Energy Efficiency of Massive Random Access in MIMO Quasi-Static Rayleigh Fading Channels With Finite Blocklength

Junyuan Gao, Yongpeng Wu, *Senior Member, IEEE*, Shuo Shao, *Member, IEEE*, Wei Yang, *Member, IEEE*, and H. Vincent Poor, *Life Fellow, IEEE*

***Abstract*—This paper considers the massive random access problem in multiple-input multiple-output (MIMO) quasi-static Rayleigh fading channels. Specifically, we derive achievability and converse bounds on the minimum energy-per-bit required for each active user to transmit J bits with blocklength n, power P, and L receive antennas under a per-user probability of error (PUPE) constraint, in the cases with and without *a priori* channel state information at the receiver (CSIR and no-CSI). In the case of no-CSI, we consider both the settings with and without the knowledge of the number K_a of active users at the receiver. Numerical evaluation shows that the gap between achievability and converse bounds is less than 2.5 dB for the CSIR case and less than 4 dB for the no-CSI case in most considered regimes. Under the condition that the distribution of K_a is known in advance, the uncertainty of the exact value of K_a entails only a small penalty in terms of energy efficiency. Our results show the significance of MIMO for the massive random access problem. As an example, we show that the spectral efficiency grows approximately linearly with the number of receive antennas in the case of CSIR, whereas the growth rate decreases in the case of no-CSI. Moreover, in the case of no-CSI, we demonstrate the suboptimality of the pilot-assisted scheme, especially when the number of active users is large. Building on non-asymptotic results, assuming all users are active and $J = \Theta(1)$, we obtain scaling laws of the number of supported users as follows: when $L = \Theta\left(n^2\right)$ and $P = \Theta\left(\frac{1}{n^2}\right)$, one can reliably serve $K = \mathcal{O}(n^2)$ users in the case of no-CSI; under mild conditions in the case of CSIR, the PUPE requirement is satisfied if and only if $\frac{nL\ln KP}{K} = \Omega\left(1\right)$.**

***Index Terms*—Energy efficiency, finite blocklength, massive random access, MIMO, scaling law.**

I. Introduction

THE design of uplink communication systems in many contemporary wireless networks is influenced by four issues: the rapidly expanding number of users with random activity patterns; the relatively small quantity of information bits to transmit; the strict requirement in communication latency; and the stringent demand on communication energy efficiency. Notably, these issues are present in many Internet-of-Things (IoT) applications, in which a very large number of sensors are deployed, but only a fraction of them are active at any given time. Active sensors often transmit hundreds of bits describing the parameters they have sensed to the base station (BS) within latency and energy constraints. To address these issues, massive random access technologies have been proposed recently, the study of which includes the information-theoretic analysis and the development of transmission and reception strategies for massive numbers of users with sporadic activity patterns in the regime of finite blocklength.

A. Previous Work

Some subsets of these issues have been discussed in recent years. The classical multiuser information theory in [1], [2], and [3] studied the fundamental limits of the conventional multiple access channel (MAC), where the number of users is fixed and the blocklength is taken to infinity. To characterize the massive user population in IoT applications, a new model called the many-access channel (MnAC) was proposed in [4], which allows the number of users to grow unboundedly with the blocklength. Based on this model, a new notion of capacity was introduced and characterized with random user activity [4]. Since the publication of [4], MnACs have been studied in various works with different settings, where a common assumption is that the number of users grows linearly and unboundedly with the blocklength [5], [6], [7], [8], [9]. However, the work in [4] relies on the assumption of infinite

Manuscript received 20 December 2021; revised 14 September 2022; accepted 20 October 2022. Date of publication 7 November 2022; date of current version 16 February 2023. The work of Yongpeng Wu was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1801102, in part by the National Science Foundation of China (NSFC) under Grant 62122052 and Grant 62071289, in part by the 111 Project under Grant BP0719010, and in part by the Science and Technology Commission of Shanghai Municipality (STCSM) under Grant 18DZ2270700. The work of Shuo Shao was supported in part by the NSFC under Grant 61872149, Grant 61901261, and Grant 12031011. The work of H. Vincent Poor was supported in part by the U.S. National Science Foundation under Grant CCF-1908308. *(Corresponding author: Yongpeng Wu.)*

Junyuan Gao and Yongpeng Wu are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Minhang 200240, China (e-mail: sunflower0515@sjtu.edu.cn; yongpeng.wu@sjtu.edu.cn).

Shuo Shao is with the School of Cyber Security, Shanghai Jiao Tong University, Minhang 200240, China (e-mail: shuoshao@sjtu.edu.cn).

Wei Yang is with Qualcomm Technologies, Inc., San Diego, CA 92121 USA (e-mail: weiyang@qti.qualcomm.com).

H. Vincent Poor is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Communicated by C. Murthy, Associate Editor for Communications.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2022.3220261.

Digital Object Identifier 10.1109/TIT.2022.3220261

payload size and blocklength, which cannot capture stringent energy requirements in massive access systems.

In addition to the massive user population, finite payload size and even finite blocklength should be taken into consideration to make the setting more relevant in practice. On this topic, Polyanskiy introduced the per-user probability of error (PUPE) criterion to measure the fraction of transmitted messages that are missing from the list of decoded messages, instead of utilizing the traditional joint error probability criterion, which results in another crucial departure from the classical MAC model [6].

Under the PUPE criterion, some works considered the regime with finite payload size, finite energy-per-bit, and infinite blocklength [7], [8], [9]. In particular, based on the MnAC model with the linear scaling mentioned above, under the assumptions of individual codebooks[1] and a single BS antenna, Zadik et al. [7] and Kowshik and Polyanskiy [8] presented bounds on the tradeoff between user density and energy-per-bit for reliable transmission in additive white Gaussian noise (AWGN) channels and quasi-static fading channels, respectively. In both models, it was observed that in the low user density regime, the multi-user interference (MUI) can be almost perfectly canceled with good coded access schemes.

Finite blocklength considerations have also been studied to address transmission within latency constraints. For point-to-point channels, Polyanskiy et al. [11] developed a tight approximation to the maximal achievable rate for various channels with positive Shannon capacity, and this approximation was extended to quasi-static fading channels by Yang et al. [12]. For the K-user Gaussian MAC, achievability bounds and normal approximations with a joint error probability criterion were studied in [13]. Yavas et al. [14], [15] improved the achievable third-order term in [13] for Gaussian MAC, and extended this result to Gaussian random access channels under the assumption that the number K of users does not grow with the blocklength n. For the massive random access problem with finite blocklength, the works in [6] and [16] derived non-asymptotic bounds for Gaussian and Rayleigh fading channels, respectively, under the PUPE criterion and the assumption that the number K_a of active users is known *a priori*. It was pointed out in [17] that the number K_a of active users can be detected with high success probability in Rayleigh fading channels when both uplink and downlink transmissions are exploited to mitigate fading uncertainty, which supports the assumption of known K_a in [6] and [16].

When only the uplink transmission is utilized, the success probability of detecting K_a can be reduced [17]. The performance penalty in AWGN channels, suffering from the lack of knowledge of K_a, was analysed in [4], [18], and [19]. Specifically, in the asymptotic regime with infinite number of users, it was pointed out in [4] that the message-length capacity penalty due to unknown user activity on each of the K_a active users is $H_2(p_a)/p_a$ under the joint error probability criterion and the assumption that each user becomes active independently with probability p_a. Moreover, in [18], Lancho et al. derived non-asymptotic achievability and converse bounds for the single-user random access scenario, and numerical results for the binary-input Gaussian channel indicated that the bound with unknown user activity approaches the one with known K_a as the blocklength and signal-to-noise ratio (SNR) increase. Following from the maximum likelihood (ML) principle, a non-asymptotic achievability bound was derived in [19] for massive random access in Gaussian channels with unknown K_a, whereas a matching converse bound was not provided. As a result, it is significant to construct tight non-asymptotic bounds in both achievability and converse sides to characterize the performance loss caused by unknown K_a for massive random access. This problem will be addressed in this paper.

It should be noted that the above-mentioned non-asymptotic works on the massive random access communication problem [6], [16], [18], [19] rely on the assumption of a single BS antenna. In practice, equipping multiple antennas at the BS can bring great benefits in massive random access systems. Specifically, for the user activity detection problem, it was demonstrated in [20] that, with n channel uses and a sufficiently large number L of BS antennas satisfying $K_a/L = o(1)$, up to $K_a = \mathcal{O}\left(n^2\right)$ active users can be identified among K potential users when $\frac{K_a}{K} = \Theta(1)$; it overcomes the fundamental limitation of the single-receive-antenna system, in which the number K_a of active users that can be identified is at most linear with the blocklength n. Given the great potential of multiple receive antennas for the activity detection problem as revealed by the scaling law in [20], it is natural to expect that multiple receive antennas could bring similar benefits for the joint activity and data detection problem in massive random access channels. An important goal of this paper is to characterize the impact of multiple BS antennas on the performance of joint activity and data detection in both the non-asymptotic regime and the asymptotic regime.

From the perspective of channel state information (CSI) availability, the above mentioned works can be divided into two categories: the case in which CSI is known at the receiver in advance (CSIR) [6], [7], [8], [9], [13], [14], [15], [19] (the AWGN channel without fading is a special case of CSIR), and the case in which there is no *a priori* CSI at the receiver (no-CSI) [8], [12], [20]. In the no-CSI case (i.e. the so called noncoherent setting), the communication scheme suggested by the capacity result makes no effort to estimate channel coefficients [21]. Thus, the scheme without explicit channel estimation is adopted in many works, such as [8], [12], and [20]. In the no-CSI case, the receiver is also allowed to gain channel knowledge, where channel estimation can be simply viewed as a specific form of coding [22], [23], [24]. In practical wireless systems, pilot-assisted schemes are widely adopted, in which users first send pilots for channel estimation, and then the estimated channels are utilized to decode the signals for each user. The performance of this scheme has been investigated in some works. In the single-user case, it was proved in [21] that the pilot-assisted scheme is optimal at a high SNR in terms of degrees of freedom for block-fading channels, and non-asymptotic bounds on the maximum coding

[1] It should be noted that individual codebook and common codebook assumptions correspond to different massive access models in practice [10]. In essence, the detection problem under these two assumptions reduces to the block sparse support recovery problem and the sparse support recovery problem, respectively.

rate with finite blocklength were derived in [25]. For the scenario with multiple users, the large-antenna limit of the pilot-assisted scheme was studied in [26], where the achievable error probability was derived at finite blocklength, assuming channels were estimated based on the minimum mean-square error (MMSE) criterion and both the MMSE and maximum ratio criteria were utilized for mismatched combining. After combining, the complicated problem of jointly detecting K transmitted codewords based on the received signals among L BS antennas, is converted to the problem of separately detecting K codewords in the single-receive-antenna fading channel, which, however, can result in a performance loss.

B. Our Contributions

In this paper, we consider the joint activity and data detection problem for massive random access in multiple-input multiple-output (MIMO) quasi-static Rayleigh fading channels. We analyze the fundamental tradeoff in this setup between the energy efficiency, latency, user density, and number of receive antennas using information-theoretic tools. Specifically, in both cases of CSIR and no-CSI, we derive achievability and converse bounds on the minimum energy-per-bit required for each active user to transmit $J = \log_2 M$ information bits with blocklength n, power P, L receive antennas, and PUPE less than a constant, under the assumption that the number K_a of active users is known *a priori*. To characterize the performance loss caused by the uncertainty of user activities in the non-asymptotic regime, we also derive achievability and converse results for the no-CSI case in the setting where K_a is random and unknown but its distribution $K_a \sim \text{Binom}(K, p_a)$ is known at the receiver in advance. Moreover, we study the performance of a pilot-assisted scheme. The derived non-asymptotic bounds provide theoretical benchmarks to evaluate practical transmission schemes. Building on these non-asymptotic bounds, we obtain scaling laws of the number of reliably served users in a special case where all users are active. These results reveal the great potential of multiple receive antennas for the massive access problem. Meanwhile, they show a significant difference in the required number of BS antennas between utilizing the PUPE criterion and the joint error probability criterion.

Non-Asymptotic Analysis: There are some twists in deriving non-asymptotic achievability bounds for massive random access in MIMO quasi-static Rayleigh fading channels. First, compared with traditional MAC, the number of users is greatly increased in massive random access channels, leading to a considerable increase in the number of error events. As a consequence, the simple union bound can be substantially loosened if not applied with care, and we need to resort to more efficient tools. Second, it is difficult to extend the bounds derived for the single-receive-antenna setting in [8] to the multiple-receive-antenna setting. Specifically, a projection decoder was used in [8] to derive the achievability bound in the no-CSI case, which measures the angle between the received signals and the subspace spanned by candidate transmitted codewords. Although the projection decoder has the advantage of not requiring fading distributions, it can be ineffectual when applied to the setting with multiple receive antennas. As an example, the use of large antenna arrays allows the number of reliably served active users to be much larger than the blocklength. In this case, the dimension of the subspace spanned by the transmitted codewords of active users is limited by the blocklength. Thus, the subspace spanned by K_a transmitted codewords can be the same as that spanned by another set of K_a codewords, which prevents the projection decoder from distinguishing them. Moreover, it is challenging (although not impossible) to jointly deal with the signals received over L BS antennas based on the projection decoder, because the analysis of the angle between the subspace spanned by the received signals and the subspace spanned by K_a transmitted codewords is quite involved.

To alleviate the problems mentioned above, for massive random access in MIMO quasi-static Rayleigh fading channels, new techniques are utilized in this paper to derive non-asymptotic achievability bounds on the minimum required energy-per-bit. Specifically, in both cases of CSIR and no-CSI, we leverage the ML-based decoder when K_a is known *a priori*. Note that, in contrast to the projection decoder mentioned above, the ML decoder is applicable regardless of whether K_a is less than the blocklength or not, but at the price of requiring *a priori* fading distribution. Moreover, when K_a is unknown in advance, we first obtain an estimate of K_a via an energy-based estimator; then, we output a set of decoded messages following the maximum *a posteriori* (MAP) principle, which incorporates prior distributions in users' messages of various sizes. For the pilot-assisted coded access scheme, in a special case where all users are active, we leverage the MMSE criterion to estimate channels in the first stage, and utilize the mismatched nearest neighbor criterion [27], [28] to decode in the second stage. The signals received over L receive antennas can be jointly dealt with in aforementioned cases.

To evaluate the probability of the union of extremely many error events, we resort to standard bounding techniques proposed by Fano [29] and Gallager [30]. Gallager's ρ-trick bound is used for a special case in which both the user activity and CSI are known at the receiver, considering that this bound is difficult to evaluate by the Monte Carlo method when random access is taken into consideration. The Fano's bound is used to establish non-asymptotic achievability bounds in massive random access channels for the case of CSIR and no-CSI. Its performance relies on the choice of a region around the linear combination of the transmitted signals, which is interpreted as the "good region" [31]. In this work, we select an appropriate "good region" for massive random access channels, which is parameterized by two parameters ω and ν. Our "good region" reduces to the one used in [8] if the parameter ν is set to 0. Using two parameters instead of one gives us great flexibility and accuracy in tuning the shape of the "good region", thereby leading to tighter achievability results.

Numerical results demonstrate the tightness of our bounds. Specifically, the gap between the achievability bound and the converse bound is less than 2.5 dB for the CSIR case and less than 4 dB for the no-CSI case in most considered regimes (the Fano type converse bound for the no-CSI case relies

TABLE I

COMPARISON OF SCALING LAWS FOR MASSIVE ACCESS IN QUASI-STATIC RAYLEIGH FADING CHANNELS

result	K	L	P	M	CSIR / no-CSI	error criterion	achievability / converse
Theorem 5	$\mathcal{O}(n^2)$	$\Theta\left(n\right)$	$\Theta\left(\frac{1}{n^2}\right)$	$\Theta(1)$	CSIR	PUPE	both
Theorem 5	$\mathcal{O}(n^2)$	$\Theta\left(\frac{n}{\ln n}\right)$	$\Theta\left(\frac{1}{n}\right)$	$\Theta(1)$	CSIR	PUPE	both
Theorem 11[1]	$\mathcal{O}(n^2)$	$\Theta\left(n^2\right)$	$\Theta\left(\frac{1}{n^2}\right)$	$\Theta(1)$	no-CSI	PUPE	both
extended from [20]	$\mathcal{O}(n^2)$	$\Theta\left(n^2 \ln n\right)$	$\Theta\left(\frac{1}{n^2}\right)$	$\Theta(1)$	no-CSI	joint error probability	achievability
[8]	$\mathcal{O}(n)$	1	$\Theta\left(\frac{1}{n}\right)$	$\Theta(1)$	both	PUPE	both
[33]	$o(n)$	1	$\Theta\left(\frac{1}{n}\right)$	$\Theta(1)$	CSIR (AWGN)	PUPE (vanish)[2]	both

[1] In the case of no-CSI, the converse bound relies on the assumption of i.i.d. Gaussian codebooks.

[2] The PUPE is required to vanish for the scaling law in [33], and a positive constant PUPE is acceptable for other cases in Table I.

on the assumption of i.i.d. Gaussian codebooks). Compared to the case where the number K_a of active users is known, the performance loss caused by unknown K_a is small. For example, in the setup with blocklength $n = 1000$, payload $J = 100$ bits, active probability $p_a = 0.4$, error requirement $\epsilon = 0.001$, and $L = 128$ receive antennas, the extra required energy-per-bit due to the uncertainty of the exact value of K_a is less than 0.3 dB on the converse side and less than 1.1 dB on the achievability side. Similar to AWGN channels [7] and single-receive-antenna quasi-static fading channels [8], the MUI can be almost perfectly cancelled in multiple-receive-antenna quasi-static fading channels when the number of active users is below a critical threshold. Additionally, in our considered regime, the spectral efficiency grows approximately linearly with the number of BS antennas for the CSIR case, but the lack of CSI at the receiver causes a slowdown in the growth rate. Furthermore, our results for the no-CSI case reveal that the orthogonal-pilot-assisted coded access scheme is suboptimal, especially when the number of active users is large, even if the power allocation between pilot and data symbols is optimized. Overall, we believe our non-asymptotic bounds provide theoretical benchmarks to evaluate practical transmission schemes, which are of considerable importance in massive random access systems.

Asymptotic Analysis: Building on these non-asymptotic results, in a special case where all users are assumed to be active, we obtain scaling laws of the number of reliably served users under the PUPE criterion. For the CSIR case, assuming $n, K \to \infty$, $M = \Theta(1)$, $\ln K = o(n)$, and $KP = \Omega\left(1\right)$ (P denotes the transmitting power per channel use), the PUPE requirement is satisfied if and only if $\frac{nL \ln KP}{K} = \Omega\left(1\right)$. It can be divided into the following two regimes: 1) $\frac{nL}{K} = \Omega\left(1\right)$ and $KP = \Theta\left(1\right)$; 2) $\frac{nL \ln KP}{K} = \Omega\left(1\right)$ and $KP \to \infty$. The first regime is power-limited, where the number of degrees of freedom grows linearly with the number of users. As a result, by allocating orthogonal resources to users, the minimum received energy-per-user can be $nLP = \Theta\left(1\right)$, which is as low as that in the single-user case [32]. The second regime is degrees-of-freedom-limited, where the number of degrees of freedom, i.e. nL, is far less than the number of users, and the minimum received energy-per-user $nLP \to \infty$. Two special scaling laws in the CSIR case are presented in Table I. We can observe that, in order to reliably serve $K = \mathcal{O}(n^2)$ users, when the number of BS antennas is increased from $L = \Theta\left(\frac{n}{\ln n}\right)$ to $L = \Theta\left(n\right)$, the minimum required power can be considerably decreased from $P = \Theta\left(\frac{1}{n}\right)$ to $P = \Theta\left(\frac{1}{n^2}\right)$, which indicates the great potential of multiple receive antennas for the data detection problem. Moreover, our scaling laws reveal the tightness of the derived bounds in asymptotic cases since they are proved from both the achievability side and the converse side. The scaling law for the scenario with a single BS antenna is also presented in Table I for comparison: one can reliably serve $K = \mathcal{O}(n)$ users when a positive constant PUPE is acceptable [8]; however, the number of users is only allowed to grow sublinearly with n even in AWGN channels when the PUPE is required to vanish [33].

For the no-CSI case, the scaling law from our result is shown in Table I, together with the result extended from [20], which is based on the joint error probability criterion. We observe from Table I a significant difference in the number of receive antennas to reliably serve K users between utilizing the PUPE criterion and the joint error probability criterion. Specifically, in order to obtain the scaling law on the achievability side, both the activity detection problem considered in [20] and the data detection problem of interest in this work can be formulated as similar sparse support recovery problems. Thus, the scaling law of the activity detection problem in [20] can be extended to that of the data detection problem as follows: under the joint error probability criterion, with a coherence block of dimension $n \to \infty$ and a sufficient number of BS antennas $L = \Theta\left(n^2 \ln n\right)$, one can reliably serve up to $K = \mathcal{O}\left(n^2\right)$ users when the payload $J = \Theta(1)$ and the power $P = \Theta\left(\frac{1}{n^2}\right)$ in the case of no-CSI. In this work, we consider the PUPE criterion, which is more appropriate for the consideration of massive access [6]. Our result shows that the required number of receive antennas can be reduced from $L = \Theta\left(n^2 \ln n\right)$ to $L = \Theta\left(n^2\right)$ when we change from the joint error probability criterion to the PUPE criterion. In addition, it should be noted that, the case of $nP = \Theta\left(1\right)$ and the case of $n^2 P = \Theta\left(1\right)$ in Table I imply that the energy-per-bit is finite and goes to 0, respectively, which are crucial in practical communication systems with stringent energy constraints.

The remainder of this paper is organized as follows. Section II introduces the system model. In Section III, we provide our main results, including achievability and converse bounds in both cases of CSIR and no-CSI, and corresponding scaling laws. Section IV presents numerical results. Conclusions are drawn in Section V. Most proofs are included in the appendices.

Notation: Throughout this paper, uppercase and lowercase boldface letters denote matrices and column vectors, respectively. We use $[\mathbf{x}]_m$ to denote the m-th element of a vector $\mathbf{x}$, and use $[\mathbf{A}]_{m,n}$, $[\mathbf{A}]_{m,:}$, and $[\mathbf{A}]_{:,n}$ to denote the (m,n)-th element, the m-th row vector, and the n-th column vector of a matrix $\mathbf{A}$, respectively. The notation $\mathbf{I}_n$ denotes an $n \times n$ identity matrix, and $\mathbf{I}_{(t)} \in \{0,1\}^{n\times n}$ denotes a diagonal matrix with the first $t \leq n$ diagonal entries being ones and all of the rest being 0. We use $(\cdot)^T$, $(\cdot)^H$, $\mathrm{vec}\,(\mathbf{X})$, $|\mathbf{X}|$, $\|\mathbf{x}\|_p$, and $\|\mathbf{X}\|_F$ to denote transpose, conjugate transpose, vectorization of a matrix $\mathbf{X}$, determinant of a matrix $\mathbf{X}$, ℓ_p-norm of a vector $\mathbf{x}$, and Frobenius norm of a matrix $\mathbf{X}$, respectively. The notations $\lceil\cdot\rceil$ and $k!$ depict the ceiling function and factorial function, respectively. Given any complex variable, vector or matrix, the notations $\Re\,(\cdot)$ and $\Im\,(\cdot)$ return its real and imaginary parts, respectively. We use $\mathrm{diag}\,\{\mathbf{x}\}$ to denote a diagonal matrix with vector $\mathbf{x}$ comprising its diagonal elements, and use $\mathrm{diag}\,\{\mathbf{A},\mathbf{B}\}$ to denote a block diagonal matrix with $\mathbf{A}$ and $\mathbf{B}$ in diagonal blocks. We use $\cdot\backslash\cdot$ and $|\mathcal{A}|$ to denote set subtraction and the cardinality of a set $\mathcal{A}$, respectively. We use $\mathbf{b}_{[\mathcal{A}]} = \{\mathbf{b}_i : i \in \mathcal{A}\}$ to denote a set of vectors. We denote the set of nonnegative natural numbers by $\mathbb{N}_+$. For an integer $k > 0$, the notation $[k]$ denotes $\{1,2,\ldots,k\}$; for integers $k_2 \geq k_1 > 0$, the notation $[k_1 : k_2]$ denotes $\{k_1, k_1+1, \ldots, k_2\}$. We denote $x^+ = \max\{x,0\}$. We denote the projection matrix onto the subspace spanned by $S \subset \mathbb{C}^n$ and its orthogonal complement as $\mathcal{P}_S$ and $\mathcal{P}_S^{\perp}$, respectively. The notation $\mathcal{G}^c$ denotes the complement of the event $\mathcal{G}$. We use $\mathcal{N}(\cdot,\cdot)$, $\mathcal{CN}(\cdot,\cdot)$, $\chi^2(d)$, $\chi^2(d,\lambda)$, and $\mathcal{W}_m(n,\mathbf{A})$ to denote the standard Gaussian distribution, circularly symmetric complex Gaussian distribution, central chi-squared distribution with d degrees of freedom, noncentral chi-squared distribution with d degrees of freedom and noncentrality parameter λ, and Wishart distribution with n degrees of freedom and covariance matrix $\mathbf{A}$ of size $m \times m$, respectively. The functions $\gamma\,(\cdot,\cdot)$ and $\Gamma\,(\cdot)$ denote the lower incomplete gamma function and gamma function, respectively, with the assumption that $\gamma\,(\cdot,a) = 0$ if $a \leq 0$. For $0 \leq p \leq 1$, we denote $h(p) = -p\ln(p) - (1-p)\ln(1-p)$ and $h_2(p) = h(p)/\ln 2$ with $0\ln 0$ defined to be 0. Let $f(x)$ and $g(x)$ be positive. The notation $f(x) = o\,(g(x))$ means that $\lim_{x\to\infty} f(x)/g(x) = 0$, $f(x) = \mathcal{O}\,(g(x))$ means that $\limsup_{x\to\infty} f(x)/g(x) < \infty$, $f(x) = \Theta\,(g(x))$ means that $f(x) = \mathcal{O}\,(g(x))$ and $g(x) = \mathcal{O}\,(f(x))$, and $f(x) = \Omega\,(g(x))$ means that $g(x) = \mathcal{O}\,(f(x))$.

II. System Model

We consider a massive random access system consisting of a BS equipped with L receive antennas and K potential users each equipped with a single transmit antenna. We assume that the user traffic is sporadic, i.e., only $K_a \leq K$ users are active at any given time. Each active user transmits J information bits with blocklength n. The user set and active user set are denoted as $\mathcal{K}$ and $\mathcal{K}_a$, respectively.

We assume each user has an individual codebook of size $M = 2^J$ and blocklength n. The matrix $\mathbf{X}_k = [\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \ldots, \mathbf{x}_{k,M}] \in \mathbb{C}^{n\times M}$ consists of the codewords of the k-th user and the matrix $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_K] \in \mathbb{C}^{n\times MK}$ is obtained by concatenating all codebooks.

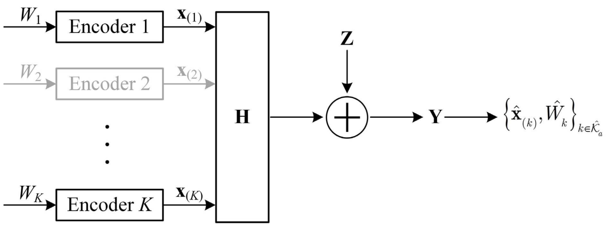


Fig. 1. Massive random access in MIMO quasi-static Rayleigh fading channels.

We consider a quasi-static Rayleigh fading channel model, where the channel stays constant during the transmission of a codeword. We assume synchronous transmission. The l-th antenna of the BS observes $\mathbf{y}_l \in \mathbb{C}^n$ given by

$$\mathbf{y}_l = \sum_{k\in\mathcal{K}} h_{k,l}\mathbf{x}_{(k)} + \mathbf{z}_l, \tag{1}$$

where $h_{k,l} \sim \mathcal{CN}(0,1)$ denotes the fading coefficient between the k-th user and the l-th antenna of the BS, which is i.i.d. across different users and different BS antennas; the noise vector $\mathbf{z}_l$ is distributed as $\mathcal{CN}(\mathbf{0},\mathbf{I}_n)$, which is i.i.d. across L BS antennas; the transmitted codeword of the k-th user is denoted as $\mathbf{x}_{(k)} = \mathbf{x}_{k,W_k}$. Here, if the k-th user is active, its message $W_k \in [M]$ is chosen uniformly at random; if it is inactive, we denote $W_k = 0$ and $\mathbf{x}_{(k)} = \mathbf{0}$. Denote $\mathbf{\Phi} \in \{0,1\}^{MK\times K}$ the binary selection matrix that satisfies $[\mathbf{\Phi}]_{(k-1)M+W_k,k} = 1$ if the k-th user is active and the W_k-th codeword is transmitted by this user, and $[\mathbf{\Phi}]_{(k-1)M+W_k,k} = 0$ otherwise. As presented in Fig. 1, the received signal over L antennas of the BS can be written as

$$\mathbf{Y} = \mathbf{X}\mathbf{\Phi}\mathbf{H} + \mathbf{Z}, \tag{2}$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_L] \in \mathbb{C}^{n\times L}$, $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_L] \in \mathbb{C}^{K\times L}$, $\mathbf{h}_l = [h_{1,l}, \ldots, h_{K,l}]^T \in \mathbb{C}^K$, and $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_L] \in \mathbb{C}^{n\times L}$.

The decoder aims to find the estimated set $\hat{\mathcal{K}}_a$ of active users, and find the estimate $\hat{\mathbf{x}}_{(k)}$ of $\mathbf{x}_{(k)}$ and corresponding message $\hat{W}_k$ of W_k for $k \in \hat{\mathcal{K}}_a$. We denote $\hat{W}_k = 0$ and $\hat{\mathbf{x}}_{(k)} = \mathbf{0}$ for $k \notin \hat{\mathcal{K}}_a$. As noted previously, in this work, we consider two scenarios: CSIR (the decoder knows the realization of the fading channel beforehand) and no-CSI (the decoder does not have *a priori* knowledge of the realization of the fading channel but it knows its distribution in advance). In the case of CSIR, we assume the number K_a of active users is fixed and known to the receiver in advance as in [6]; in the case of no-CSI, we consider two settings: 1) K_a is fixed and known to the receiver *a priori*; 2) K_a is random and unknown to the receiver, but its distribution is known in advance.

Based on the PUPE criterion in [6] and [8], we introduce the notion of a massive random access code for the case of CSIR and no-CSI with known K_a as follows:

Definition 1 (CSIR and known K_a): Let $\mathcal{X}_k$, $\mathcal{H}_k$, and $\mathcal{Y}$ denote the input alphabet of user k, the channel fading coefficient alphabet of user k, and the output alphabet, respectively. An $(n, M, \epsilon, P)_{\text{CSIR},K_a}$ massive random access code consists of

1) An encoder $f_{\text{en},k} : [M] \mapsto \mathcal{X}_k$ that maps the message $W_k \in [M]$ to a codeword $\mathbf{x}_{(k)} \in \mathcal{X}_k$ for $k \in \mathcal{K}_a$. The codewords in $\{\mathcal{X}_k : k \in \mathcal{K}\}$ satisfy the power constraint

$$\|\mathbf{x}_{k,m}\|_2^2 \leq nP, \quad k \in \mathcal{K}, \;\; m \in [M]. \tag{3}$$

We assume that W_k is equiprobable on $[M]$ for $k \in \mathcal{K}_a$.

2) A decoder $g_{\text{de,CSIR},K_a} : \mathcal{Y} \times \prod_{k\in\mathcal{K}} \mathcal{H}_k \mapsto [M]^{K_a}$ that satisfies the PUPE constraint

$$P_e = \frac{1}{K_a} \sum_{k\in\mathcal{K}_a} \mathbb{P}\left[W_k \neq \hat{W}_k\right] \leq \epsilon, \tag{4}$$

where $\hat{W}_k = \left(g_{\text{de,CSIR},K_a}(\mathbf{Y},\mathbf{H})\right)_k$ denotes the decoded message for user k in the case of CSIR with known K_a to the receiver in advance.

Definition 2 (No-CSI and Known K_a): Let $\mathcal{X}_k$ and $\mathcal{Y}$ denote the input alphabet of user k and the output alphabet, respectively. An $(n, M, \epsilon, P)_{\text{no-CSI},K_a}$ massive random access code consists of

1) An encoder $f_{\text{en},k} : [M] \mapsto \mathcal{X}_k$ that maps the message $W_k \in [M]$ to a codeword $\mathbf{x}_{(k)} \in \mathcal{X}_k$ for $k \in \mathcal{K}_a$. The codewords satisfy the power constraint in (3). We assume that W_k is equiprobable on $[M]$ for $k \in \mathcal{K}_a$.

2) A decoder $g_{\text{de,no-CSI},K_a} : \mathcal{Y} \mapsto [M]^{K_a}$ that satisfies the PUPE constraint in (4) for the case of no-CSI with known K_a to the receiver in advance. The decoded message for user k is denoted as $\hat{W}_k = \left(g_{\text{de,no-CSI},K_a}(\mathbf{Y})\right)_k$.

In the following, we introduce the notion of a massive random access code for the no-CSI case when the number K_a of active users is random and unknown. Specifically, we assume that each user becomes active independently with identical probability p_a during any given block. In this case, the number K_a of active users is random and distributed as $K_a \sim \text{Binom}(K, p_a)$, which is assumed to be known to the receiver as in [34], [35], and [36]. The probability of the event that $K_a = \mathrm{K}_a$, i.e., there are exactly $\mathrm{K}_a \in \{0, 1, \ldots, K\}$ active users among K potential users, is given by

$$P_{K_a}(\mathrm{K}_a) = \binom{K}{\mathrm{K}_a} {p_a}^{\mathrm{K}_a} (1-p_a)^{K-\mathrm{K}_a}. \tag{5}$$

Based on the per-user probability of misdetection/false-alarm in [19], we introduce the notion of a massive random access code for the no-CSI case with random and unknown K_a as follows:

Definition 3 (No-CSI and Unknown K_a): Let $\mathcal{X}_k$ and $\mathcal{Y}$ denote the input alphabet of user k and output alphabet, respectively. An $(n, M, \epsilon_{\text{MD}}, \epsilon_{\text{FA}}, P)_{\text{no-CSI, no-}K_a}$ massive random access code consists of

1) An encoder $f_{\text{en},k} : [M] \mapsto \mathcal{X}_k$ that maps the message $W_k \in [M]$ to a codeword $\mathbf{x}_{(k)} \in \mathcal{X}_k$ for $k \in \mathcal{K}_a$. The codewords satisfy the power constraint in (3). We assume that W_k is equiprobable on $[M]$ for $k \in \mathcal{K}_a$.

2) A decoder $g_{\text{de,no-CSI,no-}K_a} : \mathcal{Y} \mapsto [M]^{|\hat{\mathcal{K}}_a|}$ that satisfies the per-user probability of misdetection constraint and the per-user probability of false-alarm constraint as follows:

$$P_{e,\text{MD}} = \mathbb{E}_{K_a}\left[1\left[K_a > 0\right] \frac{1}{K_a} \sum_{k\in\mathcal{K}_a} \mathbb{P}\left[W_k \neq \hat{W}_k\right]\right] \tag{6}$$

$$\leq \epsilon_{\text{MD}}, \tag{7}$$

$$P_{e,\text{FA}} = \mathbb{E}_{|\hat{\mathcal{K}}_a|}\left[1\left[|\hat{\mathcal{K}}_a| > 0\right] \frac{1}{|\hat{\mathcal{K}}_a|} \sum_{k\in\hat{\mathcal{K}}_a} \mathbb{P}\left[\hat{W}_k \neq W_k\right]\right] \tag{8}$$

$$\leq \epsilon_{\text{FA}}, \tag{9}$$

where the decoded message $\hat{W}_k$ for user k is given by $\hat{W}_k = \left(g_{\text{de,no-CSI,no-}K_a}(\mathbf{Y})\right)_k$ in the case of no-CSI with unknown K_a at the decoder.

Let $S_e = \frac{K_a J}{n}$ denote the spectral efficiency and $E_b = \frac{nP}{J}$ denote the energy-per-bit. The minimum energy-per-bit in the case of CSIR and no-CSI with known K_a is defined as

$$E_{b,i}^{*}(n, M, \epsilon) \triangleq \inf\left\{E_b : \exists (n, M, \epsilon, P)_i \text{ code }\right\}, \quad i \in \{\{\text{CSIR},K_a\}, \{\text{no-CSI},K_a\}\}. \tag{10}$$

The minimum energy-per-bit in the case of no-CSI with unknown K_a is defined as

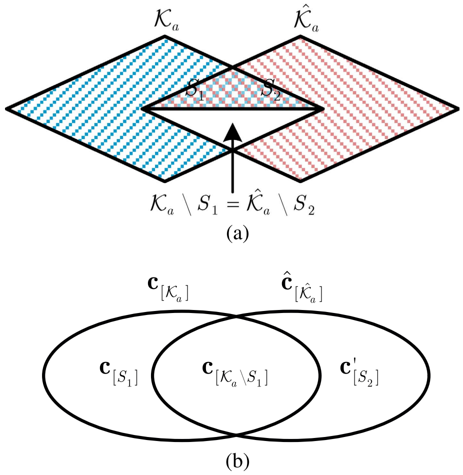
$$E_{b,\text{no-CSI,no-}K_a}^{*}(n, M, \epsilon_{\text{MD}}, \epsilon_{\text{FA}}) \triangleq \inf\left\{E_b : \exists (n, M, \epsilon_{\text{MD}}, \epsilon_{\text{FA}}, P)_{\text{no-CSI,no-}K_a} \text{ code }\right\}. \tag{11}$$

III. Main Results

In this section, we aim to bound the minimum energy-per-bit for ensuring reliable communication in MIMO quasi-static Rayleigh fading massive random access channels with finite blocklength and finite payload size, and to provide corresponding scaling laws. In Section III-A, we first introduce the main proof technique used to derive non-asymptotic achievability bounds for both the CSIR and no-CSI cases, where an appropriate "good region" is selected for massive random access channels. Next, we provide non-asymptotic bounds and scaling laws for the case of CSIR in Section III-B and for the case of no-CSI in Section III-C, respectively. Then, in Section III-D, we derive a non-asymptotic achievability bound for a pilot-assisted scheme. Several possible generalizations of our results are provided in Section III-E.

A. "Good Region" for Massive Random Access Channels

In this subsection, we consider a special case where the number K_a of active users is known *a priori*. A crucial step to derive an achievability bound on the minimum required energy-per-bit is to establish an upper bound on the probability $\mathbb{P}\left[\mathcal{F}_{t,S_1}\right]$ with fixed blocklength n, payload J, power P, and L BS antennas. Here, $\mathcal{F}_{t,S_1}$ denotes the event that there are exactly t misdecoded codewords transmitted by users in the set $S_1 \subset \mathcal{K}_a$. In massive random access channels, a major challenge lies in that the event $\mathcal{F}_{t,S_1}$ is the union of a massive number of error events and most of them are not disjoint.


$\mathcal{K}_a$
$\hat{\mathcal{K}}_a$
S_1
S_2
$\mathcal{K}_a \setminus S_1 = \hat{\mathcal{K}}_a \setminus S_2$
$\mathbf{c}_{[\mathcal{K}_a]}$
$\hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}$
$\mathbf{c}_{[S_1]}$
$\mathbf{c}_{[\mathcal{K}_a \setminus S_1]}$
$\mathbf{c}^{'}_{[S_2]}$

(a)

(b)

Fig. 2. The set relationship: (a) users: S_1 (in blue) denotes the set of active users whose transmitted codewords are misdecoded, S_2 (in red) denotes the set of identified users with false alarm codewords, $S_1 \cap S_2$ includes users that are correctly identified but incorrectly decoded, and $\mathcal{K}_a \backslash S_1 = \hat{\mathcal{K}}_a \backslash S_2$ (in white) includes users that are correctly identified and correctly decoded; (b) codewords: for $S \in \{S_1, \mathcal{K}_a, \mathcal{K}_a \backslash S_1\}$, the set $\mathbf{c}_{[S]}$ includes codewords transmitted by users in the set S, $\hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}$ denotes the set of decoded codewords for users in the set $\hat{\mathcal{K}}_a$, and the set $\mathbf{c}^{'}_{[S_2]}$ includes false alarm codewords corresponding to users in the set S_2.

Specifically, we have $\mathcal{F}_{t,S_1} = \bigcup_{S_2} \bigcup_{\mathbf{c}^{'}_{[S_2]}} \mathcal{F}_{t,S_1,S_2,\mathbf{c}^{'}_{[S_2]}}$. Here, the set $S_2 \subset \mathcal{K} \backslash \mathcal{K}_a \cup S_1$ of size t includes identified users with false alarm codewords, and it is worth noting that S_2 can also take values in S_1 because some users that are correctly identified can still be incorrectly decoded; the set $\mathbf{c}^{'}_{[S_2]}$ includes t false alarm codewords corresponding to users in the set S_2. As a result, the event $\mathcal{F}_{t,S_1}$ is the union of about $\binom{K-K_a+t}{t} M^t$ events, which is considerably large for the massive random access communication problem. The set relationship is presented in Fig. 2.

