On the High-SNR Capacity of Noncoherent Networks

Amos Lapidoth, *Fellow, IEEE*

Dedicated to "VJ"
יהודה וולפרט
(1958–2005)

***Abstract*—We obtain the first term in the high signal-to-noise ratio (SNR) asymptotic expansion of the sum-rate capacity of noncoherent fading networks, i.e., networks where the transmitters and receivers—while fully cognizant of the fading *law*—have no access to the fading *realization*. This term is an integer multiple of log log SNR with the coefficient having a simple combinatorial characterization. It can be interpreted as the effective number of parallel channels that can be supported by the network, i.e., as the maximal number of point-to-point single-user scalar channels that can be supported by the network in a manner that will allow, with proper power allocation, negligible cross interference. The results hold irrespective of whether the transmitters can cooperate or must operate in an multiple-access regime; irrespective of whether feedback from the receivers to the transmitters is available or not; and irrespective of whether the receivers can cooperate or not.**

***Index Terms*—Channel capacity, fading, high signal-to-noise ratio (SNR), memory, multiple-antenna network, noncoherent.**

I. INTRODUCTION

In this paper, we consider a discrete-time vector fading channel, where the transmitted vector suffers from both multiplicative and additive noises. The multiplicative noise takes the form of a matrix-valued stationary and ergodic process that multiplies the transmitted vector, and the additive noise takes the form of independent and identically distributed (i.i.d.) isotropic Gaussian vectors. We only consider the case where the realization of neither the additive noise nor of the multiplicative noise are known to the transmitter and receiver; only their probability laws are given. The mathematical model that we address is thus very similar to the "noncoherent" flat-fading multiple-antenna channel model.

There is, however, an important difference. In the multiple-antenna channel model we think of the components of the transmitted vector as being the signals transmitted by colocated antennas. Similarly, the components of the received vectors are viewed as the signals received at colocated antennas. Our model is more general. We can think of the different components of the input vector as being controlled by a single user as in a single-user multiple-antenna communication scenario, but we can also think of each component as being controlled by different geographically separated users as, for example, in a multiple-user network where each of the users employs a single transmit antenna. We can also envision that the components of the transmitted vector are partitioned into disjoint groups where the different groups are controlled by geographically separated users. This corresponds to a network where the different geographically separated users may employ multiple transmit antennas of various numbers. Finally, in our setup, the different components of the input vector need not correspond to physically different transmit antennas. We can also envision a scenario where they correspond to transmissions taking place at different frequencies and/or times as in a network employing a slotted protocol. Analogous scenarios can be envisioned for the received vector.

The various scenarios mentioned above differ not only in the allowed dependencies between the different components of the transmitted vector. It turns out that, at high signal-to-noise ratio (SNR), far more important is the structure of the multiplicative noise that they imply. For example, if a certain receive antenna and a certain transmit antenna operate at different time/frequency slots, then the corresponding component in the multiplicative noise matrix will be deterministically zero. A similar situation occurs when a given transmitter is geographically very far apart from a given receiver as could, for example, be the case in a cellular system. For example, in Wyner's linear cellular model [1], [2] the transmitters are assumed to be uniformly spaced on a line, and each transmitter is received by only two base stations: the base station to its left and the base station to its right.

As we shall see, rather than the cooperation restrictions, it is these deterministic zeros (and the interference that their lack implies) that will determine the high-SNR asymptotic behavior of channel capacity. Very roughly, the main result of this paper is that, irrespective of the cooperation allowed, at high SNR the capacity C of the channel is given approximately by

$$C \approx \kappa^* \cdot \log\log \mathrm{SNR} \tag{1}$$

where the nonnegative integer κ^* can be computed combinatorially from the zeros of the multiplicative noise. Thus, rather than the number of antennas or the fine structure of the fading correlations, it is the network's topology and its frequency/time reuse pattern that determine—via the zeros in the fading matrix that they induce—the high-SNR asymptotics of the network's capacity.

Manuscript received November 16, 2004; revised April 14, 2005. The material in this papers was presented in part at the Third Joint Workshop on Communications and Coding (JWCC'04), Donnini–Firenze, Italy, October 2004.

The author is with the Department of Information Technology and Electrical Engineering, Swiss Federal Institute of Technology (ETH), CH-8092 Zurich, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch).

Communicated by Y. Steinberg, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2005.853332

This asymptotic expansion is proved by establishing an upper bound on the network sum–rate capacity ("converse") and an achievable lower bound ("direct part"). The upper bound is proved assuming that the transmitters can cooperate in full, i.e., that they are controlled by a single user and assuming that a decoder has access to all the received signals. The lower bound is proved using κ^* independent scalar transmitters and κ^* single-user scalar detectors. The bounds are fairly general and only require that the nonzero components of the multiplicative noise be stationary and ergodic with finite variances and finite joint differential entropy rate. They need not be Gaussian. See Theorem 1 for details.

