Villacrés, G.; Koch, T.; Vázquez-Vilar, G. Bursty Wireless Networks of Bounded Capacity, *2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 21-26 June 2020. IEEE, 2020, Pp. 2959-2964*
DOI: https://doi.org/10.1109/ISIT44484.2020.9174034

Bursty Wireless Networks of Bounded Capacity

Grace Villacrés, Tobias Koch and Gonzalo Vazquez-Vilar
Universidad Carlos III de Madrid, Spain & Gregorio Marañón Health Research Institute, Madrid, Spain
Emails:{gvillacres,koch,gvazquez}@tsc.uc3m.es

***Abstract*—The channel capacity of wireless networks is often studied under the assumption that the communicating nodes have perfect channel-state information and that interference is always present. In this paper, we study the channel capacity of a wireless network without these assumptions, i.e., a bursty noncoherent wireless network where the users are grouped in cells and the base-station features several receive antennas. We demonstrate that the channel capacity is bounded in the *signal-to-noise ratio (*SNR*)* when the number of receive antennas is finite and the probability of presence of interference is strictly positive.**

I. Introduction

According to the Ericsson mobility report [1], it is expected that, in 2025, 90 percent of broadband subscriptions will be mobile broadband. This will cause an increase in the data traffic and a demand for higher data rates. Fulfillment of such requirements in future wireless networks will be mainly based on the densification of the radio access network. Indeed, the next generations of wireless networks enable a more effective share of network resources through femtocells and macrocells. Such network densification, with the associated increase in the number of users that are communicating, implies higher interference from surrounding cells (inter-cell interference). Since interference is one of the main limiting factors to achieve higher data rates, its management has been the object of several studies; see, e.g., [2] and references therein.

To better understand the impact of interference on the throughput of wireless networks, the information-theoretic limits of such networks have been extensively studied in the past; see, e.g., [3]–[5]. Most of these works consider that i) interference is always present and ii) the nodes in the network have perfect *channel-state information (CSI)* in the sense that they have perfect knowledge of the fading coefficients. Regarding assumption i), we observe that in certain scenarios assuming that interference is always present is overly pessimistic. For instance, intermittent user activity or opportunistic frequency reuse among cells [6] may cause interference to be intermittent/bursty. Interference burstiness may be exploited to achieve higher data rates at the different cells. Regarding assumption ii), it is *prima facie* unclear whether perfect CSI can actually be obtained if the number of users is large. Thus, assuming that the receivers have perfect CSI may be overly optimistic.

This work has been supported in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 714161), and by the Spanish Ministerio de Economía y Competitividad under Grants RYC-2014-16332 and TEC2016-78434-C3-3-R (AEI/FEDER, EU).

The possible weaknesses in the assumptions i) and ii) motivate the present work. While bursty interference has been studied in the information theory literature [7]–[10], these works consider either a Gaussian channel or a linear deterministic model, with a finite number of users. Hence, their relevance for future wireless networks is unclear. Among the works that study the channel capacity of wireless networks without perfect CSI, the most relevant for this paper are [11] and [12]. Lozano *et al.* [11] consider a wireless network without perfect CSI and study the maximum rate achievable with channel inputs of the form $\sqrt{\text{SNR}}U$, where the distribution of U does not depend on the SNR, and demonstrate that this rate is bounded in the SNR. The authors in [12] model a wireless network without perfect CSI by a flat-fading channel where only two users communicate with each other (SISO scenario) and an infinite number of users interferes this communication. They show that, under some simplifying assumptions, the channel capacity of this channel remains bounded in the SNR even if the channel inputs normalized by $\sqrt{\text{SNR}}$ are allowed to depend on the SNR.

In the present work, we generalize [12] by considering a) a MIMO scenario where a number n_T of users communicates with a base-station equipped with a number n_R of receive antennas, b) this communication is affected by the interference from a large number of cells, and c) this interference is bursty. Similar to [12], the input distribution is allowed to depend arbitrarily on the SNR. We believe that this model better approximates cellular communication systems since our assumptions do not prohibit common access or multiplexing methods such as *frequency-division-multiple-access (FDMA)*, *orthogonal-frequency-division-multiple-access (OFDMA)*, *time-division-multiple-access (TDMA)* or *frequency hoping*, which can be modeled through the burstiness of the interference. We show that under assumptions a)–c), and under the assumption that the distances between the interfering cells and the intended cell grow at most exponentially, the channel capacity of such networks is bounded in the SNR, unless either the number of receive antennas tends to infinity or the interference is absent with probability one.

II. Channel Model

In a cellular network users are grouped inside cells and communicate with a base station. Users inside each cell are assumed to cooperate, hence there is no intra-cell interference, but they do not cooperate with the users in other cells.

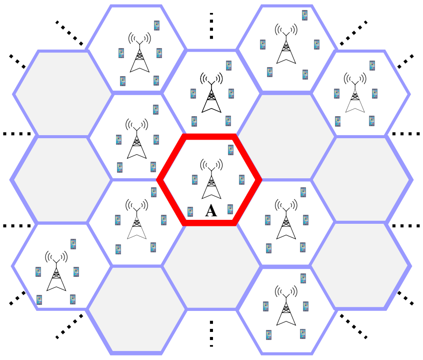

Figure 1. Channel model.