A classical method to upper-bound the probability $\mathbb{P}[\mathcal{F}_{t,S_1}]$ is applying the union bound, which yields $\mathbb{P}[\mathcal{F}_{t,S_1}] \leq \sum_{S_2} \sum_{\mathbf{c}^{'}_{[S_2]}} \mathbb{P}\left[\mathcal{F}_{t,S_1,S_2,\mathbf{c}^{'}_{[S_2]}}\right]$. However, it may be very loose when the number of terms in the summation is large, as in the massive random access scenario considered in this paper. In order to tightly upper-bound the probability of the union of extremely many events, a standard bounding technique was proposed by Fano [29], which upper-bounds $\mathbb{P}[\mathcal{F}_{t,S_1}]$ as follows:

$$\mathbb{P}[\mathcal{F}_{t,S_1}] \leq \mathbb{P}[\mathcal{F}_{t,S_1}, \mathbf{Y} \in \mathcal{R}_{t,S_1}] + \mathbb{P}[\mathbf{Y} \notin \mathcal{R}_{t,S_1}], \quad (12)$$

where $\mathbf{Y}$ denotes the received signal and $\mathcal{R}_{t,S_1}$ represents a region around the linear combination of the transmitted signals, also known as the "good region" [31]. The union bound is only applied on the first term on the right-hand side (RHS) of (12), and the second term on the RHS of (12) can be tightly bounded and even accurately computed if $\mathcal{R}_{t,S_1}$ is chosen appropriately. With this technique, the probability of the union of many events can be tightly bounded.

To get a tight non-asymptotic achievability bound in massive random access channels, we select an appropriate "good region" $\mathcal{R}_{t,S_1}$ in the remainder of this subsection. Assuming that there is no power constraint, let $\mathbf{c}_{(k)}$ denote the transmitted codeword of the k-th user, which is chosen uniformly at random from its codebook $\mathcal{C}_k$. For a given received signal $\mathbf{Y}$, the decoder searches for the estimated set of active users, i.e. $\hat{\mathcal{K}}_a \subset \mathcal{K}$ of size K_a, and the estimated set of transmitted codewords, i.e. $\hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]} = \left\{\hat{\mathbf{c}}_{(k)} \in \mathcal{C}_k : k \in \hat{\mathcal{K}}_a\right\}$, to minimize the decoding metric $g\left(\mathbf{Y}, \hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right)$. Note that an error event $\mathcal{F}_{t,S_1}$ occurs, if there exists a set of codewords $\mathbf{c}_{[\mathcal{K}_a \backslash S_1]} \cup \mathbf{c}^{'}_{[S_2]}$ satisfying $g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]} \cup \mathbf{c}^{'}_{[S_2]}\right) \leq g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a]}\right)$, where $\mathbf{c}^{'}_{[S]} = \left\{\mathbf{c}^{\prime}_{(k)} \in \mathcal{C}_k : k \in S, \mathbf{c}^{\prime}_{(k)} \neq \mathbf{c}_{(k)}\right\}$ and $\mathbf{c}_{[S]} = \left\{\mathbf{c}_{(k)} \in \mathcal{C}_k : k \in S\right\}$ for the set $S \subset \mathcal{K}$. Roughly speaking, the more similar the "good region" $\mathcal{R}_{t,S_1}$ is to the Voronoi region $\mathcal{V}_{t,S_1}$, the tighter the upper bound on the RHS of (12) is but the higher the complexity is to compute this bound [31], where $\mathcal{V}_{t,S_1}$ is given by

$$\mathcal{V}_{t,S_1} = \Big\{\mathbf{Y} : g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a]}\right) \leq g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]} \cup \mathbf{c}^{'}_{[S_2]}\right), \\ \forall S_2 \subset \mathcal{K} \backslash \mathcal{K}_a \cup S_1, \forall \mathbf{c}^{'}_{[S_2]}\Big\}. \quad (13)$$

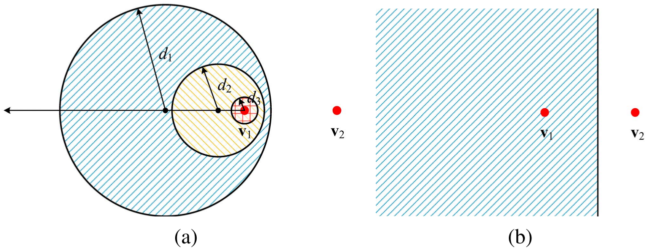
For massive random access in MIMO fading channels, the *"good region"* $\mathcal{R}_{t,S_1}$ used for deriving a tight upper bound on the probability $\mathbb{P}[\mathcal{F}_{t,S_1}]$ in (12) is selected as follows:

$$\mathcal{R}_{t,S_1} = \left\{\mathbf{Y} : g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a]}\right) \leq \omega g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]}\right) + \nu n L\right\}, \quad (14)$$

where $0 \leq \omega \leq 1$ and $\nu \geq 0$. By adjusting ω and ν, we can find a "good region" $\mathcal{R}_{t,S_1}$ similar to the Voronoi region $\mathcal{V}_{t,S_1}$. As a result, when the received signal $\mathbf{Y}$ falls inside $\mathcal{R}_{t,S_1}$, K_a transmitted codewords are likely to be correctly decoded rather than with t misdecoded codewords corresponding to users in the set S_1.

In the following, we take the case of CSIR as an example to clearly illustrate the "good region" $\mathcal{R}_{t,S_1}$. Based on the ML decoding metric, the "good region" $\mathcal{R}_{t,S_1}$ in (14) can be expressed as

$$\mathcal{R}_{t,S_1} = \Bigg\{\mathbf{Y} : \sum_{l=1}^{L} \left\|\mathbf{y}_l - \sum_{k \in \mathcal{K}_a} h_{k,l} \mathbf{c}_{(k)}\right\|_2^2 \\ \leq \omega \sum_{l=1}^{L} \left\|\mathbf{y}_l - \sum_{k \in \mathcal{K}_a \backslash S_1} h_{k,l} \mathbf{c}_{(k)}\right\|_2^2 + \nu n L\Bigg\}. \quad (15)$$



(a)

(b)

Fig. 3. A geometric illustration of the cross section of the "good region" $\mathcal{R}_{t,S_1}$ in the CSIR case: (a) $0 = \omega_3 < \omega_2 < \omega_1 < 1$, $\nu > 0$; (b) $\omega = 1$, $\nu > 0$.

In a special case of $0 \leq \omega < 1$, by straightforward manipulations, the "good region" $\mathcal{R}_{t,S_1}$ in (15) can be rewritten as

$$\mathcal{R}_{t,S_1} = \left\{ \mathbf{Y} : \sum_{l=1}^{L} \left\| \mathbf{y}_l - \frac{\sum_{k\in\mathcal{K}_a} h_{k,l}\mathbf{c}_{(k)} - \omega \sum_{k\in\mathcal{K}_a\setminus S_1} h_{k,l}\mathbf{c}_{(k)}}{1-\omega} \right\|_2^2 \leq \frac{\omega}{(1-\omega)^2} \sum_{l=1}^{L} \left\| \sum_{k\in S_1} h_{k,l}\mathbf{c}_{(k)} \right\|_2^2 + \frac{\nu n L}{1-\omega} \right\}. \tag{16}$$

The region $\mathcal{R}_{t,S_1}$ in (16) can be regarded as a sphere with flexible center and radius for different fading coefficients and codewords. For convenience, we denote $\mathbf{v}_1 = \left[\sum_{k\in\mathcal{K}_a} h_{k,1}\mathbf{c}_{(k)}, \ldots, \sum_{k\in\mathcal{K}_a} h_{k,L}\mathbf{c}_{(k)}\right] \in \mathbb{C}^{n\times L}$ and $\mathbf{v}_2 = \left[\sum_{k\in\mathcal{K}_a\setminus S_1} h_{k,l}\mathbf{c}_{(k)}, \ldots, \sum_{k\in\mathcal{K}_a\setminus S_1} h_{k,L}\mathbf{c}_{(k)}\right] \in \mathbb{C}^{n\times L}$. The center of this sphere can be any point in the ray with endpoint $\mathbf{v}_1$ and direction $\mathbf{v}_1 - \mathbf{v}_2$; the radius of this sphere is $\sqrt{\frac{\omega}{(1-\omega)^2}\left\|\mathbf{v}_1 - \mathbf{v}_2\right\|_F^2 + \frac{\nu n L}{1-\omega}}$. When $\omega = 0$, the region $\mathcal{R}_{t,S_1}$ becomes a sphere with center $\mathbf{v}_1$ and radius $\sqrt{\nu n L}$. We illustrate the cross section of the region $\mathcal{R}_{t,S_1}$ with $0 \leq \omega < 1$ in Fig. 3a. As mentioned above, the region $\mathcal{R}_{t,S_1}$ (the shaded area) is around the sum of the faded codewords transmitted from active users, i.e., around $\mathbf{v}_1$. As in Fig. 3a, for a given ν, as ω increases, the radius of the sphere gradually increases and its center (located in the ray with endpoint $\mathbf{v}_1$ and direction $\mathbf{v}_1 - \mathbf{v}_2$) gradually moves away from $\mathbf{v}_1$. In the special case of $\omega = 1$, the region $\mathcal{R}_{t,S_1}$ becomes a halfspace as shown in Fig. 3b. In other words, the upper bound based on the region $\mathcal{R}_{t,S_1}$ reduces to the commonly used sphere bound [37] and tangential bound [38] in some special cases.

In general, the "good region" $\mathcal{R}_{t,S_1}$ in (14) has some properties as follows:

- When the "good region" $\mathcal{R}_{t,S_1}$ in (14) is the whole observation space, such as in the case of $\omega = 0$ and $\nu = \infty$, the upper bound on $\mathbb{P}\left[\mathcal{F}_{t,S_1}\right]$ in (12) based on $\mathcal{R}_{t,S_1}$ reduces to that obtained by straightforwardly applying the union bound to $\mathbb{P}\left[\mathcal{F}_{t,S_1}\right]$ as provided above.
- In the special case of $\omega = 0$, the "good region" $\mathcal{R}_{t,S_1}$ is independent of the set S_1 and reduces to $\mathcal{R} = \left\{\mathbf{Y} : g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a]}\right) \leq \nu n L\right\}$. In essence, the transmitted signals from active users are treated as a whole for $\mathcal{R}$ with $\omega = 0$, which is equivalent to the case of a single user. However, in the case of $0 < \omega \leq 1$, the region $\mathcal{R}_{t,S_1}$ relies on the set of misdecoded users, which incorporates more details of the massive access model.
- Our "good region" in (14) is parameterized by two parameters ω and ν, which reduces to the one used in [8] if ν is set to 0. In general, in order to derive non-asymptotic achievability bounds, using our "good region" in (14) is better than using the one in [8] for two reasons:
  - Using two parameters instead of one gives us great flexibility and accuracy in tuning the shape of the "good region", thereby leading to tighter achievability results. Specifically, when $\omega = 0$, the region in (14) reduces to $\mathcal{R}$ as explained above, but the upper bound in [8] diverges in this case. Moreover, as in (16), in the CSIR case with $0 \leq \omega < 1$, the region $\mathcal{R}_{t,S_1}$ is essentially a sphere, where its center is determined by ω and its radius is controlled by both ω and ν. However, for the region with $\nu = 0$, both the center and the radius are controlled by ω. Thus, the value of the radius depends on the position of the center for the region in [8], whereas the radius of our "good region" can be flexibly changed by adjusting ν. As a result, it is more likely to find a "good region" similar to the Voronoi region by simultaneously adjusting ω and ν.
  - The problem of finding an appropriate "good region" $\mathcal{R}_{t,S_1}$ can be expressed as $\arg\min_{\omega,\nu} f_{t,S_1}(\omega,\nu)$, where $f_{t,S_1}(\omega,\nu)$ denotes an upper bound on the probability $\mathbb{P}[\mathcal{F}_{t,S_1}]$. Since it is difficult to obtain closed-form solutions for the optimal values of ω and ν, we resort to numerical evaluations with exhaustive search to find ω and ν that yield a tight bound. Note that since the dependency of $f_{t,S_1}(\omega,\nu)$ on ω is more complicated than its dependency on ν, there is a much higher complexity when searching for ω than ν (see Theorem 6 for instance). In contrast to the case of $\nu = 0$, the feasible region of ω, in which the error requirement is satisfied, is enlarged when both ω and ν are taken into consideration. As a result, by introducing ν in (14), we can reduce the number of sampling points when searching for ω, thereby reducing the complexity of finding an appropriate "good region".

B. CSIR

In this subsection, we consider the case where CSI and the number of active users are available at the receiver, and establish non-asymptotic bounds for the massive random access model described in Section II. Specifically, we establish an upper bound on the PUPE in Theorem 1. On the basis of it, Corollary 2 is obtained, which presents an achievability bound (upper bound) on the minimum required energy-per-bit for massive random access. In a special case where all users are assumed to be active, we obtain a simplified achievability bound in Corollary 3. Then, in Theorem 4, we establish a converse bound (lower bound) on the minimum required energy-per-bit assuming user activity is known, and thus it can also be regarded as a converse bound for massive random access. Finally, on the basis of Corollary 3 and Theorem 4, we establish scaling laws in Theorem 5 for a special case where all users are assumed to be active.

1) Achievability Bound: An upper bound on the PUPE for massive random access in MIMO quasi-static Rayleigh fading channels with CSIR and known K_a is given in Theorem 1.

Theorem 1: Assume that there are K_a active users among K potential users each equipped with a single antenna and the number of BS antennas is L. Each user has an individual codebook with size $M = 2^J$ and length n satisfying the maximum power constraint in (3). For massive random access in MIMO quasi-static Rayleigh fading channels with CSIR and known K_a, the PUPE can be upper-bounded as

$$P_e \le \min_{0<P'<P} \left\{ p_0 + \sum_{t=1}^{K_a} \frac{t}{K_a} \min\left\{1, p_{1,t}, p_{2,t}\right\} \right\}, \tag{17}$$

where

$$p_0 = K_a \left(1 - \frac{\gamma\left(n, \frac{nP}{P'}\right)}{\Gamma\left(n\right)} \right), \tag{18}$$

$$p_{1,t} = \min_{0\le\omega\le1, 0\le\nu} \left\{ q_{1,t}(\omega,\nu) + q_{2,t}(\omega,\nu) \right\}, \tag{19}$$

$$q_{1,t}(\omega,\nu) = \sum_{t_0=0}^{t} C_{t_0,t}\, \mathbb{E}_{\tilde{\mathbf{A}}_{S_1}, \tilde{\mathbf{A}}'_{S_2}} \Bigg[\min_{u\ge0, r\ge0, \lambda_{\min}\left(\tilde{\mathbf{B}}\right)>-1} \exp\Big\{ -L \cdot \Big(n\ln(1+r\left(1-\omega\right)) + \ln\Big|\mathbf{I}_K + \tilde{\mathbf{B}}\Big| - rn\nu \Big) \Big\} \Bigg], \tag{20}$$

$$C_{t_0,t} = \binom{K_a}{t}\binom{t}{t_0}\binom{K-K_a}{t-t_0} (M-1)^{t_0} M^{t-t_0}, \tag{21}$$

$$\tilde{\mathbf{B}} = \frac{(1+r-u)(u-r\omega)}{1+r\left(1-\omega\right)} \left(\tilde{\mathbf{A}}_{S_1} - \frac{u}{u-r\omega} \tilde{\mathbf{A}}'_{S_2} \right)^H \cdot \left(\tilde{\mathbf{A}}_{S_1} - \frac{u}{u-r\omega} \tilde{\mathbf{A}}'_{S_2} \right) - \frac{r\omega u}{u-r\omega} \left(\tilde{\mathbf{A}}'_{S_2} \right)^H \tilde{\mathbf{A}}'_{S_2}, \tag{22}$$

$$q_{2,t}(\omega,\nu) = \begin{cases} \min\limits_{\eta\ge0,\delta\ge0} \binom{K_a}{t} \mathbb{E}_{\tilde{\mathbf{A}}_{S_1}} \left[\frac{\gamma\left(tL, L\frac{t(1+\eta)-n\nu+n(1+\delta)(1-\omega)}{\left|\mathbf{I}_n + \omega\tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H\right|^{1/t}} \right)}{\Gamma(tL)} \right] + 2\binom{K_a}{t} \\ \quad - \binom{K_a}{t} \left(\frac{\gamma(tL, tL\ (1+\eta))}{\Gamma(tL)} + \frac{\gamma(nL, nL\ (1+\delta))}{\Gamma(nL)} \right), & t<n, \omega\in(0,1] \\ \min\limits_{\eta\ge0} \binom{K_a}{t} \mathbb{E}_{\tilde{\mathbf{A}}_{S_1}} \left[\frac{\gamma\left(nL, \frac{nL(1+\eta-\nu)}{\omega} \left|\mathbf{I}_n + \tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H\right|^{-1/n} \right)}{\Gamma(nL)} \right] \\ \quad + 1 - \frac{\gamma(nL, nL\ (1+\eta))}{\Gamma(nL)}, & t\ge n, \omega\in(0,1] \\ 1 - \frac{\gamma(nL, nL\ \nu)}{\Gamma(nL)}, & \omega=0 \end{cases} \tag{23}$$

$$p_{2,t} = \min_{0\le\rho\le1, 0\le\beta<\frac{1}{\rho}} \sum_{t_0=0}^{t} \binom{K_a}{t}\binom{t}{t_0}\binom{K-K_a}{t-t_0} M^{\rho t} \cdot \mathbb{E}_{\mathbf{H}_1,\mathbf{H}_2} \left[\exp\left\{ (1-\rho)n\ln\left|\mathbf{I}_L + \beta P'\mathbf{H}_2^H\mathbf{H}_2\right| - n\ln\left|\mathbf{I}_L + \beta(1-\rho\beta)P'\left(\rho\mathbf{H}_1^H\mathbf{H}_1 + \mathbf{H}_2^H\mathbf{H}_2\right)\right| \right\} \right]. \tag{24}$$

Here, $\mathbf{H}_1$ and $\mathbf{H}_2$ are $t \times L$ submatrices of $\mathbf{H} \in \mathbb{C}^{K\times L}$ formed by rows corresponding to the support of S_1 and S_2, respectively; $\mathbf{H} \in \mathbb{C}^{K\times L}$ has i.i.d. $\mathcal{CN}(0,1)$ entries; S_1 is an arbitrary t-subset of $\mathcal{K}_a$; $S_2 = S_{2,1} \cup S_{2,2}$, where $S_{2,1}$ is an arbitrary t_0-subset of S_1 and $S_{2,2}$ is an arbitrary $(t-t_0)$-subset of $\mathcal{K}\backslash\mathcal{K}_a$; $\tilde{\mathbf{A}}_{S_1} = \mathbf{A}\mathbf{\Phi}_{S_1}$ and $\tilde{\mathbf{A}}'_{S_2} = \mathbf{A}\mathbf{\Phi}'_{S_2}$; the matrix $\mathbf{A} \in \mathbb{C}^{n\times MK}$ is the concatenation of codebooks of the K users without power constraint, which has i.i.d. $\mathcal{CN}\left(0, P'\right)$ entries; the binary selection matrix $\mathbf{\Phi}_{S_1} \in \{0,1\}^{MK\times K}$ indicates which codewords are transmitted by users in the set S_1, where $[\mathbf{\Phi}_{S_1}]_{(k-1)M+W_k,k} = 1$ if user $k \in S_1$ is active and transmits the W_k-th codeword, and $[\mathbf{\Phi}_{S_1}]_{(k-1)M+W_k,k} = 0$ otherwise; and similarly, $\mathbf{\Phi}'_{S_2} \in \{0,1\}^{MK\times K}$ indicates which codewords are not transmitted but decoded for users in the set S_2.

Proof Sketch: We use a random coding scheme and an ML decoder, which searches for all possible support sets and finds the one that maximizes the likelihood function. As shown in (17), the upper bound on the PUPE comprises of two terms: the first term p_0 upper-bounds the total variation distance between the measure with power constraint and the one without power constraint, whose expression is given in (18) relying on a straightforward utilization of the union bound; the second term $\sum_{t=1}^{K_a} \frac{t}{K_a} \min\left\{1, p_{1,t}, p_{2,t}\right\}$ upper-bounds the PUPE assuming there is no power constraint. Here, $p_{1,t}$ and $p_{2,t}$ denote two upper bounds on $\mathbb{P}\left[\mathcal{F}_t\right]$, which indicates the probability of the event that there are exactly t misdecoded users. We have $\mathbb{P}\left[\mathcal{F}_t\right] \le \binom{K_a}{t}\mathbb{P}\left[\mathcal{F}_{t,S_1}\right]$. As mentioned in Section III-A, upper-bounding $\mathbb{P}\left[\mathcal{F}_{t,S_1}\right]$ is involved since $\mathcal{F}_{t,S_1}$ is essentially the union of a massive number of events. Two upper bounds on $\mathbb{P}\left[\mathcal{F}_t\right]$, i.e. $p_{1,t}$ and $p_{2,t}$, are obtained as follows:

- In Appendix A, we derive a general upper bound on the PUPE based on Fano's bounding technique [29]. We obtain $p_{1,t}$ by particularizing this general bound to the CSIR case and performing additional manipulations as introduced in Appendix B-A. Specifically, we upper-bound $\mathbb{P}\left[\mathcal{F}_t\right]$ by the sum of two terms as presented in (19). The first term $q_{1,t}(\omega,\nu)$ denotes an upper bound on the probability of the joint event that the decoder yields exactly t misdecoded users and the received signal falls inside the "good region". The expression of $q_{1,t}(\omega,\nu)$ is given in (20), which is obtained by applying the union bound, Chernoff bound, and moment generating function of quadratic forms [39]. The second term $q_{2,t}(\omega,\nu)$ upper-bounds the probability of the event that the received signal falls outside this region, whose expression is given in (23).
- The expression of $p_{2,t}$ is given in (24), which is derived relying on Gallager's ρ-trick [30] as introduced in Appendix B-B. Specifically, given a set S_1 including t misdecoded users and a set S_2 including t detected users with false alarm codewords, Gallager's ρ-trick is applied to the union of about M^t events, corresponding to different sets of false alarm codewords.

See Appendix B for the complete proof. ■

The following corollary of Theorem 1 provides an achievability bound on the minimum required energy-per-bit for the massive random access problem with CSIR and known K_a at the receiver.

Corollary 2: Assume that there are K_a active users among K potential users each equipped with a single antenna and the number of BS antennas is L. Each user has an individual codebook with size $M = 2^J$ and length n satisfying the maximum power constraint in (3). For massive random access in MIMO quasi-static Rayleigh fading channels with CSIR and known K_a, the minimum energy-per-bit $E^*_{b,\mathrm{CSIR},K_a}(n,M,\epsilon)$

for satisfying the PUPE requirement in (4) can be upper-bounded as

$$E^{*}_{b,\text{CSIR},K_a}(n,M,\epsilon) \leq \inf \frac{nP}{J}, \tag{25}$$

where the $\inf$ is taken over all $P>0$ satisfying that

$$\epsilon \geq \min_{0<P'<P} \left\{ p_0 + \sum_{t=1}^{K_a} \frac{t}{K_a} \min\left\{1, p_{1,t}, p_{2,t}\right\} \right\}. \tag{26}$$

Here, p_0, $p_{1,t}$, and $p_{2,t}$ are the same as those in Theorem 1.

In a special case where all users are assumed to be active, Corollary 2 reduces to the following Corollary 3. In essence, the achievability bound for the case where all users are active is equivalent to that with knowledge of the active user set.

Corollary 3: Assume that all users are active, i.e. $K_a = K$. Suppose each user is equipped with a single antenna and the number of BS antennas is L. Each user has an individual codebook with size $M = 2^J$ and length n satisfying the maximum power constraint in (3). In MIMO quasi-static Rayleigh fading channels with CSIR, the minimum energy-per-bit $E^{*}_{b,\text{CSIR},K_a}(n,M,\epsilon)$ for satisfying the PUPE requirement in (4) can be upper-bounded as

$$E^{*}_{b,\text{CSIR},K_a}(n,M,\epsilon) \leq \inf \frac{nP}{J}, \tag{27}$$

where the $\inf$ is taken over all $P>0$ satisfying that

$$\epsilon \geq \min_{0<P'<P} \left\{ \tilde{p}_0 + \sum_{t=1}^{K} \frac{t}{K} \min\left\{1, \tilde{p}_{1,t}, \tilde{p}_{2,t}\right\} \right\}. \tag{28}$$

Here, $\tilde{p}_0$ follows from p_0 in (18) by allowing $K_a = K$; $\tilde{p}_{1,t}$ is obtained by assuming $S_1 = S_2$ and $K_a = K$ in (19), (20), (21), (22), and (23); and $\tilde{p}_{2,t}$ is given by

$$\tilde{p}_{2,t} = \min_{0\leq\rho\leq 1,\rho n\in\mathbb{N}_+} \binom{K}{t} M^{\rho t}\, \mathbb{E}_{\mathbf{G}}\left[\left| \mathbf{I}_t + \frac{P'}{1+\rho}\mathbf{G}\mathbf{G}^H \right|^{-L} \right], \tag{29}$$

where each element of $\mathbf{G} \in \mathbb{C}^{t\times \rho n}$ is i.i.d. $\mathcal{CN}\left(0,1\right)$ distributed. To simplify simulation complexities, $\tilde{p}_{2,t}$ can be further upper-bounded as

$$\tilde{p}_{2,t} \leq \tilde{p}^{\text{u}}_{2,t} = \min_{0\leq\rho\leq 1,\rho n\in\mathbb{N}_+} q(\rho), \tag{30}$$

$$q(\rho) = \begin{cases} \binom{K}{t} M^{\rho t} \left(\frac{P'}{1+\rho}\right)^{-Lt} \prod\limits_{i=\rho n-t+1}^{\rho n} \frac{\Gamma(i-L)}{\Gamma(i)}, & \rho n \geq t+L \\ \binom{K}{t} M^{\rho t} \left(\frac{P'}{1+\rho}\right)^{-L\rho n} \prod\limits_{i=t-\rho n+1}^{t} \frac{\Gamma(i-L)}{\Gamma(i)}, & \rho n \leq t-L \\ 1, & t-L < \rho n < t+L \end{cases}. \tag{31}$$

Proof: See Appendix C. ■

When the BS is equipped with a single antenna, assuming all users are active and the number of users grows linearly and unboundedly with the blocklength, an achievability bound on the minimum required energy-per-bit was derived in [8, Th. IV.4] for the case of CSIR. In contrast, we consider a more practical communication system with random access, multiple BS antennas, and finite blocklength. In general, there are two major differences in the proof ideas of our achievability bounds and the result in [8, Th. IV.4]. First, we utilize standard bounding techniques proposed by Fano [29] and by Gallager [30] (corresponding to (19) and (24) in Theorem 1, respectively), whereas only the latter one, namely Gallager's ρ-trick, is used in [8, Th. IV.4]. When random access is taken into consideration, in contrast to the "good region"-based bound (19), more samples are required by Gallager's ρ-trick bound (24) to obtain a good estimate, which can be observed from numerical simulation. Gallager's ρ-trick bound (24) is easy-to-evaluate only for a special case with knowledge of the active user set. Thus, for the massive random access problem, we resort to the bounding technique proposed by Fano [29] and the "good region" selected in (14). Second, when the BS is equipped with a single antenna, a key idea used in [8, Th. IV.4] is to drop a subset of users (less than ϵK_a) with very bad channel gains and decode the rest [40]. However, this idea is not applicable in our regime for two reasons: 1) as introduced in Section IV, ϵK_a is very small and even less than 1 in most of our considered settings, thereby making this decoding technique useless; 2) the channel quality imbalance between different users is greatly reduced when multiple antennas are equipped at the receiver, and it is not necessary to drop some users.

2) Converse Bound: Apart from the achievability bound, we provide a converse bound on the minimum required energy-per-bit for the massive random access problem in MIMO quasi-static Rayleigh fading channels with CSIR in the following theorem.

Theorem 4: Assume that there are K_a active users among K potential users each equipped with a single antenna and the number of BS antennas is L. Let $M = 2^J$ be the codebook size and n be the blocklength. For massive random access in MIMO quasi-static Rayleigh fading channels with CSIR, the minimum energy-per-bit $E^{*}_{b,\text{CSIR},K_a}(n,M,\epsilon)$ required for satisfying the PUPE requirement in (4) can be lower-bounded as

$$E^{*}_{b,\text{CSIR},K_a}(n,M,\epsilon) \geq \inf \frac{nP}{J}. \tag{32}$$

The $\inf$ is taken over all $P>0$ satisfying that

$$\left(\frac{t}{K_a} - \epsilon\right) J - h_2\left(\epsilon\right) \leq \frac{n}{K_a} \mathbb{E}_{\mathbf{H}_t}\left[\log_2\left|\mathbf{I}_L + P\mathbf{H}_t^H\mathbf{H}_t\right|\right], \quad \forall t \in [K_a], \tag{33}$$

where $\mathbf{H}_t \in \mathbb{C}^{t\times L}$ has i.i.d. $\mathcal{CN}(0,1)$ entries. The condition in (33) can be loosened to:

$$\begin{aligned}&\left(\frac{t}{K_a} - \epsilon\right) J - h_2\left(\epsilon\right) \\ &\leq \frac{n}{K_a} \min\{L\log_2(1+Pt)\,, t\log_2(1+PL)\}\,, \forall t\in[K_a].\end{aligned} \tag{34}$$

The minimum required energy-per-bit $E^{*}_{b,\text{CSIR},K_a}(n,M,\epsilon)$ should also satisfy the meta-converse bound for the single-user multiple-receive-antenna channel with CSIR [41, Th. 1].

Proof Sketch: For the converse bound with multiple users, we first utilize the Fano inequality and then bound the mutual information therein under the assumption of CSIR, which contributes to (33). In order to simplify calculations, we further upper-bound the RHS of (33) and obtain (34) by applying the concavity of the $\log_2|\cdot|$ function. Moreover, the minimum required energy-per-bit $E^*_{b,\mathrm{CSIR},K_a}(n, M, \epsilon)$ should also satisfy the converse bound for the single-user multiple-antenna channels in the CSIR case [41, Th. 1], which is based on the meta-converse theorem in [11]. See Appendix D for the complete proof. ■

3) Asymptotic Analysis: On the basis of the achievability bound in Corollary 3 and the converse bound in Theorem 4, we establish scaling laws of the number of reliably served users in Theorem 5 for a special case where all users are assumed to be active.

Theorem 5: Assume that all users are active, i.e. $K_a = K$. Each user is equipped with a single antenna and the number of BS antennas is L. The channel is assumed to be Rayleigh distributed. Each user has an individual codebook with size M and length n satisfying the maximum power constraint in (3). Let $n, K \to \infty$, $M = \Theta(1)$, $\ln K = o(n)$, and $KP = \Omega(1)$. In the case of CSIR, the PUPE requirement in (4) is satisfied if and only if $\frac{nL \ln KP}{K} = \Omega(1)$.

Proof: See Appendix E. ■

Remark 1: In the case of CSIR, under the assumptions in Theorem 5, the sufficient and necessary condition $\frac{nL \ln KP}{K} = \Omega(1)$ for satisfying the PUPE requirement can be divided into the following two regimes: 1) $\frac{nL}{K} = \Omega(1)$ and $KP = \Theta(1)$; 2) $\frac{nL \ln KP}{K} = \Omega(1)$ and $KP \to \infty$. The first regime is power-limited, where the number of degrees of freedom, i.e., $n \min\{K, L\} = nL$, grows linearly with the number of users. It was pointed out in [32] that, in the single-user case, the minimum received energy-per-user required to transmit a finite number of information bits is given by $nLP = \Theta(1)$. By allocating orthogonal resources to K users, the minimum required energy-per-user nLP in the first regime can be as low as that in the single-user case. The second regime is degrees-of-freedom-limited, where the number of degrees of freedom, i.e. nL, is far less than the number of users, and the minimum received energy-per-user $nLP \to \infty$.

Remark 2: In the case of CSIR, under the maximum power constraint in (3) and the PUPE requirement in (4), the number of reliably served users is in the order of $K = \mathcal{O}(n^2)$ in two regimes: 1) the number of BS antennas is $L = \Theta(n)$ and the power satisfies $P = \Theta\left(\frac{1}{n^2}\right)$; 2) the number of BS antennas is $L = \Theta\left(\frac{n}{\ln n}\right)$ and the power satisfies $P = \Theta\left(\frac{1}{n}\right)$.

Proof: See Appendix E. ■

Our scaling law in Theorem 5 is proved from both the achievability side and the converse side, which reveals the tightness of our bounds in Corollary 3 and Theorem 4 in asymptotic cases. Moreover, it indicates the great potential of multiple receive antennas for the data detection problem. Specifically, we can observe from the condition $\frac{nL \ln KP}{K} = \Omega(1)$ that, when the number L of BS antennas is increased, the maximum number K of reliably served users can be greatly increased and the required blocklength n and power P can be greatly decreased. As in Remark 2, in order to reliably serve $K = \mathcal{O}(n^2)$ users, when the number of BS antennas is increased from $L = \Theta\left(\frac{n}{\ln n}\right)$ to $L = \Theta(n)$, the minimum required power can be considerably decreased from $P = \Theta\left(\frac{1}{n}\right)$ to $P = \Theta\left(\frac{1}{n^2}\right)$. Notably, the case of $P = \Theta\left(\frac{1}{n}\right)$ and the case of $P = \Theta\left(\frac{1}{n^2}\right)$ imply that the energy-per-bit is finite and goes to 0, respectively, which are crucial in practical communication systems with stringent energy constraints.

C. No-CSI

In this subsection, we consider the case where neither the transmitters nor the decoder knows the realization of fading coefficients, but they both know the distribution. In this noncoherent setting, we establish non-asymptotic bounds for the massive random access model described in Section II, where both the cases with known K_a and unknown K_a are considered. Specifically, in Theorem 6, we establish an upper bound on the PUPE for massive random access with known K_a. On the basis of it, Corollary 7 is established, which presents an achievability bound (upper bound) on the minimum required energy-per-bit. For a general setting where the number of active users is random and unknown at the receiver, we establish an achievability bound on the minimum required energy-per-bit in Theorem 8. Then, we present the converse bounds (lower bounds) on the minimum required energy-per-bit in the cases with and without the knowledge of the number K_a of active users at the receiver in Theorem 9 and Theorem 10, respectively, where the multiple-user Fano type bounds are established under the assumption of i.i.d. Gaussian codebooks. Finally, on the basis of Corollary 7 and Theorem 9, we establish scaling laws in Theorem 11 for a special case where all users are assumed to be active.

1) Achievability Bound With Known K_a: An upper bound on the PUPE for massive random access in MIMO quasi-static Rayleigh fading channels in the case with no-CSI and known K_a is given in Theorem 6.

Theorem 6: Assume that there are K_a active users among K potential users each equipped with a single antenna and the number of BS antennas is L. Each user has an individual codebook with size $M = 2^J$ and length n satisfying the maximum power constraint in (3). For massive random access in MIMO quasi-static Rayleigh fading channels with known K_a but unknown CSI at the receiver, the PUPE is upper-bounded as

$$P_e \leq \min_{0<P'<P} \left\{ p_0 + \sum_{t=1}^{K_a} \frac{t}{K_a} \min\{1, p_t\} \right\}, \tag{35}$$

where

$$p_0 = K_a \left(1 - \frac{\gamma\left(n, \frac{nP}{P'}\right)}{\Gamma(n)} \right), \tag{36}$$

$$p_t = \min_{0\leq\omega\leq1, 0\leq\nu} \left\{ q_{1,t}(\omega,\nu) + q_{2,t}(\omega,\nu) \right\}, \tag{37}$$

$$q_{1,t}(\omega,\nu) = \binom{K_a}{t}\binom{K-K_a+t}{t} M^t$$

$$\cdot\, \mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}, \mathbf{A}'_{S_2}} \Bigg[\min_{u\geq0, r\geq0, \lambda_{\min}(\mathbf{B})>0} \exp\{L(rn\nu - u\ln|\mathbf{F}'|$$

$$+(u-r)\ln|\mathbf{F}|+r\omega\ln|\mathbf{F}_1|-\ln|\mathbf{B}|)\}\Big], \tag{38}$$

$$\mathbf{B}=(1-u+r)\mathbf{I}_n+u\left(\mathbf{F}'\right)^{-1}\mathbf{F}-r\omega\mathbf{F}_1^{-1}\mathbf{F}, \tag{39}$$

$$\mathbf{F}=\mathbf{I}_n+\mathbf{A}_{\mathcal{K}_a}\mathbf{A}_{\mathcal{K}_a}^H, \tag{40}$$

$$\mathbf{F}'=\mathbf{I}_n+\mathbf{A}_{\mathcal{K}_a\backslash S_1}\mathbf{A}_{\mathcal{K}_a\backslash S_1}^H+\mathbf{A}'_{S_2}\left(\mathbf{A}'_{S_2}\right)^H, \tag{41}$$

$$\mathbf{F}_1=\mathbf{I}_n+\mathbf{A}_{\mathcal{K}_a\backslash S_1}\mathbf{A}_{\mathcal{K}_a\backslash S_1}^H, \tag{42}$$

$$q_{2,t}(\omega,\nu)=\binom{K_a}{t}\min_{\delta\geq 0}\left\{1-\frac{\gamma\left(nL,nL\ \ (1+\delta)\right)}{\Gamma\left(nL\right)}\right.$$
$$\left.+\mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}}\left[\frac{\gamma\left(Lm,L\prod_{i=1}^{m}\lambda_i^{-\frac{1}{m}}\frac{n(1+\delta)(1-\omega)-\omega\ln|\mathbf{F}_1|+\ln|\mathbf{F}|-n\nu}{\omega}\right)}{\Gamma\left(Lm\right)}\right]\right\}. \tag{43}$$

Here, S_1 is an arbitrary t-subset of $\mathcal{K}_a$; S_2 is an arbitrary t-subset of $\mathcal{K}\backslash\mathcal{K}_a\cup S_1$; $\mathbf{A}_S$ denotes an $n\times|S|$ submatrix of $\mathbf{A}$ including transmitted codewords of active users in the set $S\subset\mathcal{K}_a$; $\mathbf{A}'_{S_2}$ denotes an $n\times|S_2|$ submatrix of $\mathbf{A}$ including false-alarm codewords for users in the set S_2; the matrix $\mathbf{A}\in\mathbb{C}^{n\times MK}$ is the concatenation of codebooks of the K users without power constraint, which has i.i.d. $\mathcal{CN}\left(0,P'\right)$ entries; and $\lambda_1,\ldots,\lambda_m$ denote non-zero eigenvalues of $\mathbf{F}_1^{-1}\mathbf{A}_{S_1}\mathbf{A}_{S_1}^H$ with $m=\min\{n,t\}$.

Proof Sketch: Similar to the CSIR case, we use the random coding scheme and the ML decoder in the no-CSI case with known K_a. The PUPE can be upper-bounded as the sum of two terms as in (35): the first term p_0 upper-bounds the total variation distance between the measures with and without power constraint; the second term $\sum_{t=1}^{K_a}\frac{t}{K_a}\min\{1,p_t\}$ upper-bounds the PUPE assuming there is no power constraint. Here, p_t denotes an upper bound on $\mathbb{P}\left[\mathcal{F}_t\right]$, which indicates the probability of the event that there are exactly t misdecoded users. There are some differences in bounding $\mathbb{P}\left[\mathcal{F}_t\right]$ between in the case of CSIR and no-CSI. First, Gallager's ρ-trick is difficult to apply in the no-CSI case. Specifically, for a set S_2 including t detected users with false alarm codewords, there are about M^t (extremely large) events corresponding to different sets of false alarm codewords. Gallager's ρ-trick can be useful only if it is applied to the union of these events at first. However, the probability over false alarm codewords conditioned on other variables is difficult to handle in the no-CSI case because they exist in many terms including $\ln|\cdot|$ and $\mathrm{tr}\left(\cdot\right)$. Therefore, in this case, we only utilize the bounding technique proposed by Fano [29] and apply the "good region" selected in (14). Second, different from the CSIR case, both the channel and noise are unknown to the receiver in the no-CSI case. As a result, the effects due to noise and channel are coupled together, and it is difficult to separate these two effects in the analysis. Fortunately, when channels are Rayleigh distributed, conditioned on $\mathbf{X}$, the received signal $\mathbf{Y}$ is Gaussian distributed, making the analysis easier. The main techniques used for bounding $\mathbb{P}\left[\mathcal{F}_t\right]$ in the CSIR case, such as the Chernoff bound and moment generating function of quadratic forms, are applied in the no-CSI case. See Appendix F for the complete proof. ■

The following corollary of Theorem 6 provides an achievability bound on the minimum required energy-per-bit for massive random access with known K_a and no-CSI at the receiver.

Corollary 7: Assume that there are K_a active users among K potential users each equipped with a single antenna and the number of BS antennas is L. Each user has an individual codebook with size $M=2^J$ and length n satisfying the maximum power constraint in (3). For massive random access in MIMO quasi-static Rayleigh fading channels with known K_a but unknown CSI at the receiver, the minimum energy-per-bit $E_{b,\text{no-CSI},K_a}^{*}(n,M,\epsilon)$ for satisfying the PUPE requirement in (4) can be upper-bounded as

$$E_{b,\text{no-CSI},K_a}^{*}(n,M,\epsilon)\leq\inf\frac{nP}{J}, \tag{44}$$

where the inf is taken over all $P>0$ satisfying that

$$\epsilon\geq\min_{0<P'<P}\left\{p_0+\sum_{t=1}^{K_a}\frac{t}{K_a}\min\{1,p_t\}\right\}. \tag{45}$$

Here, p_0 and p_t are the same as those in Theorem 6.

