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): Fix input distributions P_{X_1} and P_{X_2}. Let $P_{X_1,\bar{X}_1,X_2,\bar{X}_2,Y_2}(x_1,\bar{x}_1,x_2,\bar{x}_2,y)=P_{X_1}(x_1)P_{X_1}(\bar{x}_1)P_{X_2}(x_2)P_{X_2}(\bar{x}_2)P_{Y_2|X_1X_2}(y|x_1,x_2)$. There exists an (M_1,M_2,ϵ)-MAC code for $P_{Y_2|X_1X_2}$ such that

$$\epsilon\leq\mathbb{E}\Big[\min\Big\{1,(M_1-1)\mathbb{P}\big[\imath_1(\bar{X}_1;Y_2|X_2)\geq\imath_1(X_1;Y_2|X_2)\mid X_1,X_2,Y_2\big]+(M_2-1)\mathbb{P}\big[\imath_2(\bar{X}_2;Y_2|X_1)\geq\imath_2(X_2;Y_2|X_1)\mid X_1,X_2,Y_2\big]+(M_1-1)(M_2-1)\mathbb{P}\big[\imath_{1,2}(\bar{X}_1,\bar{X}_2;Y_2)\geq\imath_{1,2}(X_1,X_2;Y_2)\mid X_1,X_2,Y_2\big]\Big\}\Big]. \quad (7)$$

Proof: The proof follows an argument similar to [5, Th. 16] (for point-to-point channels) and [23] (for multiple access source coding). The codewords $X_1(m_1)$, $m_1\in[M_1]$ and $X_2(m_2)$, $m_2\in[M_2]$ are drawn i.i.d. from P_{X_1} and P_{X_2}, respectively, and independently of each other. At the receiver, a maximum likelihood decoder chooses the message pair (m_1,m_2) with the maximum information density $\imath_{1,2}(X_1(m_1),X_2(m_2);Y_2)$. We bound the average probability of error from above as

$$\epsilon\leq\mathbb{P}\Bigg[\bigcup_{(j,k)\neq(1,1)}\{\imath_{1,2}(X_1(j),X_2(k);Y_2)\geq\imath_{1,2}(X_1(1),X_2(1);Y_2)\}\Bigg] \quad (8)$$

$$=\mathbb{E}\Bigg[\mathbb{P}\Bigg[\bigcup_{(j,k)\neq(1,1)}\{\imath_{1,2}(X_1(j),X_2(k);Y_2)\geq\imath_{1,2}(X_1(1),X_2(1);Y_2)\}\,\Bigg|\,X_1(1),X_2(1),Y_2\Bigg]\Bigg] \quad (9)$$

$$\leq\mathbb{E}\Big[\min\Big\{1,(M_1-1)\big[\imath_{1,2}(\bar{X}_1,X_2;Y_2)\geq\imath_{1,2}(X_1,X_2;Y_2)\mid X_1,X_2,Y_2\big]+(M_2-1)\mathbb{P}\big[\imath_{1,2}(X_1,\bar{X}_2;Y_2)\geq\imath_{1,2}(X_1,X_2;Y_2)\mid X_1,X_2,Y_2\big]+(M_1-1)(M_2-1)\mathbb{P}\big[\imath_{1,2}(\bar{X}_1,\bar{X}_2;Y_2)\geq\imath_{1,2}(X_1,X_2;Y_2)\mid X_1,X_2,Y_2\big]\Big\}\Big], \quad (10)$$

where (10) follows by choosing the tighter bound on the probability term between 1 and the union bound. Notice that the right-hand side of (10) is equal to the right-hand side of (7), since we can expand the mutual information density $\imath_{1,2}(x_1,x_2;y)$ as

$$\imath_{1,2}(x_1,x_2;y)=\imath_1(x_1;y|x_2)+\imath_2(x_2;y)=\imath_2(x_2;y|x_1)+\imath_1(x_1;y), \quad (11)$$

where $\imath_i(x_i;y)\triangleq\log\frac{P_{Y_2|X_i}(y|x_i)}{P_{Y_2}(y)}$, $i\in\{1,2\}$. Since the average error probability of randomly generated codewords is bounded by the right-hand side of (7), there exists a code satisfying (7). ∎

Remark 1: Theorem 1 generalizes to the K-transmitter MAC in a straightforward manner. Define the conditional mutual information densities for the K-transmitter MAC as

$$\imath_\mathcal{S}(x_\mathcal{S};y|x_{\mathcal{S}^c})=\log\frac{P_{Y_K|X_{[K]}}(y|x_{[K]})}{P_{Y_K|X_{\mathcal{S}^c}}(y|x_{\mathcal{S}^c})}, \quad (12)$$

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

$$\imath_{[K]}(x_{[K]};y)=\log\frac{P_{Y_K|X_{[K]}}(y|x_{[K]})}{P_{Y_K}(y)}. \quad (13)$$

Following identical arguments to those in the proof of Theorem 2, the inequality in (7) is extended to the K-transmitter MAC as

$$\epsilon \leq \mathbb{E}\Bigg[\min\Bigg\{1, \sum_{\mathcal{S}\in\mathcal{P}([K])} \Bigg(\prod_{s\in\mathcal{S}}(M_s-1)\Bigg)\mathbb{P}\big[\imath_{\mathcal{S}}(\bar{X}_{\mathcal{S}};Y_K|X_{\mathcal{S}^c}) \geq \imath_{\mathcal{S}}(X_{\mathcal{S}};Y_K|X_{\mathcal{S}^c}) \mid X_{[K]},Y_K\big]\Bigg\}\Bigg]. \tag{14}$$

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 two 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

$$\mathbf{Y}_k = \mathbf{X}_{\langle[k]\rangle} + \mathbf{Z}. \tag{15}$$

The channel transition law induced by (15) can be written as

$$P_{\mathbf{Y}_k|\mathbf{X}_{[k]}}(\mathbf{y}|\mathbf{x}_{[k]}) = \prod_{i=1}^n P_{Y_k|X_{[k]}}(y_i|x_{1i},\ldots,x_{ki}), \tag{16}$$

where

$$P_{Y_k|X_{[k]}}(y|x_{[k]}) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{\left(y-x_{\langle[k]\rangle}\right)^2}{2}\right\}. \tag{17}$$

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

$$Q_{\rm inv}(\mathsf{V},\epsilon) = \left\{\mathbf{z}\in\mathbb{R}^d : \mathbb{P}\left[\mathbf{Z}\leq\mathbf{z}\right]\geq 1-\epsilon\right\}. \tag{18}$$

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

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

$$\mathbf{C}(P_1,P_2) \triangleq \begin{bmatrix} C(P_1)\\ C(P_2)\\ C(P_{\langle[2]\rangle})\end{bmatrix}. \tag{19}$$

The dispersion matrix for the two-transmitter Gaussian MAC is defined as

$$\mathsf{V}(P_1,P_2) \triangleq \begin{bmatrix} V(P_1) & V_{1,2}(P_1,P_2) & V_{1,12}(P_1,P_2)\\ V_{1,2}(P_1,P_2) & V(P_2) & V_{2,12}(P_1,P_2)\\ V_{1,12}(P_1,P_2) & V_{2,12}(P_1,P_2) & V(P_{\langle[2]\rangle})+V_{12}(P_1,P_2)\end{bmatrix} \tag{20}$$

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

$$V_{1,2}(P_1,P_2) = \frac{1}{2}\frac{P_1P_2}{(1+P_1)(1+P_2)}, \tag{21}$$

$$V_{i,12}(P_1,P_2) = \frac{1}{2}\frac{P_i(2+P_{\langle[2]\rangle})}{(1+P_i)(1+P_{\langle[2]\rangle})}, \quad i\in\{1,2\}, \tag{22}$$

$$V_{12}(P_1,P_2) = \frac{P_1P_2}{(1+P_{\langle[2]\rangle})^2}. \tag{23}$$

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

$$\begin{bmatrix}\log M_1\\ \log M_2\\ \log M_1M_2\end{bmatrix} \in n\mathbf{C}(P_1,P_2) - \sqrt{n}Q_{\rm inv}(\mathsf{V}(P_1,P_2),\epsilon) + \frac{1}{2}\log n\mathbf{1} + O(1)\mathbf{1}. \tag{24}$$

Proof: The proof employs a refined asymptotic analysis of the bound in Theorem 1 with uniform distributions P_{X_1} and P_{X_2} on $\mathbb{S}^{n-1}\left(\sqrt{nP_1}\right)$ and $\mathbb{S}^{n-1}\left(\sqrt{nP_2}\right)$, respectively. ∎

Our third-order achievability result extends to the general K-transmitter Gaussian MAC. An $(n, M_{[K]}, \epsilon, P_{[K]})$-MAC code for the K-transmitter Gaussian MAC with the message set sizes $M_1,\ldots,M_K$, and the power constraints $P_1,\ldots,P_K$ is a natural extension of the two-transmitter MAC code given in Definition 2. The following theorem formally states the achievable region for the K-transmitter Gaussian MAC.

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

$$\left(\sum_{s\in\mathcal{S}}\log M_s\colon \mathcal{S}\in\mathcal{P}([K])\right) \in n\mathbf{C}(P_{[K]}) - \sqrt{n}Q_{\rm inv}(\mathsf{V}(P_{[K]}),\epsilon) + \frac{1}{2}\log n\mathbf{1} + O(1)\mathbf{1}, \tag{25}$$

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

$$\mathbf{C}(P_{[K]}) = \left(C(P_{\langle\mathcal{S}\rangle})\colon \mathcal{S}\in\mathcal{P}([K])\right)\in\mathbb{R}^{2^K-1}, \tag{26}$$

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

$$\mathsf{V}_{\mathcal{S}_1,\mathcal{S}_2}(P_{[K]}) \triangleq \frac{P_{\langle\mathcal{S}_1\rangle}P_{\langle\mathcal{S}_2\rangle}+2P_{\langle\mathcal{S}_1\cap\mathcal{S}_2\rangle}+\left(P_{\langle\mathcal{S}_1\cap\mathcal{S}_2\rangle}\right)^2-P^2_{\langle\mathcal{S}_1\cap\mathcal{S}_2\rangle}}{2(1+P_{\langle\mathcal{S}_1\rangle})(1+P_{\langle\mathcal{S}_2\rangle})}. \tag{27}$$

Proof: Section VI. ∎

Before concluding this section, we make several remarks on our achievability results in 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 [7] and Scarlett et al. [8], and improves the third-order term from $O\left(n^{1/4}\right)\mathbf{1}$ in [7] and $O\left(n^{1/4}\log n\right)\mathbf{1}$ in [8] to $\frac{1}{2}\log n\mathbf{1}+O(1)\mathbf{1}$.

2) Our proof technique in Theorem 2 differs from the technique in [7] in two key ways. First, we use a maximum likelihood decoder while [7] relies on a set of simultaneous threshold rules based on unconditional and conditional mutual information densities; 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 mutual information 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-Esséen theorem. However, when the inner product of the inputs $\langle \mathbf{X}_1, \mathbf{X}_2 \rangle$ equals a fixed constant, the mutual information random vector $\boldsymbol{\imath}_2$ can be written as a sum of independent random vectors. Therefore, we apply the Berry-Esséen 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, and 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.
In [7], a bound is obtained on the probability that the mutual information 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, they apply a central limit theorem for functions of sums [7, Proposition 1]. The resulting rate of convergence to normality in [7] is $O\left(\frac{1}{n^{1/4}}\right)$. 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 [7] in fact 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 [6, Th. 53] and for the point-to-point Gaussian channel [4]. In [6, Th. 53], the application of the RCU bound uses a refined large deviations result [5, Lemma 47]. However, using non-i.i.d. input distribution for the Gaussian channel prevents the direct application of the large deviation result given in [5, Lemma 47]. In [4, eq. (52)], Tan and Tomamichel derive a bound that replaces the large deviations result [5, Lemma 47] for the point-to-point Gaussian channel in order to accommodate the codewords drawn uniformly on an n-dimensional sphere. While evaluating the RCU bound in this paper, we extend the bound in [4, eq. (52)] to the Gaussian MAC.
4) For the symmetric setting, that is $P_i = P$ and $M_i = M$ for $i \in [K]$, Theorem 3 reduces to the scalar inequality below. This result refines the result in [7, 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

$$K\log M \leq nC(KP) - \sqrt{n(V(KP) + V_{\text{cr}}(K,P))}Q^{-1}(\epsilon) + \frac{1}{2}\log n + O(1). \tag{28}$$

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

$$V_{\text{cr}}(K,P) \triangleq \frac{K(K-1)P^2}{2(1+KP)^2}. \tag{29}$$

Proof: Appendix D. ■
5) In [24], Fong and Tan derive a converse for the Gaussian MAC with second-order term $O(\sqrt{n\log n})\mathbf{1}$. This converse does not match the second-order term in the achievability bounds proven in this paper. The gap in the second-order analyses of current MAC achievability and converse results is a challenging open problem, as discussed in [25].

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

A. System Model

Channel model: In order to capture the scenario of a memoryless Gaussian channel with K possible transmitters, a single receiver, and an unknown activity pattern $\mathcal{A} \subseteq [K]$ describing which transmitters are active, we describe the Gaussian RAC by a family of Gaussian MACs $\{P_{Y_k|X_{[k]}}\}_{k=0}^{K}$ (17), each indexed by the number of active transmitters $k \in \{0, \dots, K\}$. We choose a *compound channel* model in order to avoid the need to assign an activity probability to each transmitter.

Communication strategy: We adapt the epoch-based *rateless* communication strategy we put forth in [20] to achieve the fundamental limits of the Gaussian RAC. Each transmitter is either active or silent during a whole epoch. When the decoder can decode, it broadcasts a positive acknowledgment bit (ACK) to all transmitters, thereby ending the current epoch and staring the next. The length of the epoch is n_t, where t is the decoder's estimate of the number of transmitters and $n_0, n_1, \dots, n_K$ is a fixed collection of possible decoding times. As in [16], [20], 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. Messages $W_{\mathcal{A}}$ are independent and uniformly distributed over 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 transmitters k, and therefore that the decoding time n_k increases with k. We prove that $n_0 < \cdots < n_K$ is optimal for the Gaussian RAC. (See [20, Lemma 1] for more general sufficient conditions under which $n_0 < \cdots < n_K$.) At decoding time n_K, the decoder sees

$$\mathbf{Y}_k = \mathbf{X}_{\langle [k] \rangle} + \mathbf{Z} \in \mathbb{R}^{n_K} \quad \text{for } k \in [K], \tag{30}$$

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, $\mathbf{Y}_0 = \mathbf{Z}$. At times $n_t < n_K$, the decoder has access to the first n_t dimensions of $\mathbf{Y}_k$.

As in [20], 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: Given $0 < n_0 < n_1 < \cdots < n_K$, an $\left(\{n_j, \epsilon_j\}_{j=0}^K, M, P\right)$-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 $k = 0, \ldots, K$. The codewords satisfy the maximal-power constraints

$$\left\| \mathsf{f}(u, m)^{[n_j]} \right\|^2 \le n_j P \text{ for } m \in [M], u \in \mathcal{U}, j \in [K]. \tag{31}$$

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

$$\frac{1}{M^k} \sum_{m_{[k]} \in [M]^k} \mathbb{P}\Bigg[\Bigg\{ \bigcup_{t<k} \left\{ \mathsf{g}_t(U, \mathbf{Y}_k^{[n_t]}) \neq \mathsf{e} \right\} \Bigg\} \bigcup \\ \left\{ \mathsf{g}_k(U, \mathbf{Y}_k^{[n_k]}) \overset{\pi}{\neq} m_{[k]} \right\} \Bigg| \mathbf{X}_{[k]}^{[n_k]} = \mathsf{f}(U, m_{[k]})^{[n_k]} \Bigg] \le \epsilon_k, \tag{32}$$

where $\mathsf{f}(U, m_i)$ is the codeword for the message $m_i \in [M]$, U is the common randomness random variable[1], and the output $\mathbf{Y}_k$ is generated according to (30). If no transmitters are active, then the unique message $\{0\}$ is decoded with probability of error bounded as

$$\mathbb{P}\left[\mathsf{g}_0(U, \mathbf{Y}_0^{[n_0]}) \neq 0 \right] \le \epsilon_0. \tag{33}$$

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 $k \in \{0, 1, \ldots, K\}$, and M. An $\left(\{n_j, \epsilon_j\}_{j=0}^K, M, P\right)$-RAC code exists for the Gaussian RAC with a total of K transmitters provided that

$$k \log M \le n_k C(kP) - \sqrt{n_k (V(kP) + V_{\text{cr}}(k, P))} Q^{-1}(\epsilon_k) \\ + \frac{1}{2} \log n_k + O(1) \tag{34}$$

for $k \in [K]$, and

$$n_0 \ge c \log n_1 + o(\log n_1) \tag{35}$$

for some constant $c > 0$, where $C(\cdot)$ and $V(\cdot)$ are the capacity (1) and the dispersion (3) functions, respectively, and $V_{\text{cr}}(\cdot, \cdot)$ is the cross dispersion term (29).