Since a characterization of all achievable rates in the network is unfeasible when the number of cells and users is large, it is common to study the *symmetric rate* of the network, i.e., the rate that can be achieved when all users communicate at the same rate. For the sake of tractability, we simplify the original problem as follows: Firstly, we consider the case where the users in a given cell are communicating with a base-station and the interfering cells emit symbols that interfere with this communication. To model a large network, we assume that there are infinitely many interfering cells. Secondly, to ensure that all nodes (transmitting and interfering) communicate at the same rate, we invoke a random-coding argument and assume that all nodes use the same distribution to draw their codebooks. This can be viewed as a generalization of the assumption of Gaussian codebooks, which is often encountered in the literature.

The presence of interference is modeled using a Bernoulli-distributed random variable that indicates whether the interference links from the different cells are present or not. We assume that the interference state remains constant during the whole transmission of the message. Our performance measure is the capacity of the channel between the users and the base-station inside a cell.

Specifically, we consider a system where n_T users are located in a cell and communicate with a base-station with n_R receive antennas. This transmission is affected by inter-cell interference. For each cell, the channel input-output relation at time $k \in \mathbb{Z}$ is

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

where $\mathbf{X}_{0,k} = [X_{0,1,k}, \dots, X_{0,n_T,k}] \in \mathbb{C}^{n_T \times 1}$ corresponds to a vector of transmitted symbols and $\mathbf{X}_{\ell,k} = [X_{\ell,1,k}, \dots, X_{\ell,n_T,k}] \in \mathbb{C}^{n_T \times 1}$ denotes the vector of interfering symbols. The vector $\mathbf{Z}_k \in \mathbb{C}^{n_R \times 1}$ models the time-k additive noise vector; $\mathbb{H}_{0,k} \in \mathbb{C}^{n_R \times n_T}$ denotes the matrix of fading coefficients of the links from the transmitters in the intended cell; $\mathbb{H}_{\ell,k} \in \mathbb{C}^{n_R \times n_T}$, $\ell = 1, 2, \dots$ denotes the matrix of fading coefficients of the interfering links from the transmitters inside the ℓ-th interfering cell; and $B_\ell \in \{0, 1\}$ denotes the state of the ℓ-th interfering cell. In Figure 1, **A** (in red) corresponds to the intended cell, the dotted lines indicate the infinite number of interfering cells. The shadowed cells correspond to cells that do not interfere due to interference burstiness and the non-shadowed cells are interfering the communication. We assume that $\{\mathbf{Z}_k,\, k \in \mathbb{Z}\}$, $\{B_\ell,\, \ell = 1, 2, \dots\}$, and $\{\mathbb{H}_{\ell,k},\, k \in \mathbb{Z}\}$, $\ell = 0, 1, \dots$ are independent sequences of *independent and identically distributed (IID)* random variables. We further assume that $B_\ell \sim Ber(p)$, $\mathbf{Z}_k \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2\mathbb{I})$, $\mathbb{H}_{0,k} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbb{I})$, and $\mathbb{H}_{\ell,k} \sim \mathcal{N}_{\mathbb{C}}(0, \alpha_\ell\mathbb{I})$ for some $\alpha_\ell > 0$, $\ell = 1, 2, \dots$ Here $\mathbb{I}$ denotes the identity matrix, $Ber(p)$ is the Bernoulli distribution with parameter p, and $\mathcal{N}_{\mathbb{C}}(\mu, K)$ denotes the circularly-symmetric complex Gaussian distribution with mean μ and covariance matrix K.

We consider a noncoherent scenario where transmitter and receiver only know the statistics of the fading coefficients but not their realizations. We assume that the interferers do neither cooperate with each other nor with the transmitters in the intended cell. Hence, the sequences $\{\mathbf{X}_{\ell,k},\, k \in \mathbb{Z}\}$, $\ell = 0, 1, \dots$ are independent. We further assume that each such sequence has the same distribution. Finally, we assume that the transmitters are not aware of which cells interfere and which do not, i.e., the input sequences are independent of the interference state $\{B_\ell, \ell = 1, 2, \dots\}$. However, the receiver of the intended cell may have access to $\{B_\ell, \ell = 1, 2, \dots\}$.

The locations of the interfering cells enter the channel model through the variance α_ℓ of the fading coefficients corresponding to the paths between the interfering cells and the intended cell. For simplicity, we assume that the autocovariance matrix of $H_{\ell,k}$ has identical diagonal elements.[1] Without loss of generality, we order the interfering cells according to the variances of the corresponding fading coefficients, i.e., $\alpha_\ell \geq \alpha_{\ell'}$ for any $\ell < \ell'$. We further assume that the total power of the interference received at the intended cell is finite, i.e., $\sum_{\ell=1}^{\infty} \alpha_\ell < \infty$, and that there exists $0 < \rho < 1$ such that

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

If we suppose that the path loss grows polynomially with the distance, then (2) implies that the distances from the interfering cells to the intended cell grow at most exponentially with the cell index ℓ.

