This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2026.3701621

Fundamental Limits of Noncoherent Massive Random Access Networks

Grace Villacrés, *Member, IEEE,* Tobias Koch, *Senior Member, IEEE,* and Gonzalo Vazquez-Vilar, *Member, IEEE*

***Abstract*—This paper studies the capacity of massive random-access networks modeled as a multiple-input multiple-output fading channel with infinitely many interfering users. The network is assumed to operate in a noncoherent regime, where transmitters and receivers know the fading statistics but not their realizations. Users access the network via random activation with a given probability. To characterize the symmetric sum rate, a random-coding argument is invoked together with the assumption that users and interferers draw their codebooks according to the same distribution. For this channel model, rigorous upper and lower bounds on the network capacity are derived. The behavior of these bounds depends critically on the spatial decay of the large-scale fading statistics from interfering users. In particular, if the large-scale fading coefficients of the interferers (ordered according to their distance to the receiver) decay exponentially or more slowly, then the capacity is bounded in the transmit power. This occurs because the aggregate interference scales with the transmit power, and reveals an inherent saturation effect in interference-limited networks. Moreover, in this regime, random user activity cannot fundamentally eliminate the resulting capacity ceiling. In contrast, if the large-scale fading coefficients of the interferers decay faster than double-exponentially, then the capacity becomes unbounded in the transmit power. Note that proving an unbounded capacity is nontrivial even if the number of interfering users is finite, since the condition that the users' codebooks follow the same distribution prevents interference-avoiding strategies such as time-, frequency-, or code-division multiple access, and cooperation among users associated with different access nodes is not allowed. An unbounded coding rate is achieved by using bursty signaling together with treating interference as noise.**

***Index Terms*—Channel capacity, fading channel, massive multiple access, noncoherent, random access, wireless networks.**

G. Villacrés has been partially supported by the Comunidad de Madrid within the 2023-2026 agreement with Universidad Rey Juan Carlos for the granting of direct subsidies for the promotion, encouragement of research and technology transfer, line of Action A Emerging Doctors, under Project OrdeNGN (Ref. F1177). T. Koch has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 714161), from the Spanish Ministerio de Ciencia, Innovación y Universidades under Grants PID2024-159557OB-C21 (MICIU/AEI/10.13039/501100011033 and ERDF/UE), PID2020-116683GB-C21 (AEI/10.13039/501100011033), TEC2016-78434-C3-3-R (AEI/FEDER, EU), and RYC-2014-16332, and from the Comunidad de Madrid under Grant IDEA-CM (TEC-2024/COM-89). An earlier version of this paper was presented in part at the 2016 IEEE International Symposium on Information Theory, Barcelona, Spain, July 2016 [DOI: 10.1109/ISIT.2016.7541766] and in part at the 2020 IEEE International Symposium on Information Theory, Los Angeles, CA, USA, June 2020 [DOI: 10.1109/ISIT44484.2020.9174034].

G. Villacrés is with the Department of Signal Theory and Communications, Universidad Rey Juan Carlos, 28942 Fuenlabrada, Spain (email: grace.villacres@urjc.es).

T. Koch and G. Vazquez-Vilar are with the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Spain, and also with the Gregorio Marañón Health Research Institute, 28007 Madrid, Spain (emails: tkoch@ing.uc3m.es, gonzalo.vazquez@uc3m.es).

I. Introduction

THE rapid evolution of modern and future wireless communication technologies is driven by the growing demand for enhanced data rates, extended coverage, and improved scalability. Emerging applications, such as high-speed mobile broadband, the *Internet of Things (IoT)*, and *machine-type communications (MTC)*, require networks to support both human users and a vast number of interconnected devices. As a result, technological advancements are not only improving user experiences but also addressing the increasing complexity of managing heterogeneous devices and services [1]–[3].

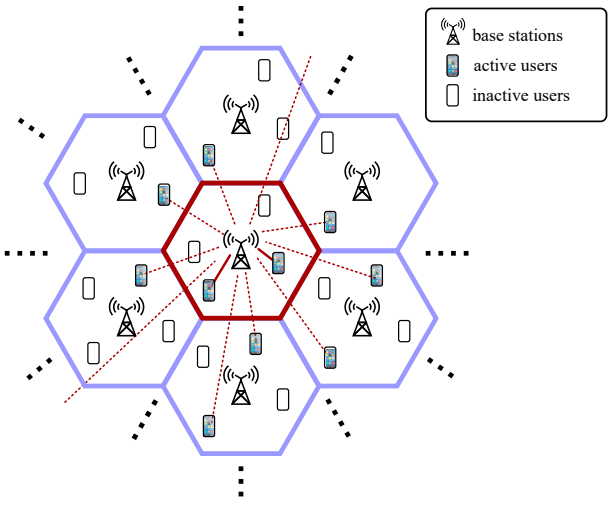
A key factor in this transformation is the exponential growth of IoT devices. According to the IoT Analytics Report [4], the number of connected IoT devices is expected to reach 40 billion by 2030. This growth, coupled with the rising number of network users, is expected to significantly increase data traffic and place higher demands on data rates. To meet these demands, future wireless networks will rely heavily on radio access network densification. However, this densification also introduces new challenges, particularly in handling traffic bursts. These bursts, often caused by sporadic user activity or the unpredictable nature of IoT and MTC devices, can overload network resources and degrade performance if not properly managed [5].

In this context, wireless communication technologies are evolving along two main paradigms:

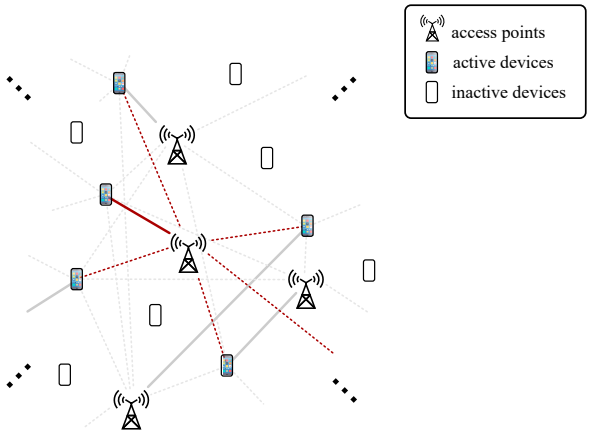
Enhanced mobile broadband (eMBB) supports applications that require high data rates and depend on dense network deployments and advanced interference management. These systems are typically deployed using a cell-based network architecture (see Fig. 1a) and use, for example, femtocells and macrocells to share network resources effectively. However, this densification, while enhancing capacity, also increases the risk of inter-cell interference as the number of communicating users grows. Since interference is one of the main limiting factors for achieving higher data rates, its management has been the subject of several studies; see, e.g., [6] and references therein.

Massive machine-type communications (mMTC) in turn, is tailored for massive connectivity, focusing on supporting a large number of devices with sporadic activity, as seen in IoT systems. This approach typically employs a cell-free network architecture, allowing devices to communicate directly with multiple access points (see Fig. 1b). Access to the network is often based on grant-free random access, enabling devices to transmit data without waiting for scheduling or resource allo-

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2026.3701621



(a) eMBB: cell-based network.



(b) mMTC: cell-free network.

Fig. 1. Comparison of eMBB and mMTC network architectures.

cation [2], [7], [8]. In scenarios involving hundreds of millions of potentially connected devices, Polyansky [9] introduced the concept of *unsourced random access (URA)*. In URA, all users share a common codebook, and the active users select codewords from it to transmit their messages. The receiver focuses on decoding the content of the messages without needing to identify the individual users, thereby reducing signaling overhead. In this scenario, inter-user interference also becomes a critical challenge due to the large number of potentially transmitting devices.

A. Research Scope and Problem Statement

There are several lines of research that have studied different aspects of wireless networks. For example, the impact of interference in wireless networks was investigated under the framework of the *interference channel*; see, e.g., [10]–[12]. While, traditionally, it was assumed that in these networks interference is always present, more recent works have also considered intermittent/bursty interference due to intermittent user activity (as in IoT networks) or opportunistic frequency reuse among cells [13]; see, e.g., [14]–[16]. More recently, the problem of a massive number of users accessing sporadically a wireless network was studied under the frameworks of the *many-access channel* [17] or *massive random access* [9]. In the many-access channel, the number of users accessing the network grows to infinity with the blocklength to model a massive number of users. This problem has attracted significant attention in the information theory literature, with contributions ranging from achievability and converse bounds for Gaussian channels [18]–[22] and fading channels [23]–[25] to the development of practical coding schemes for this setting [21], [22], [26]–[32]. When the number of users grows without bound alongside the blocklength n, the number of bits each user can reliably transmit scales sublinearly in n [17]. In fact, many of the aforementioned works assume that each user transmits a fixed number of bits. This implies that the number of bits that each user can transmit per channel use vanishes as the blocklength tends to infinity.

A different approach to studying the fundamental limits of wireless networks was followed by Lozano, Heath, and Andrews [33], who modeled the wireless network as a multiple-input multiple-output (MIMO) block-fading channel and then studied its capacity. They showed that the maximal coding rate achievable with channel inputs of the form $\sqrt{\mathsf{P}}X$ (where X is normalized, so that P represents the transmit power, and the distribution of X does not depend on P) is bounded in the transmit power P. This suggests that there is a saturation regime where the channel capacity hits a ceiling that is independent of the transmit power. However, while Gaussian inputs can be written in the above form, limiting the analysis to such inputs may be overly restrictive and could potentially be the reason for a bounded coding rate. In fact, it was shown in [34, Th. 4.3] that, for noncoherent point-to-point fading channels, the rate achievable by such inputs is bounded in the transmit power. For a more detailed discussion, see Section III-C.

In this paper, we also model the overall wireless network as a MIMO fading channel. To enable analytical tractability and gain insight into fundamental performance limits, we adopt a number of idealized modeling assumptions. In particular, we assume that there is an infinite number of interferers to model a massive number of users. More precisely, we consider a network where n_T users are associated with an access node (AN) equipped with n_R antennas, which may represent a base station in a cellular network or access points in a cell-free architecture. Inter-AN interference arises from users associated with neighboring ANs. Users associated with the same access node (AN) are allowed to cooperate. This allows us to model the users as a virtual multi-antenna transmitter and enables analytically-tractable converse bounds. In practical systems, user cooperation may incur significant overhead, e.g., for user-activity detection and scheduling [35]–[40], which we do not account for. Nevertheless, since user cooperation can only increase the achievable rates, the derived converse bounds remain valid even when such cooperation is only partial or absent in practice. In contrast to the vast literature on massive random access, we assume that there is an infinite number of interferers (rather than a number of interferers that grows with

the blocklength), but we account for path loss to ensure that the interference from distant users diminishes, thereby permitting the transmission at a positive coding rate. Furthermore, we consider intermittent user activity, i.e., users access the network at random with a given activity probability. We invoke a random-coding argument and assume that the interferers draw their codebooks according to the same distribution as the n_T users associated with the intended AN. This can be viewed as a generalization of the commonly-adapted Gaussian codebook model and aligns with large-system analyses. Apart from the power constraint, we do not impose any further restrictions on the channel inputs. In particular, we do not assume that the inputs are of the form $\sqrt{\mathsf{P}}X$ for a random variable X independent of P. We believe our model captures the essence of both eMBB and mMTC scenarios. Specifically, the eMBB scenario corresponds to the above model for general n_T and n_R, allowing multiple simultaneously active users and multiple receive antennas at the base stations (see Fig. 1a). In the mMTC regime, the massive number of devices is modeled through an infinite number of interferers. Sporadic user activity across the network ensures that the number of simultaneously active users per access point and per resource block is typically small (see Fig. 1b). In contrast to an earlier version of this work [41], we treat interference at a per-user level rather than a per-cell level, thus capturing the impact of user activity more accurately. This model better approximates new-generation wireless communication systems, which are expected to support emerging IoT and *machine-to-machine (M2M)* communication services, typically characterized, as mentioned before, by a massive number of IoT devices with different types of data traffic and sporadic, bursty activity of the individual users.

B. Contributions and Organization

We analyze the behavior of channel capacity in noncoherent wireless networks at high transmit power, focusing on the scenario where users and receivers only know the statistics of the fading coefficients, but not their realizations—an assumption that we believe is realistic when the number of interferers is large. We establish explicit conditions under which the capacity remains bounded in the transmit power, revealing a severe power inefficiency in these networks. Our key findings can be summarized as follows:

1) *Bounded capacity in noncoherent wireless networks:* We rigorously establish an upper bound on the channel capacity, showing that capacity remains finite under practical fading conditions. Specifically, we prove that if the variances of the fading coefficients of the interferers decay at an exponential rate or slower, then the channel capacity is bounded in the transmit power. Thus, the presence of interference fundamentally limits the achievable rate, regardless of how high the transmission power is. This insight is validated through well-known propagation models, including the free-space path loss model, the two-ray model, and the Okumura-Hata model.
2) *Unbounded capacity under faster-than-double-exponential fading decay:* We establish that, if the variances of the fading coefficients of the interferers decay faster than double-exponentially, then the channel capacity is unbounded in the transmit power. This result highlights the critical role of the decay rate of fading variances in determining fundamental limits on information transmission.
3) *Effect of interference burstiness:* To establish the unbounded capacity-growth for faster than double-exponential fading decays, we propose an intermittent signaling scheme that artificially induces interference burstiness, creating opportunities where the channel is effectively interference-free. The proposed construction is deliberately simplified and does not rely on user cooperation. While suboptimal, it suffices to demonstrate that unbounded capacity growth can be achieved for a sufficiently fast decay rate of fading variances and sparse user activity. We therefore show the impact of interference burstiness on the network capacity, and demonstrate that under a sufficiently sparse user activity pattern, unbounded capacity growth can be achieved.

The rest of this paper is organized as follows. Section II introduces the cellular network model, detailing the system setup, key assumptions, and the fading model considered in our analysis. In Section III, we rigorously establish an upper bound on the channel capacity, demonstrating that, under practical fading conditions, capacity remains bounded in the transmit power. We derive explicit conditions under which this result holds and validate it through representative propagation models. Section IV shifts focus to achievable rates, showing that bursty signaling schemes can effectively mitigate interference and, under specific conditions, enable unbounded capacity growth. Finally, Section V concludes the paper with a summary and discussion of our results. Proofs and technical derivations are deferred to the appendices.

II. Network Model

We consider a wireless network in which users are associated with access nodes (ANs). Users associated with the same AN are allowed to cooperate. In contrast, users associated with different ANs do not cooperate, and their transmissions generate inter-AN interference. User activity in the network is assumed to be intermittent, so that only a subset of users is active at any given time, a model that captures both enhanced mobile broadband (eMBB) and massive machine-type communications (mMTC) scenarios. Fig. 1 illustrates the considered network model for both cell-based and cell-free architectures. In Fig. 1a, the cell in the center denotes the intended cell, while an infinite number of surrounding cells generate interference. In each cell, only a fraction of the users is active simultaneously. Solid lines represent desired communication links, whereas dotted lines correspond to interfering signals. Fig. 1b depicts a cell-free architecture in which each active device communicates with an associated access point. The AN in the center is the intended AN, with solid lines for the intended signal and dotted lines for interference from other active users.

Our performance measure is the capacity of the channel between the users and the intended AN (uplink transmission).

Since a complete characterization of the achievable rate region becomes analytically intractable as the number of cells and users grows, we focus on the *symmetric sum rate* of the network, i.e., the sum rate achievable by the users associated with the intended AN when all ANs communicate at the same sum rate. This metric captures the fundamental interference-limited behavior of large networks while ensuring fairness across users.

To isolate these fundamental effects and obtain tractable yet insightful capacity bounds, we adopt the following assumptions:

A1. *Infinitely many interfering users:* We model a large-scale network by assuming infinitely many interfering users. This assumption enables us to capture aggregate interference effects without imposing artificial network boundaries and does not restrict generality, since interfering users are parametrized by fading variances α_ℓ, $\ell = 1, 2, \ldots$, which can be set to $\alpha_\ell = 0$ to represent inactive or sufficiently distant users.

A2. *No cooperation:* We assume that users associated with different ANs do not cooperate. This reflects practical cellular and cell-free deployments where coordination across cells is limited and, more importantly, allows us to study intrinsic interference-limited phenomena. In particular, this restriction excludes the use of coordination strategies such as *time-division multiple access (TDMA)* or *frequency-division multiple access (FDMA)*, which could otherwise enable unbounded capacity in specific time or frequency slots.

A3. *Uniform coding across ANs:* To ensure that all users (intended and interfering) operate at the same rate, we invoke a random-coding argument and assume that all users and interferers draw their codebooks from the same distribution. This assumption generalizes the commonly-adopted Gaussian codebook model and aligns with large-system analyses where users are statistically homogeneous. It is also reminiscent of the unsourced random access (URA) framework [9], in which all nodes employ a common codebook in massive access scenarios.

A. Channel Model and User Activity

Let us denote by n_T the number of users associated with an AN, and by n_R the number of receive antennas of the AN itself. Throughout the paper, the indices ℓ, u, and k denote the AN, the user associated with that AN, and the time instant, respectively, so that $X_{\ell,u,k} \in \mathbb{C}$ represents the symbol transmitted by user u associated with ℓ-th AN at time k.

We consider an uplink transmission scenario in which the AN with $\ell = 0$ is the intended receiver, while the users associated with ANs $\ell = 1, 2, \ldots$ generate interference. The channel input-output relation at time $k \in \mathbb{Z}$ is given by

$$\mathbf{Y}_k = \mathbb{H}_{0,k}\mathbf{X}_{0,k} + \sum_{\ell=1}^{\infty} \mathbb{H}_{\ell,k}\mathbf{X}_{\ell,k} + \mathbf{Z}_k, \tag{1}$$

where

- $\mathbf{Y}_k \in \mathbb{C}^{n_R \times 1}$ is the vector of received symbols at the intended receiver at time k;
- $\mathbb{H}_{\ell,k} \in \mathbb{C}^{n_R \times n_T}$ is the fading matrix corresponding to the channel realizations between the users of the ℓ-th AN and the intended receiver;
- $\mathbf{X}_{\ell,k} = [X_{\ell,1,k}, \ldots, X_{\ell,n_T,k}]^T \in \mathbb{C}^{n_T \times 1}$ is the vector of symbols transmitted at time k by the users of the ℓ-th AN;
- $\mathbf{Z}_k \in \mathbb{C}^{n_R \times 1}$ is the additive noise vector at time k.

To simplify terminology, we shall sometimes refer to the set of users associated with the intended receiver as the *intended cell*, and to the set of users associated with AN $\ell = 1, 2, \ldots$ as an *intefering cell*. While this terminology is motivated by cell-based networks, it also remains valid for cell-free networks.

We assume that the additive noise $\{\mathbf{Z}_k, k \in \mathbb{Z}\}$ is a sequence of *independent and identically distributed (i.i.d.)* random variables with a circularly-symmetric, complex Gaussian distribution of zero mean and covariance matrix $\sigma^2\mathsf{I}$, i.e., $\{\mathbf{Z}_k, k \in \mathbb{Z}\} \sim$ i.i.d. $\mathcal{N}_{\mathbb{C}}(0, \sigma^2\mathsf{I})$, where I denotes the identity matrix and $\mathcal{N}_{\mathbb{C}}(\mu, \mathsf{K})$ denotes the circularly-symmetric, complex Gaussian distribution with mean μ and covariance matrix K. For simplicity, we further assume that the fading coefficients $\{\mathbb{H}_{0,k}, k \in \mathbb{Z}\}$ and $\{\mathbb{H}_{\ell,k}, k \in \mathbb{Z}\}$, $\ell = 1, 2, \ldots$ are sequences of i.i.d. random matrices, the former with i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$ entries and the latter with i.i.d. $\mathcal{N}_{\mathbb{C}}(0, \alpha_\ell)$ entries for some α_ℓ, $\ell = 1, 2, \ldots$ We consider a noncoherent scenario where transmitter and receiver only know the statistics of the fading coefficients but not their realizations.

The locations of the interfering users relative to the intended receiver enter the channel model through the variances α_ℓ, $\ell = 1, 2, \ldots$ of the fading coefficients $\{\mathbb{H}_{\ell,k}, k \in \mathbb{Z}\}$. For analytical simplicity, we assume that the entries of $\mathbb{H}_{\ell,k}$ corresponding to a given interfering cell have identical variances. This corresponds to the idealized case in which all nodes associated with the ℓ-th interfering cell are at the same distance from the intended receiver. This assumption can be relaxed to allow for user-dependent variances that account for different distances within the same interfering cell, provided that these variances scale proportionally with the large-scale attenuation factor α_ℓ. Moreover, frequency reuse schemes can be naturally incorporated in this framework by appropriately modifying the set of interfering cells to those sharing the same frequency band, which results in a faster effective decay of the coefficients $\{\alpha_\ell\}$.