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

Theorem 5: Fix constants $K < \infty$, $\epsilon_k \in (0, 1)$, $\lambda_k > 0$ for $k \in \{0, \ldots, K\}$, $n_0 < n_1 < \cdots < n_K$, $P > 0$, M, and a distribution $P_{\mathbf{X}}$ on $\mathbb{R}^{n_K}$. Then, there exists an $\left(\{n_j, \epsilon_j\}_{j=0}^K, M, P\right)$-RAC code with

$$\epsilon_0 \le \mathbb{P}\left[\left| \left\| \mathbf{Y}_0^{[n_0]} \right\|^2 - n_0 \right| > n_0 \lambda_0 \right] \tag{36}$$

$$\epsilon_k \le \frac{k(k-1)}{2M} + \mathbb{P}\left[\cup_{i=1}^k \cup_{j=1}^k \left\{ \left\| \mathbf{X}_i^{[n_j]} \right\|^2 > n_j P \right\} \right] \tag{37a}$$

$$+ \mathbb{P}\Big[\cup_{t=0}^{k-1} \left\{ \left| \left\| \mathbf{Y}_k^{[n_t]} \right\|^2 - n_t(1 + tP) \right| \le n_t \lambda_t \right\} \\ \bigcup \left\{ \left| \left\| \mathbf{Y}_k^{[n_k]} \right\|^2 - n_k(1 + kP) \right| > n_k \lambda_k \right\} \Big] \tag{37b}$$

$$+ \mathbb{E}\Bigg[\min \Bigg\{ 1, \sum_{s=1}^k \binom{k}{s} \binom{M-k}{s} \\ \mathbb{P}\Big[\imath_{[s]}(\bar{\mathbf{X}}_{[s]}^{[n_k]}; \mathbf{Y}_k^{[n_k]} | \mathbf{X}_{[s+1:k]}^{[n_k]}) \\ \ge \imath_{[s]}(\mathbf{X}_{[s]}^{[n_k]}; \mathbf{Y}_k^{[n_k]} | \mathbf{X}_{[s+1:k]}^{[n_k]}) \mid \mathbf{X}_{[k]}^{[n_k]}, \mathbf{Y}_k^{[n_k]} \Big] \Bigg\} \Bigg] \tag{37c}$$

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 $P_{\mathbf{X}_{[K]}, \bar{\mathbf{X}}_{[K]}, \mathbf{Y}_k}(\mathbf{x}_{[K]}, \bar{\mathbf{x}}_{[K]}, \mathbf{y}_k) = \left(\prod_{j \in [K]} P_{\mathbf{X}}(\mathbf{x}_j) P_{\mathbf{X}}(\bar{\mathbf{x}}_j) \right) P_{\mathbf{Y}_k | \mathbf{X}_{[k]}}(\mathbf{y}_k | \mathbf{x}_{[k]})$, and $P_{\mathbf{Y}_k | \mathbf{X}_{[k]}}$ is given in (30).

Proof: The terms in (37a) correspond to the probability that at least two transmitters send the same message, and the probability of a power violation, respectively. The probability in (37b) corresponds to the probability that the decoder decodes at a wrong decoding time, and the expectation in (37c) corresponds to the probability that the decoder decodes a wrong message list at the right 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 communication scheme when the set of active transmitters is known (Corollary 1). In other words, the first three terms on the right-hand side of (34) for k active transmitters match the first three terms of the largest achievable sum-rate in our achievability bound in (28) 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 $\mathbb{S}^{n_1 - 1}(\sqrt{n_1 P})$, the symbols indexed from $n_{j-1}+1$ to n_j are drawn uniformly from $\mathbb{S}^{n_j - n_{j-1} - 1}(\sqrt{(n_j - n_{j-1})P})$ for $j = 2, \ldots, K$, and these K spherically distributed sub-codewords are independent. Under such $P_{\mathbf{X}}$, the maximal-power constraints for each number of active transmitters given in (31) are satisfied with equality.

[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 [25, Appendix C] that the alphabet size of U is bounded by $K + 1$.

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 used.

3) As noted in [12], our achievability proofs leverage the fact that the number of active transmitters can be reliably estimated from the total received power. This is because the average received power at time n_k when k transmitters are active, i.e. $\frac{1}{n_k}\mathbb{E}\left[\left\|\mathbf{Y}_k^{[n_k]}\right\|^2\right]$, concentrates around its mean value, $1+kP$, which is distinct for each $k \in \{0, \ldots, K\}$. 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. For each k, if the average received power at time n_k lies on a small interval around $1+kP$, the decoder runs the maximum likelihood decoding rule, decodes a list of k messages, and sends a positive ACK to the transmitters; otherwise the decoder does not decode at time n_k, and sends a negative ACK to the transmitters informing them that they must keep transmitting until the next decoding time.
4) Theorem 5 applies verbatim to non-Gaussian RACs with power constraints satisfying the conditions in [20, 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 (34) depends on the number of active transmitters k, but not the total number of transmitters K. Therefore, even in the case of unbounded K, for every finite number of active transmitters k, the performance in (34) is still achieved by our proposed code. Not requiring to decode transmitter identity is crucial for this result to hold.
6) 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. In other words, requiring the power constraints on each sub-block of the codewords as
$$\left\|\mathsf{f}_i(m_i)^{[n_j]}\right\|^2 \leq n_j P_i \text{ for } m_i \in [M_i], i \in [K], j \in [K], \tag{38}$$
does not result in a performance loss in terms of the first three terms in the expansion in Theorem 3. The number of blocks K can be any positive integer as long as $n_j - n_{j-1} \geq c n_K$ for some constant $c > 0$ and all j. The supports of the distributions from which the codewords are drawn for the Gaussian MAC and RAC are illustrated in Fig. 1 for a small blocklength ($n = 3, K = 2$).
7) The coding strategy we propose in [20, Th. 1] requires an i.i.d. input distribution. One can employ the coding strategy in [20, Th. 1] to the Gaussian MAC drawing codewords i.i.d. from $\mathcal{N}(0, P')$ for some $P' = P - \delta$ and δ sufficiently small, and discarding codewords violating the maximal-power P constraint. However, [22, eq. (5.113)] shows that the resulting achievable second-order term is inferior to that achieved by the spherically distributed codewords.
8) The decoding rule used in Theorems 4 and 5 first checks if the received power at time n_i lies in a predetermined interval for $i = 0, 1, \ldots, K$. If the received power lies in that interval, then the decoder proceeds to run a maximum likelihood decoding rule for i transmitters; otherwise it does not decode any messages and waits until the next decoding time n_{i+1}. This process goes on until a decision is made or the largest decoding time n_K is reached. In this way, the number of active transmitters is estimated via a sequence of binary tests during an epoch. Another strategy for this purpose 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, so that they can modify their encoding function based on t. We show in Appendix E-A that employing this modified coding strategy only affects the $O(1)$ term in the expansion given in (34).
9) As in [25, Sec. V], by employing distinct encoders at the transmitters, the decoder can associate the transmitter identities with the decoded messages. We show that the first three terms of the expansion in (34) are still achievable in this setting. This scenario is discussed in Appendix E-B.

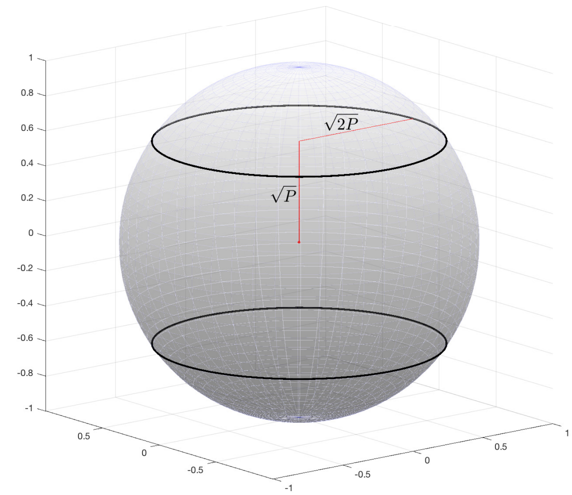

Fig. 1. Let $K = 2$, $n_1 = 2, n_2 = 3$, and $P_1 = P_2 = P = \frac{1}{3}$. The support of the input distribution for the Gaussian RAC is the Cartesian product of $\mathbb{S}^{n_1-1}(\sqrt{n_1 P})$ (here a circle with radius $\sqrt{2P}$) and $\mathbb{S}^{n_2-n_1-1}(\sqrt{(n_2-n_1)P})$ (here the set $\{-\sqrt{P},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

V. Proof of Theorem 2

A. Tools

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, $\mathbf{Y}_2 = \mathbf{X}_{\langle[2]\rangle} + \mathbf{Z}$, does not belong to its typical interval around $1 + 2P$.

Lemma 1, stated next, uniformly bounds the Radon-Nikodym derivative of the conditional and unconditional output distributions of the Gaussian MAC (16) in response to the spherical inputs with respect to the corresponding output distributions in response to the 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 [7, Proposition 2]): Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be independent and distributed uniformly over $\mathbb{S}^{n-1}(\sqrt{nP_1})$ and $\mathbb{S}^{n-1}(\sqrt{nP_2})$, respectively. Let $\tilde{\mathbf{X}}_i \sim \mathcal{N}(\mathbf{0}, P_i \mathsf{I}_n)$ for $i \in [2]$, 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 (16) with $k = 2$ transmitters. Then $\exists n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$, $\forall\, (\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \in \mathbb{R}^{n\otimes 3}$, it holds that

$$\frac{P_{\mathbf{Y}_2|\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_2)}{P_{\tilde{\mathbf{Y}}_2|\tilde{\mathbf{X}}_2}(\mathbf{y}|\mathbf{x}_2)} \leq \kappa_1(P_1) = 27\sqrt{\frac{\pi}{8}}\frac{1+P_1}{\sqrt{1+2P_1}} \tag{39}$$

$$\frac{P_{\mathbf{Y}_2}(\mathbf{y})}{P_{\tilde{\mathbf{Y}}_2}(\mathbf{y})} \leq \kappa_2(P_1, P_2) = \frac{9}{2\pi\sqrt{2}}\frac{P_{\langle[2]\rangle}}{\sqrt{P_1P_2}}. \tag{40}$$

If there is no additive noise $\mathbf{Z}$ in (16), (40) continues to hold.

Remark 2: Lemma 1 is generalized to the K-transmitter Gaussian MAC in [22, eq. (5.138)] as follows. Let $\mathbf{X}_i$ be independent and distributed uniformly over $\mathbb{S}^{n-1}(\sqrt{nP_i})$, and let $\tilde{\mathbf{X}}_i \sim \mathcal{N}(\mathbf{0}, P_i\mathsf{I}_n)$ for $i \in [K]$, 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 (16) with K transmitters. Then $\exists n_K \in \mathbb{N}$ such that $\forall n \geq n_K$, for any $\mathbf{x}_{[K]} \in \mathbb{R}^{n\otimes K}$, $\mathbf{y} \in \mathbb{R}^n$, and any non-empty $\mathcal{S} \in \mathcal{P}([K])$, it holds that

$$\frac{P_{\mathbf{Y}_K|\mathbf{X}_{\mathcal{S}^c}}(\mathbf{y}|\mathbf{x}_{\mathcal{S}^c})}{P_{\tilde{\mathbf{Y}}_K|\tilde{\mathbf{X}}_{\mathcal{S}^c}}(\mathbf{y}|\mathbf{x}_{\mathcal{S}^c})} \leq \kappa_{|\mathcal{S}|}(P_s : s \in \mathcal{S}), \tag{41}$$

where $\kappa_{|\mathcal{S}|}(P_s : s \in \mathcal{S})$ is a constant depending only on the set of powers $(P_s : s \in \mathcal{S})$. The proof of (41) relies on a recursive formula for the distribution of $\mathbf{Y}_K$.

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

Lemma 2 (Laurent and Massart [26, Lemma 1]): Let χ_n^2 be a chi-squared distributed random variable with n degrees of freedom. Then for $t > 0$,

$$\mathbb{P}\left[\chi_n^2 - n \geq 2\sqrt{nt} + 2t\right] \leq \exp\{-t\}, \tag{42}$$

$$\mathbb{P}\left[\chi_n^2 - n \leq -2\sqrt{nt}\right] \leq \exp\{-t\}. \tag{43}$$

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

Lemma 3 (Tan and Tomamichel [4, 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,

$$\mathbb{P}\left[Z_1 \in \left[\frac{a}{\sqrt{nP}}, \frac{a+\mu}{\sqrt{nP}}\right] \middle| \|\mathbf{x} + \mathbf{Z}\|^2 = ns\right] \leq \frac{L(P,s)\mu}{\sqrt{n}}, \tag{44}$$

where

$$L(P,s) \triangleq \frac{8(Ps)^{3/2}}{\sqrt{2\pi}}\sqrt{\frac{1+4Ps-\sqrt{1+4Ps}}{(\sqrt{1+4Ps}-1)^5}}. \tag{45}$$

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

Theorem 6 (Bentkus [27]): 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 $\mathrm{Cov}\,[\mathbf{S}] = \mathsf{I}_d$. Then, there exists a constant $c > 0$ such that

$$\sup_{\mathcal{A}\in\mathfrak{C}_d} |\mathbb{P}\left[\mathbf{S} \in \mathcal{A}\right] - \mathbb{P}[\mathbf{Z} \in \mathcal{A}]| \leq cd^{1/4}T, \tag{46}$$

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

Raic [28, Th. 1.1] establishes that the constant $cd^{1/4}$ in (46) can be replaced by $42d^{1/4} + 16$. For the case of general nonsingular $\mathrm{Cov}\,[\mathbf{S}]$, the following corollary to Theorem 6 is given by Tan and Kosut [29].

Corollary 2 (Tan and Kosut [29, Corollary 8]): For the setup in Theorem 6, assume that $\mathrm{Cov}\,[\mathbf{S}] = n\mathsf{V}$, 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

$$\sup_{\mathcal{A}\in\mathfrak{C}_d} \left|\mathbb{P}\left[\frac{1}{\sqrt{n}}\mathbf{S} \in \mathcal{A}\right] - \mathbb{P}[\mathbf{Z} \in \mathcal{A}]\right| \leq \frac{cd^{1/4}T}{\sqrt{n}\lambda_{\min}(\mathsf{V})^{3/2}}. \tag{47}$$

Lemma 4 and 5, below, are used to bound the probability that the mutual information 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

$$\begin{aligned}\mathrm{TV}(P_X, P_Y) &\triangleq \sup_{\mathcal{D}\in\mathbb{R}^d} |\mathbb{P}\left[X \in \mathcal{D}\right] - \mathbb{P}\left[Y \in \mathcal{D}\right]| \\ &= \frac{1}{2}\int_{x\in\mathbb{R}^d} |dP_X(x) - dP_Y(x)|\,. \end{aligned} \tag{48}$$

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, the total variation distance $\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}_1, \Sigma_1), \mathcal{N}(\boldsymbol{\mu}_2, \Sigma_2))$ is bounded as

$$\begin{aligned}&\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}_1, \Sigma_1), \mathcal{N}(\boldsymbol{\mu}_2, \Sigma_2)) \\ &\leq \frac{2+\sqrt{6}}{4}\left\|\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathsf{I}_d\right\|_F \\ &\quad + \frac{1}{2}\sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\Sigma_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}, \end{aligned} \tag{49}$$

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

Proof: Appendix B. ■

A weaker version of the bound in Lemma 4 has been recently shown by Devroye et al. [30, Th. 1.1] with a factor of 1.5 in front of the Frobenius norm. Like our proof, the proof of [30, Th. 1.1] relies on Pinsker's inequality. We improve the factor from 1.5 to $\frac{2+\sqrt{6}}{4} \approx 1.113$ by using the result in

[31, 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 (49).

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 over $\mathbb{S}^{n-1}(\sqrt{n})$ and the k-dimensional standard Gaussian random vector.

Lemma 5 (Stam [32, Th. 2]): Let $\tilde{\mathbf{Q}} \sim \mathcal{N}(\mathbf{0}, \mathsf{I}_k)$. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be distributed uniformly over $\mathbb{S}^{n-1}\sqrt{n}$. Let $\mathbf{X}^k = (X_1, \ldots, X_k)$ contain the first k coordinates of $\mathbf{X}$. Then

$$\mathrm{TV}(P_{\mathbf{X}^k}, \mathcal{N}(\mathbf{0}, \mathsf{I}_k)) \leq n^{\frac{1}{2}k}(n-k-2)^{-\frac{1}{2}k} - 2, \quad n > k+2. \tag{50}$$

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-Esséen 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 [4].

