

Massive Uncoordinated Access With Random User Activity

Khac-Hoang Ngo, Alejandro Lancho, Giuseppe Durisi, and Alexandre Graell i Amat
Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden

Abstract—We extend the seminal work by Polyanskiy (2017) on massive uncoordinated access to the case where the number of active users is random and unknown *a priori*. We define a random-access code accounting for both misdetection (MD) and false-alarm (FA), and derive a random-coding achievability bound for the Gaussian multiple access channel. Our bound captures the fundamental trade-off between MD and FA probabilities. It suggests that lack of knowledge of the number of active users entails a small penalty in power efficiency. For a typical scenario, to achieve both MD and FA probabilities below 0.1, the required energy per bit predicted by our bound is 0.5–0.7 dB higher than that predicted by the bound in Polyanskiy (2017) for a known number of active users. Taking both MD and FA into account, we use our bound to benchmark the energy efficiency of some recently proposed massive random access schemes.

I. INTRODUCTION

Under the paradigm of the Internet of Things (IoT), the number of connecting devices is increasing dramatically. IoT devices are mostly battery limited and transmit short packets in a sporadic and uncoordinated manner [1], [2]. This calls for new theoretical frameworks that help to understand the fundamental limits of massive random access and provide guidelines for system design. Polyanskiy [3] proposed a novel formulation for the massive uncoordinated access problem with three key assumptions: i) all users employ a common codebook and the decoder only aims to return a list of messages without recovering users' identities; ii) the error event is defined per user and the error probability is averaged over the users; iii) each user sends a fixed amount of information bits within a finite frame length. Under this formulation, traditional as well as novel random access protocols [4] yield achievability bounds. In [3], an achievability bound for the Gaussian multiple access channel (MAC) was derived and it was shown that modern random access schemes exhibit a large gap to this bound. This gap has been later reduced in, e.g., [5]–[10]. Polyanskiy's framework has been extended to the quasi-static fading channel [11], multiple-antenna channel [12], and a setup with common alarm messages [13].

In Polyanskiy's achievability bound, the number of active users is fixed and known to the receiver, an assumption that has practical shortcomings. Since IoT devices access the channel at random times and in a grant-free manner, the number of active users varies over time, and hence, it is typically unknown to the receiver. Therefore, the bound in [3] may be an overoptimistic benchmark for random-access schemes that are designed to work with unknown number of active users. Moreover, when the number of active users

is unknown, the decoder needs to determine the list size. Choosing a list size smaller than the number of active users will result in misdetections (MDs)—i.e., transmitted messages that are not included in the decoded list—whereas choosing it larger than the number of active users will result in false alarms (FAs)—i.e., decoded messages that were not transmitted. Furthermore, additional MDs and FAs may occur in the decoding process. There is a trade-off between MD and FA probabilities. A decoder that always outputs the whole codebook will never misdetect, but has FA probability close to one; similarly, a decoder that always outputs an empty set will never raise a FA but always misdetects. Characterizing the MD–FA trade-off is a fundamental engineering challenge that was not addressed in [3]. An achievability bound for the Gaussian MAC with unknown number of active users was presented in [14]. However, the authors consider the joint-user error event instead of the per-user error event, and thus, MD and FA are not explicitly considered. In short, a random-coding bound accounting for both MD and FA, which can serve as a benchmark for common-codebook massive uncoordinated random access with random user activity, is still missing.

Most of the practical algorithms that have been proposed for common-codebook massive random access require knowledge of the number of active users. Advanced ALOHA schemes, such as irregular repetition slotted ALOHA (IRSA) [15], can also operate when the number of active users is unknown. However, research on modern random access protocols [4], such as IRSA, has traditionally focused on characterizing and minimizing the packet loss rate, which accounts only for MD. The scheme proposed in [6] also addressed MD only. Minimizing the MD probability alone can entail a high FA probability. In [16], a tensor-based communication scheme was proposed, and both MD and FA probabilities are reported in the performance evaluation. Another scheme for which both MD and FA probabilities are reported was recently proposed in [17] for the quasi-static fading MAC and for the case in which the receiver has a large number of antennas.

In this work, we extend Polyanskiy's bound to the case where the number of active users is *random* and *unknown*. To this end, we first extend the definition of a random-access code provided in [3] to account for both MD and FA probabilities. Then, we derive a random-coding bound for the Gaussian MAC. Unlike the scheme in [3], our decoder does not assume knowledge of the number of active users, and thus cannot use it to set the decoded list size. Instead, we let our decoder decide the best list size within a predetermined interval around an

estimated value of the number of active users. Our decoding metric is similar to that used in [13]. However, different from [13], we limit the decoded list size to be in an interval to avoid overfitting.

Compared with the bound in [3], our bound suggests that the lack of knowledge of the number of active users entails a small penalty in power efficiency. Furthermore, we apply our bound to characterize MD and FA in slotted ALOHA with multi-packet reception (SA-MPR). Using our bound, we benchmark the energy efficiency of SA-MPR and of the massive random access schemes proposed in [7], [8]. For instance, for a system with 300 active users in average, to achieve both MD and FA probabilities below 10^{-1} , the required energy per bit predicted by our achievability bound is 0.65 dB higher than that predicted by the bound for a known number of active users [3]. In the same setting, the required energy per bit predicted by our bound is 9 dB, 4.1 dB, and 3.6 dB lower than that of the SA-MPR bound, the scheme based on sparse regression code (SPARC) [7], and an enhancement of SPARC [8], respectively.

Notation: Random quantities are denoted with non-italic letters with sans-serif font, e.g., a scalar x and a vector \mathbf{v} . Deterministic quantities are denoted with italic letters, e.g., a scalar x and a vector \mathbf{v} . The Euclidean norm is denoted by $\|\cdot\|$. We use $\mathfrak{P}(\mathcal{A})$ to denote the set of all subsets of \mathcal{A} ; $[n]$ denotes the set of integers $\{1, \dots, n\}$ if $n \geq 1$ and $[n] = \emptyset$ if $n = 0$; $[m : n] = \{m, m+1, \dots, n\}$ if $m \leq n$ and $[m : n] = \emptyset$ if $m > n$; $x^+ = \max\{x, 0\}$; $\mathbb{1}\{\cdot\}$ is the indicator function. The set of natural and complex numbers are denoted by \mathbb{N} and \mathbb{C} , respectively. We denote the Gamma function by $\Gamma(x) = \int_0^\infty z^{x-1} e^{-z} dz$, and the lower and upper incomplete Gamma functions by $\gamma(x, y) = \int_0^y z^{x-1} e^{-z} dz$ and $\Gamma(x, y) = \int_y^\infty z^{x-1} e^{-z} dz$, respectively.

II. RANDOM-ACCESS CHANNEL

We consider a MAC in which a random set of K_a users transmit their messages to a receiver over n uses of a stationary memoryless channel. Let $\mathbf{x}_k \in \mathcal{X}$ be the transmitted signal of user k in a channel use. Given $K_a = K_a$, the channel law is given by $P_{\mathbf{y}|\mathbf{x}_1, \dots, \mathbf{x}_{K_a}}$. Thus this random-access channel is characterized by the probability mass function (PMF) P_{K_a} of K_a and by the set of conditional probabilities $\{P_{\mathbf{y}|\mathbf{x}_1, \dots, \mathbf{x}_{K_a}} : \mathcal{X}^{K_a} \rightarrow \mathcal{Y}\}_{K_a \in \mathbb{N}}$. As in [3], we assume that the channel law is permutation invariant. We further assume that the receiver does not know the realizations of K_a .

As in [3], our model differs from the classical MAC in that the total number of users is not limited, all users employ the same codebook, and the receiver decodes up to a permutation of messages. However, as opposed to [3], where the number of active users is assumed to be fixed and known, we assume that K_a is random and unknown. We therefore need to account for both MD and FA. We next rigorously define the MD and FA probabilities, as well as the notion of a random-access code.

Definition 1 (Random-access code): Consider a random-access channel characterized by $\{P_{K_a}, \{P_{\mathbf{y}|\mathbf{x}_1, \dots, \mathbf{x}_{K_a}}\}_{K_a \in \mathbb{N}}\}$. An $(M, n, \epsilon_{\text{MD}}, \epsilon_{\text{FA}})$ random-access code for this channel,

where M and n are positive integers and $\epsilon_{\text{MD}}, \epsilon_{\text{FA}} \in (0, 1)$, consists of:

- A random variable \mathbf{U} defined on a set \mathcal{U} that is revealed to both the transmitter and the receiver before the start of the transmission. This random variable acts as common randomness and allows for the use of randomized coding strategies.
- An encoder mapping $f: \mathcal{U} \times [M] \rightarrow \mathcal{X}^n$ defining the transmitted codeword $\mathbf{x}_i = f(\mathbf{U}, w_i)$ of user i for a given message w_i , which is assumed to be uniformly distributed over $[M]$.
- A decoding function $g: \mathcal{U} \times \mathcal{Y}^n \rightarrow \mathfrak{P}([M])$ providing an estimate $\widehat{\mathcal{W}} = \{\hat{w}_1, \dots, \hat{w}_{|\widehat{\mathcal{W}}|}\} = g(\mathbf{U}, \mathbf{y})$ of the list of transmitted messages, where $\mathbf{y} = [y(1) \dots y(n)]^\top$ denotes the channel output sequence.

Let $\widetilde{\mathcal{W}} = \{\tilde{w}_1, \dots, \tilde{w}_{|\widetilde{\mathcal{W}}|}\}$ denotes the set of distinct elements of $\{\mathbf{w}_1, \dots, \mathbf{w}_{K_a}\}$. We assume that the decoding function satisfies the following constraints on the MD and FA probabilities:

$$P_{\text{MD}} = \mathbb{E} \left[\mathbb{1}\{|\widetilde{\mathcal{W}}| \neq 0\} \cdot \frac{1}{|\widetilde{\mathcal{W}}|} \sum_{i=1}^{|\widetilde{\mathcal{W}}|} \mathbb{P}[\tilde{w}_i \notin \widehat{\mathcal{W}}] \right] \leq \epsilon_{\text{MD}}, \quad (1)$$

$$P_{\text{FA}} = \mathbb{E} \left[\mathbb{1}\{|\widehat{\mathcal{W}}| \neq 0\} \cdot \frac{1}{|\widehat{\mathcal{W}}|} \sum_{i=1}^{|\widehat{\mathcal{W}}|} \mathbb{P}[\hat{w}_i \notin \mathcal{W}] \right] \leq \epsilon_{\text{FA}}, \quad (2)$$

The expectations in (1) and (2) are with respect to the size of \mathcal{W} and $\widehat{\mathcal{W}}$, respectively.

In the random-access code defined in [3], the decoder outputs a list of messages of size equal to the number of active users, which is assumed to be known. In such a setup, a MD implies a FA, and vice versa. Hence, the two types of errors become indistinguishable. In our setup, the decoded list size can be different from the number of transmitted messages. Hence, we introduce explicitly the MD and FA probabilities. This allows us to characterize the MD–FA trade-off.