III. Channel Capacity and Main Result

We denote sequences such as $A_n, A_{n+1}, \dots, A_m$ by A_n^m. We define two capacities of the channel (1), depending on the level of knowledge of the interference states at the receiver side:[2]

1) Receiver does not have access to the interference states:

$$C(\mathsf{P}) \triangleq \lim_{N\to\infty} \frac{1}{N} \sup_{Q^N} I(\mathbf{X}_{0,1}^N; \mathbf{Y}_1^N). \tag{3}$$

[1] This corresponds to the case where all nodes in the ℓ-th interfering cell are at the same distance from the receiver of the intended cell. This assumption can be relaxed to different diagonal elements of the autocovariance matrix, provided that they are finite and bounded away from zero. See Remark 1.

[2] The logarithms used in this paper are natural logarithms. The capacity has thus the dimension “nats per channel use”.

2) Receiver has access to the interference states:

$$C_B(\mathsf{P}) \triangleq \lim_{N\to\infty} \frac{1}{N} \sup_{Q^N} \lim_{L\to\infty} I(\mathbf{X}_{0,1}^N; \mathbf{Y}_1^N, B_1^L). \quad (4)$$

The suprema in (3) and (4) are over all N-dimensional probability distributions Q^N satisfying the power constraint

$$\frac{1}{N}\sum_{k=1}^{N} \int |x_{0,j,k}|^2 dQ^N(x^N) \leq \mathsf{P},\; j = 1, 2, \ldots, n_T. \quad (5)$$

As we shall show below, $C(\mathsf{P})$ and $C_B(\mathsf{P})$ are bounded uniformly in P. By Fano's inequality [13, Sec. 7.9], any encoding and decoding scheme with a rate above capacity has an error probability that is bounded away from zero as $N \to \infty$. By demonstrating that $C(\mathsf{P})$ and $C_B(\mathsf{P})$ are bounded uniformly in P, we therefore demonstrate that there exists no encoding and decoding scheme that has a rate that tends to infinity as $\mathsf{P} \to \infty$ and for which the decoding error probability vanishes as $N \to \infty$.

Our main result is presented in the following theorem.

Theorem 1: Consider the channel model introduced in Section II. For every $\mathsf{P} > 0$ and $0 < p \leq 1$, irrespective of the knowledge of the sequence B_ℓ, $\ell = 1, 2, \ldots$ at the receiver, the channel capacity is upper-bounded by

$$C(\mathsf{P}) \leq C_B(\mathsf{P}) \leq n_R \tfrac{1-p}{p} \log\left(\rho^{-\frac{3}{2}}\right) + \log \frac{\pi}{n_R \Gamma(n_R)} + n_R\left(\log\frac{n_R}{e} + \frac{1}{2}\log\eta_{\max} + \log\left(1+\eta_{\max}\right)\right) \quad (6)$$

where $\eta_{\max} \triangleq \max\left(\frac{1}{\alpha_1}, \frac{1}{\rho}\right)$.

Proof: See Section V. ■

Remark 1: Theorem 1 can be generalized to the case where the autocovariance matrices of the fading coefficients $\{H_{\ell,k},\, k \in \mathbb{Z}\}$ are not scaled identity matrices. Instead, these coefficients have diagonal covariance matrices D_ℓ with arbitrary diagonal elements in $[\underline{d}, \bar{d}]$. If $\bar{d}/\underline{d} < \infty$, then the upper bound (6) is increased by $\frac{3}{2} n_R \log\left(\bar{d}/\underline{d}\right)$, but it remains bounded and independent of P.

IV. Discussion

Lozano, Heath, and Andrews demonstrated in [11] that, without perfect CSI, the information rates of wireless networks achievable with inputs of the form $\sqrt{\mathrm{SNR}}U$ (where U is independent of the SNR) is bounded in the SNR. In [12], it was further shown that, under some additional simplifying assumptions, the capacity remains bounded in the SNR even if the input distribution normalized by $\sqrt{\mathrm{SNR}}$ is allowed to depend on the SNR. Theorem 1 demonstrates that the same is true under less stringent assumptions and even when the interference is bursty. Intuitively, to avoid a bounded capacity, a carefully chosen frequency reuse scheme is crucial to ensure that the interfering cells are super-exponentially far away so that (2) is violated. However, interference burstiness increases the value of the upper bound (6). For example, for $n_R = n_T = 1$ and $\rho = 0.5$, if the interference is always present ($p = 1$), the upper bound (6) becomes 1.59 nats/ch. use. In contrast, for $n_R = n_T = 1$, $\rho = p = 0.5$, it becomes 2.63 nats/ch. use. This also suggests that bursty signaling strategies, which artificially enhance interference burstiness, may increase the transmission rate of noncoherent wireless networks.

Observe that the upper bound presented in Theorem 1 decreases monotonically with the probability of presence of interference p. For $p = n_R = 1$, we recover [12, Th. 1]. Further observe that the upper bound (6) is independent of n_T. The impact of enabling n_T users to cooperate vanishes as the number of interfering cells tends to infinity.