We assume that the total power of the interference received at the intended receiver is finite, i.e.,

$$\sum_{\ell=1}^{\infty} \alpha_\ell < \infty. \tag{2}$$

Without loss of generality, we order the interfering users according to the fading variances of the corresponding ANs, i.e., $\alpha_\ell \geq \alpha_{\ell'}$ for any $\ell < \ell'$. See also Assumption A1.

To summarize, the number of users per cell n_T and the number of receive antennas per cell n_R are finite. In contrast, the number of interfering cells, indexed by $\ell = 1, 2, \ldots$, is infinite, which gives rise to the infinite sum in (1). The large-scale fading variances α_ℓ decay sufficiently fast to ensure that the total interference (2) is finite.

The activity of the users both inside the intended cell and the interfering cells is assumed to be intermittent. We model this activity as

$$X_{\ell,u,k} = B_{\ell,u}\tilde{X}_{\ell,u,k}, \quad \ell = 0, 1, \ldots \tag{3}$$

where $\tilde{X}_{\ell,u,k}$ denotes the symbol transmitted at time k by user u within cell ℓ and $B_{\ell,u}$ is a random variable that captures the activity of this user. We shall model this activity by a Bernoulli random variable $B_{\ell,u} \sim \text{Ber}(\delta)$, which remains constant during the entire transmission, i.e., for $k = 1, 2, \ldots, n$. Thus, users are active in bursts with probability δ for some $0 \leq \delta \leq 1$. We further assume that the user activities $B_{\ell,u}$ of different users (within the same cell or in different cells) are independent, i.e., the random variables $\{B_{\ell,u},\, u = 1, \ldots, n_T,\, \ell = 0, 1, \ldots\}$ are i.i.d.

Let $\tilde{\mathbf{X}}_{\ell,k} = [\tilde{X}_{\ell,1,k}, \ldots, \tilde{X}_{\ell,n_T,k}]^T \in \mathbb{C}^{n_T \times 1}$. We assume that the interferers do not cooperate with the transmitters in the intended cell (Assumption A2). We further invoke a random-coding argument and assume that the users within a cell draw their codebooks from the same distribution (Assumption A3). This implies that the sequences $\{\tilde{\mathbf{X}}_{\ell,k},\, k \in \mathbb{Z}\}$, $\ell = 0, 1, \ldots$ are independent and each such sequence has the same distribution. Finally, the activity patterns $\{B_{\ell,u},\, u = 1, \ldots n_T\}$, the additive noise $\{\mathbf{Z}_k,\, k \in \mathbb{Z}\}$, and the fading coefficients $\{\mathbb{H}_{\ell,k},\, k \in \mathbb{Z}\}$, $\ell = 0, 1, \ldots$ are assumed to be independent of each other.

B. Channel Capacity

For any time-indexed sequence $S_1, \ldots, S_n$, we use the shorthand $S_1^n \triangleq (S_1, \ldots, S_n)$. This notation also applies to vector-valued quantities such as $\mathbf{Y}_1^n \triangleq (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$. For vectors indexed by access point and time, we shall write $\mathbf{X}_{0,1}^n \triangleq (\mathbf{X}_{0,1}, \ldots, \mathbf{X}_{0,n})$.

We define the capacity of the channel model (1) as[1]

$$C(\mathsf{P}) \triangleq \lim_{n\to\infty} \frac{1}{n} \sup_{Q^n} I(\mathbf{X}_{0,1}^n; \mathbf{Y}_1^n). \tag{4}$$

Here, for $\mathbf{X}_{\ell,1}^n$ defined in (3), the supremum is over all n-dimensional probability distributions Q^n of $\tilde{\mathbf{X}}_{\ell,1}^n$, $\ell = 0, 1, 2, \ldots$ satisfying the per-user average power constraint

$$\int \|\tilde{x}_{\ell,u,1}^n\|^2 dQ^n(\tilde{x}_{\ell,u,1}^n) \leq n\mathsf{P},\ u = 1, 2, \ldots, n_T. \tag{5}$$

The logarithms used in this paper are natural logarithms. The capacity $C(\mathsf{P})$ has thus the dimension "nats per channel use". Intuitively, $C(\mathsf{P})$ characterizes the sum rate at which information can be transmitted to the intended receiver. In order to obtain the rate per user, one would need to divide $C(\mathsf{P})$ by n_T.

We do not claim that there is a coding theorem associated with (4), i.e., we do not claim that $C(\mathsf{P})$ represents the largest achievable rate in the sense that, for any rate below $C(\mathsf{P})$, there exists an encoding and decoding scheme for which the decoding error probability tends to zero as n tends to infinity. Nevertheless, by Fano's inequality [43, Sec. 7.9], any encoding-decoding scheme operating at a rate exceeding $C(\mathsf{P})$ must have a decoding error probability that is bounded away from zero as n tends to infinity. Consequently, any upper bound on (4) also serves as an upper bound on the largest achievable rate. Moreover, it can be shown that any rate obtained by evaluating (4) for an i.i.d. distribution of $(\tilde{\mathbf{X}}_{\ell,1}, \ldots, \tilde{\mathbf{X}}_{\ell,n})$ is achievable. It follows that the upper bounds derived in Section III, as well as the lower bound derived in Section IV for an i.i.d. distribution of $(\tilde{\mathbf{X}}_{\ell,1}, \ldots, \tilde{\mathbf{X}}_{\ell,n})$, also bound the largest achievable rate.

[1]It can be shown that the sequence $\{\sup_{Q^n} I(\mathbf{X}_{0,1}^n; \mathbf{Y}_1^n),\, n = 1, 2, \ldots\}$ is superadditive, so the limit in (4) exists by Fekete's lemma [42].

III. Bounded Capacity

A. Main Result

We next present the main result of this paper: an upper bound on the channel capacity $C(\mathsf{P})$ that applies when the variances α_ℓ, $\ell = 1, 2, \ldots$ decay exponentially or more slowly. Since this upper bound does not depend on P, it implies that, for such α_ℓ's, the capacity is bounded in the signal-to-noise ratio (SNR) of the system, defined as $\text{SNR} \triangleq \mathsf{P}/\sigma^2$.

Theorem 1 (Converse Bound): Consider the channel model introduced in Section II. Assume that, for some $0 < \rho < 1$,

$$\frac{\alpha_{\ell+1}}{\alpha_\ell} \geq \rho, \quad \ell = 1, 2, \ldots \tag{6}$$

Then, for every $\mathsf{P} > 0$ and $0 < \delta \leq 1$, the channel capacity is upper-bounded by

$$\begin{aligned} C(\mathsf{P}) \leq\ & n_R\big((2-\delta)^{n_T} - 1\big) \log\left(\rho^{-\frac{3}{2}}\right) \\ & + \big(1 - (1-\delta)^{n_T}\big)\left(\log \frac{\pi}{n_R \Gamma(n_R)} + n_R \log \frac{n_R}{e}\right. \\ & \left. + \frac{n_R}{2}\log(\eta_{\max}) + n_R \log(1 + \eta_{\max})\right) \end{aligned} \tag{7}$$

where $\Gamma(\cdot)$ denotes the gamma function, and where we define

$$\eta_{\max} \triangleq \max\left(\tfrac{1}{\alpha_1}, \tfrac{1}{\rho}\right). \tag{8}$$

Proof: See Appendix I. ∎

Observe that the upper bound (7) does not depend on the transmit power P and scales roughly linearly with the number of receive antennas at the AN n_R. Further observe that, for a fixed activity probability δ, the bound grows exponentially with the number of users n_T within the cell. However, this growth depends on δ: when $\delta \to 0$, the bound is roughly proportional to 2^{n_T}, whereas when $\delta \to 1$, the advantage of user cooperation within a cell vanishes.

Theorem 1 applies to the capacity definition in (4), which optimizes over all possible joint input distributions across users within each cell. This formulation allows us to establish that the bounded-capacity phenomenon is intrinsic to noncoherent interference networks and does not rely on specific implementation constraints. While this level of coordination provides the strongest possible performance benchmark, realizing such gains would require transmitters to adapt their joint signaling strategies to the instantaneous realizations of active-user sets across cells, which may be unrealistic in networks with sporadic and uncoordinated user activity. Motivated by this observation, we restrict attention to transmission strategies that are invariant with respect to the user index, i.e., strategies in

which users employ statistically identical signaling whenever they are active. Under this constraint, the coordination gains described above are no longer available. We next show that the resulting upper bound scales linearly with n_T, in contrast to the exponential growth in n_T obtained in Theorem 1.

We define the *exchangeable capacity* as

$$C_E(\mathsf{P}) \triangleq \lim_{n\to\infty} \frac{1}{n} \sup_{Q^n} I(\mathbf{X}_{0,1}^n; \mathbf{Y}_1^n), \tag{9}$$

where the supremum is taken over all n-dimensional probability distributions Q^n of $\tilde{\mathbf{X}}_{\ell,1}^n$, $\ell = 0, 1, 2, \ldots$ that satisfy the per-user average power constraint (5) and are *exchangeable* in the sense that, for every permutation π of $\{1, \ldots, n_T\}$,

$$(\tilde{X}_{\ell,1,1}^n, \ldots, \tilde{X}_{\ell,n_T,1}^n) \stackrel{d}{=} (\tilde{X}_{\ell,\pi(1),1}^n, \ldots, \tilde{X}_{\ell,\pi(n_T),1}^n). \tag{10}$$

Note that the property of exchangeability is closely related to the concept of i.i.d. random variables: any sequence of random variables that are i.i.d. across users within a cell is indeed exchangeable.

Proposition 1 (Exchangeable Capacity): Assume that (6) holds. Then, for every $\mathsf{P} > 0$ and $0 < \delta \leq 1$, the exchangeable channel capacity is upper-bounded by

$$\begin{aligned} C_E(\mathsf{P}) \leq\ & n_R n_T (1-\delta) \log\left(\rho^{-\frac{3}{2}}\right) \\ & + \left(1-(1-\delta)^{n_T}\right)\Bigg(\log\frac{\pi}{n_R\Gamma(n_R)} + n_R \log\frac{n_R}{e} \\ & + \frac{n_R}{2}\log(\eta_{\max}) + n_R\log(1+\eta_{\max})\Bigg). \end{aligned} \tag{11}$$

Proof: See Appendix II. ∎

We observe that, in contrast to Theorem 1, the upper bound on the exchangeable capacity (11) scales linearly with the number of transmitters n_T. To illustrate the impact of user activity and the decay pattern of the variances of the fading coefficients on the capacity upper bounds, Fig. 2 presents the upper bounds on the channel capacity (Theorem 1) and on the exchangeable channel capacity (Proposition 1) for $n_T = n_R = 2$. Specifically, Fig. 2a shows the upper bounds as a function of the user activity probability δ, while Fig. 2b depicts the upper bounds as a function of ρ, which is related to the decay of the variances of fading coefficients. This example is intentionally presented for a small number of transmitters in order to clearly expose the qualitative behavior and scaling trends of the derived bounds. The same insights apply to larger values of n_T and n_R as established by the analytical results.

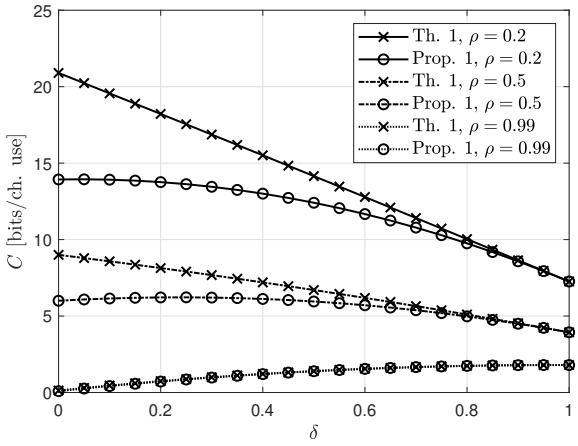
Observe that the impact of the user activity on the capacity upper bounds depends on ρ. In fact, the upper bounds suggest that, when ρ is close to 1, a small user activity is detrimental, whereas for small values of ρ, a small user activity is beneficial. Intuitively, a small user activity reduces the interference from other cells, but also reduces the time when users in the intended cell transmit information. Since the benefits of reduced interference are more pronounced when the variances of the fading coefficients decay faster, it follows that in this case, a reduced user activity is the most beneficial.

Further observe that, when ρ is close to 1, the upper bounds are below 2.5 bits per channel use, irrespective of the user activity δ. Intuitively, the assumption (6) requires that the variances $\{\alpha_\ell\}$ decay at most exponentially. Indeed, letting $\alpha_0 = 1$, we have for every integer L that

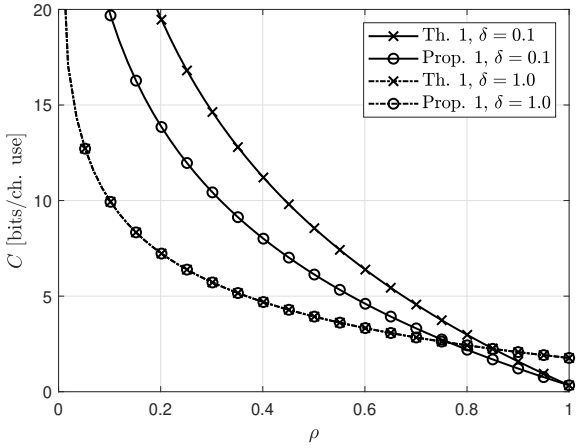
$$\frac{1}{L}\log\frac{1}{\alpha_L} = \frac{1}{L}\sum_{\ell=0}^{L-1}\log\frac{\alpha_\ell}{\alpha_{\ell+1}}. \tag{12}$$

For any sequence $\{\alpha_\ell\}$ satisfying (6), the right-hand side (RHS) of (12) is upper-bounded by $\log(\eta_{\max})$, so the left-hand side of (12) must be finite. This in turn implies that the sequence $\{\alpha_\ell\}$ decays exponentially or more slowly. In the limiting case as $\rho \to 1$, the decay of $\{\alpha_\ell\}$ becomes sub-exponential. As we shall argue in Section III-B, such a decay is consistent with most common propagation models. This suggests that, for $n_T = n_R = 2$, the maximum rate achievable in noncoherent wireless networks does not exceed 2.5 bits per channel use.

Finally, observe that, as ρ vanishes, our upper bounds grow to infinity. The case $\rho \to 0$ corresponds to the case where (6) is not satisfied, i.e., where $\{\alpha_\ell\}$ decays faster than exponentially.



(a) Upper bounds as a function of the user activity probability δ.



(b) Upper bounds as a function of the interference decay factor ρ.

Fig. 2. Upper bounds on the channel capacity (Theorem 1) and the exchangeable channel capacity (Proposition 1) for $n_T = n_R = 2$.

In this scenario, Theorem 1 does not apply and, consequently, the channel capacity may be unbounded. In Section IV, we show that this is the case when $\{\alpha_\ell\}$ decays faster than double-exponentially. Nevertheless, as we shall argue next, the assumption of an exponential or sub-exponential decay of $\{\alpha_\ell\}$ is reasonably mild and satisfied in most cases.

B. Decay Patterns of Common Path Loss Models

In practical wireless communication scenarios, most propagation models assume that the path loss increases polynomially with distance, capturing the attenuation of signal power as it propagates through space. Examples include the *Free-Space Path Loss (FSPL)* model, applicable in idealized line-of-sight conditions; the two-ray model, which extends FSPL by incorporating ground-reflected signals; the single-slope path loss exponent model, which generalizes the polynomial decay to accommodate different propagation environments using an empirically determined exponent; and the Okumura-Hata model, based on extensive measurement data, among others [44], [45].

Assume that the locations of interfering nodes can be modeled as a spatially random process over the plane. If the network cells follow a homogeneous Poisson point process with intensity η, the number of nodes in any region of area A follows a Poisson distribution with mean ηA, and the node locations are independent and uniformly distributed within that area. This statistical model accurately reflects scenarios where base stations or users are randomly deployed, such as in large-scale wireless networks or unplanned urban deployments. The following proposition establishes a relation between the expected ordered distances d_ℓ and the corresponding interfering cell index $\ell = 1, 2, \ldots$

Proposition 2 (Ordered Distances): Let x_0 be a reference point in a 2-dimensional plane where points are distributed according to a homogeneous Poisson point process Φ with intensity η. Denote by D_ℓ the distance from x_0 to the ℓ-th closest point. Then, the expectation of this distance with respect to Φ satisfies

$$\mathbb{E}_\Phi[D_\ell] = \sqrt{\frac{\ell}{\eta\pi}}. \tag{13}$$

Proof: The number of points within a radius r follows a Poisson distribution with mean $\eta\pi r^2$. Therefore, the expected number of points $N(r)$ inside a ball of radius r is given by

$$\mathbb{E}_\Phi[N(r)] = \eta\pi r^2. \tag{14}$$

Since D_ℓ is the distance to the ℓ-th closest point, on average there should be ℓ points within a radius $r = \mathbb{E}_\Phi[D_\ell]$. Substituting this value in (14) leads to the identity

$$\ell = \eta\pi\mathbb{E}_\Phi[D_\ell]^2. \tag{15}$$

Solving for $\mathbb{E}_\Phi[D_\ell]$ yields the desired result. ∎

By Proposition 2, the distance from the ℓ-th nearest interfering node to the intended node grows proportionally to $\sqrt{\ell}$. This scaling behavior arises from the expected spacing between points for a homogeneous Poisson point process with a certain intensity η. It establishes a direct relationship between the spatial distribution of interfering nodes and the decay rate of the associated fading coefficients. Since path loss typically follows an inverse polynomial dependence on distance, the ordered distances d_ℓ determine how fast the variances of the fading coefficients $\{\alpha_\ell\}$ decay with ℓ. In particular, if the path-loss function $\mathrm{PL}(d)$ follows a power-law behavior of the form

$$\mathrm{PL}(d) \propto d^\beta, \tag{16}$$

where β is the path-loss exponent, then using Proposition 2, we obtain that, for an average cell deployment,

$$\mathbb{E}_\Phi[\alpha_\ell] \approx \frac{1}{\mathrm{PL}(\mathbb{E}_\Phi[D_\ell])} \propto \ell^{-\beta/2}. \tag{17}$$

For instance, the free-space path loss model has $\beta = 2$. Therefore, the expected fading variances $\{\alpha_\ell\}$ behave linearly with the cell-index ℓ, $\mathbb{E}_\Phi[\alpha_\ell] \propto \ell^{-1}$. We conclude that, for typical path loss models, the variances of the fading coefficients $\{\alpha_\ell\}$ decay slower than exponentially for an average cell deployment, and Theorem 1 implies that the network has bounded capacity.

A similar argument applies when considering frequency-reuse schemes, where only a subset of cells share the same frequency band and therefore act as co-channel interferers. In this case, the number of interfering cells is effectively reduced by a fixed fraction, but the ordering of distances from the intended node remains unchanged. Consequently, the overall scaling behavior of the fading variances $\{\alpha_\ell\}$ with ℓ is preserved, and the asymptotic decay regime (slower than exponential, exponential, etc.) remains the same. The only effect is a potential modification of the effective path-loss exponent: if a fraction of the nearest interferers are removed due to frequency reuse, the average distance to the ℓ-th co-channel interferer increases, which can be captured as a larger effective β in the power-law model $\alpha_\ell \propto \ell^{-\beta/2}$.

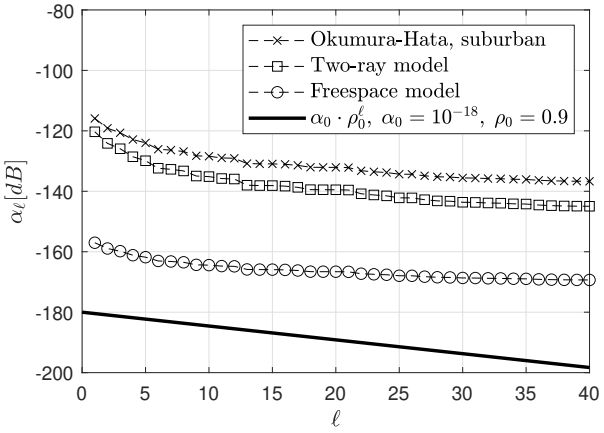


Fig. 3. Decay patterns with respect to the cell index ℓ under typical path-loss models for one realization of the Poisson point process Φ with intensity $\eta = 3.2$ cells per square km.