Hereafter, we consider the Gaussian MAC with $\{P_{\mathbf{y}|\mathbf{x}_1, \dots, \mathbf{x}_{K_a}}\}$ imposed by $\mathbf{y} = \sum_{i=1}^{K_a} \mathbf{x}_i + \mathbf{z}$, where $\{\mathbf{x}_i\}_{i=1}^{K_a}$ are the transmitted signals over n channel uses and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_n)$ is the Gaussian noise independent of $\{\mathbf{x}_i\}_{i=1}^{K_a}$. We consider the power constraint $\|\mathbf{x}_i\|^2 \leq nP, \forall i \in [K_a]$.

III. RANDOM-CODING BOUND

The random-coding bound in [3, Th. 1] is derived by constructing a random-coding scheme as follows. Let $\mathcal{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_{K_a}\} \subset [M]$ be the set of transmitted messages. Each active user picks randomly a codeword \mathbf{c}_{w_i} from a common codebook containing M codewords $\mathbf{c}_1, \dots, \mathbf{c}_M$ drawn independently from the distribution $\mathcal{CN}(\mathbf{0}, P' \mathbf{I}_n)$ for a fixed $P' < P$. To convey message w_i , user i transmits \mathbf{c}_{w_i} provided that $\|\mathbf{c}_{w_i}\|^2 \leq nP$. Otherwise, it transmits the all-zero codeword. The receiver employs a minimum distance decoder where the decoded list is $\widehat{\mathcal{W}} = \arg \min_{\widehat{\mathcal{W}} \subset [M], |\widehat{\mathcal{W}}|=K_a} \|c(\widehat{\mathcal{W}}) - \mathbf{y}\|^2$, with $c(\mathcal{W}) = \sum_{i \in \mathcal{W}} \mathbf{c}_i$. The error analysis involves manipulations of unions of the pairwise error events via a change of measure and the application

of the Chernoff bound combined with Gallager's ρ -trick [18, p. 136]. An alternative bound is also obtained by writing the pairwise error event as an inequality involving information densities, and by applying a property of the information density given in [19, Cor. 17.1].

In the following, we derive a similar random-coding bound for the case in which K_a is random and unknown to the receiver. Specifically, we consider a random-coding scheme with the same encoder as in [3]. However, since the receiver does not know K_a , we let the decoder estimate K_a from \mathbf{y} , then decide the best list size within an interval around the initially detected value of K_a . Specifically, given the channel output \mathbf{y} , the receiver estimates K_a as

$$K'_a = \arg \min_{K \in [K_l, K_u]} m(\mathbf{y}, K), \quad (3)$$

where $m(\mathbf{y}, K)$ is a suitably chosen metric, and K_l and K_u are suitably chosen lower and upper limits on K'_a , respectively. Then, given K'_a , the receiver decodes the list of messages as

$$\widehat{\mathcal{W}} = \arg \min_{\widehat{\mathcal{W}} \subset [M], |\widehat{\mathcal{W}}| \leq \overline{K'_a}} \|c(\widehat{\mathcal{W}}) - \mathbf{y}\|^2, \quad (4)$$

where $\underline{K'_a} = \max\{K_l, K'_a - r\}$ and $\overline{K'_a} = \min\{K_u, K'_a + r\}$ with a chosen nonnegative integer r . We refer to r as the *decoding radius*. An error analysis of this random-coding scheme conducted along similar lines as in [3] leads to the following result.

Theorem 1 (Random-coding bound, K_a random and unknown): Fix $P' < P$, r , K_l , and K_u ($K_l \leq K_u$). For the K_a -user Gaussian MAC with $K_a \sim P_{K_a}$, there exists an $(M, n, \epsilon_{\text{MD}}, \epsilon_{\text{FA}})$ random-access code satisfying the power constraint P and

$$\epsilon_{\text{MD}} = \sum_{K_a = \max\{K_l, 1\}}^{K_u} \left(P_{K_a}(K_a) \sum_{K'_a = K_l}^{K_u} \sum_{t \in \mathcal{T}} \frac{t + (K_a - \overline{K'_a})^+}{K_a} \cdot \min\{p_t, q_t, \xi(K_a, K'_a)\} \right) + p_0, \quad (5)$$

$$\epsilon_{\text{FA}} = \sum_{K_a = K_l}^{K_u} \left(P_{K_a}(K_a) \sum_{K'_a = K_l}^{K_u} \sum_{t \in \mathcal{T}} \sum_{t' \in \mathcal{T}_t} \frac{t' + (\underline{K'_a} - K_a)^+}{K_a - t - (K_a - \overline{K'_a})^+ + t' + (\underline{K'_a} - K_a)^+} \cdot \min\{p_{t,t'}, q_{t,t'}, \xi(K_a, K'_a)\} \right) + p_0, \quad (6)$$

where

$$p_0 = 2 - \sum_{K_a = K_l}^{K_u} P_{K_a}(K_a) - \mathbb{E}_{K_a} \left[\frac{M!}{M^{K_a} (M - K_a)!} \right] + \mathbb{E}_{K_a} \left[\frac{\Gamma(n, nP/P')}{\Gamma(n)} \right], \quad (7)$$

$$p_t = \sum_{t' \in \mathcal{T}_t} p_{t,t'}, \quad (8)$$

$$p_{t,t'} = e^{-nE(t,t')}, \quad (9)$$

$$E(t, t') = \max_{\rho, \rho_1 \in [0, 1]} -\rho \rho_1 t' R_1 - \rho_1 R_2 + E_0(\rho, \rho_1), \quad (10)$$

$$E_0(\rho, \rho_1) = \max_{\lambda} \rho_1 a + \ln(1 - \rho_1 P_2 b), \quad (11)$$

$$a = \rho \ln(1 + P' t' \lambda) + \ln(1 + P' t \mu), \quad (12)$$

$$b = \rho \lambda - \frac{\mu}{1 + P' t \mu}, \quad (13)$$

$$\mu = \frac{\rho \lambda}{1 + P' t' \lambda}, \quad (14)$$

$$P_2 = 1 + ((K_a - \overline{K'_a})^+ + (\underline{K'_a} - K_a)^+) P', \quad (15)$$

$$R_1 = \frac{1}{nt'} \ln \left(\frac{M - \max\{K_a, \underline{K'_a}\}}{t'} \right), \quad (16)$$

$$R_2 = \frac{1}{n} \ln \left(\frac{\min\{K_a, \overline{K'_a}\}}{t} \right), \quad (17)$$

$$q_t = \inf_{\gamma} \left(\mathbb{P}[l_t \leq \gamma] + \sum_{t' \in \mathcal{T}_t} \exp(n(t' R_1 + R_2) - \gamma) \right), \quad (18)$$

$$q_{t,t'} = \inf_{\gamma} \left(\mathbb{P}[l_t \leq \gamma] + \exp(n(t' R_1 + R_2) - \gamma) \right), \quad (19)$$

$$\mathcal{T} = [0 : \min\{\overline{K'_a}, K_a, M - \underline{K'_a} - (K_a - \overline{K'_a})^+\}], \quad (20)$$

$$\mathcal{T}_t = [((K_a - \overline{K'_a})^+ - (\underline{K'_a} - K_a)^+ + \max\{\underline{K'_a}, 1\} - K_a + t)^+ : u_t], \quad (21)$$

$$\overline{\mathcal{T}}_t = [((K_a - \overline{K'_a})^+ - (K_a - \underline{K'_a})^+ + t)^+ : u_t], \quad (22)$$

$$u_t = \min \{ (\overline{K'_a} - K_a)^+ - (\underline{K'_a} - K_a)^+ + t, \overline{K'_a} - (\underline{K'_a} - K_a)^+, M - \max\{\underline{K'_a}, K_a\} \}, \quad (23)$$

$$\xi(K_a, K'_a) = \min_{K: K \neq K'_a} \mathbb{P}[m(\mathbf{y}_0, K'_a) < m(\mathbf{y}_0, K)], \quad (24)$$

in (24), $\mathbf{y}_0 \sim \mathcal{CN}(\mathbf{0}, (1 + K_a P') \mathbf{I}_n)$. The random variable l_t in (18) and (19) is defined as

$$l_t = \min_{\substack{\mathcal{W}_{02} \subset [(\underline{K_a} - \underline{K'_a})^+ + 1 : K_a] \\ |\mathcal{W}_{02}| = t}} \iota_t(c(\mathcal{W}'_{01}) + c(\mathcal{W}_{02}); \mathbf{y} | c([K_a] \setminus \mathcal{W}_0)), \quad (25)$$

where $\mathcal{W}'_{01} = [K_a + 1 : \underline{K'_a}]$, $\mathcal{W}_0 = [(K_a - \overline{K'_a})^+] \cup \mathcal{W}_{02}$, and

$$\begin{aligned} & \iota_t(c(\mathcal{W}_0); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0)) \\ &= n \ln(1 + (t + (K_a - \overline{K'_a})^+) P') + \frac{\|\mathbf{y} - c(\mathcal{W} \setminus \mathcal{W}_0)\|^2}{1 + (t + (K_a - \overline{K'_a})^+) P'} \\ & \quad - \|\mathbf{y} - c(\mathcal{W}_0) - c(\mathcal{W} \setminus \mathcal{W}_0)\|^2. \end{aligned} \quad (26)$$

Some remarks are in order.

- i) The parameters K_l and K_u can be taken to be the essential infimum and the essential supremum of K_a , respectively. In numerical evaluation, it is often convenient to set K_l to be the largest value and K_u the smallest value for which $\sum_{K_a = K_l}^{K_u} P_{K_a}(K_a)$ exceeds a predetermined threshold.
- ii) The term $1 - \mathbb{E}_{K_a} \left[\frac{M!}{M^{K_a} (M - K_a)!} \right]$ in p_0 can be upper-bounded by $\mathbb{E}_{K_a} \left[\frac{\binom{K_a}{2}}{M} \right]$ as in [3].
- iii) The term R_1 in (16) can be upper-bounded by $\frac{1}{n} \ln(M - \max\{K_a, \underline{K'_a}\}) - \frac{1}{nt'} \ln t'!$, which allows for a stable computation when $M - \max\{K_a, \underline{K'_a}\}$ is large.

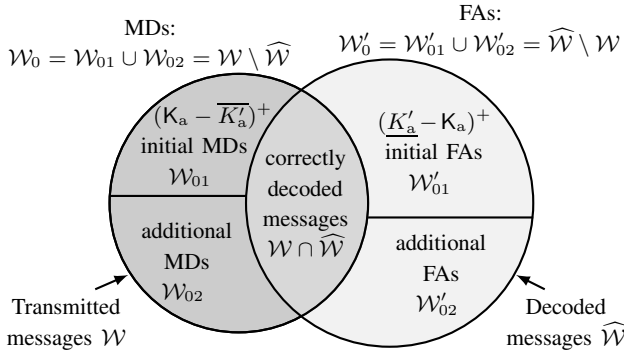


Fig. 1. A diagram depicting the relation between the defined sets of messages.