The preceding result can be viewed as an extension of a result of [3] on multiple-antenna fading channels. In the multiple-antenna scenario, where the components of the transmitted vector are geographically colocated and where the components of the received vector are also colocated, there are typically no deterministic zeros in the fading matrix. In this case, it can be readily verified that our combinatorial expression for κ^* yields the value of 1, thus recovering the $1 \cdot \log\log \mathrm{SNR}$ asymptotics of [3]. Similarly, the "joint isotropic fading" assumption of [4] implies that deterministic zeros in the fading matrix cannot appear in isolation. If a component is deterministically zero, then so must be its entire row. Under this condition $\kappa^* = 1$ as well.

The rest of the this paper is organized as follows. In the next section, we describe the channel model, state the main result, and discuss some examples. In Section III, we provide a proof, and in the final section, Section IV, we summarize our results and discuss some possible extensions.

II. Channel Model and Main Result

The channel we consider is a discrete-time channel where the time-k channel input $\boldsymbol{x}_k \in \mathbb{C}^{n_\mathrm{T}}$ is an n_T-dimensional complex vector, where $k \in \mathbb{Z}$ is a discrete-time index taking value in the integers $\mathbb{Z}$; n_T is a positive integer; $\mathbb{C}$ denotes the complex field; and $\mathbb{C}^{n_\mathrm{T}}$ denotes the n_T-dimensional complex Euclidean space. We refer to n_T as the number of transmitters, and to the set

$$\mathcal{T} = \{1, \ldots, n_\mathrm{T}\} \tag{2}$$

as the set of transmitters. For every $t \in \mathcal{T}$ we denote the tth component of the time-k input vector $\boldsymbol{x}_k$ by $x_k(t)$. This corresponds to the signal transmitted at time k by transmitter t. The time-k channel output $\boldsymbol{Y}_k \in \mathbb{C}^{n_\mathrm{R}}$ corresponding to the input $\boldsymbol{x}_k$ is given by

$$\boldsymbol{Y}_k = \mathbb{H}_k \boldsymbol{x}_k + \boldsymbol{Z}_k \tag{3}$$

where n_R is a positive integer that denotes the number of receive antennas and where

$$\mathcal{R} = \{1, \ldots, n_\mathrm{R}\} \tag{4}$$

denotes the set of receivers. In the above, $\{\mathbb{H}_k\}$ is a matrix-valued stochastic process such that at every time instant k the random matrix $\mathbb{H}_k$ is an $n_\mathrm{R} \times n_\mathrm{T}$ complex random matrix, and the random vectors $\{\boldsymbol{Z}_k\}$ are i.i.d., each taking value in $\mathbb{C}^{n_\mathrm{R}}$ according to an isotropic circularly symmetric multivariate complex Gaussian law

$$\boldsymbol{Z}_k \sim \mathcal{N}_\mathbb{C}(\mathbf{0}, \mathsf{I}_{n_\mathrm{R}}) \tag{5}$$

where $\mathsf{I}_{n_\mathrm{R}}$ denotes the $n_\mathrm{R} \times n_\mathrm{R}$ identity matrix. (In general, $\boldsymbol{W} \sim \mathcal{N}_\mathbb{Z}(\boldsymbol{\mu}, \Lambda)$ indicates that $\boldsymbol{W} - \boldsymbol{\mu}$ is a zero-mean circularly symmetric complex Gaussian random vector of covariance matrix Λ.) We assume throughout that the processes $\{\mathbb{H}_k\}$ and $\{\boldsymbol{Z}_k\}$ are independent and that their joint law does not depend on the input sequence $\{\boldsymbol{x}_k\}$. Denoting by $H_k(r,t)$ the row-r column-t entry of the matrix $\mathbb{H}_k$, and denoting by $Z_k(r)$ the rth component of the time-k additive noise vector $\boldsymbol{Z}_k$, we can rewrite (3) as

$$Y_k(r) = \sum_{t \in \mathcal{T}} H_k(r,t) x_k(t) + Z_k(r), \qquad r \in \mathcal{R}. \tag{6}$$

To account for the possibility that some of the components of the fading matrices might be deterministically zero we introduce the set $\mathcal{Z}$

$$\mathcal{Z} \subset \mathcal{R} \times \mathcal{T} \tag{7}$$

where if $(r,t) \in \mathcal{Z}$ then $H_k(r,t)$ is deterministically zero at all times $k \in \mathbb{Z}$

$$(r,t) \in \mathcal{Z} \Rightarrow (H_k(r,t) = 0, \forall k \in \mathbb{Z}). \tag{8}$$