In the single-receive-antenna setting with known active user set, an asymptotic achievability bound on the minimum required energy-per-bit was derived in [8, Th. IV.1] for the no-CSI case. In the multiple-receive-antenna setting, a non-asymptotic achievability bound is provided in Corollary 7. There are some differences between the proof ideas of [8, Th. IV.1] and Theorem 6 in our work. First, we utilize the "good region" given in (14), which is better than the one used in [8, Th. IV.1] and reduces to it if ν is set to 0 as mentioned in Section III-A. Second, the projection decoder is used in [8, Theorem IV.1] for the single-receive-antenna setting, but we leverage the ML decoder in the multiple-receive-antenna setting in this work. As mentioned in the introduction, the projection decoder has an advantage of requiring no knowledge of the fading distribution, which can be observed from the decoding criterion of the projection decoder given by

$$\left[\hat{\mathcal{K}}_a,\{\hat{W}_k:k\in\hat{\mathcal{K}}_a\}\right]=$$
$$\underset{\hat{\mathcal{K}}_a\subset[K],|\hat{\mathcal{K}}_a|=K_a}{\operatorname{argmax}}\ \max_{\{\hat{W}_k\in[M]:k\in\hat{\mathcal{K}}_a\}}\ \max_{\mathbf{H}}\mathbb{P}\left[\mathbf{Y}\middle|\mathbf{X},\{\hat{W}_k:k\in\hat{\mathcal{K}}_a\},\mathbf{H}\right], \tag{46}$$

where $\mathbf{X}\in\mathbb{C}^{n\times MK}$ denotes the concatenation of codebooks of the K users, $\mathbf{H}$ contains the channel fading coefficients, $\mathbf{Y}$ denotes the received signal, $\hat{\mathcal{K}}_a$ denotes the estimated set of active users, and $\hat{W}_k$ denotes the decoded message for user k. However, the projection decoder can be ineffectual when it is applied to the framework with multiple receive antennas and the number of active users is larger than the blocklength. Meanwhile, it is challenging (although not impossible) to jointly deal with the signals received over L antennas based on the projection decoder, because the analysis of the angle between the subspace spanned by L received signals and the one spanned by K_a codewords is quite involved. Thus, we leverage the ML decoder in this work, which is efficient in the multiple-receive-antenna setting no matter whether K_a is less than n or not, at the price of requiring *a priori* distribution on $\mathbf{H}$, which can be observed from the ML decoding criterion given by

$$\Big[\hat{\mathcal{K}}_a, \{\hat{W}_k : k \in \hat{\mathcal{K}}_a\}\Big] = \underset{\hat{\mathcal{K}}_a \subset [K], |\hat{\mathcal{K}}_a| = K_a}{\operatorname{argmax}} \max_{\{\hat{W}_k \in [M] : k \in \hat{\mathcal{K}}_a\}} \mathbb{P}\Big[\mathbf{Y}\Big|\mathbf{X}, \{\hat{W}_k : k \in \hat{\mathcal{K}}_a\}\Big], \tag{47}$$

$$\mathbb{P}\left[\mathbf{Y}\middle|\mathbf{X}, \{\hat{W}_k : k \in \hat{\mathcal{K}}_a\}\right] = \mathbb{E}_{\mathbf{H}}\left\{\mathbb{P}\left[\mathbf{Y}\middle|\mathbf{X}, \{\hat{W}_k : k \in \hat{\mathcal{K}}_a\}, \mathbf{H}\right]\right\}. \tag{48}$$

2) Achievability Bound With Random and Unknown K_a: In Theorem 6 and Corollary 7, we assume K_a is known at the receiver in advance and the decoder outputs K_a messages. In such a setup, a misdetection for a user implies a false-alarm for another user, and vice versa. In this part, we consider a general case in which the number of active users is random and unknown to the receiver. In this case, we need to account for both the per-user probability of misdetection and the per-user probability of false-alarm. The following theorem provides an achievability bound on the minimum required energy-per-bit for the no-CSI case when the number K_a of active users is random and unknown.

Theorem 8: Assume that there are K potential users each equipped with a single antenna and the number of BS antennas is L. The number of active users is assumed to be random and unknown, which is distributed as $K_a \sim \mathrm{Binom}(K, p_a)$. Each user has an individual codebook with size $M = 2^J$ and length n satisfying the maximum power constraint in (3). For massive random access in MIMO quasi-static Rayleigh fading channels with no-CSI, the minimum energy-per-bit $E^*_{b,\text{no-CSI,no-}K_a}(n, M, \epsilon_{\mathrm{MD}}, \epsilon_{\mathrm{FA}})$ for satisfying the per-user probability of misdetection and the per-user probability of false-alarm requirements in (7) and (9) can be upper-bounded as

$$E^*_{b,\text{no-CSI,no-}K_a}(n, M, \epsilon_{\mathrm{MD}}, \epsilon_{\mathrm{FA}}) \leq \inf \frac{nP}{J}. \tag{49}$$

The inf is taken over all $P > 0$ satisfying

$$\epsilon_{\mathrm{MD}} \geq \min_{0<P'<P} \left\{ p_0 + \sum_{\mathrm{K}_a=1}^{K} P_{K_a}(\mathrm{K}_a) \sum_{\mathrm{K}'_a=0}^{K} \sum_{t\in\mathcal{T}_{\mathrm{K}'_a}} \frac{t + (\mathrm{K}_a - \mathrm{K}'_{a,u})^+}{\mathrm{K}_a} \cdot \min\left\{1, \sum_{t'\in\bar{\mathcal{T}}_{\mathrm{K}'_a,t}} p_{\mathrm{K}'_a,t,t'}, p_{\mathrm{K}_a\to\mathrm{K}'_a}\right\}\right\}, \tag{50}$$

$$\epsilon_{\mathrm{FA}} \geq \min_{0<P'<P} \left\{ p_0 + \sum_{\mathrm{K}_a=0}^{K} P_{K_a}(\mathrm{K}_a) \cdot \sum_{\mathrm{K}'_a=0}^{K} \sum_{t\in\mathcal{T}_{\mathrm{K}'_a}} \sum_{t'\in\mathcal{T}_{\mathrm{K}'_a,t}} \frac{t' + (\mathrm{K}'_{a,l} - \mathrm{K}_a)^+}{\hat{\mathrm{K}}_a} \cdot \min\left\{1, p_{\mathrm{K}'_a,t,t'}, p_{\mathrm{K}_a\to\mathrm{K}'_a}\right\}\right\}, \tag{51}$$

where

$$p_0 = p_a K \left(1 - \frac{\gamma\left(n, \frac{nP}{P'}\right)}{\Gamma(n)}\right), \tag{52}$$

$$\mathcal{T}_{\mathrm{K}'_a} = \left[0 : \min\{\mathrm{K}_a, \mathrm{K}'_{a,u}\}\right], \tag{53}$$

$$\bar{\mathcal{T}}_{\mathrm{K}'_a,t} = \Big[\left((\mathrm{K}_a - \mathrm{K}'_{a,u})^+ - (\mathrm{K}_a - \mathrm{K}'_{a,l})^+ + t\right)^+ : (\mathrm{K}'_{a,u} - \mathrm{K}_a)^+ - (\mathrm{K}'_{a,l} - \mathrm{K}_a)^+ + t\Big], \tag{54}$$

$$\mathcal{T}_{\mathrm{K}'_a,t} = \Big[\left((\mathrm{K}_a - \mathrm{K}'_{a,u})^+ - (\mathrm{K}'_{a,l} - \mathrm{K}_a)^+ + \max\{\mathrm{K}'_{a,l}, 1\} - \mathrm{K}_a + t\right)^+ : (\mathrm{K}'_{a,u} - \mathrm{K}_a)^+ - (\mathrm{K}'_{a,l} - \mathrm{K}_a)^+ + t\Big], \tag{55}$$

$$\hat{\mathrm{K}}_a = \mathrm{K}_a - t - (\mathrm{K}_a - \mathrm{K}'_{a,u})^+ + t' + (\mathrm{K}'_{a,l} - \mathrm{K}_a)^+, \tag{56}$$

$$\mathrm{K}'_{a,l} = \max\{0, \mathrm{K}'_a - r'\}, \tag{57}$$

$$\mathrm{K}'_{a,u} = \min\{K, \mathrm{K}'_a + r'\}, \tag{58}$$

$$p_{\mathrm{K}'_a,t,t'} = \min_{0\leq\omega\leq1, 0\leq\nu} \Big\{q_{1,\mathrm{K}'_a,t,t'}(\omega,\nu) + 1\left[t + (\mathrm{K}_a - \mathrm{K}'_{a,u})^+ > 0\right] q_{2,\mathrm{K}'_a,t}(\omega,\nu) + 1\left[t + (\mathrm{K}_a - \mathrm{K}'_{a,u})^+ = 0\right] q_{2,\mathrm{K}'_a,t,0}(\omega,\nu)\Big\}, \tag{59}$$

$$q_{1,\mathrm{K}'_a,t,t'}(\omega,\nu) = C_{\mathrm{K}'_a,t,t'} \cdot \mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}, \mathbf{A}_{\mathcal{K}_a\setminus S_1}, \mathbf{A}_{\mathcal{K}_a\setminus S_{1,1}}, \mathbf{A}'_{S_2}, \mathbf{A}'_{S_{2,1}}}\Bigg[\min_{u\geq0, r\geq0, \lambda_{\min}(\mathbf{B})>0} \exp\{Lrn\nu + b_{u,r} + L(u\ln|\mathbf{F}''| - r\ln|\mathbf{F}| - u\ln|\mathbf{F}'| + r\omega\ln|\mathbf{F}_1| - \ln|\mathbf{B}|)\}\Bigg], \tag{60}$$

$$C_{\mathrm{K}'_a,t,t'} = \binom{\mathrm{K}_a}{t + (\mathrm{K}_a - \mathrm{K}'_{a,u})^+} \binom{K - \min\{\mathrm{K}_a, \mathrm{K}'_{a,u}\} + t}{t' + (\mathrm{K}'_{a,l} - \mathrm{K}_a)^+} \cdot M^{t' + (\mathrm{K}'_{a,l} - \mathrm{K}_a)^+}, \tag{61}$$

$$\mathbf{F}'' = \mathbf{I}_n + \mathbf{A}_{\mathcal{K}_a\setminus S_{1,1}} \mathbf{A}^H_{\mathcal{K}_a\setminus S_{1,1}} + \mathbf{A}'_{S_{2,1}} (\mathbf{A}'_{S_{2,1}})^H, \tag{62}$$

$$\mathbf{B} = (1 + r)\mathbf{I}_n - u\left(\mathbf{F}''\right)^{-1}\mathbf{F} + u\left(\mathbf{F}'\right)^{-1}\mathbf{F} - r\omega\mathbf{F}_1^{-1}\mathbf{F}, \tag{63}$$

$$b_{u,r} = -ub'' + rb + ub' - r\omega b_1, \tag{64}$$

$$b = \ln\left(P_{K_a}(\mathrm{K}_a)\right) - \mathrm{K}_a \ln M, \tag{65}$$

$$b_1 = \ln\left(P_{K_a}\left(\mathrm{K}_a - t - (\mathrm{K}_a - \mathrm{K}'_{a,u})^+\right)\right) - \left(\mathrm{K}_a - t - (\mathrm{K}_a - \mathrm{K}'_{a,u})^+\right)\ln M, \tag{66}$$

$$b' = \ln\left(P_{K_a}\left(\hat{\mathrm{K}}_a\right)\right) - \hat{\mathrm{K}}_a \ln M, \tag{67}$$

$$b'' = \ln\left(P_{K_a}\left(\mathrm{K}_a - (\mathrm{K}_a - \mathrm{K}'_{a,u})^+ + (\mathrm{K}'_{a,l} - \mathrm{K}_a)^+\right)\right) - \left(\mathrm{K}_a - (\mathrm{K}_a - \mathrm{K}'_{a,u})^+ + (\mathrm{K}'_{a,l} - \mathrm{K}_a)^+\right)\ln M, \tag{68}$$

$$q_{2,\mathrm{K}'_a,t}(\omega,\nu) = \binom{\mathrm{K}_a}{t + (\mathrm{K}_a - \mathrm{K}'_{a,u})^+} \cdot \min_{\delta\geq0}\Bigg\{1 - \frac{\gamma\left(nL, nL\ \ (1+\delta)\right)}{\Gamma(nL)} + \mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}}\Bigg[\frac{\gamma\left(Lm, \frac{nL(1+\delta)(1-\omega) - \omega(L\ln|\mathbf{F}_1| - b_1) + L\ln|\mathbf{F}| - b - nL\nu}{\omega\prod_{i=1}^m \lambda_i^{1/m}}\right)}{\Gamma(Lm)}\Bigg]\Bigg\}, \tag{69}$$

$$q_{2,\mathrm{K}'_a,t,0} = \mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}}\left[1 - \frac{\gamma\left(nL, \frac{nL\nu}{1-\omega} - L\ln|\mathbf{F}| + b\right)}{\Gamma(nL)}\right], \tag{70}$$

$$p_{\mathrm{K}_a\to\mathrm{K}'_a} = \min_{\tilde{\mathrm{K}}_a\in[0:K], \tilde{\mathrm{K}}_a\neq\mathrm{K}'_a} \Big\{1\left[\mathrm{K}'_a < \tilde{\mathrm{K}}_a\right] p_{\mathrm{K}_a\to\mathrm{K}'_a,1} + 1\left[\mathrm{K}'_a > \tilde{\mathrm{K}}_a\right] p_{\mathrm{K}_a\to\mathrm{K}'_a,2}\Big\}, \tag{71}$$

$$p_{\mathrm{K}_a \to \mathrm{K}'_a,1} = \min \Bigg\{ \min_{\eta>0} \Bigg\{ \mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}} \left[\frac{\gamma\Big(Lm', \prod_{i=1}^{m'} (\lambda_i^{'})^{-1/m'} nL\Big(1+\frac{\mathrm{K}'_a+\tilde{\mathrm{K}}_a}{2}P'-\eta\Big)\Big)}{\Gamma\left(Lm'\right)} \right] + \frac{\gamma\left(nL, nL\eta\right)}{\Gamma\left(nL\right)} \Bigg\}, \mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}} \left[\min_{\rho\geq 0} \exp\left\{ \rho nL\left(1+\frac{\mathrm{K}'_a+\tilde{\mathrm{K}}_a}{2}P'\right) - L\ln|\mathbf{I}_n + \rho\mathbf{F}| \right\} \right] \Bigg\}, \tag{72}$$

$$p_{\mathrm{K}_a \to \mathrm{K}'_a,2} = \min \Bigg\{ \min_{\eta>0} \Bigg\{ 2 - \mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}} \left[\frac{\gamma\left(Lm', \frac{nL}{\lambda'_1}\left(1+\frac{\mathrm{K}'_a+\tilde{\mathrm{K}}_a}{2}P'-\eta\right)\right)}{\Gamma\left(Lm'\right)} \right] - \frac{\gamma\left(nL, nL\eta\right)}{\Gamma\left(nL\right)} \Bigg\}, \mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}} \left[\min_{0\leq\rho<1/(1+\lambda'_1)} \exp\left\{ -\rho nL\left(1+\frac{\mathrm{K}'_a+\tilde{\mathrm{K}}_a}{2}P'\right) - L\ln|\mathbf{I}_n - \rho\mathbf{F}| \right\} \right] \Bigg\}. \tag{73}$$

Here, $\mathbf{F}$, $\mathbf{F}'$, and $\mathbf{F}_1$ are defined in (40), (41), and (42), respectively; r' denotes a nonnegative integer referred to as the decoding radius; S_1 is an arbitrary subset of $\mathcal{K}_a$ of size $t + (\mathrm{K}_a - \mathrm{K}'_{a,u})^+$, which denotes the set of users whose codewords are misdecoded and can be divided into two subsets $S_{1,1}$ and $S_{1,2}$ of size $(\mathrm{K}_a - \mathrm{K}'_{a,u})^+$ and t, respectively; S_2 is an arbitrary subset of $\mathcal{K}\backslash\mathcal{K}_a \cup S_1$ of size $t' + (\mathrm{K}'_{a,l} - \mathrm{K}_a)^+$, which denotes the set of detected users with false-alarm codewords; $S_{2,1}$ is an arbitrary subset of S_2 of size $(\mathrm{K}'_{a,l} - \mathrm{K}_a)^+$; $\mathbf{A}_S$ denotes an $n \times |S|$ submatrix of $\mathbf{A}$ including transmitted codewords of users in the set $S \subset \mathcal{K}_a$; $\mathbf{A}'_S$ denotes an $n \times |S|$ submatrix of $\mathbf{A}$ including false-alarm codewords for users in the set $S \subset \mathcal{K}$; the matrix $\mathbf{A} \in \mathbb{C}^{n\times MK}$ is the concatenation of codebooks of all users without power constraint, which has i.i.d. $\mathcal{CN}\left(0, P'\right)$ entries; $\lambda_1^{'}, \ldots, \lambda_{m'}^{'}$ are non-zero eigenvalues of $\mathbf{A}_{\mathcal{K}_a}\mathbf{A}_{\mathcal{K}_a}^H$ in decreasing order with $m' = \min\left\{n, \mathrm{K}_a\right\}$; and $\lambda_1, \ldots, \lambda_m$ denote non-zero eigenvalues of $\mathbf{F}_1^{-1}\mathbf{A}_{S_1}\mathbf{A}_{S_1}^H$ with $m = \min\left\{n, t + (\mathrm{K}_a - \mathrm{K}'_{a,u})^+\right\}$.

Proof Sketch: The receiver first estimates the number of active users via an energy-based estimator, which is denoted as K'_a, and then outputs a set of decoded messages of size $\hat{\mathrm{K}}_a \in [\mathrm{K}'_{a,l} : \mathrm{K}'_{a,u}]$ via an MAP-based decoder. The quantity p_0 upper-bounds the total variation distance between the measures with and without power constraint. When there is no power constraint, $p_{\mathrm{K}_a \to \mathrm{K}'_a}$ upper-bounds the probability of the event that the estimation of K_a is K'_a, which is obtained based on the Chernoff bound and moment generating function of quadratic forms. Moreover, $p_{\mathrm{K}'_a,t,t'}$ upper-bounds the probability of the event that there are exactly $t + (\mathrm{K}_a - \mathrm{K}'_{a,u})^+$ misdetected codewords and $t' + (\mathrm{K}'_{a,l} - \mathrm{K}_a)^+$ false-alarm codewords, which is derived along similar lines as in the case of known K_a. See Appendix G for the complete proof. ■

Theorem 8 presents an achievability bound on the minimum required energy-per-bit for the case in which the number K_a of active users is random and unknown. Specifically, we first estimate the number of active users via an energy-based estimator, which is denoted as K'_a; then, we obtain a set of decoded messages of size $\hat{\mathrm{K}}_a$ via an MAP-based decoder, where $\hat{\mathrm{K}}_a$ is selected from the interval $[\mathrm{K}'_{a,l} : \mathrm{K}'_{a,u}]$ determined by K'_a and r'. The decoding radius r' can be optimized according to the target misdetection and false-alarm probabilities. In general, a large decoding radius r' can reduce the error probabilities suffering from inaccurate estimation of the number of active users; however, increasing r' may increase the chance that the decoder returns a set of codewords whose posterior probability is larger than that of the transmitted codewords, especially when P is small [19].

Compared with [19], where a random-coding achievability bound was derived for Gaussian massive random access channels assuming K_a is unknown *a priori*, there are two main changes in this work. First, we employ the MAP-based decoder rather than the ML-based decoder used in [19]. When K_a is unknown, the number of decoded messages is not given in advance. In this case, it is more advantageous to use the MAP-based decoder since it incorporates prior distributions in users' messages of various sizes, at the price of requiring the knowledge of the distribution of K_a. Indeed, knowing the distribution of K_a is a common assumption in many works such as [19], [34], [35], and [36]. Second, compared with Gaussian channels considered in [19], we further consider the massive random access problem in MIMO quasi-static Rayleigh fading channels, which increases the difficulties of upper-bounding the error probabilities. For example, the probability of the event that the number of active users is estimated as K'_a is obtained by straightforward manipulation in [19], whereas more techniques, such as the Chernoff bound, "good region"-trick, and moment generating function of quadratic forms, are employed in quasi-static Rayleigh fading channels.

3) Converse Bound With Known K_a: In Theorem 9, we provide a converse bound on the minimum required energy-per-bit for massive random access in MIMO quasi-static Rayleigh fading channels with no-CSI and known K_a. This converse bound contains two parts, namely the multiple-user Fano type bound and the single-user bound, where the former relies on the assumption of i.i.d Gaussian codebooks (i.e., the converse is a weaker ensemble converse), but the single-user bound holds for all codes.

Theorem 9: Assume that there are K_a active users among K potential users each equipped with a single antenna and the number of BS antennas is L. Each user has an individual codebook with size $M = 2^J$ and length n. For massive random access in MIMO quasi-static Rayleigh fading channels with no-CSI and known K_a, the minimum energy-per-bit required for satisfying the PUPE requirement in (4) can be lower-bounded as

$$E^{*}_{b,\text{no-CSI},K_a}(n,M,\epsilon) \geq \inf \frac{nP}{J}. \tag{74}$$

The $\inf$ is taken over all $P > 0$ satisfying the following two conditions:

1) Under the assumption that codewords have i.i.d. Gaussian entries, it should be satisfied that

$$b_1 \leq \frac{LC}{K_a} - \frac{L}{K_a}\mathbb{E}_{\mathbf{X}_{K_a}}\left[\log_2\left|\mathbf{I}_{K_a} + \mathbf{X}^H_{K_a}\mathbf{X}_{K_a}\right|\right], \tag{75}$$

$$C = \min\left\{n\log_2\left(1+K_aP\right), K_aM\log_2\left(1+\frac{nP}{M}\right)\right\}, \tag{76}$$

where $b_1 = J(1-\epsilon) - h_2(\epsilon)$ and $\mathbf{X}_{K_a}$ is an $n \times K_a$ matrix with each entry i.i.d. from $\mathcal{CN}(0,P)$. The condition in (75) can be loosened to

$$b_1 \leq \begin{cases} \frac{LC}{K_a} - \frac{L}{K_a}\sum\limits_{i=0}^{K_a-1}\Big(\psi(n-i)\log_2 e \\ \qquad + \log_2\Big(P + \frac{1}{n-i}\Big)\Big), \quad 1 \leq K_a \leq n \\ \frac{LC}{K_a} - \frac{L}{K_a}\sum\limits_{i=0}^{n-1}\Big(\psi(K_a-i)\log_2 e \\ \qquad + \log_2\Big(P + \frac{1}{K_a-i}\Big)\Big), \; K_a > n \end{cases}, \tag{77}$$

where $\psi(\cdot)$ denotes Euler's digamma function.

2) The single-user finite-blocklength bound shows that

$$M \leq \frac{1}{\mathbb{P}\left[\chi^2(2L) \geq (1+(n+1)P)r\right]}, \tag{78}$$

where r is the solution of

$$\mathbb{P}\left[\chi^2(2L) \leq r\right] = \epsilon. \tag{79}$$

Proof Sketch: Similar to the CSIR case, we first utilize Fano's inequality; then, we follow the idea in [42] to deal with the mutual information therein. Under the assumption of i.i.d. Gaussian codebooks, we obtain (75) for the scenario with multiple BS antennas and finite blocklength, which reduces to an easy-to-evaluate bound in (77). Moreover, the minimum required energy-per-bit $E^{*}_{b,\text{no-CSI},K_a}(n,M,\epsilon)$ should also satisfy the single-user meta-converse bound in [43, Th. 3] with three changes as follows: 1) both the number of transmitting antennas and the number of subcodewords are set to be 1; 2) the blocklength is changed from n to $n+1$ because we consider the maximum power constraint in (3), which can be replaced by the equal power constraint in [43] following from the standard $n \rightarrow n+1$ trick [11, Lemma 39]; 3) to reduce the simulation complexity of the meta-converse bound in the single-user case, we choose the auxiliary distribution as $Q_{\mathbf{Y}^{(n+1)\times L}} = \prod_{l=1}^{L}\mathcal{CN}(0,\mathbf{I}_{n+1})$, rather than the output distribution induced by the input distribution as considered in [43]. See Appendix H for the complete proof of the Fano type bound. ■

Under the assumption that the entries of codebooks are i.i.d. with mean zero and variance P, a converse bound was established in [8] and [42], in which the number of users is assumed to grow linearly and unboundedly with the blocklength and the BS is assumed to be equipped with a single antenna. In the scenario with multiple BS antennas and finite blocklength, some useful techniques used in [8] and [42], such as some results from random matrix theory, are not applicable, and it becomes more involved to obtain an easy-to-evaluate converse bound. Instead, in Theorem 9, we make stronger assumptions, i.e., we assume codebooks have i.i.d. $\mathcal{CN}(0,P)$ entries, which makes the analysis easier. This raises an interesting open question of whether an easy-to-evaluate non-asymptotic converse bound can be obtained for the massive access problem in the multiple-receive-antenna setting under more general assumptions on the codebooks.

4) Converse Bound With Random and Unknown K_a: In Theorem 10, we provide a converse bound on the minimum required energy-per-bit for massive random access in MIMO quasi-static Rayleigh fading channels with no-CSI and unknown number of active users. Similar to the case of known K_a, the converse bound in Theorem 10 contains two parts, namely the multiple-user Fano type bound and the single-user bound, where the former relies on the assumption of i.i.d Gaussian codebooks (i.e., the converse is a weaker ensemble converse), but the single-user bound holds for all codes.

Theorem 10: Assume that there are K potential users each equipped with a single antenna and the number of BS antennas is L. The number of active users is random and unknown, which is distributed as $K_a \sim \text{Binom}(K,p_a)$. Each user has an individual codebook with size $M = 2^J$ and length n. For massive random access in MIMO quasi-static Rayleigh fading channels with no-CSI, the minimum energy-per-bit required for satisfying the error requirements in (7) and (9) can be lower-bounded as

$$E^{*}_{b,\text{no-CSI},\text{no-}K_a}(n,M,\epsilon_{\text{MD}},\epsilon_{\text{FA}}) \geq \inf \frac{nP}{J}. \tag{80}$$

The $\inf$ is taken over all $P > 0$ satisfying the following two conditions:

1) Under the assumptions that each codebook has i.i.d. $\mathcal{CN}(0,P)$ entries and $\epsilon_{\text{MD}} + \epsilon_{\text{FA}} \leq 1 - \frac{1}{1+2^{h_2(p_a)+p_aJ}}$, it should be satisfied that

$$b_1 \leq \frac{LC}{K} - \frac{L}{K}\sum_{\mathrm{K}_a=0}^{K} P_{K_a}(\mathrm{K}_a)\mathbb{E}_{\mathbf{X}_{\mathrm{K}_a}}\left[\log_2\left|\mathbf{I}_n + \mathbf{X}_{\mathrm{K}_a}\mathbf{X}^H_{\mathrm{K}_a}\right|\right], \tag{81}$$

$$b_1 = (1-\epsilon_{\text{MD}}-\epsilon_{\text{FA}})\left(h_2(p_a)+p_aJ\right) - h_2(\epsilon_{\text{MD}}+\epsilon_{\text{FA}}), \tag{82}$$

$$C = \min\left\{n\log_2(1+p_aKP), K\ M\log_2\left(1+\frac{p_a}{M}nP\right)\right\}, \tag{83}$$

where $P_{K_a}(\mathrm{K}_a)$ denotes the probability of the event that there are exactly K_a active users given in (5) and $\mathbf{X}_{\mathrm{K}_a}$ denotes an $n \times \mathrm{K}_a$ matrix with each entry i.i.d. from $\mathcal{CN}(0,P)$.

2) The single-user finite-blocklength bound shows that

$$M \leq \frac{\epsilon_1}{\mathbb{P}\left[\chi^2(2L) \geq (1+(n+1)P)r\right]}, \tag{84}$$

where r is the solution of

$$\mathbb{P}\left[\chi^2(2L) \leq r\right] = \epsilon_2, \tag{85}$$

$$\epsilon_1 = \min\left\{1, \frac{\epsilon_{\mathrm{FA}}}{1-p_a}\right\}, \tag{86}$$

$$\epsilon_2 = \min\left\{1, \frac{\epsilon_{\mathrm{MD}}}{p_a}\right\}. \tag{87}$$

Proof Sketch: Both Condition 1 and Condition 2 take the uncertainty of user activities into consideration. Inspired by [4], condition 1 is established for the massive random access problem applying Fano's inequality, under the assumption that codebooks have i.i.d. $\mathcal{CN}(0, P)$ entries. Condition 2 is established based on the single-user random access converse result in [18, Th. 2] with a properly selected auxiliary distribution (motivated by [44]). See Appendix I for the complete proof. ■

In [4], a Fano type converse bound was established for Gaussian massive random access channels under the joint error probability criterion. In this case, it was pointed out in [4] that Fano's converse bound matches the achievability result well in terms of the message-length capacity, and the capacity penalty due to unknown user activities on each of the K_a active users is $H_2(p_a)/p_a$ in the asymptotic regime with infinite number of users. In this work, under the assumption of Gaussian codebooks, we extend the Fano type converse result in [4] to the multiple-receive-antenna fading channels under the PUPE criterion. Moreover, based on the result in [18], we establish a finite-blocklength converse bound for the single-user random access problem in multiple-receive-antenna fading channels with unknown user activity, which can also be regarded as a converse bound for the massive random access problem.

5) Asymptotic Analysis: On the basis of the achievability bound in Theorem 6 and the converse bound in Theorem 9, we establish scaling laws of the number of reliably served users in Theorem 11 for a special case in which all users are assumed to be active.

Theorem 11: Assume that all users are active, i.e. $K_a = K$. Each user is equipped with a single antenna and the number of BS antennas is L. The channel is assumed to be Rayleigh distributed. Each user has an individual codebook with size M and length n satisfying the maximum power constraint in (3). Let $n, L \to \infty$ and $M = \Theta(1)$. In the case of no-CSI, when the number of BS antennas is in the order of $L = \Theta\left(n^2\right)$ and the power satisfies $P = \Theta\left(\frac{1}{n^2}\right)$, one can reliably serve up to $K = \mathcal{O}(n^2)$ users. A matching converse result is established assuming codebooks have i.i.d. Gaussian entries.

Proof: See Appendix J. ■

In order to obtain the scaling law on the achievability side, both the activity detection problem considered in [20] and the data detection problem of interest in this work can be formulated as similar sparse support recovery problems. This is because one can immediately obtain a data detection scheme from an activity detection scheme by assigning to each user a unique set of codewords, such that a user can transmit the codeword corresponding to its information message. Thus, by expanding the number of users from K to KM and expanding the number of active users from K_a to K, the scaling law of the activity detection problem in [20] can be extended to that of the data detection problem as presented in Table I: under the joint error probability criterion, with blocklength $n \to \infty$ and a sufficient number of BS antennas $L = \Theta\left(n^2 \ln n\right)$, one can reliably serve up to $K = \mathcal{O}\left(n^2\right)$ users when the payload is $J = \Theta(1)$ and the power is $P = \Theta\left(\frac{1}{n^2}\right)$. Notably, there are some differences between this result and our scaling law in Theorem 11. First, the joint error probability criterion is used in [20], but we utilize the PUPE criterion in this work, which is more appropriate for massive access channels [6]. We point out that the required number of BS antennas can be reduced from $L = \Theta\left(n^2 \ln n\right)$ to $L = \Theta\left(n^2\right)$ when we change from the joint error probability criterion to the PUPE criterion. Second, the result in [20] is on the achievability side; Theorem 11 is proved from both the achievability and converse sides, in which the converse result relies on the assumption that the codebooks have i.i.d. Gaussian entries. Notably, in our regime, it is satisfied that $n^2 P = \Theta(1)$, i.e., the energy-per-bit goes to 0, which is attractive for IoT settings with stringent energy constraints.

In this subsection, without assuming *a priori* CSI at the receiver, we focus on the regime of $K = \mathcal{O}(n^2)$, because this is the maximum number of users that can be reliably served in the sparse support recovery problem to the best of our knowledge. Theorem 11 shows that, when the power is $P = \Theta\left(\frac{1}{n^2}\right)$, one can reliably serve up to $K = \mathcal{O}\left(n^2\right)$ users with $L = \Theta\left(n^2\right)$ BS antennas. However, it is still unknown how the number of reliably served users increases as the number of BS antennas further increases.

D. Pilot-Assisted Scheme in the No-CSI Case

The pilot-assisted coded access scheme is widely used in practical wireless systems when there is no *a priori* CSI at the receiver. This scheme consists of two stages: 1) users transmit dedicated pilots for channel estimation; 2) users transmit codewords and the receiver utilizes the channel estimates to decode. This methodology falls into the general framework of the mismatched decoder [28]. From an information-theoretic perspective, channel estimation can be viewed as a specific form of coding in the no-CSI case as explained in the introduction. In this subsection, we only consider a special case where all users are active for simplicity, and establish an upper bound on the PUPE in Theorem 12. Indeed, the achievability bound for the case where all users are active is equivalent to that with knowledge of the active user set.

Theorem 12: Assume that all users are active, i.e. $K_a = K$. Each user is equipped with a single antenna and the number of BS antennas is L. Each user has a dedicated pilot with length $n_p \leq \min\{n, K\}$ and power $n_p P_p \leq nP$. The matrix $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_K] \in \mathbb{C}^{n_p \times K}$ comprises of pilots of all users, which are drawn uniformly at random on an n_p-dimensional sphere of radius $\sqrt{n_p P_p}$. Each user also has an individual codebook with size $M = 2^J$ and length $n_d = n - n_p$, satisfying that the power of each codeword is no more than $nP - n_p P_p$. For the pilot-assisted coded access scheme in MIMO quasi-static

Rayleigh fading channels, the PUPE can be upper-bounded as

$$P_e \leq \min_{0<P'<P} \left\{ p_0 + \sum_{t=1}^{K} \frac{t}{K} \min\left\{1, p_t\right\} \right\}, \tag{88}$$

where

$$p_0 = K \left(1 - \frac{\gamma\left(n_d, \frac{nP - n_p P_p}{P'}\right)}{\Gamma\left(n_d\right)} \right), \tag{89}$$

$$p_t = \min_{0\leq\nu} \left\{ q_{1,t}\left(\nu\right) + q_{2,t}\left(\nu\right) \right\}, \tag{90}$$

$$q_{1,t}\left(\nu\right) = \binom{K}{t} M^t \cdot \mathbb{E}_{\tilde{\mathbf{A}}_{\mathcal{K}}, \tilde{\mathbf{A}}_{S_1}, \tilde{\mathbf{A}}'_{S_1}, \mathbf{B}} \left[\min_{u\geq 0, r\geq 0, \lambda_{\min}(\mathbf{D})>0} \exp\left\{ r n_d L \nu - \frac{L}{2} \ln|\mathbf{D}| \right\} \right], \tag{91}$$

$$q_{2,t}\left(\nu\right) = \min\left\{ \mathbb{E}_{\tilde{\mathbf{A}}_{\mathcal{K}}, \mathbf{B}} \left[\min_{0\leq\delta<\frac{1}{1+\lambda_{max}\left(\tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{A}}_{\mathcal{K}}^H\right)}} \exp\left\{-\delta n_d L \nu\right\} \cdot \left| \left(1-\delta\right) \mathbf{I}_{n_d} - \delta \tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{A}}_{\mathcal{K}}^H \right|^{-L} \right], \min_{0\leq\eta\leq\nu} \left\{ 2 - \frac{\gamma\left(n_d L, n_d L \eta\right)}{\Gamma\left(n_d L\right)} - \mathbb{E}_{\tilde{\mathbf{A}}_{\mathcal{K}}, \mathbf{B}} \left[\frac{\gamma\left(L n^*, \frac{n_d L (\nu - \eta)}{\lambda_{max}\left(\tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{A}}_{\mathcal{K}}^H\right)}\right)}{\Gamma\left(L n^*\right)} \right] \right\} \right\}, \tag{92}$$

$$\mathbf{D} = \left(1+r\right) \mathbf{I}_{2n_d} + u\left(1-u+r\right) \bar{\mathbf{\Sigma}}_1 + r \bar{\mathbf{\Sigma}}_2 - u\left(u-r\right) \bar{\mathbf{\Sigma}}_1 \bar{\mathbf{\Sigma}}_2, \tag{93}$$

$$\bar{\mathbf{\Sigma}}_1 = \begin{bmatrix} \Re\left(\tilde{\mathbf{D}}_{S_1}\right) & -\Im\left(\tilde{\mathbf{D}}_{S_1}\right) \\ \Im\left(\tilde{\mathbf{D}}_{S_1}\right) & \Re\left(\tilde{\mathbf{D}}_{S_1}\right) \end{bmatrix}, \tag{94}$$

$$\tilde{\mathbf{D}}_{S_1} = \left(\tilde{\mathbf{A}}_{S_1} - \tilde{\mathbf{A}}'_{S_1}\right) \hat{\mathbf{\Sigma}} \left(\tilde{\mathbf{A}}_{S_1} - \tilde{\mathbf{A}}'_{S_1}\right)^H, \tag{95}$$

$$\bar{\mathbf{\Sigma}}_2 = \begin{bmatrix} \Re\left(\tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{A}}_{\mathcal{K}}^H\right) & -\Im\left(\tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{A}}_{\mathcal{K}}^H\right) \\ \Im\left(\tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{A}}_{\mathcal{K}}^H\right) & \Re\left(\tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{A}}_{\mathcal{K}}^H\right) \end{bmatrix}, \tag{96}$$

$$\hat{\mathbf{\Sigma}} = \mathbf{I}_K - \left(\mathbf{I}_K + \mathbf{B}^H \mathbf{B}\right)^{-1}, \tag{97}$$

$$\tilde{\mathbf{\Sigma}} = \left(\mathbf{I}_K + \mathbf{B}^H \mathbf{B}\right)^{-1}, \tag{98}$$

$$n^* = \min\left\{K, n_d\right\}. \tag{99}$$

Here, in a special case where pilots are orthogonal with $n_p = K$, $\hat{\mathbf{\Sigma}}$ in (97) and $\tilde{\mathbf{\Sigma}}$ in (98) reduce to $\hat{\mathbf{\Sigma}} = \frac{n_p P_p}{1+n_p P_p} \mathbf{I}_K$ and $\tilde{\mathbf{\Sigma}} = \frac{1}{1+n_p P_p} \mathbf{I}_K$, respectively; we have $\tilde{\mathbf{A}}_{S_1} = \mathbf{A}\mathbf{\Phi}_{S_1}$, $\tilde{\mathbf{A}}'_{S_1} = \mathbf{A}\mathbf{\Phi}'_{S_1}$, and $\tilde{\mathbf{A}}_{\mathcal{K}} = \mathbf{A}\mathbf{\Phi}_{\mathcal{K}}$; the matrix $\mathbf{A} \in \mathbb{C}^{n_d \times MK}$ is the concatenation of codebooks of the K users without power constraint, which has i.i.d. $\mathcal{CN}\left(0, P'\right)$ entries; S_1 is an arbitrary t-subset of $\mathcal{K}$; the binary selection matrix $\mathbf{\Phi}_{S_1} \in \{0,1\}^{MK\times K}$ indicates which codewords are transmitted by users in the set S_1, where $\left[\mathbf{\Phi}_{S_1}\right]_{(k-1)M+W_k,k} = 1$ if user k in the set S_1 is active and the W_k-th codeword is transmitted, and $\left[\mathbf{\Phi}_{S_1}\right]_{(k-1)M+W_k,k} = 0$ otherwise; and similarly, $\mathbf{\Phi}'_{S_1} \in \{0,1\}^{MK\times K}$ indicates which codewords are not transmitted but decoded for users in the set S_1.

Proof Sketch: The power of each pilot is $n_p P_p$ and the power of each codeword is no more than $nP - n_p P_p$, thereby satisfying the power constraint in (3). In the pilot transmission phase, users transmit dedicated pilots and the receiver estimates channels based on the MMSE criterion. In the data transmission phase, we use the random coding scheme and assume that users transmit codewords uniformly selected from their own codebooks. For the pilot-assisted scheme, the decoder has an incorrect estimate of the channel but uses the estimate as if it were perfect, which is different from the case of CSIR. Due to the channel estimation error, bounding $\mathbb{P}\left[\mathcal{F}_t\right]$ is more involved for the pilot-assisted scheme than in the case of CSIR. Thus, in this subsection, we only utilize the bounding technique proposed by Fano in [29] and simplify the "good region" given in (14) with $\omega = 0$. See Appendix K for the complete proof. ■

The following corollary of Theorem 12 provides an achievability bound on the minimum required energy-per-bit for the pilot-assisted coded access scheme.

Corollary 13: Assume that all users are active. Each user is equipped with a single antenna and the number of BS antennas is L. Assume each user has a dedicated pilot with length $n_p < n$ and power $n_p P_p < nP$. Each user also has an individual codebook with size $M = 2^J$ and length $n_d = n - n_p$ satisfying that the power of each codeword is no more than $nP - n_p P_p$. For the pilot-assisted scheme in MIMO quasi-static Rayleigh fading channels, the minimum energy-per-bit $E^*_{b,\text{no-CSI},K_a}(n, M, \epsilon)$ for satisfying the PUPE requirement in (4) can be upper-bounded as

$$E^*_{b,\text{no-CSI},K_a}(n, M, \epsilon) \leq \inf \frac{nP}{J}, \tag{100}$$

where the inf is taken over all $P > 0$ satisfying that

$$\epsilon \geq \min_{0<P'<P} \left\{ p_0 + \sum_{t=1}^{K} \frac{t}{K} \min\left\{1, p_t\right\} \right\}. \tag{101}$$

Here, p_0 and p_t are the same as those in Theorem 12.

In Corollary 13, we derive an achievability bound on the minimum required energy-per-bit for the pilot-assisted transmission scheme. As we can see from the result, there exists a tradeoff between the accuracy of the estimated CSI and the blocklength available for data transmission. That is, a longer pilot is beneficial to improve the channel estimation performance, but at the price of reducing the number of channel uses available for data transmission. More results on this can be found in Section IV.