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 n-dimensional spheres centered at the origin, with radii $\sqrt{nP_1}$ and $\sqrt{nP_2}$, respectively:

$$P_{\mathbf{X}_1}(\mathbf{x}_1)P_{\mathbf{X}_2}(\mathbf{x}_2) = \frac{\delta(\|\mathbf{x}_1\|^2 - nP_1)}{S_n(\sqrt{nP_1})}\frac{\delta(\|\mathbf{x}_2\|^2 - nP_2)}{S_n(\sqrt{nP_2})}, \tag{51}$$

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

$$S_n(r) = \frac{2\pi^{n/2}}{\Gamma(n/2)}r^{n-1} \tag{52}$$

is the surface area of an n-dimensional sphere with radius r. We draw M_1 and M_2 independent codewords from $P_{\mathbf{X}_1}$ and $P_{\mathbf{X}_2}$, respectively. These codewords are denoted 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 (16). Upon receiving the output sequence $\mathbf{y}$, the decoder employs a maximum likelihood decoding rule, given by

$$\mathsf{g}(\mathbf{y}) = \begin{cases} (m_1, m_2) & \text{if } \imath_{1,2}(\mathsf{f}_1(m_1), \mathsf{f}_2(m_2); \mathbf{y}) \\ & > \imath_{1,2}(\mathsf{f}_1(m_1'), \mathsf{f}_2(m_2'); \mathbf{y}) \\ & \text{for all } (m_1', m_2') \neq (m_1, m_2), \\ & (m_1', m_2') \in [M_1] \times [M_2], \\ \text{error} & \text{otherwise.} \end{cases} \tag{53}$$

We disregard the ties in (53) because the event 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 (16) and the spherical input distributions (51) into (5a)–(5c), we compute for any $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \in \mathbb{R}^{n \otimes 3}$

$$\imath_1(\mathbf{x}_1; \mathbf{y}|\mathbf{x}_2) = \frac{n}{2}\log\frac{1}{2\pi} + \langle \mathbf{y} - \mathbf{x}_2, \mathbf{x}_1 \rangle - \frac{\|\mathbf{y} - \mathbf{x}_2\|^2}{2} - \frac{nP_1}{2} - \log P_{\mathbf{Y}_2|\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_2), \tag{54}$$

$$\imath_2(\mathbf{x}_2; \mathbf{y}|\mathbf{x}_1) = \frac{n}{2}\log\frac{1}{2\pi} + \langle \mathbf{y} - \mathbf{x}_1, \mathbf{x}_2 \rangle - \frac{\|\mathbf{y} - \mathbf{x}_1\|^2}{2} - \frac{nP_2}{2} - \log P_{\mathbf{Y}_2|\mathbf{X}_1}(\mathbf{y}|\mathbf{x}_1), \tag{55}$$

$$\imath_{1,2}(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}) = \frac{n}{2}\log\frac{1}{2\pi} + \langle \mathbf{y}, \mathbf{x}_1 + \mathbf{x}_2 \rangle - \frac{\|\mathbf{y}\|^2}{2} - \frac{\|\mathbf{x}_1 + \mathbf{x}_2\|^2}{2} - \log P_{\mathbf{Y}_2}(\mathbf{y}). \tag{56}$$

Observe that $\imath_1(\mathbf{x}_1; \mathbf{y}|\mathbf{x}_2)$ depends on $\mathbf{x}_1$ only through the inner product $\langle \mathbf{y} - \mathbf{x}_2, \mathbf{x}_1 \rangle$, and $\imath_{1,2}(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y})$ 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 (15), the conditional mutual information density for two transmitters, $\imath_1(\mathbf{x}_1; \mathbf{y}|\mathbf{x}_2)$, can be reduced to the unconditional mutual information density for a single transmitter as

$$\imath_1(\mathbf{x}_1; \mathbf{y}|\mathbf{x}_2) = \imath_1(\mathbf{x}_1; \mathbf{y} - \mathbf{x}_2) = \log\frac{P_{\mathbf{Y}_1|\mathbf{X}_1}(\mathbf{y} - \mathbf{x}_2|\mathbf{x}_1)}{P_{\mathbf{Y}_1}(\mathbf{y} - \mathbf{x}_2)}, \tag{57}$$

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 independent from 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 $n(1+P_{\langle[2]\rangle})$, respectively. We define a typicality set for the triplet $(\mathbf{X}_1 + \mathbf{Z}, \mathbf{X}_2 + \mathbf{Z}, \mathbf{Y}_2)$ by

$$\mathcal{F} \triangleq \bigtimes_{\mathcal{S} \in \mathcal{P}([2])} \mathcal{F}(\mathcal{S}) \subseteq \mathbb{R}^{3n}, \tag{58}$$

where

$$\mathcal{F}(\mathcal{S}) \triangleq \left\{ \mathbf{x}_{\langle \mathcal{S} \rangle} + \mathbf{z} \in \mathbb{R}^n \colon \frac{1}{n}\left\|\mathbf{x}_{\langle \mathcal{S} \rangle} + \mathbf{z}\right\|^2 \in \mathcal{I}(\mathcal{S}) \right\}, \tag{59}$$

$$\mathcal{I}(\mathcal{S}) \triangleq [1 + P_{\langle \mathcal{S} \rangle} - n^{-1/3}, 1 + P_{\langle \mathcal{S} \rangle} + n^{-1/3}]. \tag{60}$$

We will show that for a large enough n

$$\mathbb{P}\left[(\mathbf{X}_1 + \mathbf{Z}, \mathbf{X}_2 + \mathbf{Z}, \mathbf{Y}_2) \notin \mathcal{F}\right] \leq \exp\{-c_2 n^{1/3}\}, \tag{61}$$

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 a multiple of chi-squared distributed random variables with n degrees of freedom, and then use Lemma 2 to upper bound the two-sided tail probability of these chi-squared distributed random variables. Weakening the upper bound (42) 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 $\mathbf{Y}_2$

$$\mathbb{P}\left[\left|\|\mathbf{X}_1 + \mathbf{Z}\|^2 - n(1 + P_1)\right| > nt_1\right] \leq 2\kappa_1(P_1)\exp\left\{-\frac{nt_1^2}{8(1+P_1)^2}\right\}, \tag{62}$$

$$\mathbb{P}\left[\left|\|\mathbf{Y}_2\|^2 - n(1+P_{\langle[2]\rangle})\right| > nt_2\right] \leq 2\kappa_2(P_1,P_2)\exp\left\{-\frac{nt_2^2}{8(1+P_{\langle[2]\rangle})^2}\right\}, \quad (63)$$

for $t_1 \in (0, 1+P_1)$, and $t_2 \in (0, 1+P_{\langle[2]\rangle})$, where $\kappa_1(P_1)$ and $\kappa_2(P_1,P_2)$ are constants defined in Lemma 1. We deduce (61) by the union bound and setting $t_1 = t_2 = n^{-1/3}$ in (62)–(63).

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 [4, eq. (27)]

$$g_1(t;\mathbf{y},\mathbf{x}_2) \triangleq \mathbb{P}\left[\imath_1(\bar{\mathbf{X}}_1;\mathbf{Y}_2|\mathbf{X}_2) \geq t \mid \mathbf{X}_2 = \mathbf{x}_2, \mathbf{Y}_2 = \mathbf{y}\right] \quad (64)$$

$$g_2(t;\mathbf{y},\mathbf{x}_1) \triangleq \mathbb{P}\left[\imath_2(\bar{\mathbf{X}}_2;\mathbf{Y}_2|\mathbf{X}_1) \geq t \mid \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y}_2 = \mathbf{y}\right] \quad (65)$$

$$g_{1,2}(t;\mathbf{y}) \triangleq \mathbb{P}\left[\imath_{1,2}(\bar{\mathbf{X}}_1,\bar{\mathbf{X}}_2;\mathbf{Y}_2) \geq t \mid \mathbf{Y}_2 = \mathbf{y}\right], \quad (66)$$

where

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

The following lemma, which is a generalization of [4, eq. (53)] for the Gaussian MAC, gives upper bounds on these functions, and is used 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 (58). Then, for a large enough n,

$$g_1(t;\mathbf{y},\mathbf{x}_2) \leq \frac{G_1\exp\{-t\}}{\sqrt{n}}, \quad (67a)$$

$$g_2(t;\mathbf{y},\mathbf{x}_1) \leq \frac{G_2\exp\{-t\}}{\sqrt{n}}, \quad (67b)$$

$$g_{1,2}(t;\mathbf{y}) \leq \frac{G_{1,2}\exp\{-t\}}{\sqrt{n}}, \quad (67c)$$

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 (67a) and (67b) follow from the equivalence of the conditional mutual information density for two transmitters and the unconditional mutual information density for a single transmitter stated in (57), and the analysis in [4, Sec. IV-E]. The resulting constants in (67a) and (67b) are

$$G_i = (3\log 2)L(P_i, 1+P_i), \quad i \in \{1,2\}, \quad (68)$$

where $L(\cdot,\cdot)$ is the function defined in (45).

Bounding the function $g_{1,2}(t;\mathbf{y})$ is more challenging than bounding $g_1(t;\mathbf{y},\mathbf{x}_2)$; while $\|\mathbf{X}_1\|^2$ is a constant under a spherical input distribution, $\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2$ is not. The proof of (67c) follows from similar steps as [4, 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

$$g_{1,2}(t;\mathbf{y}) = \mathbb{P}[\exp\{-\imath_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2)\} 1\{\imath_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2) \geq t\} \mid \mathbf{Y}_2 = \mathbf{y}]. \quad (69)$$

In order to bound (69), we define the following function for any constants $a \in \mathbb{R}$ and $\mu > 0$

$$h_{1,2}(\mathbf{y};a,\mu) \triangleq \mathbb{P}\left[\imath_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2) \in [a, a+\mu] \middle| \mathbf{Y}_2 = \mathbf{y}\right] \quad (70)$$

$$= \mathbb{P}\left[\langle\mathbf{X}_{\langle[2]\rangle},\mathbf{Y}_2\rangle - \frac{\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2}{2} \in [a', a'+\mu] \middle| \mathbf{Y}_2 = \mathbf{y}\right] \quad (71)$$

where a' is some other real constant that is shifted from a by some amount depending on $\mathbf{y}$, and (71) follows from (56). Due to the spherical symmetry of the distribution of $\mathbf{Y}_2$, we observe that (71) depends on $\mathbf{y}$ only through its norm $\|\mathbf{y}\|$. Therefore,

$$h_{1,2}(s;a,\mu) \triangleq h_{1,2}(\mathbf{y};a,\mu) = \mathbb{P}\left[\langle\mathbf{X}_{\langle[2]\rangle},\mathbf{Y}_2\rangle - \frac{\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2}{2} \in [a', a'+\mu] \middle| \|\mathbf{Y}_2\|^2 = ns\right] \quad (72)$$

$$= \mathbb{E}\left[\mathbb{P}\left[\langle\mathbf{X}_{\langle[2]\rangle},\mathbf{Y}_2\rangle - \frac{\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2}{2} \in [a', a'+\mu] \middle| \|\mathbf{Y}_2\|^2 = ns, \left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2\right]\right], \quad (73)$$

where $\|\mathbf{y}\|^2 = ns$. Recall that the support of the norm $\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2$ is $[n(\sqrt{P_1}-\sqrt{P_2})^2, n(\sqrt{P_1}+\sqrt{P_2})^2]$. To avoid the cases where $\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2$ is too small, we separate the probability term (73) according to whether the event

$$\mathcal{B} = \left\{\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2 < n(P_{\langle[2]\rangle} - \sqrt{P_1P_2})\right\} \quad (74)$$

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 $(0, 2\sqrt{P_1P_2})$.

Conditioning on the event $\mathcal{B}$ in (73), and upper bounding the corresponding probability terms by 1 gives

$$h_{1,2}(s;a,\mu) \leq \mathbb{P}\left[\mathcal{B} \middle| \|\mathbf{Y}_2\|^2 = ns\right] + \mathbb{P}\left[\langle\mathbf{X}_{\langle[2]\rangle},\mathbf{Y}_2\rangle - \frac{\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2}{2} \in [a', a'+\mu] \middle| \|\mathbf{Y}_2\|^2 = ns, \mathcal{B}^c\right]. \quad (75)$$

We upper bound the first term in the right-hand side of (75) for a large enough n as

$$\mathbb{P}\left[\mathcal{B} \middle| \left\|\mathbf{X}_{\langle[2]\rangle} + \mathbf{Z}\right\|^2 = ns\right] \leq \exp\{-nC\}, \quad (76)$$

where $C > 0$ is a constant. The proof of (76) is given in Appendix A.

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

$$\mathbb{P}\left[\langle\mathbf{X}_{\langle[2]\rangle}, \mathbf{X}_{\langle[2]\rangle} + \mathbf{Z}\rangle - \frac{nu}{2} \in [a', a'+\mu] \middle| \left\|\mathbf{X}_{\langle[2]\rangle} + \mathbf{Z}\right\|^2 = ns, \left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2 = nu\right]$$

$$= \mathbb{P}\left[Z_1 + \frac{\sqrt{nu}}{2} \in \left[\frac{a'}{\sqrt{nu}}, \frac{a'+\mu}{\sqrt{nu}}\right] \middle| \|\mathbf{x}+\mathbf{Z}\|^2 = ns\right] \tag{77}$$

$$\leq \frac{L(u,s)\mu}{\sqrt{n}}, \tag{78}$$

$$\leq \frac{3}{2}\frac{L(u,1+P_{\langle[2]\rangle})\mu}{\sqrt{n}} \tag{79}$$

where (78) follows by Lemma 3, and (79) holds by the continuity of the map $s \mapsto L(u,s)$ using $n^{-1/3} \to 0$. Using (79), we bound the second term in (75) as

$$\mathbb{P}\Big[\langle \mathbf{X}_{\langle[2]\rangle}, \mathbf{Y}_2\rangle - \frac{\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2}{2} \in [a', a'+\mu] \Big| \left\|\mathbf{X}_{\langle[2]\rangle} + \mathbf{Z}\right\|^2 = ns, \mathcal{B}^c\Big] \leq \max_{u\in[P_{\langle[2]\rangle}-\sqrt{P_1P_2},(\sqrt{P_1}+\sqrt{P_2})^2]} \frac{3}{2}L(u,1+P_{\langle[2]\rangle})\frac{\mu}{\sqrt{n}}. \tag{80}$$

By (75), (76), (80), and because the map $u \mapsto L(u, 1+P_{\langle[2]\rangle})$ is bounded above in the interval $[P_{\langle[2]\rangle} - \sqrt{P_1P_2}, (\sqrt{P_1} + \sqrt{P_2})^2]$, for a large enough n, there exists a constant $K_2(P_1,P_2) > 0$ such that

$$h_{1,2}(s;a,\mu) \leq K_2(P_1,P_2)\frac{\mu}{\sqrt{n}}. \tag{81}$$

By following the same steps as [4, eq. (55)-(57)], we conclude that

$$g_{1,2}(t;\mathbf{y}) \leq \frac{G_{1,2}\exp\{-t\}}{\sqrt{n}}, \tag{82}$$

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

■

E. Evaluating the RCU Bound for the MAC

We now upper bound the right-hand side of (7) in Theorem 1. Define the typical events

$$\mathcal{E}(\mathcal{S}) = \left\{\mathbf{X}_{\langle\mathcal{S}\rangle} + \mathbf{Z} \in \mathcal{F}(\mathcal{S})\right\}, \tag{83}$$

$$\mathcal{E} = \bigcap_{\mathcal{S}\in\mathcal{P}([2])} \mathcal{E}(\mathcal{S}), \tag{84}$$

$$\mathcal{A} = \left\{\boldsymbol{\imath}_2 \geq \log\begin{bmatrix} M_1G_1 \\ M_2G_2 \\ M_1M_2G_{1,2}\end{bmatrix} - \frac{1}{2}\log n\mathbf{1}\right\}, \tag{85}$$

where G_1, G_2 and $G_{1,2}$ are the constants given in (67a)–(67c), and $\mathcal{F}(\mathcal{S})$ is defined in (59). Denote for brevity

$$g_1 \triangleq g_1(\imath_1(\mathbf{X}_1;\mathbf{Y}_2|\mathbf{X}_2);\mathbf{Y}_2,\mathbf{X}_2) \tag{86a}$$

$$g_2 \triangleq g_2(\imath_2(\mathbf{X}_2;\mathbf{Y}_2|\mathbf{X}_1);\mathbf{Y}_2,\mathbf{X}_1) \tag{86b}$$

$$g_{1,2} \triangleq g_{1,2}(\imath_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2);\mathbf{Y}_2). \tag{86c}$$

The right-hand side of (7) is bounded in (87)–(91) at the top of the next page. Here, c_2 is the positive constant defined in (61). The equality (87) 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 according to whether the event $\{\mathcal{A}^c \cup \mathcal{E}^c\}$ occurs or not. The inequality (88) follows by upper bounding the minimum inside the first expectation in (87) by 1; upper bounding the minimum inside the second expectation in (87) 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 (84) and (85), and distributing that multiplication over the summation. The inequality (89) follows from Lemma 6, and by upper bounding the probability terms by 1. The inequality (90) is obtained by applying the union bound to $\mathbb{P}[\mathcal{A}^c \cup \mathcal{E}^c]$, and by using Lemma 6 with $t = \log\frac{M_1G_1}{\sqrt{n}}$, $t = \log\frac{M_2G_2}{\sqrt{n}}$, and $t = \log\frac{M_1M_2G_{1,2}}{\sqrt{n}}$ to bound the three remaining terms, respectively. The inequality (91) follows from (61).