As for the other components, we shall assume a finite second moment

$$\mathsf{E}[|H_k(r,t)|^2] < \infty, \qquad (r,t) \in \mathcal{R} \times \mathcal{T} \tag{9}$$

and a finite differential entropy rate condition that we next describe. But first we introduce some notation. Given a collection of random variables $\{W(\alpha)\}_{\alpha \in \mathcal{A}}$ indexed by a set $\mathcal{A}$, we denote, for any subset $\mathcal{B} \subseteq \mathcal{A}$, by $W(\mathcal{B})$ the unordered collection $\{W(\alpha)\}_{\alpha \in \mathcal{B}}$. With this notation and (7) we have that $H_k(\mathcal{Z}^\mathrm{c})$ is the collection of $|\mathcal{Z}^\mathrm{c}|$ $(= n_\mathrm{R} \cdot n_\mathrm{T} - |\mathcal{Z}|)$ random variables

$$H_k(\mathcal{Z}^\mathrm{c}) = \{H_k(r,t) : (r,t) \notin \mathcal{Z}\} \tag{10}$$

where we use $\mathcal{Z}^\mathrm{c}$ to denote the set complement of $\mathcal{Z}$ in $\mathcal{R} \times \mathcal{T}$ and we use $|\cdot|$ to denote set cardinality. The finite differentiable entropy rate condition that we require can be now stated as

$$h(\{H_k(\mathcal{Z}^\mathrm{c})\}_{k \in \mathbb{Z}}) > -\infty. \tag{11}$$

In the case where $\{\mathbb{H}_k\}$ is i.i.d., this condition translates to the joint differential entropy of the $(n_\mathrm{R} \cdot n_\mathrm{T} - |\mathcal{Z}|)$ random variables $\{H_k(r,t), (r,t) \notin \mathcal{Z}\}$ being finite. In the more general case, (11) can be written as

$$\lim_{n \to \infty} \frac{1}{n} h(H_1(\mathcal{Z}^\mathrm{c}), \ldots, H_n(\mathcal{Z}^\mathrm{c})) > -\infty \tag{12}$$

or even more explicitly as

$$\lim_{n \to \infty} \frac{1}{n} h(\{H_k(r,t)\}, 1 \le k \le n, (r,t) \notin \mathcal{Z}) > -\infty. \tag{13}$$

Notice that a stationary process $\{\mathbb{H}_k\}$ simultaneously satisfies (9) and (11) if, and only if, it simultaneously satisfies (9) and the two conditions

$$h(\mathbb{H}_1(\mathcal{Z}^c)) > -\infty \quad \text{and} \quad \lim_{k\to\infty} I(\mathbb{H}_1,\ldots,\mathbb{H}_{k-1};\mathbb{H}_k) < \infty. \tag{14}$$

We denote by $C_{\mathrm{SU}}(\mathcal{E})$ the capacity of this channel under full cooperation conditions when the input is subjected to the average-power constraint $\mathcal{E}$. That is,

$$C_{\mathrm{SU}}(\mathcal{E}) = \lim_{n\to\infty} \frac{1}{n} \sup I(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n;\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n) \tag{15}$$

where the supremum is over all joint distributions on $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ satisfying

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}[\,\|\boldsymbol{X}_k\|^2] \leq \mathcal{E}$$

where $\|\boldsymbol{X}_k\|$ is the Euclidean norm of the vector $\boldsymbol{X}_k$. This is thus the capacity when a single user controls the input vector $\boldsymbol{x}_k \in \mathbb{C}^{n_{\mathrm{T}}}$, and when a "super-receiver" has access to all the components of the received vector $\boldsymbol{Y}_k$. Similarly, we define $C_{\mathrm{SU,FB}}(\mathcal{E})$ as the single-user capacity but when there is a noiseless feedback link so that the time-k transmitted signal $\boldsymbol{X}_k$ is allowed to depend not only on the message to be transmitted but also on all the past channel outputs. Clearly

$$C_{\mathrm{SU}}(\mathcal{E}) \leq C_{\mathrm{SU,FB}}(\mathcal{E})$$

because the feedback link can always be ignored.

At the other extreme, we define $C_{\mathrm{MAC}}(\mathcal{E})$ as the sum-rate capacity for this channel when it is viewed as a multiple-access channel (MAC) where the different components of the input vector are viewed as separate users who wish to communicate independent messages. Each user is assumed to be allowed a peak power of $\mathcal{E}$ and no feedback link is available. The assumption of a "super-receiver" continues to hold. (We shall later see that this assumption can be significantly relaxed.) We thus have

$$C_{\mathrm{MAC}}(\mathcal{E}) \leq C_{\mathrm{SU,FB}}(n_{\mathrm{T}}\cdot\mathcal{E}). \tag{16}$$

To state the paper's main theorem we need to introduce the notion of a "power chain." To define this concept we introduce the following notation. For any transmitter $t\in\mathcal{T}$ let $\mathcal{R}_t$ be the set of receivers that can "hear" it, i.e.,

$$\mathcal{R}_t = \{r\in\mathcal{R} : (r,t)\notin\mathcal{Z}\}\,. \tag{17}$$

Analogously, for any receiver $r\in\mathcal{R}$, let $\mathcal{T}_r$ denote the set of transmitters that r "hears"

$$\mathcal{T}_r = \{t\in\mathcal{T} : (r,t)\notin\mathcal{Z}\}. \tag{18}$$

Definition 1: We shall say that the κ-tuple $(t_1,\ldots,t_\kappa)\in\mathcal{T}^\kappa$ is a κ-length power chain with respect to the set $\mathcal{Z}$ if

$$\mathcal{R}_{t_1} \neq \emptyset \tag{19}$$

and

$$\mathcal{R}_{t_\nu} \Big\backslash \bigcup_{1\leq\eta<\nu} \mathcal{R}_{t_\eta} \neq \emptyset\,, \qquad \nu=2,\ldots,\kappa. \tag{20}$$

We can now state the paper's main result.