E. Generalizations

In this subsection, we introduce several possible generalizations of the results in this paper.

First, we have focused on MIMO quasi-static Rayleigh fading channels in this work. Note that the results can be extended to other types of fading channels, such as Rician fading. Specifically, for the CSIR case, the derivations of the achievability bound based on Gallager's ρ-trick and the converse bound are independent of the fading distribution (i.e., these bounds can be general). The fading distribution only kicks in when evaluating them numerically, and the Rayleigh distribution assumption could simplify the computation. In both CSIR and no-CSI cases, the "good region"-based achievability bounds for Rayleigh fading channels can be extended to Rician fading channels because the main techniques used to derive them, such as Fano's bounding technique, the union bound, Chernoff bound, and moment generating function of quadratic forms, are also applicable when channels are subject to Rician fading. In addition, in the case of no-CSI, converse bounds derived in [42] are applicable to a general fading model in the single-receive-antenna setting. Applying similar ideas in [42], we can extend the converse bound for Rayleigh fading to various types of fading in MIMO channels.

Second, we have considered the joint activity and data detection problem in MIMO quasi-static Rayleigh fading channels in this work, where each user is assumed to have an individual codebook. Note that the results can be extended to the framework of a common codebook. A similar extension with AWGN channels can be found in [6] and [15].

IV. Numerical Results

In this section, we validate our theoretical results in Section III through numerical simulations. We consider quasi-static fading channels with L BS antennas. The channel between each transmit-receive antenna pair is independently Rayleigh-distributed. We assume the blocklength is $n = 1000$, payload is $J = 100$ bits, and target PUPE is $\epsilon = 0.001$. The required memory space to compute the bounds is $\mathcal{O}(W^2)$ with $W = \max\{n, K\}$. In Section IV-A, we present the number of reliably served active users versus the energy-per-bit when the number of BS antennas is given. In Section IV-B, we present the spectral efficiency versus the number of BS antennas for fixed energy-per-bit. We use the Monte Carlo method with 500 samples to evaluate expectations in the converse bounds. For the achievability bounds, the parameters outside the expectations are optimized by sampling and exhaustively searching, with the expectations therein evaluated by the Monte Carlo method using 500 samples; once these parameters are determined, we generate 10000 samples to obtain ultimate achievability bounds.

A. The Number of Users Versus the Energy-per-bit

In Fig. 4, we present our achievability and converse bounds on the minimum required energy-per-bit with known K_a, together with the achievability bounds on the orthogonalization scheme time division multiple access (TDMA) [11], [12] and the performance of the scheme proposed in [20]. We assume there are $L = 32$ BS antennas. Next, we explain how each curve is obtained:

1) The achievability bound for the case of CSIR with knowledge of the active user set $\mathcal{K}_a$ is based on Corollary 3, where only Gallager's ρ-trick bound $\tilde{p}_{2,t}$ in (29) is utilized because it is tighter than the "good region"-based bound $\tilde{p}_{1,t}$ in our considered regime.

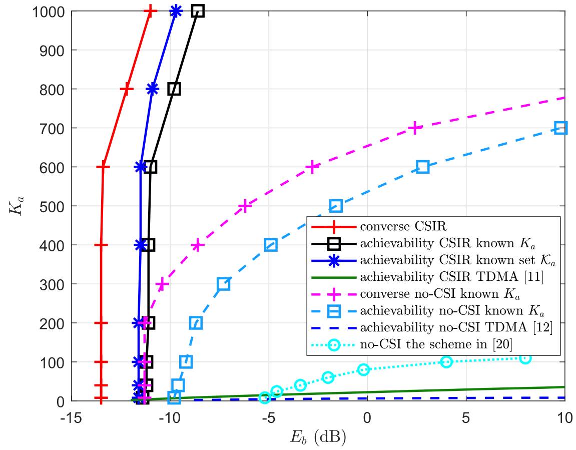


Fig. 4. The number K_a of active users versus the energy-per-bit E_b with $n = 1000$, $J = 100$ bits, $K_a = 0.4K$, $\epsilon = 0.001$, and $L = 32$.

2) The achievability bound for the case of CSIR with known K_a but unknown $\mathcal{K}_a$ is based on the "good region" bound $p_{1,t}$ in Corollary 2. We set $u = \frac{1+r}{2}$ to reduce searching complexity, which is optimal when $\omega = 0$. Gallager's ρ-trick bound $p_{2,t}$ in Corollary 2 is not used because we observe from numerical simulation that it requires an extremely large number of samples to get a good estimate for the massive random access problem.
3) The converse bound for the case of CSIR is Theorem 4.
4) The achievability bound for the no-CSI case with known K_a is Corollary 7, where the "good region"-based bound p_t is provided in Theorem 6. To reduce simulation complexity, we set the parameter u in p_t to be $u = \frac{1+r}{2}$. In this case, the term inside the expectation in (38) is a convex function of r, which is optimized by Newton's method.
5) The converse bound for the case of no-CSI with known K_a is Theorem 9.
6) For TDMA, to achieve the spectral efficiency $S_e = \frac{K_a J}{n}$, we compute the smallest P ensuring the access of an active user with rate $\frac{KJ}{n}$, blocklength $\frac{n}{K}$, target PUPE ϵ, and L BS antennas. Specifically, we utilize the $\kappa\beta$ bound [11, Th. 25] for the case of CSIR and the bound in [12, Eq. (67)] for the case of no-CSI, respectively.
7) For comparison, we present the joint activity and data detection performance of the scheme proposed in [20] for the case of no-CSI. We follow the concatenated coding scheme in [20, Sec. V], suitably adapted to our case. Specifically, we equally divide a coherence block with length $n = 1000$ into $D = 10$ slots. Let each user transmit $J_D = 10$ bits over a slot with $n_D = 100$ dimensions, yielding an overall payload $J = 100$. In each slot, we choose the columns of each coding matrix uniformly i.i.d. from the sphere with radius Pn_D. For the inner code, we assume user k sends the $i_{k,d}$-th column of the coding matrix, where $i_{k,d} \in \left[2^{J_D}\right]$ denotes the message produced by user k in slot d. For the inner decoder, we use the non-Bayesian approach in

[20, Algorithm 1], which is proposed for the unsourced random access model (i.e., the framework of a common codebook). To cater for the framework of individual codebooks, we utilize a hard decision on the support of the estimated vector $\hat{\gamma}$ with the threshold 0.08, and importantly, we restrict that at most one codeword can be decoded in each codebook. Moreover, since each user has unique codebook known at the receiver in advance, the decoded messages across different slots can be stitched based on this prior knowledge. Thus, there is no need to utilize the tree code as the outer code. We obtain the average of the misdetection error probability and the false-alarm error probability, i.e., $P_e = (P_{e,\mathrm{MD}} + P_{e,\mathrm{FA}})/2$, and plot the minimum required energy-per-bit to satisfy $P_e \leq \epsilon$ for different numbers of active users.

As shown in Fig. 4, the gap between our achievability and converse bounds is less than 2.5 dB in all K_a regimes for the CSIR case and less than 4 dB for K_a less than 500 in the case of no-CSI with known K_a. Thus, our non-asymptotic bounds provide relatively accurate theoretical benchmarks to evaluate practical transmission schemes, which are of considerable importance in massive random access systems. In the case of CSIR with known K_a, we can observe that the lack of knowledge of the active user set entails a penalty less than 1.2 dB in terms of energy efficiency. As expected, it is more costly to communicate in the no-CSI case than in the CSIR case, especially for a large number of active users. Additionally, similar to AWGN channels [7] and single-receive-antenna quasi-static fading channels [8], the almost perfect MUI cancellation effect is observed in multiple-receive-antenna quasi-static Rayleigh fading channels. Specifically, when the number of active users is below a critical threshold, the minimum required energy-per-bit is almost a constant in the case of CSIR, although there is a slow growth of the energy-per-bit as K_a increases within this range for the no-CSI case. Moreover, we observe that the scheme in [20] is inferior to the achievability bound in the case of no-CSI, especially when $K_a > 80$. This is because, although the concatenated coding scheme in [20] contributes to the manageability of the coding matrix with the dimension as small as $100 \times 1024K$, it leads to a performance loss since the dimension of a slot is greatly reduced. In addition, the orthogonalization scheme TDMA does not have the perfect MUI cancellation effect. TDMA is shown to be energy-inefficient for large user densities when user activity is known [8], and it becomes more energy-inefficient for the random access model since some resources allocated for inactive users are not utilized.

In Fig. 5, we compare the achievability and converse bounds on the minimum required energy-per-bit in the following two settings with no-CSI: 1) the number of active users is $K_a = 0.4K$, which is fixed and known in advance; 2) the number of active users is random and unknown, and its distribution $K_a \sim \mathrm{Binom}(K, 0.4)$ is known *a priori*. In the case of unknown K_a, it is required that $(\epsilon_{\mathrm{MD}} + \epsilon_{\mathrm{FA}})/2 = \epsilon$. Moreover, the achievability bound for a pilot-assisted scheme is also computed. Next, we explain how each curve is obtained:

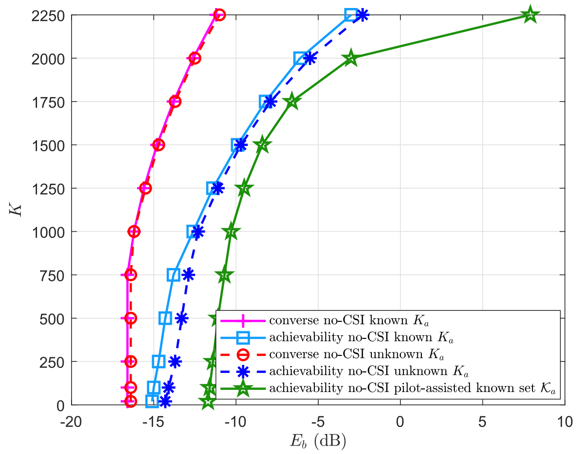


Fig. 5. The number K of potential users versus the energy-per-bit E_b with $n = 1000$, $J = 100$ bits, $\epsilon = 0.001$, and $L = 128$ in two cases: 1) the number of active users is $K_a = 0.4K$, which is fixed and known in advance; 2) the number of active users is random and unknown, and its distributed $K_a \sim \mathrm{Binom}(K, p_a)$ is known *a priori* with $p_a = 0.4$ and mean $\bar{K}_a = 0.4K$.

1) The achievability bounds for the no-CSI case are presented for the settings with and without the knowledge of the number K_a of active users at the receiver. The bound with known K_a is based on Corollary 7, which is computed in a similar way to that in Fig. 4. The bound for the setting with unknown K_a is based on Theorem 8, where the decoding radius r' is determined by brute-force searching from the set $\{0, 1, \ldots, 25\}$.
2) The converse bound for the setting with and without the knowledge of the number K_a of active users is based on Theorem 9 and Theorem 10, respectively.
3) The achievability bound for the pilot-assisted coded access scheme is based on Corollary 13 under the assumption that the active user set $\mathcal{K}_a$ is known *a priori*, wherein the power allocation between the pilot and data symbols is optimized and orthogonal pilots of length $n_p = K_a$ are utilized.

Our results reveal that the pilot-assisted coded access scheme is suboptimal in the no-CSI case, even if the power allocation between the pilot and data symbols is optimized. Specifically, the gap between the achievability bounds of the pilot-assisted scheme and the scheme without explicit channel estimation is less than 3.5 dB when the number of users is less than 800 but sees a dramatic increase when the number of users exceeds this. Moreover, from the achievability and converse bounds with and without the knowledge of the number K_a of active users at the receiver, we can observe that once the distribution $K_a \sim \mathrm{Binom}(K, p_a)$ is known in advance, the uncertainty of the exact value of K_a entails only a small penalty in terms of energy efficiency, with the extra required energy-per-bit less than 0.3 dB on the converse side and less than 1.1 dB on the achievability side.

In Fig. 6, considering the setup with blocklength $n = 1000$, payload $J = 100$ bits, PUPE requirement $\epsilon = 0.001$, $L \in \{128, 500\}$ BS antennas, and known active user set $\mathcal{K}_a$, we compare the non-orthogonal-pilot-based scheme with pilot

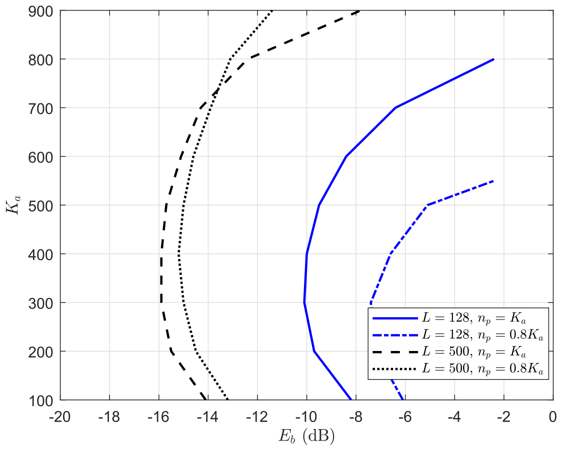
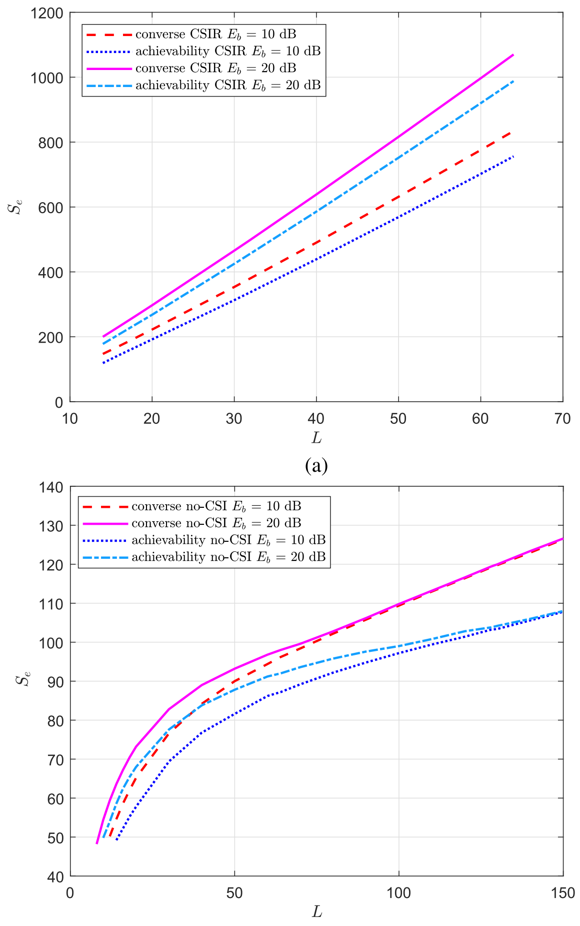


Fig. 6. The number K_a of active users versus the energy-per-bit E_b for the pilot-assisted scheme with $n = 1000$, $J = 100$, $\epsilon = 0.001$, $n_p \in \{K_a, 0.8K_a\}$, $L \in \{128, 500\}$, $P_p = P$, and $P_d \leq P$.

length $n_p = 0.8K_a$ (i.e. $n_d = n - 0.8K_a$ channel uses for data transmission) and the orthogonal-pilot-based scheme with $n_p = K_a$ (i.e. $n_d = n - K_a$ channel uses for data transmission). The non-orthogonal pilots are generated using a sub-sampled discrete Fourier transform matrix. As opposed to Fig. 5, the power allocation between the pilot and data symbols is not optimized in Fig. 6 due to simulation complexity. Specifically, we assume the transmitting power of the pilot per channel use is $P_p = P$ and the transmitting power of the data per channel use is $P_d \leq P$ to satisfy the maximum power constraint in (3). As shown in Fig. 6, the achievability bound for the scheme based on non-orthogonal pilots is inferior to the orthogonal-pilot-based one in the setup with $L = 128$ BS antennas. However, when the number of BS antennas increases to $L = 500$ and the number of users is above 800, the scheme based on non-orthogonal pilots of length $n_p = 0.8K_a$ outperforms the orthogonal-pilot-based one. As a result, for the pilot-assisted scheme, there exists a tradeoff between the channel estimation performance and the blocklength used for data transmission. In particular, for a fixed blocklength n, when the numbers of BS antennas and users are large, it is more reasonable to use non-orthogonal pilots to set aside more channel uses for data transmission, instead of allocating an orthogonal pilot to each user. This is because when the number of users is large, allocating orthogonal pilots results in little time left for data transmission; meanwhile, a large number of BS antennas can mitigate the effects of noise and fast fading, which allows us to reduce the length of pilots.

B. The Spectral Efficiency Versus the Number of BS Antennas

As illustrated in Fig. 7, we present bounds on the maximum spectral efficiency S_e against the number L of BS antennas. Specifically, the looser converse bounds (34) in Theorem 4 and (77) in Theorem 9 are utilized in the case of CSIR and no-CSI, respectively. For the achievability bound in the CSIR case, we utilize the "good region" bound $p_{1,t}$ in Corollary 2, where ω is set to be 0 to reduce simulation complexity.



(a)

(b)

Fig. 7. The spectral efficiency S_e versus the number L of BS antennas with $n = 1000$, $J = 100$ bits, $K_a = 0.4K$, and $\epsilon = 0.001$: (a) CSIR; (b) no-CSI.

In this case, the optimal value of u is given by $u = \frac{1+r}{2}$ and the term inside the expectation in (20) becomes a convex function of r. Thus, the optimal solution of r can be generated by Newton's method. The achievability bound in the no-CSI case is Corollary 7, which is computed similar to that in Fig. 4. We observe from Fig. 7 that, as L increases, the spectral efficiency S_e can exceed 100, i.e., the number of active users that are reliably served can exceed the blocklength n, regardless of whether CSIR is available or not. In the case of CSIR, the spectral efficiency increases with L at an approximately constant speed, whereas the increasing speed gradually reduces in the no-CSI case due to the increased channel uncertainty. Additionally, as shown in Fig. 7a, in the case of CSIR, increasing energy-per-bit E_b is beneficial for different values of BS antennas, where the gap between the spectral efficiency for $E_b = 10$ dB and $E_b = 20$ dB increases as L increases. However, as observed in Fig. 7b, in the case of no-CSI, increasing energy-per-bit contributes only when K_a (or S_e) is small, in line with the results in Fig. 4. For both the achievability and the converse bounds, the gap between the spectral efficiency for $E_b = 10$ dB and $E_b = 20$ dB vanishes to zero as K_a grows large, suffering from channel uncertainty in such a worse interference environment.

In the case of no-CSI, the gap between achievability and converse bounds on the spectral efficiency per antenna is less than 0.13 bit/s/Hz, regardless of whether $E_b = 10$ dB or $E_b = 20$ dB.

V. Conclusion

Supporting the transmission of short packets under stringent latency and energy constraints is critically required for next-generation wireless communication networks. In this paper, we have considered such a communication system with finite blocklength and payload size. Under the PUPE criterion, we have established non-asymptotic achievability and converse bounds on the minimum required energy-per-bit for massive random access in MIMO quasi-static Rayleigh fading channels, with and without *a priori* CSI at the receiver. In the case of no-CSI, we consider both the settings with and without the knowledge of the number K_a of active users at the receiver. One key ingredient of the achievability bounds is the selection of an appropriate "good region". Numerical results demonstrate the tightness of our bounds. Specifically, the gap between the achievability and converse bounds is less than 2.5 dB for the CSIR case and less than 4 dB for the no-CSI case in most considered regimes. The no-CSI achievability and converse bounds show that the extra required energy-per-bit due to the uncertainty of the exact value of K_a is small in the considered regime, under the condition that the distribution of K_a is known *a priori*. The almost perfect MUI cancellation effect for the number of active users below a certain threshold, which was previously observed in AWGN channels [7] and single-receive-antenna quasi-static fading channels [8], is prominent in multiple-receive-antenna quasi-static Rayleigh fading channels with CSIR, although there is a slow growth of the energy-per-bit as the number of active users increases within this range in the no-CSI case. Additionally, our results show the significance of MIMO for the massive random access problem. As an example, in our considered regime, the spectral efficiency grows approximately linearly with the number of BS antennas in the CSIR case, but the lack of CSI at the receiver causes a slowdown in the growth rate. Furthermore, in the case of no-CSI, we have demonstrated the suboptimality of the pilot-assisted scheme, especially when there are many users. Overall, we believe our non-asymptotic bounds provide theoretical benchmarks to evaluate practical schemes, and are of considerable importance in massive random access systems.

Building on these non-asymptotic bounds, assuming $n \to \infty$ and $J = \Theta(1)$, we have obtained scaling laws of the number of reliably served users for a special case where all users are active. For the CSIR case, assuming $K \to \infty$, $\ln K = o(n)$, and $KP = \Omega(1)$, the PUPE requirement is satisfied if and only if $\frac{nL\ln KP}{K} = \Omega(1)$, i.e., if and only if one of the following two relations is satisfied: 1) $\frac{nL}{K} = \Omega(1)$ and $KP = \Theta(1)$; 2) $\frac{nL\ln KP}{K} = \Omega(1)$ and $KP \to \infty$. The first regime is power-limited and the second regime is degrees-of-freedom-limited. The condition $\frac{nL\ln KP}{K} = \Omega(1)$ shows the great potential of multiple receive antennas to considerably increase the number of reliably served users and reduce the required power P and blocklength n. For the no-CSI case, we observe a significant difference in the required number of BS antennas between utilizing the PUPE criterion and the joint error probability criterion. Specifically, in order to reliably serve $K = \mathcal{O}(n^2)$ users with power $P = \Theta\left(\frac{1}{n^2}\right)$, the required number of BS antennas is reduced from $L = \Theta\left(n^2 \ln n\right)$ to $L = \Theta\left(n^2\right)$ when we change from the joint error probability criterion to the PUPE criterion. Notably, as presented in Table I, our scaling laws consider the regime in which the energy-per-bit is finite or goes to 0, which are crucial in practical communication systems with stringent energy constraints.

Appendix A
A General Upper Bound on the PUPE Based on Fano's Bounding Technique

In this appendix, we provide a general upper bound on the PUPE applying Fano's "good region" technique, which is applicable for both CSIR and no-CSI cases. This bound is derived under the assumption that K_a is known at the receiver beforehand, and it can be extended to the case without known K_a as introduced in Appendix G.

We use a random coding scheme. Specifically, we generate a Gaussian codebook of size M and length n for each user independently. Let $\mathcal{C}_k = \{\mathbf{c}_{k,1}, \mathbf{c}_{k,2}, \ldots, \mathbf{c}_{k,M}\}$ denote the codebook of user k without power constraint, where $\mathbf{c}_{k,m} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, P'\mathbf{I}_n)$ for $m \in [M]$ and $k \in \mathcal{K}$. We choose $P' < P$ to ensure that we can control the maximum power constraint violation events. Let $\mathbf{A} \in \mathbb{C}^{n \times MK}$ denote the concatenation of codebooks of the K users without power constraint. If user k is active, let its transmitted codeword be $\mathbf{x}_{(k)} = \mathbf{c}_{(k)} 1\left\{\left\|\mathbf{c}_{(k)}\right\|_2^2 \leq nP\right\}$, where $\mathbf{c}_{(k)} = \mathbf{c}_{k,W_k}$ with the message $W_k \in [M]$ chosen uniformly at random; if user k is inactive, let $\mathbf{x}_{(k)} = \mathbf{c}_{(k)} = \mathbf{0}$.

The decoder aims to find the estimated set $\hat{\mathcal{K}}_a$ of active users, and find the estimate $\hat{\mathbf{c}}_{(k)}$ of $\mathbf{c}_{(k)}$ and corresponding message $\hat{W}_k$ of W_k for $k \in \hat{\mathcal{K}}_a$. Let $\hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]} = \left\{\hat{\mathbf{c}}_{(k)} \in \mathcal{C}_k : k \in \hat{\mathcal{K}}_a\right\}$. The outputs of the decoder are given by

$$\left[\hat{\mathcal{K}}_a, \hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right] = \arg \min_{\hat{\mathcal{K}}_a \subset \mathcal{K}, \left|\hat{\mathcal{K}}_a\right| = K_a} \min_{\left(\hat{\mathbf{c}}_{(k)} \in \mathcal{C}_k\right)_{k \in \hat{\mathcal{K}}_a}} g\left(\mathbf{Y}, \hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right), \tag{102}$$

$$\hat{W}_k = f_{\text{en},k}^{-1}\left(\hat{\mathbf{c}}_{(k)}\right), \quad k \in \hat{\mathcal{K}}_a, \tag{103}$$

where $g\left(\mathbf{Y}, \hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right)$ denotes the decoding metric. We have $\hat{W}_k = 0$ and $\hat{\mathbf{c}}_{(k)} = \mathbf{0}$ for $k \notin \hat{\mathcal{K}}_a$.

The PUPE in (4) can be upper-bounded as

$$P_e \leq p_0 + \mathbb{E}\left[\frac{1}{K_a} \sum_{k \in \mathcal{K}_a} 1\left[W_k \neq \hat{W}_k\right]\right]_{\text{no power constraint}} \tag{104}$$

$$= p_0 + \sum_{t=1}^{K_a} \frac{t}{K_a} \mathbb{P}\left[\mathcal{F}_t\right]_{\text{no power constraint}}, \tag{105}$$

where (104) follows because we change the measure $\mathbf{x}_{(k)} = \mathbf{c}_{(k)} 1\left\{ \left\| \mathbf{c}_{(k)} \right\|_2^2 \leq nP \right\}$ with power constraint to $\mathbf{x}_{(k)} = \mathbf{c}_{(k)}$ without power constraint by adding a total variation distance upper-bounded by p_0 [8]. Here, p_0 is given by

$$p_0 = K_a \mathbb{P}\left[\left\| \mathbf{c}_{(k)} \right\|_2^2 > nP \right] \tag{106}$$

$$= K_a \left(1 - \frac{\gamma\left(n, \frac{nP}{P'}\right)}{\Gamma(n)} \right), \tag{107}$$

which follows from the fact that $\left\| \mathbf{c}_{(k)} \right\|_2^2 \sim \frac{P'}{2}\chi^2(2n)$; $\mathcal{F}_t = \left\{ \sum_{k \in \mathcal{K}_a} 1\left\{ W_k \neq \hat{W}_k \right\} = t \right\}$ indicates the event that there are exactly t misdecoded users. In what follows, we omit the subscript "no power constraint" for simplicity and upper-bound $\mathbb{P}\left[\mathcal{F}_t\right]$ applying Fano's "good region" technique [29].

Let the set $S_1 \subset \mathcal{K}_a$ of size t denote the set of users whose codewords are misdecoded. Let the set $S_2 \subset \mathcal{K} \backslash \mathcal{K}_a \cup S_1$ of size t denote the set of detected users with false alarm codewords. For the sake of simplicity, we rewrite "$\bigcup_{S_1 \subset \mathcal{K}_a, |S_1| = t}$" to "$\bigcup_{S_1}$" and "$\bigcup_{S_2 \subset \mathcal{K} \backslash \mathcal{K}_a \cup S_1, |S_2| = t}$" to "$\bigcup_{S_2}$"; and similarly for $\sum$ and $\bigcap$. We use $\mathbf{c}_{[S]} = \left\{ \mathbf{c}_{(k)} : k \in S \right\}$ to denote the set of transmitted codewords corresponding to users in the set $S \subset \mathcal{K}_a$, and use $\mathbf{c}'_{[S_2]} = \left\{ \mathbf{c}'_{(k)} \in \mathcal{C}_k : k \in S_2, \mathbf{c}'_{(k)} \neq \mathbf{c}_{(k)} \right\}$ to denote the set of false alarm codewords corresponding to users in the set $S_2 \subset \mathcal{K}$. Recall that for massive random access in MIMO fading channels, the "good region" $\mathcal{R}_{t,S_1}$ is given in (14) for any subset $S_1 \subset \mathcal{K}_a$ of size t. We define the event $\mathcal{G}_{\omega,\nu} = \bigcap_{S_1} \left\{ \mathbf{Y} \in \mathcal{R}_{t,S_1} \right\}$. Then, we obtain

$$\mathbb{P}\left[\mathcal{F}_t\right]$$

$$\leq \mathbb{P}\left[\bigcup_{S_1} \bigcup_{S_2} \bigcup_{\mathbf{c}'_{[S_2]}} \left\{ g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]} \cup \mathbf{c}'_{[S_2]} \right) \leq g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a]} \right) \right\} \right] \tag{108}$$

$$\leq \min_{\substack{0 \leq \omega \leq 1 \\ \nu \geq 0}} \left\{ \mathbb{P}\left[\bigcup_{S_1} \bigcup_{S_2} \bigcup_{\mathbf{c}'_{[S_2]}} \left\{ g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]} \cup \mathbf{c}'_{[S_2]} \right) \leq g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a]} \right) \right\} \right.\right.$$
$$\left.\left. \bigcap \mathcal{G}_{\omega,\nu} \right] + \mathbb{P}\left[\mathcal{G}_{\omega,\nu}^c \right] \right\}, \tag{109}$$

where (109) follows from Fano's bounding technique given in (12).

The first probability on the RHS of (109) can be upper-bounded as

$$\mathbb{P}\left[\bigcup_{S_1} \bigcup_{S_2} \bigcup_{\mathbf{c}'_{[S_2]}} \left\{ g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]} \cup \mathbf{c}'_{[S_2]} \right) \leq g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a]} \right) \right\} \bigcap \mathcal{G}_{\omega,\nu} \right]$$
$$\leq \sum_{S_1} \sum_{S_2} \sum_{\mathbf{c}'_{[S_2]}} \mathbb{E}_{\mathbf{c}_{[\mathcal{K}_a]}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]}, \mathbf{c}'_{[S_2]}} \left[\min_{u \geq 0, r \geq 0} \mathbb{P}\left[(u - r) g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a]} \right) \right.\right.$$
$$- u g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]} \cup \mathbf{c}'_{[S_2]} \right) + r \omega g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]} \right)$$
$$\left.\left. + r \nu n L \geq 0 \,\middle|\, \mathbf{c}_{[\mathcal{K}_a]}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]}, \mathbf{c}'_{[S_2]} \right] \right] \tag{110}$$

$$\leq \sum_{S_1} \sum_{S_2} \sum_{\mathbf{c}'_{[S_2]}} \mathbb{E}_{\mathbf{c}_{[\mathcal{K}_a]}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]}, \mathbf{c}'_{[S_2]}} \left[\min_{u \geq 0, r \geq 0} \exp\left\{ r \nu n L \right\} \right.$$
$$\cdot \mathbb{E}_{\mathbf{H},\mathbf{Z}}\left[\exp\left\{ (u - r) g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a]} \right) - u g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]} \cup \mathbf{c}'_{[S_2]} \right) \right.\right.$$
$$\left.\left.\left. + r \omega g\left(\mathbf{Y}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]} \right) \right\} \,\middle|\, \mathbf{c}_{[\mathcal{K}_a]}, \mathbf{c}_{[\mathcal{K}_a \backslash S_1]}, \mathbf{c}'_{[S_2]} \right] \right], \tag{111}$$

where (110) follows from the union bound and the fact that $\mathbb{P}[\{a \geq 0\} \cap \{b \geq 0\}] \leq \mathbb{P}[a + b \geq 0]$; (111) follows by applying the Chernoff bound in Lemma 14 shown below to the conditional probability in (110).

Lemma 14 (Section 3.2.4 in [31]): Let Z and W be any random variables. Then we have

$$\mathbb{P}\left[Z \geq 0, W \leq 0 \right] \leq \mathbb{E}\left[\exp\left\{ sZ - rW \right\} \right], \quad \forall s \geq 0, \;\; r \geq 0, \tag{112}$$

and

$$\mathbb{P}\left[W > 0 \right] \leq \mathbb{E}\left[\exp\left\{ sW \right\} \right], \quad \forall s \geq 0. \tag{113}$$

We obtain an upper bound on $\mathbb{P}\left[\mathcal{F}_t\right]$ by substituting (111) into (109). Together with (105), we obtain a general upper bound on the PUPE. Note that in both cases of CSIR and no-CSI, the expectations in (111) and the probability $\mathbb{P}\left[\mathcal{G}_{\omega,\nu}^c\right]$ on the RHS of (109) can be further bounded.

Appendix B
Proof of Theorem 1

In this appendix, we prove Theorem 1 to derive an upper bound on the PUPE with known K_a and CSIR. Based on the notation in Appendix A, the ML decoding metric in this case is given by

$$g\left(\mathbf{Y}, \hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]} \right) = \sum_{l=1}^{L} \left\| \mathbf{y}_l - \sum_{k \in \hat{\mathcal{K}}_a} h_{k,l} \hat{\mathbf{c}}_{(k)} \right\|_2^2. \tag{114}$$

As introduced in Appendix A, the PUPE can be upper-bounded by (105). The probability $\mathbb{P}\left[\mathcal{F}_t\right]$ therein, i.e. the probability of the event that there are exactly t misdecoded users, is upper-bounded in (109) applying Fano's "good region" technique. In the following Appendix B-A, we particularize the "good region"-based bound on $\mathbb{P}\left[\mathcal{F}_t\right]$ given in Appendix A to the case of CSIR; then, in Appendix B-B, we derive another upper bound on $\mathbb{P}\left[\mathcal{F}_t\right]$ applying Gallager's error exponent analysis [30]. The two upper bounds are denoted as $p_{1,t}$ and $p_{2,t}$, respectively.

A. Upper-Bounding $\mathbb{P}\left[\mathcal{F}_t\right]$ Based on Fano's Bounding Technique

In this subsection, we particularize the "good region"-based bound on $\mathbb{P}\left[\mathcal{F}_t\right]$ in (109) to the CSIR case, followed by further manipulations on the two probabilities on the RHS of (109).

Based on the notation in Appendix A, we have $|S_1 \cap S_2| = t_0 \in [0, t]$. Let the binary selection matrix $\mathbf{\Phi}_{S_1} \in \{0,1\}^{MK \times K}$ indicate which codewords are transmitted by users in the set $S_1 \subset \mathcal{K}_a$. Let the binary selection matrix $\mathbf{\Phi}'_{S_2} \in \{0,1\}^{MK \times K}$ indicate which codewords are not transmitted but decoded for users in the set $S_2 \subset \mathcal{K}\backslash\mathcal{K}_a \cup S_1$. It is satisfied that $[\mathbf{\Phi}_{S_1}]_{(k-1)M+W_k,k} = 1$ if user $k \in S_1$ is active and the W_k-th codeword is transmitted by it, and $[\mathbf{\Phi}_{S_1}]_{(k-1)M+W_k,k} = 0$ otherwise; and similarly for $\mathbf{\Phi}^{'}_{S_2}$. Let $\tilde{\mathbf{A}}_{S_1} = \mathbf{A}\mathbf{\Phi}_{S_1} \in \mathbb{C}^{n\times K}$ and $\tilde{\mathbf{A}}^{'}_{S_2} = \mathbf{A}\mathbf{\Phi}^{'}_{S_2} \in \mathbb{C}^{n\times K}$. The conditional expectation in (111) can be written as

$$\begin{aligned}
&\mathbb{E}_{\mathbf{H},\mathbf{Z}}\Big[\exp\Big\{(u-r)g\big(\mathbf{Y},\mathbf{c}_{[\mathcal{K}_a]}\big) - ug\Big(\mathbf{Y},\mathbf{c}_{[\mathcal{K}_a\backslash S_1]}\cup\mathbf{c}^{'}_{[S_2]}\Big) \\
&\qquad +r\omega g\big(\mathbf{Y},\mathbf{c}_{[\mathcal{K}_a\backslash S_1]}\big)\Big\}\Big|\,\mathbf{c}_{[\mathcal{K}_a]},\mathbf{c}_{[\mathcal{K}_a\backslash S_1]},\mathbf{c}^{'}_{[S_2]}\Big] \\
&= \mathbb{E}_{\mathbf{H},\mathbf{Z}}\Big[\exp\Big\{(u-r)\left\|\mathbf{Z}\right\|_F^2 - u\left\|\mathbf{Z}+\left(\tilde{\mathbf{A}}_{S_1}-\tilde{\mathbf{A}}^{'}_{S_2}\right)\mathbf{H}\right\|_F^2 \\
&\qquad +r\omega\left\|\mathbf{Z}+\tilde{\mathbf{A}}_{S_1}\mathbf{H}\right\|_F^2\Big\}\Big|\,\tilde{\mathbf{A}}_{S_1},\tilde{\mathbf{A}}^{'}_{S_2}\Big] &(115) \\
&= (1+r(1-\omega))^{-nL}\,\mathbb{E}_{\mathbf{H}}\Big[\exp\Big\{-u\left\|\left(\tilde{\mathbf{A}}_{S_1}-\tilde{\mathbf{A}}^{'}_{S_2}\right)\mathbf{H}\right\|_F^2 \\
&\qquad +\frac{1}{1+r(1-\omega)}\left\|\left((u-r\omega)\tilde{\mathbf{A}}_{S_1}-u\tilde{\mathbf{A}}^{'}_{S_2}\right)\mathbf{H}\right\|_F^2 \\
&\qquad +r\omega\left\|\tilde{\mathbf{A}}_{S_1}\mathbf{H}\right\|_F^2\Big\}\Big|\,\tilde{\mathbf{A}}_{S_1},\tilde{\mathbf{A}}^{'}_{S_2}\Big] &(116) \\
&= (1+r(1-\omega))^{-nL}\exp\Big\{-L\ln\left|\mathbf{I}_K+\tilde{\mathbf{B}}\right|\Big\}. &(117)
\end{aligned}$$

Here, (116) follows from Lemma 15 provided below by taking the expectation over $\mathbf{Z}$; (117) also follows from Lemma 15 by taking the expectation over $\mathbf{H}$, under the condition that the minimum eigenvalue of $\tilde{\mathbf{B}}$ satisfies $\lambda_{\min}\left(\tilde{\mathbf{B}}\right) > -1$, where $\tilde{\mathbf{B}}$ is given by

$$\begin{aligned}
\tilde{\mathbf{B}} =& \frac{(1+r-u)(u-r\omega)}{1+r(1-\omega)}\left(\tilde{\mathbf{A}}_{S_1}-\frac{u}{u-r\omega}\tilde{\mathbf{A}}^{'}_{S_2}\right)^H \\
&\cdot\left(\tilde{\mathbf{A}}_{S_1}-\frac{u}{u-r\omega}\tilde{\mathbf{A}}^{'}_{S_2}\right) - \frac{r\omega u}{u-r\omega}\left(\tilde{\mathbf{A}}^{'}_{S_2}\right)^H\tilde{\mathbf{A}}^{'}_{S_2}. \qquad (118)
\end{aligned}$$

Lemma 15 (Corollary 3.2a.2 in [45], Result 4.4.1 in [46]): Let the vector $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu},\mathbf{\Sigma})$ with $\boldsymbol{\mu} \in \mathbb{R}^{p\times 1}$ and $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$. Let $\mathbf{D} \in \mathbb{R}^{p\times p}$ be a symmetric matrix. For any γ, if the eigenvalues of the matrix $\mathbf{I}_p - 2\gamma\mathbf{\Sigma}\mathbf{D}$ are positive, the expectation $\mathbb{E}\left[\exp\left\{\gamma\mathbf{x}^T\mathbf{D}\mathbf{x}\right\}\right]$ is given by

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{\gamma\mathbf{x}^T\mathbf{D}\mathbf{x}\right\}\right] \\
&= \left|\mathbf{I}_p - 2\gamma\mathbf{\Sigma}\mathbf{D}\right|^{-\frac{1}{2}}\exp\left\{\gamma\boldsymbol{\mu}^T\mathbf{D}(\mathbf{I}_p-2\gamma\mathbf{\Sigma}\mathbf{D})^{-1}\boldsymbol{\mu}\right\}. \qquad (119)
\end{aligned}$$

In particular, if $\mathbf{x} \sim \mathcal{N}(\mathbf{0},\mathbf{\Sigma})$, we have

$$\mathbb{E}\left[\exp\left\{\gamma\mathbf{x}^T\mathbf{D}\mathbf{x}\right\}\right] = \left|\mathbf{I}_p - 2\gamma\mathbf{\Sigma}\mathbf{D}\right|^{-\frac{1}{2}}. \qquad (120)$$

Let $\bar{\mathbf{x}} \in \mathbb{C}^{p\times 1}$ be a complex random vector distributed as $\bar{\mathbf{x}} \sim \mathcal{CN}\left(\mathbf{0},\bar{\mathbf{\Sigma}}\right)$. Let $\mathbf{B} \in \mathbb{C}^{p\times p}$ be a Hermitian matrix. For any γ, if the eigenvalues of the matrix $\mathbf{I}_p - \gamma\bar{\mathbf{\Sigma}}\mathbf{B}$ are positive, the expectation $\mathbb{E}\left[\exp\left\{\gamma\bar{\mathbf{x}}^H\mathbf{B}\bar{\mathbf{x}}\right\}\right]$ is given by

$$\mathbb{E}\left[\exp\left\{\gamma\bar{\mathbf{x}}^H\mathbf{B}\bar{\mathbf{x}}\right\}\right] = \left|\mathbf{I}_p - \gamma\bar{\mathbf{\Sigma}}\mathbf{B}\right|^{-1}. \qquad (121)$$

If $\bar{\mathbf{x}} \sim \mathcal{CN}\left(\bar{\boldsymbol{\mu}},\mathbf{I}_p\right)$ with $\bar{\boldsymbol{\mu}} \in \mathbb{C}^{p\times 1}$ and $\gamma < 1$, the expectation $\mathbb{E}\left[\exp\left\{\gamma\bar{\mathbf{x}}^H\bar{\mathbf{x}}\right\}\right]$ is given by

$$\mathbb{E}\left[\exp\left\{\gamma\bar{\mathbf{x}}^H\bar{\mathbf{x}}\right\}\right] = (1-\gamma)^{-p}\exp\left\{\frac{\gamma}{1-\gamma}\bar{\boldsymbol{\mu}}^H\bar{\boldsymbol{\mu}}\right\}. \qquad (122)$$

Substituting (117) into (111), we have

$$\begin{aligned}
&\mathbb{P}\Bigg[\bigcup_{S_1}\bigcup_{S_2}\bigcup_{\mathbf{c}^{'}_{[S_2]}}\Big\{g\Big(\mathbf{Y},\mathbf{c}_{[\mathcal{K}_a\backslash S_1]}\cup\mathbf{c}^{'}_{[S_2]}\Big)\leq g\big(\mathbf{Y},\mathbf{c}_{[\mathcal{K}_a]}\big)\Big\}\bigcap\mathcal{G}_{\omega,\nu}\Bigg] \\
&\leq \sum_{t_0=0}^{t} C_{t_0,t}\,\mathbb{E}_{\tilde{\mathbf{A}}_{S_1},\tilde{\mathbf{A}}^{'}_{S_2}}\Bigg[\min_{u\geq 0,r\geq 0,\lambda_{\min}\left(\tilde{\mathbf{B}}\right)>-1}(1+r(1-\omega))^{-nL} \\
&\qquad\qquad \cdot\exp\Big\{r\nu nL - L\ln\left|\mathbf{I}_K+\tilde{\mathbf{B}}\right|\Big\}\Bigg], \qquad (123)
\end{aligned}$$

where $C_{t_0,t} = \binom{K_a}{t}\binom{t}{t_0}\binom{K-K_a}{t-t_0}(M-1)^{t_0}M^{t-t_0}$. Here, (123) follows because the expectation over $\tilde{\mathbf{A}}_{S_1}$ and $\tilde{\mathbf{A}}^{'}_{S_2}$ is unchanged for different S_1, S_2, and $\mathbf{c}^{'}_{[S_2]}$ once t_0 and t are fixed, considering that the codebook matrix $\mathbf{A}$ has i.i.d. $\mathcal{CN}(0,P')$ entries. As a result, the first probability on the RHS of (109) is upper-bounded by (123), which is denoted as $q_{1,t}(\omega,\nu)$ as presented in (20).