- iv) The optimal λ in (11) is given by the largest real root of the cubic function $c_1x^3 + c_2x^2 + c_3x + c_4$ with

$$c_1 = -\rho\rho_1(\rho\rho_1 + 1)t'P'P_2P_3^2, \quad (27)$$

$$c_2 = \rho\rho_1t'P'P_3^2 - \rho\rho_1(3 - \rho_1)t'P'P_2P_3 - \rho\rho_1(\rho_1 + 1)P_2P_3^2, \quad (28)$$

$$c_3 = (2\rho - 1)\rho_1t'P'P_3 + \rho_1P_3^2 - 2\rho\rho_1P_2P_3, \quad (29)$$

$$c_4 = (\rho - 1)\rho_1t'P' + \rho_1P_3, \quad (30)$$

where P_2 is given by (15) and $P_3 = (t' + \rho t)P'$.

- v) If the number of active users is fixed to K_a , by letting $K'_a = K_a$ with probability 1 and by setting the decoding radius r to 0, one obtains from Theorem 1 a trivial generalization of [3, Th. 1] to the complex case.

Proof of Theorem 1: We next present a sketch of the proof. The full proof can be found in Appendix A.

Denote by \mathcal{W}_0 the set of misdetected messages, i.e., $\mathcal{W}_0 = \mathcal{W} \setminus \widehat{\mathcal{W}}$, and by \mathcal{W}'_0 the set of falsely alarmed messages, i.e., $\mathcal{W}'_0 = \widehat{\mathcal{W}} \setminus \mathcal{W}$. The MD and FA probabilities, given in (5) and (6), respectively, can be expressed as $P_{\text{MD}} = \mathbb{E}[\mathbb{1}\{|\mathcal{W}| \neq 0\} \cdot \frac{|\mathcal{W}_0|}{|\mathcal{W}|}]$ and $P_{\text{FA}} = \mathbb{E}[\mathbb{1}\{|\widehat{\mathcal{W}}| \neq 0\} \cdot \frac{|\mathcal{W}'_0|}{|\widehat{\mathcal{W}}|}]$. At a cost of adding a constant bounded by p_0 given in (7), we first replace the measure over which the expectation is taken by the one under which: i) there are at least K_l and at most K_u active users; ii) $\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_{K_a}$ are sampled uniformly without replacement from $[M]$; iii) $\mathbf{x}_i = \mathbf{c}_{w_i} \forall i$, instead of $\mathbf{x}_i = \mathbf{c}_{w_i} \mathbb{1}\{\|\mathbf{c}_{w_i}\|^2 \leq nP\}$.

Let $K_a \rightarrow K'_a$ denote the event that the estimation step outputs K'_a while K_a users are active. Given $K_a \rightarrow K'_a$, note that if $\overline{K'_a} < K_a$, the decoder commits at least $K_a - \overline{K'_a}$ MDs; if $\overline{K'_a} > K_a$, the decoder commits at least $\overline{K'_a} - K_a$ FAs. We let $\mathcal{W}_0 = \mathcal{W}_{01} \cup \mathcal{W}_{02}$ where \mathcal{W}_{01} denotes the list of $(K_a - \overline{K'_a})^+$ initial MDs due to insufficient decoded list size, and \mathcal{W}_{02} the additional MDs occurring during decoding. Similarly, let $\mathcal{W}'_0 = \mathcal{W}'_{01} \cup \mathcal{W}'_{02}$ where \mathcal{W}'_{01} denotes the list of $(\overline{K'_a} - K_a)^+$ initial FAs due to excessive decoded list size, and \mathcal{W}'_{02} the additional FAs. Fig. 1 depicts the relation between these sets.

Using the above definitions, the set of transmitted messages is $\mathcal{W} = \mathcal{W}_{01} \cup \mathcal{W}_{02} \cup (\mathcal{W} \setminus \mathcal{W}_0)$, and the received signal is

$\mathbf{y} = c(\mathcal{W}_{01}) + c(\mathcal{W}_{02}) + c(\mathcal{W} \setminus \mathcal{W}_0) + \mathbf{z}$. Since the messages in \mathcal{W}_{01} are always misdetected and the messages in \mathcal{W}'_{01} are always falsely alarmed, the best approximation of \mathcal{W} that the decoder can produce is $\mathcal{W}_{02} \cup (\mathcal{W} \setminus \mathcal{W}_0) \cup \mathcal{W}'_{01}$. However, under the considered error event $\mathcal{W} \rightarrow \widehat{\mathcal{W}}$, the actual decoded list is $\mathcal{W}'_{02} \cup (\mathcal{W} \setminus \mathcal{W}_0) \cup \mathcal{W}'_{01}$. Therefore, $\mathcal{W} \rightarrow \widehat{\mathcal{W}}$ implies the event $F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_{01}, \mathcal{W}'_{02}) = \{\|c(\mathcal{W}_{01}) + c(\mathcal{W}_{02}) - c(\mathcal{W}'_{01}) - c(\mathcal{W}'_{02}) + \mathbf{z}\|^2 < \|c(\mathcal{W}_{01}) - c(\mathcal{W}'_{01}) + \mathbf{z}\|^2\}$.

It follows that, after the change of measure, P_{MD} and P_{FA} can be bounded as

$$P_{\text{MD}} \leq \sum_{K_a=\max\{K_l, 1\}}^{K_u} \left(P_{K_a}(K_a) \sum_{K'_a=K_l}^{K_u} \sum_{t \in \mathcal{T}} \frac{t + (K_a - \overline{K'_a})^+}{K_a} \cdot \mathbb{P}[|\mathcal{W}_{02}| = t, K_a \rightarrow K'_a] \right) + p_0, \quad (31)$$

$$P_{\text{FA}} \leq \sum_{K_a=K_l}^{K_u} \left(P_{K_a}(K_a) \sum_{K'_a=K_l}^{K_u} \sum_{t \in \mathcal{T}} \sum_{t' \in \mathcal{T}_t} \frac{t + (\overline{K'_a} - K_a)^+}{K_a - t - (K_a - \overline{K'_a})^+ + t' + (\overline{K'_a} - K_a)^+} \cdot \mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', K_a \rightarrow K'_a] \right) + p_0, \quad (32)$$

where \mathcal{T} and \mathcal{T}_t are given by (20) and (21), respectively. The constraint $t \in \mathcal{T}$ holds because the number of MDs, given by $t + (K_a - \overline{K'_a})^+$, is upper-bounded by the total number of transmitted messages K_a , and by $M - \overline{K'_a}$ (since at least $\overline{K'_a}$ messages are returned). The constraint $t' \in \mathcal{T}_t$ holds because: i) the decoded list size, given by $K_a - t - (K_a - \overline{K'_a})^+ + t' + (\overline{K'_a} - K_a)^+$, must be in $[\overline{K'_a} : \overline{K'_a}]$ and must be positive since the event $|\widehat{\mathcal{W}}| = 0$ results in no FA by definition; ii) the number of FAs, given by $t' + (\overline{K'_a} - K_a)^+$, is upper-bounded by the number of messages that are not transmitted $M - K_a$, and by the maximal number of decoded messages $\overline{K'_a}$.

Let $A(K_a, K'_a) = \{\mathbf{y} : K'_a = \arg \min_{K \in [K_l, K_u]} m(\mathbf{y}, K)\}$. Since the event $K_a \rightarrow K'_a$ implies $|\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}]$ and $A(K_a, K'_a)$, we have that

$$\begin{aligned} & \mathbb{P}[|\mathcal{W}_{02}| = t, K_a \rightarrow K'_a] \\ & \leq \mathbb{P}[|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}], A(K_a, K'_a)] \\ & \leq \min \{ \mathbb{P}[|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}]], \mathbb{P}[A(K_a, K'_a)] \}. \end{aligned}$$

Similarly, $\mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', K_a \rightarrow K'_a] \leq \min \{ \mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', |\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}]], \mathbb{P}[A(K_a, K'_a)] \}.$

Under the new measure, $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, (1 + K_a P') \mathbf{I}_n)$. Thus, we can show that $\mathbb{P}[A(K_a, K'_a)]$ is upper-bounded by $\xi(K_a, K'_a)$ given by (24). To establish (5) and (6), we proceed as in [3], write the events $\{|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}]\}$ and $\{|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', |\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}]\}$ as the union of $F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_{01}, \mathcal{W}'_{02})$ events, and bound their probabilities by $\min\{p_t, q_t\}$ and $\min\{p_{t,t'}, q_{t,t'}\}$, respectively.

Finally, to guarantee that *both* (5) and (6) are satisfied, we allow for randomized coding strategies, by introducing the variable \mathbf{U} , which acts as common randomness. Proceeding as in [20, Th. 19], one can show that it is sufficient to perform randomization across (at most) three deterministic codes, i.e., $|\mathcal{U}| \leq 3$. ■

In the following proposition, we derive $\xi(K_a, K'_a)$ for two different estimators of K_a .

Proposition 1: For the maximum likelihood (ML) estimation of K_a , i.e., $m(\mathbf{y}, K) = -\ln p_{\mathbf{y}|K_a}(\mathbf{y}|K)$, $\xi(K_a, K'_a)$ is given by

$$\xi(K_a, K'_a) = \min_{K: K \neq K'_a} \left(\mathbb{1}\{K < K'_a\} \frac{\Gamma(n, \zeta(K, K_a, K'_a))}{\Gamma(n)} + \mathbb{1}\{K > K'_a\} \frac{\gamma(n, \zeta(K, K_a, K'_a))}{\Gamma(n)} \right), \quad (33)$$

with

$$\zeta(K, K_a, K'_a) = n \ln \left(\frac{1 + KP'}{1 + K'_a P'} \right) (1 + K_a P')^{-1} \cdot \left(\frac{1}{1 + K'_a P'} - \frac{1}{1 + KP'} \right)^{-1}. \quad (34)$$

For an energy-based estimation of K_a with $m(\mathbf{y}, K) = \|\mathbf{y}\|^2 - n(1 + KP')$, $\xi(K_a, K'_a)$ is given by (33) with

$$\zeta(K, K_a, K'_a) = \frac{n}{1 + K_a P'} \left(1 + \frac{K + K'_a P'}{2} \right). \quad (35)$$

Proof: See Appendix B. ■

The decoding radius r can be optimized according to the target MD and FA probabilities. A large decoding radius reduces the initial MDs and FAs at the cost of overfitting, especially at low SNR values. Specifically, when the noise dominates, increasing r increases the chance that the decoder (4) returns a list whose sum is closer in Euclidean distance to the noise than to the sum of the transmitted codewords.

Our random-coding bound can also be applied to SA-MPR to investigate the resulting MD-FA trade-off. Consider an SA-MPR scheme where a length- n frame is divided into L slots and each user chooses randomly a slot to transmit. For K_a active users, the number of users transmitting in a slot follows a Binomial distribution with parameter $(K_a, 1/L)$. The PMF of the number of active users per slot, denoted by K_{SA} , is given by $P_{K_{SA}}(K_{SA}) = \sum_{K_a} P_{K_a}(K_a) \binom{K_a}{K_{SA}} L^{-K_{SA}} (1 - \frac{1}{L})^{K_a - K_{SA}}$. Existing analyses of slotted ALOHA usually assume that the decoder can detect perfectly if no signal, one signal, or more than one signal have been transmitted in a slot. Furthermore, it is usually assumed that a collision-free slot leads to successful message decoding. However, the more slots, the shorter the slot length over which a user transmits its signal. To account for both detection and decoding errors, in Corollary 1 below, we apply our decoder in a slot-by-slot manner, and obtain a random-coding bound similar to Theorem 1.