Fig. 3 illustrates the fading variances (in dB) predicted by different path loss models for one realization of cells that are randomly deployed on a two-dimensional surface with a density of 3.2 cells per square kilometer. The parameters for

the Okumura-Hata model (suburban environment) are set as follows: frequency $f = 1.5$ GHz, base-station height $h_b = 50$ meters, and mobile user height $h_m = 1.5$ meters. For the free-space and two-ray models, we use the same antenna heights, a link gain of $G = 10^{-4}$, and a transmission frequency of $f = 2.4$ GHz. For comparison, we further illustrate the fading variances $\alpha_\ell = 10^{-18}(0.9)^\ell$, having an exponential decay of $\{\alpha_\ell\}$ with $\eta_{\max} = 10^{18}$. We observe that, due to the polynomial behavior of the path loss in different models, the fading variances $\{\alpha_\ell\}$ decay more slowly than exponential.

C. Fundamental Limits of Cooperation

Lozano, Heath, and Andrews [33] studied the fundamental limits of cooperation in wireless networks across a variety of scenarios. Among these, the case most relevant to our work is their analysis of the block-fading MIMO channel model. In this model, a cell comprises n_T transmitters and n_R receivers, and the channel remains constant over T symbol durations. When $n_T > T$, there are insufficient degrees of freedom within each coherence interval to estimate all channel coefficients independently, so this regime falls under the scope of noncoherent communication. They showed that, under these assumptions, the channel capacity is bounded in the SNR.

In order to compare our results with those of [33], we focus on the specific case where $T = 1$. In this case, the channel model considered in [33] has essentially the input-output relation

$$\mathbf{Y}_k = \sqrt{\mathsf{P}}\mathbb{H}_k\mathbf{X}_k + \mathbf{Z}_k \tag{18}$$

where $\mathbf{X}_k \in \mathbb{C}^{n_T\times 1}$ corresponds to the vector of symbols transmitted at time k by the users inside the intended cell, $\mathbb{H}_k \in \mathbb{C}^{n_R\times n_T}$ denotes the matrix of fading coefficients, and $\mathbf{Z}_k \in \mathbb{C}^{n_R\times 1}$ denotes the additive noise vector. The sequence of noise vectors $\{\mathbf{Z}_k,\, k \in \mathbb{Z}\}$ is i.i.d. with independent, zero-mean, unit-variance, circularly-symmetric, Gaussian entries. Similarly, $\{\mathbb{H}_k,\, k \in \mathbb{Z}\}$ is a sequence of i.i.d. matrices with independent, zero-mean, circularly-symmetric, Gaussian entries having variance $g_{r,u} > 0$, $r = 1,\ldots,n_R$, $u = 1,\ldots,n_T$; cf. [33, Eq. (4)]. The vector of transmitted symbols $\mathbf{X}_k$ is assumed to be independent of P and to have unit-energy entries. As can be observed by comparing (18) with (1), in the channel model considered in [33], out-of-cluster interference is incorporated into the additive noise $\mathbf{Z}_k$. Furthermore, the channel input $\sqrt{\mathsf{P}}\mathbf{X}_k$ in (18) corresponds to $\mathbf{X}_k$ in (1). Lozano, Heath, and Andrews obtained the following upper bound on the information rate $I(\mathbf{X}_k;\mathbf{Y}_k)$.

Proposition 3 (Lozano, Heath, and Andrews [33]): Consider the channel model (18) of a cellular system with $n_T > 1$ transmitting users. For each $r = 1,\ldots,n_R$, define the diagonal matrix $\mathsf{G}_r = \mathrm{diag}(g_{r,1},\ldots,g_{r,n_T})$, and assume that G_r and $\mathbf{X}_k$ satisfy $g_{r,1} + \ldots + g_{r,n_T} = 1$ and $\mathsf{E}[\log(\mathbf{X}_k^\dagger\mathsf{G}_r\mathbf{X}_k)] > -\infty$. Then

$$\overline{\lim_{\mathsf{P}\to\infty}}\, I(\mathbf{X}_k;\mathbf{Y}_k) \le -\sum_{r=1}^{n_R}\mathsf{E}\left[\log(\mathbf{X}_k^\dagger\mathsf{G}_r\mathbf{X}_k)\right] \tag{19}$$

where $\overline{\lim}$ denotes the *limit superior*.

Proof: See [33, App. A]. ∎

The bound (19) has a similar flavor as our upper bound (7) in Theorem 1: both bounds imply that channel capacity (under some constraints on the channel inputs) is bounded in the transmit power. However, (19) requires more stringent conditions on $\mathbf{X}_k$ than (7). Among other things, the condition that $\mathbf{X}_k$ is independent of P is rather restrictive. In fact, it was shown by Lapidoth and Moser [34, Th. 4.3] that the information rates achievable over (18) with inputs of the form $\sqrt{\mathsf{P}}\mathbf{X}_k$ (where $\mathbf{X}_k$ is independent of $\sqrt{P}$) are bounded in the transmit power. Thus, such inputs result in a bounded information rate even for a single-user channel. In other words, the boundedness of the information rate $I(\mathbf{X}_k;\mathbf{Y}_k)$ may be a consequence of the suboptimal input distribution rather than of the interference from other cells. Furthermore, the condition that $\mathsf{E}[\log(\mathbf{X}_k^\dagger\mathsf{G}_r\mathbf{X}_k)] > -\infty$ is not compatible with the random-access scenario considered in this paper, where users may be inactive with a positive probability. In contrast, our upper bound (7) in Theorem 1 proves the boundedness of channel capacity even without imposing the aforementioned conditions on $\mathbf{X}_k$.

Intuitively, when users are inactive with a positive probability, the upper bound (19) becomes infinite, and channel capacity is potentially unbounded. To see this, let us model the random user activity again by writing the channel inputs as

$$X_{u,k} = B_u\tilde{X}_{u,k}, \quad u = 1,\ldots,n_T, \tag{20}$$

where $\tilde{X}_{u,k}$ denotes the symbol transmitted by user u at time k, and B_u is a Bernoulli random variable that is 1 with probability δ and 0 otherwise. The random variable B_u captures the activity of user u and remains constant during the entire transmission. By following the same steps as in the proof of Proposition 3, we can derive the following upper bound analogous to (19).

Proposition 4 (Random-Access Upper Bound): Consider the channel model (18) of a cellular system with $n_T > 1$ transmitting users. Let $\mathbf{X}_k = [X_{1,k},\ldots,X_{n_T,k}]^T$ be given by (20), and assume that $\tilde{\mathbf{X}}_k = [\tilde{X}_{1,k},\ldots,\tilde{X}_{n_T,k}]^T$ is independent of P and its entries have unit energy and satisfy $\mathsf{E}[\log|\tilde{X}_{t,k}|^2] > -\infty$ for every $t = 1,\ldots,n_T$. Further assume that the fading variances are strictly positive and satisfy $g_{r,1}+\ldots+g_{r,n_T} = 1$ for every $r = 1,\ldots,n_R$. Then,

$$\begin{aligned} I(\mathbf{X}_k;\mathbf{Y}_k) &\le n_R(1-\delta)^{n_T}\log\mathsf{P} + n_T H_b(\delta) + n_R\log\left(1+\frac{1}{\mathsf{P}}\right) \\ &\quad - (1-(1-\delta)^{n_T})\sum_{r=1}^{n_R}\mathsf{E}\left[\log(|\tilde{X}_{u_\star,k}|^2 g_{r,u_\star})\right], \end{aligned} \tag{21}$$

where $H_b(\cdot)$ denotes the binary entropy function and

$$u_\star \triangleq \arg\min_{u=1,\ldots,n_T}\mathsf{E}\left[\log(|\tilde{X}_{u,k}|^2 g_{r,u})\right]. \tag{22}$$

Proof: See Appendix III. ∎

For $\delta < 1$, the upper bound (21) grows logarithmically in P and fails to capture the boundedness of channel capacity in P. However, in contrast to (7), it does not depend on the decay rate of the variances of the fading coefficients $\{\alpha_\ell\}$. Furthermore, it suggests that, by artificially reducing the

user activity, one may achieve an unbounded capacity when Theorem 1 does not apply, i.e., when $\{\alpha_\ell\}$ decay faster than exponentially. We shall refer to this transmit strategy as *bursty signaling* and show in the next section that it indeed achieves an unbounded information rate when $\{\alpha_\ell\}$ decay faster than double-exponentially.

IV. Bursty Signaling

In the previous section, we have seen that whenever the fading variances $\{\alpha_\ell\}$ decay exponentially or sub-exponentially, noncoherent wireless networks have bounded capacity. In this section, we explore how to attain an unbounded coding rate if we relax this restriction.

Recall that the signal transmitted by user u at time k is modeled as

$$X_{\ell,u,k} = B_{\ell,u}\tilde{X}_{\ell,u,k}, \quad \ell = 0, 1, \dots \tag{23}$$

where $B_{\ell,u}$ captures the activity of the user. We next follow a bursty signaling strategy to achieve an unbounded rate. To this end, consider an input distribution of the form

$$\tilde{X}_{\ell,u,k} = \hat{B}_{\ell,u}\hat{X}_{\ell,u,k}, \quad \ell = 0, 1, \dots \tag{24}$$

where, for $u = 1$, $\{\hat{B}_{\ell,u}\}$ are i.i.d. Bernoulli random variables with activation probability ξ and $\{\hat{X}_{\ell,u,k}\}$ are i.i.d. circularly-symmetric random variables with $\log|\hat{X}_{\ell,u,k}|^2$ uniformly distributed over the interval $[0, \log \mathsf{P}]$; for $u = 2, \dots, n_T$, $\hat{B}_{\ell,u}$ and $\hat{X}_{\ell,u,k}$ are set to zero. Intuitively, the random variable $\hat{B}_{\ell,u}$ artificially reduces the user activity. This, in turn, reduces the interference from interfering cells to the intended cell. The distribution of $\hat{X}_{\ell,u,k}$ is the one that achieves the high-SNR asymptotic capacity of noncoherent single-user fading channels; cf. [34].

Let $L_{\mathsf{P}}^\star \in \mathbb{N}$ denote the smallest integer L_{P} such that

$$\sum_{\ell=L_{\mathsf{P}}+1}^{\infty} \alpha_\ell \mathsf{P} \le \sigma^2. \tag{25}$$

Thus, $L_{\mathsf{P}}^\star$ represents the minimum number of dominant interfering cells that must be retained so that the residual interference remains below the noise level.

If there are only finitely many interfering ANs, say L, then $L_{\mathsf{P}}^\star = \mathsf{L}$ satisfies (25) independently of P. Otherwise, since the aggregate interference from the remaining cells scales with P, $L_{\mathsf{P}}^\star$ must grow unboundedly with P in order to satisfy (25). The growth rate of $L_{\mathsf{P}}^\star$ is determined by the decay of the fading variances $\{\alpha_\ell\}$. For example, when $\{\alpha_\ell\}$ decays exponentially as $\alpha_\ell = \rho^\ell$, $\ell = 1, 2, \dots$ for some $0 < \rho < 1$, then the left-hand side of (25) becomes $\mathsf{P}\rho^{L_{\mathsf{P}}}/(1-\rho)$ and $L_{\mathsf{P}}^\star$ is given by

$$L_{\mathsf{P}}^\star = \left\lceil \frac{\log P/\sigma^2}{\log(1/\rho)} + \frac{\log(1-\rho)}{\log(\rho)} - 1 \right\rceil. \tag{26}$$

The behavior of $L_{\mathsf{P}}^\star$ for other decays of $\{\alpha_\ell\}$ can be obtained along similar lines. Of particular interest in this section is the double-exponential decay, which is considered in the following lemma.

Lemma 1 ($L_{\mathsf{P}}^\star$ for Double-Exponential Decay): For some $a \ge 1$, let the fading variances $\{\alpha_\ell\}$ satisfy

$$\alpha_\ell \le \frac{1}{\exp(\exp(\ell^a))}, \quad \ell = 1, 2, \dots \tag{27}$$

Then, any integer L_{P} satisfying

$$L_{\mathsf{P}} > \left(\log\log\left(\frac{\mathsf{P}}{\sigma^2} \frac{e^{-a}}{1-e^{-a}} \right) \right)^{\frac{1}{a}} \tag{28}$$

will also satisfy (25).

Proof: See Appendix IV. ∎

Intuitively, introducing $L_{\mathsf{P}}^\star$ allows us to model the network as an interference network with only $L_{\mathsf{P}}^\star$ interfering cells by treating the remaining cells as noise. Following this strategy together with the bursty signaling scheme described in (24), we obtain the following lower bound on $C(\mathsf{P})$.

Theorem 2 (Bursty-Signaling Lower Bound): Consider the channel model from Section II. Then, for every $\mathsf{P} > 0$ and $0 < \delta \le 1$, the bursty signaling scheme (24) with activation probability ξ achieves the lower bound

$$C(\mathsf{P}) \ge \delta\xi(1-\delta\xi)^{L_{\mathsf{P}}} \times \times \Big(\log\log \mathsf{P} - \gamma - \log(e) - 2\log\big(1+\sqrt{2}\sigma\big) \Big), \tag{29}$$

where L_{P} is an arbitrary integer satisfying (25) and γ denotes Euler's constant.

Proof: See Appendix V. ∎

The result in Theorem 2 is based on a bursty signaling scheme that exploits favorable network activity realizations in which all nearby interferers remain silent. In such realizations, the intended cell effectively operates as a noncoherent point-to-point channel, whose capacity is known to grow double-logarithmically with the transmit power [34]. The lower bound (29) is obtained by considering only these favorable transmission opportunities and discarding all other activity realizations. While this approach is not expected to yield tight bounds, it highlights the fundamental mechanisms that allow the sum rate to become unbounded. Moreover, this strategy is reminiscent of classical contention-based random access schemes, such as ALOHA, where users exploit interference-free transmission opportunities created by sporadic activity patterns.

When there are only a finite number of interfering cells, $L_{\mathsf{P}}^\star$ is bounded. In this case, the lower bound in Theorem 2 implies that $C(\mathsf{P})$ is unbounded and grows at least double-logarithmically with P. However, in general $L_{\mathsf{P}}^\star$ grows with P, in which case the term $(1-\delta\xi)^{n_T L_{\mathsf{P}}^\star}$ vanishes. It then depends on the order of $L_{\mathsf{P}}^\star$ whether the RHS of (29) is unbounded. For example, when $\{\alpha_\ell\}$ decays exponentially, $L_{\mathsf{P}}^\star$ grows logarithmically with P. It can then be shown that (29) vanishes as $\mathsf{P} \to \infty$, even if, for every P, we optimize the lower bound over ξ. This is consistent with the upper bound presented in Theorem 1, which applies to this case and is bounded.

In contrast, when the sequence $\{\alpha_\ell\}$ decays faster than double-exponentially, the lower bound (29), and hence also $C(\mathsf{P})$, are unbounded, as shown in the next corollary. To this end, we need to choose the parameter ξ, which specifies the extra burstiness introduced by the signal scheme (24), to be a decaying function of P. More precisely, we choose

$$\xi = \big(\log\log \mathsf{P}\big)^{-\frac{1+\varepsilon}{a}}, \tag{30}$$

where $a > 1$ is as in (27) and determines the rate of $\{\alpha_\ell\}$, and $0 < \varepsilon < a - 1$ is an arbitrary parameter that determines the growth of the lower bound (29) with P.

Table I
ASYMPTOTIC CAPACITY BEHAVIOR AS $\mathsf{P} \to \infty$ UNDER DIFFERENT FADING-DECAY REGIMES

Fading-Decay Regime	Capacity	Reference
Exponential or slower (6)	Bounded	Th. 1 and Prop. 1
Between exponential and doubly-exponential	Unknown	–
Faster than doubly-exponential (27)	Unbounded	Th. 2 and Cor. 1

Corollary 1 (Achievable Rate for Double-Exponential Decay): Consider the channel model introduced in Section II. Further assume that the fading variances $\{\alpha_\ell\}$ satisfy (27) for some $a > 1$. Then, for every $\mathsf{P} > 0$ and $0 < \delta \leq 1$, the bursty signaling scheme (24) achieves

$$\varliminf_{\mathsf{P}\to\infty} \frac{C(\mathsf{P})}{\delta\big(\log\log \mathsf{P}\big)^{1-\frac{1+\varepsilon}{a}}} > 0 \tag{31}$$

for every $0 < \varepsilon < a - 1$, where $\varliminf$ denotes the *limit inferior*.

Proof: See Appendix VI. ∎

Corollary 1 implies that the capacity $C(\mathsf{P})$ is asymptotically lower-bounded by

$$C(\mathsf{P}) \gtrapprox \delta\big(\log\log \mathsf{P}\big)^{\eta} \tag{32}$$

for every $0 < \eta < (a-1)/a$. Clearly, this lower bound is unbounded, even though it grows only double-logarithmically with P. In any case, the double-logarithmic growth is inherent to the memoryless noncoherent fading model and does not stem from the bursty signaling scheme (24). In fact, even in the absence of interfering cells, the capacity grows only double-logarithmically with P; e.g., see [34]. We believe the approach of introducing signal burstiness to reduce inter-cell interference may be a promising transmission strategy at high SNR, where interference is the most harmful.

V. CONCLUSIONS

This work provides a comprehensive analysis of the capacity of interference-limited, noncoherent networks under general fading conditions. We modeled the network as a fading channel with an infinite number of interfering users, where users associated with different ANs do not cooperate but are constrained to use codebooks generated according to the same distribution. We considered a noncoherent scenario in which users and receivers know only the statistics of the fading coefficients, but not their realizations—an assumption that we believe is realistic when the number of users and ANs is large. We further assumed that users access the network at random with a positive activity probability δ. This model combines the network model of Lozano, Heath, and Andrews [33] with the assumptions of random user activity and an infinite number of interferers, as considered in the massive random access literature. To avoid that the number of bits per user and channel use vanishes as the blocklength tends to infinity, we accounted for the path loss of interfering cells, ensuring that interference from distant cells diminishes. Overall, this model captures the fundamental features of both eMBB and mMTC scenarios.

A. Key Results

For the considered channel model, we derived rigorous upper and lower bounds on the capacity, i.e., the maximum rate at which information can be transmitted reliably. The behavior of these bounds depends critically on the decay of the fading variances $\{\alpha_\ell\}$, which represent the path loss from interfering cells:

- When the fading variances decay at an exponential rate or slower, the capacity remains bounded in the transmit power.
- When the fading variances decay faster than double-exponentially, the capacity becomes unbounded with increasing transmit power.
- In the intermediate regime, where the decay is faster than exponential but slower than double-exponential, the behavior of the capacity remains an open problem.

These results are summarized in Table I. The analysis of the upper bounds in Theorem 1 and Proposition 1 further indicates that any super-linear scaling of achievable rates with the number of users would require strong forms of intra-cell coordination, which are unlikely to be realizable in large-scale random access networks. Finally, we investigated the effect of intermittent user activity. While reduced activity lowers interference, it does not overcome boundedness in the slow-decay regime. Nevertheless, it can significantly lower effective interference and increase the absolute value of achievable rates. In particular, bursty signaling creates favorable transmission opportunities that raise the operational capacity ceiling. This effect is illustrated by our numerical example with $\rho = 0.5$ and $n_T = n_R = 2$, where the upper bound increases from 3.94 to 6.70 bits per channel use when the user activity probability decreases from $\delta = 1$ to $\delta = 0.5$.

The bounds derived in Theorems 1 and 2 and Proposition 1 are not meant to be tight in a quantitative sense. In fact, a direct numerical comparison between the converse and achievability bounds is not meaningful, since they apply to fundamentally different interference-decay regimes: the upper bounds characterize scenarios with slower-than-exponential decay of the fading variances, whereas the achievability results rely on doubly-exponential decay to demonstrate unbounded capacity. Nevertheless, taken together, these results provide a coherent picture of how interference decay and user activity jointly shape the fundamental limits of large-scale noncoherent cellular networks, as discussed above.