Theorem 1: Consider a vector fading channel (3) whose input takes value in $\mathbb{C}^{n_{\mathrm{T}}}$ and whose output takes value in $\mathbb{C}^{n_{\mathrm{R}}}$. Let the set $\mathcal{Z}\subset\mathcal{R}\times\mathcal{T}$ be given, where $\mathcal{R}$ and $\mathcal{T}$ are defined in (4) and (2), respectively. Assume that the stationary and ergodic matrix-valued fading process $\{\mathbb{H}_k\}$ satisfies (8), (9), and (11). Further assume that $\{\boldsymbol{Z}_k\}$ are i.i.d. according to (5), that the process $\{\boldsymbol{Z}_k\}$ is independent of $\{\mathbb{H}_k\}$, and that their joint law does not depend on the channel input sequence $\{\boldsymbol{x}_k\}$. Let $C_{\mathrm{SU,FB}}(\mathcal{E})$, and $C_{\mathrm{MAC}}(\mathcal{E})$ be defined as above. Then

$$\overline{\lim_{\mathcal{E}\to\infty}}\,\{C_{\mathrm{SU,FB}}(\mathcal{E}) - \kappa^*\log\log\mathcal{E}\} < \infty \tag{21}$$

and

$$\overline{\lim_{\mathcal{E}\to\infty}}\,\{\kappa^*\log\log\mathcal{E} - C_{\mathrm{MAC}}(\mathcal{E})\} < \infty \tag{22}$$

where $\kappa^* = \kappa^*(n_{\mathrm{T}}, n_{\mathrm{R}}, \mathcal{Z})$ is the length of the longest power chain with respect to $\mathcal{Z}$.

Moreover, (22) is achievable with κ^* single-user scalar detectors. That is, there exist transmitters $t_1,\ldots,t_{\kappa^*}\in\mathcal{T}$; receivers $r_1,\ldots,r_{\kappa^*}\in\mathcal{R}$; and distributions for $\boldsymbol{X}$ under which the components of $\boldsymbol{X}$ are independent, under which the peak constraints $|X(t)|^2\leq\mathcal{E}$, $t\in\mathcal{T}$ are satisfied almost surely, and such that

$$\overline{\lim_{\mathcal{E}\to\infty}}\left\{\kappa^*\log\log\mathcal{E} - \sum_{\nu=1}^{\kappa^*} I(X(t_\nu);Y(r_\nu))\right\} < \infty. \tag{23}$$

Note that since $\log\log(a\xi) - \log\log\xi$ converges to zero as $\xi\to\infty$ with $a>0$ held fixed, it follows from (16) and from the theorem that, under the theorem's assumptions

$$\overline{\lim_{\mathcal{E}\to\infty}}\,\{|C_{\mathrm{SU,FB}}(\mathcal{E}) - C_{\mathrm{MAC}}(\mathcal{E})|\} < \infty. \tag{24}$$

Consequently, we can loosely say that, at high SNR, the capacity of a fading network is given by (1), where $\kappa^* = \kappa^*(n_{\mathrm{T}}, n_{\mathrm{R}}, \mathcal{Z})$, irrespective of whether we impose individual peak-power constraints or whether we impose combined average-power constraints, irrespective of whether we allow cooperation between the transmitters or not, irrespective of whether feedback is available or not, irrespective of whether the receivers can cooperate or not, and irrespective of the precise law of the fading process (subject to (8), (9), and (11)).

Before proceeding to prove this theorem in the next section, we next present some examples. In these examples, we denote all generic nonzero entries of the fading matrix by H. It should, however, be understood that the entries that are denoted by H are different and that they have finite joint differential entropy rate.

1) A single-input single-output (SISO) channel can be described by the 1×1 matrix

$$\mathbb{H} = (H). \tag{25}$$

In this case, the only power chain is the length-one chain (1), and the pre-loglog κ^* is thus one.

2) Similarly, a 2×2 single-user multiple-input multiple-output (MIMO) channel can be described by the 2×2 matrix

$$\mathbb{H} = \begin{pmatrix} H & H \\ H & H \end{pmatrix}. \tag{26}$$

In this case, there are two power chains, namely, the power chain (1), and the power chain (2). Both are of length one and hence $\kappa^* = 1$.

3) The 1×2 case

$$\mathbb{H} = (H \quad H) \tag{27}$$

can describe a multiple-input single-output (MISO) channel or a two-user scalar MAC. In either case, there are only two power chains: the chain (1) and the chain (2), so that $\kappa^* = 1$.

4) The case of a "dangling transmitter" can be described by

$$\mathbb{H} = \begin{pmatrix} H & 0 \\ H & 0 \end{pmatrix}. \tag{28}$$

Here, the signal transmitted by Transmitter 2 is not heard by any receiver. The only power chain is (1), which is of length one.