Corollary 1: For the Gaussian MAC with the number of active users following P_{K_a} and frame length n , an SA-MPR scheme with L slots can achieve the MD and FA probabilities

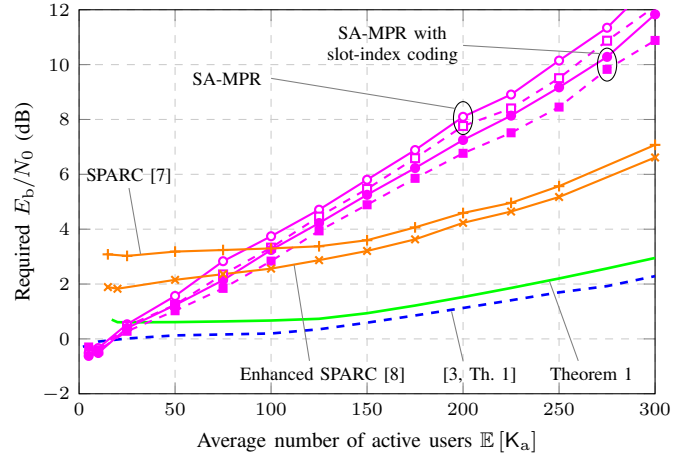


Fig. 2. The required E_b/N_0 to achieve $\max\{P_{MD}, P_{FA}\} \leq 0.1$ vs. $\mathbb{E}[K_a]$ for $k = 128$ bits, $n = 19200$ channel uses, and $K_a \sim \text{Poisson}(\mathbb{E}[K_a])$. We compare our random coding bound (Theorem 1) and SA-MPR bound (Corollary 1) with the bound in [3, Th. 1] for K_a known, and two practical schemes, namely, the SPARC scheme [7], and its enhancement [8]. Solid lines represent schemes/bounds with K_a unknown; dashed lines represent schemes/bounds with K_a known.

given in (5) and (6), respectively, with codebook size M , codeword length n/L , power constraint PL , and per-slot number of active users following $P_{K_{SA}}$.

IV. NUMERICAL EVALUATION

In this section, we numerically evaluate the proposed random-coding bound and compare it with the random-access bound/schemes in [3], [7], [8]. We assume that K_a follows a Poisson distribution. We consider $k = \log_2 M = 128$ bits and $n = 19200$ complex channel uses (i.e., 38400 real degrees of freedom). The power is given in terms of average bit energy $E_b/N_0 = nP/k$.

In Fig. 2, we compare our random-coding bound with that of Polyanskiy [3] in terms of the required E_b/N_0 so that neither P_{MD} nor P_{FA} exceed 0.1. For our bound, we consider the ML estimator of K_a and zero decoding radius, i.e., $\bar{K}'_a = K'_a = K'_a$. Numerical evaluation suggests that this choice is optimal for these target MD and FA probabilities. We choose K_l to be the largest value and K_u the smallest value for which $\mathbb{P}[K_a \notin [K_l, K_u]] < 10^{-9}$. The q_t and $q_{t,t'}$ terms are evaluated for $t = 1$ and $K_a \leq 50$ only. For the bound in [3, Th. 1], we average over the Poisson distribution of K_a . This corresponds to the scenario where K_a is random but known. As can be seen, the extra required E_b/N_0 due to the lack of knowledge of K_a is about 0.5–0.7 dB.

In Fig. 2, we also show the performance of the SA-MPR bound given in Corollary 1, where we optimize L and the decoding radius for each $\mathbb{E}[K_a]$. We also consider the possibility to encode $\lfloor \log_2 L \rfloor$ extra bits for each user in the slot index, and assume perfect decoding of these bits. We refer to this scheme as SA-MPR with slot-index coding. We also evaluate the performance of two practical schemes, namely:

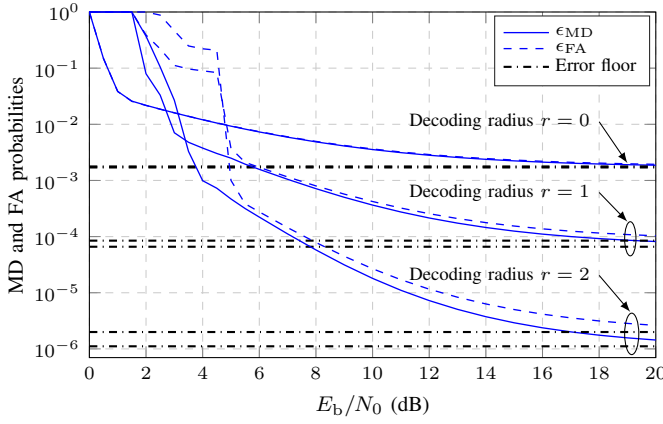


Fig. 3. The bounds on the MD and FA probabilities vs. E_b/N_0 for $k = 128$ bits, $n = 19200$ channel uses, and $K_a \sim \text{Poisson}(50)$.

- the SPARC scheme proposed in [7], which employs a concatenated coding framework with an inner approximate message passing (AMP) decoder followed by an outer tree decoder.
- an enhancement of the SPARC scheme proposed in [8], which we refer to as enhanced SPARC. This scheme introduces belief propagation between the inner AMP decoder and the outer tree decoder in an iterative manner.

Note that the SPARC and enhanced SPARC schemes were proposed for the Gaussian MAC with *known* number of active users. To adapt these schemes to the case where K_a is unknown, we employ an energy-based estimation of K_a , then treat this estimate as the true K_a in the decoding process. From Fig. 2, we see that SA-MPR, even with slot-index coding, becomes power inefficient as $\mathbb{E}[K_a]$ increases. The enhanced SPARC scheme achieves the closest performance to our bound for $\mathbb{E}[K_a] \geq 100$. It outperforms the original SPARC scheme by about 0.5 dB for large $\mathbb{E}[K_a]$.

In Fig. 3, we plot the bounds on the MD and FA probabilities in Theorem 1 (with ML estimation of K_a) as a function of E_b/N_0 for different decoding radii. We observe that decoding with a small radius performs better in the low E_b/N_0 regime, where noise overfitting is the bottleneck. Increasing the decoding radius improves the performance in the moderate and high E_b/N_0 regime, where setting $r = 0$ results in a high error floor due to the initial MDs and FAs. The error floor can be characterized analytically (see Appendix C).

V. CONCLUSIONS

We proposed a formulation for massive uncoordinated access where both the identity and the number of active users are unknown. We derived a random-coding bound for the Gaussian multiple access channel that reveals a trade-off between misdetection and false alarm. Our bound provides an estimate of the penalty in terms of energy efficiency due to the lack of knowledge of the number of active users, and serves as a benchmark to assess the performance of practical

schemes. Possible future works include extending our bound to the MAC with fading and multiple antennas.

ACKNOWLEDGEMENT

This work has been supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP).

APPENDIX A PROOF OF THEOREM 1

The following well-known results will be used in our proof.

Lemma 1 (Change of measure [21, Lemma 4]): Let p and q be two probability measures. Consider a random variable x supported on \mathcal{H} and a function $f: \mathcal{H} \rightarrow [0, 1]$. It holds that

$$\mathbb{E}_p[f(x)] \leq \mathbb{E}_q[f(x)] + d_{TV}(p, q) \quad (36)$$

where $d_{TV}(p, q)$ denotes the total variation distance between p and q .

Lemma 2 (Chernoff bound [22, Th. 6.2.7]): For a random variable x with moment-generating function $\mathbb{E}[e^{tx}]$ defined for all $|t| \leq b$, it holds for all $\lambda \in [0, b]$ that

$$\mathbb{P}[x \leq x] \leq e^{\lambda x} \mathbb{E}[e^{-\lambda x}]. \quad (37)$$

Lemma 3 (Gallager's ρ -trick [18, p. 136]): It holds that $\mathbb{P}[\cup_i A_i] \leq (\sum_i \mathbb{P}[A_i])^\rho$ for every $\rho \in [0, 1]$.

Lemma 4: Let $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$. For all $\gamma > -\frac{1}{\sigma^2}$, it holds that

$$\mathbb{E}[e^{-\gamma \|\mathbf{x}\|^2}] = (1 + \gamma \sigma^2)^{-n} \exp\left(-\frac{\gamma \|\boldsymbol{\mu}\|^2}{1 + \gamma \sigma^2}\right). \quad (38)$$

We present next an error analysis of the random-coding scheme introduced in Section III. Denote by \mathcal{W}_0 the set of misdetected messages, i.e., $\mathcal{W}_0 = \mathcal{W} \setminus \widehat{\mathcal{W}}$, and by \mathcal{W}'_0 the set of falsely alarmed messages, i.e., $\mathcal{W}'_0 = \widehat{\mathcal{W}} \setminus \mathcal{W}$. The MD and FA probabilities, defined respectively in (5) and (6), can be expressed as the average fraction of misdetected and falsely alarmed messages, respectively, i.e.,

$$P_{MD} = \mathbb{E}\left[\mathbb{1}\{|\mathcal{W}| \neq 0\} \cdot \frac{|\mathcal{W}_0|}{|\mathcal{W}|}\right], \quad (39)$$

$$P_{FA} = \mathbb{E}\left[\mathbb{1}\{|\widehat{\mathcal{W}}| \neq 0\} \cdot \frac{|\mathcal{W}'_0|}{|\widehat{\mathcal{W}}|}\right]. \quad (40)$$

A. A Change of Measure

Recall that $|\mathcal{W}|$ is the number of *distinct* transmitted messages. Since multiple transmitters may pick the same codeword to transmit, $|\mathcal{W}|$ can be smaller than K_a . Since both $\mathbb{1}\{|\mathcal{W}| \neq 0\} \cdot \frac{|\mathcal{W}_0|}{|\mathcal{W}|}$ and $\mathbb{1}\{|\widehat{\mathcal{W}}| \neq 0\} \cdot \frac{|\mathcal{W}'_0|}{|\widehat{\mathcal{W}}|}$ are nonnegative and upper-bounded by one, we can apply Lemma 1 to these random quantities. Specifically, we replace the measure over which the expectation is taken by the one under which: i) there are at least K_l and at most $K_u \geq \overline{K}_a$ active users, i.e., $K_l \leq K_a \leq K_u$; ii) $\tilde{w}_1, \dots, \tilde{w}_{K_a}$ are sampled uniformly without replacement from $[M]$, i.e., $|\mathcal{W}| = K_a$; iii) $\mathbf{x}_i = \mathbf{c}_{w_i}, \forall i$ (instead of $\mathbf{x}_i = \mathbf{c}_{w_i} \mathbb{1}\{\|\mathbf{c}_{w_i}\|^2 \leq nP\}$).