Our analysis complements the work of Lozano, Heath, and Andrews [33], who analyzed a different network model and showed that capacity is also bounded in transmit power under certain constraints on the input distributions. In particular, [33] requires that the channel inputs are from a scale family—i.e., the inputs can be written as $\sqrt{P}\mathbf{X}_k$ for a vector $\mathbf{X}_k$ that is independent of P—and the entries of $\mathbf{X}_k$ have unit energy and a finite logarithmic moment. This excludes bursty signaling strategies, such as the ones we employed in the proof of Theorem 2 to show unbounded capacity when the variances of the fading coefficients decay faster than exponentially. In fact, it was shown in [34, Th. 4.3] that, for noncoherent point-to-point fading channels, the rate achievable by any scale

family of input distributions is bounded in the transmit power. We therefore argue that the constraints on the channel inputs imposed in [33] are rather restrictive and could potentially be the reason for a bounded coding rate. Our results show that this is not the case, unless the variances of the fading coefficients decay faster than exponentially: our upper bound presented in Theorem 1 does not require these constraints and is still bounded in the transmit power. This confirms the existence of a saturation regime in interference-limited networks, where capacity reaches a ceiling that is independent of power.

B. Implications and Insights

Our findings have several important implications for interference-limited and massive-access networks. First, for networks where the fading variances decay slowly, capacity saturates at high power, confirming a fundamental ceiling independent of the transmit power. While this result is based on the key assumption that inter-cell cooperation is not allowed, frequency or time reuse across cells can be incorporated in our framework by adjusting the large-scale fading statistics to account for co-channel interference. Therefore, unless reuse schemes modify the decay of fading variances significantly, the bounded-capacity result from Theorem 1 still holds.

Second, user activity and intra-cell cooperation strongly affects performance: bursty signaling and well-designed frequency or time reuse can reduce interference and increase achievable rates, especially when fading variances decay rapidly. Although capacity remains asymptotically bounded at high power, these strategies can raise the effective ceiling to levels that make the bound practically irrelevant. Exploiting frequency or time reuse schemes and activity-aware scheduling further enhances performance without requiring complex inter-cell coordination. Our results also suggest that any super-linear scaling of rates with the number of users would require strong forms of intra-cell coordination, which are unlikely to be feasible in large-scale, random-access deployments.

Finally, unlike prior work [33], which restricts inputs to a scale family, our analysis allows general input distributions and demonstrates that bursty signaling can achieve unbounded rates in scenarios with fast-decaying fading coefficients. Together, these insights clarify the practical limitations and opportunities in dense and massive-access networks, and provide guidance for designing interference-aware and network-aware transmission strategies.

Appendix I
Proof of Theorem 1

A. Proof Outline

The proof of Theorem 1 consists of various parts, as we shall outline below:

1) Introducing $\mathcal{J}$-interfering cells: We begin by defining $\mathcal{J}$ as the random set of active users in the intended cell, i.e.,

$$\mathcal{J} \triangleq \big\{u = 1, \ldots, n_T \colon B_{0,u} = 1\big\}. \tag{33}$$

To facilitate the derivation of a converse bound, we restrict attention to a subset of interfering cells, termed *$\mathcal{J}$-interfering cells*. Specifically, the ℓ-th interfering cell is said to be a $\mathcal{J}$-interfering cell if the set of its active users contains $\mathcal{J}$, i.e.,

$$\mathcal{G}_\ell \triangleq \big\{u = 1, \ldots, n_T \colon B_{\ell,u} = 1\big\}, \tag{34}$$

and $\mathcal{J} \subseteq \mathcal{G}_\ell$.

As an example, consider a system with $n_T = 3$ users per cell. Suppose that the set of active users in the intended cell is $\mathcal{J} = \{1, 3\}$. An interfering cell ℓ is said to be $\mathcal{J}$-interfering if its set of active users $\mathcal{G}_\ell$ contains both users 1 and 3. Thus, cell ℓ is $\mathcal{J}$-interfering if $\mathcal{G}_\ell = \{1, 2, 3\}$ or $\mathcal{G}_\ell = \{1, 3\}$.

In the proof, we retain only the interference generated by users in $\mathcal{J}$ from $\mathcal{J}$-interfering cells, and we assume that a genie provides side information allowing the receiver to remove all remaining interference components. This construction is used solely as an analytical tool: by selectively discarding interference, the resulting channel is less noisy than the original one, which ensures that any capacity upper bound derived under this assumption remains valid for the original network model. To this end, we introduce the indicator variable A_ℓ, which equals 1 if the ℓ-th cell is $\mathcal{J}$-interfering and equals 0 otherwise, i.e.,

$$A_\ell = \begin{cases} 1, & \mathcal{J} \subseteq \mathcal{G}_\ell, \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

To derive an upper bound, we assume that a "genie" provides side information about the interference caused by users not in $\mathcal{J}$ in every interfering cell, as well as by users in $\mathcal{J}$ of non-$\mathcal{J}$-interfering cells. This side information allows the receiver to remove these interference components, yielding the following initial upper bound:

$$I(\mathbf{X}_{0,1}^n; \mathbf{Y}_1^n) \le H(\mathcal{J}) + I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \,|\, A_1^L, \mathcal{J}), \tag{36}$$

where L is an arbitrary integer that we will let go to infinity at the end of the proof, and $\hat{\mathbf{Y}}_1^n$ represents the received signal when in each cell either all users in $\mathcal{J}$ are active or none of them is active, i.e.,

$$\hat{\mathbf{Y}}_k = \mathbb{H}_{0,k}\mathbf{X}_{0,k} + \sum_{\ell=1}^{\infty} A_\ell \mathbb{H}_{\ell,\mathcal{J},k}\mathbf{X}_{\ell,\mathcal{J},k} + \mathbf{Z}_k. \tag{37}$$

Here and throughout this appendix, we use the notation $\mathbb{H}_{\ell,\mathcal{A},k}$ to denote a matrix with the columns $a \in \mathcal{A}$ of $\mathbb{H}_{\ell,k}$, and we use $\mathbf{X}_{\ell,\mathcal{A},k}$ to denote the vector with the components $a \in \mathcal{A}$ of the vector $\mathbf{X}_{\ell,k}$.

2) Decomposition into differential entropies: We express the conditional mutual information $I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \,|\, A_1^L, \mathcal{J})$ using differential entropies and lower-bound $h(\hat{\mathbf{Y}}_1^n \,|\, \mathbf{X}_{0,1}^n, A_1^L, \mathcal{J})$ by conditioning on $\mathbb{H}_{0,1}^n$. We then use that, conditioned on $(\mathbb{H}_{0,1}^n, \mathbf{X}_{0,1}^n, A_1^L, \mathcal{J})$, $\hat{\mathbf{Y}}_1^n$ has the same distribution as

$$\begin{aligned} \bar{\mathbf{Y}}_k = &\sum_{\ell=1}^{L} A_\ell \bar{\mathbb{H}}_{\ell,\mathcal{J},k}\mathbf{X}_{\ell,\mathcal{J},k} \\ &+ \sum_{\ell=L+1}^{\infty} \bar{A}_\ell \bar{\mathbb{H}}_{\ell,\mathcal{J},k}\mathbf{X}_{\ell,\mathcal{J},k} + \bar{\mathbf{Z}}_k, \end{aligned} \tag{38}$$

where the sequences $\{\bar{\mathbb{H}}_{\ell,k}, k \in \mathbb{Z}\}$, $\{\bar{A}_\ell, \ell = 1, 2, \ldots\}$, and $\{\bar{\mathbf{Z}}_k, k \in \mathbb{Z}\}$ have the same distributions as $\{\mathbb{H}_{\ell,k}, k \in \mathbb{Z}\}$,

$\{A_\ell, \ell = 1, 2, \ldots\}$, and $\{\mathbf{Z}_k, k \in \mathbb{Z}\}$ but are independent of them. It follows that

$$I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \mid A_1^L, \mathcal{J}) \leq h(\hat{\mathbf{Y}}_1^n \mid A_1^L, \mathcal{J}) - h(\bar{\mathbf{Y}}_1^n \mid A_1^L, \mathcal{J}). \tag{39}$$

3) Pairing activity patterns: The interference term $\bar{\mathbf{Y}}_1^n$ may have inactive leading interferers due to bursty activity. To compare and cancel terms between $h(\hat{\mathbf{Y}}_1^n \mid A_1^L, \mathcal{J})$ and $h(\bar{\mathbf{Y}}_1^n \mid A_1^L, \mathcal{J})$ in (39), we introduce a bijective mapping f_L that pairs each activity pattern a_1^L with a "shifted" pattern $\tilde{a}_1^L = f_L(a_1^L)$. Specifically, let m denote the number of leading zeros before the first active interfering component in a_1^L. Then, $\tilde{a}_1^L$ shifts the non-zero part of a_1^L by m positions, ensuring that the first active component aligns with the desired signal. Formally, $\tilde{a}_1^L = [0_1^{m-1}, 1, b_1^{L-m}]$, where b_1^{L-m} captures the remaining activity pattern.

This construction allows $\bar{\mathbf{Y}}_1^n$ to be expressed as a shifted version of $\hat{\mathbf{Y}}_1^n$, with the shift determined by the position m of the first active interferer. Moreover, the mapping preserves the Hamming weight and hence the probability of each activity pattern, which justifies the rigorous application of the following variational bound to the differential entropy terms in (39).

4) Applying a variational bound based on the information inequality: We combine the two aligned entropy terms using the identity $h(U) - h(V) = h(U|V) - h(V|U)$. To upper-bound the resulting expression, we apply the following variational bound that follows from the nonnegativity of relative entropy:

Lemma 2: Let f and g be two arbitrary *probability density functions (pdfs)*. If $-\int f(x) \log f(x) \mathrm{d}x$ is finite, then $-\int f(x) \log g(x) \mathrm{d}x$ exists and

$$-\int f(x) \log f(x) \mathrm{d}x \leq -\int f(x) \log g(x) \mathrm{d}x. \tag{40}$$

Proof: See [46, Lemma 8.3.1]. ∎

A clever choice of $g(\cdot)$ allows us then to establish an upper bound on the difference of conditional differential entropies based on the expected power of the received signals.

5) Average and limits: The final result is obtained by averaging the derived upper bound over all possible realizations of the activity patterns A_1^L. Subsequently, we take the limits as the number of interfering cells L and the blocklength n tend to infinity, thereby establishing an upper bound (7) in Theorem 1.

We now provide the detailed steps of each of the 5 parts described in this outline.

B. Introducing $\mathcal{J}$-Interfering Cells

Using the compact notation S_1^n for the sequence $S_1, S_2, \ldots, S_n$, we can write the channel model (1) as

$$\mathbf{Y}_1^n = \mathbb{H}_{0,1}^n \mathbf{X}_{0,1}^n + \sum_{\ell=1}^{\infty} \mathbb{H}_{\ell,1}^n \mathbf{X}_{\ell,1}^n + \mathbf{Z}_1^n. \tag{41}$$

We begin by upper-bounding the mutual information as

$$\begin{aligned} I(\mathbf{X}_{0,1}^n; \mathbf{Y}_1^n) &\overset{(a)}{\leq} I(\mathcal{J}, \mathbf{X}_{0,1}^n; \mathbf{Y}_1^n) \\ &\overset{(b)}{=} I(\mathcal{J}; \mathbf{Y}_1^n) + I(\mathbf{X}_{0,1}^n; \mathbf{Y}_1^n \mid \mathcal{J}) \\ &\overset{(c)}{\leq} H(\mathcal{J}) + I(\mathbf{X}_{0,1}^n; \mathbf{Y}_1^n \mid \mathcal{J}) \end{aligned} \tag{42}$$

where (a) follows by giving the set of active users in the intended cell $\mathcal{J}$ as extra information; (b) follows by applying the chain rule of mutual information; and (c) follows by upper-bounding the mutual information $I(\mathcal{J}; \mathbf{Y}_1^n)$ by the entropy of the random variable $\mathcal{J}$. Note that the entropy $H(\mathcal{J})$ does not scale with the blocklength n.

We next upper-bound the conditional mutual information in (42) by giving extra information to the receiver. Specifically, for each $\mathcal{J}$-interfering cell, i.e., for every ℓ satisfying $A_\ell = 1$, we provide the extra information $\{\mathbb{H}_{\ell,\mathcal{J}^C,1}^n, \mathbf{X}_{\ell,\mathcal{J}^C,1}^n\}$. For the remaining cells, we provide the extra information $\{\mathbb{H}_{\ell,1}^n, \mathbf{X}_{\ell,1}^n\}$. This allows us to upper-bound the conditional mutual information as shown in (43) at the top of the next page, where $\hat{\mathbf{Y}}_1^n$ was defined in (37). In (43), step (a) follows because giving extra information increases mutual information; step (b) follows by applying the chain rule of mutual information; and step (c) follows since the first mutual information is equal to zero given the independence of $\mathbf{X}_{0,1}^n$, $\mathbb{H}_{\ell,1}^n$ and $\mathbf{X}_{\ell,1}^n$, $\ell > 0$, and since the second mutual information is equal to $I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \mid \mathcal{J})$.

The conditional mutual information in (43) can be further upper-bounded by giving the activity information A_1^L of the first L interfering cells, where L is an arbitrary integer that we will let go to infinity at the end of the proof. This yields

$$\begin{aligned} I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \mid \mathcal{J}) &\leq I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n, A_1^L \mid \mathcal{J}) \\ &\overset{(a)}{=} I(\mathbf{X}_{0,1}^n; A_1^L \mid \mathcal{J}) + I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \mid A_1^L, \mathcal{J}) \\ &\overset{(b)}{=} I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \mid A_1^L, \mathcal{J}), \end{aligned} \tag{44}$$

where (a) follows from the chain rule of mutual information and (b) follows because A_1^L, which is a function of $\mathbf{X}_{\ell,1}^n$, $\ell = 1, \ldots, L$, is independent of $\mathbf{X}_{0,1}^n$.

C. Decomposition into Differential Entropies

We write

$$I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \mid A_1^L, \mathcal{J}) = \sum_{\jmath} \Pr\{\mathcal{J} = \jmath\} I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \mid A_1^L, \mathcal{J} = \jmath), \tag{45}$$

where the sum is over all sets of active users $\jmath \subseteq \{1, \ldots, n_T\}$. When $\jmath$ is the empty set, $\mathbf{X}_{0,1}^n = \mathbf{0}_1^n$ and the mutual information $I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \mid A_1^L, \mathcal{J} = \jmath)$ is zero. We next upper-bound this mutual information when $\jmath$ is not the empty set. As we shall see, this bound depends only on the number of active users $|\jmath|$, i.e., on the cardinality of $\jmath$. Specifically, we

$$
\begin{aligned}
I(\mathbf{X}_{0,1}^n;\mathbf{Y}_1^n\,|\,\mathcal{J}) &\overset{(a)}{\le} I(\mathbf{X}_{0,1}^n;\mathbf{Y}_1^n,\{\mathbb{H}_{\ell,\mathcal{J}^C,1}^n,\ \mathbf{X}_{\ell,\mathcal{J}^C,1}^n\}_{\ell:A_\ell=1},\{\mathbb{H}_{\ell,1}^n,\ \mathbf{X}_{\ell,1}^n\}_{\ell:A_\ell=0}\,|\,\mathcal{J})\\
&\overset{(b)}{=} I(\mathbf{X}_{0,1}^n;\{\mathbb{H}_{\ell,\mathcal{J}^C,1}^n,\ \mathbf{X}_{\ell,\mathcal{J}^C,1}^n\}_{\ell:A_\ell=1},\{\mathbb{H}_{\ell,1}^n,\ \mathbf{X}_{\ell,1}^n\}_{\ell:A_\ell=0}\,|\,\mathcal{J})\\
&\quad+\ I(\mathbf{X}_{0,1}^n;\mathbf{Y}_1^n\,|\,\mathcal{J},\{\mathbb{H}_{\ell,\mathcal{J}^C,1}^n,\ \mathbf{X}_{\ell,\mathcal{J}^C,1}^n\}_{\ell:A_\ell=1},\{\mathbb{H}_{\ell,1}^n,\ \mathbf{X}_{\ell,1}^n\}_{\ell:A_\ell=0})\\
&\overset{(c)}{=} I(\mathbf{X}_{0,1}^n;\hat{\mathbf{Y}}_1^n\,|\,\mathcal{J})
\end{aligned}
\tag{43}
$$

write the mutual information in terms of differential entropies and upper-bound the resulting expression as

$$
\begin{aligned}
&I(\mathbf{X}_{0,1}^n;\hat{\mathbf{Y}}_1^n\,|\,A_1^L,\mathcal{J}=\jmath)\\
&= h(\hat{\mathbf{Y}}_1^n\,|\,A_1^L,\mathcal{J}=\jmath)-h(\hat{\mathbf{Y}}_1^n\,|\,\mathbf{X}_{0,1}^n,A_1^L,\mathcal{J}=\jmath)\\
&\overset{(a)}{\le} h(\hat{\mathbf{Y}}_1^n\,|\,A_1^L,\mathcal{J}=\jmath)-h(\hat{\mathbf{Y}}_1^n\,|\,\mathbb{H}_{0,1}^n,\mathbf{X}_{0,1}^n,A_1^L,\mathcal{J}=\jmath)\\
&\overset{(b)}{=} h(\hat{\mathbf{Y}}_1^n\,|\,A_1^L,\mathcal{J}=\jmath)\\
&\qquad-h(\hat{\mathbf{Y}}_1^n-\mathbb{H}_{0,1}^n\mathbf{X}_{0,1}^n\,|\,A_1^L,\mathcal{J}=\jmath),
\end{aligned}
\tag{46}
$$

where (a) follows because conditioning reduces entropy, and (b) follows because, conditioned on $(A_1^L,\mathcal{J})$, the sequence $\hat{\mathbf{Y}}_1^n-\mathbb{H}_{0,1}^n\mathbf{X}_{0,1}^n$ is independent of $(\mathbb{H}_{0,1}^n,\mathbf{X}_{0,1}^n)$.

For a given realization $A_1^L=a_1^L$, and setting $a_0\triangleq 1$ (because when $\jmath$ is non-empty, the intended cell is always active), we define

$$
\begin{aligned}
&\hat{\mathbf{Y}}_k\big(a_1^L,\jmath\big)\\
&\quad\triangleq\sum_{\ell=0}^{L}a_\ell\mathbb{H}_{\ell,\jmath,k}\mathbf{X}_{\ell,\jmath,k}+\sum_{\ell=L+1}^{\infty}A_\ell\mathbb{H}_{\ell,\jmath,k}\mathbf{X}_{\ell,\jmath,k}+\mathbf{Z}_k,
\end{aligned}
\tag{47}
$$

$$
\begin{aligned}
&\bar{\mathbf{Y}}_k\big(a_1^L,\jmath\big)\\
&\quad\triangleq\sum_{\ell=1}^{L}a_\ell\bar{\mathbb{H}}_{\ell,\jmath,k}\mathbf{X}_{\ell,\jmath,k}+\sum_{\ell=L+1}^{\infty}\bar{A}_\ell\bar{\mathbb{H}}_{\ell,\jmath,k}\mathbf{X}_{\ell,\jmath,k}+\bar{\mathbf{Z}}_k.
\end{aligned}
\tag{48}
$$

As in (38), the sequences $\{\bar{\mathbb{H}}_{\ell,k},k\in\mathbb{Z}\}$, $\{\bar{A}_\ell,\ell=1,2,\ldots\}$, and $\{\bar{\mathbf{Z}}_k,k\in\mathbb{Z}\}$ have the same distributions as $\{\mathbb{H}_{\ell,k},k\in\mathbb{Z}\}$, $\{A_\ell,\ell=1,2,\ldots\}$, and $\{\mathbf{Z}_k,k\in\mathbb{Z}\}$ but are independent of them. We recall that $\mathbb{H}_{\ell,\jmath,k}$ denotes the submatrix of $\mathbb{H}_{\ell,k}$ obtained by selecting the columns indexed by $j\in\jmath$, while $\mathbf{X}_{\ell,\jmath,k}$ denotes the corresponding subvector of $\mathbf{X}_{\ell,k}$ consisting of the components indexed by $j\in\jmath$.