5) Similarly, the case of a "dangling receiver" can be described by

$$\mathbb{H} = \begin{pmatrix} H & H \\ 0 & 0 \end{pmatrix}. \tag{29}$$

Here, Receiver 2 does not hear any of the transmitted signals. Its signal is additive noise only, and it is thus irrelevant. Consequently, we expect κ^* to equal one. This is indeed the case because the only power chains are the two length-one power chains (1) and (2).

6) An example of a network consisting of two noninterfering single-user channels (one of which is SISO and the other MIMO) is

$$\mathbb{H} = \begin{pmatrix} H & 0 & 0 \\ 0 & H & H \\ 0 & H & H \end{pmatrix}. \tag{30}$$

Here, the longest power chains are of length two: $(1, 2)$ and $(1, 3)$.

7) An example that is good to have in mind in studying the direct part of the theorem is one where

$$\mathbb{H} = \begin{pmatrix} H & H \\ 0 & H \end{pmatrix} \tag{31}$$

is upper triangular. In fact, one can view the the different power chains as different ways of upper triangularizing the channel's fading matrix.

A scenario that might yield such a matrix is one where the transmitters and receivers are placed on the integer lattice as follows. Transmitter 1 is at 1, Receiver 1 is at 2, Transmitter 2 is at 3, and Receiver 2 is at 4. The first receiver thus hears both transmitters whereas the second only hears the second transmitter. The longest power chain here is the chain $(1, 2)$. As we shall see in the proof of achievability, this power chain corresponds to power-$\mathcal{E}$ transmission by Transmitter 1 and power-$\sqrt{\mathcal{E}}$ transmission by Transmitter 2. The signal $X(1)$ is decoded by Receiver 1 and the signal $X(2)$ is decoded by Receiver 2. When decoding $X(1)$, Receiver 1 treats $X(2)$ as interference.[1] It can overcome this interference because $X(1)$ is of power $\mathcal{E}$ whereas $X(2)$ is of power $\sqrt{\mathcal{E}}$ so that the SNR is roughly $\mathcal{E}/(\sqrt{\mathcal{E}} + 1)$. The weaker signal, $X(2)$, is decoded by Receiver 2. The zero in the matrix guarantees that this receiver suffers no interference from $X(1)$. Thus, it operates at an SNR of $\sqrt{\mathcal{E}}$.

8) An example that is particularly useful for demonstrating the converse is one where

$$\mathbb{H} = \begin{pmatrix} H & H & 0 \\ 0 & H & H \\ H & 0 & H \end{pmatrix}. \tag{32}$$

This may correspond to a circular geometry where the transmitters and receivers are located on a circle, say a clock. The three transmitters are located at 12, 4, and 8, and the three receivers at 2, 6, and 10. Each receiver only receives the signals transmitted by the transmitters adjacent to it. This circular model was introduced (in the absence of fading) by [5] for the study of cellular systems. Here the longest power chains are of length 2 and are given by $(1, 2)$, $(1, 3)$, and $(2, 3)$.

9) A variation on a theme of Wyner's [1] is a cellular telephony model where M cells are arranged uniformly along a circle. A transmission in a given cell is received by three antennas: by the cell's base station and by the base stations of the two cells adjacent to it. An example of the resulting matrix for $M = 6$ is

$$\mathbb{H} = \begin{pmatrix} H & H & 0 & 0 & 0 & H \\ H & H & H & 0 & 0 & 0 \\ 0 & H & H & H & 0 & 0 \\ 0 & 0 & H & H & H & 0 \\ 0 & 0 & 0 & H & H & H \\ H & 0 & 0 & 0 & H & H \end{pmatrix}. \tag{33}$$

A longest power chain for this case is, for example, $(1, 2, \ldots, M - 2)$, which yields $\kappa^* = M - 2$.

III. Proof of Theorem 1

In this section, we provide a proof of Theorem 1. We shall begin by showing that it suffices to prove the theorem in the case where the fading $\{\mathbb{H}_k\}$ is memoryless, i.e., when $\{\mathbb{H}_k\}$ are i.i.d. We shall then separately prove the "converse" (21) and the "direct" (22) parts in the two corresponding subsections.

Let then $\{\mathbb{H}_k\}$ be some stationary and ergodic fading process with memory satisfying (9) and (14), and let $\{\tilde{\mathbb{H}}_k\}$ be an i.i.d. fading process of equal marginal so that the law of $\tilde{\mathbb{H}}_k$ is the

[1]This interference is non-Gaussian so that some care must be exercised in analyzing it. Moreover, because we allow for dependence among the components of $\mathbb{H}$, this interference is not independent of the fading that $X(1)$ experiences at Receiver 1.

same as the law of $\mathbb{H}_1$ (which is the same, by stationarity, as the law of $\mathbb{H}_k$ for any $k \in \mathbb{Z}$).

That it suffices to prove the converse in the memoryless case follows because, as shown by Moser [6, Ch. 8], the difference between the feedback capacity of the channel with fading $\{\mathbb{H}_k\}$ and the capacity of the channel with i.i.d. fading $\{\tilde{\mathbb{H}}_k\}$ is bounded in the SNR. For the sake of completeness we repeat Moser's result in the Appendix.