By applying the Stirling series to approximate $\log\Gamma(n_R)$, it can be observed that, for large values of n_R, the upper bound (6) grows linearly with the number of receive antennas n_R. A numerical evaluation of (6) reveals that increasing the number of receive antennas has a larger impact if either the interference burstiness level p or the value of ρ are small. Finally, as $\rho \to 1$, the impact of interference burstiness vanishes.

To summarize, Theorem 1 provides an upper bound on the capacity that is independent of the SNR. For example, for $n_R = n_T = 1$, $\rho = 0.5$, and $p = 1$, the upper bound (6) becomes 1.59 nats/ch. use. If one wishes to communicate above this rate, one has three possibilities: i) increase the number of receive antennas, ii) use bursty signaling to increase interference burstiness or iii) decrease ρ by adapting the frequence reuse scheme.

V. Proof of Theorem 1

The left-most inequality in (6) follows because providing the receiver with information about $\{B_\ell\}$ does not reduce capacity. The proof of the right-most inequality in (6) follows the proofs for the non-bursty SISO case [12, Sec. V] and the bursty SISO case [14, Ch. 5]. We begin by upper-bounding

$$\begin{aligned} &I(\mathbf{X}_{0,1}^N; \mathbf{Y}_1^N, B_1^L) \\ &\stackrel{(a)}{=} h(\mathbf{Y}_1^N | B_1^L) - h(\mathbf{Y}_1^N | \mathbf{X}_{0,1}^N, B_1^L) \\ &\stackrel{(b)}{\leq} h(\mathbf{Y}_1^N | B_1^L) - h(\mathbf{Y}_1^N | \mathbf{X}_{0,1}^N, \mathbb{H}_{0,1}^N, B_1^L) \\ &\stackrel{(c)}{=} h(\mathbf{Y}_1^N | B_1^L) - h(\mathbf{Y}_1^N - \mathbb{H}_{0,1}^N \mathbf{X}_{0,1}^N | B_1^L) \end{aligned} \quad (7)$$

where (a) follows because $\mathbf{X}_{0,1}^N$ and B_1^L are independent, (b) follows because conditioning reduces entropy, and (c) follows because, conditioned on B_1^L, $\mathbf{Y}_1^N - \mathbb{H}_{0,1}^N \mathbf{X}_{0,1}^N$ is independent of $(\mathbb{H}_{0,1}^N, \mathbf{X}_{0,1}^N)$.

For $B_1^L = b_1^L$ and $b_0 \triangleq 1$, we define the random variables

$$\mathbf{Y}_k(b_1^L) \triangleq \sum_{\ell=0}^{L} b_\ell \mathbb{H}_{\ell,k} \mathbf{X}_{\ell,k} + \sum_{\ell=L+1}^{\infty} B_\ell \mathbb{H}_{\ell,k} \mathbf{X}_{\ell,k} + \mathbf{Z}_k \quad (8)$$

$$\bar{\mathbf{Y}}_k(b_1^L) \triangleq \sum_{\ell=1}^{L} b_\ell \bar{\mathbb{H}}_{\ell,k} \mathbf{X}_{\ell,k} + \sum_{\ell=L+1}^{\infty} \bar{B}_\ell \bar{\mathbb{H}}_{\ell,k} \mathbf{X}_{\ell,k} + \bar{\mathbf{Z}}_k. \quad (9)$$

In (9), for every $\ell = 1, 2, \ldots$ the fading coeffxicients $\{\bar{\mathbb{H}}_{\ell,k}, k \in \mathbb{Z}\}$ have the same distribution as $\{\mathbb{H}_{\ell,k}, k \in \mathbb{Z}\}$ but are independent of $\{\mathbb{H}_{\ell,k},\; k \in \mathbb{Z}\}$. Likewise, the random variables $\{\bar{B}_\ell, \ell = 1, 2, \ldots\}$ have the same distribution as $\{B_\ell, \ell = 1, 2, \ldots\}$ but are independent of $\{B_\ell, \ell = 1, 2 \ldots\}$, and the additive noise terms $\{\bar{\mathbf{Z}}_k, k\in\mathbb{Z}\}$ have the same distribution as $\{\mathbf{Z}_k, k\in\mathbb{Z}\}$ but are independent of $\{\mathbf{Z}_k, k\in\mathbb{Z}\}$. It follows that, conditioned on $B_1^L = b_1^L$, the random variable $\mathbf{Y}_1^N - \mathbb{H}_{0,1}^N \mathbf{X}_{0,1}^N$ has the same distribution as $\bar{\mathbf{Y}}_1^N(b_1^L)$. Furthermore, $\mathbf{Y}_1^N(b_1^L)$ and $\bar{\mathbf{Y}}_1^N(b_1^L)$ are both independent of B_1^L. Using the definitions (8)–(9), we can thus rewrite (7) as

$$I(\mathbf{X}_{0,1}^N;\mathbf{Y}_1^N, B_1^L) \le \sum_{b_1^L\in\mathcal{B}_L} \Pr\{B_1^L = b_1^L\} h\big(\mathbf{Y}_1^N(b_1^L)\big) - \sum_{\tilde{b}_1^L\in\mathcal{B}_L} \Pr\{B_1^L = \tilde{b}_1^L\} h\big(\bar{\mathbf{Y}}_1^N(\tilde{b}_1^L)\big) \quad (10)$$

where $\mathcal{B}_L \triangleq \{0,1\}^L$ denotes the set of all binary sequences of length L. We next consider a partition of $\mathcal{B}_L$ based on the position of the leading 1 in each sequence. In particular, for $m = 1,\dots,L+1$, we define

$$\mathcal{B}_L(m) \triangleq \begin{cases} \{b_1^L \in \mathcal{B}_L : b_1^m = [0_1^{m-1}, 1]\}, & 1\le m\le L,\\ \{0_1^L\}, & m = L+1. \end{cases} \quad (11)$$