We recall that the first sum in (47) starts at $\ell=0$, while the the first sum in (48) starts at $\ell=1$. The random variable $\hat{\mathbf{Y}}_k(a_1^L,\jmath)$ is precisely $\hat{\mathbf{Y}}_1^n$ conditioned on $A_1^L=a_1^L$ and $\mathcal{J}=\jmath$. Moreover, conditioned on $A_1^L=a_1^L$ and $\mathcal{J}=\jmath$, $\hat{\mathbf{Y}}_1^n-\mathbb{H}_{0,1}^n\mathbf{X}_{0,1}^n$ has the same distribution as $\bar{\mathbf{Y}}_k\left(a_1^L,\jmath\right)$. Thus, the differential entropy terms in (46) can be written as

$$
\begin{aligned}
&h(\hat{\mathbf{Y}}_1^n\,|\,A_1^L,\mathcal{J}=\jmath)\\
&\quad=\sum_{a_1^L\in\mathcal{B}_L}\Pr\{A_1^L=a_1^L\}\,h\big(\hat{\mathbf{Y}}_1^n(a_1^L,\jmath)\big),
\end{aligned}
\tag{49}
$$

$$
\begin{aligned}
&h(\hat{\mathbf{Y}}_1^n-\mathbb{H}_{0,1}^n\mathbf{X}_{0,1}^n\,|\,A_1^L,\mathcal{J}=\jmath)\\
&\quad=\sum_{\tilde{a}_1^L\in\mathcal{B}_L}\Pr\{A_1^L=\tilde{a}_1^L\}\,h\big(\bar{\mathbf{Y}}_1^n(\tilde{a}_1^L,\jmath)\big),
\end{aligned}
\tag{50}
$$

where $\mathcal{B}_L\triangleq\{0,1\}^L$ denotes the set of all binary sequences of length L.

D. Pairing Activity Patterns

We pair the sequences a_1^L and $\tilde{a}_1^L$ appearing in the sums in (49) and (50) through a deterministic bijective mapping f_L. This construction ensures that each paired sequence has the same Hamming weight and that the two sequences differ only by a shift in the position of their leading nonzero entry. To this end, we first partition the set $\mathcal{B}_L$ of all binary sequences of length L according to the position of their leading 1. For $m=1,\ldots,L+1$, we define

$$
\mathcal{B}_L(m)\triangleq\begin{cases}\{a_1^L\in\mathcal{B}_L:a_1^m=[0_1^{m-1},1]\}, & 1\le m\le L,\\ \{0_1^L\}, & m=L+1.\end{cases}
\tag{51}
$$

Thus, $\mathcal{B}_L(m)$ contains all sequences whose first nonzero entry occurs at position m. The sets $\{\mathcal{B}_L(m)\}_{m=1}^{L+1}$ are disjoint and form a partition of $\mathcal{B}_L$.

Proposition 5: Let $f_L\colon\mathcal{B}_L\to\mathcal{B}_L$ be the mapping defined by Algorithm 1. Then, f_L is bijective and satisfies:

1) For every $a_1^L\in\mathcal{B}_L$, the image $\tilde{a}_1^L=f_L(a_1^L)$ belongs to $\mathcal{B}_L(m)$ for an m uniquely determined by a_1^L and has the form
$$\tilde{a}_1^L=[0_1^{m-1},\,1,\,a_1^{L-m}].\tag{52}$$
2) The mapping preserves the ℓ_1-norm, i.e.,
$$\|\tilde{a}_1^L\|_1=\|a_1^L\|_1.\tag{53}$$

Proof: The mapping f_L is constructed explicitly by Algorithm 1, which repositions the right-most active entry of a_1^L to the first active position while preserving the relative order of the remaining entries. By construction, the output $\tilde{a}_1^L$ has the form (52), hence $\tilde{a}_1^L\in\mathcal{B}_L(m)$, and preserves identical Hamming weight. Bijectivity follows from Algorithm 2, which inverts this operation by uniquely recovering a_1^L from $\tilde{a}_1^L$. Hence, f_L is one-to-one and onto. ∎

Using the bijective mapping f_L from Proposition 5, we can re-index the sum on the RHS of (50) as

$$
\begin{aligned}
&\sum_{\tilde{a}_1^L\in\mathcal{B}_L}\Pr\{A_1^L=\tilde{a}_1^L\}\,h\big(\bar{\mathbf{Y}}_1^n(\tilde{a}_1^L,\jmath)\big)\\
&\quad=\sum_{a_1^L\in\mathcal{B}_L}\Pr\{A_1^L=f_L(a_1^L)\}\,h\big(\bar{\mathbf{Y}}_1^n(f_L(a_1^L),\jmath)\big),
\end{aligned}
\tag{54}
$$

where the equality follows from the bijectivity of f_L. Substituting (49) and (50) back in (46), and using (54), we obtain

$$
\begin{aligned}
&I(\mathbf{X}_{0,1}^n;\hat{\mathbf{Y}}_1^n\,|\,A_1^L,\mathcal{J}=\jmath)\\
&\quad\le\sum_{a_1^L\in\mathcal{B}_L}\Pr\{A_1^L=a_1^L\}\Big[h\big(\hat{\mathbf{Y}}_1^n(a_1^L,\jmath)\big)\\
&\qquad\qquad-h\big(\bar{\mathbf{Y}}_1^n(f_L(a_1^L),\jmath)\big)\Big],
\end{aligned}
\tag{55}
$$

Algorithm 1: Mapping between $\mathbf{b}$ and $\tilde{\mathbf{b}}$.

Data: Binary sequence $\mathbf{b}$ of length L with Hamming weight $\|\mathbf{b}\|_1 = W$
Result: Binary sequence $\tilde{\mathbf{b}} = [0_1^{m-1}, 1, b_1^{L-m}]$ of length L and Hamming weight $\|\tilde{\mathbf{b}}\|_1 = W$.
if $\mathbf{b} = 0_1^L$ **then**
  $\tilde{\mathbf{b}} = \mathbf{b}$
else
  $i \leftarrow$ take the position of the right-most 1 in sequence $\mathbf{b}$
  $m \leftarrow L - i + 1$ (length of b_i^L)
  $\tilde{\mathbf{b}} = [0_1^{m-1}, 1, b_1^{L-m}]$
end

Algorithm 2: Mapping between $\tilde{\mathbf{b}}$ and $\mathbf{b}$.

Data: Binary sequence $\tilde{\mathbf{b}}$ of length L with Hamming weight $\|\tilde{\mathbf{b}}\|_1 = W$
Result: Binary sequence $\mathbf{b} = [b_{m+1}^L, 1, 0_1^{m-1}]$ of length L and Hamming weight $\|\mathbf{b}\|_1 = W$.
if $\tilde{\mathbf{b}} = 0_1^L$ **then**
  $\mathbf{b} = \tilde{\mathbf{b}}$
else
  $i \leftarrow$ take the position of the left-most 1 in sequence $\mathbf{b}$
  $m \leftarrow i$: (length of b_1^i)
  $\mathbf{b} = [\tilde{b}_{m+1}^L, 1, 0_1^{m-1}]$
end

where we used that the activity indicators $\{A_k\}_{k=1}^L$ are i.i.d. and that f_L preserves the Hamming weight, so $\Pr\{A_1^L = f_L(a_1^L)\} = \Pr\{A_1^L = a_1^L\}$.

Next, recalling that the sets $\{\mathcal{B}_L(m)\}_{m=1}^{L+1}$ form a partition of $\mathcal{B}_L$, we group the terms in (55) according to the index m associated with $f_L(a_1^L)$:

$$\begin{aligned} &I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \,|\, A_1^L, \mathcal{J} = \jmath) \\ &\le \sum_{m=1}^{L+1} \sum_{a_1^L : f_L(a_1^L) \in \mathcal{B}_L(m)} \Pr\{A_1^L = a_1^L\} \Big[h\big(\hat{\mathbf{Y}}_1^n(a_1^L, \jmath)\big) \\ &\qquad - h\big(\bar{\mathbf{Y}}_1^n(f_L(a_1^L), \jmath)\big) \Big]. \quad (56) \end{aligned}$$

By Assumption A3, the distribution of the random variables $\mathbf{X}_{\ell,1}^n$ does not depend on ℓ. Consequently, for any $f_L(a_1^L) \in \mathcal{B}_L(m)$, the random variable $\bar{\mathbf{Y}}_k(f_L(a_1^L), \jmath)$ has the same distribution as

$$\begin{aligned} \bar{\mathbf{Y}}_k(a_1^L, \jmath, m) \triangleq &\sum_{\ell=0}^{L-m} a_\ell \bar{\mathbb{H}}_{\ell+m,\jmath,k} \mathbf{X}_{\ell,\jmath,k} \\ &+ \sum_{\ell=L-m+1}^{\infty} \bar{A}_\ell \bar{\mathbb{H}}_{\ell+m,\jmath,k} \mathbf{X}_{\ell,\jmath,k} + \bar{\mathbf{Z}}_k. \quad (57) \end{aligned}$$

Comparing (47) and (57), we observe that, as $L \to \infty$, the two random variables differ only by a finite shift $\ell \mapsto \ell + m$ in the fading matrices. This observation will be exploited next to bound the differential entropies in (56).

E. Variational Bound Based on the Information Inequality

We focus on the bracketed term in (56). Using the equivalence in distribution of $\bar{\mathbf{Y}}_k(f_L(a_1^L), \jmath)$ and $\bar{\mathbf{Y}}_k(a_1^L, \jmath, m)$ for $f_L(a_1^L) \in \mathcal{B}_L(m)$, and using the identity $h(U) - h(V) = h(U|V) - h(V|U)$, it follows that

$$\begin{aligned} &h\big(\hat{\mathbf{Y}}_1^n(a_1^L, \jmath)\big) - h\big(\bar{\mathbf{Y}}_1^n(f_L(a_1^L), \jmath)\big) \\ &= h\big(\hat{\mathbf{Y}}_1^n(a_1^L, \jmath)\big) - h\big(\bar{\mathbf{Y}}_1^n(a_1^L, \jmath, m)\big) \\ &= h\big(\hat{\mathbf{Y}}_1^n(a_1^L, \jmath) \,\big|\, \bar{\mathbf{Y}}_1^n(a_1^L, \jmath, m)\big) \\ &\quad - h\big(\bar{\mathbf{Y}}_1^n(a_1^L, \jmath, m) \,\big|\, \hat{\mathbf{Y}}_1^n(a_1^L, \jmath)\big) \\ &\le \sum_{k=1}^{n} \big[h\big(\hat{\mathbf{Y}}_k(a_1^L, \jmath) \,\big|\, \bar{\mathbf{Y}}_k(a_1^L, \jmath, m)\big) \\ &\qquad - h\big(\bar{\mathbf{Y}}_k(a_1^L, \jmath, m) \,\big|\, \bar{\mathbf{Y}}_1^{k-1}(a_1^L, \jmath, m), \hat{\mathbf{Y}}_1^n(a_1^L, \jmath)\big)\big], \quad (58) \end{aligned}$$

where the inequality follows from the chain rule and because conditioning reduces entropy.

To upper-bound (58), we next apply Lemma 2 with a cleverly chosen $g(\cdot)$. Since we bound the conditional entropy given $\bar{\mathbf{Y}}_k(a_1^L, \jmath, m)$, we can choose a conditional pdf that depends on $\bar{\mathbf{Y}}_k(a_1^L, \jmath, m)$. Specifically, let $f_{\hat{\mathbf{Y}}_k|\bar{\mathbf{Y}}_k}$ denote the true conditional pdf of $\hat{\mathbf{Y}}_k(a_1^L, \jmath)$ given $\bar{\mathbf{Y}}_k(a_1^L, \jmath, m)$. Lemma 2 allows us to upper-bound the conditional differential entropy of $\hat{\mathbf{Y}}_k(a_1^L, \jmath)$ given $\bar{\mathbf{Y}}_k(a_1^L, \jmath, m)$ by replacing $f_{\hat{\mathbf{Y}}_k|\bar{\mathbf{Y}}_k}$ by an auxiliary pdf $g_{\hat{\mathbf{Y}}_k|\bar{\mathbf{Y}}_k}$. For every $\bar{\mathbf{Y}}_k(a_1^L, \jmath, m) = \bar{\mathbf{y}}_k$, we choose

$$g_{\hat{\mathbf{Y}}_k|\bar{\mathbf{Y}}_k}(\hat{\mathbf{y}}_k|\bar{\mathbf{y}}_k) = \frac{n_R \sqrt{\beta} \Gamma(n_R)}{\pi^{n_R+1} \|\hat{\mathbf{y}}_k\|_2^{n_R}} \frac{1}{1 + \beta \|\hat{\mathbf{y}}_k\|_2^{2n_R}} \quad (59)$$

with $\beta = 1/\|\bar{\mathbf{y}}_k\|_2^{2n_R}$. This is the density of a circularly-symmetric complex random variable whose magnitude is Cauchy distributed. A similar pdf has been used in [47] to analyze frequency-dispersive fading channels.

Applying Lemma 2 with the auxiliary distribution $g_{\hat{\mathbf{Y}}_k|\bar{\mathbf{Y}}_k}(\cdot)$ from (59), we obtain

$$\begin{aligned} &h\big(\hat{\mathbf{Y}}_k(a_1^L, \jmath) \,\big|\, \bar{\mathbf{Y}}_k(a_1^L, \jmath, m)\big) \\ &\quad = -\mathsf{E}\Big[\log f_{\hat{\mathbf{Y}}_k|\bar{\mathbf{Y}}_k}\big(\hat{\mathbf{Y}}_k(a_1^L, \jmath) \,|\, \bar{\mathbf{Y}}_k(a_1^L, \jmath, m)\big)\Big] \\ &\quad \overset{(a)}{\le} -\mathsf{E}\Big[\log g_{\hat{\mathbf{Y}}_k|\bar{\mathbf{Y}}_k}\big(\hat{\mathbf{Y}}_k(a_1^L, \jmath) \,|\, \bar{\mathbf{Y}}_k(a_1^L, \jmath, m)\big)\Big] \\ &\quad \overset{(b)}{=} (n_R + 1)\log\pi - \log\big(n_R \Gamma(n_R)\big) \\ &\qquad + \frac{n_R}{2} \mathsf{E}\Big[\log \|\hat{\mathbf{Y}}_k(a_1^L, \jmath)\|_2^2\Big] \\ &\qquad - \frac{n_R}{2} \mathsf{E}\big[\log \|\bar{\mathbf{Y}}_k(a_1^L, \jmath, m)\|_2^2\big] \\ &\qquad + \mathsf{E}\Big[\log\big(\|\hat{\mathbf{Y}}_k(a_1^L, \jmath)\|_2^{2n_R} + \|\bar{\mathbf{Y}}_k(a_1^L, \jmath, m)\|_2^{2n_R}\big)\Big] \\ &\quad \overset{(c)}{\le} (n_R + 1)\log\pi - \log\big(n_R \Gamma(n_R)\big) \\ &\qquad + \frac{n_R}{2} \mathsf{E}\Big[\log \|\hat{\mathbf{Y}}_k(a_1^L, \jmath)\|_2^2\Big] \\ &\qquad - \frac{n_R}{2} \mathsf{E}\big[\log \|\bar{\mathbf{Y}}_k(a_1^L, \jmath, m)\|_2^2\big] \\ &\qquad + n_R \mathsf{E}\Big[\log\big(\|\hat{\mathbf{Y}}_k(a_1^L, \jmath)\|_2^2 + \|\bar{\mathbf{Y}}_k(a_1^L, \jmath, m)\|_2^2\big)\Big], \quad (60) \end{aligned}$$

where (a) follows from Lemma 2; in (b) we substituted the auxiliary distribution $g_{\hat{\mathbf{Y}}_k|\bar{\mathbf{Y}}_k}(\cdot)$ in (59) with $\beta = 1/\|\bar{\mathbf{y}}_k\|_2^{2n_R}$; and (c) follows since $a^{n_R} + b^{n_R} \le (a+b)^{n_R}$ for $a, b \ge 0$.

Next, we consider the second conditional entropy in (58). By conditioning on $\{\mathbf{X}_{\ell,\jmath,k}\}_{\ell=1}^{\infty}$ and $\bar{A}_{L-m+1}^{\infty}$, the random variable $\bar{\mathbf{Y}}_k(a_1^L,\jmath,m)$ is independent of $(\bar{\mathbf{Y}}_1^{k-1}(a_1^L,\jmath,m),\hat{\mathbf{Y}}_1^n(a_1^L,\jmath))$ and has a Gaussian distribution. Hence,

$$\begin{aligned}
&h\big(\bar{\mathbf{Y}}_k(a_1^L,\jmath,m)\,\big|\,\bar{\mathbf{Y}}_1^{k-1}(a_1^L,\jmath,m),\hat{\mathbf{Y}}_1^n(a_1^L,\jmath)\big)\\
&\geq h\big(\bar{\mathbf{Y}}_k(a_1^L,\jmath,m)\,\big|\,\{\mathbf{X}_{\ell,\jmath,k}\}_{\ell=1}^{\infty},\bar{A}_{L-m+1}^{\infty}\big)\\
&= n_R\log(\pi e)+n_R\mathsf{E}\big[\log\bar{\mathsf{K}}(a_1^{L-m},\bar{A}_{L-m+1}^{\infty},\mathbf{X}_{0,\jmath,k}^{\infty})\big]\,, \quad (61)
\end{aligned}$$

where

$$\begin{aligned}
&\bar{\mathsf{K}}(a_1^{L-m},\bar{A}_{L-m+1}^{\infty},\mathbf{X}_{0,\jmath,k}^{\infty})\\
&\quad\triangleq\sum_{\ell=0}^{L-m}a_\ell\alpha_{\ell+m}\|\mathbf{X}_{\ell,\jmath,k}\|_2^2\\
&\qquad+\sum_{\ell=L-m+1}^{\infty}\bar{A}_\ell\alpha_{\ell+m}\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sigma^2. \quad (62)
\end{aligned}$$

The inequality in (61) follows because conditioning reduces entropy.

Using (60) and (61), we obtain that (58) can be upper-bounded by

$$\begin{aligned}
&h\big(\hat{\mathbf{Y}}_1^n(a_1^L,\jmath)\big)-h\big(\bar{\mathbf{Y}}_1^n(f_L(a_1^L),\jmath)\big)\\
&\leq\sum_{k=1}^{n}\Bigg(\log\frac{\pi}{n_R\Gamma(n_R)}-n_R\log e\\
&\qquad+\frac{n_R}{2}\mathsf{E}\Big[\log\|\hat{\mathbf{Y}}_k(a_1^L,\jmath)\|_2^2\Big]\\
&\qquad-\frac{n_R}{2}\mathsf{E}\big[\log\|\bar{\mathbf{Y}}_k(a_1^L,\jmath,m)\|_2^2\big]\\
&\qquad+n_R\mathsf{E}\Big[\log(\|\hat{\mathbf{Y}}_k(a_1^L,\jmath)\|_2^2+\|\bar{\mathbf{Y}}_k(a_1^L,\jmath,m)\|_2^2)\Big]\\
&\qquad-n_R\mathsf{E}\Big[\log\bar{\mathsf{K}}(a_1^{L-m},\bar{A}_{L-m+1}^{\infty},\mathbf{X}_{0,\jmath,k}^{\infty})\Big]\Bigg). \quad (63)
\end{aligned}$$

To upper-bound the the third and fourth term on the RHS of (63), we note that, conditioned on $\mathbf{X}_{\ell,\jmath,k}=\mathbf{x}_\ell$, $A_\ell=a_\ell$, and $\bar{A}_\ell=\bar{a}_\ell$, $\ell=0,1,\ldots$, both $\|\hat{\mathbf{Y}}_k(a_1^L,\jmath)\|_2^2$ and $\|\bar{\mathbf{Y}}_k(a_1^L,\jmath,m)\|_2^2$ have a chi-square distribution with $2n_R$ degrees of freedom. Using [48, Eq. 4.352], we thus obtain that

$$\begin{aligned}
&\mathsf{E}\Big[\log\|\hat{\mathbf{Y}}_k(a_1^L,\jmath)\|_2^2\Big]-\mathsf{E}\big[\log\|\bar{\mathbf{Y}}_k(a_1^L,\jmath,m)\|_2^2\big]\\
&\qquad=\mathsf{E}\left[\log\left(\frac{\mathsf{K}(a_1^L,A_{L+1}^{\infty},\mathbf{X}_{0,\jmath,k}^{\infty})}{\bar{\mathsf{K}}(a_1^{L-m},\bar{A}_{L-m+1}^{\infty},\mathbf{X}_{0,\jmath,k}^{\infty})}\right)\right], \quad (64)
\end{aligned}$$

where

$$\begin{aligned}
&\mathsf{K}(a_1^L,A_{L+1}^{\infty},\mathbf{X}_{0,\jmath,k}^{\infty})\\
&\quad\triangleq\sum_{\ell=0}^{L}a_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sum_{\ell=L+1}^{\infty}A_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sigma^2. \quad (65)
\end{aligned}$$

The theorem's assumption (6) implies that

$$\alpha_{\ell+m}\geq\frac{\rho^{m-1}}{\eta_{\max}}\alpha_\ell,\quad \ell=0,1,\ldots \quad (66)$$

This allows us to lower-bound

$$\begin{aligned}
&\bar{\mathsf{K}}(a_1^{L-m},\bar{A}_{L-m+1}^{\infty},\mathbf{X}_{0,\jmath,k}^{\infty})\\
&\quad\geq\frac{\rho^{m-1}}{\eta_{\max}}\sum_{\ell=0}^{L-m}a_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2\\
&\qquad+\frac{\rho^{m-1}}{\eta_{\max}}\sum_{\ell=L-m+1}^{\infty}\bar{A}_\ell\alpha_{\ell+m}\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sigma^2\\
&\quad\geq\frac{\rho^{m-1}}{\eta_{\max}}\sum_{\ell=0}^{L-m}a_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2\\
&\qquad+\frac{\rho^{m-1}}{\eta_{\max}}\sum_{\ell=L-m+1}^{\infty}\bar{A}_\ell\alpha_{\ell+m}\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\frac{\rho^{m-1}}{\eta_{\max}}\sigma^2, \quad (67)
\end{aligned}$$

where the second inequality follows because $\rho^{m-1}/\eta_{\max}\leq 1$, since $\eta_{\max}\geq 1/\rho$ by definition and $0<\rho<1$ by assumption. Substituting the resulting bound in (64), we thus obtain (68) at the top of the next page, where

$$\begin{aligned}
&\zeta(a_{L-m+1}^L,A_{L+1}^{\infty},\mathbf{X}_{L-m+1,\jmath,k}^{\infty})\\
&\quad\triangleq\sum_{\ell=L-m+1}^{L}\frac{a_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2}{\sigma^2}+\sum_{\ell=L+1}^{\infty}\frac{\alpha_\ell A_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2}{\sigma^2}. \quad (69)
\end{aligned}$$

In (68), (a) follows by writing the fraction as the sum of two fractions, the first with numerator $\sum_{\ell=0}^{L-m}a_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sigma^2$ and the second with the remaining terms, and by lower-bounding the denominator of the first fraction by $\sum_{\ell=0}^{L-m}a_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sigma^2$ and the denominator of the second fraction by σ^2; (b) follows because $\log(1+x)\leq x$, $x\geq 0$.