In the following, we derive an upper bound $q_{2,t}(\omega,\nu)$ on the second term $\mathbb{P}\left[\mathcal{G}^c_{\omega,\nu}\right]$ on the RHS of (109). To obtain $q_{2,t}(\omega,\nu)$, we upper-bound $\mathbb{P}\left[\mathcal{G}^c_{\omega,\nu}\right]$ for the case of $t < n$ and $\omega \in (0,1]$, the case of $t \geq n$ and $\omega \in (0,1]$, and the case of $\omega = 0$, respectively.

Case 1: $t < n, \omega \in (0,1]$.

Let $\tilde{\mathbf{y}}_l = \mathbf{z}_l + \tilde{\mathbf{A}}_{S_1}\mathbf{h}_l$. We define the event $\mathcal{G}_\eta = \bigcap_{S_1}\left\{\sum_{l=1}^{L}\left\|\mathcal{P}_{\mathbf{c}_{[S_1]}}\mathbf{z}_l\right\|_2^2 \leq tL(1+\eta)\right\}$ for $\eta \geq 0$. We can bound $\mathbb{P}\left[\mathcal{G}^c_{\omega,\nu}\right]$ as

$$\begin{aligned}
&\mathbb{P}\left[\mathcal{G}^c_{\omega,\nu}\right] \\
&= \mathbb{P}\left[\bigcup_{S_1}\left\{\sum_{l=1}^{L}\left\|\mathbf{z}_l\right\|_2^2 > \omega\sum_{l=1}^{L}\left\|\tilde{\mathbf{y}}_l\right\|_2^2 + nL\nu\right\}\right] &(124) \\
&= \mathbb{P}\Bigg[\bigcup_{S_1}\Bigg\{\sum_{l=1}^{L}\left\|\mathcal{P}_{\mathbf{c}_{[S_1]}}\mathbf{z}_l\right\|_2^2 + \sum_{l=1}^{L}\left\|\mathcal{P}^{\perp}_{\mathbf{c}_{[S_1]}}\tilde{\mathbf{y}}_l\right\|_2^2 \\
&\qquad > \omega\sum_{l=1}^{L}\left\|\tilde{\mathbf{y}}_l\right\|_2^2 + nL\nu\Bigg\}\Bigg] &(125) \\
&\leq \sum_{S_1}\mathbb{P}\left[\sum_{l=1}^{L}\tilde{\mathbf{y}}_l^H\left(\mathcal{P}^{\perp}_{\mathbf{c}_{[S_1]}}-\omega\mathbf{I}_n\right)\tilde{\mathbf{y}}_l > nL\nu - tL(1+\eta)\right] \\
&\quad + \mathbb{P}\left[\mathcal{G}^c_\eta\right], &(126)
\end{aligned}$$

where (125) follows due to $\left\|\mathbf{z}_l\right\|_2^2 = \left\|\mathcal{P}_{\mathbf{c}_{[S_1]}}\mathbf{z}_l\right\|_2^2 + \left\|\mathcal{P}^{\perp}_{\mathbf{c}_{[S_1]}}\mathbf{z}_l\right\|_2^2$ and $\mathcal{P}^{\perp}_{\mathbf{c}_{[S_1]}}\tilde{\mathbf{y}}_l = \mathcal{P}^{\perp}_{\mathbf{c}_{[S_1]}}\mathbf{z}_l$; (126) follows from the bounding technique in (12) and the union bound.

Next, we focus on two terms on the RHS of (126). We have $\mathcal{P}_{\mathbf{c}_{[S_1]}}\mathbf{z}_l \overset{i.i.d.}{\sim} \mathcal{CN}\left(\mathbf{0},\mathcal{P}_{\mathbf{c}_{[S_1]}}\right)$ for $l \in [L]$ conditioned on $\mathbf{c}_{[S_1]}$. Let $\mathbf{U}$ be a unitary matrix satisfying $\mathbf{U}\mathcal{P}_{\mathbf{c}_{[S_1]}}\mathbf{U}^H = \mathbf{I}_{(t)}$. Conditioned on $\mathbf{c}_{[S_1]}$, we have $\mathbf{U}\mathcal{P}_{\mathbf{c}_{[S_1]}}\mathbf{z}_l \sim \mathcal{CN}\left(\mathbf{0},\mathbf{I}_{(t)}\right)$, which implies that $\left\|\mathcal{P}_{\mathbf{c}_{[S_1]}}\mathbf{z}_l\right\|_2^2 = \left\|\mathbf{U}\mathcal{P}_{\mathbf{c}_{[S_1]}}\mathbf{z}_l\right\|_2^2 \sim \frac{1}{2}\chi^2(2t)$ and $\sum_{l=1}^{L}\left\|\mathcal{P}_{\mathbf{c}_{[S_1]}}\mathbf{z}_l\right\|_2^2 \sim \frac{1}{2}\chi^2(2tL)$. Hence, we can obtain

$$\mathbb{P}\left[\mathcal{G}^c_\eta\right] \leq \sum_{S_1}\mathbb{P}\left[\sum_{l=1}^{L}\left\|\mathcal{P}_{\mathbf{c}_{[S_1]}}\mathbf{z}_l\right\|_2^2 > tL(1+\eta)\right] \qquad (127)$$

$$= \binom{K_a}{t} \left(1 - \frac{\gamma\left(tL, tL\ \ (1+\eta)\right)}{\Gamma\left(tL\right)}\right). \tag{128}$$

We can bound the first term on the RHS of (126) as follows. Let $\mathbf{F}_{S_1} = \mathbf{I}_n + \tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H$ which can be decomposed as $\mathbf{F}_{S_1} = \mathbf{F}_{S_1}^{\frac{1}{2}}\mathbf{F}_{S_1}^{\frac{H}{2}}$. Conditioned on $\tilde{\mathbf{A}}_{S_1}$, we have $\tilde{\mathbf{y}}_l = \mathbf{F}_{S_1}^{\frac{1}{2}}\tilde{\mathbf{w}}_l \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{F}_{S_1}\right)$ where $\tilde{\mathbf{w}}_l \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_n\right)$. Define the event $\mathcal{G}_\delta = \left\{\sum_{l=1}^{L} \tilde{\mathbf{w}}_l^H \tilde{\mathbf{w}}_l < (1+\delta)nL\right\}$ for $\delta \geq 0$. Applying the bounding technique in (12), we can bound the first probability on the RHS of (126) as

$$\begin{aligned}
&q_{3,t}\left(\omega,\nu\right)\\
&= \mathbb{P}\left[\sum_{l=1}^{L}\tilde{\mathbf{w}}_l^H\mathbf{F}_{S_1}^{\frac{H}{2}}\left(\mathcal{P}_{\mathbf{c}_{[S_1]}}^{\perp} - \omega\mathbf{I}_n\right)\mathbf{F}_{S_1}^{\frac{1}{2}}\tilde{\mathbf{w}}_l > nL\nu - tL(1+\eta)\right] &(129)\\
&\leq \mathbb{P}\left[\left\{\sum_{l=1}^{L}\tilde{\mathbf{w}}_l^H\mathbf{F}_{S_1}^{\frac{H}{2}}\left(\mathcal{P}_{\mathbf{c}_{[S_1]}}^{\perp} - \omega\mathbf{I}_n\right)\mathbf{F}_{S_1}^{\frac{1}{2}}\tilde{\mathbf{w}}_l > nL\nu - tL(1+\eta)\right\}\right.\\
&\qquad \left.\cap\ \mathcal{G}_\delta\right] + \mathbb{P}\left[\mathcal{G}_\delta^c\right] &(130)\\
&= q_{4,t}\left(\omega,\nu\right) + 1 - \frac{\gamma\left(nL, nL\ \ (1+\delta)\right)}{\Gamma\left(nL\right)}. &(131)
\end{aligned}$$

The probability $q_{4,t}\left(\omega,\nu\right)$ in (131) can be further upper-bounded as

$$\begin{aligned}
&q_{4,t}\left(\omega,\nu\right)\\
&= \mathbb{E}_{\tilde{\mathbf{A}}_{S_1}}\left[\mathbb{P}\left[\left\{\sum_{l=1}^{L}\tilde{\mathbf{w}}_l^H\left((1-\omega)\mathbf{I}_n - \mathcal{P}_{\mathbf{c}_{[S_1]}} - \omega\tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H\right)\tilde{\mathbf{w}}_l\right.\right.\right.\\
&\qquad\qquad \left.\left.\left. > nL\nu - tL(1+\eta)\right\} \cap \mathcal{G}_\delta \middle| \tilde{\mathbf{A}}_{S_1}\right]\right] &(132)\\
&\leq \mathbb{E}_{\tilde{\mathbf{A}}_{S_1}}\left[\mathbb{P}\left[\sum_{l=1}^{L}\tilde{\mathbf{w}}_l^H\left(\mathcal{P}_{\mathbf{c}_{[S_1]}} + \omega\tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H\right)\tilde{\mathbf{w}}_l < Lt(1+\eta)\right.\right.\\
&\qquad\qquad \left.\left. + L(-n\nu + (1+\delta)n(1-\omega)) \middle| \tilde{\mathbf{A}}_{S_1}\right]\right], &(133)
\end{aligned}$$

where (132) holds because the eigenvalues of the matrix $(1-\omega)\mathbf{I}_n - \mathcal{P}_{\mathbf{c}_{[S_1]}} - \omega\tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H$ are the same as that of the matrix $\mathbf{F}_{S_1}^{\frac{H}{2}}\left(\mathcal{P}_{\mathbf{c}_{[S_1]}}^{\perp} - \omega\mathbf{I}_n\right)\mathbf{F}_{S_1}^{\frac{1}{2}}$. Denote $c' = t(1+\eta) - n\nu + (1+\delta)n(1-\omega)$. Then, the conditional probability in (133) can be upper-bounded as

$$\begin{aligned}
&\mathbb{P}\left[\sum_{l=1}^{L}\tilde{\mathbf{w}}_l^H\left(\mathcal{P}_{\mathbf{c}_{[S_1]}} + \omega\tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H\right)\tilde{\mathbf{w}}_l < Lc' \middle| \tilde{\mathbf{A}}_{S_1}\right]\\
&= \mathbb{P}\left[\sum_{i=1}^{t}\frac{\lambda_i\chi_i^2(2L)}{2L} < c' \middle| \tilde{\mathbf{A}}_{S_1}\right] &(134)\\
&\leq \mathbb{P}\left[\frac{\chi^2(2tL)}{2tL} < \frac{c'}{t\prod_{i=1}^{t}\lambda_i^{\frac{1}{t}}} \middle| \tilde{\mathbf{A}}_{S_1}\right] &(135)\\
&= \frac{\gamma\left(tL, L\ c'\left|\mathbf{I}_n + \omega\tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H\right|^{-\frac{1}{t}}\right)}{\Gamma\left(tL\right)}, &(136)
\end{aligned}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_t$ are non-zero eigenvalues of the matrix $\mathcal{P}_{\mathbf{c}_{[S_1]}} + \omega\tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H$. Here, (134) follows because conditioned on $\tilde{\mathbf{A}}_{S_1}$, the random vectors $\mathcal{U}\tilde{\mathbf{w}}_l$ and $\tilde{\mathbf{w}}_l$ have the same distribution as $\mathcal{CN}\left(\mathbf{0}, \mathbf{I}_n\right)$ for a unitary matrix $\mathcal{U}$ satisfying $\mathcal{U}^H\left(\mathcal{P}_{\mathbf{c}_{[S_1]}} + \omega\tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H\right)\mathcal{U} = \operatorname{diag}\{\lambda_1, \ldots, \lambda_t, 0, \ldots, 0\} \in \mathbb{R}^{n\times n}$; (135) follows from Lemma 16 shown below; (136) follows because $\prod_{i=1}^{t}\lambda_i = \left|\mathbf{I}_n + \omega\tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H\right|$. Together with (133), we can obtain

$$q_{4,t}(\omega,\nu) \leq \mathbb{E}_{\tilde{\mathbf{A}}_{S_1}}\left[\frac{\gamma\left(tL, L\ c'\left|\mathbf{I}_n + \omega\tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H\right|^{-\frac{1}{t}}\right)}{\Gamma\left(tL\right)}\right]. \tag{137}$$

Lemma 16 ([47]): Assume $x_1, \ldots, x_s$ are independently distributed chi-square variables with m degrees of freedom. Assume $\tilde{x} \sim \chi^2\left(sm\right)$. Let the constant $\gamma_j > 0$. Then, for every constant c,

$$\mathbb{P}\left(\sum\nolimits_{j=1}^{s}\gamma_j x_j < c\right) \leq \mathbb{P}\left(\prod\nolimits_{j=1}^{s}\gamma_j^{\frac{1}{s}}\tilde{x} < c\right). \tag{138}$$

Substituting (127), (131), and (137) into (126), we can obtain an upper bound on $\mathbb{P}\left[\mathcal{G}_{\omega,\nu}^c\right]$ in the case of $t < n$ and $\omega \in (0, 1]$ as presented in (23).

Case 2: $t \geq n, \omega \in (0, 1]$.

Next, we upper-bound $\mathbb{P}\left[\mathcal{G}_{\omega,\nu}^c\right]$ when $t \geq n$ and $\omega \in (0, 1]$. Recall that $\tilde{\mathbf{y}}_l = \mathbf{z}_l + \tilde{\mathbf{A}}_{S_1}\mathbf{h}_l$. We define the event $\mathcal{G}_\eta = \left\{\sum_{l=1}^{L}\|\mathbf{z}_l\|_2^2 \leq nL(1+\eta)\right\}$ for $\eta \geq 0$. We can bound $\mathbb{P}\left[\mathcal{G}_{\omega,\nu}^c\right]$ as

$$\begin{aligned}
&\mathbb{P}\left[\mathcal{G}_{\omega,\nu}^c\right]\\
&\leq \mathbb{P}\left[\bigcup_{S_1}\left\{\sum_{l=1}^{L}\|\mathbf{z}_l\|_2^2 > \omega\sum_{l=1}^{L}\|\tilde{\mathbf{y}}_l\|_2^2 + nL\nu\right\}\bigcap\mathcal{G}_\eta\right] + \mathbb{P}\left[\mathcal{G}_\eta^c\right] &(139)\\
&\leq \mathbb{P}\left[\bigcup_{S_1}\left\{\sum_{l=1}^{L}\|\tilde{\mathbf{y}}_l\|_2^2 < nLC_{\omega,\nu,\eta}\right\}\right] + 1 - \frac{\gamma(nL, nL\ \ (1+\eta))}{\Gamma\left(nL\right)}, &(140)
\end{aligned}$$

where $C_{\omega,\nu,\eta} = \frac{1+\eta-\nu}{\omega}$. The first probability in the RHS of (140) can be further bounded as

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{S_1}\left\{\sum_{l=1}^{L}\|\tilde{\mathbf{y}}_l\|_2^2 < nLC_{\omega,\nu,\eta}\right\}\right]\\
&\leq \binom{K_a}{t}\mathbb{E}_{\tilde{\mathbf{A}}_{S_1}}\left[\frac{\gamma\left(nL, nLC_{\omega,\nu,\eta}\left|\mathbf{I}_n + \tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H\right|^{-\frac{1}{n}}\right)}{\Gamma\left(nL\right)}\right], &(141)
\end{aligned}$$

where (141) follows by applying Lemma 16 and from the fact that $\tilde{\mathbf{y}}_l \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_n + \tilde{\mathbf{A}}_{S_1}\tilde{\mathbf{A}}_{S_1}^H\right)$ conditioned on $\tilde{\mathbf{A}}_{S_1}$.

Case 3: $\omega = 0$.

In the case of $\omega = 0$, we have

$$\mathbb{P}\left[\mathcal{G}_{\omega,\nu}^c\right] = \mathbb{P}\left[\sum_{l=1}^{L}\|\mathbf{z}_l\|_2^2 > nL\nu\right] \tag{142}$$

$$= 1 - \frac{\gamma\left(nL, nL\ \nu\right)}{\Gamma\left(nL\right)}. \tag{143}$$

In conclusion, based on Fano's bounding technique, we have obtained $q_{1,t}\left(\omega,\nu\right)$ in (20) (i.e. an upper bound on the first term in (109)) and $q_{2,t}\left(\omega,\nu\right)$ in (23) (i.e. an upper bound on the second term in (109)), which contributes to an upper bound $p_{1,t}$ in (19) on the probability $\mathbb{P}\left[\mathcal{F}_t\right]$.

B. Upper-Bounding $\mathbb{P}\left[\mathcal{F}_t\right]$ Based on Gallager's Bounding Technique

Let $\mathbf{H}_1$ and $\mathbf{H}_2$ be $t \times L$ submatrices of $\mathbf{H}$ formed by rows corresponding to the support of S_1 and S_2, respectively. Let $\mathbf{A}_{S_1}$ and $\mathbf{A}'_{S_2}$ be $n \times t$ submatrices of $\mathbf{A}$ formed by columns corresponding to the codewords transmitted by users in the set S_1 and the codewords not transmitted but decoded for users in the set S_2, respectively. Then, we have

$$\begin{aligned}
&\mathbb{P}\left[\mathcal{F}_t \middle| \mathbf{Z},\mathbf{H}_1,\mathbf{H}_2,\mathbf{A}_{S_1}\right] \\
&\leq \mathbb{P}\left[\bigcup_{S_1}\bigcup_{S_2}\bigcup_{\mathbf{c}'_{[S_2]}}\left\{\sum_{l\in[L]}\left\|\mathbf{z}_l + \sum_{k\in S_1} h_{k,l}\mathbf{c}_{(k)} - \sum_{k\in S_2} h_{k,l}\mathbf{c}'_{(k)}\right\|_2^2\right.\right. \\
&\qquad\qquad \left.\left.\leq \sum_{l\in[L]}\left\|\mathbf{z}_l\right\|_2^2\right\}\middle| \mathbf{Z},\mathbf{H}_1,\mathbf{H}_2,\mathbf{A}_{S_1}\right] \qquad (144)\\
&\leq \sum_{S_1}\sum_{S_2} M^{\rho t}\left(\mathbb{P}\left[\left\|\mathbf{Z}+\mathbf{A}_{S_1}\mathbf{H}_1 - \mathbf{A}'_{S_2}\mathbf{H}_2\right\|_F^2\right.\right. \\
&\qquad\qquad \left.\left.\leq \left\|\mathbf{Z}\right\|_F^2 \middle| \mathbf{Z},\mathbf{H}_1,\mathbf{H}_2,\mathbf{A}_{S_1}\right]\right)^{\rho}, \qquad (145)
\end{aligned}$$

where (145) follows by applying Gallager's ρ-trick, i.e., $\mathbb{P}\left[\cup_j B_j\right] \leq \left(\sum_j \mathbb{P}\left[B_j\right]\right)^{\rho}$ for any $\rho \in [0,1]$ [30], [48, Sec. 5.6].

The probability on the RHS of (145) can be upper-bounded as

$$\begin{aligned}
&\mathbb{P}\left[\left\|\mathbf{Z}+\mathbf{A}_{S_1}\mathbf{H}_1 - \mathbf{A}'_{S_2}\mathbf{H}_2\right\|_F^2 \leq \left\|\mathbf{Z}\right\|_F^2 \middle| \mathbf{Z},\mathbf{H}_1,\mathbf{H}_2,\mathbf{A}_{S_1}\right] \\
&\leq \mathbb{E}_{\mathbf{A}'_{S_2}}\left[\exp\left\{-\beta\left\|\mathbf{Z}+\mathbf{A}_{S_1}\mathbf{H}_1 - \mathbf{A}'_{S_2}\mathbf{H}_2\right\|_F^2\right.\right. \\
&\qquad\qquad \left.\left. +\beta\left\|\mathbf{Z}\right\|_F^2\right\}\middle| \mathbf{Z},\mathbf{H}_1,\mathbf{H}_2,\mathbf{A}_{S_1}\right] \qquad (146)\\
&= \exp\left\{\beta\left\|\mathbf{Z}\right\|_F^2\right\}\left|\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2\right|^{-\frac{n}{2}} \\
&\quad \cdot \prod_{i=1}^{n}\exp\left\{-\beta\tilde{\boldsymbol{\mu}}_i^T\left(\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2\right)^{-1}\tilde{\boldsymbol{\mu}}_i\right\}, \qquad (147)
\end{aligned}$$

where (146) follows from the Chernoff bound in Lemma 14 with $\beta \geq 0$, and (147) is obtained as follows. Let $\boldsymbol{\mu}_i = \left([\mathbf{Z}]_{i,:} + [\mathbf{A}_{S_1}]_{i,:}\mathbf{H}_1\right)^H$ and $\boldsymbol{\nu}_i = \boldsymbol{\mu}_i - \left([\mathbf{A}'_{S_2}]_{i,:}\mathbf{H}_2\right)^H$. Conditioned on $\mathbf{Z},\mathbf{H}_1,\mathbf{H}_2$, and $\mathbf{A}_{S_1}$, we can obtain that $\boldsymbol{\nu}_i \sim \mathcal{CN}\left(\boldsymbol{\mu}_i, P'\mathbf{H}_2^H\mathbf{H}_2\right)$ for $i \in [n]$. Let $\tilde{\boldsymbol{\nu}}_i = \begin{bmatrix}\Re\left(\boldsymbol{\nu}_i\right)\\ \Im\left(\boldsymbol{\nu}_i\right)\end{bmatrix}$, $\tilde{\boldsymbol{\mu}}_i = \begin{bmatrix}\Re\left(\boldsymbol{\mu}_i\right)\\ \Im\left(\boldsymbol{\mu}_i\right)\end{bmatrix}$, and $\tilde{\mathbf{\Sigma}}_2 = \begin{bmatrix}\Re(P'\mathbf{H}_2^H\mathbf{H}_2) & -\Im(P'\mathbf{H}_2^H\mathbf{H}_2)\\ \Im(P'\mathbf{H}_2^H\mathbf{H}_2) & \Re(P'\mathbf{H}_2^H\mathbf{H}_2)\end{bmatrix}$. We have $\tilde{\boldsymbol{\nu}}_i \sim \mathcal{N}\left(\tilde{\boldsymbol{\mu}}_i, \frac{1}{2}\tilde{\mathbf{\Sigma}}_2\right)$ conditioned on $\mathbf{Z},\mathbf{H}_1,\mathbf{H}_2$, and $\mathbf{A}_{S_1}$. Then, applying Lemma 15, we can obtain

$$\begin{aligned}
&\mathbb{E}_{\mathbf{A}'_{S_2}}\left[\exp\left\{-\beta\boldsymbol{\nu}_i^H\boldsymbol{\nu}_i\right\}\middle| \mathbf{Z},\mathbf{H}_1,\mathbf{H}_2,\mathbf{A}_{S_1}\right] \\
&= \mathbb{E}_{\mathbf{A}'_{S_2}}\left[\exp\left\{-\beta\tilde{\boldsymbol{\nu}}_i^T\tilde{\boldsymbol{\nu}}_i\right\}\middle| \mathbf{Z},\mathbf{H}_1,\mathbf{H}_2,\mathbf{A}_{S_1}\right] \qquad (148)\\
&= \left|\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2\right|^{-\frac{1}{2}}\exp\left\{-\beta\tilde{\boldsymbol{\mu}}_i^T\left(\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2\right)^{-1}\tilde{\boldsymbol{\mu}}_i\right\}, \\
&\qquad\qquad (149)
\end{aligned}$$

which yields (147). Substituting (147) into (145) and taking the expectation over $\mathbf{A}_{S_1}$ and $\mathbf{Z}$, we can obtain

$$\begin{aligned}
&\mathbb{P}\left[\mathcal{F}_t \middle| \mathbf{H}_1,\mathbf{H}_2\right] \leq \sum_{S_1}\sum_{S_2} M^{\rho t}\left|\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2\right|^{-\frac{\rho n}{2}} \\
&\underbrace{\cdot\left|\begin{bmatrix}\mathbf{I}_{2L}+\rho\beta(\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2)^{-1}(\mathbf{I}_{2L}+\tilde{\mathbf{\Sigma}}_1) & \rho\beta(\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2)^{-1}\\ -\rho\beta\mathbf{I}_{2L} & (1-\rho\beta)\,\mathbf{I}_{2L}\end{bmatrix}\right|^{-\frac{n}{2}},}_{C_{S_1,S_2}} \\
&\qquad\qquad (150)
\end{aligned}$$

where $\tilde{\mathbf{\Sigma}}_1 = \begin{bmatrix}\Re(P'\mathbf{H}_1^H\mathbf{H}_1) & -\Im(P'\mathbf{H}_1^H\mathbf{H}_1)\\ \Im(P'\mathbf{H}_1^H\mathbf{H}_1) & \Re(P'\mathbf{H}_1^H\mathbf{H}_1)\end{bmatrix}$. Here, (150) follows by applying Lemma 15 and Sylvester's determinant theorem under the condition that $0 \leq \beta < \frac{1}{\rho}$. Then, we have

$$\begin{aligned}
&\left|\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2\right|^{-\frac{\rho n}{2}} C_{S_1,S_2} \\
&= \left|(1-\rho\beta)\,\mathbf{I}_{2L}\right|^{-\frac{n}{2}}\left|\rho\beta\left(\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2\right)^{\rho-1}\left(\mathbf{I}_{2L}+\tilde{\mathbf{\Sigma}}_1\right)\right. \\
&\qquad \left.+\left(\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2\right)^{\rho}+\frac{\rho^2\beta^2}{1-\rho\beta}\left(\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2\right)^{\rho-1}\right|^{-\frac{n}{2}} \qquad (151)\\
&= \left|\mathbf{I}_{2L}+\beta\tilde{\mathbf{\Sigma}}_2\right|^{-\frac{(-1+\rho)n}{2}} \\
&\quad \cdot\left|\mathbf{I}_{2L}+\rho\beta\left(1-\rho\beta\right)\tilde{\mathbf{\Sigma}}_1+\beta\left(1-\rho\beta\right)\tilde{\mathbf{\Sigma}}_2\right|^{-\frac{n}{2}} \qquad (152)\\
&= \left|\mathbf{I}_L+\beta P'\mathbf{H}_2^H\mathbf{H}_2\right|^{(1-\rho)n}\left|\mathbf{I}_L+\rho\beta\left(1-\rho\beta\right)P'\mathbf{H}_1^H\mathbf{H}_1\right. \\
&\qquad\qquad \left.+\beta(1-\rho\beta)P'\mathbf{H}_2^H\mathbf{H}_2\right|^{-n}, \qquad (153)
\end{aligned}$$

where (151) holds because $\left|\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}\right| = \left|\mathbf{D}\right|\left|\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right|$ when $\mathbf{D}$ is nonsingular, and (153) holds because $\left|\mathbf{D}\right|^2 = \left|\begin{bmatrix}\Re(\mathbf{D}) & -\Im(\mathbf{D})\\ \Im(\mathbf{D}) & \Re(\mathbf{D})\end{bmatrix}\right|$.

Substituting (153) into (150) and taking the expectation over $\mathbf{H}_1$ and $\mathbf{H}_2$, we have

$$\begin{aligned}
&\mathbb{P}\left[\mathcal{F}_t\right] \leq \sum_{t_0=0}^{t}\binom{K_a}{t}\binom{t}{t_0}\binom{K-K_a}{t-t_0}M^{\rho t} \\
&\quad \cdot\mathbb{E}_{\mathbf{H}_1,\mathbf{H}_2}\left[\exp\left\{(1-\rho)n\ln\left|\mathbf{I}_L+\beta P'\mathbf{H}_2^H\mathbf{H}_2\right|\right.\right. \\
&\qquad \left.\left.-n\ln\left|\mathbf{I}_L+\beta\left(1-\rho\beta\right)P'\left(\rho\mathbf{H}_1^H\mathbf{H}_1+\mathbf{H}_2^H\mathbf{H}_2\right)\right|\right\}\right], \\
&\qquad\qquad (154)
\end{aligned}$$

where (154) holds because the expectation is unchanged for different S_1 and S_2 once t_0 and t are fixed, considering that channel coefficients are i.i.d. for different users. Taking the minimum over ρ and β on the RHS of (154), we can obtain an upper bound on $\mathbb{P}\left[\mathcal{F}_t\right]$ based on Gallager's ρ-trick, which is denoted as $p_{2,t}$ given in (24). This completes the proof of Theorem 1.

APPENDIX C
PROOF OF COROLLARY 3

In a special case where all users are active (i.e. $K_a = K$), the set S_1 of misdecoded users is the same as the set S_2 including detected users with false alarm codewords. Thus, $\tilde{p}_{1,t}$ can be easily obtained from Theorem 1 and its proof is omitted here for the sake of brevity. Moreover, the proof of $\tilde{p}_{2,t}$ is provided in Appendix C-A; the upper bound $\tilde{p}_{2,t}^{\mathrm{u}}$ on $\tilde{p}_{2,t}$ is derived in Appendix C-B.

A. Proof of (29)

In a special case where all users are active, i.e., $K_a = K$, $\mathbf{H}_1 = \mathbf{H}_2$, and $S_1 = S_2$, it is easy to see that the optimum value of β minimizing (154) is given by $\beta^* = 1/(1+\rho)$. Then, we have

$$\mathbb{P}\left[\mathcal{F}_t \middle| \mathbf{H}_1\right] \leq \sum_{S_1} M^{\rho t} \exp\left\{-\rho n \ln\left|\mathbf{I}_L + \frac{P'}{1+\rho}\mathbf{H}_1^H\mathbf{H}_1\right|\right\}. \tag{155}$$

Taking the expectation over $\mathbf{H}_1$, we have

$$\begin{aligned}&\mathbb{P}\left[\mathcal{F}_t\right] \\ &\leq \min_{0\leq\rho\leq1} \sum_{S_1} M^{\rho t}\, \mathbb{E}_{\mathbf{H}_1}\left[\exp\left\{-\rho n \ln\left|\mathbf{I}_L + \frac{P'}{1+\rho}\mathbf{H}_1^H\mathbf{H}_1\right|\right\}\right] \quad (156)\\ &= \min_{0\leq\rho\leq1} \binom{K}{t} M^{\rho t}\, \mathbb{E}_{\mathbf{G}}\left[\exp\left\{-L \ln\left|\mathbf{I}_t + \frac{P'}{1+\rho}\mathbf{G}\mathbf{G}^H\right|\right\}\right], \quad (157)\end{aligned}$$

where (157) holds when ρn is an integer and each element of $\mathbf{H}_1 \in \mathbb{C}^{t\times L}$ and $\mathbf{G} \in \mathbb{C}^{t\times \rho n}$ is i.i.d. $\mathcal{CN}(0,1)$ distributed. This is because

$$\begin{aligned}&\mathbb{E}_{\mathbf{H}_1,\mathbf{G}}\left[\exp\left\{-\frac{P'}{1+\rho}\left\|\mathbf{G}^H\mathbf{H}_1\right\|_F^2\right\}\right]\\ &= \mathbb{E}_{\mathbf{H}_1}\left[\prod_{i=1}^{\rho n}\mathbb{E}\left[\exp\left\{-\frac{P'}{1+\rho}\left\|\left([\mathbf{G}]_{:,i}\right)^H\mathbf{H}_1\right\|_2^2\right\}\middle|\mathbf{H}_1\right]\right] \quad (158)\\ &= \mathbb{E}_{\mathbf{H}_1}\left[\exp\left\{-\rho n \ln\left|\mathbf{I}_L + \frac{P'}{1+\rho}\mathbf{H}_1^H\mathbf{H}_1\right|\right\}\right] \quad (159)\\ &= \mathbb{E}_{\mathbf{G}}\left[\prod_{l=1}^{L}\mathbb{E}\left[\exp\left\{-\frac{P'}{1+\rho}\left\|\mathbf{G}^H\left[\mathbf{H}_1\right]_{:,l}\right\|_2^2\right\}\middle|\mathbf{G}\right]\right] \quad (160)\\ &= \mathbb{E}_{\mathbf{G}}\left[\exp\left\{-L \ln\left|\mathbf{I}_t + \frac{P'}{1+\rho}\mathbf{G}\mathbf{G}^H\right|\right\}\right], \quad (161)\end{aligned}$$

where (159) and (161) follows from Lemma 15. Denote the RHS of (157) as $\tilde{p}_{2,t}$. This completes the proof of (29).

B. Proof of the Upper Bound $\tilde{p}_{2,t}^{\mathrm{u}}$ on $\tilde{p}_{2,t}$

Recall that each element of $\mathbf{G} \in \mathbb{C}^{t\times\rho n}$ is i.i.d. $\mathcal{CN}(0,1)$ distributed. In the case of $\rho n \geq t + L$, the expectation in (29) can be upper-bounded as

$$\begin{aligned}&\mathbb{E}_{\mathbf{G}}\left[\left|\mathbf{I}_t + \frac{P'}{1+\rho}\mathbf{G}\mathbf{G}^H\right|^{-L}\right]\\ &\leq \left(\frac{P'}{1+\rho}\right)^{-Lt}\mathbb{E}_{\mathbf{G}}\left[\left|\mathbf{G}\mathbf{G}^H\right|^{-L}\right] \quad (162)\\ &= \left(\frac{P'}{1+\rho}\right)^{-Lt}\mathbb{E}\left[\prod_{i=\rho n-t+1}^{\rho n}\left(\frac{\chi^2(2i)}{2}\right)^{-L}\right] \quad (163)\\ &= \left(\frac{P'}{1+\rho}\right)^{-Lt}\prod_{i=\rho n-t+1}^{\rho n}\frac{\Gamma(i-L)}{\Gamma(i)}, \quad (164)\end{aligned}$$

where (162) follows because $|\mathbf{I}+\mathbf{A}| \geq |\mathbf{A}|$ when $\mathbf{A}$ is a positive semidefinite matrix; (163) follows because the determinant of the Wishart matrix $\left|\mathbf{G}\mathbf{G}^H\right|$ has the same distribution as the product of independent random variables with chi-square distributions, i.e., $\prod_{i=\rho n-t+1}^{\rho n}\frac{\chi^2(2i)}{2}$ [49, Sec. 3.5]; (164) follows from the moments of chi-square random variables.

Applying similar ideas, we can upper-bound this term in the case of $\rho n \leq t - L$ as follows:

$$\begin{aligned}&\mathbb{E}_{\mathbf{G}}\left[\left|\mathbf{I}_t + \frac{P'}{1+\rho}\mathbf{G}\mathbf{G}^H\right|^{-L}\right]\\ &\leq \left(\frac{P'}{1+\rho}\right)^{-L\rho n}\prod_{i=t-\rho n+1}^{t}\frac{\Gamma(i-L)}{\Gamma(i)}. \quad (165)\end{aligned}$$

Substituting (164) and (165) into (29) and considering $\tilde{p}_{2,t} \leq 1$, we can obtain (30).

APPENDIX D
PROOF OF THEOREM 4

In this appendix, we prove Theorem 4 to establish a converse bound on the minimum required energy-per-bit for the CSIR case. We assume a genie G reveals the set $\mathcal{K}_a$ of active users and a subset $S_1 \subset \mathcal{K}_a$ for messages $\mathcal{W}_{S_1} = \{W_k : k \in S_1\}$ and corresponding fading coefficients to the decoder. It is evident that a converse bound in the genie case is a converse bound for the problem without genie. Let $S_2 = \mathcal{K}_a \backslash S_1$ of size t. Let $\mathbf{\Phi}_{S_2} \in \{0,1\}^{MK\times K}$ denote which codewords are transmitted by users in the set $S_2 \subset \mathcal{K}_a$, where $[\mathbf{\Phi}_{S_2}]_{(k-1)M+W_k,k} = 1$ if the k-th user belonging to the set S_2 is active and the W_k-th codeword is transmitted, and $[\mathbf{\Phi}_{S_2}]_{(k-1)M+W_k,k} = 0$ otherwise. The equivalent received signal of the l-th antenna at the BS is given by

$$\mathbf{y}_l^G = \sum_{k\in S_2} h_{k,l}\mathbf{x}_{(k)} + \mathbf{z}_l = \mathbf{X}\mathbf{\Phi}_{S_2}\mathbf{h}_l + \mathbf{z}_l \in \mathbb{C}^n. \tag{166}$$

The equivalent received message over all antennas is given by

$$\mathbf{Y}^G = \mathbf{X}\mathbf{\Phi}_{S_2}\mathbf{H} + \mathbf{Z}, \tag{167}$$

where $\mathbf{Y}^G = [\mathbf{y}_1^G, \mathbf{y}_2^G, \ldots, \mathbf{y}_L^G] \in \mathbb{C}^{n\times L}$, and $\mathbf{H}$ and $\mathbf{Z}$ are defined in Section II. Denote the decoded signal for the k-th user with genie as $\hat{W}_k^G$. Let $\mathcal{M}_k = 1\left[W_k \neq \hat{W}_k^G\right]$ and $P_{e,k}^G = \mathbb{E}\left[\mathcal{M}_k\right]$. We have $P_{e,k}^G = 0$ for $k \in S_1$. The averaged PUPE is $P_e^G = \frac{1}{K_a}\sum_{k\in S_2} P_{e,k}^G \leq \epsilon$.

Based on the Fano inequality, we have

$$\frac{t}{K_a}J - P_e^G \log_2\left(2^J - 1\right) - \frac{1}{K_a}\sum_{k\in S_2} h_2\left(P_{e,k}^G\right)$$

$$\le \frac{1}{K_a}\sum_{k\in S_2} I_2\left(W_k;\hat{W}_k^G\right). \tag{168}$$

Considering the concavity of $h_2(\cdot)$ and the inequality that $P_e^G \le \epsilon \le 1-\frac{1}{2^J}$, we have

$$P_e^G\log_2\left(2^J-1\right)+\frac{1}{K_a}\sum_{k\in S_2} h_2\left(P_{e,k}^G\right) \le \epsilon J + h_2\left(\epsilon\right). \tag{169}$$

Denote $\mathcal{W}_{S_2}=\{W_k : k\in S_2\}$, $\mathcal{X}_{S_2}=\left\{\mathbf{x}_{(k)} : k\in S_2\right\}$, and $\hat{\mathcal{W}}_{S_2}^G=\left\{\hat{W}_k^G : k\in S_2\right\}$. The matrix $\mathbf{H}_t$ is a $t\times L$ submatrix of $\mathbf{H}$ corresponding to fading coefficients of users in the set S_2. We can upper-bound $\sum_{k\in S_2} I_2\left(W_k;\hat{W}_k^G\right)$ as

$$\sum_{k\in S_2} I_2\left(W_k;\hat{W}_k^G\right) = H_2(\mathcal{W}_{S_2}) - \sum_{k\in S_2} H_2\left(W_k\left|\hat{W}_k^G\right.\right) \tag{170}$$

$$\le H_2\left(\mathcal{W}_{S_2}\right) - H_2\left(\mathcal{W}_{S_2}\left|\hat{\mathcal{W}}_{S_2}^G\right.\right) \tag{171}$$

$$= I_2\left(\mathcal{W}_{S_2};\hat{\mathcal{W}}_{S_2}^G\right) \tag{172}$$

$$\le I_2\left(\mathcal{X}_{S_2};\mathbf{Y}^G\right) \tag{173}$$

$$\le n\,\mathbb{E}_{\mathbf{H}_t}\left[\log_2\left|\mathbf{I}_L + P\mathbf{H}_t^H\mathbf{H}_t\right|\right], \tag{174}$$

where $H_2(x)$ denotes the entropy of a random variable x. Here, (171) follows because

$$H_2\left(\mathcal{W}_{S_2}\left|\hat{\mathcal{W}}_{S_2}^G\right.\right) = \sum_{k\in S_2} H_2\left(W_k\left|\hat{\mathcal{W}}_{S_2}^G, W_1,\ldots,W_{k-1}\right.\right) \tag{175}$$

$$\le \sum_{k\in S_2} H_2\left(W_k\left|\hat{W}_k^G\right.\right), \tag{176}$$

(173) follows due to the data processing inequality and the Markov chain: $\mathcal{W}_{S_2}\to\mathcal{X}_{S_2}\to\mathbf{Y}^G\to\hat{\mathcal{W}}_{S_2}^G$, and (174) holds because both $\mathbf{X}$ and $\mathbf{Z}$ are independent for n channel uses and the normal distribution of codewords maximizes the entropy for a given variance [3, Th. 8.6.5].