In words, $\mathcal{B}_L(m)$ is the set of all sequences of length L whose leading 1 is in the m-th position. The sets $\mathcal{B}_L(m)$, $m = 1,\dots,L+1$ are disjoint and define a partition of $\mathcal{B}_L$.

To upper-bound (10), we will pair two sequences b_1^L and $\tilde{b}_1^L$ according to the mapping described in the next proposition.

Proposition 1: There exists a one-to-one and onto mapping $f_L\colon \mathcal{B}_L \to \mathcal{B}_L$ such that, for every $b_1^L \in \mathcal{B}_L$, the vector $\tilde{b}_1^L = f_L(b_1^L)$ lies in $\mathcal{B}_L(m)$ for some m and satisfies $\tilde{b}_1^L = [0_1^{m-1}, 1, b_1^{L-m}]$ and $\|\tilde{b}_1^L\|_1 = \|b_1^L\|_1$. Here $\|\cdot\|_p$, $p\ge 1$ denotes the p-norm.

Proof: See [14, Ch. 5, Prop. 1]. ■

Since $\|b_1^L\|_1 = \|\tilde{b}_1^L\|_1$, it follows that $\Pr\{B_1^L = b_1^L\} = \Pr\{B_1^L = f_L(b_1^L)\}$. Consequently, (10) can be written as

$$I(\mathbf{X}_{0,1}^N;\mathbf{Y}_1^N, B_1^L) \le \sum_{m=1}^{L} \sum_{b_1^L : f_L(b_1^L)\in\mathcal{B}_L(m)} \Pr\{B_1^L = b_1^L\} \times\Big[h\big(\mathbf{Y}_1^N(b_1^L)\big) - h\big(\bar{\mathbf{Y}}_1^N(f_L(b_1^L))\big)\Big]. \quad (12)$$

We next focus on the bracketed term in (12). Let

$$\bar{\mathbf{Y}}_k(b_1^L, m) \triangleq \sum_{\ell=0}^{L-m} b_\ell \bar{\mathbb{H}}_{\ell+m,k}\mathbf{X}_{\ell,k} + \sum_{\ell=L-m+1}^{\infty} \bar{B}_\ell \bar{\mathbb{H}}_{\ell+m,k}\mathbf{X}_{\ell,k} + \bar{\mathbf{Z}}_k. \quad (13)$$

Since $\mathbf{X}_{0,1}^N$ and $\mathbf{X}_{\ell,1}^N$, $\ell = 1,2,\dots$ have the same distribution, it follows that $\bar{\mathbf{Y}}_k(b_1^L, m)$ and $\bar{\mathbf{Y}}_k(f_L(b_1^L))$ have the same distribution for every $f_L(b_1^L)\in\mathcal{B}_L(m)$. Hence, using the identity $h(U) - h(V) = h(U|V) - h(V|U)$, we obtain

$$\begin{aligned} &h\big(\mathbf{Y}_1^N(b_1^L)\big) - h\big(\bar{\mathbf{Y}}_1^N(f_L(b_1^L))\big)\\ &= h\big(\mathbf{Y}_1^N(b_1^L) \,\big|\, \bar{\mathbf{Y}}_1^N(b_1^L,m)\big) - h\big(\bar{\mathbf{Y}}_1^N(b_1^L,m) \,\big|\, \mathbf{Y}_1^N(b_1^L)\big)\\ &\le \sum_{k=1}^{N}\Big[h\big(\mathbf{Y}_k(b_1^L) \,\big|\, \bar{\mathbf{Y}}_k(b_1^L,m)\big) - h\big(\bar{\mathbf{Y}}_k(b_1^L,m) \,\big|\, \bar{\mathbf{Y}}_1^{k-1}(b_1^L,m), \mathbf{Y}_1^N(b_1^L)\big)\Big] \end{aligned} \quad (14)$$

where the inequality follows from the chain rule and because conditioning reduces entropy. To upper-bound (14), we next apply the following lemma.