To upper-bound the fifth term on the RHS of (63), we use the law of iterated expectations and upper-bound the conditional expectation using Jensen's inequality:

$$\begin{aligned}
&\mathsf{E}\Big[\log\Big(\|\hat{\mathbf{Y}}_k(a_1^L,\jmath)\|_2^2+\|\bar{\mathbf{Y}}_k(a_1^L,\jmath,m)\|_2^2\Big)\Big]\\
&\leq\mathsf{E}\Big[\log\Big(\mathsf{E}\Big[\|\hat{\mathbf{Y}}_k(a_1^L,\jmath)\|_2^2+\|\bar{\mathbf{Y}}_k(a_1^L,\jmath,m)\|_2^2\,\Big|\\
&\qquad\qquad\{\mathbf{X}_{\ell,\jmath,k}\}_{\ell=1}^{\infty},A_1^{\infty},\bar{A}_1^{\infty}\Big]\Big)\Big]\\
&=\mathsf{E}\Big[\log\Big(n_R\mathsf{K}(a_1^L,A_{L+1}^{\infty},\mathbf{X}_{0,\jmath,k}^{\infty})\\
&\qquad\qquad+n_R\bar{\mathsf{K}}(a_1^{L-m},\bar{A}_{L-m+1}^{\infty},\mathbf{X}_{0,\jmath,k}^{\infty})\Big)\Big]. \quad (70)
\end{aligned}$$

We next use again (66) to upper-bound

$$\begin{aligned}
&\mathsf{K}(a_1^L,A_{L+1}^{\infty},\mathbf{X}_{0,\jmath,k}^{\infty})\\
&\leq\frac{\eta_{\max}}{\rho^{m-1}}\sum_{\ell=0}^{L-m}a_\ell\alpha_{\ell+m}\|\mathbf{X}_{\ell,\jmath,k}\|_2^2\\
&\qquad+\sum_{\ell=L-m+1}^{L}a_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sum_{\ell=L+1}^{\infty}A_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sigma^2\\
&\leq\frac{\eta_{\max}}{\rho^{m-1}}\sum_{\ell=0}^{L-m}a_\ell\alpha_{\ell+m}\|\mathbf{X}_{\ell,\jmath,k}\|_2^2\\
&\qquad+\frac{\eta_{\max}}{\rho^{m-1}}\sum_{\ell=L-m+1}^{\infty}\bar{A}_\ell\alpha_{\ell+m}\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2\\
&\qquad+\sum_{\ell=L-m+1}^{L}a_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sum_{\ell=L+1}^{\infty}A_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2
\end{aligned}$$

$$\begin{aligned}
&\mathsf{E}\left[\log\|\mathbf{Y}_k(a_1^L,\jmath)\|_2^2\right]-\mathsf{E}\left[\log\|\bar{\mathbf{Y}}_k(a_1^L,\jmath,m)\|_2^2\right]\\
&\quad\le \mathsf{E}\left[\log\left(\frac{\sum_{\ell=0}^{L}a_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sum_{\ell=L+1}^{\infty}A_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sigma^2}{\sum_{\ell=0}^{L-m}a_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sum_{\ell=L-m+1}^{\infty}\bar{A}_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sigma^2}\right)\right]+\log\left(\frac{\eta_{\max}}{\rho^{m-1}}\right)\\
&\quad\overset{(a)}{\le} \mathsf{E}\left[\log\left(1+\frac{\sum_{\ell=L-m+1}^{L}a_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2+\sum_{\ell=L+1}^{\infty}A_\ell\alpha_\ell\|\mathbf{X}_{\ell,\jmath,k}\|_2^2}{\sigma^2}\right)\right]+\log\left(\frac{\eta_{\max}}{\rho^{m-1}}\right)\\
&\quad\overset{(b)}{\le} \log\left(\frac{\eta_{\max}}{\rho^{m-1}}\right)+\mathsf{E}\left[\zeta(a_{L-m+1}^L,A_{L+1}^\infty,\mathbf{X}_{L-m+1,\jmath,k}^\infty)\right] \qquad (68)
\end{aligned}$$

$$\begin{aligned}
&=\frac{\eta_{\max}}{\rho^{m-1}}\bar{\mathsf{K}}(a_1^{L-m},\bar{A}_{L-m+1}^\infty,\mathbf{X}_{0,k}^\infty)\\
&\quad+\sigma^2\zeta(a_{L-m+1}^L,A_{L+1}^\infty,\mathbf{X}_{L-m+1,\jmath,k}^\infty) \qquad (71)
\end{aligned}$$

where the second inequality follows because the second sum on the second line is nonnegative and because $\eta_{\max}/\rho^{m-1}\ge 1$. It follows that (70) can be further upper-bounded as

$$\begin{aligned}
&\mathsf{E}\left[\log\left(\|\hat{\mathbf{Y}}_k(a_1^L,\jmath)\|_2^2+\|\bar{\mathbf{Y}}_k(a_1^L,\jmath,m)\|_2^2\right)\right]\\
&\le\log\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right)+\log n_R\\
&\quad+\mathsf{E}\left[\log\bar{\mathsf{K}}(a_1^{L-m},\bar{A}_{L-m+1}^\infty,\mathbf{X}_{0,k}^\infty)\right]\\
&\quad+\mathsf{E}\left[\log\left(1+\frac{\sigma^2\zeta(a_{L-m+1}^L,A_{L+1}^\infty,\mathbf{X}_{L-m+1,\jmath,k}^\infty)}{\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right)\bar{\mathsf{K}}(a_1^{L-m},\bar{A}_{L-m+1}^\infty,\mathbf{X}_{0,k}^\infty)}\right)\right]\\
&\le\log\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right)+\log n_R\\
&\quad+\mathsf{E}\left[\log\bar{\mathsf{K}}(a_1^{L-m},\bar{A}_{L-m+1}^\infty,\mathbf{X}_{0,k}^\infty)\right]\\
&\quad+\mathsf{E}\left[\zeta(a_{L-m+1}^L,A_{L+1}^\infty,\mathbf{X}_{L-m+1,\jmath,k}^\infty)\right], \qquad (72)
\end{aligned}$$

where the last step follows by bounding $\log(1+x)\le x$, $x\ge 0$ and $(1+\eta_{\max}/\rho^{m-1})\bar{\mathsf{K}}(a_1^{L-m},\bar{A}_{L-m+1}^\infty,\mathbf{X}_{0,k}^\infty)\ge\sigma^2$.

Combining (68) and (72) with (63), we obtain

$$\begin{aligned}
&h\big(\hat{\mathbf{Y}}_1^n(a_1^L,\jmath)\big)-h\big(\bar{\mathbf{Y}}_1^n(f_L(a_1^L),\jmath)\big)\\
&\le n\Big[n_R(m-1)\log\left(\rho^{-\frac{3}{2}}\right)+\log\frac{\pi}{n_R\Gamma(n_R)}+n_R\log\frac{n_R}{e}\\
&\qquad+\frac{n_R}{2}\log(\eta_{\max})+n_R\log(\rho^{m-1}+\eta_{\max})\Big]\\
&\quad+\frac{3}{2}n_R\sum_{k=1}^{n}\mathsf{E}\left[\zeta(a_{L-m+1}^L,A_{L+1}^\infty,\mathbf{X}_{L-m+1,\jmath,k}^\infty)\right]\\
&\le n\Big[n_R(m-1)\log\left(\rho^{-\frac{3}{2}}\right)+\log\frac{\pi}{n_R\Gamma(n_R)}+n_R\log\frac{n_R}{e}\\
&\qquad+\frac{n_R}{2}\log(\eta_{\max})+n_R\log(1+\eta_{\max})+\frac{3}{2}n_R\bar{\zeta}_{L,m}\Big], \qquad (73)
\end{aligned}$$

where in the second step we upper-bounded ρ^{m-1} in the fifth term by 1, and we used the power constraint (5) together with $\mathsf{E}[A_\ell]=p_{|\jmath|}$ to upper-bound

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left[\zeta(a_{L-m+1}^L,A_{L+1}^\infty,\mathbf{X}_{L-m+1,\jmath,k}^\infty)\right]\\
&\qquad\le\sum_{\ell=L-m+1}^{L}\frac{\alpha_\ell n_T\mathsf{P}}{\sigma^2}+\sum_{\ell=L+1}^{\infty}\frac{\alpha_\ell p_{|\jmath|}n_T\mathsf{P}}{\sigma^2}\triangleq\bar{\zeta}_{L,m}. \qquad (74)
\end{aligned}$$

Back to (56), we note that the upper bound (73) depends on the position m of the leading 1 of $f_L(a_1^L)$, but not on the specific activity pattern a_1^L. The probability that a length-L binary sequence has the leading 1 in the m-th position is equal to

$$\begin{aligned}
\sum_{a_1^L:f_L(a_1^L)\in\mathcal{B}_L(m)}\Pr\{A_1^L=a_1^L\}&=\Pr\{f_L(A_1^L)\in\mathcal{B}_L(m)\}\\
&=p_{|\jmath|}(1-p_{|\jmath|})^{m-1}, \qquad (75)
\end{aligned}$$

since $A_1,\dots,A_L$ follow an independent Bernoulli distribution with parameter $p_{|\jmath|}$, and since $f_L(A_1^L)$ has the same number of ones as A_1^L. Therefore, using (73)–(75) in (56), we obtain that

$$\begin{aligned}
&\frac{1}{n}I(\mathbf{X}_{0,1}^n;\hat{\mathbf{Y}}_1^n\mid A_1^L,\mathcal{J}=\jmath)\\
&\quad\le\sum_{m=1}^{L+1}p_{|\jmath|}(1-p_{|\jmath|})^{m-1}\Big[n_R(m-1)\log\left(\rho^{-\frac{3}{2}}\right)\\
&\qquad+\log\frac{\pi}{n_R\Gamma(n_R)}+n_R\log\frac{n_R}{e}\\
&\qquad+\frac{n_R}{2}\log(\eta_{\max})+n_R\log(1+\eta_{\max})\Big]\\
&\quad+\frac{3}{2}n_R\sum_{m=1}^{L+1}p_{|\jmath|}(1-p_{|\jmath|})^{m-1}\bar{\zeta}_{L,m}. \qquad (76)
\end{aligned}$$

F. Average and Limits

Using that

$$\sum_{m=1}^{\infty}p_{|\jmath|}(1-p_{|\jmath|})^{m-1}=1 \qquad (77)$$

and [48, Eq. 0.231]

$$\sum_{m=1}^{\infty}p_{|\jmath|}(1-p_{|\jmath|})^{m-1}(m-1)=\frac{1-p_{|\jmath|}}{p_{|\jmath|}} \qquad (78)$$

the first sum on the RHS of (76) converges to

$$\begin{aligned}
\Upsilon_{|\jmath|}&\triangleq n_R\tfrac{1-p_{|\jmath|}}{p_{|\jmath|}}\log\left(\rho^{-\frac{3}{2}}\right)+\log\frac{\pi}{n_R\Gamma(n_R)}+n_R\log\frac{n_R}{e}\\
&\quad+\frac{n_R}{2}\log\eta_{\max}+n_R\log\left(1+\eta_{\max}\right) \qquad (79)
\end{aligned}$$

as $L\to\infty$.

We next show that the second sum on the RHS of (76) vanishes as L tends to infinity. Indeed, after some algebraic manipulations, we obtain that

$$\begin{aligned}
&\sum_{m=1}^{L+1} p_{|\jmath|}(1-p_{|\jmath|})^{m-1}\bar{\zeta}_{L,m} \\
&\quad= \frac{n_T\mathsf{P}}{\sigma^2}\sum_{m=1}^{L+1} p_{|\jmath|}(1-p_{|\jmath|})^{m-1}\sum_{\ell=L-m+1}^{L}\alpha_\ell \\
&\qquad+ \frac{n_T\mathsf{P}}{\sigma^2}\sum_{m=1}^{L+1} p_{|\jmath|}^2(1-p_{|\jmath|})^{m-1}\sum_{\ell=L+1}^{\infty}\alpha_\ell \\
&\quad\overset{(a)}{=} \frac{n_T\mathsf{P}}{\sigma^2}\sum_{\ell=0}^{L}\alpha_\ell\sum_{m=L-\ell+1}^{L+1} p_{|\jmath|}(1-p_{|\jmath|})^{m-1} \\
&\qquad+ \frac{n_T\mathsf{P}}{\sigma^2}p_{|\jmath|}\sum_{\ell=L+1}^{\infty}\alpha_\ell \\
&\quad\overset{(b)}{\leq} \frac{n_T\mathsf{P}}{\sigma^2}\sum_{\ell=0}^{L}\alpha_\ell(1-p_{|\jmath|})^{L-\ell} + \frac{n_T\mathsf{P}}{\sigma^2}p_{|\jmath|}\sum_{\ell=L+1}^{\infty}\alpha_\ell, \quad (80)
\end{aligned}$$

where (a) follows by changing the order of the sums in the first term, and by evaluating the sum over m in the second term using (77); to obtain (b), we upper-bound the first term by letting the sum over m run up to infinity, and by then evaluating the sum of a geometric series.

The second term on the RHS of (80) vanishes as $L\to\infty$ since $\sum_{\ell=0}^{\infty}\alpha_\ell<\infty$ by assumption. To bound the first term on the RHS of (80), we fix an arbitrary ν and bound the sum as

$$\begin{aligned}
&\sum_{\ell=0}^{L}\alpha_\ell(1-p_{|\jmath|})^{L-\ell} \\
&\quad= \sum_{\ell=0}^{\nu}\alpha_\ell(1-p_{|\jmath|})^{L-\ell} + \sum_{\ell=\nu+1}^{L}\alpha_\ell(1-p_{|\jmath|})^{L-\ell} \\
&\quad\leq (1-p_{|\jmath|})^{L-\nu}\sum_{\ell=0}^{\infty}\alpha_\ell + \sum_{\ell=\nu+1}^{\infty}\alpha_\ell, \quad (81)
\end{aligned}$$

where we upper-bound the first sum by bounding $(1-p_{|\jmath|})^{L-\ell}\leq(1-p_{|\jmath|})^{L-\nu}$, the second sum by bounding $(1-p_{|\jmath|})^{L-\ell}\leq 1$, and both sums by changing the upper limit of the sum over ℓ to ∞. The first sum on the RHS of (81) vanishes as $L\to\infty$, whereas the second sum vanishes as $\nu\to\infty$.

Combining (81) with (80), we thus obtain that the first term on the RHS of (80) vanishes as we first let $L\to\infty$ and then $\nu\to\infty$. As a result, the RHS of (76) tends to (79) as $L\to\infty$. Observe that (79) only depends on the set of active users in the intended cell $\jmath$ via its cardinality $|\jmath|$. We next use that, for $j\triangleq|\jmath|$, the activity pattern in the intended cell satisfies

$$\Pr\{\mathcal{J}=\jmath\}=\delta^j(1-\delta)^{n_T-j} \quad (82)$$

and that there are $\binom{n_T}{j}$ of distinct sets $\jmath$ with cardinality j. Also, the probability of a cell being $\mathcal{J}$-interfering is

$$p_{|\jmath|}=\delta^j. \quad (83)$$

Then, using (82) and (83), we obtain from (76), (79), and (45) that

$$\begin{aligned}
&\overline{\lim_{L\to\infty}}\frac{1}{n}I(\mathbf{X}_{0,1}^n;\hat{\mathbf{Y}}_1^n\,|\,A_1^L,\mathcal{J}) \\
&\leq \sum_{\jmath}\Pr\{\mathcal{J}=\jmath\}\Upsilon_{|\jmath|} \\
&= \sum_{j=1}^{n_T}\binom{n_T}{j}\delta^j(1-\delta)^{n_T-j}n_R\frac{1-\delta^j}{\delta^j}\log\left(\rho^{-\frac{3}{2}}\right) \\
&\quad+\left(1-(1-\delta)^{n_T}\right)\Bigg(\log\frac{\pi}{n_R\Gamma(n_R)}+n_R\log\frac{n_R}{e} \\
&\qquad+\frac{n_R}{2}\log(\eta_{\max})+n_R\log(1+\eta_{\max})\Bigg) \quad (84)
\end{aligned}$$

Evaluating the first term on the RHS of (84) using the binomial theorem, and combining (84) with (42)–(44), we obtain

$$\begin{aligned}
&\frac{1}{n}I(\mathbf{X}_{0,1}^n;\mathbf{Y}_1^n) \\
&\leq \frac{1}{n}H(\mathcal{J})+n_R\left((2-\delta)^{n_T}-1\right)\log\left(\rho^{-\frac{3}{2}}\right) \\
&\quad+\left(1-(1-\delta)^{n_T}\right)\Bigg(\log\frac{\pi}{n_R\Gamma(n_R)}+n_R\log\frac{n_R}{e} \\
&\qquad+\frac{n_R}{2}\log(\eta_{\max})+n_R\log(1+\eta_{\max})\Bigg) \quad (85)
\end{aligned}$$

from which the desired bound (7) follows by noting that, as $n\to\infty$, the term $\frac{1}{n}H(\mathcal{J})$ vanishes.

Appendix II
Proof of Proposition 1

The proof of Proposition 1 closely follows the proof of Theorem 1 in Appendix I, but relies on a slightly different definition of interfering cells. Specifically, in the proof of Proposition 1, we say that a cell ℓ is *J-interfering* if

$$|\mathcal{J}|\leq|\mathcal{G}_\ell|, \quad (86)$$

that is, if the number of active users in $\mathcal{G}_\ell$ is at least the number of active users in the intended cell $\mathcal{J}$. We adopt the notation J-interfering (instead of $\mathcal{J}$-interfering) to emphasize that the definition depends only on the cardinality $J=|\mathcal{J}|$, rather than on the specific set $\mathcal{J}$. In words, instead of restricting the analysis to cells whose active-user set $\mathcal{G}_\ell$ contains the entire set $\mathcal{J}$, we consider all interfering cells with at least $|\mathcal{J}|$ active users.