It only remains to evaluate the probability $\mathbb{P}[\mathcal{A}^c]$ in (91) to complete the proof of Theorem 2. We note that if the operational rate pair $\left(\frac{\log M_1}{n}, \frac{\log M_2}{n}\right)$ is not 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 $O\left(\frac{1}{\sqrt{n}}\right)$. For the corner points, $\mathbb{P}[\mathcal{A}^c]$ needs to be upper bounded without using the union bound in order to obtain a tighter achievability bound as discussed in [22, Sec. 5.1.1].

F. A Multidimensional Berry-Esséen Type Inequality

In this section, we will upper bound the probability $\mathbb{P}[\mathcal{A}^c]$ in (91). 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 mutual information densities whose denominators have the corresponding Gaussian distributions as

$$\tilde{\imath}_1(\mathbf{x}_1;\mathbf{y}|\mathbf{x}_2) \triangleq \sum_{i=1}^n \log\frac{P_{Y_2|X_1X_2}(y_i|x_{1i},x_{2i})}{P_{\tilde{Y}_2|\tilde{X}_2}(y_i|x_{2i})}, \tag{92a}$$

$$\tilde{\imath}_2(\mathbf{x}_2;\mathbf{y}|\mathbf{x}_1) \triangleq \sum_{i=1}^n \log\frac{P_{Y_2|X_1X_2}(y_i|x_{1i},x_{2i})}{P_{\tilde{Y}_2|\tilde{X}_1}(y_i|x_{1i})}, \tag{92b}$$

$$\tilde{\imath}_{1,2}(\mathbf{x}_1,\mathbf{x}_2;\mathbf{y}) \triangleq \sum_{i=1}^n \log\frac{P_{Y_2|X_1X_2}(y_i|x_{1i},x_{2i})}{P_{\tilde{Y}_2}(y_i)}, \tag{92c}$$

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

$$\tilde{\boldsymbol{\imath}}_2 \triangleq \frac{1}{\sqrt{n}}\left(\begin{bmatrix}\tilde{\imath}_1(\mathbf{X}_1;\mathbf{Y}_2|\mathbf{X}_2) \\ \tilde{\imath}_2(\mathbf{X}_2;\mathbf{Y}_2|\mathbf{X}_1) \\ \tilde{\imath}_{1,2}(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_2)\end{bmatrix} - n\mathbf{C}(P_1,P_2)\right), \tag{93}$$

where $\mathbf{C}(P_1,P_2) = \frac{1}{n}\mathbb{E}[\boldsymbol{\imath}_2]$ is the capacity vector defined in (19). Define the threshold vector

$$\boldsymbol{\tau} = \log\begin{bmatrix} M_1G_1\kappa_1(P_1) \\ M_2G_2\kappa_1(P_2) \\ M_1M_2G_{1,2}\kappa_2(P_1,P_2)\end{bmatrix} - \frac{1}{2}\log n\mathbf{1} - n\mathbf{C}(P_1,P_2). \tag{94}$$

We will explain our method to upper bound the probability $\mathbb{P}[\mathcal{A}^c]$ in 5 steps.

Step 1: To upper bound $\mathbb{P}[\mathcal{A}^c]$, 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. This step appears in [4, eq. (65)] for the point-to-point channel and [7, eq. (2)] for the MAC. Then we bound $\mathbb{P}[\mathcal{A}^c]$

$$
\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] \qquad (87)\\
&\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_1G_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_2G_2}{\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_2G_{1,2}}{\sqrt{n}}\right\}\,\middle|\,\mathcal{E}(\{1,2\})\right] \qquad (88)\\
&\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_1G_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_2G_2}{\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_2G_{1,2}}{\sqrt{n}}\right\}\,\middle|\,\mathcal{E}(\{1,2\})\right] \qquad (89)\\
&\leq \mathbb{P}\left[\mathcal{A}^c\right]+\mathbb{P}\left[\mathcal{E}^c\right]+\frac{G_1+G_2+G_{1,2}}{\sqrt{n}} \qquad (90)\\
&\leq \mathbb{P}\left[\mathcal{A}^c\right]+\exp\left\{-c_2 n^{1/3}\right\}+\frac{G_1+G_2+G_{1,2}}{\sqrt{n}} \qquad (91)
\end{aligned}
$$

---

in terms of the modified mutual information random vector $\tilde{\boldsymbol{\imath}}_2$. By (85) and Lemma 1, we have that

$$\mathbb{P}\left[\mathcal{A}^c\right]=1-\mathbb{P}\left[\boldsymbol{\imath}_2-\mathbb{E}\left[\boldsymbol{\imath}_2\right]\geq\left(\boldsymbol{\tau}-\log\begin{bmatrix}\kappa_1(P_1)\\ \kappa_1(P_2)\\ \kappa_2(P_1,P_2)\end{bmatrix}\right)\right] \tag{95}$$

$$\leq 1-\mathbb{P}\left[\tilde{\boldsymbol{\imath}}_2\geq\frac{1}{\sqrt{n}}\boldsymbol{\tau}\right]. \tag{96}$$

Our goal in Steps 1-5 is to upper bound the right-hand side of (96). From (92a)–(92c), we see that $\tilde{\boldsymbol{\imath}}_2$ is distributed the same as

$$\tilde{\boldsymbol{\imath}}_2\sim\frac{1}{\sqrt{n}}\begin{bmatrix}\frac{(n-\|\mathbf{Z}\|^2)P_1+2\langle\mathbf{X}_1,\mathbf{Z}\rangle}{2(1+P_1)}\\ \frac{(n-\|\mathbf{Z}\|^2)P_2+2\langle\mathbf{X}_2,\mathbf{Z}\rangle}{2(1+P_2)}\\ \frac{(n-\|\mathbf{Z}\|^2)(P_{\langle[2]\rangle})+2\langle\mathbf{X}_1,\mathbf{X}_2\rangle+2\langle\mathbf{Z},\mathbf{X}_{\langle[2]\rangle}\rangle}{2(1+P_{\langle[2]\rangle})}\end{bmatrix}. \tag{97}$$

The key observation here is that although (97) is not a sum of n independent random vectors, the conditional distribution of $\tilde{\boldsymbol{\imath}}_2$ given $(\mathbf{X}_1,\mathbf{X}_2)$ can be written as such a sum. Therefore, the multidimensional Berry-Esséen 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 the squared norms satisfy $\|\mathbf{x}_i\|^2=nP_i$ for $i\in[2]$. Define the normalized inner product random variable

$$Q\triangleq\frac{\langle\mathbf{X}_1,\mathbf{X}_2\rangle}{\sqrt{nP_1P_2}}, \tag{98}$$

and set

$$\mathbf{x}_1=(\sqrt{nP_1},0,\ldots,0), \tag{99}$$

$$\mathbf{x}_2=(q\sqrt{P_2},\sqrt{(n-q^2)P_2},0,\ldots,0), \tag{100}$$

which satisfy

$$\frac{\langle\mathbf{x}_1,\mathbf{x}_2\rangle}{\sqrt{nP_1P_2}}=q. \tag{101}$$

Putting (99)–(100) into (97), we get

$$\tilde{\boldsymbol{\imath}}_2|Q=q\sim\tilde{\boldsymbol{\imath}}_2|(\mathbf{X}_1,\mathbf{X}_2)=(\mathbf{x}_1,\mathbf{x}_2) \tag{102}$$

$$\sim\boldsymbol{\mu}(q)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{J}_i(q), \tag{103}$$

where

$$\boldsymbol{\mu}(q)\triangleq\mathbb{E}\left[\tilde{\boldsymbol{\imath}}_2|Q=q\right]=q\begin{bmatrix}0\\0\\ \frac{\sqrt{P_1P_2}}{1+P_{\langle[2]\rangle}}\end{bmatrix}, \tag{104}$$

$$\mathbf{J}_i(q)\triangleq\begin{bmatrix}\frac{(1-Z_i^2)P_1+2x_{1i}Z_i}{2(1+P_1)}\\ \frac{(1-Z_i^2)P_2+2x_{2i}Z_i}{2(1+P_2)}\\ \frac{(1-Z_i^2)(P_{\langle[2]\rangle})+2(x_{1i}+x_{2i})Z_i}{2(1+P_{\langle[2]\rangle})}\end{bmatrix},\quad i\in[n]. \tag{105}$$

$\mathbf{J}_i(q)$ depends on q through the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ given in (99)–(100). In (103), the modified mutual information random vector conditioned on Q is written as a sum of independent, but not identical random vectors.

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

$$Q \sim \frac{X_{21}}{\sqrt{P_2}}, \tag{106}$$

where X_{21} denotes the first coordinate of $\mathbf{X}_2$. Therefore, Q is distributed according to the marginal distribution of the first coordinate of a random vector distributed uniformly over $\mathbb{S}^{n-1}(\sqrt{n})$. The distribution of Q is computed as (e.g., [32, Th. 1])

$$P_Q(q) = \frac{\Gamma(\frac{n}{2})}{\sqrt{\pi n}\Gamma(\frac{n-1}{2})}\left(1 - \frac{q^2}{n}\right)_+^{\frac{n-3}{2}}, \tag{107}$$

where $\Gamma(\cdot)$ denotes the Gamma function, and $x_+ \triangleq \max\{0, x\}$ for all $x \in \mathbb{R}$. Clearly, the support of Q is $[-\sqrt{n}, \sqrt{n}]$. From (107), we compute

$$\mathbb{E}[Q] = 0, \quad \mathrm{Var}[Q] = 1. \tag{108}$$

By Sterling's approximation, it can be shown that $Q \to \mathcal{N}(0, 1)$ in distribution as $n \to \infty$ (e.g., [32, Th. 1]). Recall that an upper bound on the total variation distance between P_Q and $\mathcal{N}(0, 1)$ is given in Lemma 5.

From (103), we find the conditional covariance matrix of the modified mutual information random vector as

$$\Sigma(q) \triangleq \mathsf{Cov}[\tilde{\boldsymbol{\imath}}_2 | Q = q] = \Sigma + \frac{q}{\sqrt{n}}\mathsf{B}, \tag{109}$$

where

$$\Sigma \triangleq \begin{bmatrix} V(P_1) & V_{1,2}(P_1, P_2) & V_{1,12}(P_1, P_2) \\ V_{1,2}(P_1, P_2) & V(P_2) & V_{2,12}(P_1, P_2), \\ V_{1,12}(P_1, P_2) & V_{2,12}(P_1, P_2) & V(P_{\langle[2]\rangle}) \end{bmatrix}, \tag{110}$$

$$\mathsf{B} \triangleq \frac{\sqrt{P_1 P_2}}{(1+P_1)(1+P_2)(1+P_{\langle[2]\rangle})} \cdot \begin{bmatrix} 0 & 1+P_{\langle[2]\rangle} & 1+P_2 \\ 1+P_{\langle[2]\rangle} & 0 & 1+P_1 \\ 1+P_2 & 1+P_1 & \frac{(1+P_1)(1+P_2)}{(1+P_{\langle[2]\rangle})} \end{bmatrix}, \tag{111}$$

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), (21) and (22), respectively. Note that Σ and B depend only on P_1 and P_2. Using (104), (108), (109), by the law of total expectation and variance, we compute

$$\mathbb{E}[\tilde{\boldsymbol{\imath}}_2] = 0, \tag{112}$$

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

where $\mathsf{V}(P_1, P_2)$ is the dispersion matrix defined in (20).

Step 2: The goal of step 2 is to delineate how to approximate the distribution of $\tilde{\boldsymbol{\imath}}_2$ by an appropriate Gaussian. Toward that end, we consider some auxiliary random variables. Based on our observation in (103), we express the probability in the right-hand side of (96) by conditioning on Q and taking the expectation with respect to P_Q. Let $\mathcal{D}$ be any convex, Borel-measurable subset of $\mathbb{R}^3$. Define the probability measure $P_{\tilde{Q}}$, and the transition probability kernels $P_{\mathbf{V}|Q}$ and $P_{\mathbf{W}|Q}$ as

$$\tilde{Q} \sim \mathcal{N}(0, 1) \tag{114}$$

$$\mathbf{V}|Q = q \sim \begin{cases} \mathcal{N}(\boldsymbol{\mu}(q), \Sigma(q)) & \text{if } |q| < \sqrt{n} \\ \mathcal{N}(\boldsymbol{\mu}(q), \Sigma) & \text{if } |q| \geq \sqrt{n} \end{cases} \tag{115}$$

$$\mathbf{W}|Q = q \sim \mathcal{N}(\boldsymbol{\mu}(q), \Sigma) \qquad \text{for } q \in (-\infty, \infty). \tag{116}$$

Similar to $P_{\mathbf{V}|Q}$, we extend the definition of the kernel $P_{\tilde{\boldsymbol{\imath}}_2|Q}$ given in (103) to the outside of the support of P_Q by choosing $\tilde{\boldsymbol{\imath}}_2 | Q = q \sim \mathcal{N}(\boldsymbol{\mu}(q), \Sigma)$ for $|q| \geq \sqrt{n}$ in order for the joint distribution $P_{\tilde{Q}} P_{\tilde{\boldsymbol{\imath}}_2|Q}$ to be valid. Recall that $\tilde{Q}$ is Gaussian distributed with the same mean and variance as Q, and that $\mathbf{V}|Q = q$ has the same mean vector and covariance matrix as $\tilde{\boldsymbol{\imath}}_2 | Q = q$. The Gaussian kernel $P_{\mathbf{W}|Q}$ is obtained from $P_{\mathbf{V}|Q}$ by replacing its covariance matrix $\Sigma(Q)$ by its mean value with respect to Q, i.e. Σ.

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

$$P_{Q\,\tilde{\boldsymbol{\imath}}_2} = P_Q P_{\tilde{\boldsymbol{\imath}}_2|Q} \tag{117a}$$

$$P_{\tilde{Q}\,\boldsymbol{\imath}_2^*} = P_{\tilde{Q}} P_{\tilde{\boldsymbol{\imath}}_2|Q} \tag{117b}$$

$$P_{\tilde{Q}\,\mathbf{V}} = P_{\tilde{Q}} P_{\mathbf{V}|Q} \tag{117c}$$

$$P_{\tilde{Q}\,\mathbf{W}} = P_{\tilde{Q}} P_{\mathbf{W}|Q}. \tag{117d}$$

We realize that $\mathbf{W}$ is zero mean Gaussian distributed

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

The distribution (118) is the desired Gaussian distribution in our Berry-Esséen type bound. We want to upper bound the absolute difference

$$|\mathbb{P}[\tilde{\boldsymbol{\imath}}_2 \in \mathcal{D}] - \mathbb{P}[\mathbf{W} \in \mathcal{D}]| \tag{119a}$$

$$\leq |\mathbb{P}[\tilde{\boldsymbol{\imath}}_2 \in \mathcal{D}] - \mathbb{P}[\boldsymbol{\imath}_2^* \in \mathcal{D}]| \tag{119b}$$

$$+ |\mathbb{P}[\boldsymbol{\imath}_2^* \in \mathcal{D}] - \mathbb{P}[\mathbf{V} \in \mathcal{D}]| \tag{119c}$$

$$+ |\mathbb{P}[\mathbf{V} \in \mathcal{D}] - \mathbb{P}[\mathbf{W} \in \mathcal{D}]|, \tag{119d}$$

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_Q to $P_{\tilde{Q}}$, the change of the transition probability kernel from $P_{\tilde{\boldsymbol{\imath}}_2|Q}$ to $P_{\mathbf{V}|Q}$, and the change of the transition probability kernel from $P_{\mathbf{V}|Q}$ to $P_{\mathbf{W}|Q}$, respectively. We are going to bound (119a) by showing that the absolute differences in each of the terms in (119b)–(119d) is $O\left(\frac{1}{\sqrt{n}}\right)$. In the next three steps, we upper bound these absolute differences in the given order.