Lemma 1: Let f and g be two arbitrary pdf. If $-\int f(x)\log f(x)\mathrm{d}x$ is finite, then $-\int f(x)\log g(x)\mathrm{d}x$ exists and

$$-\int f(x)\log f(x)\mathrm{d}x \le -\int f(x)\log g(x)\mathrm{d}x. \quad (15)$$

Proof: See [15, Lemma 8.3.1]. ■

Let $f_{\mathbf{Y}_k|\bar{\mathbf{Y}}_k}$ denote the true conditional pdf of $\mathbf{Y}_k(b_1^L)$ given $\bar{\mathbf{Y}}_k(b_1^L, m)$. Lemma 1 allows us to upper-bound the conditional differential entropy of $\mathbf{Y}_k(b_1^L)$ given $\bar{\mathbf{Y}}_k(b_1^L,m)$ by replacing $f_{\mathbf{Y}_k|\bar{\mathbf{Y}}_k}$ by an auxiliary pdf $g_{\mathbf{Y}_k|\bar{\mathbf{Y}}_k}$. For every $\bar{\mathbf{Y}}_k(b_1^L,m) = \bar{\mathbf{y}}_k$, we choose

$$g_{\mathbf{Y}_k|\bar{\mathbf{Y}}_k}(\mathbf{y}_k|\bar{\mathbf{y}}_k) = \frac{n_R\sqrt{\beta}\,\Gamma(n_R)}{\pi^{n_R+1}\|\mathbf{y}_k\|_2^{n_R}} \frac{1}{1+\beta\|\mathbf{y}_k\|_2^{2n_R}} \quad (16)$$

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 [12] for the SISO case and in [16] to analyze frequency-dispersive fading channels.

Using (16) in (15), and since $a^n + b^n \le (a+b)^n$ for $a, b, \ge 0$,

$$\begin{aligned} &h\big(\mathbf{Y}_k(b_1^L) \,\big|\, \bar{\mathbf{Y}}_k(b_1^L,m)\big)\\ &\le (n_R+1)\log\pi - \log n_R - \log\Gamma(n_R)\\ &\quad + \frac{n_R}{2}\Big(\mathsf{E}\big[\log\|\mathbf{Y}_k(b_1^L)\|_2^2\big] - \mathsf{E}\big[\log\|\bar{\mathbf{Y}}_k(b_1^L,m)\|_2^2\big]\Big)\\ &\quad + n_R\mathsf{E}\big[\log(\|\mathbf{Y}_k(b_1^L)\|_2^2 + \|\bar{\mathbf{Y}}_k(b_1^L,m)\|_2^2)\big]. \end{aligned} \quad (17)$$

Next, we consider the second conditional entropy in (14). By conditioning on $\{\mathbf{X}_{\ell,k}\}_{\ell=1}^{\infty}$ and $\{\bar{B}_\ell\}_{\ell=L-m+1}^{\infty}$, the random variable $\bar{\mathbf{Y}}_k(b_1^L,m)$ is independent of $(\bar{\mathbf{Y}}_1^{k-1}(b_1^L,m), \mathbf{Y}_1^N(b_1^L))$ and has a Gaussian distribution. Hence,

$$\begin{aligned} &h\big(\bar{\mathbf{Y}}_k(b_1^L,m) \,\big|\, \bar{\mathbf{Y}}_1^{k-1}(b_1^L,m), \mathbf{Y}_1^N(b_1^L)\big)\\ &\ge h\big(\bar{\mathbf{Y}}_k(b_1^L,m) \,\big|\, \{\mathbf{X}_{\ell,k}\}_{\ell=1}^{\infty}, \bar{B}_{L-m+1}^{\infty}\big)\\ &= n_R\log(\pi e) + n_R\mathsf{E}\big[\log\bar{\mathsf{K}}(b_1^L, \bar{B}_{L-m+1}^{\infty}, \mathbf{X}_{0,k}^{\infty})\big] \end{aligned} \quad (18)$$

where

$$\bar{\mathsf{K}}(b_1^L, \bar{B}_{L-m+1}^{\infty}, \mathbf{X}_{0,k}^{\infty}) \triangleq \sum_{\ell=0}^{L-m} b_\ell\alpha_{\ell+m}\|\mathbf{X}_{\ell,k}\|_2^2 + \sum_{\ell=L-m+1}^{\infty} |\bar{B}_\ell|^2\alpha_{\ell+m}\|\mathbf{X}_{\ell,k}\|_2^2 + \sigma^2. \quad (19)$$

Using (17) and (18), and simplifying terms, we obtain that (14) can be upper-bounded by

$$\begin{aligned} \sum_{k=1}^{N}\bigg(&\log\frac{\pi}{n_R\Gamma(n_R)} - n_R\log e\\ &+ \frac{n_R}{2}\Big(\mathsf{E}\big[\log\|\mathbf{Y}_k(b_1^L)\|_2^2\big] - \mathsf{E}\big[\log\|\bar{\mathbf{Y}}_k(b_1^L,m)\|_2^2\big]\Big)\\ &+ n_R\mathsf{E}\big[\log(\|\mathbf{Y}_k(b_1^L)\|_2^2 + \|\bar{\mathbf{Y}}_k(b_1^L,m)\|_2^2)\big]\\ &- n_R\mathsf{E}\Big[\log\bar{\mathsf{K}}(b_1^L, \bar{B}_{L-m+1}^{\infty}, \mathbf{X}_{0,k}^{\infty})\Big]\bigg). \end{aligned} \quad (20)$$