As in the proof of Theorem 1, we define the corresponding indicator variable A_ℓ, which is 1 if the ℓ-th cell is J-interfering and zero otherwise, i.e.,

$$A_\ell=\begin{cases}1, & |\mathcal{J}|\leq|\mathcal{G}_\ell|,\\ 0, & \text{otherwise.}\end{cases} \quad (87)$$

For the J-interfering cells, we apply a permutation on the user indices $u=1,\ldots,n_T$ on both $\mathbf{X}_{\ell,1}^n$ and $\mathbb{H}_{\ell,1}^n$ so that the active indices in cell ℓ are aligned with the active users in the intended cell $\mathcal{J}$. This permutation does not affect the mutual information $I(\mathbf{X}_{0,1}^n;\mathbf{Y}_1^n\,|\,\mathcal{J})$, since the transmitted symbols $(\tilde{X}_{\ell,1,1}^n,\ldots,\tilde{X}_{\ell,n_T,1}^n)$ are exchangeable by the proposition's assumption, and since $(B_{\ell,1},\ldots,B_{\ell,n_T})$ are i.i.d. and $\mathbb{H}_{\ell,1}^n$

has i.i.d. entries for all $\ell = 0, 1, \ldots$ according to the channel model.

After applying this permutation, all the steps in Appendix I apply for the new J-interfering cells until the final average of the bound in Section I-F. Indeed, let $\jmath$ be the set of active users in the intended cell and let $j = |\jmath|$. Then, the probability of the ℓ-th cell being J-interfering is

$$p_{|\jmath|} = \Pr[G \geq j] = \sum_{k=j}^{n_T} \binom{n}{k} \delta^k (1-\delta)^{n-k}, \qquad (88)$$

where $G \sim \mathrm{Bin}(n_T, \delta)$ follows a binomial distribution with n_T trials and activation probability $0 < \delta < 1$. Using (88) in (80), instead of $p_{|\jmath|} = \delta^j$, the steps (80)–(84) then yield that

$$\begin{aligned}
&\varlimsup_{L\to\infty} \frac{1}{n} I(\mathbf{X}_{0,1}^n; \hat{\mathbf{Y}}_1^n \,|\, A_1^L, \mathcal{J}) \\
&\leq n_R \sum_{j=1}^{n_T} \Pr[G=j] \frac{1-\Pr[G\geq j]}{\Pr[G\geq j]} \log\left(\rho^{-\frac{3}{2}}\right) \\
&\quad + \left(1-(1-\delta)^{n_T}\right)\Bigg(\log\frac{\pi}{n_R\Gamma(n_R)} + n_R \log\frac{n_R}{e} \\
&\qquad + \frac{n_R}{2}\log(\eta_{\max}) + n_R\log(1+\eta_{\max})\Bigg). \qquad (89)
\end{aligned}$$

Therefore, to prove (11) in Proposition 1, it only remains to show that $\sum_{j=1}^{n_T} \Pr[G=j] \frac{1-\Pr[G\geq j]}{\Pr[G\geq j]}$ is upper bounded by $n_T(1-\delta)$. To this end, note that

$$\begin{aligned}
&\sum_{j=1}^{n_T} \Pr[G=j] \frac{1-\Pr[G\geq j]}{\Pr[G\geq j]} \\
&\overset{(a)}{\leq} \sum_{j=1}^{n_T} \Pr[G=j] \frac{1-\Pr[G\geq j]}{\Pr[G=j]} \\
&\overset{(b)}{=} \sum_{j=1}^{n_T} \Pr[G<j] \\
&\overset{(c)}{=} \sum_{k=0}^{n_T} (n_T-k)\,\Pr[G=k] \\
&\overset{(d)}{=} n_T - n_T\delta, \qquad (90)
\end{aligned}$$

where in (a) we used that $\Pr[G \geq j] \geq \Pr[G=j]$; in (b) we used that $1-\Pr[G\geq j] = \Pr[G<j]$; (c) follows by reindexing, since each $\Pr[G=k]$ appears exactly $n_T - k$ times in the sum; and (d) follows since $\sum_{k=0}^{n_T} \Pr[G=k] = 1$ and $\sum_{k=0}^{n_T} \Pr[G=k]k = \mathbb{E}[G] = n_T\delta$.

Using (90) in (89), combining the resulting expression with (42)–(44), and taking the limit as $n \to \infty$, we obtain the desired upper bound (11) on the exchangeable capacity.

Appendix III
Proof of Proposition 4

Let

$$X_{u,k} = B_u \tilde{X}_{u,k}, \quad u = 1, \ldots, n_T, \qquad (91)$$

where $B_u \sim \mathrm{Ber}(\delta)$ and $\tilde{X}_{u,k}$ is independent of P, has unit energy, and is non-zero almost surely. Further let $\mathbf{X}_k = [X_{1,k}, \ldots, X_{n_T,k}]^T$, $\mathbf{B} = [B_1, \ldots, B_{n_T}]^T$, and $\tilde{\mathbf{X}}_k = [\tilde{X}_{1,k}, \ldots, \tilde{X}_{n_T,k}]^T$. The mutual information can then be upper-bounded as

$$\begin{aligned}
I(\mathbf{X}_k; \mathbf{Y}_k) &\overset{(a)}{\leq} I(\mathbf{X}_k; \mathbf{Y}_k, \mathbf{B}) \\
&\overset{(b)}{=} I(\mathbf{X}_k; \mathbf{B}) + I(\mathbf{X}_k; \mathbf{Y}_k|\mathbf{B}) \\
&\overset{(c)}{\leq} n_T H_b(\delta) + I(\mathbb{H}_k, \mathbf{X}_k; \mathbf{Y}_k|\mathbf{B}) \\
&\quad - I(\mathbf{Y}_k; \mathbb{H}_k|\mathbf{X}_k, \mathbf{B}), \qquad (92)
\end{aligned}$$

where (a) follows by giving the extra information $\mathbf{B}$; (b) follows from the chain rule of mutual information; and (c) follows by upper-bounding $I(\mathbf{X}_k; \mathbf{B})$ by the entropy of the independent binary random variables $B_1, \ldots, B_{n_T}$ and by using the chain rule of mutual information to add and subtract the extra information $\mathbb{H}_k$.

We continue by bounding the first mutual information on the RHS of (92) as

$$\begin{aligned}
I(\mathbb{H}_k, \mathbf{X}_k; \mathbf{Y}_k|\mathbf{B}) &= h(\mathbf{Y}_k|\mathbf{B}) - h(\mathbf{Y}_k|\mathbb{H}_k, \mathbf{X}_k, \mathbf{B}) \\
&\overset{(a)}{\leq} \sum_{r=1}^{n_R} \log\left(1 + \mathsf{P}\sum_{u=1}^{n_T} g_{r,u}\right) \\
&\overset{(b)}{=} n_R \log(1+\mathsf{P}) \qquad (93)
\end{aligned}$$

where (a) follows by upper-bounding the first entropy by that of a Gaussian random vector with the same second moment as $\mathbf{Y}_k$ [43, Th. 8.6.5] and by noting that the second entropy is the entropy of $\mathbf{Z}_k$, which has a Gaussian distribution; and (b) follows because, by the proposition's assumption, $g_{r,1} + \ldots + g_{r,n_T} = 1$.

We next analyze the second mutual information on the RHS of (92) as follows:

$$\begin{aligned}
&I(\mathbf{Y}_k; \mathbb{H}_k|\mathbf{X}_k, \mathbf{B}) \\
&\quad = \sum_{r=1}^{n_R} I(Y_{r,k}; \mathbb{H}_k|\mathbf{X}_k, \mathbf{B}) \\
&\quad = \sum_{r=1}^{n_R} \mathsf{E}\left[\log\left(1 + \mathsf{P}\sum_{u=1}^{n_T} B_t |\tilde{X}_u|^2 g_{r,u}\right)\right], \qquad (94)
\end{aligned}$$

where we used that, conditioned on $\mathbf{X}_k$ and $\mathbf{B}$, the entries in $\mathbf{Y}_k$ are independent of each other and have a Gaussian distribution, hence the mutual information can be evaluated in closed form.

We lower-bound the RHS of (94) by distinguishing between the case where $B_1 = \ldots = B_{n_T} = 0$ and the case where at least one B_t is 1. In the former case, the logarithm inside the expectation is zero. In the latter case, we can lower-bound it as

$$\log\left(1 + \mathsf{P}\sum_{u=1}^{n_T} B_t |\tilde{X}_u|^2 g_{r,u}\right) \geq \log\left(\mathsf{P}|\tilde{X}_{u'}|^2 g_{r,u'}\right) \qquad (95)$$

for any arbitrary $u' = 1, \ldots, n_T$ for which $B_{u'} = 1$. It follows that

$$\begin{aligned}
&I(\mathbf{Y}_k; \mathbb{H}_k|\mathbf{X}_k, \mathbf{B}) \\
&\quad \geq \left(1-(1-\delta)^{n_T}\right) \min_{u=1,\ldots,n_T} \mathsf{E}\left[\log\left(\mathsf{P}|\tilde{X}_u|^2 g_{r,u}\right)\right]. \qquad (96)
\end{aligned}$$

Combining (93) and (96) with (92), we obtain the upper bound

$$\begin{aligned} I(\mathbf{X}_k;\mathbf{Y}_k) &\le n_R(1-\delta)^{n_T}\log\mathsf{P} + n_T H_b(\delta) + n_R\log\left(1+\frac{1}{\mathsf{P}}\right) \\ &\quad - \left(1-(1-\delta)^{n_T}\right)\sum_{r=1}^{n_R}\mathsf{E}\left[\log(|\tilde{X}_{u_\star}|^2 g_{r,u_\star})\right], \end{aligned} \tag{97}$$

where

$$u_\star \triangleq \arg\min_{u=1,\dots,n_T}\mathsf{E}\left[\log\left(|\tilde{X}_u|^2 g_{r,u}\right)\right]. \tag{98}$$

This is (21) in Proposition 4.

Appendix IV
Proof of Lemma 1

By (27), we have

$$\sum_{\ell=L_\mathsf{P}+1}^{\infty}\alpha_\ell \le \sum_{\ell=L_\mathsf{P}+1}^{\infty}\frac{1}{\exp(\exp(\ell^a))}. \tag{99}$$

Any L_P satisfying

$$\sum_{\ell=L_\mathsf{P}+1}^{\infty}\frac{1}{\exp(\exp(\ell^a))} \le \frac{\sigma^2}{\mathsf{P}} \tag{100}$$

will therefore also satisfy (25). To find such an L_P, we first rewrite the sum in (100) as

$$\begin{aligned} &\sum_{\ell=L_\mathsf{P}+1}^{\infty}\frac{1}{\exp(\exp(\ell^a))} \\ &\quad = \frac{1}{\exp\left(\exp\left(L_\mathsf{P}^a\right)\right)}\sum_{\ell=L_\mathsf{P}+1}^{\infty}\exp\left(\exp(L_\mathsf{P}^a)-\exp(\ell^a)\right). \end{aligned} \tag{101}$$

Since the function $x \mapsto e^{x^a}-x^a$, $a \ge 1$ is monotonically increasing, we next note that, for every $\ell \ge L_\mathsf{P}+1$,

$$\exp\left(\exp\left(L_\mathsf{P}^a\right)-\exp(\ell^a)\right) \le \exp\left(L_\mathsf{P}^a-\ell^a\right). \tag{102}$$

Using a first-order Taylor series of ℓ^a around $\ell = L_\mathsf{P}$, and that $x \mapsto x^a$ is a convex function for $a \ge 1$, we obtain

$$\ell^a \ge L_\mathsf{P}^a + aL_\mathsf{P}^{a-1}\left(\ell - L_\mathsf{P}\right). \tag{103}$$

It follows that

$$\begin{aligned} &\sum_{\ell=L_\mathsf{P}+1}^{\infty}\exp\left(\exp(L_\mathsf{P}^a)-\exp(\ell^a)\right) \\ &\quad \le \sum_{\ell=L_\mathsf{P}+1}^{\infty}\exp\left(aL_\mathsf{P}^a - \ell a L_\mathsf{P}^{a-1}\right) \\ &\quad = \frac{\exp\left(-aL_\mathsf{P}^{a-1}\right)}{1-\exp\left(-aL_\mathsf{P}^{a-1}\right)} \\ &\quad \le \frac{e^{-a}}{1-e^{-a}} \end{aligned} \tag{104}$$

by the expression of the geometric sum, and because the function $x \mapsto e^{-x}/(1-e^{-x})$ is monotonically decreasing in x and $L_\mathsf{P} \ge 1$. We thus have that

$$\sum_{\ell=L_\mathsf{P}+1}^{\infty}\frac{1}{\exp(\exp(\ell^a))} \le \frac{1}{\exp\left(\exp\left(L_\mathsf{P}^a\right)\right)}\frac{e^{-a}}{1-e^{-a}}. \tag{105}$$

We finish the proof by noting that any integer L_P satisfying

$$L_\mathsf{P} \ge \left(\log\log\left(\frac{\mathsf{P}}{\sigma^2}\frac{e^{-a}}{1-e^{-a}}\right)\right)^{\frac{1}{a}} \tag{106}$$

yields

$$\frac{1}{\exp\left(\exp\left(L_\mathsf{P}^a\right)\right)}\frac{e^{-a}}{1-e^{-a}} \le \frac{\sigma^2}{\mathsf{P}}. \tag{107}$$

By (99) and (105), the same L_P also satisfies (25). This proves Lemma 1.

Appendix V
Proof of Theorem 2

Consider the bursty signaling scheme introduces in Section IV, where the signal transmitted by user u at time k is given by

$$X_{\ell,u,k} = B_{\ell,1}\hat{B}_{\ell,u}\hat{X}_{\ell,u,k}, \quad \ell = 0,1,\dots \tag{108}$$

Here, $B_{\ell,u} \sim \mathrm{Ber}(\delta)$ models the user activity, $\hat{B}_{\ell,1} \sim \mathrm{Ber}(\xi)$ is a random variable that artificially introduces burstiness, $\{\hat{X}_{\ell,1,k},\, k \in \mathbb{Z}\}$ are i.i.d. circularly-symmetric random variables with $\log|\hat{X}_{\ell,1,k}|^2$ uniformly distributed over the interval $[0, \log P]$, and $\hat{B}_{\ell,u}$ and $\{\hat{X}_{\ell,u,k},\, k \in \mathbb{Z}\}$ are zero for $u = 2,\dots,n_T$; cf. (23) and (24).

Since the only active user is $u = 1$, we will omit the subscript u in the remainder of this proof. To simplify notation, let $\tilde{B}_\ell \triangleq B_\ell\hat{B}_\ell$. Since B_ℓ and $\hat{B}_\ell$ are independent, $\tilde{B}_\ell$ has a Bernoulli distribution with activation probability $\delta\xi$. By (4), a lower bound on $C(\mathsf{P})$ follows by lower-bounding the mutual information $I(\mathbf{X}_{0,1}^n;\mathbf{Y}_1^n)$ as

$$\begin{aligned} &I(X_{0,1}^n;\mathbf{Y}_1^n) \\ &\quad \overset{(a)}{=} I(X_{0,1}^n;\tilde{B}_1^{L_\mathsf{P}}) + I(X_{0,1}^n;\mathbf{Y}_1^n|\tilde{B}_1^{L_\mathsf{P}}) - I(X_{0,1}^n;\tilde{B}_1^{L_\mathsf{P}}|\mathbf{Y}_1^n) \\ &\quad \overset{(b)}{=} I(X_{0,1}^n;\tilde{B}_1^{L_\mathsf{P}}) + I(X_{0,1}^n;\mathbf{Y}_1^n|\tilde{B}_1^{L_\mathsf{P}}) \\ &\qquad - H(\tilde{B}_1^{L_\mathsf{P}}|\mathbf{Y}_1^n) + H(\tilde{B}_1^{L_\mathsf{P}}|\mathbf{Y}_1^n, X_{0,1}^n) \\ &\quad \overset{(c)}{\ge} I(X_{0,1}^n;\mathbf{Y}_1^n|\tilde{B}_1^{L_\mathsf{P}}) - L_\mathsf{P}\log 2 \\ &\quad \overset{(d)}{\ge} I(\hat{X}_{0,1}^n,\tilde{B}_0;\mathbf{Y}_1^n|\tilde{B}_1^{L_\mathsf{P}}) - L_\mathsf{P}\log 2 \\ &\quad \overset{(e)}{=} I(\tilde{B}_0;\mathbf{Y}_1^n|\tilde{B}_1^{L_\mathsf{P}}) + I(\hat{X}_{0,1};\mathbf{Y}_1^n|\tilde{B}_1^{L_\mathsf{P}},\tilde{B}_0) - L_\mathsf{P}\log 2 \\ &\quad \overset{(f)}{\ge} I(\hat{X}_{0,1}^n;\mathbf{Y}_1^n|\tilde{B}_1^{L_\mathsf{P}},\tilde{B}_0) - L_\mathsf{P}\log 2, \end{aligned} \tag{109}$$

where L_P is an integer satisfying (25). Here, (a) follows from the chain rule for mutual information; (b) follows by expressing the last mutual information as a difference of conditional entropies of the discrete random variables $\tilde{B}_1^{L_\mathsf{P}}$; (c) follows by the nonnegativity of mutual information and entropy, and by upper-bounding the entropy $H(\tilde{B}_1^{L_\mathsf{P}}|\mathbf{Y}_1^n)$ by the logarithm of the number of possible values $\tilde{B}_1^{L_\mathsf{P}}$ can take; (d) follows because $X_{0,1}^n$ is a function of $\hat{X}_{0,1}^n$ and $\tilde{B}_0$ and by the data processing inequality [43, Th. 2.8.1]; (e) follows again from the chain rule of mutual information; (f) follows again from the nonnegativity of mutual information.

To lower-bound the mutual information on the RHS of (109), we write the mutual information $I(\hat{X}_{0,1}^n;\mathbf{Y}_1^n|\tilde{B}_1^{L_\mathsf{P}},\tilde{B}_0)$ as an average of the conditional mutual informations

$I(\hat{X}_{0,1}^n; \mathbf{Y}_1^n | \tilde{B}_1^{L_\mathsf{P}} = \tilde{b}_1^{L_\mathsf{P}}, \tilde{B}_0 = \tilde{b}_0)$, and lower-bound these mutual informations as follows: If $\tilde{b}_0 = 0$ or $\tilde{b}_1^{L_\mathsf{P}} \neq \mathbf{0}$, then we use the nonnegativity of mutual information to bound

$$I(\hat{X}_{0,1}^n; \mathbf{Y}_1^n | \tilde{B}_1^{L_\mathsf{P}} = \tilde{b}_1^{L_\mathsf{P}}, \tilde{B}_0 = \tilde{b}_0) \geq 0. \tag{110}$$

If $\tilde{b}_0 = 1$ and $\tilde{b}_1^{L_\mathsf{P}} = \mathbf{0}$, then we use the following lemma:

Lemma 3: We have that

$$\begin{aligned} &I(\hat{X}_{0,1}^n; \mathbf{Y}_1^n | \tilde{B}_1^{L_\mathsf{P}} = \mathbf{0}, \tilde{B}_0 = 1) \\ &\quad \geq n \left(\log\log \mathsf{P} - \gamma - \log(e) - 2\log\left(1 + \sqrt{2}\sigma\right)\right) \\ &\quad \triangleq n\underline{R}(\mathsf{P}). \end{aligned} \tag{111}$$

Proof: The result is a lower bound on the capacity of a noncoherent point-to-point fading channel with non-Gaussian additive noise and is based on a lower bound presented in [49, Lemma 4]. For completeness, we provide the proof in the subsection at the end of this appendix. ∎

It follows from (110) and (111) that

$$I(\hat{X}_{0,1}^n; \mathbf{Y}_1^n | \tilde{B}_1^{L_\mathsf{P}}, \tilde{B}_0) \geq n\delta\xi(1-\delta\xi)^{L_\mathsf{P}} \underline{R}(\mathsf{P}), \tag{112}$$

where we also used that $\Pr(\tilde{B}_0 = 1) = \delta\xi$ and

$$\Pr\left(\tilde{B}_1^{L_\mathsf{P}} = \mathbf{0}\right) = (1-\delta\xi)^{L_\mathsf{P}}. \tag{113}$$

Replacing (112) in (109), and dividing both sides of the inequality by n, we finally obtain

$$\frac{1}{n} I(\mathbf{X}_{0,1}^n; \mathbf{Y}_1^n) \geq \delta\xi(1-\delta\xi)^{L_\mathsf{P}} \underline{R}(\mathsf{P}) - \frac{1}{n} L_\mathsf{P} \log 2. \tag{114}$$

The lower bound (29) in Theorem 2 follows then from (114) and (4) upon letting n tend to infinity.