Substituting (169) and (174) into (168), we can obtain (33) in Theorem 4. Then, applying the concavity of $\log_2|\cdot|$ function, we obtain (34), which completes the proof of Theorem 4.

APPENDIX E
PROOF OF THEOREM 5

To prove Theorem 5, we first establish an achievability result in Appendix E-A and then prove a converse result in Appendix E-B for the CSIR case assuming all users are active.

A. Achievability

In a special case where all users are active, the PUPE can be upper-bounded as

$$P_e \le \mathbb{E}\left[\frac{1}{K}\sum_{k\in\mathcal{K}} 1\left[W_k\neq\hat{W}_k\right]\right]_{\text{no power constraint}} + \tilde{p}_0 \tag{177}$$

$$\le \epsilon_1 + \mathbb{P}\left[\frac{1}{K}\sum_{k\in\mathcal{K}} 1\left[W_k\neq\hat{W}_k\right] \ge \epsilon_1\right]_{\text{no power constraint}} + \tilde{p}_0 \tag{178}$$

$$= \epsilon_1 + \sum_{t=\lceil\epsilon_1 K\rceil}^{K} \mathbb{P}\left[\mathcal{F}_t\right]_{\text{no power constraint}} + \tilde{p}_0, \tag{179}$$

where ϵ_1 denotes a positive constant less than ϵ; $\mathcal{F}_t = \left\{\sum_{k\in\mathcal{K}} 1\left\{W_k\neq\hat{W}_k\right\}=t\right\}$ denotes the event that there are exactly t misdecoded users; $\tilde{p}_0$ upper-bounds the power constraint violation probability given by

$$\tilde{p}_0 = K\,\mathbb{P}\left[\frac{x}{2n} > \frac{P}{P'}\right] \le K\exp\left\{-n\left(\frac{P}{P'}-\sqrt{2\frac{P}{P'}-1}\right)\right\}, \quad x\sim\chi^2(2n), \tag{180}$$

which follows from Lemma 17 presented below. It is easy to see that $c_P = \frac{P}{P'}-\sqrt{2\frac{P}{P'}-1}$ is a positive finite constant, provided that $\frac{P}{P'}-1$ is a positive finite constant. In the case of $\ln K = o(n)$, we have $\tilde{p}_0 \le \exp\{o(n)-c_P n\}\to 0$ as $n\to\infty$.

Lemma 17 ([50]): Let $x\sim\chi^2\left(m\right)$ be a central chi-square distributed variable with m degrees of freedom. For $\forall a>0$,

$$\mathbb{P}\left[x-m\ge a\right] \le \exp\left\{-\frac{1}{2}\left(a+m-\sqrt{m}\sqrt{2a+m}\right)\right\}. \tag{181}$$

An upper bound $\tilde{p}_{1,t}$ on $\mathbb{P}\left[\mathcal{F}_t\right]_{\text{no power constraint}}$ is given in Corollary 3. Next, we pay attention to upper-bounding $\tilde{p}_{1,t}$, thereby finding the condition under which $\sum_{t=\lceil\epsilon_1 K\rceil}^{K}\tilde{p}_{1,t}\to 0$ and thus $P_e\le\epsilon$. In the case of $\epsilon_1 K \ge n+L$, for $t=\lceil\epsilon_1 K\rceil, \lceil\epsilon_1 K\rceil+1,\ldots,K$, we have

$$\tilde{p}_{1,t} \le \binom{K}{t} M^t \left(\frac{P'}{2}\right)^{-Ln} \prod_{i=t-n+1}^{t}\frac{\Gamma(i-L)}{\Gamma(i)} \tag{182}$$

$$\le \binom{K}{t} M^t \left(\frac{P'\left(t-n+1-L\right)}{2}\right)^{-Ln}, \tag{183}$$

where (182) follows from (30) and (31) by allowing $\rho=1$, and (183) follows from the equality that $\Gamma(x)=(x-1)!$ for any positive integer x.

Let $t=\theta K$ with $\theta\in S_\theta=\left\{\frac{1}{K},\frac{2}{K},\ldots,1\right\}\cap[\epsilon_1,1]$. We have

$$\sum_{t=\lceil\epsilon_1 K\rceil}^{K}\tilde{p}_{1,t}$$

$$\le \sum_{t=\lceil\epsilon_1 K\rceil}^{K}\binom{K}{t} M^t \left(\frac{P'\left(t-n+1-L\right)}{2}\right)^{-Ln} \tag{184}$$

$$\le \exp\left\{o(K)+K\max_{\theta\in S_\theta}\left\{h(\theta)+\theta\ln M - \frac{Ln}{K}\ln\left(\frac{P'(\theta K-n+1-L)}{2}\right)\right\}\right\} \tag{185}$$

$$\le \exp\left\{o(K)+K\left(h\left(\frac{1}{2}\right)+\ln M - \frac{Ln}{K}\ln\left(\frac{P'(\epsilon_1 K-n+1-L)}{2}\right)\right)\right\}, \tag{186}$$

where (185) follows from the inequality that [3, Example 11.1.3]

$$\binom{K}{t} \le \exp\left\{Kh(\theta)\right\}. \tag{187}$$

Therefore, in the case of $K \to \infty$, we have $\sum_{t=\lceil \epsilon_1 K \rceil}^{K} \tilde{p}_{1,t} \to 0$ provided that the finite constant

$$c_1 = \frac{Ln}{K} \ln \left(\frac{P' \left(\epsilon_1 K - n + 1 - L \right)}{2} \right) - h \left(\frac{1}{2} \right) - \ln M > 0. \tag{188}$$

Assume $M = \Theta(1)$, K and $n \to \infty$, and $K \gg \frac{n+L-1}{\epsilon_1}$. In the case of $KP' = \Omega(1)$, we can obtain that (188) is satisfied if and only if $\frac{nL \ln KP'}{K} = \Omega(1)$, which can be divided into the following two relations:

1) We assume $P'K$ is a finite positive constant satisfying $P'K > \frac{2}{\epsilon_1}$. In this case, we have

$$c_1 = c_3 \frac{nL}{K} - c_2, \tag{189}$$

where $c_2 = h\left(\frac{1}{2}\right) + \ln M$ and c_3 is a finite positive constant. In order to satisfy the condition in (188), it is possible to choose $\frac{nL}{K} = \Omega(1)$ and $P'K = \Theta(1)$. An example for this case is that the number of BS antennas satisfies $L = \Theta(n)$, the power satisfies $P' = \Theta\left(\frac{1}{n^2}\right)$ and the number of users satisfies $K = \Theta(n^2)$.

2) In the case of $P'K \to \infty$, we have

$$c_1 = Ln \frac{\ln KP'}{K} - Ln \frac{\mathcal{O}(1)}{K} - c_2, \tag{190}$$

where $c_2 = h\left(\frac{1}{2}\right) + \ln M$. Applying (190), in order to satisfy the condition in (188), it is possible to choose $\frac{nL \ln KP'}{K} = \Omega(1)$ with $KP' \to \infty$. An example for this case satisfies $L = \Theta\left(\frac{n}{\ln n}\right)$, $P' = \Theta\left(\frac{1}{n}\right)$, and $K = \Theta(n^2)$.

Combining (180) and (188), we conclude that assuming $K, n \to \infty$, $\ln K = o(n)$, $KP = \Omega(1)$, $M = \Theta(1)$, and $K \geq \frac{n+L-1}{\epsilon_1}$, the PUPE requirement $P_e \leq \epsilon$ is satisfied provided that $\frac{nL \ln KP}{K} = \Omega(1)$. In particular, the PUPE requirement is satisfied for $K = \frac{n+L-1}{\epsilon_1}$ users when $KP = \Omega(1)$ (the condition $\frac{nL \ln KP}{K} = \Omega(1)$ is satisfied directly in this case). It was proved in [8, Appendix A-C] that, if one can achieve a certain PUPE for K users, it will also be possible to achieve the same PUPE for less than K users. Thus, one can reliably serve $K \leq \frac{n+L-1}{\epsilon_1}$ users provided that $\frac{n+L-1}{\epsilon_1} P = \Omega(1)$, or under a stricter condition that $KP = \Omega(1)$. As a result, assuming $K, n \to \infty$, $\ln K = o(n)$, $KP = \Omega(1)$, and $M = \Theta(1)$, the PUPE requirement $P_e \leq \epsilon$ is satisfied if $\frac{nL \ln KP}{K} = \Omega(1)$. That is, it is possible to choose the following two regimes: $\frac{nL}{K} = \Omega(1)$ and $KP = \Theta(1)$; $\frac{nL \ln KP}{K} = \Omega(1)$ and $KP \to \infty$. In particular, when the number of BS antennas is $L = \Theta(n)$ (resp. $L = \Theta\left(\frac{n}{\ln n}\right)$) and the power satisfies $P = \Theta\left(\frac{1}{n^2}\right)$ (resp. $P = \Theta\left(\frac{1}{n}\right)$), we can reliably serve $K = \mathcal{O}(n^2)$ users.

B. Converse

Assume that $t = \theta K$ with $\theta \in S'_\theta = \left\{\frac{1}{K}, \frac{2}{K}, \ldots, 1\right\}$. We consider the case where ϵ and J are finite positive constants. Following from Theorem 4, the minimum required energy-per-bit is larger than $\inf \frac{nP}{J}$, where the infimum is taken over all $P > 0$ satisfying that

$$\left(\theta - \epsilon\right) J - h_2\left(\epsilon\right) \leq \frac{nL}{K} \log_2 \left(1 + \theta PK\right), \forall \theta \in S'_\theta. \tag{191}$$

When $\theta \leq \theta' = \frac{h_2(\epsilon)}{J} + \epsilon$, (191) is satisfied for any positive P, n, L, and K.

Next, assuming $n, K \to \infty$, $KP = \Omega(1)$, and $J = \Theta(1)$, the inequality in (191) holds for any $\theta \in S'_\theta \cap (\theta', 1]$ if and only if $\frac{nL \ln PK}{K} = \Omega(1)$, which can be divided into the following two relations: 1) $KP = \Theta(1)$ and $\frac{nL}{K} = \Omega(1)$; 2) $KP \to \infty$ and $\frac{nL \ln KP}{K} = \Omega(1)$. Moreover, since the RHS of (191) is a monotonically decreasing function of K, when the number of BS antennas is $L = \Theta(n)$ (resp. $L = \Theta\left(\frac{n}{\ln n}\right)$) and the power satisfies $P = \Theta\left(\frac{1}{n^2}\right)$ (resp. $P = \Theta\left(\frac{1}{n}\right)$), K users can be reliably served only if $K = \mathcal{O}(n^2)$.

Together with the case of $\theta \leq \theta'$, assuming $n, K \to \infty$, $KP = \Omega(1)$, and $J = \Theta(1)$, the inequality in (191) holds for any $\theta \in S'_\theta$ if and only if $\frac{nL \ln PK}{K} = \Omega(1)$, i.e., if and only if one of the two relations mentioned above is satisfied.

Together with the achievability result in Appendix E-A, we conclude that assuming $n, K \to \infty$, $\ln K = o(n)$, $KP = \Omega(1)$, and $J = \Theta(1)$, the PUPE requirement $P_e \leq \epsilon$ is satisfied if and only if $\frac{nL \ln PK}{K} = \Omega(1)$, i.e., if and only if one of the following two relations is satisfied: 1) $\frac{nL}{K} = \Omega(1)$ and $KP = \Theta(1)$; 2) $\frac{nL \ln KP}{K} = \Omega(1)$ and $KP \to \infty$. In particular, when the number of BS antennas is $L = \Theta(n)$ (resp. $L = \Theta\left(\frac{n}{\ln n}\right)$) and the power satisfies $P = \Theta\left(\frac{1}{n^2}\right)$ (resp. $P = \Theta\left(\frac{1}{n}\right)$), the number of users that can be reliably served is in the order of $K = \mathcal{O}(n^2)$.

Appendix F
Proof of Theorem 6

In this appendix, we prove Theorem 6 to establish an achievability bound on the PUPE in the case of no-CSI with known K_a. As introduced in Appendix A, the PUPE can be upper-bounded by (105). The probability $\mathbb{P}\left[\mathcal{F}_t\right]$ therein, i.e. the probability of the event that there are exactly t misdecoded users, is upper-bounded in (109) applying Fano's "good region" technique [29]. In the following, we particularize the "good region"-based bound on $\mathbb{P}\left[\mathcal{F}_t\right]$ given in Appendix A to the no-CSI case, followed by further manipulations on the two probabilities on the RHS of (109).

Based on the notation introduced in Appendix A, the ML decoding metric $g\left(\mathbf{Y}, \hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right)$ in the case of no-CSI is given by [20]

$$\begin{aligned} g&\left(\mathbf{Y}, \hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right) \\ &= L \ln \left| \mathbf{I}_n + \sum\nolimits_{k \in \hat{\mathcal{K}}_a} \hat{\mathbf{c}}_{(k)} \hat{\mathbf{c}}_{(k)}^H \right| \\ &\quad + \operatorname{tr}\left(\left(\mathbf{I}_n + \sum\nolimits_{k \in \hat{\mathcal{K}}_a} \hat{\mathbf{c}}_{(k)} \hat{\mathbf{c}}_{(k)}^H \right)^{-1} \mathbf{Y}\mathbf{Y}^H \right) \end{aligned} \tag{192}$$

$$= L \ln \left| \mathbf{I}_n + \mathbf{A} \mathbf{\Gamma}'_{\hat{\mathcal{K}}_a} \mathbf{A}^H \right| + \operatorname{tr}\left(\left(\mathbf{I}_n + \mathbf{A} \mathbf{\Gamma}'_{\hat{\mathcal{K}}_a} \mathbf{A}^H \right)^{-1} \mathbf{Y}\mathbf{Y}^H \right). \tag{193}$$

Here, the matrix $\mathbf{A} \in \mathbb{C}^{n \times MK}$ denotes the concatenation of codebooks of the K users, which has i.i.d. $\mathcal{CN}\left(0, P'\right)$ entries; the matrix $\mathbf{\Gamma}'_S = \operatorname{diag}\left\{\boldsymbol{\gamma}'_S\right\} \in \{0,1\}^{KM \times KM}$, where $\left[\boldsymbol{\gamma}'_S\right]_{(k-1)M + W_k} = 1$ if $k \in S$ and the W_k-th codeword is decoded for this user, and $\left[\boldsymbol{\gamma}^{'}_{S}\right]_{(k-1)M+W_k} = 0$ otherwise. Similarly, let $\boldsymbol{\Gamma}_S = \operatorname{diag}\{\boldsymbol{\gamma}_S\} \in \{0,1\}^{KM\times KM}$ be a diagonal matrix, where $[\boldsymbol{\gamma}_S]_{(k-1)M+W_k} = 1$ if $k \in S$ and the W_k-th codeword is transmitted by this user, and $[\boldsymbol{\gamma}_S]_{(k-1)M+W_k} = 0$ otherwise. In the following, we denote $g\left(\mathbf{Y}, \hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right)$ as $g\left(\boldsymbol{\Gamma}^{'}_{\hat{\mathcal{K}}_a}\right)$ for simplicity.

Let $\mathbf{A}_S \in \mathbb{C}^{n\times|S|}$ denote the concatenation of transmitted codewords of active users in the set $S \subset \mathcal{K}_a$ and let $\mathbf{A}^{'}_{S_2} \in \mathbb{C}^{n\times|S_2|}$ denote the concatenation of false-alarm codewords for users in the set $S_2 \subset \mathcal{K}\backslash\mathcal{K}_a \cup S_1$. Denote $\mathbf{A}_{all} = \left\{\mathbf{A}_{\mathcal{K}_a}, \mathbf{A}_{\mathcal{K}_a\backslash S_1}, \mathbf{A}^{'}_{S_2}\right\}$. Define $\mathbf{F}$, $\mathbf{F}^{'}$, and $\mathbf{F}_1$ as in (40), (41), and (42), respectively. The conditional expectation in (111) can be written as

$$\begin{aligned}
&\mathbb{E}_{\mathbf{H},\mathbf{Z}}\left[\exp\left\{(u-r)g\left(\boldsymbol{\Gamma}_{\mathcal{K}_a}\right) - ug\left(\boldsymbol{\Gamma}^{'}_{\mathcal{K}_a\backslash S_1\cup S_2}\right)\right.\right.\\
&\qquad\qquad \left.\left. + r\omega g\left(\boldsymbol{\Gamma}_{\mathcal{K}_a\backslash S_1}\right)\right\}\middle| \mathbf{c}_{[\mathcal{K}_a]}, \mathbf{c}_{[\mathcal{K}_a\backslash S_1]}, \mathbf{c}^{'}_{[S_2]}\right]\\
&= \exp\left\{(u-r)L\ln|\mathbf{F}| - uL\ln\left|\mathbf{F}^{'}\right| + r\omega L\ln|\mathbf{F}_1|\right\}\\
&\quad \cdot \mathbb{E}_{\mathbf{H},\mathbf{Z}}\left[\exp\left\{\operatorname{tr}\left(\mathbf{Y}^H\left((u-r)\mathbf{F}^{-1} - u\left(\mathbf{F}^{'}\right)^{-1}\right.\right.\right.\right.\\
&\qquad\qquad \left.\left.\left.\left. + r\omega\mathbf{F}_1^{-1}\right)\mathbf{Y}\right)\right\}\middle| \mathbf{A}_{all}\right] \qquad (194)\\
&= \exp\left\{L\left((u-r)\ln|\mathbf{F}| - u\ln\left|\mathbf{F}^{'}\right| + r\omega\ln|\mathbf{F}_1| - \ln|\mathbf{B}|\right)\right\}. \qquad (195)
\end{aligned}$$

Here, (195) follows from Lemma 15 by taking the expectation over $\mathbf{H}$ and $\mathbf{Z}$ provided that the minimum eigenvalue of $\mathbf{B}$ satisfies $\lambda_{\min}(\mathbf{B}) > 0$, where the matrix $\mathbf{B}$ is given by

$$\mathbf{B} = (1-u+r)\mathbf{I}_n + u\left(\mathbf{F}'\right)^{-1}\mathbf{F} - r\omega\mathbf{F}_1^{-1}\mathbf{F}. \qquad (196)$$

Then, we have

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{S_1}\bigcup_{S_2}\bigcup_{\mathbf{c}^{'}_{[S_2]}}\left\{g\left(\boldsymbol{\Gamma}^{'}_{\mathcal{K}_a\backslash S_1\cup S_2}\right) \le g\left(\boldsymbol{\Gamma}_{\mathcal{K}_a}\right)\right\}\bigcap \mathcal{G}_{\omega,\nu}\right]\\
&\le C_t\,\mathbb{E}_{\mathbf{A}_{all}}\left[\min_{u\ge0,r\ge0,\lambda_{\min}(\mathbf{B})>0}\exp\left\{L(rn\nu + (u-r)\ln|\mathbf{F}|\right.\right.\\
&\qquad\qquad \left.\left. - u\ln\left|\mathbf{F}^{'}\right| + r\omega\ln|\mathbf{F}_1| - \ln|\mathbf{B}|)\right\}\right], \qquad (197)
\end{aligned}$$

where $C_t = \binom{K_a}{t}\binom{K-K_a+t}{t}M^t$. Here, (197) follows by substituting (195) into (111) and follows from the fact that the expectation in (197) is unchanged for different S_1, S_2, and $\mathbf{c}^{'}_{[S_2]}$ once t is fixed, considering that the codebook matrix $\mathbf{A}$ has i.i.d. $\mathcal{CN}(0, P')$ entries. As a result, the first probability on the RHS of (109) is upper-bounded by (197), denoted as $q_{1,t}(\omega,\nu)$ in (38).

Next, we proceed to upper-bound the second term $\mathbb{P}\left[\mathcal{G}^c_{\omega,\nu}\right]$ on the RHS of (109). Denote $\mathbf{A}_{all} = \left\{\mathbf{A}_{\mathcal{K}_a}, \mathbf{A}_{\mathcal{K}_a\backslash S_1}\right\}$ and define the event $\mathcal{G}_\delta = \left\{\sum_{i=1}^{n}\frac{\chi_i^2(2L)}{2} \le nL(1+\delta)\right\}$ for $\delta \ge 0$. We have

$$\begin{aligned}
&\mathbb{P}\left[\mathcal{G}^c_{\omega,\nu}\right]\\
&= \mathbb{P}\left[\bigcup_{S_1}\left\{g\left(\boldsymbol{\Gamma}_{\mathcal{K}_a}\right) > \omega g\left(\boldsymbol{\Gamma}_{\mathcal{K}_a\backslash S_1}\right) + nL\nu\right\}\right] \qquad (198)\\
&\le \sum_{S_1}\mathbb{E}_{\mathbf{A}_{all}}\left[\mathbb{P}\left[\sum_{l=1}^{L}\tilde{\mathbf{y}}_l^H\left(\mathbf{I}_n - \omega\mathbf{F}^{\frac{H}{2}}\mathbf{F}_1^{-1}\mathbf{F}^{\frac{1}{2}}\right)\tilde{\mathbf{y}}_l > C_F\middle|\mathbf{A}_{all}\right]\right]\\
&\qquad (199)\\
&\le \min_{\delta\ge0}\sum_{S_1}\left\{\mathbb{E}_{\mathbf{A}_{all}}\left[\mathbb{P}\left[\left\{\sum_{i=1}^{n}(1-\omega-\omega\lambda_i)\frac{\chi_i^2(2L)}{2} > C_F\right\}\right.\right.\right.\\
&\qquad\qquad \left.\left.\left.\bigcap\mathcal{G}_\delta\middle|\mathbf{A}_{all}\right]\right] + \mathbb{P}\left[\mathcal{G}_\delta^c\right]\right\} \qquad (200)\\
&= \min_{\delta\ge0}\left\{\sum_{S_1}q_{3,t}(\omega,\nu) + \binom{K_a}{t}\left(1 - \frac{\gamma\left(nL, nL\ (1+\delta)\right)}{\Gamma(nL)}\right)\right\},\\
&\qquad (201)
\end{aligned}$$

where $C_F = \omega L\ln|\mathbf{F}_1| - L\ln|\mathbf{F}| + nL\nu$; $\lambda_1,\ldots,\lambda_n$ are eigenvalues of $\mathbf{F}_1^{-1}\mathbf{A}_{S_1}\mathbf{A}_{S_1}^H$ in decreasing order with the first $m = \min\{n,t\}$ eigenvalues being positive and all of the rest being 0. Here, (199) follows from the union bound and the fact that $\mathbf{y}_l = \mathbf{F}^{\frac{1}{2}}\tilde{\mathbf{y}}_l \overset{i.i.d.}{\sim} \mathcal{CN}(\mathbf{0},\mathbf{F})$ conditioned on $\mathbf{A}_{\mathcal{K}_a}$, where $\tilde{\mathbf{y}}_l \overset{i.i.d.}{\sim} \mathcal{CN}(\mathbf{0},\mathbf{I}_n)$ for $l \in [L]$; (200) follows from the bounding technique in (12), and the fact that conditioned on $\mathbf{A}_{all}$, $\mathcal{U}\tilde{\mathbf{y}}_l$ and $\tilde{\mathbf{y}}_l$ have the same distribution as $\mathcal{CN}(\mathbf{0},\mathbf{I}_n)$ for the unitary matrix $\mathcal{U}$ satisfying $\mathcal{U}^H\left(\mathbf{I}_n - \omega\mathbf{F}^{\frac{H}{2}}\mathbf{F}_1^{-1}\mathbf{F}^{\frac{1}{2}}\right)\mathcal{U} = \boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda} = \operatorname{diag}\{1-\omega-\omega\lambda_1,\ldots,1-\omega-\omega\lambda_n\}$; (201) holds because $\sum_{i=1}^{n}\chi_i^2(2L)$ has the same distribution as $\chi^2(2nL)$ considering that $\chi_i^2(2L)$, $i = 1,\ldots,n$, are independent.

The first term on the RHS of (201) can be upper-bounded as

$$\begin{aligned}
&\sum_{S_1}q_{3,t}(\omega,\nu)\\
&\le \sum_{S_1}\mathbb{E}_{\mathbf{A}_{all}}\left[\mathbb{P}\left[\sum_{i=1}^{n}\lambda_i\frac{\chi_i^2(2L)}{2} < \frac{nL(1+\delta)(1-\omega)}{\omega}\right.\right.\\
&\qquad\qquad \left.\left. - \frac{C_F}{\omega}\middle|\mathbf{A}_{all}\right]\right] \qquad (202)\\
&\le \binom{K_a}{t}\mathbb{E}_{\mathbf{A}_{all}}\left[\frac{\gamma\left(Lm, L\frac{n(1+\delta)(1-\omega)-\omega\ln|\mathbf{F}_1|+\ln|\mathbf{F}|-n\nu}{\omega\prod_{i=1}^{m}\lambda_i^{1/m}}\right)}{\Gamma(Lm)}\right],\\
&\qquad (203)
\end{aligned}$$

where (203) follows from Lemma 16 and the fact that the number of non-zero eigenvalues of $\mathbf{F}_1^{-1}\mathbf{A}_{S_1}\mathbf{A}_{S_1}^H$ is $m = \min\{n,t\}$, which are denoted as $\lambda_1,\ldots,\lambda_m$ in decreasing order as aforementioned. Substituting (203) into (201), we can obtain an upper bound on $\mathbb{P}\left[\mathcal{G}^c_{\omega,\nu}\right]$, which is denoted as $q_{2,t}(\omega,\nu)$ in (43).

In conclusion, based on Fano's bounding technique, we have obtained $q_{1,t}(\omega,\nu)$ in (38) (i.e. an upper bound on the first probability in (109)) and $q_{2,t}(\omega,\nu)$ in (43) (i.e. an upper bound on the second probability in (109)), which contributes to an upper bound p_t given in (37) on the probability $\mathbb{P}\left[\mathcal{F}_t\right]$. This completes the proof of Theorem 6.

Appendix G
Proof of Theorem 8

In this appendix, we prove Theorem 8 to establish an achievability bound on the PUPE for the scenario in which the number K_a of active users is random and unknown. In this case, the decoder first obtains an estimate K_a' of K_a via an energy-based estimator. Then, given K_a', the decoder produces a set of decoded codewords, which is denoted as $\hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}$. The number of codewords in the set $\hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}$ belongs to an interval around K_a', i.e., it is satisfied that $|\hat{\mathcal{K}}_a| \in [\mathrm{K}_{a,l}^{'}, \mathrm{K}_{a,u}^{'}]$, where $\mathrm{K}_{a,l}' = \max\left\{0, \mathrm{K}_a' - r'\right\}$, $\mathrm{K}_{a,u}' = \min\left\{K, \mathrm{K}_a' + r'\right\}$, and r' denotes a nonnegative integer referred to as the decoding radius. Based on the notation introduced in Appendix A, the per-user probability of misdetection in (7) can be upper-bounded as

$$P_{e,\mathrm{MD}} = \mathbb{E}\left[\frac{1}{K_a}\sum_{k\in\mathcal{K}_a} 1\left[W_k \neq \hat{W}_k\right]\right] \tag{204}$$

$$\leq \sum_{\mathrm{K}_a=1}^{K} P_{K_a}(\mathrm{K}_a) \sum_{\mathrm{K}_a'=0}^{K} \sum_{t\in\mathcal{T}_{\mathrm{K}_a'}} \frac{t + (\mathrm{K}_a - \mathrm{K}_{a,u}')^+}{\mathrm{K}_a} \cdot \mathbb{P}\left[\mathcal{F}_t \cap \{\mathrm{K}_a \to \mathrm{K}_a'\}\right]_{\text{no power constraint}} + p_0. \tag{205}$$

Here, the integer t takes value in $\mathcal{T}_{\mathrm{K}_a'}$ defined in (53) because the number of misdetected codewords, given by $t + (\mathrm{K}_a - \mathrm{K}_{a,u}')^+$, is lower-bounded by $(\mathrm{K}_a - \mathrm{K}_{a,u}')^+$ and upper-bounded by the total number K_a of transmitted messages; $\mathcal{F}_t$ denotes the event that there are exactly $t + (\mathrm{K}_a - \mathrm{K}_{a,u}')^+$ misdetected codewords; $\left\{\mathrm{K}_a \to \mathrm{K}_a^{'}\right\}$ denotes the event that the estimation of K_a results in K_a'; p_0 denotes an upper bound on the total variation distance between the measures with and without power constraint given by

$$p_0 = \mathbb{E}[K_a]\left(1 - \frac{\gamma\left(n, \frac{nP}{P'}\right)}{\Gamma\left(n\right)}\right). \tag{206}$$

Likewise, the per-user probability of false-alarm in (9) can be upper-bounded as

$$P_{e,\mathrm{FA}} = \mathbb{E}\left[\frac{1}{|\hat{\mathcal{K}}_a|}\sum_{k\in\hat{\mathcal{K}}_a} 1\left[\hat{W}_k \neq W_k\right]\right] \tag{207}$$

$$\leq \sum_{\mathrm{K}_a=0}^{K} P_{K_a}(\mathrm{K}_a) \sum_{\mathrm{K}_a'=0}^{K} \sum_{t\in\mathcal{T}_{\mathrm{K}_a'}} \sum_{t'\in\mathcal{T}_{\mathrm{K}_a',t}} \frac{t' + (\mathrm{K}_{a,l}^{'} - \mathrm{K}_a)^+}{\hat{\mathrm{K}}_a} \cdot \mathbb{P}\left[\mathcal{F}_{t,t'} \cap \{\mathrm{K}_a \to \mathrm{K}_a'\}\right]_{\text{no power constraint}} + p_0, \tag{208}$$

where $\hat{\mathrm{K}}_a$ denotes the number of detected codewords as given in (56); $\mathcal{F}_{t,t'}$ denotes the event that there are exactly $t + (\mathrm{K}_a - \mathrm{K}_{a,u}')^+$ misdetected codewords and $t' + (\mathrm{K}_{a,l}' - \mathrm{K}_a)^+$ falsely alarmed codewords; the integer t' takes value in $\mathcal{T}_{\mathrm{K}_a'}$ defined in (55) because: i) $\hat{\mathrm{K}}_a$ must be in $[\mathrm{K}_{a,l}' : \mathrm{K}_{a,u}']$; ii) the number of falsely alarmed codewords is lower-bounded by $(\mathrm{K}_{a,l}' - \mathrm{K}_a)^+$; iii) there exist falsely alarmed codewords only when $\hat{\mathrm{K}}_a \geq 1$.

Next, we omit the subscript "no power constraint" for the sake of brevity. The probability in the RHS of 205 can be bounded as

$$\mathbb{P}\left[\mathcal{F}_t \cap \{\mathrm{K}_a \to \mathrm{K}_a'\}\right]$$

$$= \mathbb{P}\left[\mathcal{F}_t \cap \left\{|\hat{\mathcal{K}}_a| \in [\mathrm{K}_{a,l}', \mathrm{K}_{a,u}']\right\} \cap \{\mathrm{K}_a \to \mathrm{K}_a'\}\right] \tag{209}$$

$$\leq \min\left\{\mathbb{P}\left[\mathcal{F}_t \cap \left\{|\hat{\mathcal{K}}_a| \in [\mathrm{K}_{a,l}', \mathrm{K}_{a,u}']\right\}\right], \mathbb{P}\left[\mathrm{K}_a \to \mathrm{K}_a'\right]\right\} \tag{210}$$

$$\leq \min\left\{\sum_{t'\in\bar{\mathcal{T}}_{\mathrm{K}_a',t}} \mathbb{P}\left[\mathcal{F}_{t,t'}\middle| |\hat{\mathcal{K}}_a| \in [\mathrm{K}_{a,l}', \mathrm{K}_{a,u}']\right], \mathbb{P}\left[\mathrm{K}_a \to \mathrm{K}_a'\right]\right\}. \tag{211}$$

Here, $\bar{\mathcal{T}}_{\mathrm{K}_a',t}$ is defined in (54), which is obtained similar to $\mathcal{T}_{\mathrm{K}_a',t}$ with the difference that the number $\hat{\mathrm{K}}_a$ of detected codewords can be 0; (209) follows because the event $\mathrm{K}_a \to \mathrm{K}_a'$ implies that $|\hat{\mathcal{K}}_a| \in [\mathrm{K}_{a,l}', \mathrm{K}_{a,u}']$ [19]; (210) follows from the fact that the joint probability is upper-bounded by each of the individual probabilities. Similarly, the probability in the RHS of 208 can be bounded as

$$\mathbb{P}\left[\mathcal{F}_{t,t'} \cap \{\mathrm{K}_a \to \mathrm{K}_a'\}\right]$$

$$\leq \min\left\{\mathbb{P}\left[\mathcal{F}_{t,t'}\middle| |\hat{\mathcal{K}}_a| \in [\mathrm{K}_{a,l}', \mathrm{K}_{a,u}']\right], \mathbb{P}\left[\mathrm{K}_a \to \mathrm{K}_a'\right]\right\}. \tag{212}$$

In the remainder of this part, we proceed to upper-bound $\mathbb{P}\left[\mathcal{F}_{t,t'}\middle| |\hat{\mathcal{K}}_a| \in [\mathrm{K}_{a,l}', \mathrm{K}_{a,u}']\right]$ and $\mathbb{P}\left[\mathrm{K}_a \to \mathrm{K}_a'\right]$, respectively, which are two ingredients of upper-bounding $P_{e,\mathrm{MD}}$ and $P_{e,\mathrm{FA}}$.

A. Upper-Bounding $\mathbb{P}[\mathrm{K}_a \to \mathrm{K}_a']$

Given the channel output $\mathbf{Y}$, the receiver estimates the number K_a of active users as

$$\mathrm{K}_a' = \arg\min_{\tilde{\mathrm{K}}_a \in [\mathrm{K}_l, \mathrm{K}_u]} m(\mathbf{Y}, \tilde{\mathrm{K}}_a), \tag{213}$$

where $m(\mathbf{Y}, \tilde{\mathrm{K}}_a)$ denotes the energy-based estimation metric given by

$$m\left(\mathbf{Y}, \tilde{\mathrm{K}}_a\right) = \left| \|\mathbf{Y}\|_F^2 - nL\left(1 + \tilde{\mathrm{K}}_a P'\right)\right|. \tag{214}$$

Denote $C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a} = \frac{\mathrm{K}_a' + \tilde{\mathrm{K}}_a}{2}$. In the case of $\tilde{\mathrm{K}}_a \neq \mathrm{K}_a'$, the event $m\left(\mathbf{Y}, \mathrm{K}_a'\right) \leq m(\mathbf{Y}, \tilde{\mathrm{K}}_a)$ is equivalent to

$$\begin{cases} \|\mathbf{Y}\|_F^2 \leq nL\left(1 + C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a} P'\right), & \text{if } \mathrm{K}_a' < \tilde{\mathrm{K}}_a \\ \|\mathbf{Y}\|_F^2 \geq nL\left(1 + C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a} P'\right), & \text{if } \mathrm{K}_a' > \tilde{\mathrm{K}}_a \end{cases}. \tag{215}$$

As a result, the probability of the event that K_a is estimated as K_a' is upper-bounded as

$$\mathbb{P}\left[\mathrm{K}_a \to \mathrm{K}_a'\right]$$

$$\leq \mathbb{P}\left[m\left(\mathbf{Y}, \mathrm{K}_a'\right) \leq m(\mathbf{Y}, \tilde{\mathrm{K}}_a), \forall \tilde{\mathrm{K}}_a \neq \mathrm{K}_a'\right] \tag{216}$$

$$\leq \min_{\tilde{\mathrm{K}}_a \in [0:K], \tilde{\mathrm{K}}_a \neq \mathrm{K}_a'} \mathbb{P}\left[m\left(\mathbf{Y}, \mathrm{K}_a'\right) \leq m(\mathbf{Y}, \tilde{\mathrm{K}}_a)\right] \tag{217}$$

$$= \min_{\substack{\tilde{\mathrm{K}}_a \in [0:K] \\ \tilde{\mathrm{K}}_a \neq \mathrm{K}_a'}} 1\left[\mathrm{K}_a' < \tilde{\mathrm{K}}_a\right] \mathbb{P}\left[\|\mathbf{Y}\|_F^2 \leq nL\left(1 + C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a} P'\right)\right]$$

$$+\,1\Big[\mathrm{K}_a'\!>\!\tilde{\mathrm{K}}_a\Big]\,\mathbb{P}\Big[\|\mathbf{Y}\|_F^2\geq nL\Big(1+C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'\Big)\Big]. \tag{218}$$

The probability $\mathbb{P}\left[\|\mathbf{Y}\|_F^2 \leq nL\left(1+C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'\right)\right]$ on the RHS of (218) can be upper-bounded following two approaches. First, applying the Chernoff bound and Lemma 15, we can obtain

$$\begin{aligned}
&\mathbb{P}\left[\|\mathbf{Y}\|_F^2 \leq nL\left(1+C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'\right)\right]\\
&\leq \mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}}\left[\min_{\rho\geq 0}\exp\Big\{\rho nL\Big(1+C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'\Big)-L\ln|\mathbf{I}_n+\rho\mathbf{F}|\Big\}\right],
\end{aligned} \tag{219}$$

where $\mathbf{A}_{\mathcal{K}_a} \in \mathbb{C}^{n\times \mathrm{K}_a}$ denotes the concatenation of the transmitted codewords of K_a active users and $\mathbf{F} = \mathbf{I}_n + \mathbf{A}_{\mathcal{K}_a}\mathbf{A}_{\mathcal{K}_a}^H$. Define the event $\mathcal{G}_\eta = \left\{\sum_{i=1}^n \frac{\chi_i^2(2L)}{2} \geq nL\eta\right\}$ for $\eta \geq 0$. Then, we can obtain another upper bound on $\mathbb{P}\left[\|\mathbf{Y}\|_F^2 \leq nL\left(1+C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'\right)\right]$ as follows:

$$\begin{aligned}
&\mathbb{P}\left[\|\mathbf{Y}\|_F^2 \leq nL\left(1+C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'\right)\right]\\
&= \mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}}\left[\mathbb{P}\left[\sum_{l=1}^{L}\tilde{\mathbf{y}}_l^H\left(\mathbf{I}_n+\mathbf{A}_{\mathcal{K}_a}\mathbf{A}_{\mathcal{K}_a}^H\right)\tilde{\mathbf{y}}_l \leq nL\right.\right.\\
&\qquad\qquad\left.\left. +nLC_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'\right|\mathbf{A}_{\mathcal{K}_a}\right]\Bigg]
\end{aligned} \tag{220}$$

$$\begin{aligned}
&\leq \min_{\eta>0}\left\{\mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}}\left[\mathbb{P}\left[\left\{\sum_{i=1}^{n}(1+\lambda_i')\frac{\chi_i^2(2L)}{2}\leq nL\Big(1+C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'\Big)\right\}\right.\right.\right.\\
&\qquad\qquad\left.\left.\left. \cap\mathcal{G}_\eta\right|\mathbf{A}_{\mathcal{K}_a}\right]\Bigg]+\mathbb{P}\left[\mathcal{G}_\eta^c\right]\right\}
\end{aligned} \tag{221}$$

$$\begin{aligned}
&\leq \min_{\eta>0}\left\{\mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}}\left[\mathbb{P}\left[\prod_{i=1}^{m'}(\lambda_i')^{\frac{1}{m'}}\frac{\chi^2(2Lm')}{2}\leq nL\,(1-\eta)\right.\right.\right.\\
&\qquad\qquad\left.\left.\left. +nLC_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'\right|\mathbf{A}_{\mathcal{K}_a}\right]\Bigg]+\mathbb{P}\big[\mathcal{G}_\eta^c\big]\right\}
\end{aligned} \tag{222}$$

$$= \min_{\eta>0}\left\{\mathbb{E}_{\mathbf{A}_{\mathcal{K}_a}}\left[\frac{\gamma\left(Lm', \frac{nL\left(1+C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'-\eta\right)}{\prod_{i=1}^{m'}(\lambda_i')^{\frac{1}{m'}}}\right)}{\Gamma\left(Lm'\right)}\right]+\frac{\gamma(nL,nL\eta)}{\Gamma\left(nL\right)}\right\}, \tag{223}$$

where $\lambda_1',\ldots,\lambda_n'$ are eigenvalues of $\mathbf{A}_{\mathcal{K}_a}\mathbf{A}_{\mathcal{K}_a}^H$ in decreasing order with the first $m' = \min\{n,\mathrm{K}_a\}$ eigenvalues being positive and others being 0. Here, (220) holds because $\mathbf{y}_l = \mathbf{F}^{\frac{1}{2}}\tilde{\mathbf{y}}_l \sim \mathcal{CN}\left(\mathbf{0},\mathbf{F}\right)$ conditioned on $\mathbf{A}_{\mathcal{K}_a}$, where $\tilde{\mathbf{y}}_l \sim \mathcal{CN}\left(\mathbf{0},\mathbf{I}_n\right)$; (221) follows from the "good region" technique in (12); (222) follows from Lemma 16. Taking the minimum of (219) and (223), we obtain the ultimate upper bound on $\mathbb{P}\left[\|\mathbf{Y}\|_F^2 \leq nL\left(1+C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'\right)\right]$ denoted as $p_{\mathrm{K}_a\to\mathrm{K}_a',1}$ in (72).