We upper-bound the third line in (20) by applying Jensen's inequality to the concave function $\log(\cdot)$:

$$\begin{aligned}&\mathsf{E}\left[\log\left(\|\mathbf{Y}_k(b_1^L)\|_2^2+\|\bar{\mathbf{Y}}_k(b_1^L,m)\|_2^2\right)\right]\\&\leq\mathsf{E}\Big[\log\Big(\mathsf{E}\Big[\|\mathbf{Y}_k(b_1^L)\|_2^2\\&\qquad+\|\bar{\mathbf{Y}}_k(b_1^L,m)\|_2^2\Big|\{\mathbf{X}_{\ell,k},\ell=0,1,\ldots\},B_1^\infty,\bar{B}_1^\infty\Big]\Big)\Big]\\&=\mathsf{E}\left[\log\left(\sum_{\ell=0}^{L}b_\ell\alpha_\ell\|\mathbf{X}_{\ell,k}\|_2^2+\sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell\|\mathbf{X}_{\ell,k}\|_2^2\right.\right.\\&\qquad+\sum_{\ell=0}^{L-m}b_\ell\alpha_{\ell+m}\|\mathbf{X}_{\ell,k}\|_2^2\\&\qquad\left.\left.+\sum_{\ell=L-m+1}^{\infty}|\bar{B}_\ell|^2\alpha_{\ell+m}\|\mathbf{X}_{\ell,k}\|_2^2+2\sigma^2\right)\right]+\log n_R\end{aligned}\tag{21}$$

where $\alpha_0\triangleq 1$. We then use that assumption (2) implies that

$$\alpha_\ell\leq\frac{\eta_{\max}}{\rho^{m-1}}\alpha_{\ell+m},\quad \ell=0,1,\ldots\tag{22}$$

This gives

$$\begin{aligned}\sum_{\ell=0}^{L-m}b_\ell\alpha_\ell\|\mathbf{X}_{\ell,k}\|_2^2\leq\frac{\eta_{\max}}{\rho^{m-1}}&\left(\sum_{\ell=0}^{L-m}b_\ell\alpha_{\ell+m}\|\mathbf{X}_{\ell,k}\|_2^2\right.\\&\left.+\sum_{\ell=L-m+1}^{\infty}|\bar{B}_\ell|^2\alpha_{\ell+m}\|\mathbf{X}_{\ell,k}\|_2^2\right).\end{aligned}\tag{23}$$

It follows that

$$\begin{aligned}&\mathsf{E}\left[\log\left(\|\mathbf{Y}_k(b_1^L)\|_2^2+\|\bar{\mathbf{Y}}_k(b_1^L,m)\|_2^2\right)\right]\\&\leq\mathsf{E}\left[\log\left(\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right)\bar{\mathsf{K}}(b_1^L,\bar{B}_{L-m+1}^\infty,\mathbf{X}_{0,k}^\infty)\right)\right]\\&\quad+\log n_R+\zeta_{L,m}\end{aligned}\tag{24}$$

where

$$\begin{aligned}\zeta_{L,m}\triangleq&\sum_{\ell=L-m+1}^{L}\frac{\alpha_\ell b_\ell\mathsf{E}[\|\mathbf{X}_{\ell,k}\|_2^2]}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\\&+\sum_{\ell=L+1}^{\infty}\frac{\alpha_\ell\mathsf{E}[|B_\ell|^2\|\mathbf{X}_{\ell,k}\|_2^2]}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\end{aligned}\tag{25}$$

and where $\bar{\mathsf{K}}(b_1^L,\bar{B}_{L-m+1}^\infty,\mathbf{X}_{0,k}^\infty)$ was defined in (19).

To upper-bound the second line in (20), we note that, conditioned on $\mathbf{X}_{\ell,k}=\mathbf{x}_\ell$, $B_\ell=b_\ell$, and $\bar{B}_\ell=\bar{b}_\ell$, $\ell=0,1,\ldots$, both $\|\mathbf{Y}_k(b_1^L)\|_2^2$ and $\|\bar{\mathbf{Y}}_k(b_1^L,m)\|_2^2$ have a chi-square distribution with $2n_R$ degrees of freedom. Using [17, Eq. 4.352], we thus obtain that

$$\begin{aligned}&\mathsf{E}\left[\log\|\mathbf{Y}_k(b_1^L)\|_2^2\right]-\mathsf{E}\left[\log\|\bar{\mathbf{Y}}_k(b_1^L,m)\|_2^2\right]\\&=\mathsf{E}\left[\log\left(\frac{\mathsf{K}(b_1^L,B_{L+1}^\infty,\mathbf{X}_{0,k}^\infty)}{\bar{\mathsf{K}}(b_1^L,\bar{B}_{L-m+1}^\infty,\mathbf{X}_{0,k}^\infty)}\right)\right]\end{aligned}\tag{26}$$

where

$$\begin{aligned}&\mathsf{K}(b_1^L,B_{L+1}^\infty,\mathbf{X}_{0,k}^\infty)\\&\triangleq\sum_{\ell=0}^{L}b_\ell\alpha_\ell\|\mathbf{X}_{\ell,k}\|_2^2+\sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell\|\mathbf{X}_{\ell,k}\|_2^2+\sigma^2.\end{aligned}\tag{27}$$