It then follows from [23, Eq. (41)] that the total variation between the true measure and the new one is upper-bounded

by $\mathbb{P}[K_a \notin [K_l : K_u]] + \mathbb{P}[|\mathcal{W}| < K_a] + \mathbb{P}[\overline{U}]$, where $U = \{\|\mathbf{c}_{w_i}\|^2 \leq nP, \forall i \in [K_a]\}$ and \overline{U} denotes the complement of U . We compute these probabilities as follows:

- To compute the first probability, we simply use that $\mathbb{P}[K_a \notin [K_l : K_u]] = 1 - \sum_{K_a=K_l}^{K_u} P_{K_a}(K_a)$.
- Consider a given $K_a = K_a$. Since $\tilde{w}_1, \dots, \tilde{w}_{K_a}$ are drawn uniformly and independently from $[M]$, there are M^{K_a} possible K_a -tuples. Among them, $\frac{M!}{(M-K_a)!}$ tuples have nonduplicate elements. Therefore, $\mathbb{P}[|\mathcal{W}| = K_a | K_a = K_a] = \frac{M!}{(M-K_a)!} \frac{1}{M^{K_a}}$. As a consequence, $\mathbb{P}[|\mathcal{W}| < K_a] = 1 - \mathbb{P}[|\mathcal{W}| = K_a] = 1 - \mathbb{E}_{K_a} \left[\frac{M!}{M^{K_a} (M-K_a)!} \right]$.
- The probability $\mathbb{P}[\overline{U}]$ can be finally evaluated as

$$\mathbb{P}[\overline{U}] = \mathbb{E}_{K_a} \left[\mathbb{P} \left[\bigcup_{i=1}^{K_a} \|\mathbf{c}_{w_i}\|^2 > nP \right] \right] \quad (41)$$

$$\leq \mathbb{E}_{K_a} \left[\sum_{i=1}^{K_a} \mathbb{P} [\|\mathbf{c}_{w_i}\|^2 > nP] \right] \quad (42)$$

$$= \mathbb{E}_{K_a} \left[\frac{\Gamma(n, nP/P')}{\Gamma(n)} \right], \quad (43)$$

where (42) follows from the union bound and (43) holds since $\|\mathbf{c}_{w_i}\|^2$ follows the Gamma distribution with shape n and scale P' .

From the above calculations, we deduce that the total variation between the two measures is upper-bounded by p_0 defined in (7). Applying Lemma 1 to the random quantities $\mathbb{1}\{|\widehat{\mathcal{W}}| \neq 0\} \cdot \frac{|\mathcal{W}_0|}{|\widehat{\mathcal{W}}|}$ and $\mathbb{1}\{|\mathcal{W}| \neq 0\} \cdot \frac{|\mathcal{W}'_0|}{|\widehat{\mathcal{W}}|}$, we consider implicitly the new measure from now on at a cost of adding p_0 to their original expectations. It remains to bound the MD and FA probabilities given in (39) and (40), respectively, under the new measure. For the sake of clarity, in Appendix A-B, we shall prove a bound on P_{MD} and P_{FA} for a special case where i) K_a and K'_a are fixed and $r = 0$, i.e., there are always K_a users transmitting and the decoder always outputs a list of size K'_a ; ii) $K'_a < \min\{K_a, M - K_a\}$. Then, in Appendix A-C, we shall show how to extend the proof for the general case where K_a and K'_a are random and $r \geq 0$.

B. A Special Case

In the aforementioned special case, (5) and (6) become

$$\epsilon_{MD} = \sum_{t=0}^{K'_a} \frac{t + K_a - K'_a}{K_a} \min\{p_{t,t}, q_{t,t}\} + p_0, \quad (44)$$

$$\epsilon_{FA} = \sum_{t=0}^{K'_a} \frac{t}{K'_a} \min\{p_{t,t}, q_{t,t}\} + p_0, \quad (45)$$

where $p_{t,t}$ and $q_{t,t}$ will be derived next. We next show that ϵ_{MD} and ϵ_{FA} are indeed upper-bounds of P_{MD} and P_{FA} , respectively, in this special case.

Observe that since the decoded list size K'_a is smaller than the number of transmitted messages K_a , at least $K_a - K'_a$

messages are misdetected by default, and there can be $t \in [0 : K'_a]$ additional MDs occurring during the decoding process. Due to symmetry, we can assume without loss of generality that $\mathcal{W} = [K_a]$ and that the list of messages that are initially misdetected due to insufficient decoded list size is $\mathcal{W}_{01} = [K_a - K'_a]$.² Furthermore, let $\mathcal{W}_{02} = \mathcal{W}_0 \setminus \mathcal{W}_{01}$ denote the set of t additional MDs. Note that \mathcal{W}_{02} is a generic subset of $[K_a - K'_a + 1 : K_a]$. Note also that t is the number of FAs, i.e., $|\mathcal{W}'_0| = t$. The set of transmitted messages can thus be expressed as $\mathcal{W} = \mathcal{W}_{01} \cup \mathcal{W}_{02} \cup (\mathcal{W} \setminus \mathcal{W}_0)$, and the received signal is $\mathbf{y} = c(\mathcal{W}_{01}) + c(\mathcal{W}_{02}) + c(\mathcal{W} \setminus \mathcal{W}_0) + \mathbf{z}$. Since the messages in \mathcal{W}_{01} are always misdetected, the best approximation of \mathcal{W} that the decoder can produce is $\mathcal{W}_{02} \cup (\mathcal{W} \setminus \mathcal{W}_0)$. However, under the considered error event $\mathcal{W} \rightarrow \widehat{\mathcal{W}}$, the actual decoded list is $\mathcal{W}'_0 \cup (\mathcal{W} \setminus \mathcal{W}_0)$. Therefore, $\mathcal{W} \rightarrow \widehat{\mathcal{W}}$ implies that $\|\mathbf{y} - c(\mathcal{W}'_0) - c(\mathcal{W} \setminus \mathcal{W}_0)\|^2 < \|\mathbf{y} - c(\mathcal{W}_{02}) - c(\mathcal{W} \setminus \mathcal{W}_0)\|^2$, which is equivalent to

$$\|c(\mathcal{W}_{01}) + c(\mathcal{W}_{02}) - c(\mathcal{W}'_0) + \mathbf{z}\|^2 < \|c(\mathcal{W}_{01}) + \mathbf{z}\|^2. \quad (46)$$

We denote by $F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_0)$ the set of $\{\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_0\}$ that satisfy (46).

We now compute the expectations in (39) and (40). Recall that, under assumptions just stated, we have $|\mathcal{W}_0| = t + K_a - K'_a$, $|\mathcal{W}'_0| = |\mathcal{W}_{02}| = t$, and $|\widehat{\mathcal{W}}| = K'_a$. It follows from (39) and (40) that, after the change of measure in Appendix A-A, P_{MD} and P_{FA} can be bounded as

$$P_{MD} \leq \sum_{t=0}^{K'_a} \frac{t + K_a - K'_a}{K_a} \mathbb{P}[|\mathcal{W}_{02}| = t] + p_0, \quad (47)$$

$$P_{FA} \leq \sum_{t=0}^{K'_a} \frac{t}{K'_a} \mathbb{P}[|\mathcal{W}_{02}| = t] + p_0, \quad (48)$$

Next, we proceed to bound $\mathbb{P}[|\mathcal{W}_{02}| = t]$. This is done following two approaches. The first approach is based on error exponent analyses, resulting in the term $p_{t,t}$ in (44). The second approach is a variation of the dependence testing (DT) bound [24, Th. 17], resulting in $q_{t,t}$ in (44).

1) *The Error-Exponent-Based Approach:* By writing the event $|\mathcal{W}_{02}| = t$ as the union of the pairwise error events $F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}_0)$, we have that

$$\begin{aligned} \mathbb{P}[|\mathcal{W}_{02}| = t] \\ = \mathbb{P} \left[\bigcup_{\substack{\mathcal{W}_{02} \subset [K_a - K'_a + 1 : K_a] \\ |\mathcal{W}_{02}| = t}} \bigcup_{\substack{\mathcal{W}'_0 \subset [K_a + 1 : M] \\ |\mathcal{W}'_0| = t}} F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_0) \right], \end{aligned} \quad (49)$$

Next, given $c(\mathcal{W}_{01})$, $c(\mathcal{W}_{02})$, and \mathbf{z} , it holds for every $\lambda > -\frac{1}{tP'}$ that

$$\begin{aligned} \mathbb{P}[F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_0)] \\ \leq e^{\lambda \|c(\mathcal{W}_{01}) + \mathbf{z}\|^2} \end{aligned}$$

¹In [3], $\mathbb{P}[|\mathcal{W}| \leq K_a]$ is upper-bounded by $\binom{K_a}{2}/M$ using the union bound.

²Note that due to user's identity ambiguity, this does not imply that the messages from a set of specific users are always misdetected.

$$\begin{aligned}
& \cdot \mathbb{E}_{c(\mathcal{W}'_0)} \left[e^{-\lambda \|c(\mathcal{W}_{01}) + c(\mathcal{W}_{02}) - c(\mathcal{W}'_0) + \mathbf{z}\|^2} \right] \quad (50) \\
& = e^{\lambda \|c(\mathcal{W}_{01}) + \mathbf{z}\|^2} (1 + \lambda t P')^{-n} \\
& \cdot \exp \left(- \frac{\lambda \|c(\mathcal{W}_{01}) + c(\mathcal{W}_{02}) + \mathbf{z}\|^2}{1 + \lambda t P'} \right), \quad (51)
\end{aligned}$$

where (50) follows from the Chernoff bound in Lemma 2, and (51) follows by computing the expectation in (50) using Lemma 4. Next, we apply Gallager's ρ -trick in Lemma 3 to get that, given $c(\mathcal{W}_{01})$, $c(\mathcal{W}_{02})$, and \mathbf{z} , it holds for every $\rho \in [0, 1]$ that

$$\begin{aligned}
& \mathbb{P} \left[\bigcup_{\substack{\mathcal{W}'_0 \subset [K_a+1:M] \\ |\mathcal{W}'_0|=t}} F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_0) \right] \quad (52) \\
& \leq \binom{M - K_a}{t}^\rho (1 + \lambda t P')^{-n\rho} \\
& \cdot \exp \left(\lambda \rho \left(\|c(\mathcal{W}_{01}) + \mathbf{z}\|^2 - \frac{\|c(\mathcal{W}_{01}) + c(\mathcal{W}_{02}) + \mathbf{z}\|^2}{1 + \lambda t P'} \right) \right). \quad (53)
\end{aligned}$$