Step 3: We bound the absolute difference due to the change of input measure as follows:

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

$$= \left| \int_{-\infty}^{\infty} \mathbb{P}[\tilde{\boldsymbol{\imath}}_2 \in \mathcal{D} | Q = q] \left(P_Q(q) - P_{\tilde{Q}}(q)\right) dq \right| \tag{120}$$

$$\leq \int_{-\infty}^{\infty} \left| P_Q(q) - P_{\tilde{Q}}(q) \right| dq \tag{121}$$

$$= 2\,\mathrm{TV}(P_Q, P_{\tilde{Q}}) \tag{122}$$

$$\leq 2\frac{\sqrt{n}}{\sqrt{n-3}} - 2 \tag{123}$$

$$\leq \frac{C_{\mathrm{Q}}}{n}, \tag{124}$$

where $C_{\mathrm{Q}} = 8$. The inequality (121) follows by moving the absolute value to the inside of the integral and upper bounding the conditional probability by 1 for all q, (123) holds for any $n \geq 4$, which follows from Lemma 5. The inequality in (124) holds for $n \geq 4$. 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 $P_{\tilde{\boldsymbol{\imath}}_2|Q}$ to the Gaussian kernel $P_{\mathbf{V}|Q}$ as follows:

$$\begin{aligned}
&|\mathbb{P}[\boldsymbol{\imath}_2^* \in \mathcal{D}] - \mathbb{P}[\mathbf{V} \in \mathcal{D}]| \\
&= \left|\mathbb{E}\left[\mathbb{P}\left[\boldsymbol{\imath}_2^* \in \mathcal{D}\middle|\tilde{Q}\right] - \mathbb{P}\left[\mathbf{V} \in \mathcal{D}\middle|\tilde{Q}\right]\right]\right| && (125)\\
&\leq \mathbb{E}\left[\left|\mathbb{P}\left[\boldsymbol{\imath}_2^* \in \mathcal{D}\middle|\tilde{Q}\right] - \mathbb{P}\left[\mathbf{V} \in \mathcal{D}\middle|\tilde{Q}\right]\right| 1\left\{\left|\tilde{Q}\right| \leq \frac{\sqrt{n}}{2}\right\}\right] \\
&\quad + \mathbb{P}\left[\left|\tilde{Q}\right| > \frac{\sqrt{n}}{2}\right] && (126)\\
&\leq \max_{q \in \left[-\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right]} \frac{C(q)}{\sqrt{n}} + \mathbb{P}\left[\left|\tilde{Q}\right| > \frac{\sqrt{n}}{2}\right] && (127)\\
&\leq \frac{C_{\text{BE}}}{\sqrt{n}} + 2\exp\left\{-\frac{n}{8}\right\} && (128)\\
&\leq \frac{C_{\text{BE}} + C_{\text{Ch}}}{\sqrt{n}}, && (129)
\end{aligned}$$

where

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

$$C(q) = \frac{c\, 3^{1/4} T(q)}{\lambda_{\min}(\Sigma(q))^{3/2}}, \tag{131}$$

$$C_{\text{BE}} \triangleq \max_{q \in \left[-\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right]} C(q), \tag{132}$$

$$C_{\text{Ch}} \triangleq 4\exp\left\{-\frac{1}{2}\right\}, \tag{133}$$

and $\mathbf{J}_i(q)$ are as in (105), and c is the Berry-Esséen constant given in Theorem 6. Here, in (126), we first move the absolute value in (125) to the inside of the expectation, and then we separate the expectation into two parts according to whether $\left|\tilde{Q}\right| \leq \frac{\sqrt{n}}{2}$ holds to guarantee that we apply the Berry-Esséen theorem for values of q such that $\Sigma(q)$ is positive-definite. The inequality (127) follows from Corollary 2, and (128) follows from the Chernoff bound applied to a standard Gaussian random variable. The inequality (129) holds for any n. Since for every $q \in \left[-\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2}\right]$, $\Sigma(q)$ is a non-degenerate covariance matrix, and $T(q) < \infty$ is satisfied, we conclude that $C_{\text{BE}} < \infty$.

Step 5: In this step, we upper bound the probability in (119d), which is the absolute difference due to changing the covariance matrix of the Gaussian kernel from $\Sigma(q)$ to Σ, by using the upper bound on the total variation distance between two Gaussian vectors in Lemma 4. 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 $\lambda_i(\cdot)$ denotes the i-th largest eigenvalue of a matrix, and

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

We have

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

where (137) follows by moving the absolute value to the inside of the expectation in (136), and (139) follows from Lemma 4.

We observe that the matrices $\Sigma + \mathsf{B}$ and $\Sigma - \mathsf{B}$ are both positive semidefinite. Hence $\Sigma^{-1/2}(\Sigma + \mathsf{B})\Sigma^{-1/2}$ and $\Sigma^{-1/2}(\Sigma - \mathsf{B})\Sigma^{-1/2}$ are also positive semidefinite, and $\rho(\mathsf{A}) \leq 1$. Using the fact that $\|\mathsf{M}\|_F \leq \sqrt{d}\rho(\mathsf{M})$ for any $d \times d$ symmetric matrix M, and substituting the value of the expectation in (139), we conclude

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

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

By combining the bounds in (124), (129) and (140), we conclude the following Berry-Esséen-type inequality for the modified mutual information random vector:

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

G. Completion

We particularize the set $\mathcal{D} = \left\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} \geq \frac{1}{\sqrt{n}}\boldsymbol{\tau}\right\}$ in (141), where $\boldsymbol{\tau}$ is given in (94). Combining (96) and (141), we conclude that the probability $\mathbb{P}[\mathcal{A}^c]$ in (91) is upper bounded as

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

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

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

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

Here, the equality (143) follows since $\mathbf{W} \sim -\mathbf{W}$. Suppose that $\boldsymbol{\tau}$ satisfies

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

$$\gamma_n \triangleq \exp\left\{-c_2 n^{1/3}\right\} + \frac{G_1 + G_2 + G_{1,2} + C_{\text{Out}}}{\sqrt{n}}, \tag{146}$$

where the constants $c_2, G_1, G_2, G_{1,2}$ are as in (91). Then, the right-hand side of (91) is upper bounded by ϵ. From a Taylor series expansion of $Q_{\text{inv}}(\mathsf{V}, \cdot)$, we conclude that (145) is equivalent to the inequality in (24), which completes the proof.

VI. Proof of Theorem 3

In this section, we sketch the proof of Theorem 3, applicable to the K-transmitter MAC, by detailing the appropriate modifications in the proof of Theorem 2. Assume that $\mathcal{S} \in \mathcal{P}([K])$. Define the mutual information densities as

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

where $\mathcal{S}^c = [K]\backslash\mathcal{S}$, and the mutual information random vector for K transmitters as

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

where $\mathbf{X}_k$ is distributed uniformly over $\mathbb{S}^{n-1}(\sqrt{nP_k})$ for $k \in [K]$, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{I}_n)$, $\mathbf{X}_1, \dots, \mathbf{X}_K, \mathbf{Z}$ are independent, and $\mathbf{Y}_K = \mathbf{X}_{\langle [K] \rangle} + \mathbf{Z}$.

We will use the generalizations of our Lemma 1 given in Remark 2 and Lemma 6 given in term (3) below. The following lemma, which generalizes Lemma 5 to K transmitters, is the critical part of the proof of Theorem 3.

Lemma 7: Let $\mathbf{X}_i = (X_{i1}, \dots, X_{in})$, $i = 1, \dots, r$, be r independent random vectors, distributed uniformly over $\mathbb{S}^{n-1}(1)$. Let $Q_{ij} = \sqrt{n}\langle \mathbf{X}_i, \mathbf{X}_j \rangle$ for $1 \le i < j \le r$, and $\mathbf{Q} = (Q_{ij} \colon 1 \le i < j \le r)$. Then

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

for some constant C_r depending only on r.

Proof: See Appendix C. ■

The modifications in Section V are as follows:

1) The maximum likelihood decoder given in (53) is replaced by the decoder that chooses the message vector $m_{[K]} = (m_1, \dots, m_K)$ corresponding to the maximal mutual information density $\imath_{[K]}(\mathsf{f}_{[K]}(m_{[K]}); \mathbf{y})$.
2) The typical set $\mathcal{F}$ defined in (58) is replaced by

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

where $\mathcal{F}(\mathcal{S})$ is defined in (59). The inequality in (61) extends to $\mathcal{F}_K$ by Remark 2.
3) The functions given in (64)–(66) are extended as

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

In the proof of Lemma 6, we replace $P_{\langle [2] \rangle}$ by $P_{\langle \mathcal{S} \rangle}$, and $P_1 P_2$ by $\sum_{\substack{i,j \in [K] \\ i<j}} P_i P_j$. The inequality in (76) generalizes to the K-transmitter MAC by inspecting its proof in Appendix A, and applying Remark 2 in Section V-A. Hence, Lemma 6 generalizes as

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

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 (84) and (85) are replaced by

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

$$\mathcal{A}_K = \left\{ \boldsymbol{\imath}_K \ge \left(\log\left(\prod_{s \in \mathcal{S}} M_s G(\mathcal{S}) \right) \colon \mathcal{S} \in \mathcal{P}([K]) \right) - \frac{1}{2} \log n \mathbf{1} \right\}, \tag{154}$$

respectively. By using the extension of the RCU bound for K transmitters given in Remark 1, and following the same steps as Section V-E, we conclude that the right-hand side of the inequality in (91) is replaced by

$$\mathbb{P}\left[\mathcal{A}_K^c\right] + \exp\left\{-c_K n^{1/3}\right\} + \frac{\sum_{\mathcal{S} \in \mathcal{P}([K])} G(\mathcal{S})}{\sqrt{n}}, \tag{155}$$

where c_K is a constant.
5) We here explain the differences between bounding $\mathbb{P}\left[\mathcal{A}_K^c\right]$ and $\mathbb{P}\left[\mathcal{A}^c\right]$. We naturally extend the definition of the modified and centered mutual information random vector to K transmitters by introducing

$$\tilde{\imath}_{\mathcal{S}}(\mathbf{x}_{\mathcal{S}}; \mathbf{y}_K|\mathbf{x}_{\mathcal{S}^c}) \triangleq \sum_{i=1}^n \log \frac{P_{Y_K|X_{[K]}}(y_i|x_{[K]i})}{P_{\tilde{Y}_K|\tilde{X}_{\mathcal{S}^c}}(y_i|x_{\mathcal{S}^c i})}, \tag{156}$$

$$\tilde{\boldsymbol{\imath}}_K \triangleq \frac{1}{\sqrt{n}}\big[(\tilde{\imath}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}; \mathbf{Y}_K|\mathbf{X}_{\mathcal{S}^c}) \colon \mathcal{S} \in \mathcal{P}([K])) - n\mathbf{C}(P_{[K]})\big], \tag{157}$$

where $\mathbf{C}(P_{[K]})$ is the capacity vector defined in (26), $\tilde{X}_k \sim \mathcal{N}(0, P_k)$ for $k \in [K]$, and $\prod_{k=1}^K P_{\tilde{X}_k} \to P_{Y_K|X_{[K]}} \to P_{\tilde{Y}_K} = \mathcal{N}(0, 1 + P_{[K]})$.
The threshold value in (94) is replaced by

$$\boldsymbol{\tau} \triangleq \log\left(\frac{\left(\prod_{s \in \mathcal{S}} M_s\right) G(\mathcal{S}) \kappa(\mathcal{S})}{\sqrt{n}} \colon \mathcal{S} \in \mathcal{P}([K]) \right) - n\mathbf{C}(P_{[K]}). \tag{158}$$

By using the joint distribution of $(\mathbf{X}_{[K]}, \mathbf{Y}_K)$, we get

$$\tilde{\boldsymbol{\imath}}_K \sim \frac{1}{\sqrt{n}}\left(\frac{(n - \|\mathbf{Z}\|^2) P_{\langle \mathcal{S} \rangle}}{2(1 + P_{\langle \mathcal{S} \rangle})} + \frac{\sum_{\substack{i,j \in \mathcal{S} \\ i<j}} \langle \mathbf{X}_i, \mathbf{X}_j \rangle + \langle \mathbf{Z}, \mathbf{X}_{\langle \mathcal{S} \rangle} \rangle}{1 + P_{\langle \mathcal{S} \rangle}} \colon \mathcal{S} \in \mathcal{P}([K]) \right). \tag{159}$$

Define the random vector that consists of the inner products of all different pairs in $\mathbf{X}_{[K]}$ as

$$\mathbf{Q} \triangleq (Q_{ij} \colon i, j \in [K], i < j) \in \mathbb{R}^{\binom{K}{2}}, \tag{160}$$

where $Q_{ij} = \frac{\langle \mathbf{X}_i, \mathbf{X}_j \rangle}{\sqrt{nP_iP_j}}$ denotes the normalized inner product of $\mathbf{X}_i$ and $\mathbf{X}_j$. The inner product random vector $\mathbf{Q}$ replaces Q in (106). Observe that for all different (i_1, j_1) and (i_2, j_2) pairs, $Q_{i_1 j_1}$ and $Q_{i_2 j_2}$ are independent of each other, which follows by independence of $\mathbf{X}_1, \dots, \mathbf{X}_K$. However, $\mathbf{Q}$ does not have a product distribution due to the fact that any triplets in $\mathbf{Q}$ are not jointly independent[2]. Despite not being a product

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

distribution, by Lemma 7, the joint distribution $P_{\mathbf{Q}}$ converges to the distribution of $\binom{K}{2}$ i.i.d. standard Gaussian random variables in total variation, allowing us to use the Berry-Esséen theorem in the same manner as for the two-transmitter MAC.

As for the two-transmitter MAC, the distribution in (159) depends on $\mathbf{X}_{[K]}$ only through the inner product random vector $\mathbf{Q}$. We obtain that

$$\tilde{\boldsymbol{\imath}}_K|\mathbf{Q} = \mathbf{q} \sim \boldsymbol{\imath}_K|\mathbf{X}^n_{[K]} = \mathbf{x}_{[K]} \tag{161}$$

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

where

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

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

for $i \in [n]$, and $\mathbf{x}_{[K]}$ are any vectors on the n-dimensional spheres with the corresponding radii, satisfying $\frac{\langle\mathbf{x}_i,\mathbf{x}_j\rangle}{\sqrt{nP_iP_j}} = q_{ij}$ for all $i < j \in [K]$. The conditional covariance matrix given in (109) is extended to K transmitters as

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

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 $Q_{i_1j_1}$, $Q_{i_2j_2}$ for all different (i_1,j_1) and (i_2,j_2) pairs, using the law of total expectation and variance, we compute

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

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

where the covariance matrix $\mathsf{V}(P_{[K]})$ is defined in (27). The rest of the proof follows identically to Section V-F, where we replace Q by $\mathbf{Q}$, $\tilde{Q}$ by the $\binom{K}{2}$-dimensional standard Gaussian random vector $\tilde{\mathbf{Q}}$, $\tilde{\boldsymbol{\imath}}_2|Q = q$ by $\tilde{\boldsymbol{\imath}}_K|\mathbf{Q} = \mathbf{q}$, $\mathbf{V}|Q = q$ by $\mathbf{V}|\mathbf{Q} = \mathbf{q}$, and $\mathbf{W}|Q = q$ by $\mathbf{W}|\mathbf{Q} = \mathbf{q}$. For the probability transition kernels $P_{\mathbf{V}|\mathbf{Q}}$ and $P_{\mathbf{W}|\mathbf{Q}}$, we replace $\boldsymbol{\mu}(q)$ by $\boldsymbol{\mu}(\mathbf{q})$, Σ by Σ_K, and $\Sigma(q)$ by $\Sigma(\mathbf{q})$. In the proof, whenever a condition in the form of $|q| < t$ is used, it is replaced by $|\mathbf{q}| < t\mathbf{1}$.

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

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

for some constant $C_K > 0$, where $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathsf{V}(P_{[K]}))$. By combining (155) 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: 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, the largest possible decoding time for an epoch. By the time $n_k < n_K$, each active transmitter will have transmitted a sub-codeword of length n_k. If decoding happens at time n_k, the rest of the codeword is not used. We generate M length-n_K i.i.d. codewords according to some distribution $P_{\mathbf{X}}$. In other words, the encoding function has the distribution

$$\mathsf{f}(U, W_m) \sim \text{i.i.d. } P_{\mathbf{X}} \text{ for } m\in[M]. \tag{171}$$

Here, U is the common randomness that initializes the encoders and the decoder.