To upper-bound (26), we use (22) and follow similar steps as in (23)–(24) to obtain that

$$\begin{aligned}&\mathsf{E}\left[\log\|\mathbf{Y}_k(b_1^L)\|_2^2\right]-\mathsf{E}\left[\log\|\bar{\mathbf{Y}}_k(b_1^L,m)\|_2^2\right]\\&\leq\log\left(\frac{\eta_{\max}}{\rho^{m-1}}\right)+\zeta_{L,m}.\end{aligned}\tag{28}$$

Combining (24) and (28) with (20) and (14), it follows that

$$\begin{aligned}&h\big(\mathbf{Y}_1^N(b_1^L)\big)-h\big(\bar{\mathbf{Y}}_1^N(f_L(b_1^L))\big)\\&\leq N\Big[n_R(m-1)\log\left(\rho^{-\frac{3}{2}}\right)+\log\frac{\pi}{n_R\Gamma(n_R)}\\&\qquad+n_R\log\frac{n_R}{e}+\frac{n_R}{2}\log(\eta_{\max})+n_R\log(1+\eta_{\max})\Big]\\&\quad+\frac{3}{2}Nn_R\zeta_{L,m}.\end{aligned}\tag{29}$$

The term $\zeta_{L,m}$, defined in (25), can be upper-bounded as

$$\begin{aligned}\zeta_{L,m}&\leq\sum_{\ell=L-m+1}^{L}\left(\frac{\alpha_\ell n_T\mathsf{P}}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right)+\sum_{\ell=L+1}^{\infty}\left(\frac{\alpha_\ell p n_T\mathsf{P}}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right)\\&\triangleq\bar{\zeta}_{L,m}\end{aligned}\tag{30}$$

by using that $\mathsf{E}\left[|B_\ell|^2\right]=\mathsf{E}\left[|\bar{B}_\ell|^2\right]=p$ and $b_\ell\leq 1$, and by the power constraint (5).

Back to (12), using (29) and (30), we obtain that

$$\begin{aligned}&\frac{1}{N}I(\mathbf{X}_{0,1}^N;\mathbf{Y}_1^N,B_1^L)\\&\leq\sum_{m=1}^{L}p(1-p)^{m-1}\left[n_R(m-1)\log\left(\rho^{-\frac{3}{2}}\right)+\log\frac{\pi}{n_R\Gamma(n_R)}\right.\\&\qquad\left.+n_R\log\frac{n_R}{e}+\frac{n_R}{2}\log(\eta_{\max})+n_R\log(1+\eta_{\max})\right]\\&\quad+\frac{3}{2}\sum_{m=1}^{L}p(1-p)^{m-1}n_R\bar{\zeta}_{L,m}\end{aligned}\tag{31}$$

since the first two lines on the RHS of (29) and $\bar{\zeta}_{L,m}$ do not depend on b_1^L, and since

$$\Pr\{f_L(B_1^L)\in\mathcal{B}_L(m)\}=p(1-p)^{m-1}.\tag{32}$$

The first sum on the RHS of (31) converges to

$$\begin{aligned}&n_R\tfrac{1-p}{p}\log\left(\rho^{-\frac{3}{2}}\right)+\log\frac{\pi}{n_R\Gamma(n_R)}+n_R\log\frac{n_R}{e}\\&+\frac{n_R}{2}\log\eta_{\max}+n_R\log\left(1+\eta_{\max}\right)\end{aligned}\tag{33}$$

as $L\to\infty$, since

$$\sum_{m=1}^{\infty}p(1-p)^{m-1}=1\tag{34}$$

and [17, Eq. 0.231]

$$\sum_{m=1}^{\infty}p(1-p)^{m-1}(m-1)=\frac{1-p}{p}.\tag{35}$$

The second sum on the RHS of (31) vanishes as $L\to\infty$ by using that $\sum_{\ell=1}^{\infty}\alpha_\ell<\infty$ and by following similar steps as in [14, Ch. 5]. Since (33) does neither depend on N nor on the input distribution, Theorem 1 follows.

REFERENCES

[1] Ericsson, "Ericsson mobility report," Nov. 2019, [Online]. Available: https://www.ericsson.com/4acd7e/assets/local/mobility-report/documents/2019/emr-november-2019.pdf.

[2] 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.

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

[4] 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.

[5] 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.

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

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

[8] ——, "Harnessing bursty interference," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Jun. 2009, pp. 13–16.

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

[10] G. Villacrés, T. Koch, A. Sezgin, and G. Vazquez-Vilar, "Robust signaling for bursty interference," *Entropy*, vol. 20, no. 11:870, Nov. 2018.

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

[12] G. Villacrés and T. Koch, "Wireless networks of bounded capacity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2584–2588.

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

[14] G. Villacrés, "Capacity limits of bursty interference channels," Ph.D. dissertation, Universidad Carlos III de Madrid, Apr. 2019, [Online]. Available: https://tinyurl.com/tglva5u.

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

[16] 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.

[17] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed. Elsevier/Academic Press, Amsterdam, 2007.