Taking the expectation over $c(\mathcal{W}_{02})$ using Lemma 4, we obtain for given $c(\mathcal{W}_{01})$ and \mathbf{z} that

$$\begin{aligned}
& \mathbb{P} \left[\bigcup_{\substack{\mathcal{W}'_0 \subset [K_a+1:M] \\ |\mathcal{W}'_0|=t}} F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_0) \right] \\
& \leq \binom{M - K_a}{t}^\rho (1 + \lambda t P')^{-n\rho} \left(1 + \frac{\lambda \rho t P'}{1 + \lambda t P'} \right)^{-n} \\
& \cdot \exp \left(\lambda \rho \left(1 - \frac{1}{1 + \lambda P' t (1 + \rho)} \right) \|c(\mathcal{W}_{01}) + \mathbf{z}\|^2 \right) \quad (54) \\
& = \binom{M - K_a}{t}^\rho \exp (b_0 \|c(\mathcal{W}_{01}) + \mathbf{z}\|^2 - n a_0), \quad (55)
\end{aligned}$$

where a_0 and b_0 are given by taking $t' = t$ in (12) and (13), respectively. Now applying Gallager's ρ -trick again, we obtain that, for every $\rho_1 \in [0, 1]$,

$$\begin{aligned}
& \mathbb{P} \left[\bigcup_{\substack{\mathcal{W}_{02} \subset [K_a - K'_a + 1: K_a] \\ |\mathcal{W}_{02}|=t}} \bigcup_{\substack{\mathcal{W}'_0 \subset [K_a + 1: M] \\ |\mathcal{W}'_0|=t}} F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_0) \right] \\
& \leq \binom{K'_a}{t}^{\rho_1} \binom{M - K_a}{t}^{\rho \rho_1} \\
& \cdot \mathbb{E} [\exp (\rho_1 b_0 \|c(\mathcal{W}_{01}) + \mathbf{z}\|^2 - n \rho_1 a_0)] \quad (56) \\
& = \binom{K'_a}{t}^{\rho_1} \binom{M - K_a}{t}^{\rho \rho_1} e^{-n \rho_1 a_0} (1 - \rho_1 P_2 b_0)^{-n}, \quad (57)
\end{aligned}$$

where the last equality follows by computing the expectation in (56) jointly over $c(\mathcal{W}_{01})$ and \mathbf{z} with the help of Lemma 4, and $P_2 = 1 + (K_a - K'_a)P'$. Finally, plugging the result into (49), we obtain

$$\mathbb{P} [|\mathcal{W}_{02}| = t]$$

$$\begin{aligned}
& \leq \binom{K'_a}{t}^{\rho_1} \binom{M - K_a}{t}^{\rho \rho_1} e^{-n \rho_1 a_0} (1 - \rho_1 P_2 b_0)^{-n} \quad (58) \\
& = p_{t,t}. \quad (59)
\end{aligned}$$

2) *The DT-Based Approach:* Next, we present an alternative bound on $\mathbb{P} [|\mathcal{W}_{02}| = t]$. Consider the channel law $P_{\mathbf{y} | c(\mathcal{W}_0), c(\mathcal{W} \setminus \mathcal{W}_0)}$ with input $c(\mathcal{W}_0)$ and output \mathbf{y} where $|\mathcal{W}_{02}| = t$. The corresponding information density [19, Def. 17.1] is given by

$$\begin{aligned}
& \iota_t(c(\mathcal{W}_0); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0)) \\
& = n \ln(1 + (t + K_a - K'_a)P') + \frac{\|\mathbf{y} - c(\mathcal{W} \setminus \mathcal{W}_0)\|^2}{1 + (t + K_a - K'_a)P'} \\
& - \|\mathbf{y} - c(\mathcal{W}_0) - c(\mathcal{W} \setminus \mathcal{W}_0)\|^2. \quad (60)
\end{aligned}$$

Notice that the event $F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_0)$ defined in (46) is equivalent to $\{\iota_t(c(\mathcal{W}'_0); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0)) > \iota_t(c(\mathcal{W}_{02}); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0))\}$. Let

$$I_t = \min_{\substack{\mathcal{W}_{02} \subset [K_a - K'_a + 1: K_a] \\ |\mathcal{W}_{02}|=t}} \iota_t(c(\mathcal{W}_{02}); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0)). \quad (61)$$

For a fixed arbitrary γ , it follows that

$$\begin{aligned}
& \mathbb{P} [|\mathcal{W}_{02}| = t] \\
& = \mathbb{P} [I_t \leq \gamma] \mathbb{P} [|\mathcal{W}_{02}| = t | I_t \leq \gamma] \\
& + \mathbb{P} [I_t > \gamma] \mathbb{P} [|\mathcal{W}_{02}| = t | I_t > \gamma] \quad (62)
\end{aligned}$$

$$\begin{aligned}
& \leq \mathbb{P} [I_t \leq \gamma] + \mathbb{P} [|\mathcal{W}_{02}| = t | I_t > \gamma] \quad (63) \\
& = \mathbb{P} [I_t \leq \gamma]
\end{aligned}$$

$$\begin{aligned}
& + \mathbb{P} \left[\bigcup_{\substack{\mathcal{W}_{02} \subset [K_a - K'_a + 1: K_a] \\ |\mathcal{W}_{02}|=t}} \bigcup_{\substack{\mathcal{W}'_0 \subset [K_a + 1: M] \\ |\mathcal{W}'_0|=t}} \right. \\
& \quad \left. \{ \iota_t(c(\mathcal{W}'_0); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0)) \right. \\
& \quad \left. > \iota_t(c(\mathcal{W}_{02}); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0)) \} | I_t > \gamma \right] \quad (64)
\end{aligned}$$

$$\begin{aligned}
& \leq \mathbb{P} [I_t \leq \gamma] \\
& + \mathbb{P} \left[\bigcup_{\substack{\mathcal{W}_{02} \subset [K_a - K'_a + 1: K_a] \\ |\mathcal{W}_{02}|=t}} \bigcup_{\substack{\mathcal{W}'_0 \subset [K_a + 1: M] \\ |\mathcal{W}'_0|=t}} \right. \\
& \quad \left. \{ \iota_t(c(\mathcal{W}'_0); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0)) > \gamma \} \right]. \quad (65)
\end{aligned}$$

Here, (64) follows by writing explicitly the event $\{|\mathcal{W}_{02}| = t\}$, and (65) follows by relaxing the inequality inside the second probability. Using that $\mathbb{P} [\iota(x; \mathbf{y}) > \gamma] \leq e^{-\gamma}, \forall x$ [19, Cor. 17.1], we obtain

$$\mathbb{P} [\iota_t(c(\mathcal{W}'_0); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0)) > \gamma] \leq e^{-\gamma}. \quad (66)$$

Then, by applying the union bound and taking the infimum over γ , we conclude that

$$\begin{aligned}
& \mathbb{P} [|\mathcal{W}_{02}| = t] \\
& \leq \inf_{\gamma} \left(\mathbb{P} [I_t \leq \gamma] + \binom{K'_a}{t} \binom{M - K_a}{t} e^{-\gamma} \right) \quad (67)
\end{aligned}$$

$$= q_{t,t}. \quad (68)$$

This concludes the DT-based approach.

It follows from (59) and (68) that $\mathbb{P}[|\mathcal{W}_{02}| = t] \leq \min\{p_{t,t}, q_{t,t}\}$. Introducing this bound into (47) and (48), we obtain that the MD and FA probabilities, averaged over the Gaussian codebook ensemble, are upper-bounded by ϵ_{MD} and ϵ_{FA} given in (44) and (45), respectively.

C. The General Case

We now explain how the result in the special case considered in the previous subsection can be extended to the general case where K_a and K'_a are random and $r \geq 0$.

For random K_a and K'_a , one has to take into account all the possible combinations of the number of transmitted messages and decoded messages when computing the expectations in (39) and (40). Consider the event that K_a users are active and the estimation of K_a results in K'_a , which we denote by $K_a \rightarrow K'_a$. As in the special case, we assume without loss of generality that $\mathcal{W} = [K_a]$. Furthermore, due to symmetry, we let $\mathcal{W}_{01} = [(K_a - \overline{K'_a})^+]$ denote the list of $(K_a - \overline{K'_a})^+$ initial MDs due to insufficient decoded list size, and $\mathcal{W}_{02} = \mathcal{W}_0 \setminus \mathcal{W}_{01}$ the t additional MDs occurring during the decoding process. Note also that, if $\overline{K'_a} > K_a$, the decoder always outputs more than K_a messages. Hence, at least $\overline{K'_a} - K_a$ decoded messages are falsely alarmed. Due to symmetry, let $\mathcal{W}'_{01} = [K_a + 1 : \overline{K'_a}]$ denote the list of $(\overline{K'_a} - K_a)^+$ initial FAs due to excessive decoded list size, and $\mathcal{W}'_{02} = \mathcal{W}'_0 \setminus \mathcal{W}'_{01}$ the t' additional FAs occurring during the decoding process. See Fig. 1 for a diagram depicting the relation between these sets of messages. Under these assumptions, \mathcal{W}_{02} and \mathcal{W}'_{02} are generic subsets of $[(K_a - \overline{K'_a})^+ + 1 : K_a]$ and $[\max\{K_a, \overline{K'_a}\} + 1 : M]$, respectively.

Note that in the special case considered in Appendix A-B, t can take value from 0 to K'_a , and $t' = t$. In the general case, instead:

- The possible values of t are given by \mathcal{T} defined in (20). This is because the number of MDs, given by $t + (K_a - \overline{K'_a})^+$, is upper-bounded by the total number of transmitted messages K_a , and by $M - \overline{K'_a}$ (since at least $\overline{K'_a}$ messages are returned).
- Given t , the integer t' takes value in $\overline{\mathcal{T}}_t$ defined in (22) because: i) the decoded list size, given by $K_a - t - (K_a - \overline{K'_a})^+ + t' + (\overline{K'_a} - K_a)^+$, must be in $[\overline{K'_a} : \overline{K'_a}]$; ii) the number of FAs, given by $t' + (\overline{K'_a} - K_a)^+$, is upper-bounded by the number of messages that are not transmitted $M - K_a$, and by the maximal number of decoded messages $\overline{K'_a}$.
- If the decoded list size is further required to be strictly positive, then t' takes value in \mathcal{T}_t defined in (21).

Using the above definitions, the best approximation of \mathcal{W} that the decoder can produce is $\mathcal{W}_{02} \cup (\mathcal{W} \setminus \mathcal{W}_0) \cup \mathcal{W}'_{01}$, while the actual decoded list, under $\mathcal{W} \rightarrow \widehat{\mathcal{W}}$, is $\mathcal{W}_{02} \cup (\mathcal{W} \setminus \mathcal{W}_0) \cup \mathcal{W}'_{01}$. Therefore, $\mathcal{W} \rightarrow \widehat{\mathcal{W}}$ implies that $\|\mathbf{y} - c(\mathcal{W}'_{02}) - c(\mathcal{W} \setminus \mathcal{W}_0) - c(\mathcal{W}'_{01})\|^2 < \|\mathbf{y} - c(\mathcal{W}_{02}) - c(\mathcal{W} \setminus \mathcal{W}_0) - c(\mathcal{W}'_{01})\|^2$, which is equivalent to

$$\|c(\mathcal{W}_{01}) + c(\mathcal{W}_{02}) - c(\mathcal{W}'_{01}) - c(\mathcal{W}'_{02}) + \mathbf{z}\|^2 < \|c(\mathcal{W}_{01}) - c(\mathcal{W}'_{01}) + \mathbf{z}\|^2. \quad (69)$$

We denote by $F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_{01}, \mathcal{W}'_{02})$ the set of $\{\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_{01}, \mathcal{W}'_{02}\}$ that satisfy (69).