Decoding: Unlike the MAC, for the Gaussian RAC, we require the decoder to have an option to not decode any messages at a decoding time n_k, since the true channel is known by neither the transmitters nor the receiver. Therefore, we couple the maximum likelihood decoder given in (53) with a threshold rule, where the maximum likelihood decoder is applied only if that threshold rule is met. Here, the role of the threshold rule is to reliably determine the true channel in the communication epoch. The motivation behind the choice of the decoding rule, below, is that for any $P > 0$, under an input distribution $P_{\mathbf{X}}$ such that the power constraint (31) is met with equality on average, i.e. $\frac{1}{n_k}\mathbb{E}\left[\left\|\mathbf{X}^{[n_k]}\right\|^2\right] = P$ for all k, the normalized squared norms of the outputs $\mathbf{Y}_k^{[n_k]}$ concentrate around their mean for all k, and the expectations of the normalized squared norms of the outputs $\mathbf{Y}_k^{[n_k]}$ are distinct for all $k \in \{0, 1, \ldots, K\}$:

$$\frac{1}{n_k}\mathbb{E}\left[\left\|\mathbf{Y}_k^{[n_k]}\right\|^2\right] = 1 + kP. \tag{172}$$

Upon receiving the first n_0 symbols of the output, $\mathbf{y}^{[n_0]}$, the decoder computes the following function to decide whether there are any active transmitters

$$\mathsf{g}_0(U, \mathbf{y}^{[n_0]}) = \begin{cases} 0 & \text{if } \left|\frac{1}{n_0}\left\|\mathbf{y}^{[n_0]}\right\|^2 - 1\right| \le \lambda_0 \\ \mathsf{e} & \text{otherwise,}\end{cases} \tag{173}$$

where λ_0 is a parameter that is determined by the error criterion ϵ_0.

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

$$\mathsf{g}_k(U, \mathbf{y}^{[n_k]}) = \begin{cases} m_{[k]} & \text{if } \imath_{[k]}(\mathsf{f}(U, m_{[k]})^{[n_k]}; \mathbf{y}^{[n_k]}) \\ & \quad > \imath_{[k]}(\mathsf{f}(U, m'_{[k]})^{[n_k]}; \mathbf{y}^{[n_k]}) \\ & \quad \text{for any } m'_{[k]} \overset{\pi}{\neq} m_{[k]}, \\ & m_1 \leq \cdots \leq m_k, \\ & \left| \frac{1}{n_k} \left\| \mathbf{y}^{[n_k]} \right\|^2 - (1 + kP) \right| \leq \lambda_k, \\ \mathsf{e} & \text{otherwise}, \end{cases} \tag{174}$$

where λ_k is a parameter to satisfy the error criterion ϵ_k. Transmission stops, and a positive ACK bit is transmitted to all transmitters once a non-erasure is decoded in (174). By the permutation-invariance of the channel in terms of the inputs $\mathbf{X}_{[k]}$, and the identical encoding in (171), all permutations of the messages $m_{[k]}$ give the same mutual information density. Therefore, without loss of generality, our decoder always decodes the ordered message vector in (174). 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 to decode messages when the number of active transmitters is k at time n_k with high probability, instead of at any of the earlier decoding times $n_0, \ldots, n_{k-1}$.

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:

$$\epsilon_0 \leq \mathbb{P}\left[\left| \frac{1}{n_0} \left\| \mathbf{Y}_0^{[n_0]} \right\|^2 - 1 \right| > \lambda_0 \right]. \tag{175}$$

$k \geq 1$ active transmitters: When there are some active transmitters, according to the encoding function (171) and the decoding rule (174), an error occurs 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 (31) 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 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 will bound the probability of these events separately, and apply the union bound to upper bound the probability of error.

Since we are using a single codebook at all encoders, separating the event $\mathcal{E}_{\text{rep}}$ that at least one message among transmitted messages is repeated

$$\mathcal{E}_{\text{rep}} = \{W_i = W_j \text{ for some } i \neq j\}, \tag{176}$$

is advantageous. Given $\mathcal{E}_{\text{rep}}^c$, we can leverage the independence of the codewords for each transmitter. By the union bound, we get

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

We bound the probability of error by applying the union bound as

$$\begin{aligned} \epsilon_k &= \frac{1}{M^k} \sum_{m_{[k]} \in [M]^k} \mathbb{P}\Bigg[\bigcup_{t=0}^{k-1} \left\{ \mathsf{g}_t(U, \mathbf{Y}_k^{[n_t]}) \neq \mathsf{e} \right\} \\ &\qquad \bigcup \left\{ \mathsf{g}_k(U, \mathbf{Y}_k^{[n_k]}) \overset{\pi}{\neq} m_{[k]} \right\} \Bigg| W_{[k]} = m_{[k]} \Bigg] \quad (178) \\ &\leq \mathbb{P}\left[\mathcal{E}_{\text{rep}}\right] + \mathbb{P}\left[\mathcal{E}_{\text{rep}}^c\right] \left(\mathbb{P}\left[\mathcal{E}_{\text{codeword}} \middle| \mathcal{E}_{\text{rep}}^c\right] \right. \quad (179) \\ &\qquad \left. + \mathbb{P}\left[\mathcal{E}_{\text{time}} \middle| \mathcal{E}_{\text{rep}}^c\right] + \mathbb{P}\left[\mathcal{E}_{\text{message}} \middle| \mathcal{E}_{\text{rep}}^c\right] \right) \quad (180) \\ &\leq \mathbb{P}\left[\mathcal{E}_{\text{rep}}\right] + \mathbb{P}\left[\mathcal{E}_{\text{codeword}} \middle| \mathcal{E}_{\text{rep}}^c\right] \\ &\qquad + \mathbb{P}\left[\mathcal{E}_{\text{time}} \middle| \mathcal{E}_{\text{rep}}^c\right] + \mathbb{P}\left[\mathcal{E}_{\text{message}} \middle| \mathcal{E}_{\text{rep}}^c\right]. \quad (181) \end{aligned}$$

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

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

Wrong decoding time: According to the decoding rule in (174), decoding occurs at time n_k if and only if the output typicality criterion is not satisfied for any $t < k$, 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

$$\begin{aligned} \mathbb{P}\left[\mathcal{E}_{\text{time}} \middle| \mathcal{E}_{\text{rep}}^c\right] = \mathbb{P}\Bigg[&\bigcup_{t=0}^{k-1} \left\{ \left| \frac{1}{n_t} \left\| \mathbf{Y}_k^{[n_t]} \right\|^2 - (1 + tP) \right| \leq \lambda_t \right\} \\ &\bigcup \left\{ \left| \frac{1}{n_k} \left\| \mathbf{Y}_k^{[n_k]} \right\|^2 - (1 + kP) \right| > \lambda_k \right\} \Bigg]. \end{aligned} \tag{183}$$

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

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

Combining (175), (177) and (181)–(184) completes the proof.

VIII. Proof of Theorem 4

In this section, we will analyze the achievability bound in Theorem 5 by particulatizing the input distribution, $P_{\mathbf{X}}$ in Theorem 5, choosing the free parameters λ_k, and bounding the probability and expectation terms in (37c).

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. Define the set

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

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

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

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, $\mathbf{X}^{\mathcal{N}_j} \sim \text{Uniform}\left(\mathbb{S}^{|\mathcal{N}(j)|-1}(\sqrt{|\mathcal{N}(j)|P})\right)$, and $\mathbf{X}^{\mathcal{N}_j}$ are independent from each other for $j \in [K]$.

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 $|\mathcal{N}(j)|$. Each of these sub-codewords is distributed uniformly on the sphere with the corresponding dimension and radius according to (187). Note that the codewords chosen according to (171) satisfy the power constraints in (31) 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 in 3 steps: deriving an output typicality bound, evaluation of the RCU bound, and evaluation of a Berry-Esséen 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 (174). Since $\mathbf{Y}_k^{\mathcal{N}(j)}$ are independent for $j \in [K]$ due to the input distribution in (186), for $k \ge 1$, we have by Remark 2 and Lemma 2 that

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

for $\lambda_k \in (0, 1+kP)$, where $\kappa(j,P) \triangleq \kappa_j(P\mathbf{1})$ is the constant defined in Remark 2. For $k=0$, we have

$$\mathbb{P}\left[\left|\left\|\mathbf{Y}_0^{[n_0]}\right\|^2 - n_0\right| > n_0\lambda_0\right] \le 2\kappa(1,P)\exp\left\{-\frac{n_0\lambda_0^2}{8}\right\} \tag{190}$$

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

$$\lambda_0 = \sqrt{\frac{-8\log\frac{\epsilon_0}{2\kappa(1,P)}}{n_0}}, \tag{191}$$

to satisfy that the right-hand side of (190) is upper bounded by ϵ_0. By setting $\lambda_t = \frac{P}{2}$ for $t \ge 1$, using (189) and (190), and applying the union bound, we get that the decoding time error in (37b) is upper bounded by

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

Step 2: To bound the expectation in (37c), we first modify the definition of the typical output set $\mathcal{F}(\mathcal{S})$ in (59) as follows:

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

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

Then we show that Lemma 6 holds under the 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 have

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

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

In order to show (196), from the analysis in Section V-D, we see that we only need to verify the steps (77)–(79) for the modified input distribution in (186). Hence, we need to show that

$$\begin{aligned}&\mathbb{P}\left[\langle\mathbf{X}_{\langle[s]\rangle}^{[n_k]}, \mathbf{X}_{\langle[s]\rangle}^{[n_k]} + \mathbf{Z}^{[n_k]}\rangle - \sum_{j=1}^{k}\frac{|\mathcal{N}(j)|u_j}{2} \in [a, a+\mu]\,\middle|\,\mathcal{E}\right] \\ &\le O\left(\frac{1}{\sqrt{n_k}}\right),\end{aligned} \tag{197}$$

where

$$\mathcal{E} = \left\{\left\|\mathbf{X}_{\langle[s]\rangle}^{\mathcal{N}(j)} + \mathbf{Z}^{\mathcal{N}(j)}\right\|^2 = |\mathcal{N}(j)|s_j,\right.$$

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

and $s_j \in \mathcal{I}(j, [s])$, and $u_j > 0$. The proof of (197) follows similarly to the one in [4, Appendix A] for the parallel Gaussian channels, since we can consider our K independent sub-codewords with lengths $|\mathcal{N}(j)|$, $j \in [K]$, as K parallel channels with each having blocklength $|\mathcal{N}(j)|$, $j \in [K]$.

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

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

where a' is related to a by a constant shift, and (199) follows by setting $\mathbf{X}_{\langle [s] \rangle}^{\mathcal{N}(j)} = (\sqrt{|\mathcal{N}(j)| u_j}, 0, \ldots, 0)$, 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

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

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 mutual information random vector $\boldsymbol{\imath}_k$ and the typical events analogous to (83)–(85) as

$$\boldsymbol{\imath}_k = (\imath_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^{[n_k]}; \mathbf{Y}_k^{[n_k]} | \mathbf{X}_{\mathcal{S}^c}^{[n_k]}) \colon \mathcal{S} \in \mathcal{P}([k])), \tag{205}$$
$$\mathcal{E}(\mathcal{S})_{\text{RAC}} = \left\{ \mathbf{X}_{\langle \mathcal{S} \rangle}^{[n_k]} + \mathbf{Z}^{[n_k]} \in \mathcal{F}(\mathcal{S})_{\text{RAC}} \right\}, \tag{206}$$
$$\mathcal{E}_{\text{RAC}} = \bigcap_{\mathcal{S} \in \mathcal{P}([k])} \mathcal{E}(\mathcal{S})_{\text{RAC}}, \tag{207}$$
$$\mathcal{A}_k = \Big\{ \boldsymbol{\imath}_k \geq \Big(|\mathcal{S}| \log M + \log G'_{|\mathcal{S}|,k} \colon \mathcal{S} \in \mathcal{P}([k]) \Big) - \frac{1}{2} \log n_k \mathbf{1} \Big\}. \tag{208}$$

By Lemma 2 and the union bound, we have

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

where c_k is a positive constant. Combining (196) and (209), and following the analysis in Section V-E, we obtain that the expectation in (37c) is bounded by

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

Step 3: We set

$$k \log M = n_k C(kP) - \sqrt{n_k (V(kP) + V_{\text{cr}}(k,P))} Q^{-1}\left(\epsilon_k - \frac{D_k}{\sqrt{n_k}} \right) + \frac{1}{2} \log n_k - \log G'_{k,k} - \sum_{j \in [k]} \log \kappa(j, P), \tag{211}$$

for all $k \in [K]$, where D_k is a positive constant to be chosen later. Since $C(sP) > \frac{s}{k} C(kP)$ for $s < k$ and (211), we conclude the following:

1) There exists a constant $c_0 > 0$ such that $\min_{j \in [k]} |\mathcal{N}(j)| \geq c_0 n_k$. In other words, all $|\mathcal{N}(j)|$ are in the same order as n_k.
2) $\frac{k(k-1)}{2M}$ 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 and Chebyshev's inequality,

$$\mathbb{P}\left[\mathcal{A}_k^c\right] \leq \frac{E_k}{n_k} + \mathbb{P}\left[\imath_{[k]}(\mathbf{X}_{[k]}^{[n_k]}; \mathbf{Y}_k^{[n_k]}) < k \log M + \log G'_{k,k} - \frac{1}{2} \log n_k \right] \tag{212}$$

for some positive constant E_k.

Therefore, it only remains to evaluate the probability term in (212). Define the modified and centered mutual information random variable

$$\tilde{\imath}_k \triangleq \frac{1}{\sqrt{n_k}} \left(\sum_{i=1}^{n_k} \log \frac{P_{Y_k | X_{[k]}}(Y_i | X_{[k],i})}{P_{\tilde{Y}_k}(Y_i)} - n_k C(kP) \right), \tag{213}$$

where $\tilde{Y}_k \sim \mathcal{N}(0, 1+kP)$. By Remark 2 and (211), we get

$$\mathbb{P}\left[\imath_{[k]}(\mathbf{X}_{[k]}^{[n_k]}; \mathbf{Y}_k^{[n_k]}) < k \log M + \log G'_{k,k} - \frac{1}{2} \log n_k \right] \leq \mathbb{P}\left[\tilde{\imath}_k < -\sqrt{V(kP) + V_{\text{cr}}(k,P)} Q^{-1}\left(\epsilon_k - \frac{D_k}{\sqrt{n_k}} \right) \right]. \tag{214}$$

The conditional random variable satisfies $\tilde{\imath}_k | \mathbf{X}_{[k]}^{[n_k]} = \mathbf{x}_{[k]}^{[n_k]} \sim \imath_k | \mathbf{Q} = \mathbf{q}$, where

$$\mathbf{Q} = (Q_{ij} \colon i, j \in [k], i < j) \in \mathbb{R}^{\binom{k}{2}}, \tag{215}$$

and $Q_{ij} = \frac{\langle \mathbf{X}_i^{[n_k]}, \mathbf{X}_j^{[n_k]} \rangle}{\sqrt{n_k P^2}}$. To upper bound the right-hand side of (214), comparing with the arguments in Section VI, we only need to verify that

$$\text{TV}(P_{\mathbf{Q}}, P_{\tilde{\mathbf{Q}}}) \leq \frac{H_k}{\sqrt{n_k}} \tag{216}$$

for some constant H_k, where $\tilde{\mathbf{Q}} \sim \mathcal{N}\left(\mathbf{0}, \mathsf{I}_{\binom{k}{2}}\right)$. To show (216), we define

$$\mathbf{Q}^{(t)} \triangleq (Q_{ij}^{(t)} : i, j \in [k], i < j) \in \mathbb{R}^{\binom{k}{2}}, \tag{217}$$

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

$$\mathbf{Q} = \sum_{t=1}^{k} \frac{\sqrt{|\mathcal{N}(t)|}}{\sqrt{n_k}} \mathbf{Q}^{(t)}. \tag{218}$$

By the data processing inequality of the total variation distance and that $\mathbf{Q}^{(t)}$ are independent for $t \in [k]$, we get

$$\mathrm{TV}(P_{\mathbf{Q}}, P_{\tilde{\mathbf{Q}}}) \leq \sum_{t=1}^{k} \mathrm{TV}(P_{\mathbf{Q}^{(t)}}, P_{\tilde{\mathbf{Q}}}) \tag{219}$$

$$\leq \sum_{t=1}^{k} \frac{F_k}{\sqrt{|\mathcal{N}(j)|}} \tag{220}$$

$$\leq \frac{kF_k}{\sqrt{c_0 n_k}}, \tag{221}$$

where (220) follows from Lemma 7, F_k are constants in Lemma 7, and (221) follows from (211), which proves (216).