Likewise, we can derive two upper bounds on the probability $\mathbb{P}\left[\|\mathbf{Y}\|_F^2 \geq nL\left(1+C_{\mathrm{K}_a',\tilde{\mathrm{K}}_a}P'\right)\right]$. Taking the minimum value of them, we obtain the ultimate upper bound on it, denoted as $p_{\mathrm{K}_a\to\mathrm{K}_a',2}$ in (73).

B. Upper-Bounding $\mathbb{P}\left[\mathcal{F}_{t,t'}\,\middle|\,|\hat{\mathcal{K}}_a|\in[\mathrm{K}_{a,l}',\mathrm{K}_{a,u}']\right]$

In this subsection, we utilize the MAP decoder to upper-bound $\mathbb{P}\left[\mathcal{F}_{t,t'}\,\middle|\,|\hat{\mathcal{K}}_a|\in[\mathrm{K}_{a,l}',\mathrm{K}_{a,u}']\right]$. Under the condition that $|\hat{\mathcal{K}}_a|\in[\mathrm{K}_{a,l}',\mathrm{K}_{a,u}']$, the outputs of the decoder are given by

$$\begin{aligned}
&\left[\hat{\mathcal{K}}_a,\hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right]\\
&= \arg\min_{\hat{\mathcal{K}}_a\subset\mathcal{K},|\hat{\mathcal{K}}_a|\in[\mathrm{K}_{a,l}',\mathrm{K}_{a,u}']}\;\min_{\left(\hat{\mathbf{c}}_{(k)}\in\mathcal{C}_k\right)_{k\in\hat{\mathcal{K}}_a}} g\left(\mathbf{Y},\hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right),
\end{aligned} \tag{224}$$

$$\hat{W}_k = f_{\mathrm{en},k}^{-1}\left(\hat{\mathbf{c}}_{(k)}\right),\quad k\in\hat{\mathcal{K}}_a, \tag{225}$$

where the MAP decoding metric $g\left(\mathbf{Y},\hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right)$ is given by

$$\begin{aligned}
&g\left(\mathbf{Y},\hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right)\\
&= L\ln\left|\mathbf{I}_n+\mathbf{A}\mathbf{\Gamma}_{\hat{\mathcal{K}}_a}'\mathbf{A}^H\right|+\mathrm{tr}\left(\left(\mathbf{I}_n+\mathbf{A}\mathbf{\Gamma}_{\hat{\mathcal{K}}_a}'\mathbf{A}^H\right)^{-1}\mathbf{Y}\mathbf{Y}^H\right)\\
&\quad-\ln\left(P_{\mathrm{K}_a}(|\hat{\mathcal{K}}_a|)M^{-|\hat{\mathcal{K}}_a|}\right).
\end{aligned} \tag{226}$$

Here, $\mathbf{\Gamma}_S^{'}$ is defined in Appendix F. In the remainder of this part, $g\left(\mathbf{Y},\hat{\mathbf{c}}_{[\hat{\mathcal{K}}_a]}\right)$ is denoted as $g\left(\hat{\mathbf{\Gamma}}_{\hat{\mathcal{K}}_a}\right)$ for simplicity.

Let the set $S_1 \subset \mathcal{K}_a$ of size $t+(\mathrm{K}_a-\mathrm{K}_{a,u}')^+$ denote the set of users whose codewords are misdecoded. The set S_1 can be divided into two subsets $S_{1,1}$ and $S_{1,2}$ of size $(\mathrm{K}_a-\mathrm{K}_{a,u}')^+$ and t, respectively. Let the set $S_2\subset\mathcal{K}\backslash\mathcal{K}_a\cup S_1$ of size $t'+(\mathrm{K}_{a,l}'-\mathrm{K}_a)^+$ denote the set of detected users with false-alarm codewords. Let $S_{2,1}$ denote an arbitrary subset of S_2 of size $(\mathrm{K}_{a,l}'-\mathrm{K}_a)^+$. For the sake of simplicity, we rewrite "$\bigcup_{S_1\subset\mathcal{K}_a,|S_1|=t+(\mathrm{K}_a-\mathrm{K}_{a,u}')^+}$" to "$\bigcup_{S_1}$" and "$\bigcup_{S_2\subset\mathcal{K}\backslash\mathcal{K}_a\cup S_1,|S_2|=t'+(\mathrm{K}_{a,l}'-\mathrm{K}_a)^+}$" to "$\bigcup_{S_2}$"; similarly for $\sum$ and $\bigcap$. We rewrite $\left\{\mathbf{c}_{(k)}'\in\mathcal{C}_k : k\in S_2,\mathbf{c}_{(k)}'\neq\mathbf{c}_{(k)}\right\}$ to $\mathbf{c}_{[S_2]}^{'}$ for short, which denotes the set of false alarm codewords corresponding to users in the set S_2. Define $\mathbf{A}_S$, $\mathbf{A}_S^{'}$, $\mathbf{\Gamma}_S$ and $\mathbf{\Gamma}_S^{'}$ as in Appendix F. Define the event $\mathcal{G}_{\omega,\nu}=\bigcap_{S_1}\left\{\mathbf{Y}\in\mathcal{R}_{t,S_1}\right\}$ as in Appendix A. Following similar ideas in (109), we have

$$\begin{aligned}
&\mathbb{P}\left[\mathcal{F}_{t,t'}\,\middle|\,|\hat{\mathcal{K}}_a|\in[\mathrm{K}_{a,l}',\mathrm{K}_{a,u}']\right]\\
&\leq \min_{\substack{0\leq\omega\leq 1\\ \nu\geq 0}}\left\{\mathbb{P}\left[\bigcup_{S_1}\bigcup_{S_2}\bigcup_{\mathbf{c}_{[S_2]}'}\left\{g\left(\mathbf{\Gamma}_{\mathcal{K}_a\backslash S_1\cup S_2}^{'}\right)\leq g\left(\mathbf{\Gamma}_{\mathcal{K}_a\backslash S_{1,1}\cup S_{2,1}}^{'}\right)\right\}\right.\right.\\
&\qquad\qquad\left.\bigcap\;\mathcal{G}_{\omega,\nu}\,\middle|\,|\hat{\mathcal{K}}_a|\in[\mathrm{K}_{a,l}',\mathrm{K}_{a,u}']\right]\\
&\qquad\quad\left.+\,\mathbb{P}\Big[\mathcal{G}_{\omega,\nu}^c\,\Big|\,|\hat{\mathcal{K}}_a|\in[\mathrm{K}_{a,l}',\mathrm{K}_{a,u}']\Big]\right\}.
\end{aligned} \tag{227}$$

Similar to (111) and (197), we can obtain an upper bound on the first probability on the RHS of (227), which is denoted as $q_{1,\mathrm{K}_a',t,t'}$ in (60). In the case of $t+(\mathrm{K}_a-\mathrm{K}_{a,u}')^+>0$, the second probability on the RHS of (227) can be bounded as in Appendix F. When $t+(\mathrm{K}_a-\mathrm{K}_{a,u}')^+=0$, we have

$$\mathbb{P}[\mathcal{G}_{\omega,\nu}^c] = \mathbb{P}\left[g\left(\mathbf{\Gamma}_{\mathcal{K}_a}\right) > \frac{nL\nu}{1-\omega}\right] \tag{228}$$

$$= \mathbb{E}_{\mathbf{A}_{K_a}}\left[1-\frac{\gamma\left(nL, \frac{nL\nu}{1-\omega}-L\ln|\mathbf{F}|+b\right)}{\Gamma\left(nL\right)}\right], \quad (229)$$

where the constant b is given in (65). The RHS of (229) is denoted as $q_{2,\mathrm{K}'_a,t,0}$ in (70). This concludes the proof of Theorem 8.

Appendix H
Proof of Theorem 9

In this appendix, we prove Theorem 9 to establish a Fano type converse bound on the minimum required energy-per-bit for the no-CSI case. Let $\bar{\mathbf{y}} = \left[\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_L^T\right]^T \in \mathbb{C}^{nL\times 1}$ be a vector obtained by concatenating the received signals of L antennas at the BS. Let $\bar{\mathbf{X}}_{K_aM}$ be an $n \times K_aM$ submatrix of $\mathbf{X}$ including codebooks of K_a active users and denote $\bar{\mathbf{X}} = \mathrm{diag}\left\{\bar{\mathbf{X}}_{K_aM}, \ldots, \bar{\mathbf{X}}_{K_aM}\right\} \in \mathbb{C}^{nL\times K_aML}$. Let $\bar{\mathbf{H}}_l \in \mathbb{C}^{K_aM\times K_aM}$ be a block diagonal matrix, where block k is a diagonal $M \times M$ matrix with all diagonal entries equal to $h_{k,l} \sim \mathcal{CN}(0,1)$. Let $\bar{\mathbf{H}} = \left[\bar{\mathbf{H}}_1, \ldots, \bar{\mathbf{H}}_L\right]^T \in \mathbb{C}^{K_aML\times K_aM}$. The vector $\bar{\boldsymbol{\beta}} \in \{0,1\}^{K_aM}$ includes K_a blocks, where each block is of size M and includes one 1; we have $\left[\bar{\boldsymbol{\beta}}\right]_{(k-1)M+W_k} = 1$ if the W_k-th codeword is transmitted by user k, and $\left[\bar{\boldsymbol{\beta}}\right]_{(k-1)M+W_k} = 0$ otherwise. Then, we can model the communication system as

$$\bar{\mathbf{y}} = \bar{\mathbf{X}}\bar{\mathbf{H}}\bar{\boldsymbol{\beta}} + \bar{\mathbf{z}}, \quad (230)$$

where $\bar{\mathbf{z}} \in \mathbb{C}^{nL\times 1}$ with each entry i.i.d. from $\mathcal{CN}(0,1)$.

We assume a genie reveals the set of active users. Similar to the analysis in Appendix D, we have [8]

$$(1-\epsilon)\,J - h_2(\epsilon) \le \frac{1}{K_a} I_2\left(\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\mathbf{X}}\right). \quad (231)$$

Based on the chain rule of the mutual information, we have

$$\begin{aligned} & I_2\left(\bar{\boldsymbol{\beta}},\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\mathbf{X}}\right) \\ &= I_2\left(\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\mathbf{X}}\right) + I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\boldsymbol{\beta}},\bar{\mathbf{X}}\right) \quad (232) \\ &= I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\mathbf{X}}\right) + I_2\left(\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\mathbf{H}}\bar{\boldsymbol{\beta}},\bar{\mathbf{X}}\right). \quad (233) \end{aligned}$$

Since $\bar{\boldsymbol{\beta}} \to \bar{\mathbf{H}}\bar{\boldsymbol{\beta}} \to \left(\bar{\mathbf{y}},\bar{\mathbf{X}}\right)$ forms a Markov chain, the mutual information $I_2\left(\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\mathbf{H}}\bar{\boldsymbol{\beta}},\bar{\mathbf{X}}\right) = 0$. Hence, we have [42, Eq. (78)]

$$I_2\left(\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\mathbf{X}}\right) = I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\mathbf{X}}\right) - I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\boldsymbol{\beta}},\bar{\mathbf{X}}\right). \quad (234)$$

Next, we focus on the two terms on the RHS of (234). We have

$$\begin{aligned} & I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\mathbf{X}} = \bar{\mathbf{X}}^r\right) \\ &= I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{X}}^r\bar{\mathbf{H}}\bar{\boldsymbol{\beta}}+\bar{\mathbf{z}}\right) \quad (235) \\ &\le \sup_{\mathbf{u}} I_2\left(\mathbf{u};\bar{\mathbf{X}}^r\mathbf{u}+\bar{\mathbf{z}}\right) \quad (236) \\ &= \log_2\left|\mathbf{I}_{nL} + \frac{1}{M}\bar{\mathbf{X}}^r\left(\bar{\mathbf{X}}^r\right)^H\right| \quad (237) \\ &= L\log_2\left|\mathbf{I}_n + \frac{1}{M}\bar{\mathbf{X}}^r_{K_aM}\left(\bar{\mathbf{X}}^r_{K_aM}\right)^H\right|, \quad (238) \end{aligned}$$

where the matrix $\bar{\mathbf{X}}^r_{K_aM}$ is a realization of $\bar{\mathbf{X}}_{K_aM}$ and $\bar{\mathbf{X}}^r = \mathrm{diag}\left\{\bar{\mathbf{X}}^r_{K_aM}, \ldots, \bar{\mathbf{X}}^r_{K_aM}\right\}$ is a realization of $\bar{\mathbf{X}}$. The supremum in (236) is over $\mathbf{u}$ with $\mathbb{E}[\mathbf{u}] = \mathbf{0}$ and $\mathbb{E}\left[\mathbf{u}\mathbf{u}^H\right] = \mathbb{E}\left[\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}}\right)\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}}\right)^H\right] = \frac{1}{M}\mathbf{I}_{K_aML}$. The supremum is achieved when $\mathbf{u} \sim \mathcal{CN}\left(\mathbf{0}, \frac{1}{M}\mathbf{I}_{K_aML}\right)$ [42], which implies (237). Then, we have

$$I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\mathbf{X}}\right) \le L\mathbb{E}\left[\log_2\left|\mathbf{I}_n + \frac{1}{M}\bar{\mathbf{X}}_{K_aM}\bar{\mathbf{X}}^H_{K_aM}\right|\right]. \quad (239)$$

Under the assumption that the entries of codebooks are i.i.d. with mean zero and variance P, the expectation on the RHS of (239) can be upper-bounded as

$$\begin{aligned} & \mathbb{E}\left[\log_2\left|\mathbf{I}_n + \frac{1}{M}\bar{\mathbf{X}}_{K_aM}\bar{\mathbf{X}}^H_{K_aM}\right|\right] \\ &\le \min\left\{n\log_2(1+K_aP)\,, K_aM\log_2\left(1+\frac{nP}{M}\right)\right\}, \quad (240) \end{aligned}$$

where (240) follows from the concavity of the $\log_2|\cdot|$ function. We denote the RHS of (240) as C for simplicity.

A lower bound on $I\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\boldsymbol{\beta}},\bar{\mathbf{X}}\right)$ can be derived as follows. Let $\tilde{\mathbf{X}}_{K_a} \in \mathbb{C}^{n\times K_a}$ be a submatrix of $\mathbf{X}$ formed by columns corresponding to the support of $\bar{\boldsymbol{\beta}}$. Let $\tilde{\mathbf{H}}_{K_a}$ be a $K_a \times L$ submatrix of $\mathbf{H}$ including fading coefficients between K_a active users and L antennas of the receiver. Then, the received signal given in (2) can be rewritten as

$$\mathbf{Y} = \tilde{\mathbf{X}}_{K_a}\tilde{\mathbf{H}}_{K_a} + \mathbf{Z}. \quad (241)$$

We have

$$\begin{aligned} & I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}}^r, \bar{\mathbf{X}} = \bar{\mathbf{X}}^r\right) \\ &= I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}}^r;\bar{\mathbf{X}}^r\bar{\mathbf{H}}\bar{\boldsymbol{\beta}}^r+\bar{\mathbf{z}}\right) \quad (242) \\ &= I_2\left(\tilde{\mathbf{H}}_{K_a};\tilde{\mathbf{X}}^r_{K_a}\tilde{\mathbf{H}}_{K_a}+\mathbf{Z}\right) \quad (243) \\ &= L\log_2\left|\mathbf{I}_n + \tilde{\mathbf{X}}^r_{K_a}\left(\tilde{\mathbf{X}}^r_{K_a}\right)^H\right|, \quad (244) \end{aligned}$$

where $\bar{\boldsymbol{\beta}}^r$ is a realization of $\bar{\boldsymbol{\beta}}$ and $\tilde{\mathbf{X}}^r_{K_a}$ is a realization of $\tilde{\mathbf{X}}_{K_a}$. Hence, applying Sylvester's determinant theorem, we have

$$\begin{aligned} I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\boldsymbol{\beta}},\bar{\mathbf{X}}\right) &= L\mathbb{E}\left[\log_2\left|\mathbf{I}_n + \tilde{\mathbf{X}}_{K_a}\tilde{\mathbf{X}}^H_{K_a}\right|\right] \quad (245) \\ &= L\mathbb{E}\left[\log_2\left|\mathbf{I}_{K_a} + \tilde{\mathbf{X}}^H_{K_a}\tilde{\mathbf{X}}_{K_a}\right|\right]. \quad (246) \end{aligned}$$

Substituting (240) and (246) into (234), we obtain an upper bound on $I\left(\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\mathbf{X}}\right)$. Substituting this bound into (231), the proof of (75) in Theorem 9 is completed.

Under the assumption that K users generate their codebooks independently with each entry i.i.d. from $\mathcal{CN}(0,P)$, we further lower-bound $\mathbb{E}\left[\log_2\left|\mathbf{I}_{K_a} + \tilde{\mathbf{X}}^H_{K_a}\tilde{\mathbf{X}}_{K_a}\right|\right]$ in (246) in the remainder of this appendix. In the case of $K_a > n$, let $\alpha_i \sim \frac{\chi^2(2(K_a-i+1))}{2}$ for $i = 1, \ldots, n$. We have

$$\begin{aligned} & \mathbb{E}\left[\log_2\left|\mathbf{I}_{K_a} + \tilde{\mathbf{X}}^H_{K_a}\tilde{\mathbf{X}}_{K_a}\right|\right] \\ &\ge \sum_{i=1}^{n}\mathbb{E}\left[\log_2\left(1+P\alpha_i\right)\right] \quad (247) \\ &= \sum_{i=1}^{n}\mathbb{E}\left[\log_2\alpha_i\right] + \sum_{i=1}^{n}\mathbb{E}\left[\log_2\left(P+\frac{1}{\alpha_i}\right)\right] \quad (248) \\ &\ge \log_2 e\sum_{i=1}^{n}\psi(K_a-i+1) + \sum_{i=1}^{n}\log_2\left(P+\frac{1}{K_a-i+1}\right), \\ & \quad (249) \end{aligned}$$

where (247) follows from Lemma 18 shown below; (249) follows because $\mathbb{E}\left[\ln\frac{\chi^2(2b)}{2}\right] = \psi(b)$ with $\psi(x)$ denoting Euler's digamma function, and follows from Jensen's inequality considering $\log_2\left(P+\frac{1}{x}\right)$ is a convex function of x. Let $b_1 = J(1-\epsilon) - h_2(\epsilon)$. Substituting (249) into (75), we have

$$b_1 \leq \frac{LC}{K_a} - \frac{L}{K_a}\sum_{i=1}^{n}\Bigg(\psi(K_a-i+1)\log_2 e + \log_2\left(P+\frac{1}{K_a-i+1}\right)\Bigg). \tag{250}$$

Lemma 18 (Section 4.1.1 in [51]): For $b>0$. A central complex Wishart matrix $\mathbf{W}\sim\mathcal{W}_m(n,\mathbf{I})$, with $n\geq m$, satisfies

$$\mathbb{E}\left[\log_2\left|\mathbf{I}_m+b\mathbf{W}\right|\right] > \sum_{i=n-m+1}^{n}\mathbb{E}\left[\log_2\left(1+b\frac{\chi^2(2i)}{2}\right)\right], \tag{251}$$

where $\chi^2(2i)$ is a chi-square variate with $2i$ degrees of freedom.

Likewise, when $K_a\leq n$, we have

$$\mathbb{E}\Big[\log_2\left|\mathbf{I}_{K_a}+\tilde{\mathbf{X}}_{K_a}^H\tilde{\mathbf{X}}_{K_a}\right|\Big] \geq \log_2 e\sum_{i=1}^{K_a}\psi(n-i+1)+\sum_{i=1}^{K_a}\log_2\left(P+\frac{1}{n-i+1}\right). \tag{252}$$

Substituting (252) into (75), when $K_a\leq n$, we have

$$b_1 \leq \frac{LC}{K_a} - \frac{L}{K_a}\sum_{i=1}^{K_a}\Bigg(\psi(n-i+1)\log_2 e + \log_2\left(P+\frac{1}{n-i+1}\right)\Bigg). \tag{253}$$

Together with (250), the proof of (77) is completed, which concludes the proof of Theorem 9.

Appendix I
Proof of Theorem 10

In this appendix, we prove Theorem 10 to establish a converse bound on the minimum required energy-per-bit for the case in which there is no CSI at the receiver and the number K_a of active users is random and unknown. In Appendix I-A, we establish a converse bound for the scenario with multiple users; in Appendix I-B, we establish a converse bound for the scenario with knowledge of the activities of $K-1$ potential users and the transmitted codewords and channel coefficients of active users among them, which is also a converse bound for the massive random access problem.

A. Multiple-User Random Access Converse Bound

In this part, we use the Fano inequality to derive a converse bound on the minimum required energy-per-bit for the multiple-user case when K_a is random and unknown. Define $\bar{\mathbf{y}}$, $\bar{\mathbf{X}}$, $\bar{\mathbf{X}}_{KM}$, $\bar{\mathbf{H}}_l$, and $\bar{\mathbf{H}}$ as in Appendix H. Let the vector $\bar{\boldsymbol{\beta}}\in\{0,1\}^{KM}$ indicate which codewords are transmitted by active users, which includes K blocks with each block of size M and including at most one 1. Specifically, according to the random access model described in Section II, for $m\in[M]$ and $k\in[K]$, we have $\mathbb{P}\left[\left[\bar{\boldsymbol{\beta}}\right]_{(k-1)M+m}=0\right] = 1-\frac{p_a}{M}$ and $\mathbb{P}\left[\left[\bar{\boldsymbol{\beta}}\right]_{(k-1)M+m}=1\right] = \frac{p_a}{M}$. Then, we can model the communication system as

$$\bar{\mathbf{y}} = \bar{\mathbf{X}}\bar{\mathbf{H}}\bar{\boldsymbol{\beta}}+\bar{\mathbf{z}}, \tag{254}$$

where $\bar{\mathbf{z}}\in\mathbb{C}^{nL\times 1}$ with each entry i.i.d. from $\mathcal{CN}(0,1)$.

Let $\mathcal{M}_k = 1\left[W_k\neq\hat{W}_k\right]$ and $P_{e,k}=\mathbb{E}\left[\mathcal{M}_k\right]$. The error requirements in (7) and (9) can be loosened to

$$P_e = \frac{1}{K}\sum_{k\in\mathcal{K}}P_{e,k}\leq\epsilon_{\mathrm{MD}}+\epsilon_{\mathrm{FA}}. \tag{255}$$

For $k\in\mathcal{K}$, a Fano type argument gives

$$(1-P_{e,k})H_2(W_k|\bar{\mathbf{X}})-h_2(P_{e,k})\leq I_2(W_k;\hat{W}_k|\bar{\mathbf{X}}), \tag{256}$$

where $H_2(x)$ denotes the entropy of a random variable x and $h_2(\cdot)$ denotes the binary entropy function. The entropy $H_2(W_k|\bar{\mathbf{X}})$ can be computed as

$$H_2(W_k|\bar{\mathbf{X}}) = H_2(W_k) \tag{257}$$

$$= -(1-p_a)\log_2(1-p_a)-p_a\log_2\frac{p_a}{M} \tag{258}$$

$$= h_2(p_a)+p_aJ. \tag{259}$$

Substituting (259) into (256) and taking the summation over $k\in\mathcal{K}$ on both sides of (256), we have

$$K(1-P_e)\left(h_2(p_a)+p_aJ\right)-\sum_{k\in\mathcal{K}}h_2(P_{e,k}) \leq \sum_{k\in\mathcal{K}}I_2(W_k;\hat{W}_k\,|\,\bar{\mathbf{X}}). \tag{260}$$

Considering the concavity of $h_2(\cdot)$ and the inequality that $P_e\leq\epsilon_{\mathrm{MD}}+\epsilon_{\mathrm{FA}}\leq 1-\frac{1}{1+2^{h_2(p_a)+p_aJ}}$, we have

$$P_e\left(h_2(p_a)+p_aJ\right)+\frac{1}{K}\sum_{k\in\mathcal{K}}h_2(P_{e,k}) \leq (\epsilon_{\mathrm{MD}}+\epsilon_{\mathrm{FA}})\left(h_2(p_a)+p_aJ\right)+h_2(\epsilon_{\mathrm{MD}}+\epsilon_{\mathrm{FA}}). \tag{261}$$

Moreover, applying (173), we have $\sum_{k\in\mathcal{K}}I_2(W_k;\hat{W}_k|\bar{\mathbf{X}})\leq I_2(W_{\mathcal{K}};\bar{\mathbf{y}}\,|\,\bar{\mathbf{X}}) = I_2(\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\,|\,\bar{\mathbf{X}})$. Together with (234), (260), and (261), we can obtain

$$K(1-\epsilon_{\mathrm{MD}}-\epsilon_{\mathrm{FA}})\left(h_2(p_a)+p_aJ\right)-Kh_2(\epsilon_{\mathrm{MD}}+\epsilon_{\mathrm{FA}}) \leq I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\bar{\mathbf{X}}\right)-I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\bar{\boldsymbol{\beta}},\bar{\mathbf{X}}\right). \tag{262}$$

Next, we focus on the two terms on the RHS of (262). Following from similar ideas used in (238) and (239) with the difference that $\mathbb{E}\left[\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}}\right)\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}}\right)^H\right]=\frac{p_a}{M}\mathbf{I}_{KML}$, we have

$$I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\bar{\mathbf{X}}\right)\leq L\mathbb{E}\left[\log_2\left|\mathbf{I}_n+\frac{p_a}{M}\bar{\mathbf{X}}_{KM}\bar{\mathbf{X}}_{KM}^H\right|\right]. \tag{263}$$

Under the assumption that the entries of codebooks are i.i.d. with mean zero and variance P, the expectation on the RHS of (263) can be upper-bounded as

$$\mathbb{E}\left[\log_2\left|\mathbf{I}_n+\frac{p_a}{M}\bar{\mathbf{X}}_{KM}\bar{\mathbf{X}}_{KM}^H\right|\right]$$

$$\leq \min\left\{ n\log_2\left(1+p_a KP\right), K\ M\log_2\left(1+\frac{p_a}{M}nP\right)\right\}. \tag{264}$$

Moreover, $I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\boldsymbol{\beta}},\bar{\mathbf{X}}\right)$ can be computed as

$$\begin{aligned} & I_2\left(\bar{\mathbf{H}}\bar{\boldsymbol{\beta}};\bar{\mathbf{y}}\middle|\,\bar{\boldsymbol{\beta}},\bar{\mathbf{X}}\right) \\ & = L\,\mathbb{E}\left[\log_2\left|\mathbf{I}_n+\tilde{\mathbf{X}}_{K_a}\tilde{\mathbf{X}}_{K_a}^H\right|\right] && (265) \\ & = L\sum_{\mathrm{K}_a=0}^{K}\left(P_{K_a}(\mathrm{K}_a)\mathbb{E}\left[\log_2\left|\mathbf{I}_n+\tilde{\mathbf{X}}_{K_a}\tilde{\mathbf{X}}_{K_a}^H\right|\middle|\,K_a=\mathrm{K}_a\right]\right), && (266) \end{aligned}$$

where $\tilde{\mathbf{X}}_{K_a}\in\mathbb{C}^{n\times K_a}$ denotes a submatrix of $\mathbf{X}$ formed by columns corresponding to the support of $\bar{\boldsymbol{\beta}}$; $P_{K_a}(\mathrm{K}_a)$ denotes the probability of the event that there are exactly K_a active users given in (5); (265) follows from (246). Combining (262), (263), and (266), the proof of the converse bound for the multiple-user case is completed.

B. Single-User Random Access Converse Bound

The converse bound for the scenario with knowledge of the activities of $K-1$ potential users and the transmitted codewords and channel coefficients of active users among them, can be regarded as a converse bound for the massive random access problem. In this case, it is equivalent to assume that there is a single user in the system with active probability p_a. If this user is active, it equiprobably selects a message W from $\{1,2,\ldots,M\}$, and the corresponding codeword is denoted as $\mathbf{x}_W\in\mathbb{C}^n$ satisfying the maximum power constraint

$$\|\mathbf{x}_W\|_2^2\leq nP. \tag{267}$$

Let $\mathcal{F}\subset\mathbb{C}^n$ be a set of permissible channel inputs as specified by (267). If this user is inactive, we assume $W=0$ and $\mathbf{x}_W=\mathbf{0}$. The received signal is given by

$$\mathbf{Y}=\mathbf{x}_W\mathbf{h}^T+\mathbf{Z}\in\mathbb{C}^{n\times L}, \tag{268}$$

where the vector $\mathbf{h}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}_L)$ includes channel fading coefficients between the user and L antennas at the BS and the noise matrix $\mathbf{Z}\in\mathbb{C}^{n\times L}$ has i.i.d. $\mathcal{CN}(0,1)$ entries.

Denote the decoded message as $\hat{W}\in\{0,1,\ldots,M\}$. We define three types of error probabilities as follows: the probability of the event that the receiver detects the presence of a message even though the user is inactive is given by

$$P_{e,1}=\mathbb{P}\left[\hat{W}\neq 0|W=0\right], \tag{269}$$

the probability of the event that the receiver does not decode correctly a transmitted message is given by

$$P_{e,2}=\frac{1}{M}\sum_{m\in[M]}\mathbb{P}\left[\hat{W}\neq m|W=m\right], \tag{270}$$

and the probability of the event that the receiver erroneously decides that the user is inactive is given by

$$P_{e,3}=\frac{1}{M}\sum_{m\in[M]}\mathbb{P}\left[\hat{W}=0|W=m\right], \tag{271}$$

where $P_{e,3}\leq P_{e,2}$. Then, the error requirements in (7) and (9) can be rewritten as

$$P_{e,\mathrm{MD}}=p_aP_{e,2}\leq\epsilon_{\mathrm{MD}}, \tag{272}$$

$$P_{e,\mathrm{FA}}=(1-p_a)P_{e,1}\leq\epsilon_{\mathrm{FA}}. \tag{273}$$

An upper bound on the number of codewords that are compatible with the requirement that $P_{e,1}$, $P_{e,2}$, and $P_{e,3}$ do not exceed ϵ_1, ϵ_2, and ϵ_3, respectively, is provided in [18, Th. 2]. By changing the error requirement in [18] to (272) and (273) and by considering the multiple-receive-antenna setting, we obtain the following meta-converse result:

Proposition 19: Consider the single-user setup, where the user is active with probability p_a. Let $Q_{Y^{n\times L}}$ be an arbitrary distribution on $\mathcal{Y}^{n\times L}$. When both the CSI and the user activity are unknown, every $(n,M,\epsilon_{\mathrm{MD}},\epsilon_{\mathrm{FA}},P)_{\text{no-CSI,no-}K_a}$ code satisfies

$$M\leq\sup_{\substack{P_{X^n}:\mathbf{x}\in\mathcal{F}\\ \epsilon_1,\epsilon_2,\epsilon_3\in[0,1]}}\frac{1-\beta_{1-\epsilon_1}\left(P_{Y^{n\times L}|X^n=\mathbf{0}},Q_{Y^{n\times L}}\right)}{\beta_{1-\epsilon_2}\left(P_{X^n}P_{Y^{n\times L}|X^n},P_{X^n}Q_{Y^{n\times L}}\right)}, \tag{274}$$

where

$$\beta_{1-\epsilon_3}\left(P_{Y^{n\times L}},Q_{Y^{n\times L}}\right)\leq 1-\beta_{1-\epsilon_1}\left(P_{Y^{n\times L}|X^n=\mathbf{0}},Q_{Y^{n\times L}}\right), \tag{275}$$

$$\epsilon_3\leq\epsilon_2, \tag{276}$$

$$p_a\epsilon_2=\epsilon_{\mathrm{MD}}, \tag{277}$$

$$(1-p_a)\epsilon_1=\epsilon_{\mathrm{FA}}. \tag{278}$$

Proposition 19 presents a meta-converse bound for the single-user random access problem. However, evaluating this bound is numerically intractable because it involves an optimization over all possible input distributions. Next, we proceed to loosen Proposition 19 and obtain an easy-to-evaluate bound as provided in Theorem 10.

Following from the inequality that $M_m(n,\epsilon,P)\leq M_e(n+1,\epsilon,P)$ [11, Lemma 39], which relates the numbers of codewords under maximum power constraint and equal power constraint, the condition in (274) can be loosened to

$$\begin{aligned} & M \\ & \leq\sup_{\substack{P_{X^{n+1}}:\mathbf{x}\in\mathcal{F}^{n+1}\\ \epsilon_1,\epsilon_2,\epsilon_3\in[0,1]}}\frac{1-\beta_{1-\epsilon_1}\left(P_{Y^{(n+1)\times L}|X^{n+1}=\mathbf{0}},Q_{Y^{(n+1)\times L}}\right)}{\beta_{1-\epsilon_2}\left(P_{X^{n+1}}P_{Y^{(n+1)\times L}|X^{n+1}},P_{X^{n+1}}Q_{Y^{(n+1)\times L}}\right)} && (279) \\ & =\sup_{\epsilon_1,\epsilon_2,\epsilon_3\in[0,1]}\frac{\epsilon_1}{\beta_{1-\epsilon_2}\left(P_{Y^{(n+1)\times L}|X^{n+1}=\mathbf{x}_1},Q_{Y^{(n+1)\times L}}\right)}, && (280) \end{aligned}$$

where $\mathcal{F}^{n+1}=\left\{\mathbf{x}\in\mathbb{C}^{n+1}:\|\mathbf{x}\|_2^2=(n+1)P\right\}$, the auxiliary distribution is chosen as $Q_{Y^{(n+1)\times L}}=P_{Y^{(n+1)\times L}|X^{n+1}=\mathbf{0}}=\prod_{l=1}^{L}\mathcal{CN}(0,\mathbf{I}_{n+1})$ [44], and (280) follows from [11, Lemma 29] for any input $\mathbf{x}_1\in\mathcal{F}^{n+1}$. Meanwhile, under $Q_{Y^{(n+1)\times L}}=\prod_{l=1}^{L}\mathcal{CN}(0,\mathbf{I}_{n+1})$, the condition in (275) becomes

$$\beta_{1-\epsilon_3}\left(P_{Y^{(n+1)\times L}},Q_{Y^{(n+1)\times L}}\right)\leq\epsilon_1. \tag{281}$$

Since $\alpha \mapsto \beta_\alpha\left(P_{Y^{(n+1)\times L}}, Q_{Y^{(n+1)\times L}}\right)$ is monotonically non-decreasing, we can combine (281) and (276) as

$$\beta_{1-\epsilon_2}\left(P_{Y^{(n+1)\times L}}, Q_{Y^{(n+1)\times L}}\right) \leq \epsilon_1. \tag{282}$$

Following from [44, Lemma 6], we can obtain that

$$\begin{aligned}&\beta_{1-\epsilon_2}\left(P_{Y^{(n+1)\times L}}, Q_{Y^{(n+1)\times L}}\right)\\ &\leq M\beta_{1-\epsilon_2}\left(P_{Y^{(n+1)\times L}|X^{n+1}=\mathbf{x}_1}, Q_{Y^{(n+1)\times L}}\right).\end{aligned} \tag{283}$$

Together with (280) and (282), we observe that the condition in (282) is satisfied once (280) is satisfied.

In the remainder of this appendix, we proceed to compute $\beta_{1-\epsilon_2}\left(P_{Y^{(n+1)\times L}|X^{n+1}=\mathbf{x}_1}, Q_{Y^{(n+1)\times L}}\right)$. We have

$$\begin{aligned}&\ln\frac{dP_{Y^{(n+1)\times L}|X^{n+1}=\mathbf{x}_1}}{dQ_{Y^{(n+1)\times L}}} =\\ &-L\ln(1+(n+1)P)+\sum_{l=1}^{L}\mathbf{y}_l^H\left(\mathbf{I}_{n+1}-(\mathbf{I}_{n+1}+\mathbf{x}_1\mathbf{x}_1^H)^{-1}\right)\mathbf{y}_l.\end{aligned} \tag{284}$$

Under $P_{Y^{(n+1)\times L}|X^{n+1}=\mathbf{x}_1}$, (284) is distributed the same as

$$H = -L\ln(1+(n+1)P)+(n+1)P\frac{\chi^2(2L)}{2}. \tag{285}$$

Under $Q_{Y^{(n+1)\times L}}$, (284) is distributed the same as

$$G = -L\ln(1+(n+1)P)+\frac{(n+1)P}{1+(n+1)P}\frac{\chi^2(2L)}{2}. \tag{286}$$

Thus, we have

$$\begin{aligned}&\beta_{1-\epsilon_2}\left(P_{Y^{(n+1)\times L}|X^{n+1}=\mathbf{x}_1}, Q_{Y^{(n+1)\times L}}\right)\\ &= \mathbb{P}\left[G \geq \bar{r}\right] &(287)\\ &= \mathbb{P}\left[\chi^2(2L) \geq (1+(n+1)P)r\right], &(288)\end{aligned}$$

where $\bar{r}$ and r are chosen to satisfy

$$\mathbb{P}\left[H \leq \bar{r}\right] = \mathbb{P}\left[\chi^2(2L) \leq r\right] = \epsilon_2. \tag{289}$$

Thus, the single-user random access bound is obtained. It completes the proof of Theorem 10.

APPENDIX J
PROOF OF THEOREM 11

In this appendix, we prove Theorem 11 to establish a scaling law for the no-CSI case under the PUPE criterion and the assumption that all users are active. The achievability and converse scaling laws are established in Appendix J-A and Appendix J-B, respectively.

A. Achievability

Assume that the matrix $\mathbf{A} \in \mathbb{C}^{n\times KM}$ consists of codewords of all users, with columns drawn uniformly i.i.d. from the sphere of radius $\sqrt{nP}$. The power constraint in (3) is fulfilled in this case. Then, the PUPE can be upper-bounded as

$$\begin{aligned}P_e &\leq \epsilon_1 + \mathbb{P}\left[\frac{1}{K}\sum_{k\in\mathcal{K}}1\left[W_k \neq \hat{W}_k\right] \geq \epsilon_1\right] &(290)\\ &= \epsilon_1 + \mathbb{P}\left[\bigcup_{t=\lceil\epsilon_1 K\rceil}^{K}\mathcal{F}_t\right], &(291)\end{aligned}$$

where the positive constant $\epsilon_1 < \epsilon$ and $\mathcal{F}_t$ denotes the event that there are exactly t misdecoded users. Denote the set of codewords of K users as S_{all} of size KM and the set of the transmitted codewords of K users as $S_{\mathcal{K}}$ of size K. Let $\mathbf{\Gamma}_S = \mathrm{diag}\{\boldsymbol{\gamma}_S\} \in \{0,1\}^{KM\times KM}$, where $[\boldsymbol{\gamma}_S]_i = 1$ if the i-th codeword in the set S is transmitted by a user, and $[\boldsymbol{\gamma}_S]_i = 0$ otherwise. Similarly, let $\mathbf{\Gamma}'_S = \mathrm{diag}\left\{\boldsymbol{\gamma}'_S\right\}$, where $\left[\boldsymbol{\gamma}'_S\right]_i = 1$ if the i-th codeword in the set S is decoded for a user, and $\left[\boldsymbol{\gamma}'_S\right]_i = 0$ otherwise. Applying the decoding metric given in Appendix F, we can bound $\mathbb{P}[\mathcal{F}_t]$ as

$$\begin{aligned}&\mathbb{P}[\mathcal{F}_t]\\ &\leq \mathbb{P}\left[\bigcup_{\substack{S_1\subset S_{\mathcal{K}}\\ |S_1|=t}}\ \bigcup_{\substack{S_2\subset S_{all}\backslash S_{\mathcal{K}}\\ |S_2|=t}}\left\{g\left(\mathbf{\Gamma}'_{S_{\mathcal{K}}\backslash S_1\cup S_2}\right)\leq g(\mathbf{\Gamma}_{S_{\mathcal{K}}})\right\}\right] &(292)\\ &\leq \binom{K}{t}\binom{KM-K}{t}\mathbb{P}\left[g\left(\mathbf{\Gamma}'_{S_{\mathcal{K}}\backslash S_1\cup S_2}\right)\leq g(\mathbf{\Gamma}_{S_{\mathcal{K}}})\right] &(293)\\ &\leq \exp\left\{t\ln\left(\frac{e^2K^2M}{t^2}\right)\right\}\mathbb{P}\left[g\left(\mathbf{\Gamma}'_{S_{\mathcal{K}}\backslash S_1\cup S_2}\right)\leq g\left(\mathbf{\Gamma}_{S_{\mathcal{K}}}\right)\right], &(294)\end{aligned}$$

where (294) holds because $\binom{a}{b} \leq \left(\frac{ea}{b}\right)^b$ for $a \geq b > 0$. Denote $\mathbf{A}_{all} = \left\{\mathbf{A}, \mathbf{\Gamma}_{S_{\mathcal{K}}}, \mathbf{\Gamma}'_{S_{\mathcal{K}}\backslash S_1\cup S_2}\right\}$. We can bound the probability on the RHS of (294) as

$$\begin{aligned}&\mathbb{P}\left[g\left(\mathbf{\Gamma}'_{S_{\mathcal{K}}\backslash S_1\cup S_2}\right)\leq g(\mathbf{\Gamma}_{S_{\mathcal{K}}})\right]\\ &\leq \mathbb{E}_{\mathbf{A}_{all}}\left[\min_{u\geq 0}\mathbb{E}_{\mathbf{H},\mathbf{Z}}\left[\exp\left\{ug(\mathbf{\Gamma}_{S_{\mathcal{K}}})-ug\left(\mathbf{\Gamma}'_{S_{\mathcal{K}}\backslash S_1\cup S_2}\right)\right\}\middle|\mathbf{A}_{all}\right]\right] &(295)\\ &= \mathbb{E}_{\mathbf{A}}\left[\exp\left\{-L\left(-\frac{1}{2}\ln|\mathbf{F}|-\frac{1}{2}\ln\left|\mathbf{F}'\right|+\ln\left|\frac{1}{2}\mathbf{F}+\frac{1}{2}\mathbf{F}'\right|\right)\right\}\right], &(296)\end{aligned}$$

where $\mathbf{F} = \mathbf{I}_n + \mathbf{A}\mathbf{\Gamma}_{S_{\mathcal{K}}}\mathbf{A}^H$ and $\mathbf{F}' = \mathbf{I}_n + \mathbf{A}\mathbf{\Gamma}'_{S_{\mathcal{K}}\backslash S_1\cup S_2}\mathbf{A}^H$; (295) follows by applying Lemma 14 conditioned on $\mathbf{A}_{all}$; (296) follows from Lemma 15 by allowing $u = \frac{1}{2}$ and taking the expectation over $\mathbf{H}$ and $\mathbf{Z}$, and from the fact that the expectation is unchanged for different $\mathbf{\Gamma}_{S_{\mathcal{K}}}$ and $\mathbf{\Gamma}'_{S_{\mathcal{K}}\backslash S_1\cup S_2}$. Substituting (296) into (294), we obtain an upper bound on $\mathbb{P}[\mathcal{F}_t]$, which is a special case of the upper bound on $\mathbb{P}[\mathcal{F}_t]$ in Appendix F by allowing $\nu \to \infty$, $\omega = 1$, $r = 0$, and $u = \frac{1}{2}$.