We now compute the expectations in P_{MD} and P_{FA} . Given $|\mathcal{W}_{02}| = t$ and $|\mathcal{W}'_{02}| = t'$, we have that $|\mathcal{W}_0| = t + (K_a - \overline{K'_a})^+$, $|\mathcal{W}'_0| = t + (\overline{K'_a} - K_a)^+$, and $|\widehat{\mathcal{W}}| = K_a - t - (K_a - \overline{K'_a})^+ + t' + (\overline{K'_a} - K_a)^+$. It follows from (39) and (40) that, after the change of measure in Appendix A-A, P_{MD} and P_{FA} can be bounded as

$$P_{\text{MD}} \leq \sum_{K_a=\max\{K_l, 1\}}^{K_u} P_{K_a}(K_a) \sum_{K'_a=K_l}^{K_u} \sum_{t \in \mathcal{T}} \frac{t + (K_a - \overline{K'_a})^+}{K_a} \cdot \mathbb{P}[|\mathcal{W}_{02}| = t, K_a \rightarrow K'_a] + p_0, \quad (70)$$

$$P_{\text{FA}} \leq \sum_{K_a=K_l}^{K_u} P_{K_a}(K_a) \sum_{K'_a=K_l}^{K_u} \sum_{t \in \mathcal{T}} \sum_{t' \in \overline{\mathcal{T}}_t} \frac{t + (\overline{K'_a} - K_a)^+}{K_a - t - (K_a - \overline{K'_a})^+ + t' + (\overline{K'_a} - K_a)^+} \cdot \mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', K_a \rightarrow K'_a] + p_0. \quad (71)$$

Next, we proceed to bound the joint probabilities $\mathbb{P}[|\mathcal{W}_{02}| = t, K_a \rightarrow K'_a]$ and $\mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', K_a \rightarrow K'_a]$. Let

$$A(K_a, K'_a) = \{m(\mathbf{y}, K'_a) < m(\mathbf{y}, K), \forall K \neq K'_a\}. \quad (72)$$

Since the event $K_a \rightarrow K'_a$ implies that $|\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}]$ and $A(K_a, K'_a)$, we have

$$\begin{aligned} & \mathbb{P}[|\mathcal{W}_{02}| = t, K_a \rightarrow K'_a] \\ & \leq \mathbb{P}[|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}], A(K_a, K'_a)] \\ & \leq \min \left\{ \mathbb{P}[|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}]], \mathbb{P}[A(K_a, K'_a)] \right\}, \end{aligned} \quad (74)$$

where (74) follows from the fact that the joint probability is upper-bounded by each of the individual probabilities. Similarly, we can show that

$$\begin{aligned} & \mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', K_a \rightarrow K'_a] \\ & \leq \min \left\{ \mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', |\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}]], \right. \\ & \quad \left. \mathbb{P}[A(K_a, K'_a)] \right\}. \end{aligned} \quad (75)$$

We next present the bounds on $\mathbb{P}[A(K_a, K'_a)]$, $\mathbb{P}[|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}]]$, and

$$\mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', |\widehat{\mathcal{W}}| \in [\overline{K'_a} : \overline{K'_a}]].$$

1) *Bound on $\mathbb{P}[A(K_a, K'_a)]$:* We have

$$\mathbb{P}[A(K_a, K'_a)] = \mathbb{P}[m(\mathbf{y}, K'_a) < m(\mathbf{y}, K), \forall K \neq K'_a] \quad (76)$$

$$\leq \min_{K: K \neq K'_a} \mathbb{P}[m(\mathbf{y}, K'_a) < m(\mathbf{y}, K)] \quad (77)$$

$$= \xi(K_a, K'_a), \quad (78)$$

where $\xi(K_a, K'_a)$ is given by (24), and (78) holds since under the new measure, $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, (1 + K_a P') \mathbf{I}_n)$ distribution.

2) *Bounds of $\mathbb{P}[|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]$* : As in Appendix A-B, we follow two approaches to bound $\mathbb{P}[|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]]$. The first approach is based on error exponent analyses and the second approach is based on the DT bound. In the first approach, we write the event $\{|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]\}$ as the union of the pairwise events and obtain

$$\begin{aligned} & \mathbb{P}[|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]] \\ &= \mathbb{P}\left(\bigcup_{t' \in \overline{\mathcal{T}}_t} \bigcup_{\substack{\mathcal{W}_{02} \subset [(\underline{K}_a - \overline{K}'_a)^+ + 1 : K_a] \\ |\mathcal{W}_{02}| = t}} \bigcup_{\substack{\mathcal{W}'_{02} \subset [\max\{K_a, \underline{K}'_a\} + 1 : M] \\ |\mathcal{W}'_{02}| = t'}} F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_{01}, \mathcal{W}'_{02})\right). \end{aligned} \quad (79)$$

Then, by applying the Chernoff bound, Gallager's ρ -trick, and Lemma 4 following similar steps as in Appendix A-B1, we obtain

$$\mathbb{P}[|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]] \leq p_t \quad (80)$$

with p_t given by (8). In the second approach, we consider the channel law $P_{\mathbf{y}|c(\mathcal{W}_0), c(\mathcal{W} \setminus \mathcal{W}_0)}$ with input $c(\mathcal{W}_0)$ and output \mathbf{y} where $|\mathcal{W}_{02}| = t$. The corresponding information density $\iota_t(c(\mathcal{W}_0); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0))$ is defined in (26). Notice that the event $F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_{01}, \mathcal{W}'_{02})$ defined in (69) is equivalent to $\{\iota_t(c(\mathcal{W}'_{01}) + c(\mathcal{W}_{02}); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0)) > \iota_t(c(\mathcal{W}'_{01}) + c(\mathcal{W}_{02}); \mathbf{y} | c(\mathcal{W} \setminus \mathcal{W}_0))\}$. Then, by proceeding as in Appendix A-B2, it follows that

$$\mathbb{P}[|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]] \leq q_t \quad (81)$$

with q_t given by (18).

3) *Bounds of $\mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]]$* : First, we have that

$$\begin{aligned} & \mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]] \\ &= \mathbb{P}\left[\bigcup_{\substack{\mathcal{W}_{02} \subset [(\underline{K}_a - \overline{K}'_a)^+ + 1 : K_a] \\ |\mathcal{W}_{02}| = t}} \bigcup_{\substack{\mathcal{W}'_{02} \subset [\max\{K_a, \underline{K}'_a\} + 1 : M] \\ |\mathcal{W}'_{02}| = t'}} F(\mathcal{W}_{01}, \mathcal{W}_{02}, \mathcal{W}'_{01}, \mathcal{W}'_{02})\right]. \end{aligned} \quad (82)$$

Notice that $\mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]]$ differs from $\mathbb{P}[|\mathcal{W}_{02}| = t, |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]]$ in (79) only in the absence of the union $\bigcup_{t' \in \overline{\mathcal{T}}_t}$. By applying the Chernoff bound, Gallager's ρ -trick, and Lemma 4 following similar steps as in Appendix A-B1, we obtain that

$$\mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]] \leq p_{t,t'} \quad (83)$$

with $p_{t,t'}$ given by (9). Alternatively, bounding $\mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]]$ using the information density's property as in Appendix A-B2, we obtain

$$\mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', |\widehat{\mathcal{W}}| \in [\underline{K}'_a : \overline{K}'_a]] \leq q_{t,t'} \quad (84)$$

with $p_{t,t'}$ given by (19).

From (74), (78), (80), and (81), we obtain that

$$\mathbb{P}[|\mathcal{W}_{02}| = t, K_a \rightarrow K'_a] \leq \min\{p_t, q_t, \xi(K_a, K'_a)\}. \quad (85)$$

From (75), (78), (83), and (84), we obtain that

$$\begin{aligned} & \mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', K_a \rightarrow K'_a] \\ & \leq \min\{p_{t,t'}, q_{t,t'}, \xi(K_a, K'_a)\}. \end{aligned} \quad (86)$$

Substituting these bounds on $\mathbb{P}[|\mathcal{W}_{02}| = t, K_a \rightarrow K'_a]$ and $\mathbb{P}[|\mathcal{W}_{02}| = t, |\mathcal{W}'_{02}| = t', K_a \rightarrow K'_a]$ into (70) and (71), we deduce that the MD and FA probabilities, averaged over the Gaussian codebook ensemble, are upper-bounded by ϵ_{MD} and ϵ_{FA} given in (5) and (6), respectively. Finally, proceeding as in [20, Th. 19], one can show that there exists a randomized coding strategy that achieves (5) and (6) and involves time-sharing among three deterministic codes.

APPENDIX B PROOF OF PROPOSITION 1

The probability density function (PDF) of \mathbf{y}_0 is given by

$$p_{\mathbf{y}_0}(\mathbf{y}_0) = \frac{1}{\pi^n (1 + K_a P')^n} \exp\left(-\frac{\|\mathbf{y}_0\|^2}{1 + K_a P'}\right). \quad (87)$$

Therefore, with the ML estimation of K_a , we have that

$$m(\mathbf{y}_0, K) = -\ln p_{\mathbf{y}_0}(\mathbf{y}_0) \quad (88)$$

$$= \frac{\|\mathbf{y}_0\|^2}{1 + K P'} + n \ln(1 + K P') + n \ln \pi. \quad (89)$$

As a consequence, the event $m(\mathbf{y}_0, K'_a) < m(\mathbf{y}_0, K)$ can be written as $\frac{\|\mathbf{y}_0\|^2}{1 + K'_a P'} + n \ln(1 + K'_a P') < \frac{\|\mathbf{y}_0\|^2}{1 + K P'} + n \ln(1 + K P')$, or equivalently,

$$\|\mathbf{y}_0\|^2 \left(\frac{1}{1 + K'_a P'} - \frac{1}{1 + K P'} \right) < n \ln \left(\frac{1 + K P'}{1 + K'_a P'} \right). \quad (90)$$

Using the fact that $\|\mathbf{y}_0\|^2$ follows a Gamma distribution with shape n and scale $1 + K_a P'$, we deduce that $\xi(K_a, K'_a)$ is given by (33) with $\zeta(K, K_a, K'_a)$ given by (34).

For the energy-based estimation with $m(\mathbf{y}, K) = \|\mathbf{y}\|^2 - n(1 + K P')$, after some manipulations, the event $m(\mathbf{y}_0, K'_a) < m(\mathbf{y}_0, K)$ is equivalent to

$$\begin{cases} \|\mathbf{y}_0\|^2 > n \left(1 + \frac{K_a + K'_a}{2} P' \right), & \text{if } K'_a < K_a, \\ \|\mathbf{y}_0\|^2 < n \left(1 + \frac{K_a + K'_a}{2} P' \right), & \text{if } K'_a > K_a. \end{cases} \quad (91)$$

Thus, from the Gamma distribution of $\|\mathbf{y}_0\|^2$, we deduce that $\xi(K_a, K'_a)$ is given by (33) with $\zeta(K, K_a, K'_a)$ given by (35).

APPENDIX C
ERROR FLOOR ANALYSIS

For the decoder considered in Theorem 1, the initial MDs and FAs are unavoidable. On the other hand, the additional MDs and FAs can be reduced as the power P increases. As $P \rightarrow \infty$, by assuming that no additional MD or FA occurs on top of these initial MDs or FAs, we obtain lower bounds on ϵ_{MD} and ϵ_{FA} as follows.