A. Proof of Lemma 3

It follows from the chain rule for mutual information that

$$\begin{aligned} &I(\hat{X}_{0,1}^n; \mathbf{Y}_1^n | \tilde{B}_1^{L_\mathsf{P}} = \mathbf{0}, \tilde{B}_0 = 1) \\ &\quad = \sum_{k=1}^n I(\hat{X}_{0,k}; \mathbf{Y}_1^n | \tilde{B}_1^{L_\mathsf{P}} = \mathbf{0}, \tilde{B}_0 = 1, \hat{X}_{0,1}^{k-1}) \\ &\quad \geq \sum_{k=1}^n I(\hat{X}_{0,k}; \mathbf{Y}_k | \tilde{B}_1^{L_\mathsf{P}} = \mathbf{0}, \tilde{B}_0 = 1), \end{aligned} \tag{115}$$

where we used in the second step that the channel inputs are independent and that removing the channel outputs $\mathbf{Y}_\ell$, $\ell \neq k$ reduces mutual information.

We next use the assumptions $\tilde{B}_0 = 1$ and $\tilde{\mathbf{B}}_1^{L_\mathsf{P}} = \mathbf{0}$ to express the channel output $\mathbf{Y}_k$ as

$$\mathbf{Y}_k = \mathbf{H}_{0,k} \hat{X}_{0,k} + \mathbf{W}_k, \tag{116}$$

where

$$\mathbf{W}_k \triangleq \sum_{\ell=L_\mathsf{P}+1}^{\infty} \mathbf{H}_{\ell,k} X_{\ell,k} + \mathbf{Z}_k \tag{117}$$

and $\mathbf{H}_{\ell,k}$ are $(n_R \times 1)$-vectors containing the fading coefficients from user 1 in cell ℓ to the receive antennas in the intended cells.

With these definitions, we can lower-bound the mutual information in (115) as

$$\begin{aligned} &I(\hat{X}_{0,k}; \mathbf{Y}_k | \tilde{B}_1^{L_\mathsf{P}} = \mathbf{0}, \tilde{B}_0 = 1) \\ &\quad = I(\hat{X}_{0,k}; \mathbf{H}_{0,k} \hat{X}_{\tilde{\mathbf{b}}_{0,k}} + \mathbf{W}_k) \\ &\quad \geq I(\hat{X}_{0,k}; H_{0,1,k} \hat{X}_{0,k} + W_k), \end{aligned} \tag{118}$$

where $H_{0,1,k}$ denotes the time-k fading coefficient from user 1 to receive antenna 1, and W_k denotes the first component of $\mathbf{W}_k$. Here, the inequality follows by ignoring the received signal at receive antennas $2, \ldots, n_R$.

We next lower-bound $I(\hat{X}_{0,k}; H_{0,1,k}\hat{X}_{0,k} + W_k)$ using the following lemma:

Lemma 4 (Lapidoth [49]): Let the random variables X, H, and W have finite second moments. Assume that both X and H are of finite differential entropy. Finally, assume that X is independent of H; that X is independent of W; and that $X \multimap\!\!\circ H \multimap\!\!\circ W$ forms a Markov chain. Then,

$$\begin{aligned} I(X; HX + W) \geq h(X) - \mathsf{E}\left[\log |X|^2\right] + \mathsf{E}\left[\log |H|^2\right] \\ - \mathsf{E}\left[\log\left(\pi e \left(\sigma_H + \frac{\sigma_W}{|X|}\right)^2\right)\right], \end{aligned} \tag{119}$$

where $\sigma_H > 0$ and $\sigma_W \geq 0$ denotes the standard deviations of H and W, respectively.

Proof: See [49, Lemma 4]. ∎

It is easy to check that $X = \hat{X}_{0,k}$, $H = H_{0,1,k}$, and $W = W_k$ satisfy the lemma's condition. In particular, by the definition of L_P and our choice of $X_{\ell,k}$,

$$\mathsf{E}\left[|W_k|^2\right] \leq \sum_{\ell=L_\mathsf{P}+1}^{\infty} \alpha_\ell \mathsf{P} + \sigma^2 \leq 2\sigma^2. \tag{120}$$

It thus follows from Lemma 4 that

$$\begin{aligned} &I(\hat{X}_{0,k}; H_{0,1,k}\hat{X}_{0,k} + W_k) \\ &\quad \geq h(\hat{X}_{0,k}) - \mathsf{E}\left[\log |\hat{X}_{0,k}|^2\right] + \mathsf{E}\left[\log |H_{0,1,k}|^2\right] \\ &\qquad - \log(\pi e) - 2\mathsf{E}\left[\log\left(1 + \frac{\sqrt{\mathsf{E}[|W_k|^2]}}{|\hat{X}_{0,k}|}\right)\right] \\ &\quad \geq h(\hat{X}_{0,k}) - \mathsf{E}\left[\log |\hat{X}_{0,k}|^2\right] + \mathsf{E}\left[\log |H_{0,1,k}|^2\right] \\ &\qquad - \log(\pi e) - 2\log\left(1 + \sqrt{2}\sigma\right), \end{aligned} \tag{121}$$

where the second inequality follows from (120) and because $|\hat{X}_{0,k}| \geq 1$.

The differential entropy on the RHS of (121) can be evaluated as

$$\begin{aligned} h(\hat{X}_{0,k}) &= h(\log |\hat{X}_{0,k}|^2) + \mathsf{E}\left[\log |\hat{X}_{0,k}|^2\right] + \log \pi \\ &= \log\log \mathsf{P} + \mathsf{E}\left[\log |\hat{X}_{0,k}|^2\right] + \log \pi, \end{aligned} \tag{122}$$

where the first step follows by expressing the differential entropy of a circularly-symmetric random variable by the differential entropy of the logarithm of its absolute value [34, Eqs. (326) and (316)]; and the second step follows by computing the differential entropy of the uniformly-distributed random variable $\log |\hat{X}_{0,k}|^2$.

Similarly, the logarithmic moment of the exponentially-distributed random variable $|H_{0,k}|^2$ can be computed as (see, e.g., [34, Eqs. (209)-(212)])

$$\mathsf{E}\left[\log|H_{0,1,k}|^2\right] = -\gamma. \tag{123}$$

Combining (122) and (123) with (121), we obtain

$$\begin{aligned} &I(\hat{X}_{0,k}; H_{0,1,k}\hat{X}_{0,k} + W_k) \\ &\quad \geq \log\log \mathsf{P} - \gamma - \log(e) - 2\log\left(1+\sqrt{2}\sigma\right). \end{aligned} \tag{124}$$

Together with (118) and (115), this yields

$$\begin{aligned} &I(\hat{X}_{0,1}^n; \mathbf{Y}_1^n | \tilde{B}_1^{L_\mathsf{P}} = \mathbf{0}, \tilde{B}_0 = 1) \\ &\quad \geq n\left(\log\log \mathsf{P} - \gamma - \log(e) - 2\log\big(1+\sqrt{2}\sigma\big)\right), \end{aligned} \tag{125}$$

which proves Lemma 3.

Appendix VI
Proof of Corollary 1

By assumption, the fading variances α_ℓ satisfy (27) for some $a > 1$. It thus follows from Lemma 1 that any integer L_P satisfying

$$L_\mathsf{P} > \left(\log\log\left(\frac{\mathsf{P}}{\sigma^2}\frac{e^{-a}}{1-e^{-a}}\right)\right)^{\frac{1}{a}} \tag{126}$$

also satisfies (25). We next choose the activation probability

$$\xi_\mathsf{P} = \big(\log\log \mathsf{P}\big)^{-\frac{1+\varepsilon}{a}} \tag{127}$$

for some arbitrary $0 < \varepsilon < a-1$. We then analyze the leading term in (29), given by,

$$\delta\xi_\mathsf{P}(1-\delta\xi_\mathsf{P})^{L_\mathsf{P}} \log\log \mathsf{P} \tag{128}$$

asymptotically as $\mathsf{P} \to \infty$. To this end, we first note that, if

$$\varliminf_{\mathsf{P}\to\infty} (1-\delta\xi_\mathsf{P})^{L_\mathsf{P}} > 0 \tag{129}$$

then there exist positive constants ν and P_0 such that, for $\mathsf{P} \geq \mathsf{P}_0$, the leading term in (29) is lower-bounded as

$$\begin{aligned} \delta\xi_\mathsf{P}(1-\delta\xi_\mathsf{P})^{L_\mathsf{P}} \log\log \mathsf{P} &\geq \delta\xi_\mathsf{P}\nu \log\log \mathsf{P} \\ &= \delta\nu \left(\log\log \mathsf{P}\right)^{1-\frac{1+\varepsilon}{a}}, \end{aligned} \tag{130}$$

from which Corollary 1 follows.

It thus remains to prove (129). To this end, we consider the logarithm of $(1-\delta\xi_\mathsf{P})^{L_\mathsf{P}}$ and show that it is finite. Indeed, we have that

$$\begin{aligned} &L_\mathsf{P}\log(1-\delta\xi_\mathsf{P}) \\ &> \left(\log\log\left(\frac{\mathsf{P}}{\sigma^2}\frac{e^{-a}}{1-e^{-a}}\right)\right)^{\frac{1}{a}} \log\left(1 - \frac{\delta}{\big(\log\log \mathsf{P}\big)^{\frac{1+\varepsilon}{a}}}\right) \\ &\geq -\frac{\delta\left(\log\log\left(\frac{\mathsf{P}}{\sigma^2}\frac{e^{-a}}{1-e^{-a}}\right)\right)^{\frac{1}{a}}}{\big(\log\log \mathsf{P}\big)^{\frac{1+\epsilon}{a}}} \left(1 - \frac{\delta}{\big(\log\log \mathsf{P}\big)^{\frac{1+\varepsilon}{a}}}\right)^{-1}, \end{aligned} \tag{131}$$

where in the second step we used that, for $0 \leq z < 1$, we have $\log(1-z) \geq -\frac{z}{1-z}$.

The first term on the RHS of (131) is of order $(\log\log \mathsf{P})^{1/a}$; the second term is of order $(\log\log \mathsf{P})^{-(1+\varepsilon)/a}$; the third term tends to 1 as $\mathsf{P} \to \infty$. It follows that the RHS of (131) is of order $(\log\log \mathsf{P})^{-\varepsilon/a}$, which tends to zero as $\mathsf{P} \to \infty$. Since the limit inferior of the logarithm of $(1-\delta\xi_\mathsf{P})^{L_\mathsf{P}}$ is lower-bounded by zero, we conclude that

$$\varliminf_{\mathsf{P}\to\infty} (1-\delta\xi_\mathsf{P})^{L_\mathsf{P}} \geq 1. \tag{132}$$

In fact, we have $(1-\delta\xi_\mathsf{P})^{L_\mathsf{P}} \leq 1$ for every P, so it even holds that $(1-\delta\xi_\mathsf{P})^{L_\mathsf{P}} \to 1$ as $\mathsf{P} \to \infty$. This proves (129) and concludes the proof of Corollary 1.

References

[1] G. P. Fettweis and H. Boche, "6G: The personal tactile internet—and open questions for information theory," *IEEE BITS Inf. Theory Mag.*, vol. 1, no. 1, pp. 71–82, Sep. 2021.

[2] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133 995–134 030, 2020.

[3] M. Alsabah, M. A. Naser, B. M. Mahmmod, S. H. Abdulhussain, M. R. Eissa, A. Al-Baidhani, N. K. Noordin, S. M. Sait, K. A. Al-Utaibi, and F. Hashim, "6G wireless communications networks: A comprehensive survey," *IEEE Access*, vol. 9, pp. 148 191–148 243, 2021.

[4] I. Analytics, "State of IoT summer 2024," Aug. 2024, [Online]. Available: https://iot-analytics.com/number-connected-iot-devices/.

[5] R. Kumar, D. Sinwar, and V. Singh, "QoS aware resource allocation for coexistence mechanisms between eMBB and URLLC: Issues, challenges, and future directions in 5G," *Comput. Commun.*, vol. 213, pp. 208–235, Jan. 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140366423003894

[6] T. Zahir, K. Arshad, A. Nakata, and K. Moessner, "Interference management in femtocells," *IEEE Commun. Surv. Tutor.*, vol. 15, no. 1, pp. 293–311, 1st Quart. 2013.

[7] J. Choi, J. Ding, N.-P. Le, and Z. Ding, "Grant-free random access in machine-type communication: Approaches and challenges," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 151–158, Feb. 2022.

[8] M. Ozates, M. J. Ahmadi, M. Kazemi, D. Gündüz, and T. M. Duman, "Unsourced random access: A comprehensive survey," *IEEE Commun. Surv. Tutor.*, vol. 28, pp. 955–984, 2026.

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

[10] A. El Gamal and Y.-H. Kim, *Network Information Theory*. USA: Cambridge University Press, 2011.

[11] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[12] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the k-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[13] B. Hamdaoui, P. Venkatraman, and M. Guizani, "Opportunistic exploitation of bandwidth resources through reinforcement learning," in *Proc. IEEE Global Telecom. Conf. (GLOBECOM)*, Honolulu, HI, USA, Nov. 2009, pp. 1–6.

[14] N. Khude, V. Prabhakaran, and P. Viswanath, "Opportunistic interference management," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, South Korea, Jun. 2009, pp. 2076–2080.

[15] I. H. Wang, C. Suh, S. Diggavi, and P. Viswanath, "Bursty interference channel with feedback," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 21–25.

[16] G. Villacrés, T. Koch, A. Sezgin, and G. Vazquez-Vilar, "Robust signaling for bursty interference," *Entropy*, vol. 20, no. 11, 2018. [Online]. Available: https://www.mdpi.com/1099-4300/20/11/870

[17] 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.

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

[19] 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.

[20] K.-H. Ngo, A. Lancho, G. Durisi, and A. Graell i Amat, "Unsourced multiple access with random user activity," *IEEE Trans. Inf. Theory*, vol. 69, no. 7, pp. 4537–4558, Jul. 2023.

[21] X. Liu, P. Pascual Cobo, and R. Venkataramanan, "Many-user multiple access with random user activity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Athens, Greece, Jul. 2024, pp. 2993–2998.

[22] ——, "Many-user multiple access with random user activity: Achievability bounds and efficient schemes," *IEEE Trans. Inf. Theory*, vol. 72, no. 1, pp. 383–414, Jan. 2026.

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

[24] 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.

[25] J. Gao, Y. Wu, G. Caire, W. Yang, H. Vincent Poor, and W. Zhang, "Unsourced random access in MIMO quasi-static Rayleigh fading channels: Finite blocklength and scaling law analyses," *IEEE Trans. Inf. Theory*, vol. 71, no. 6, pp. 4342–4373, Jun. 2025.

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

[27] A. Vem, K. R. Narayanan, J. Chamberland, and J. Cheng, "A user-independent successive interference cancellation based coding scheme for the unsourced random access Gaussian channel," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8258–8272, Dec. 2019.

[28] A. Fengler, P. Jung, and G. Caire, "SPARCs for unsourced random access," *IEEE Trans. Inf. Theory*, vol. 67, no. 10, pp. 6894–6915, Oct. 2021.

[29] V. K. Amalladinne, J. F. Chamberland, and K. R. Narayanan, "A coded compressed sensing scheme for unsourced multiple access," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6509–6533, Oct. 2020.

[30] A. Fengler, P. Jung, and G. Caire, "Sparcs for unsourced random access," *IEEE Trans. Inf. Theory*, vol. 67, no. 10, pp. 6894–6915, Oct. 2021.

[31] V. K. Amalladinne, A. K. Pradhan, C. Rush, J.-F. Chamberland, and K. R. Narayanan, "Unsourced random access with coded compressed sensing: Integrating amp and belief propagation," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2384–2409, Apr. 2022.

[32] K. Hsieh, C. Rush, and R. Venkataramanan, "Near-optimal coding for many-user multiple access channels," *IEEE J. Sel. Areas Inf. Theory*, vol. 3, no. 1, pp. 21–36, Jan. 2022.

[33] A. Lozano, R. Heath, and J. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, Sep. 2013.

[34] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory,*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[35] Z. Chen, F. Sohrabi, Y.-F. Liu, and W. Yu, "Phase transition analysis for covariance-based massive random access with massive MIMO," *IEEE Trans. Inf. Theory*, vol. 68, no. 3, pp. 1696–1715, Mar. 2022.

[36] Z. Chen, F. Sohrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.

[37] A. E. Kalør, R. Kotaba, and P. Popovski, "Common message acknowledgments: Massive ARQ protocols for wireless access," *IEEE Trans. Commun.*, vol. 70, no. 8, pp. 5258–5270, Aug. 2022.

[38] R. Song, K. M. Attiah, and W. Yu, "Coded downlink massive random access and a finite de Finetti theorem," *IEEE Trans. Inf. Theory*, vol. 71, no. 9, Sep. 2025.

[39] J. Kang and W. Yu, "Minimum feedback for collision-free scheduling in massive random access," *IEEE Trans. Inf. Theory*, vol. 67, no. 12, pp. 8094–8108, Dec. 2021.

[40] ——, "Scheduling versus contention for massive random access in massive MIMO systems," *IEEE Trans. Commun.*, vol. 70, no. 9, pp. 5811–5824, Sep. 2022.

[41] G. Villacrés, T. Koch, and G. Vazquez-Vilar, "Bursty wireless networks of bounded capacity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angeles, CA, USA, Jun. 2020, pp. 2959–2964.

[42] M. Fekete, "Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten," *Mathematische Zeitschrift*, vol. 17, no. 1, pp. 228–249, Dec. 1923.

[43] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.

[44] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[45] S. Jakborvornphan, "Analysis of path loss propagation models in mobile communication," *J. Theor. Appl. Inf. Technol*, vol. 98, no. 4, pp. 725–730, Feb. 2020.

[46] R. B. Ash, *Information Theory*, ser. Dover Books on Mathematics. Mineola, NY: Courier Dover Publications, 1990.

[47] T. Koch and A. Lapidoth, "On multipath fading channels at high-SNR," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 5945–5957, Dec. 2010.

[48] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Elsevier/Academic Press, Amsterdam, 2007.

[49] A. Lapidoth, "On the high-SNR capacity of noncoherent networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3025–3036, Sep. 2005.

Grace Villacrés (Member, IEEE) received the M.Sc. and Ph.D. degrees in multimedia and communications from the Universidad Carlos III de Madrid, Spain, in 2015 and 2019, respectively. She also has an M.Sc. in information technology from the University of Applied Sciences Mannheim, Germany, in November 2009. From November 2011 until December 2013, she worked at SIKA Dr. Siebert and Kühn GmbH & Co., Kaufungen, Germany, as a development engineer. She joined the Signal Theory and Communications Department of Universidad Rey Juan Carlos in September 2023, where she is currently an Assistant Professor. Her research interests are in signal processing and information theory. She served as the Vice Chair for the Spain Chapter of the IEEE Information Theory Society from 2020 to 2023, and she is serving as the Chapter's Chair (2024–2027).

Tobias Koch (Senior Member, IEEE) received the M.Sc. (Hons.) and Ph.D. degrees in electrical engineering from ETH Zurich, Switzerland, in 2004 and 2009, respectively. From June 2010 until May 2012, he was a Marie Curie Intra-European Research Fellow with the University of Cambridge, UK. He was also a Research Intern at Bell Labs, Murray Hill, NJ, USA, in 2004, and the Universitat Pompeu Fabra, Barcelona, Spain, in 2007. He joined the Universidad Carlos III de Madrid, Spain, in 2012, where he is currently an Associate Professor. His research interests are in digital communication theory and information theory.

Dr. Koch received a Starting Grant from the European Research Council (ERC), a Ramón y Cajal Research Fellowship, a Marie Curie Intra-European Fellowship, a Marie Curie Career Integration Grant, and a Fellowship for Prospective Researchers from the Swiss National Science Foundation. He further received a medal of the 2018 Young Researchers Award "Agustín de Betancourt y Molina" by the Royal Academy of Engineering of Spain. He served as the Vice Chair for the Spain Chapter of the IEEE Information Theory Society from 2013 to 2016, and he served as the Chapter's Chair from 2020 to 2023. He was an Associate Editor for Communications for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2020 to 2026. He is also a member of the Board of Governors of the IEEE Information Theory Society from 2025 to 2027.

Gonzalo Vazquez-Vilar (Member, IEEE) received the Telecommunication Engineering degree from the University of Vigo, Spain, in 2004, the Master of Science degree from Stanford University, U.S., in 2008 and the Ph.D. in Communication Systems from the University of Vigo, Spain, in 2011.

In 2011–2014 he was a post-doctoral fellow in the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Spain, and since 2014 he has been with the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Spain. He has held appointments as visiting researcher at Stanford University, U.S., University of Cambridge, U.K., and Princeton University, U.S. His research interests lie in the field of Shannon theory, with emphasis on finite-length information theory and communications.