Then, we aim to lower-bound $f\left(\mathbf{F},\mathbf{F}'\right) = -\frac{1}{2}\ln|\mathbf{F}| - \frac{1}{2}\ln\left|\mathbf{F}'\right| + \ln\left|\frac{1}{2}\mathbf{F}+\frac{1}{2}\mathbf{F}'\right|$ in (296). We have

$$\begin{aligned}&f\left(\mathbf{F},\mathbf{F}'\right)\\ &\geq \frac{1}{8}\mathrm{tr}\left(\left(\mathbf{F}-\mathbf{F}'\right)\left(\frac{\mathbf{F}+\mathbf{F}'}{2}\right)^{-1}\left(\mathbf{F}-\mathbf{F}'\right)\left(\frac{\mathbf{F}+\mathbf{F}'}{2}\right)^{-1}\right) &(297)\\ &\geq \frac{1}{8}\sigma^2_{min}\left(\left(\frac{\mathbf{F}+\mathbf{F}'}{2}\right)^{-1}\right)\mathrm{tr}\left(\left(\mathbf{F}-\mathbf{F}'\right)\left(\mathbf{F}-\mathbf{F}'\right)\right) &(298)\end{aligned}$$

$$= \frac{\left\| \mathbf{F} - \mathbf{F}^{'} \right\|_F^2}{8\sigma_{max}^2\left(\frac{\mathbf{F}+\mathbf{F}^{'}}{2}\right)}, \tag{299}$$

where $\sigma_{min}\left(\mathbf{A}\right)$ (resp., $\sigma_{max}\left(\mathbf{A}\right)$) denotes the minimum (resp., maximum) singular value of $\mathbf{A}$; (297) follows from Lemma 20 shown below; (298) follows by applying the inequality $\operatorname{tr}\left(\mathbf{AB}\right) \geq \sigma_{min}\left(\mathbf{A}\right)\operatorname{tr}\left(\mathbf{B}\right)$ twice for positive semi-definite matrices $\mathbf{A}$ and $\mathbf{B}$, and from the cyclic property of trace; (299) follows because the matrix $\frac{\mathbf{F}+\mathbf{F}^{'}}{2}$ is positive definite, $\sigma_{min}\left(\mathbf{A}^{-1}\right) = \frac{1}{\sigma_{max}(\mathbf{A})}$ for positive definite matrix $\mathbf{A}$, the matrix $\mathbf{F} - \mathbf{F}^{'}$ is Hermitian, and $\operatorname{tr}\left(\mathbf{B}^H\mathbf{B}\right) = \left\|\mathbf{B}\right\|_F^2$.

Lemma 20 (Proposition 1 in [52]): Let p_1 and p_2 be two multivariate Gaussian distributions with zero mean and positive definite covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, respectively. Then, the $\frac{1}{2}$-Rényi divergence between p_1 and p_2 is bounded as

$$D_{\frac{1}{2}}\left(p_1, p_2\right)$$
$$= -\frac{1}{2}\ln\left|\mathbf{\Sigma}_1\right| - \frac{1}{2}\ln\left|\mathbf{\Sigma}_2\right| + \ln\left|\frac{1}{2}\mathbf{\Sigma}_1 + \frac{1}{2}\mathbf{\Sigma}_2\right| \tag{300}$$
$$\geq \frac{1}{2}\operatorname{tr}\Big(\left(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\right)\left(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right)^{-1}\left(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\right)\left(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2\right)^{-1}\Big). \tag{301}$$

In the following, we follow similar ideas in [20] to lower-bound $\left\|\mathbf{F} - \mathbf{F}^{'}\right\|_F^2$ and upper-bound $\sigma_{max}^2\left(\frac{\mathbf{F}+\mathbf{F}^{'}}{2}\right)$, respectively. Following from the Restricted Isometry Property (RIP) results in [20, Theorems 2 and 5 and Appendix A], under the condition that

$$\frac{2n(n-1)}{M} \leq K \leq \min\left\{\frac{c_1 n(n-1)}{2\ln^2\left(\frac{eM}{2c_1}\right)}, \frac{\exp\left\{\frac{\sqrt{2n(n-1)}}{c_2}\right\}}{4M}\right\}, \tag{302}$$

with probability exceeding $1 - \exp\left\{-c_\delta\sqrt{n(n-1)}\right\}$ on a draw of concatenated codebooks of K users, we have

$$\left\|\mathbf{F} - \mathbf{F}^{'}\right\|_F^2 \geq (1-\delta)n^2P^2\epsilon_1\ K, \tag{303}$$

where $0 < c_1 < 1$, $c_2 > 0$, $c_\delta > 0$, and $0 < \delta < 1$ are universal constants.

Following from the large deviation result in [53, Th. 4.6.1], an upper bound on $\sigma_{max}\left(\frac{\mathbf{F}+\mathbf{F}^{'}}{2}\right)$ can be derived as

$$\sigma_{\max}\left(\frac{\mathbf{F}+\mathbf{F}^{'}}{2}\right)$$
$$\leq PC'\left(2\ln\left(\frac{eM}{2}\right) + \frac{\ln(2/\epsilon_2)}{\max\left\{K,n\right\}}\right)\max\{K,n\} + 1, \tag{304}$$

with probability at least $1 - \exp\left(-\beta\max\left(K,n\right)\right)$ for constants $C' > 0$ and $\beta > 0$. Following from [20, Appendix A], (304) is satisfied independent of transmitted codewords and decoded codewords with probability at least $1 - \epsilon_2$ where the positive constant ϵ_2 is less than ϵ.

Denote by $\mathcal{G}$ the event that (303) and (304) hold for all possible sets of transmitted codewords and decoded codewords. If the event $\mathcal{G}$ occurs, (299) can be lower-bounded as

$$-\frac{1}{2}\ln\left|\mathbf{F}\right| - \frac{1}{2}\ln\left|\mathbf{F}^{'}\right| + \ln\left|\frac{1}{2}\mathbf{F} + \frac{1}{2}\mathbf{F}^{'}\right| \geq \frac{m^*\epsilon_1 K}{4}. \tag{305}$$

where

$$m^* \geq \frac{1-\delta}{2\left(C'\left(2\ln\left(\frac{eM}{2}\right) + \frac{\ln(2/\epsilon_2)}{\max\{K,n\}}\right)\max\left\{\frac{K}{n}, 1\right\} + \frac{1}{nP}\right)^2}. \tag{306}$$

Therefore, we have

$$\mathbb{P}\left[\bigcup_{t=\lceil\epsilon_1 K\rceil}^{K}\mathcal{F}_t \cap \mathcal{G}\right]$$
$$\leq \sum_{t=\lceil\epsilon_1 K\rceil}^{K}\mathbb{P}\left[\mathcal{F}_t\middle|\,\mathcal{G}\right] \tag{307}$$
$$\leq \sum_{t=\lceil\epsilon_1 K\rceil}^{K}\exp\left\{t\ln\left(\frac{e^2K^2M}{t^2}\right)\right\} \cdot \mathbb{P}\left[g\left(\mathbf{\Gamma}_{S_\mathcal{K}\backslash S_1\cup S_2}^{'}\right) \leq g\left(\mathbf{\Gamma}_{S_\mathcal{K}}\right)\middle|\,\mathcal{G}\right] \tag{308}$$
$$\leq \sum_{t=\lceil\epsilon_1 K\rceil}^{K}\exp\left\{t\ln\left(\frac{e^2K^2M}{t^2}\right) - L\frac{m^*\epsilon_1 K}{4}\right\} \tag{309}$$
$$\leq (1-\epsilon_1)K\exp\left\{K\left(\ln\left(\frac{e^2M}{\epsilon_1^2}\right) - \frac{m^*L\epsilon_1}{4}\right)\right\}, \tag{310}$$

where (307) follows from the union bound and the inequality that $\mathbb{P}\left[\mathcal{G}_1 \cap \mathcal{G}_2\right] = \mathbb{P}\left[\mathcal{G}_2\right]\mathbb{P}\left[\mathcal{G}_1\middle|\,\mathcal{G}_2\right] \leq \mathbb{P}\left[\mathcal{G}_1\middle|\,\mathcal{G}_2\right]$ for events $\mathcal{G}_1$ and $\mathcal{G}_2$; (308) follows from (294); (309) follows from (296) and (305). In the case of

$$c = \frac{m^*L\epsilon_1}{4} - \ln\left(\frac{e^2M}{\epsilon_1^2}\right) > 0, \tag{311}$$

we have $\mathbb{P}\left[\bigcup_{t=\lceil\epsilon_1 K\rceil}^{K}\mathcal{F}_t \cap \mathcal{G}\right] \leq \exp\left\{o(K) - cK\right\}$. Together with (291), we have

$$P_e \leq \epsilon_1 + \mathbb{P}\left[\bigcup_{t=\lceil\epsilon_1 K\rceil}^{K}\mathcal{F}_t \cap \mathcal{G}\right] + \mathbb{P}\left[\mathcal{G}^c\right] \tag{312}$$
$$\leq \epsilon_1 + \exp\{o(K) - cK\} + \exp\{-c_\delta(n-1)\} + \epsilon_2, \tag{313}$$

where $\epsilon_1 + \epsilon_2 < \epsilon$, and $c, c_\delta > 0$ are universal constants. As $n, K \to \infty$, the error requirement $P_e \leq \epsilon$ in (4) is satisfied.

Combining (302), (306), (311), and (313), we can obtain the following scaling law. Supposing $M = \Theta(1)$ and $n, K, L \to \infty$, it is possible to serve $K = \Theta(n^2)$ users with $L = \Theta\left(n^2\right)$ BS antennas and power $P = \Theta\left(\frac{1}{n^2}\right)$, such that the PUPE constraint is satisfied. It was proved in [8, Appendix A-C] that, if one can achieve a certain PUPE for K users, it will also be possible to achieve the same PUPE for less than K users. As a result, under the PUPE criterion, we can reliably serve $K = \mathcal{O}(n^2)$ users when $L = \Theta\left(n^2\right)$, $P = \Theta\left(\frac{1}{n^2}\right)$, and $M = \Theta(1)$.

B. Converse

We assume $n, L \to \infty$ and ϵ and M are positive finite constants. Recall that $b_1 = J(1-\epsilon) - h_2(\epsilon)$. Following from Theorem 9, in the case of $K > n$, the minimum required energy-per-bit is larger than $\inf \frac{nP}{J}$, where the infimum is taken over all $P > 0$ satisfying

$$b_1 \leq \frac{nL}{K}\log_2(1+KP) - \frac{L}{K}\sum_{i=1}^{n}\bigg(\psi(K-i+1)\log_2 e + \log_2\bigg(P + \frac{1}{K-i+1}\bigg)\bigg), \quad (314)$$

where $\psi(\cdot)$ is Euler's digamma function. The RHS of (314) can be further bounded and we have

$$b_1 \leq \frac{L}{K}\sum_{i=1}^{n}\left(\log_2\left(\frac{1+KP}{1+(K-i+1)P}\right) + \frac{\log_2 e}{K-i+1}\right) \quad (315)$$

$$\leq \frac{nL}{K}\left(\log_2\left(\frac{1+KP}{1+(K-n+1)P}\right) + \frac{\log_2 e}{K-n+1}\right) \quad (316)$$

$$\leq \frac{nL\log_2 e}{K}\left(\frac{(n-1)P}{1+(K-n+1)P} + \frac{1}{K-n+1}\right), \quad (317)$$

where (315) follows by applying the inequality $\psi(x) \geq \ln x - \frac{1}{x}$ for $x > 0$ into (314); (317) follows because $\log_2(1+x) \leq x\log_2 e$ for $x \geq 0$. It is evident that the RHS of (317) is a monotonically decreasing function of K. In the case of $L = \Theta(n^2)$ and $P = \Theta\left(\frac{1}{n^2}\right)$, the condition in (317) is satisfied if and only if the number K of users satisfies that $n < K \leq \Theta(n^2)$.

In the case of $1 \leq K \leq n$, following from Theorem 9, the minimum required energy-per-bit is larger than $\inf \frac{nP}{J}$, where the infimum is taken over all $P > 0$ satisfying

$$b_1 \leq ML\log_2\left(1 + \frac{nP}{M}\right) - \frac{L}{K}\sum_{i=0}^{K-1}\left(\psi(n-i)\log_2 e + \log_2\left(P + \frac{1}{n-i}\right)\right). \quad (318)$$

Similar to (317), the RHS of (318) can be further upper-bounded and we have

$$b_1 \leq ML\log_2\left(1+\frac{nP}{M}\right) - L\log_2\left(1+(n-K+1)P\right) + \frac{L\log_2 e}{n-K+1}. \quad (319)$$

In the case of $K = 1$, (319) reduces to

$$b_1 \leq ML\log_2\left(1+\frac{nP}{M}\right) - L\log_2\left(1+nP\right) + \frac{L\log_2 e}{n}. \quad (320)$$

It is evident that $ML\log_2\left(1+\frac{nP}{M}\right) - L\log_2\left(1+nP\right) \geq 0$. Thus, when $L = \Theta(n^2)$ and $P = \Theta\left(\frac{1}{n^2}\right)$, the condition in (320) is satisfied for $K = 1$. Since the RHS of (319) is a monotonically increasing function of K, (319) is satisfied for any $1 \leq K \leq n$ in this case.

Taking both the cases of $1 \leq K \leq n$ and $K > n$ into consideration, we can draw the conclusion that when $M = \Theta(1)$, $n \to \infty$, $L = \Theta(n^2)$, and $P = \Theta\left(\frac{1}{n^2}\right)$, all users can be reliably served only if $K = \mathcal{O}(n^2)$.

Together with the achievability result in Appendix J-A, we conclude that assuming $M = \Theta(1)$ and $n \to \infty$, with $L = \Theta\left(n^2\right)$ BS antennas and the power $P = \Theta\left(\frac{1}{n^2}\right)$, one can satisfy the error requirement if and only if the number of users is $K = \mathcal{O}(n^2)$ when all users are active and there is no *a priori* CSI at the receiver.

Appendix K
Proof of Theorem 12

In this appendix, we prove Theorem 12 to establish an achievability bound for the pilot-assisted coded access scheme. Specifically, we consider a special case where all users are active, i.e. $K_a = K$. We use pilots drawn uniformly at random on an n_p-dimensional sphere of radius $\sqrt{n_p P_p}$. Thus, these pilots, denoted as $\mathbf{b}_1, \ldots, \mathbf{b}_K$ with length n_p, satisfy that $\|\mathbf{b}_k\|_2^2 = n_p P_p$ for $k \in \mathcal{K}$. Denote $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_K] \in \mathbb{C}^{n_p \times K}$. The received signal of the l-th antenna at the BS in the pilot transmission phase is given by

$$\mathbf{y}_{l,p} = \sum_{k\in\mathcal{K}} h_{k,l}\mathbf{b}_k + \mathbf{z}_{l,p} = \mathbf{B}\mathbf{h}_l + \mathbf{z}_{l,p} \in \mathbb{C}^{n_p}, \quad (321)$$

where $h_{k,l} \sim \mathcal{CN}(0,1)$ denotes the fading coefficient between the k-th user and the l-th antenna of the BS, which is i.i.d. across different users and different BS antennas; the vector $\mathbf{h}_l = [h_{1,l}, \ldots, h_{K,l}]^T \in \mathbb{C}^K$; the noise vector $\mathbf{z}_{l,p} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{n_p})$, which is i.i.d. across L BS antennas.

The BS performs channel estimation based on the MMSE criterion. The estimated channel for the l-th antenna of the BS is given by

$$\hat{\mathbf{h}}_l = \left(\mathbf{I}_K + \mathbf{B}^H\mathbf{B}\right)^{-1}\mathbf{B}^H\mathbf{y}_{l,p}, \quad (322)$$

where $\hat{\mathbf{h}}_l = \left[\hat{h}_{1,l}, \ldots, \hat{h}_{K,l}\right]^T \in \mathbb{C}^K$ is distributed as $\hat{\mathbf{h}}_l \sim \mathcal{CN}\left(\mathbf{0}, \hat{\mathbf{\Sigma}}\right)$ with $\hat{\mathbf{\Sigma}} = \mathbf{I}_K - \left(\mathbf{I}_K + \mathbf{B}^H\mathbf{B}\right)^{-1}$. From the orthogonality principle of the MMSE estimation, the channel estimation error $\tilde{\mathbf{h}}_l = \mathbf{h}_l - \hat{\mathbf{h}}_l = \left[\tilde{h}_{1,l}, \ldots, \tilde{h}_{K,l}\right]^T$ is independent of $\hat{\mathbf{h}}_l$, and is distributed as $\tilde{\mathbf{h}}_l \sim \mathcal{CN}\left(\mathbf{0}, \tilde{\mathbf{\Sigma}}\right)$ with $\tilde{\mathbf{\Sigma}} = \left(\mathbf{I}_K + \mathbf{B}^H\mathbf{B}\right)^{-1}$. For fixed pilot matrix $\mathbf{B}$, both the channel estimation $\hat{\mathbf{h}}_l$ and the channel estimation error $\tilde{\mathbf{h}}_l$ are i.i.d. across L BS antennas.

Similar to Appendix A, we use a random coding scheme in the data transmission phase by generating Gaussian codebooks of size M and length $n_d = n - n_p$ without power control, which for the k-th user is denoted as $\mathcal{C}_k = \{\mathbf{c}_{k,1}, \ldots, \mathbf{c}_{k,M}\}$ with $\mathbf{c}_{k,m} \overset{i.i.d.}{\sim} \mathcal{CN}\left(0, P'\mathbf{I}_{n_d}\right)$ for $k \in \mathcal{K}$ and $m \in [M]$. We choose $P' < \frac{nP - n_pP_p}{n_d}$ to ensure that we can control the maximum power constraint violation event. The matrix $\mathbf{A} \in \mathbb{C}^{n_d \times MK}$ denotes the concatenation of codebooks of the K users. Let the transmitted codeword of user k be $\mathbf{x}_{(k)} = \mathbf{c}_{(k)} 1\left\{\left\|\mathbf{c}_{(k)}\right\|_2^2 \leq nP - n_pP_p\right\}$, where $\mathbf{c}_{(k)} = \mathbf{c}_{k,W_k}$ with the message $W_k \in [M]$ chosen uniformly at random. The received signal of the l-th antenna in the data transmission phase is given by

$$\mathbf{y}_{l,d} = \sum_{k\in\mathcal{K}} h_{k,l}\mathbf{x}_{(k)} + \mathbf{z}_{l,d} \in \mathbb{C}^{n_d}, \quad (323)$$

where $h_{k,l}$ is defined as aforementioned, and the noise vector $\mathbf{z}_{l,d} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{n_d})$, which are i.i.d. across L antennas. Denote the signals received over L antennas as $\mathbf{Y}_d = [\mathbf{y}_{1,d}, \ldots, \mathbf{y}_{L,d}] \in \mathbb{C}^{n_d \times L}$.

The decoder has an incorrect estimate of the channel, but uses the estimate as if it were perfect. Based on the notation introduced in Appendix A, the decoding metric in this case is given by

$$g\left(\mathbf{Y}_d, \hat{\mathbf{c}}_{[\mathcal{K}]}\right) = \sum_{l=1}^{L} \left\| \mathbf{y}_{l,d} - \sum_{k \in \mathcal{K}} \hat{h}_{k,l} \hat{\mathbf{c}}_{(k)} \right\|_2^2 . \tag{324}$$

The decoder outputs

$$\hat{\mathbf{c}}_{[\mathcal{K}]} = \min_{\left(\hat{\mathbf{c}}_{(k)} \in \mathcal{C}_k\right)_{k \in \mathcal{K}}} g\left(\mathbf{Y}_d, \hat{\mathbf{c}}_{[\mathcal{K}]}\right), \tag{325}$$

$$\hat{W}_k = f_{\mathrm{en},k}^{-1}\left(\hat{\mathbf{c}}_{(k)}\right), \quad k \in \mathcal{K}. \tag{326}$$

We can upper-bound the PUPE as in (105) by allowing $K_a = K$, where the total variation distance p_0 for the pilot-assisted scheme is given by

$$p_0 = K\, \mathbb{P}\left[\left\|\mathbf{c}_{(k)}\right\|_2^2 > nP - n_p P_p\right] \tag{327}$$

$$= K\left(1 - \frac{\gamma\left(n_d, \left(nP - n_p P_p\right)/P'\right)}{\Gamma\left(n_d\right)}\right). \tag{328}$$

In the remainder of this appendix, we upper-bound $\mathbb{P}\left[\mathcal{F}_t\right]$ in (105) relying on the standard bounding technique proposed by Fano [29]. Compared with the case of CSIR, upper-bounding $\mathbb{P}\left[\mathcal{F}_t\right]$ is more involved for the pilot-assisted coded access scheme due to the channel estimation error. Hence, we simplify the "good region" introduced in Section III-A by allowing $w = 0$ and obtain

$$\mathcal{R} = \left\{\mathbf{Y}_d : g\left(\mathbf{Y}_d, \mathbf{c}_{[\mathcal{K}_a]}\right) \le n_d L \nu\right\}. \tag{329}$$

Define the event $\mathcal{G}_\nu = \left\{\mathbf{Y}_d \in \mathcal{R}\right\}$. By replacing $\mathcal{G}_{\omega,\nu}$ with $\mathcal{G}_\nu$ and allowing $S_1 = S_2$, the upper bound on $\mathbb{P}\left[\mathcal{F}_t\right]$ in (109) becomes

$$\begin{aligned} &\mathbb{P}\left[\mathcal{F}_t\right] \\ &\le \min_{\nu \ge 0} \Bigg\{ \mathbb{P}\Bigg[\bigcup_{S_1} \bigcup_{\mathbf{c}'_{[S_1]}} \left\{ g\left(\mathbf{Y}_d, \mathbf{c}_{[\mathcal{K}_a \setminus S_1]} \cup \mathbf{c}'_{[S_1]}\right) \le g\left(\mathbf{Y}_d, \mathbf{c}_{[\mathcal{K}_a]}\right) \right\} \\ &\qquad\qquad \bigcap \mathcal{G}_\nu \Bigg] + \mathbb{P}\left[\mathcal{G}_\nu^c\right] \Bigg\} \qquad (330) \\ &= \min_{\nu \ge 0} \left\{ \mathbb{P}\left[\mathcal{G}_e \cap \mathcal{G}_\nu\right] + \mathbb{P}\left[\mathcal{G}_\nu^c\right] \right\}. \qquad (331) \end{aligned}$$

In the following, we upper-bound $\mathbb{P}\left[\mathcal{G}_e \cap \mathcal{G}_\nu\right]$ and $\mathbb{P}\left[\mathcal{G}_\nu^c\right]$, respectively.

Define $\tilde{\mathbf{A}}_{S_1}$, $\tilde{\mathbf{A}}'_{S_1}$, and $\tilde{\mathbf{A}}_{\mathcal{K}}$ as in Theorem 12. Denote $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_L] \in \mathbb{C}^{K \times L}$, $\hat{\mathbf{H}} = \left[\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_L\right] \in \mathbb{C}^{K \times L}$, $\tilde{\mathbf{H}} = \left[\tilde{\mathbf{h}}_1, \ldots, \tilde{\mathbf{h}}_L\right] \in \mathbb{C}^{K \times L}$, and $\mathbf{Z}_d = [\mathbf{z}_{1,d}, \ldots, \mathbf{z}_{L,d}] \in \mathbb{C}^{n_d \times L}$. Denote $\mathbf{A}_{all} = \left\{\tilde{\mathbf{A}}_{\mathcal{K}}, \tilde{\mathbf{A}}_{S_1}, \tilde{\mathbf{A}}'_{S_1}\right\}$. Using the above notation, we can obtain

$$\begin{aligned} &\mathbb{P}\left[\mathcal{G}_e \cap \mathcal{G}_\nu\right] \\ &\le \sum_{S_1} \sum_{\mathbf{c}'_{[S_1]}} \mathbb{E}_{\mathbf{A}_{all}, \mathbf{B}} \Bigg[\min_{u \ge 0, r \ge 0} \exp\left\{r n_d L \nu\right\} \\ &\quad \cdot \mathbb{E}_{\mathbf{Z}_d, \tilde{\mathbf{H}}, \hat{\mathbf{H}}} \Bigg[\exp\Bigg\{ -u \left\| \mathbf{Z}_d + \tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{H}} + \left(\tilde{\mathbf{A}}_{S_1} - \tilde{\mathbf{A}}'_{S_1}\right) \hat{\mathbf{H}} \right\|_F^2 \\ &\qquad + (u - r) \left\| \mathbf{Z}_d + \tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{H}} \right\|_F^2 \Bigg\} \Bigg| \mathbf{A}_{all}, \mathbf{B} \Bigg] \Bigg] \qquad (332) \\ &\le \binom{K}{t} M^t \mathbb{E}_{\mathbf{A}_{all}, \mathbf{B}} \left[\min_{\substack{u \ge 0, r \ge 0 \\ \lambda_{\min}(\mathbf{D}) > 0}} \exp\left\{ r n_d L \nu - \frac{L}{2} \ln |\mathbf{D}| \right\} \right], \qquad (333) \end{aligned}$$

where (332) follows by applying the Chernoff bound in Lemma 14 to the probability $\mathbb{P}\left[\mathcal{G}_e \cap \mathcal{G}_\nu\right]$ conditioned on $\mathbf{A}_{all}$ and $\mathbf{B}$; (333) follows from Lemma 15 by taking the expectation over $\tilde{\mathbf{H}}, \hat{\mathbf{H}}$, and $\mathbf{Z}_d$ provided that the eigenvalues of $\mathbf{D}$ are positive, with the expression of $\mathbf{D}$ given in (93). The term on the RHS of (333) is denoted as $q_{1,t}(\nu)$ as presented in (91).

In the remainder of this appendix, we upper-bound $\mathbb{P}\left[\mathcal{G}_\nu^c\right]$ in two ways. Let $\lambda_1, \ldots, \lambda_{n_d}$ denote the eigenvalues of $\tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{A}}_{\mathcal{K}}^H$ in decreasing order with rank $n^* = \min\{K, n_d\}$. First, applying the Chernoff bound and Lemma 15, we can bound $\mathbb{P}\left[\mathcal{G}_\nu^c\right]$ as

$$\begin{aligned} \mathbb{P}\left[\mathcal{G}_\nu^c\right] \le \mathbb{E}_{\tilde{\mathbf{A}}_{\mathcal{K}}, \mathbf{B}} \Bigg[&\min_{0 \le \delta < 1/(1+\lambda_1)} \exp\left\{-\delta n_d L \nu\right\} \\ &\cdot \left| (1 - \delta) \mathbf{I}_{n_d} - \delta \tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{\Sigma}} \tilde{\mathbf{A}}_{\mathcal{K}}^H \right|^{-L} \Bigg]. \end{aligned} \tag{334}$$

Define the event $\mathcal{G}_\eta = \left\{\frac{\chi^2(2n_d L)}{2} \le n_d L \eta\right\}$ for $\eta \ge 0$. Alternatively, we have

$$\begin{aligned} &\mathbb{P}\left[\mathcal{G}_\nu^c\right] \\ &= \mathbb{E}_{\tilde{\mathbf{A}}_{\mathcal{K}}, \mathbf{B}} \left[\mathbb{P}\left[\sum_{l \in [L]} \left\| \mathbf{z}_{l,d} + \tilde{\mathbf{A}}_{\mathcal{K}} \tilde{\mathbf{h}}_l \right\|_2^2 > n_d L \nu \,\middle|\, \tilde{\mathbf{A}}_{\mathcal{K}}, \mathbf{B} \right] \right] \qquad (335) \\ &\le \min_{0 \le \eta \le \nu} \Bigg\{ \mathbb{E}_{\tilde{\mathbf{A}}_{\mathcal{K}}, \mathbf{B}} \Bigg[\mathbb{P}\Bigg[\left\{ \frac{\chi^2(2n_d L)}{2} + \sum_{i=1}^{n^*} \frac{\lambda_i \chi_i^2(2L)}{2} > n_d L \nu \right\} \\ &\qquad \cap \mathcal{G}_\eta \Bigg| \tilde{\mathbf{A}}_{\mathcal{K}}, \mathbf{B} \Bigg] \Bigg] + \mathbb{P}\left[\mathcal{G}_\eta^c\right] \Bigg\} \qquad (336) \\ &\le \min_{0 \le \eta \le \nu} \Bigg\{ \mathbb{E}_{\tilde{\mathbf{A}}_{\mathcal{K}}, \mathbf{B}} \left[\mathbb{P}\left[\sum_{i=1}^{n^*} \frac{\lambda_i \chi_i^2(2L)}{2} > n_d L (\nu - \eta) \,\middle|\, \tilde{\mathbf{A}}_{\mathcal{K}}, \mathbf{B} \right] \right] \\ &\qquad + 1 - \frac{\gamma\left(n_d L, n_d L \eta\right)}{\Gamma\left(n_d L\right)} \Bigg\}, \qquad (337) \end{aligned}$$

where the conditional probability on the RHS of (337) can be further upper-bounded as

$$\begin{aligned} &\mathbb{P}\left[\sum_{i=1}^{n^*} \lambda_i \frac{\chi_i^2(2L)}{2} > n_d L (\nu - \eta) \,\middle|\, \tilde{\mathbf{A}}_{\mathcal{K}}, \mathbf{B} \right] \\ &\le \mathbb{P}\left[\lambda_1 \frac{\chi^2(2Ln^*)}{2} > n_d L (\nu - \eta) \,\middle|\, \tilde{\mathbf{A}}_{\mathcal{K}}, \mathbf{B} \right] \qquad (338) \end{aligned}$$

$$= 1 - \frac{\gamma\left(Ln^*, \frac{n_d L(\nu-\eta)}{\lambda_1}\right)}{\Gamma\left(Ln^*\right)}. \quad (339)$$

Taking the minimum value of (334) and (337), we can obtain the ultimate upper bound on $\mathbb{P}\left[\mathcal{G}_\nu^c\right]$, which is denoted as $q_{2,t}\left(\nu\right)$ as in (92). This concludes the proof of Theorem 12.

REFERENCES

[1] H. Liao, "A coding theorem for multiple access communications," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Pacific Grove, CA, USA, Jan. 1972, pp. 1–5.

[2] R. Ahlswede, "Multi-way communication channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Tsahkadsor, Armenia, Sep. 1971, pp. 23–52.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Hoboken, NJ, USA: Wiley, 2006.

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

[5] F. Wei, Y. Wu, W. Chen, W. Yang, and G. Caire, "On the fundamental limits of MIMO massive multiple access channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Shanghai, China, May 2019, pp. 1–6.

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

[7] 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.

[8] 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.

[9] J. Gao, Y. Wu, and W. Zhang, "Energy-efficiency of massive random access with individual codebook," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.

[10] Y. Wu, X. Gao, S. Zhou, W. Yang, Y. Polyanskiy, and G. Caire, "Massive access for future wireless communication systems," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 148–156, Aug. 2020.

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

[12] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.

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

[14] R. C. Yavas, V. Kostina, and M. Effros, "Gaussian multiple and random access channels: Finite-blocklength analysis," *IEEE Trans. Inf. Theory*, vol. 67, no. 11, pp. 6983–7009, Nov. 2021.

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

[16] 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, Aug. 2020.

[17] O. L. A. López, G. Brante, R. D. Souza, M. Juntti, and M. Latva-Aho, "Coordinated pilot transmissions for detecting the signal sparsity level in a massive IoT network under Rayleigh fading," May 2022, *arXiv:2205.00406*.

[18] A. Lancho, J. Östman, and G. Durisi, "On joint detection and decoding in short-packet communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Madrid, Spain, Dec. 2021, pp. 1–6.

[19] K.-H. Ngo, A. Lancho, G. Durisi, and A. G. I. Amat, "Unsourced multiple access with random user activity," Feb. 2022, *arXiv:2202.06365*.

[20] 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.

[21] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[22] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 117–132, Feb. 2013.

[23] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.

[24] X. Xie et al., "Massive unsourced random access: Exploiting angular domain sparsity," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2480–2498, Apr. 2022.

[25] J. Östman, G. Durisi, E. G. Ström, M. C. Coskun, and G. Liva, "Short packets over block-memoryless fading channels: Pilot-assisted or noncoherent transmission?" *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1521–1536, Feb. 2019.

[26] J. Östman, A. Lancho, G. Durisi, and L. Sanguinetti, "URLLC with massive MIMO: Analysis and design at finite blocklength," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6387–6401, Oct. 2021.

[27] H. Weingarten, Y. Steinberg, and S. S. Shitz, "Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1665–1686, Aug. 2004.

[28] A. T. Asyhari and A. G. I. Fàbregas, "Nearest neighbor decoding in MIMO block-fading channels with imperfect CSIR," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1483–1517, Mar. 2012.

[29] R. M. Fano, *Transmission of Information*. Cambridge, MA, USA: MIT Press, 1961.

[30] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 1, pp. 3–18, Jan. 1965.

[31] I. Sason and S. S. Shitz, "Performance analysis of linear codes under maximum-likelihood decoding: A tutorial," *Found. Trends Commun. Inf. Theory*, vol. 3, nos. 1–2, pp. 1–222, 2006.

[32] W. Yang, G. Durisi, and Y. Polyanskiy, "Minimum energy to send k bits over multiple-antenna fading channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6831–6853, Dec. 2016.

[33] J. Ravi and T. Koch, "Scaling laws for Gaussian random many-access channels," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2429–2459, Apr. 2022.

[34] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, Jun. 2018.

[35] G. Sun et al., "Massive grant-free OFDMA with timing and frequency offsets," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3365–3380, May 2022.

[36] T. Li et al., "Joint device detection, channel estimation, and data decoding with collision resolution for MIMO massive unsourced random access," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 5, pp. 1535–1555, May 2022.

[37] H. Herzberg and G. Poltyrev, "Techniques of bounding the probability of decoding error for block coded modulation structures," *IEEE Trans. Inf. Theory*, vol. 40, no. 3, pp. 903–911, May 1994.

[38] E. R. Berlekamp, "The technology of error correction codes," *Proc. IEEE*, vol. 68, no. 5, pp. 564–593, May 1980.

[39] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York, NY, USA: Springer, 1994.

[40] I. Bettesh and S. S. Shitz, "Outages, expected rates and delays in multiple-users fading channels," in *Proc. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, USA, Mar. 2000, pp. WA4.7–WA4.15.

[41] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static SIMO fading channels at finite blocklength," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, İstanbul, Turkey, Jul. 2013, pp. 1531–1535.

[42] G. Reeves and M. C. Gastpar, "Approximate sparsity pattern recovery: Information-theoretic lower bounds," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3451–3465, Jun. 2013.

[43] J. Östman, W. Yang, G. Durisi, and T. Koch, "Diversity versus multiplexing at finite blocklength," in *Proc. 11th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Barcelona, Spain, Aug. 2014, pp. 702–706.

[44] W. Yang, A. Collins, G. Durisi, Y. Polyanskiy, and H. V. Poor, "Beta–beta bounds: Finite-blocklength analog of the golden formula," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6236–6256, Sep. 2018.

[45] A. M. Mathai and B. P. Serge, *Quadratic Forms in Random Variables: Theory and Applications*. New York, NY, USA: Marcel Dekker, 1992.

[46] A. A. Mohsenipour, "On the distribution of quadratic expressions in various types of random vectors," Ph.D. dissertation, UWO, London, ON, Canada, Nov. 2012.

[47] M. Okamoto, "An inequality for the weighted sum of χ^2 variates," *Bull. Math. Stat.*, vol. 9, nos. 2–3, pp. 69–70, Oct. 1960.

[48] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[49] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, Dept. Math., MIT, Cambridge, MA, USA, May 1989.

[50] L. Birgé, "An alternative point of view on Lepski's method," *Lect. Notes-Monograph Ser.*, vol. 36, pp. 113–133, Jan. 2001.

[51] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[52] S. Khanna and C. R. Murthy, "On the support recovery of jointly sparse Gaussian sources using sparse Bayesian learning," Mar. 2017, *arXiv:1703.04930*.

[53] R. Vershynin, *High-Dimensional Probability: An Introduction With Applications in Data Science* (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge, U.K.: Cambridge Univ. Press, 2018.

Junyuan Gao received the B.S. degree in communication engineering from Chongqing University, Chongqing, China, in 2018. She is currently pursuing the Ph.D. degree in electronic engineering with Shanghai Jiao Tong University, Shanghai, China. Her research interests include massive random access, massive MIMO, and information theory.

Yongpeng Wu (Senior Member, IEEE) received the B.S. degree in telecommunication engineering from Wuhan University, Wuhan, China, in July 2007, and the Ph.D. degree in communication and signal processing from the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in November 2013.

He is currently a Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. Previously, he was a Senior Research Fellow with the Institute for Communications Engineering, Technical University of Munich, Germany; and the Humboldt Research Fellow and a Senior Research Fellow with the Institute for Digital Communications, University Erlangen-Nürnberg, Germany. During his doctoral studies, he conducted cooperative research at the Department of Electrical Engineering, Missouri University of Science and Technology, USA. His research interests include massive MIMO/MIMO systems, massive machine type communication, physical layer security, and signal processing for wireless communications.

Dr. Wu was awarded the IEEE Student Travel Grants for IEEE International Conference on Communications (ICC) 2010, the Alexander von Humboldt Fellowship in 2014, the Travel Grants for IEEE Communication Theory Workshop 2016, the Excellent Doctoral Thesis Awards of China Communications Society 2016, the Exemplary Editor Award of IEEE COMMUNICATION LETTERS 2017, the Young Elite Scientist Sponsorship Program by CAST 2017, and the Excellent Youth Science Fund Project of the National Natural Science Foundation of China 2021. He has been the Symposium Chair of various conferences, including Globecom, ICC, VTC, and PIMRC. He was the Lead Guest Editor of IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, and IEEE WIRELESS COMMUNICATIONS. He is currently the Editor of IEEE WIRELESS COMMUNICATIONS. He was the Editor of IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE COMMUNICATIONS LETTERS.

Shuo Shao (Member, IEEE) received the B.S. degree in information science from Southeast University, China, in 2011, the M.A.Sc. degree in electrical and computer engineering from McMaster University, Canada, in 2013, and the Ph.D. degree from Texas A&M University, USA, in 2017. He is an Associate Professor with the School of Electronics Information and Electrical Engineering, Shanghai Jiao Tong University, China. His research interests are in network information theory, algebraic code, machine learning, and semantic communications.

Wei Yang (Member, IEEE) received the B.E. degree in communication engineering and the M.E. degree in communication and information systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 2008 and 2011, respectively, and the Ph.D. degree in electrical engineering from the Chalmers University of Technology, Gothenburg, Sweden, in 2015. From 2015 to 2017, he was a Post-Doctoral Research Associate at Princeton University, Princeton, NJ, USA. In September 2017, he joined Qualcomm Research, San Diego, CA, USA. He was a recipient of the Student Paper Award at the 2012 IEEE International Symposium on Information Theory (ISIT), Cambridge, MA, USA.

H. Vincent Poor (Life Fellow, IEEE) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 to 1990, he was on the Faculty of the University of Illinois at Urbana–Champaign. Since 1990, he has been on the Faculty at Princeton, where he is currently the Michael Henry Strater University Professor. From 2006 to 2016, he served as the Dean of Princeton's School of Engineering and Applied Science. He has also held a visiting appointments at several other universities, including most recently at Berkeley and Cambridge. His research interests are in the areas of information theory, machine learning, and network science, and their applications in wireless networks, energy systems, and related fields. Among his publications in these areas is the recent book *Machine Learning and Wireless Communications* (Cambridge University Press, 2022). He is a member of the National Academy of Engineering and the National Academy of Sciences; and a Foreign Member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. He received the IEEE Alexander Graham Bell Medal in 2017.