As to the direct part, we note that the capacity of the channel with fading $\{\mathbb{H}_k\}$ cannot be smaller than that of fading $\{\tilde{\mathbb{H}}_k\}$. Indeed, if Q is any distribution on $\mathbb{C}^{n_{\mathrm{T}}}$ then the mutual information that it induces on the memoryless channel of fading $\{\tilde{\mathbb{H}}_k\}$ is achievable on the channel of fading $\{\mathbb{H}_k\}$ by considering inputs $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ that are i.i.d. according to Q. Indeed, for such i.i.d. inputs

$$\begin{aligned}\frac{1}{n} I\left(\boldsymbol{X}_1^n; \boldsymbol{Y}_1^n\right) &= \frac{1}{n} \sum_{k=1}^{n} I\left(\boldsymbol{X}_k; \boldsymbol{Y}_1^n \,\middle|\, \boldsymbol{X}_1^{k-1}\right) \\ &= \frac{1}{n} \sum_{k=1}^{n} I\left(\boldsymbol{X}_k; \boldsymbol{Y}_1^n, \boldsymbol{X}_1^{k-1}\right) \\ &\geq \frac{1}{n} \sum_{k=1}^{n} I(\boldsymbol{X}_k; \boldsymbol{Y}_k) \\ &= I(\boldsymbol{X}_1; \mathbb{H}_1 \boldsymbol{X}_1 + \boldsymbol{Z}_1) \\ &= I(\boldsymbol{X}_1; \tilde{\mathbb{H}}_1 \boldsymbol{X}_1 + \boldsymbol{Z}_1)\end{aligned}$$

where the first equality follows from the chain rule; the subsequent from the independence of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$; and the subsequent inequality because reducing observations cannot increase mutual information. Here we use $\boldsymbol{X}_\ell^m$ to denote $\boldsymbol{X}_\ell, \ldots, \boldsymbol{X}_m$.

We shall thus proceed to prove the theorem assuming that the fading is memoryless. In this case, we shall omit the time index so that our assumptions on the fading process can be now written as

$$\mathsf{E}[|H(r,t)|^2] < \infty, \qquad (r,t) \in \mathcal{R} \times \mathcal{T} \tag{34}$$

$$h(H(\mathcal{Z}^c)) > -\infty \tag{35}$$

$$(r,t) \in \mathcal{Z} \Rightarrow (H(r,t) = 0, \text{ almost surely}). \tag{36}$$

We shall further assume that none of the rows of $\mathbb{H}$ is deterministically zero, i.e.,

$$\forall r \in \mathcal{R} \quad \exists t \in \mathcal{T} : (r,t) \notin \mathcal{Z} \tag{37}$$

or equivalently

$$\mathcal{T}_r \neq \emptyset, \qquad r \in \mathcal{R}. \tag{38}$$

This corresponds to the condition that every receiver "hears" at least one transmitter. Analogously, we shall assume that none of the columns of $\mathbb{H}$ is deterministically zero, i.e.,

$$\forall t \in \mathcal{T} \quad \exists r \in \mathcal{R} : (r,t) \notin \mathcal{Z} \tag{39}$$

or equivalently

$$\mathcal{R}_t \neq \emptyset, \quad t \in \mathcal{T}. \tag{40}$$

This corresponds to the condition that every transmitter is heard by at least one receiver. The above assumptions can be made without loss of generality because a receiver that hears no signals (other than ambient additive noise) does not affect the longest power chain and can also be ignored at the detector. Similarly, a transmitter that cannot be heard by any receiver will never be an element of a power chain and there is also no point in having it transmit any signal.

A. The Converse

In this subsection, we provide a proof of (21) for i.i.d. fading $\{\mathbb{H}_k\}$ satisfying (34), (35), and (36). We begin by considering the "ordering permutation" $\sigma(\boldsymbol{x})$ of a given n_{T}-tuple $\boldsymbol{x} \in \mathbb{C}^{n_{\mathrm{T}}}$. This is the permutation that orders the components of $\boldsymbol{x}$ in descending order of their magnitudes, resolving ties with preference to lower indices. Thus, given an n_{T}-tuple $\boldsymbol{x} \in \mathbb{C}^{n_{\mathrm{T}}}$, we set $\sigma(\boldsymbol{x})$ to be the permutation $\tau : \nu \mapsto \tau_\nu$ on $\mathcal{T}$ that satisfies

$$|x(\tau_1)| \geq |x(\tau_2)| \geq \cdots \geq |x(\tau_{n_{\mathrm{T}}})| \tag{41}$$

and that resolves ties in favor of lower indices so that

$$|x(\tau_\nu)| = |x(\tau_{\nu+1})| \Rightarrow \tau_\nu < \tau_{\nu+1}. \tag{42}$$

The form in which ties are resolved does not play an important role in our analysis. It is made here explicit because it is important that $\boldsymbol{x} \in \mathbb{C}^{n_{\mathrm{T}}}$ determine the ordering permutation $\sigma(\boldsymbol{x})$ uniquely.