By (221), and following arguments similar to those in Section VI, we conclude that

$$\mathbb{P}\left[\tilde{\imath}_k < -\sqrt{V(kP) + V_{\mathrm{cr}}(k, P)} Q^{-1}\left(\epsilon_k - \frac{D_k}{\sqrt{n_k}}\right)\right]$$
$$\leq \epsilon_k - \frac{D_k}{\sqrt{n_k}} + \frac{C_k}{\sqrt{n_k}}, \tag{222}$$

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

$$\frac{D_k}{\sqrt{n_k}} \leq \frac{k(k-1)}{2M} + B + \frac{C_k}{\sqrt{n_k}} + \frac{E_k}{n_k}$$
$$+ k \exp\left\{-c_k (c_0 n_k)^{1/3}\right\} + \sum_{s=1}^{k} \frac{\binom{k}{s} G'_{s,k}}{\sqrt{n_k}}, \tag{223}$$

where B is in (192). For a large enough n_k, such a constant exists by the enumerated consequences of (211) above. From Theorem 5 and the inequalities (188), (210)–(212), (214), (222) and (223), we conclude that the probability of error is bounded by ϵ_k. By a Taylor series expansion of the function $Q^{-1}(\cdot)$, we complete the proof.

IX. Concluding Remarks

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

For the Gaussian MAC problem, we prove a third-order achievability result (Theorem 2) building on the RCU bound (Theorem 1) that is derived for general MACs. Our random encoding function uses an input distribution that is distributed uniformly on the n-dimensional sphere. At the receiver, we employ a maximum likelihood decoder. Compared to the result of MolavianJazi and Laneman [7], 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 [20] 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 spherically distributed and independent of each other. In our proposed coding strategy, the decoding occurs at finitely many time instants $n_0, \ldots, n_K$, depending on the estimate of the number of active transmitters. The receiver broadcasts a single-bit acknowledgment to all transmitters at each decoding time indicating whether a successful decoding occurs. 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 first three order terms as the best known code for the Gaussian MAC in operation (Theorem 2 and Theorem 3). Thus there is no penalty due to the unknown transmitter activity. This result also implies that for the Gaussian MAC, using length-n concatenation of sub-codewords that lies on a much smaller set than the n-dimensional sphere used in Theorem 2 nevertheless achieves the same first three order terms.

Appendix A
Proof of (76)

Let $u < P_{\langle[2]\rangle}$ be a constant. Define the interval

$$\mathcal{I} = [n(1 + P_{\langle[2]\rangle} - \epsilon), n(1 + P_{\langle[2]\rangle} + \epsilon)], \tag{224}$$

where $\epsilon = n^{-1/3}$. We would like to show that for a large enough n,

$$g(y) \triangleq \mathbb{P}\left[\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2 \leq nu \,\middle|\, \left\|\mathbf{X}_{\langle[2]\rangle} + \mathbf{Z}\right\|^2 = y\right] \tag{225}$$

$$\leq \exp\left\{-nC\right\}, \tag{226}$$

for all $y \in \mathcal{I}$, where C is a positive constant. Recall that the support of $\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2$ is

$$\mathcal{S} = [n(\sqrt{P_1} - \sqrt{P_2})^2, n(\sqrt{P_1} + \sqrt{P_2})^2]. \tag{227}$$

Hence, (226) is trivially satisfied for $u < (\sqrt{P_1} - \sqrt{P_2})^2$. To show (226) for $(\sqrt{P_1} - \sqrt{P_2})^2 \leq u < P_{\langle[2]\rangle}$, we will show two concentration results: first,

$$g(y) = g(n(1 + P_{\langle[2]\rangle})) \exp\{O(n\epsilon)\}, \tag{228}$$

for all $y \in \mathcal{I}$, and second, for a large enough n

$$p \triangleq \mathbb{P}\left[\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2 \leq nu \,\middle|\, \mathcal{A}\right] \tag{229}$$

$$\leq \exp\{-nC'\}, \tag{230}$$

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

$$\mathcal{A} \triangleq \left\{\left\|\mathbf{X}_{\langle[2]\rangle} + \mathbf{Z}\right\|^2 \in \mathcal{I}\right\}. \tag{231}$$

Using (228) and (230), we can show (226) as follows. By conditioning the probability in (229) on each value of $\left\|\mathbf{X}_{\langle[2]\rangle}+\mathbf{Z}\right\|^2$, we express p as

$$p = \int_{\mathcal{I}} g(y) f_{\|\mathbf{X}_{\langle[2]\rangle}+\mathbf{Z}\|^2|\mathcal{A}}(y)dy \tag{232}$$
$$= g(n(1+P_{\langle[2]\rangle}))\exp\{O(n\epsilon)\} \tag{233}$$
$$\leq \exp\{-nC'\} \tag{234}$$

where (233) follows from (228) and $\min_{y\in\mathcal{I}} g(y) \leq \int_{\mathcal{I}} g(y) f_{\|\mathbf{X}_{\langle[2]\rangle}+\mathbf{Z}\|^2|\mathcal{A}}(y)dy \leq \max_{y\in\mathcal{I}} g(y)$, and (234) follows from (230). The inequalities (228) and (234) imply that there exists a constant $C > 0$ such that (226) holds for all $y \in \mathcal{I}$, for a large enough n, since $O(n\epsilon) = o(n)$.

We proceed to show (230). By Bayes' rule, we have

$$p = \frac{\mathbb{P}\left[\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2 \leq nu\right]\mathbb{P}\left[\mathcal{A}\,\middle|\,\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2 \leq nu\right]}{\mathbb{P}\left[\mathcal{A}\right]}. \tag{235}$$

For $n \geq n_0$ in Lemma 1, upper-bounding the second factor in the numerator of (235) by 1, changing measure from $P_{\mathbf{X}_{\langle[2]\rangle}} P_{\mathbf{Z}}$ to $P_{\tilde{\mathbf{U}}} P_{\mathbf{Z}}$, where $\tilde{\mathbf{U}} \sim \mathcal{N}(0, (P_{\langle[2]\rangle})\mathsf{I}_n)$, and then applying Lemma 1, we get for a large enough n that

$$p \leq \frac{\kappa_2(P_1,P_2)\mathbb{P}\left[\left\|\tilde{\mathbf{U}}\right\|^2 \leq nu\right]\cdot 1}{1-\kappa_2(P_1,P_2)\mathbb{P}\left[\left|\left\|\tilde{\mathbf{U}}+\mathbf{Z}\right\|^2 - n(1+P_{\langle[2]\rangle})\right| > n\epsilon\right]} \tag{236}$$
$$\leq \kappa_2(P_1,P_2)\frac{\exp\left\{\frac{-n(P_{\langle[2]\rangle}-u)^2}{4(P_{\langle[2]\rangle})^2}\right\}}{1-2\kappa_2(P_1,P_2)\exp\{\frac{-n\epsilon^2}{8(1+P_{\langle[2]\rangle})^2}\}} \tag{237}$$
$$\leq 2\kappa_2(P_1,P_2)\exp\left\{\frac{-n(P_{\langle[2]\rangle}-u)^2}{4(P_{\langle[2]\rangle})^2}\right\} \tag{238}$$
$$\leq \exp\{-nC'\}, \tag{239}$$

where $\kappa_2(P_1,P_2)$ is the constant defined in (40), and C' is a positive constant. The inequality (237) follows from the tail bounds on the chi-squared distribution in Lemma 2, and (238) follows since the denominator in the right-hand side of (237) is greater than equal to $\frac{1}{2}$ for a large enough n. The inequality (239) holds since $u < P_{\langle[2]\rangle}$.

We proceed to prove (228). Define the events $\mathcal{B} = \{\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2 \leq nu\}$ and $\mathcal{B}(\lambda) = \{\left\|\mathbf{X}_{\langle[2]\rangle}\right\|^2 = \lambda\}$ for any $\lambda \in \mathcal{S}$. By Bayes' rule, we can express $g(y)$ as

$$g(y) = \frac{\mathbb{P}\left[\mathcal{B}\right] f_{\|\mathbf{X}_{\langle[2]\rangle}+\mathbf{Z}\|^2|\mathcal{B}}(y)}{f_{\|\mathbf{X}_{\langle[2]\rangle}+\mathbf{Z}\|^2}(y)}. \tag{240}$$

By spherical symmetry of the distribution of $\mathbf{X}_{\langle[2]\rangle}$, the conditional distribution $\left\|\mathbf{X}_{\langle[2]\rangle}+\mathbf{Z}\right\|^2|\mathcal{B}(\lambda)$ does not depend on $\mathbf{u}$ when we fix $\mathbf{X}_{\langle[2]\rangle}$ to any $\mathbf{u}$ such that $\|\mathbf{u}\|^2 = \lambda \in \mathcal{S}$. Therefore,

$$\left\|\mathbf{X}_{\langle[2]\rangle}+\mathbf{Z}\right\|^2|\mathcal{B}(\lambda) \sim \sum_{i=1}^{n}\left\|Z_i + \frac{\sqrt{\lambda}}{\sqrt{n}}\right\|^2, \tag{241}$$

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

$$f(x;n,\lambda) = \frac{1}{2}\exp\left\{-\frac{(x+\lambda)}{2}\right\}\left(\frac{x}{\lambda}\right)^{\frac{n}{4}-\frac{1}{2}} I_{\frac{n}{2}-1}(\sqrt{\lambda x}), \tag{242}$$

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

$$\frac{f(x_1;n,\lambda)}{f(x_2;n,\lambda)} = \exp\{x_2 - x_1\}\left(\frac{x_1}{x_2}\right)^{\frac{n}{4}-\frac{1}{2}}\frac{I_{\frac{n}{2}-1}(\sqrt{\lambda x_1})}{I_{\frac{n}{2}-1}(\sqrt{\lambda x_2})} \tag{243}$$

Paris [33] proves the following bounds, which hold for $0 < x < y$ and $\nu > -1/2$

$$\exp\{x-y\}\left(\frac{x}{y}\right)^{\nu} < \frac{I_\nu(x)}{I_\nu(y)} < \left(\frac{x}{y}\right)^{\nu}. \tag{244}$$

Using (244), we can lower and upper bound (243) as

$$\exp\{n\delta\}\left(1-\frac{\delta}{b+\delta}\right)^{\frac{n}{2}-1}\exp\left\{-\sqrt{n\lambda}\left(\sqrt{b+\delta}-\sqrt{b}\right)\right\}$$
$$\leq \frac{f(x_1;n,\lambda)}{f(x_2;n,\lambda)} \tag{245}$$
$$\leq \exp\{n\delta\}\left(1-\frac{\delta}{b+\delta}\right)^{\frac{n}{2}-1}. \tag{246}$$

Taylor series expansion at $\delta = 0$ gives

$$\log\left(1-\frac{\delta}{b+\delta}\right) = -\frac{\delta}{b} + O(\delta^2), \tag{247}$$
$$-\sqrt{n\lambda}\left(\sqrt{b+\delta}-\sqrt{b}\right) = -\sqrt{n\lambda}\left(\frac{\delta}{2\sqrt{b}} + O(\delta^2)\right). \tag{248}$$

Substituting (247) and (248) in (245) and (246), we get

$$\frac{f(x_1;n,\lambda)}{f(x_2;n,\lambda)} = \exp\{O(n\delta)\}. \tag{249}$$

We can also verify the validity of (249) for $\lambda = 0$ by using the probability density function of chi-squared distribution with n degrees of freedom instead of (242). Particularizing (249) to $b = 1 + P_{\langle[2]\rangle}$, we get for all $\lambda \in \mathcal{S}$ that

$$f_{\|\mathbf{X}_{\langle[2]\rangle}+\mathbf{Z}\|^2|\mathcal{B}(\lambda)}(y) = f_{\|\mathbf{X}_{\langle[2]\rangle}+\mathbf{Z}\|^2|\mathcal{B}(\lambda)}(n(1+P_{\langle[2]\rangle}))\exp\{O(n\epsilon)\}, \tag{250}$$

which together with (240) implies (228).

Appendix B
Proof of Lemma 4

Proof: Pinsker's inequality (e.g., [34, Th. 6.5]) states that for any distributions P and Q,

$$\mathrm{TV}(P,Q) \leq \sqrt{\frac{1}{2}D(P\|Q)}. \tag{251}$$

Let $\mathrm{tr}(\cdot)$ denote trace of a matrix. Relative entropy between two d-dimensional Gaussian distributions with positive covariance matrices (e.g., [34, eq. (1.18)]) is given by

$$D(\mathcal{N}(\boldsymbol{\mu}_1,\Sigma_1)\|\mathcal{N}(\boldsymbol{\mu}_2,\Sigma_2))$$

$$= \frac{1}{2}\Big(\mathrm{tr}(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathsf{I}_d) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\Sigma_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \log\det(\Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2})\Big). \tag{252}$$

Define

$$\mathsf{G} \triangleq \Sigma_1^{-1/2}\Sigma_2\Sigma_1^{-1/2} - \mathsf{I}_d, \tag{253}$$
$$a \triangleq \frac{1}{2}\sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\Sigma_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}. \tag{254}$$

Combining (251) and (252), we get

$$\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}_1,\Sigma_1),\mathcal{N}(\boldsymbol{\mu}_2,\Sigma_2)) \le a + \frac{1}{2}\sqrt{\mathrm{tr}(\mathsf{G}) - \log\det(\mathsf{I}_d + \mathsf{G})}. \tag{255}$$

To lower bound the logdeterminant term in (255), we use the following result in [31, 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

$$\exp\left\{\mathrm{tr}(\mathsf{G}) - \frac{\|\mathsf{G}\|_F^2}{2(1-\rho(\mathsf{G}))}\right\} \le \det(\mathsf{I}_d + \mathsf{G}). \tag{256}$$

For $\rho(\mathsf{G}) < 1$, we apply (256) to (255), and get

$$\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}_1,\Sigma_1),\mathcal{N}(\boldsymbol{\mu}_2,\Sigma_2)) \le \frac{1}{2\sqrt{2}}\frac{\|\mathsf{G}\|_F}{\sqrt{1-\rho(\mathsf{G})}} + a. \tag{257}$$

In addition, trivially we have that

$$\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}_1,\Sigma_1),\mathcal{N}(\boldsymbol{\mu}_2,\Sigma_2)) \le 1 \tag{258}$$
$$\le \frac{\|\mathsf{G}\|_F}{\rho(\mathsf{G})} + a. \tag{259}$$

Combining (257) and (259), we conclude that for $\rho(\mathsf{G}) < 1$,

$$\mathrm{TV}(\mathcal{N}(\boldsymbol{\mu}_1,\Sigma_1),\mathcal{N}(\boldsymbol{\mu}_2,\Sigma_2)) \le \min\left\{\frac{1}{2\sqrt{2}}\frac{1}{\sqrt{1-\rho(\mathsf{G})}}, \frac{1}{\rho(\mathsf{G})}\right\}\|\mathsf{G}\|_F + a \tag{260}$$
$$= \frac{2+\sqrt{6}}{4}\|\mathsf{G}\|_F + a. \tag{261}$$

Since the coefficient $\frac{2+\sqrt{6}}{4} > 1 \ge \frac{1}{\rho(\mathsf{G})}$ for $\rho(\mathsf{G}) \ge 1$, we conclude that (261) holds for any $\rho(\mathsf{G})$. ∎

Appendix C
Proof of Lemma 7

To prove this lemma, we are going to use the induction technique in [32, Th. 4], which shows 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 $\mathbf{Q}$ is invariant to rotations, we fix

$$\mathbf{X}_1 = (1, 0, 0, \ldots, 0). \tag{262}$$

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

$$\mathbf{Q}_1 = (Q_{1j} : 2 \le j \le r) \tag{263}$$
$$\mathbf{Q}_2 = (Q_{ij} : 2 \le i < j \le r), \tag{264}$$

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

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

$$Q_{ij} = \sqrt{n}X_{i1}X_{j1} + \frac{\sqrt{n}}{\sqrt{n-1}}(1 - X_{i1}^2)^{\frac{1}{2}}(1 - X_{j1}^2)^{\frac{1}{2}}V_{ij} \tag{265}$$

$$V_{ij} = \sqrt{n-1}\langle \mathbf{Y}_i, \mathbf{Y}_j\rangle, \tag{266}$$

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

Since $\mathbf{Y}_i$'s are independent and distributed uniformly over $\mathbb{S}^{n-2}(1)$, given $\mathbf{Q}_1$, the joint distribution of $\mathbf{V} = (V_{ij} : 2 \le i < j \le r)$ is $p_{r-1}^{(n-1)}$. By (265), we can write the joint distribution of $\mathbf{Q}_2|\mathbf{Q}_1 = \mathbf{q}_1$ as

$$Q_{ij}|\mathbf{Q}_1 = \mathbf{q}_1 \sim \frac{q_{1i}q_{1j}}{\sqrt{n}} + \frac{\sqrt{n}}{\sqrt{n-1}}\left(1 - \frac{q_{1i}^2}{n}\right)^{\frac{1}{2}}\left(1 - \frac{q_{1j}^2}{n}\right)^{\frac{1}{2}}V_{ij} \tag{267}$$

for $2 \le i < j \le r$. Also define the probability transition kernel $P_{\mathbf{Q}_2^*|\mathbf{Q}_1}$ as

$$Q_{ij}^*|\mathbf{Q}_1 = \mathbf{q}_1 \sim \frac{q_{1i}q_{1j}}{\sqrt{n}} + \frac{\sqrt{n}}{\sqrt{n-1}}\left(1 - \frac{q_{1i}^2}{n}\right)^{\frac{1}{2}}\left(1 - \frac{q_{1j}^2}{n}\right)^{\frac{1}{2}}Z_{ij} \tag{268}$$

for $2 \le i < j \le r$, where $Z_{ij} \sim \mathcal{N}(0,1)$ are i.i.d. Now, we are ready to apply the mathematical induction.