Proposition 2 (Asymptotic lower bounds on ϵ_{MD} and ϵ_{FA}): With ML or energy-based estimation of K_a , it holds that

$$\begin{aligned} & \lim_{P \rightarrow \infty} \epsilon_{\text{MD}} \\ & \geq \bar{\epsilon}_{\text{MD}} \\ & = \sum_{K_a = \max\{K_\ell, 1\}}^{K_u} \left(P_{K_a}(K_a) \sum_{K'_a = K_\ell}^{K_u} \frac{(K_a - \bar{K}'_a)^+}{K_a} \xi(K_a, K'_a) \right) + \bar{p}, \end{aligned} \quad (92)$$

$$\begin{aligned} & \lim_{P \rightarrow \infty} \epsilon_{\text{FA}} \\ & \geq \bar{\epsilon}_{\text{FA}} \\ & = \sum_{K_a = K_\ell}^{K_u} \left(P_{K_a}(K_a) \right. \\ & \quad \cdot \left. \sum_{K'_a = K_\ell}^{K_u} \frac{(\underline{K}_a - K_a)^+}{K_a - (K_a - \bar{K}'_a)^+ + (\underline{K}'_a - K_a)^+} \xi(K_a, K'_a) \right) + \bar{p}, \end{aligned} \quad (93)$$

where $\bar{p} = 2 - \sum_{K_a = K_\ell}^{K_u} P_{K_a}(K_a) - \mathbb{E}_{K_a} \left[\frac{M!}{M^{K_a}(M-K_a)!} \right]$, and $\xi(K_a, K'_a)$ is given by (33) with $\zeta(K, K_a, K'_a) = n \ln \left(\frac{K}{K_a} \right) K_a^{-1} \left(\frac{1}{K'} - \frac{1}{K} \right)^{-1}$ for ML estimation of K_a and $\zeta(K, K_a, K'_a) = n \frac{K+K'_a}{2K_a}$ for energy-based estimation of K_a .

Proof: First, the optimal value of P' minimizing the bounds must grow with P since otherwise \tilde{p} will be large. Therefore, as $P \rightarrow \infty$, we can assume without loss of optimality that $P' \rightarrow \infty$. Next, when $t = t' = 0$, we can verify that $a = b = 0$, thus $E_0(\rho, \rho_1) = 0$ and $E(0, 0) = 0$, achieved with $\rho = \rho_1 = 0$. Therefore, $p_0 = p_{0,0} = e^{-n \cdot 0} = 1$. We can also verify that q_0 and $q_{0,0}$ both converge to 1 as $P' \rightarrow \infty$. When $P' \rightarrow \infty$, $\xi(K_a, K'_a)$ given in Proposition 1 converges to the right-hand side of (33) with $\zeta(K, K_a, K'_a) = n \ln \left(\frac{K}{K_a} \right) K_a^{-1} \left(\frac{1}{K'} - \frac{1}{K} \right)^{-1}$ for ML estimation of K_a and $\zeta(K, K_a, K'_a) = n \frac{K+K'_a}{2K_a}$ for energy-based estimation of K_a . Furthermore, the last term in \tilde{p} given by (7) vanishes and thus $\tilde{p} \rightarrow \bar{p}$. Finally, the lower bounds $\bar{\epsilon}_{\text{MD}}$ and $\bar{\epsilon}_{\text{FA}}$ follows by substituting the asymptotic values of p_0 , q_0 , $p_{0,0}$, $q_{0,0}$, $\xi(K_a, K'_a)$, and \tilde{p} computed above into ϵ_{MD} and ϵ_{FA} , and by setting $\min\{p_t, q_t\}$ to zero for $t \neq 0$, and setting $\min\{p_{t,t'}, q_{t,t'}\}$ to zero for $(t, t') \neq (0, 0)$. ■

We remark that the lower bounds in (92) and (93) are tight for typical IoT settings. Indeed, equalities in (92) and (93) hold if the probability of having additional MDs and FAs vanishes, i.e., $\min\{p_t, q_t\} \rightarrow 0$ for $t \neq 0$ and $\min\{p_{t,t'}, q_{t,t'}\} \rightarrow 0$ for $(t, t') \neq (0, 0)$ as $P \rightarrow \infty$. With $\rho = \rho_1 = 1$, the optimal

λ in (11) is given by $\lambda = 1/(2P_2)$. Thus, by replacing the maximization over ρ and ρ_1 in (10) with $\rho = \rho_1 = 1$, we obtain that $E(t, t') \geq -t'R_1 - R_2 + \ln \left(1 + \frac{(t+t')P'}{4P_2} \right)$. It follows that

$$\begin{aligned} p_{t,t'} & \leq \binom{M - \max\{K_a, \underline{K}'_a\}}{t'} \binom{\min\{K_a, \bar{K}'_a\}}{t} \\ & \cdot \left(1 + \frac{(t+t')P'}{4P_2} \right)^{-n}. \end{aligned} \quad (94)$$

If $K_a \in [\underline{K}'_a, \bar{K}'_a]$, i.e., $P_2 = 1$, the right-hand side of (94) vanishes as $P' \rightarrow \infty$. Otherwise, the right-hand side of (94) converges to

$$\begin{aligned} \bar{p}_{t,t'} & = \binom{M - \max\{K_a, \underline{K}'_a\}}{t'} \binom{\min\{K_a, \bar{K}'_a\}}{t} \\ & \cdot \left(1 + \frac{t+t'}{4((K_a - \bar{K}'_a)^+ + (\underline{K}'_a - K_a)^+)} \right)^{-n} \\ & \leq M^{t'} K_a^t \left(1 + \frac{t+t'}{4((K_a - \bar{K}'_a)^+ + (\underline{K}'_a - K_a)^+)} \right)^{-n}. \end{aligned} \quad (95)$$

Observe that $\bar{p}_{t,t'}$ is small if n is relatively large compared to $\ln M$ and $\ln K_a$, which is true for relevant values of n , M and K_a in the IoT. Specifically, in typical IoT scenarios, M and K_a are in the order of 10^2 , while K_a/n is from 10^{-4} to 10^{-3} —see [3] and [25, Rem. 3]. For example, with $(M, n) = (2^{100}, 15000)$ and $K_a \leq 300$ as considered in [3] and many follow-up works, assume that $(K_a - \bar{K}'_a)^+ + (\underline{K}'_a - K_a)^+ \leq 20$, then $\bar{p}_{t,t'} < 10^{-128}$ for every $t \leq 300$ and $t' \leq 300$. As a consequence, $p_{t,t'}$ and $q_{t,t'}$ are very small. We conclude that $\lim_{P \rightarrow \infty} \epsilon_{\text{MD}}$ and $\lim_{P \rightarrow \infty} \epsilon_{\text{FA}}$ approach closely $\bar{\epsilon}_{\text{MD}}$ and $\bar{\epsilon}_{\text{FA}}$, respectively. In other words, $\bar{\epsilon}_{\text{MD}}$ and $\bar{\epsilon}_{\text{FA}}$ essentially characterize the error floors of ϵ_{MD} and ϵ_{FA} , respectively, as $P \rightarrow \infty$.

REFERENCES

- [1] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.
- [2] Y. Wu, X. Gao, S. Zhou, W. Yang, Y. Polyanskiy, and G. Caire, "Massive access for future wireless communication systems," *IEEE Wireless Commun. Mag.*, vol. 27, no. 4, pp. 148–156, Oct. 2020.
- [3] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2523–2527.
- [4] M. Berlioli, G. Cocco, G. Liva, and A. Munari, "Modern random access protocols," *Foundations and Trends in Networking*, vol. 10, no. 4, pp. 317–446, Nov. 2016.
- [5] O. Ordentlich and Y. Polyanskiy, "Low complexity schemes for the random access Gaussian channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2528–2532.
- [6] A. Vem, K. R. Narayanan, J. Chamberland, and J. Cheng, "A user-independent successive interference cancellation based coding scheme for the unsourced random access Gaussian channel," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8258–8272, Sep. 2019.
- [7] A. Fengler, P. Jung, and G. Caire, "SPARCs for unsourced random access," *IEEE Trans. Inf. Theory*, vol. 67, no. 10, pp. 6894–6915, Oct. 2021.
- [8] V. K. Amalladinne, A. K. Pradhan, C. Rush, J.-F. Chamberland, and K. R. Narayanan, "Unsourced random access with coded compressed sensing: Integrating AMP and belief propagation," *arXiv preprint arXiv:2010.04364*, Oct. 2020.

- [9] V. K. Amalladinne, J. F. Chamberland, and K. R. Narayanan, "A coded compressed sensing scheme for unsourced multiple access," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6509–6533, Jul. 2020.
- [10] A. K. Pradhan, V. K. Amalladinne, K. R. Narayanan, and J.-F. Chamberland, "Polar coding and random spreading for unsourced multiple access," in *Proc. IEEE Int. Conf. Communications (ICC)*, Dublin, Ireland, Jun. 2020.
- [11] S. S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy efficient coded random access for the wireless uplink," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4694–4708, Jun. 2020.
- [12] A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire, "Non-Bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive MIMO receiver," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, pp. 2925–2951, May 2021.
- [13] K. Stern, A. E. Kalør, B. Soret, and P. Popovski, "Massive random access with common alarm messages," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019.
- [14] M. Effros, V. Kostina, and R. C. Yavas, "Random access channel coding in the finite blocklength regime," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, Colorado, USA, Jun. 2018, pp. 1261–1265.
- [15] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted ALOHA," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 477–487, Dec. 2011.
- [16] A. Decurninge, I. Land, and M. Guillaud, "Tensor-based modulation for unsourced massive random access," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 552–556, Mar. 2021.
- [17] A. Fengler, P. Jung, and G. Caire, "Pilot-based unsourced random access with a massive MIMO receiver, MRC and polar codes," *arXiv preprint arXiv:2012.03277*, 2020.
- [18] R. G. Gallager, *Information theory and reliable communication*. New York, USA: Wiley, 1968.
- [19] Y. Polyanskiy and Y. Wu, "Lecture notes on information theory," *Massachusetts Institute of Technology (6.441), University of Illinois Urbana-Champaign (ECE563), and Yale University (STAT 664)*, 2012–2017. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/itlectures_v5.pdf
- [20] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Feedback in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4903–4925, Jul. 2011.
- [21] Y. Ohnishi and J. Honorio, "Novel change of measure inequalities with applications to PAC-Bayesian bounds and Monte Carlo estimation," *arXiv preprint arXiv:2002.10678*, Feb. 2020.
- [22] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*, 4th ed. Boston, MA, USA: Pearson, 2012.
- [23] S. S. Kowshik and Y. Polyanskiy, "Fundamental limits of many-user MAC with finite payloads and fading," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, pp. 5853–5884, Sep. 2021.
- [24] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, Apr. 2010.
- [25] I. Zadik, Y. Polyanskiy, and C. Thrampoulidis, "Improved bounds on Gaussian MAC and sparse regression via Gaussian inequalities," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 430–434.