If $\boldsymbol{X}$ is a random vector taking value in $\mathbb{C}^{n_{\mathrm{T}}}$, then its ordering permutation $\sigma(\boldsymbol{X})$ is a random permutation. Since the number of permutation on $\mathcal{T}$ is $n_{\mathrm{T}}!$, it follows that, irrespective of the distribution of $\boldsymbol{X}$, the entropy of $\sigma(\boldsymbol{X})$ is upper-bounded by

$$H(\sigma(\boldsymbol{X})) \leq \log n_{\mathrm{T}}!. \tag{43}$$

Given any channel input $\boldsymbol{X}$ we can thus expand the mutual information $I(\boldsymbol{X};\boldsymbol{Y})$ between the channel terminals as

$$\begin{aligned}I(\boldsymbol{X};\boldsymbol{Y}) &= I(\boldsymbol{X}, \sigma(\boldsymbol{X}); \boldsymbol{Y}) \\ &= I(\sigma(\boldsymbol{X}); \boldsymbol{Y}) + I(\boldsymbol{X}; \boldsymbol{Y} \,|\, \sigma(\boldsymbol{X})) \\ &\leq H(\sigma(\boldsymbol{X})) + I(\boldsymbol{X}; \boldsymbol{Y} \,|\, \sigma(\boldsymbol{X})) \\ &\leq \log n_{\mathrm{T}}! + I(\boldsymbol{X}; \boldsymbol{Y} \,|\, \sigma(\boldsymbol{X})) \\ &= \sum_{\tau : \Pr[\sigma(\boldsymbol{X}) = \tau] > 0} I(\boldsymbol{X}; \boldsymbol{Y} \,|\, \sigma(\boldsymbol{X}) = \tau) \Pr[\sigma(\boldsymbol{X}) = \tau] \\ &\quad + \log n_{\mathrm{T}}!.\end{aligned} \tag{44}$$

The proof of the converse will now focus on the terms of the form

$$I(\boldsymbol{X}; \boldsymbol{Y} \,|\, \sigma(\boldsymbol{X}) = \tau)$$

where τ is an arbitrary permutation satisfying

$$\Pr[\sigma(\boldsymbol{X}) = \tau] > 0.$$

Fix then such a permutation τ and let

$$\mathcal{E}_\tau = \mathsf{E}[\|\boldsymbol{X}\|^2 \,|\, \sigma(\boldsymbol{X}) = \tau]. \tag{45}$$

We will show that corresponding to the set $\mathcal{Z}$ and to the permutation τ there is a power chain of length $\kappa = \kappa(\mathcal{Z}, \tau)$ such that

$$\begin{aligned}I(\boldsymbol{X}; \boldsymbol{Y} \,|\, \sigma(\boldsymbol{X}) = \tau) &\leq \kappa \cdot \log(1 + \log(1 + \mathcal{E}_\tau)) + c \qquad (46) \\ &\leq \kappa^* \cdot \log(1 + \log(1 + \mathcal{E}_\tau)) + c \qquad (47)\end{aligned}$$

where the second inequality follows because, by definition, κ^* is the length of the longest power chain, and where the constant c depends only on the law of $\mathbb{H}$ and on the permutation τ but not on the power $\mathcal{E}_\tau$.

Note that once we establish (47), the converse will follow from (45) and (44) as well as Jensen's inequality by the concavity of the double-logarithmic function. We thus proceed to construct the length-$\kappa(\mathcal{Z},\tau)$ power chain and to then prove (46).

To avoid some double subscripts in the description of this length-$\kappa(\mathcal{Z},\tau)$ power chain we shall use $[\nu]$ for τ_ν. Thus, conditional on $\sigma(\boldsymbol{X}) = \tau$ we have that $X([1])$ has the maximal magnitude among all the elements of $\boldsymbol{X}$, and $X([n_\mathrm{T}])$ has the smallest magnitude.

Let $j_1 = 1$. Assume that we have defined $j_1, \ldots, j_\nu$. We then define $j_{\nu+1}$ as

$$j_{\nu+1} = \min\left\{ j_\nu < \ell \le n_\mathrm{T} : \mathcal{R}_{[\ell]} \Bigg\backslash \bigcup_{\eta=1}^{\nu} \mathcal{R}_{[j_\eta]} \neq \emptyset \right\} \quad (48)$$

where the minimum of an empty set should be understood as ∞. We then set

$$\kappa = \max\{1 \le \nu \le n_\mathrm{T} : j_\nu < \infty\} \quad (49)$$

and define

$$t_\nu = [j_\nu] \quad \mathcal{B}_\nu = \mathcal{R}_{t_\nu} \Bigg\backslash \bigcup_{\eta=1}^{\nu-1} \mathcal{R}_{t_\eta}, \qquad \nu = 1, \ldots, \kappa. \quad (50)$$