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

$$\mathrm{TV}(p_2^{(n)}, \mathcal{N}(0,1)) \le \frac{4}{n} \tag{269}$$

by Lemma 5 with $k = 1$.

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

$$\mathrm{TV}\left(p_{r-1}^{(n)}, \mathcal{N}\left(\mathbf{0}, \mathsf{I}_{\frac{1}{2}(r-1)(r-2)}\right)\right) \le \frac{C_{r-1}}{\sqrt{n}} \tag{270}$$

for some constant C_{r-1}. Let $P_{\tilde{\mathbf{Q}}_1} = \mathcal{N}(\mathbf{0}, \mathsf{I}_{r-1})$ and $P_{\tilde{\mathbf{Q}}_2} = \mathcal{N}\left(\mathbf{0}, \mathsf{I}_{\binom{r-1}{2}}\right)$. By the triangle inequality of the total variation distance, we write

$$\mathrm{TV}\left(p_r^{(n)}, \mathcal{N}\left(\mathbf{0}, \mathsf{I}_{\binom{r}{2}}\right)\right)$$
$$= \mathrm{TV}\left(P_{\mathbf{Q}_1}P_{\mathbf{Q}_2|\mathbf{Q}_1}, P_{\tilde{\mathbf{Q}}_1}P_{\tilde{\mathbf{Q}}_2}\right) \tag{271}$$
$$\le \mathrm{TV}\left(P_{\mathbf{Q}_1}P_{\mathbf{Q}_2|\mathbf{Q}_1}, P_{\tilde{\mathbf{Q}}_1}P_{\mathbf{Q}_2|\mathbf{Q}_1}\right) \tag{272}$$
$$+ \mathrm{TV}\left(P_{\tilde{\mathbf{Q}}_1}P_{\mathbf{Q}_2|\mathbf{Q}_1}, P_{\tilde{\mathbf{Q}}_1}P_{\mathbf{Q}_2^*|\mathbf{Q}_1}\right) \tag{273}$$
$$+ \mathrm{TV}\left(P_{\tilde{\mathbf{Q}}_1}P_{\mathbf{Q}_2^*|\mathbf{Q}_1}, P_{\tilde{\mathbf{Q}}_1}P_{\tilde{\mathbf{Q}}_2}\right). \tag{274}$$

Here, (272) approximates the input measure $P_{\mathbf{Q}_1}$ with the corresponding i.i.d. Gaussian measure $P_{\tilde{\mathbf{Q}}_1}$, (273) approximates the inner product random variables V_{ij} in the definition of the probability transition kernel given in (267) with i.i.d. standard Gaussian random variables, and (274) approximates the mean in (268) by 0 and the variance by 1. We will upper bound the right-hand sides of (272)–(274) in that order. We have

$$\begin{aligned}&\mathrm{TV}\left(P_{\mathbf{Q}_1}P_{\mathbf{Q}_2|\mathbf{Q}_1}, P_{\tilde{\mathbf{Q}}_1}P_{\mathbf{Q}_2|\mathbf{Q}_1}\right)\\&=\mathrm{TV}\left(P_{\mathbf{Q}_1}, P_{\tilde{\mathbf{Q}}_1}\right) &(275)\\&\leq (r-1)\mathrm{TV}\left(P_{Q_{12}}, \mathcal{N}(0,1)\right) &(276)\\&\leq \frac{4(r-1)}{n}, &(277)\end{aligned}$$

where (276) follows since $P_{\mathbf{Q}_1} = (P_{Q_{12}})^{r-1}$ is a product distribution, and (277) follows from Lemma 5. The total variation distance in (273) is bounded as

$$\begin{aligned}&\mathrm{TV}\left(P_{\tilde{\mathbf{Q}}_1}P_{\mathbf{Q}_2|\mathbf{Q}_1}, P_{\tilde{\mathbf{Q}}_1}P_{\mathbf{Q}_2^*|\mathbf{Q}_1}\right)\\&=\mathbb{E}\left[\mathrm{TV}\left(P_{\mathbf{Q}_2|\mathbf{Q}_1=\tilde{\mathbf{Q}}_1}, P_{\mathbf{Q}_2^*|\mathbf{Q}_1=\tilde{\mathbf{Q}}_1}\right)\middle|\tilde{\mathbf{Q}}_1\right] &(278)\\&=\mathrm{TV}\left(p_{r-1}^{(n-1)}, \mathcal{N}\left(\mathbf{0}, \mathsf{I}_{\binom{r-1}{2}}\right)\right) &(279)\\&\leq \frac{C_{r-1}}{\sqrt{n-1}}, &(280)\end{aligned}$$

where (279) follows from the definitions (267) and (268) since the total variation distance is shift and scale invariant, and (280) follows from the inductive assumption (270). The total variation distance in (274) is bounded as

$$\begin{aligned}&\mathrm{TV}\left(P_{\tilde{\mathbf{Q}}_1}P_{\mathbf{Q}_2^*|\mathbf{Q}_1}, P_{\tilde{\mathbf{Q}}_1}P_{\tilde{\mathbf{Q}}_2}\right)\\&=\mathbb{E}\left[\mathrm{TV}\left(P_{\mathbf{Q}_2^*|\mathbf{Q}_1=\tilde{\mathbf{Q}}_1}, P_{\tilde{\mathbf{Q}}_2}\right)\middle|\tilde{\mathbf{Q}}_1\right] &(281)\\&\leq \mathbb{E}\left[\sum_{2\leq i<j\leq r}\mathrm{TV}\left(P_{Q_{ij}^*|\mathbf{Q}_1=\tilde{\mathbf{Q}}_1}, \mathcal{N}(0,1)\right)\middle|\tilde{\mathbf{Q}}_1\right] &(282)\\&=\binom{r-1}{2}\mathbb{E}\Bigg[\mathrm{TV}\Bigg(\mathcal{N}\Bigg(\frac{\tilde{Q}_{12}\tilde{Q}_{13}}{\sqrt{n}}, \frac{n}{n-1}\left(1-\frac{\tilde{Q}_{12}^2}{n}\right)\\&\qquad\left(1-\frac{\tilde{Q}_{13}^2}{n}\right)\Bigg), \mathcal{N}(0,1)\Bigg)\Bigg] &(283)\\&\leq \binom{r-1}{2}\Bigg\{\frac{1}{2}\frac{\mathbb{E}\left[\left|\tilde{Q}_{12}\right|\right]^2}{\sqrt{n}}\\&\quad+\frac{2+\sqrt{6}}{4}\left|\frac{n}{n-1}\left(\mathbb{E}\left[1-\frac{\tilde{Q}_{12}^2}{n}\right]\right)^2-1\right|\Bigg\} &(284)\\&=\binom{r-1}{2}\left(\frac{1}{\pi\sqrt{n}}+\frac{2+\sqrt{6}}{4n}\right), &(285)\end{aligned}$$

where (282) follows since $P_{\mathbf{Q}_2^*|\mathbf{Q}_1=\mathbf{q}_1}$ is a product distribution, and $P_{\tilde{\mathbf{Q}}_2}$ is i.i.d. standard Gaussian, and (283) follows since $Q_{ij}^*|\mathbf{Q}_1 = \mathbf{q}_1$ is identically distributed for $2 \leq i < j \leq r$. The inequality (284) follows from Lemma 4 with $d = 1$ using the i.i.d. distribution of $\tilde{Q}_{12}$ and $\tilde{Q}_{13}$. Combining (277), (280), (285) and the inequality in (272) 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 $r > 2$, while it is faster $\left(O\left(\frac{1}{n}\right)\right)$ for $r = 2$.

Appendix D
Proof of Corollary 1

In order to prove Corollary 1, we will show that for any M that satisfies the inequality (28), it holds that

$$(|\mathcal{S}|\log M\colon \mathcal{S}) \in \mathcal{P}([K]) \in n\mathbf{C}(P\mathbf{1}) - \sqrt{n}Q_{\mathrm{inv}}(\mathsf{V}(P\mathbf{1}),\epsilon)\ \frac{1}{2}\log n\mathbf{1} + O(1)\mathbf{1}. \quad (286)$$

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 (28) holds, implying that

$$\mathbb{P}\Big[Z([K]) > \sqrt{n}\left(C(KP) - \frac{K\log M}{n}\right) + \frac{1}{2}\frac{\log n}{\sqrt{n}} + O\left(\frac{1}{\sqrt{n}}\right)\Big] \leq \epsilon. \quad (287)$$

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

$$\mathbb{P}\Big[Z(\mathcal{S}) > \sqrt{n}\left(C(|\mathcal{S}|P) - \frac{|\mathcal{S}|\log M}{n}\right) + \frac{1}{2}\frac{\log n}{\sqrt{n}} + O\left(\frac{1}{\sqrt{n}}\right)\Big] \leq O\left(\frac{1}{n}\right), \quad (288)$$

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

By the union bound, (287) and (288), we get

$$\mathbb{P}\Bigg[\bigcup_{\mathcal{S}\in\mathcal{P}([K])}\Big\{Z(\mathcal{S}) > \sqrt{n}\left(C(|\mathcal{S}|P) - \frac{|\mathcal{S}|\log M}{n}\right) + \frac{1}{2}\frac{\log n}{\sqrt{n}} + O\left(\frac{1}{\sqrt{n}}\right)\Big\}\Bigg] \leq \epsilon + O\left(\frac{1}{n}\right), \quad (289)$$

which is by the definition (18) equivalent to

$$(|\mathcal{S}|\log M\colon \mathcal{S} \in \mathcal{P}([K])) \in n\mathbf{C}(P\mathbf{1}) - \sqrt{n}Q_{\mathrm{inv}}\left(\mathsf{V}(P\mathbf{1}), \epsilon + O\left(\frac{1}{n}\right)\right) + \frac{1}{2}\log n\mathbf{1} + O(1)\mathbf{1}. \quad (290)$$

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

Appendix E
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 have 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 are making binary decisions at all decoding times $n_0, \ldots, n_K$ to whether decode or not decode messages. An alternative to this strategy might 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 that the squared norms of the output, $\frac{1}{n_0}\left\|\mathbf{Y}_k^{[n_0]}\right\|^2$, lie in for each $k \leq K$, the decoder estimates the number of active transmitters as t, and via feedback all parties agree that the communication epoch will end at time n_t. This strategy requires a feedback of $\lceil \log(K+1) \rceil$ bits from the receiver to transmitters at time n_0; while our strategy in the proof of Theorem 4 requires a varying length of feedback 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 of wrong decision time under this strategy can be modified as

$$\mathbb{P}\left[\mathcal{E}_{\text{time}} \middle| \mathcal{E}_{\text{rep}}^c\right] \leq 2\left(\prod_{j=1}^{k} \kappa(j,P)\right) \exp\left\{-\frac{n_0(\frac{P}{2})^2}{8(1+kP)^2}\right\}, \tag{291}$$

which decays exponentially with n_0 like (192) does, but with a smaller exponential rate than (192). Hence, this modification in the strategy only increases the constant c in (35), and affects the achievable $O(1)$ term in (34).

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. 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 $\sqrt{n_1 P}$, 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 (221) can be improved to

$$\mathrm{TV}(P_{\mathbf{Q}}, P_{\tilde{\mathbf{Q}}}) \leq \frac{F_k}{\sqrt{n_1}} + \frac{F_k}{\sqrt{n_k - n_1}}, \tag{292}$$

which also decays with the same asymptotic rate as (221). Therefore, this modification only affects the $O(1)$ term in (34), meaning that the same expansion as Theorem 4 is achieved.

B. Decoding Transmitter Identity

Another possibility in our design is to allow the decoder to decode the transmitter identities of the messages transmitted. By employing distinct encoders at each transmitter with the same input distribution $P_{\mathbf{X}}$, we can show that the coefficient $\binom{k}{s}\binom{M-k}{s}$ in (184) is replaced by $\binom{k}{s}\binom{K-(k-s)}{s}M^s$ due to the increase in the number of unions in the error event. Here, we choose $k-s$ correctly decoded messages from k active transmitters, and s wrongly decoded messages from the remaining $K-(k-s)$ many transmitters, and there are M messages for each of the wrongly decoded transmitter identities. Since K does not grow with n, decoding transmitter identities only affects the $O(1)$ term in (34). Such a modification in the encoding functions allows to decode the identities of the active transmitters in addition to the list of messages sent by active transmitters. This result holds for more general RACs as discussed in [25, Sec. V].

References

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

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

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

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

[5] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

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

[7] 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 Transactions on Information Theory*, vol. 61, no. 12, pp. 6719–6733, Dec. 2015.

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

[9] 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.

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

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

[12] O. Ordentlich and Y. Polyanskiy, "Low complexity schemes for the random access Gaussian channel," in *Proceedings 2017 IEEE International Symposium on Information Theory*, Aachen, Germany, June 2017, pp. 2528–2532.

[13] M. Ebrahimi, F. Lahouti, and V. Kostina, "Coded random access design for constrained outage," in *Proceedings 2017 IEEE International Symposium on Information Theory*, Aachen, Germany, June 2017, pp. 2732–2736.

[14] X. Chen and D. Guo, "Many-access channels: the Gaussian case with random user activities," in *2014 IEEE International Symposium on Information Theory*, Honolulu, HI, USA, June 2014, pp. 3127–3131.

[15] X. Chen, T. Y. Chen, and D. Guo, "Capacity of Gaussian many-access channels," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3516–3539, June 2017.

[16] Y. Polyanskiy, "A perspective on massive random-access," in *Proceedings 2017 IEEE International Symposium on Information Theory*, Aachen, Germany, June 2017, pp. 2523–2527.

[17] I. Zadik, Y. Polyanskiy, and C. Thrampoulidis, "Improved bounds on Gaussian MAC and sparse regression via Gaussian inequalities," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 430–434.

[18] S. S. Kowshik and Y. Polyanskiy, "Fundamental limits of many-user MAC with finite payloads and fading," *CoRR*, vol. abs/1901.06732, May 2019.

[19] S. S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy efficient random access for the quasi-static fading mac," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 2768–2772.

[20] M. Effros, V. Kostina, and R. C. Yavas, "Random access channel coding in the finite blocklength regime," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 1261–1265.

[21] Y. Liu and M. Effros, "Finite-blocklength and error exponent analyses for LDPC codes in point-to-point and multiple access communication," *arXiv Preprints*, Jan. 2020.

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

[23] S. Chen, M. Effros, and V. Kostina, "Lossless source coding in the point-to-point, multiple access, and random access scenarios," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 1692–1696.

[24] S. L. Fong and V. Y. F. Tan, "A proof of the strong converse theorem for Gaussian multiple access channels," *IEEE Transactions on Information Theory*, vol. 62, no. 8, pp. 4376–4394, Aug. 2016.

[25] M. Effros, V. Kostina, and R. C. Yavas, "Random access channel coding in the finite blocklength regime," *CoRR*, vol. abs/1801.09018, Jan. 2018, submitted to IEEE Transactions on Information Theory on July 23, 2019.

[26] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," *Ann. Statist.*, vol. 28, no. 5, pp. 1302–1338, 10 2000.

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

[28] M. Raič, "A multivariate Berry–Esséen theorem with explicit constants," *arXiv preprint arXiv:1802.06475*, 2018.

[29] V. Y. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 881–903, Feb. 2014.

[30] L. Devroye, A. Mehrabian, and T. Reddad, "The total variation distance between high-dimensional Gaussians," *arXiv preprint arXiv:1810.08693*, 2018.

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

[32] 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, 1982.

[33] 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.

[34] 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.