Thus, t_ν is the next strongest transmitter after $t_{\nu-1}$ that can be heard by some receiver that is uninfluenced by any of the stronger transmitters that are already in the chain. The set $\mathcal{B}_\nu$ is the set of receivers that can hear t_ν but not any of the stronger transmitters that are in the chain. Note that by (40) we have $\mathcal{R}_{t_1} \neq \emptyset$. In fact, $(t_1, \ldots, t_\kappa)$ is a power chain with respect to $\mathcal{Z}$, so that

$$\kappa \le \kappa^*. \quad (51)$$

(Recall that κ^* is the length of the longest power chain with respect to $\mathcal{Z}$.) Also note that the sets $\{\mathcal{B}_\nu\}$ are disjoint and that by (38) their union is $\mathcal{R}$, i.e., they form a partition of $\mathcal{R}$

$$\mathcal{R} = \bigcup_{\nu=1}^{\kappa} \mathcal{B}_\nu \quad \mathcal{B}_\nu \cap \mathcal{B}_{\nu'} = \emptyset \quad \text{whenever } 1 \le \nu \neq \nu' \le \kappa. \quad (52)$$

Finally, we define

$$\mathcal{A}_\nu = \{[j_\nu], \ldots, [j_{\nu+1} - 1]\}, \qquad \nu = 1, \ldots, \kappa - 1 \quad (53)$$

and

$$\mathcal{A}_\kappa = \{[j_\kappa], \ldots, [n_\mathrm{T}]\}. \quad (54)$$

The key properties of the constructions of $\kappa(\mathcal{Z},\tau)$, of the power chain $(t_1, \ldots, t_\kappa)$, of the collection $\{\mathcal{B}_\nu\}$, and of the collection $\{\mathcal{A}_\nu\}$ are as follows. The κ-tuple $(t_1, \ldots, t_\kappa)$ is a power chain, so that $\kappa \le \kappa^*$, (51). The collections $\{\mathcal{B}_\nu\}$ and $\{\mathcal{A}_\nu\}$ are partitions of $\mathcal{R}$ and $\mathcal{T}$, respectively. And conditional on $\sigma(\boldsymbol{X}) = \tau$, the random variables $X(\mathcal{A}_\nu)$ only influence $Y(\cup_{\eta=1}^{\nu} \mathcal{B}_\eta)$; they do not influence any receiver in $\mathcal{R} \setminus \cup_{\eta=1}^{\nu} \mathcal{B}_\eta$. That is, conditional on $\sigma(\boldsymbol{X}) = \tau$ and on the random variables $X(\mathcal{A}_\nu \cup \cdots \cup \mathcal{A}_\kappa)$, the random variables $X(\mathcal{A}_1 \cup \cdots \cup \mathcal{A}_{\nu-1})$ are independent of the random variables $Y(\mathcal{B}_\nu)$.

Using these properties, we next prove (46). The key will be the following lemma.

Lemma 2: Let $\mathbb{H}$ be a random $n_\mathrm{R} \times n_\mathrm{T}$ complex matrix whose components are all of finite second moment

$$\mathsf{E}[|H(r,t)|^2] < \infty, \qquad (r,t) \in \mathcal{R} \times \mathcal{T}$$

where $\mathcal{R} = \{1, \ldots, n_\mathrm{R}\}$ and $\mathcal{T} = \{1, \ldots, n_\mathrm{T}\}$. Let the set $\mathcal{Z} \subset \mathcal{R} \times \mathcal{T}$ be the set of pairs (r,t) such that $H(r,t)$ is deterministically zero

$$H(r,t) = 0 \quad \text{almost surely} \quad \forall (r,t) \in \mathcal{Z}.$$

Assume that the joint differential entropy of the coordinates that are not in $\mathcal{Z}$ is finite

$$h(H(\mathcal{Z}^\mathrm{c})) > -\infty. \quad (55)$$

Let $t^* \in \mathcal{T}$ be fixed. Assume that transmitter t^* influences all receivers in the sense that

$$(r, t^*) \notin \mathcal{Z}, \qquad \forall r \in \mathcal{R}. \quad (56)$$

Let $\boldsymbol{X}$ be a random vector taking value in $\mathbb{C}^{n_\mathrm{T}}$ whose component of largest magnitude is almost surely t^*

$$\max_{t \in \mathcal{T}} |X(t)| = |X(t^*)|, \quad \text{almost surely}. \quad (57)$$

Assume the average-power constraint

$$\sum_{t \in \mathcal{T}} \mathsf{E}[|X(t)|^2] \le \mathcal{E}.$$

Finally, let $\boldsymbol{Z}$ take value in $\mathbb{C}^{n_\mathrm{T}}$ according the multivariate Gaussian law $\mathcal{N}_\mathbb{C}(\mathbf{0}, \mathsf{I}_{n_\mathrm{R}})$ and assume that $\mathbb{H}$, $\boldsymbol{Z}$, and $\boldsymbol{X}$ are independent.

Then there exists some constant c, which depends on the law of $\mathbb{H}$ but not on the law of $\boldsymbol{X}$ or on its power $\mathcal{E}$, such that

$$I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X} + \boldsymbol{Z}) \le \log(1 + \log(1 + \mathcal{E})) + c. \quad (58)$$

It should be noted that this lemma cannot be applied directly to the original capacity problem at hand. There is no reason to believe that the capacity of the network will be achieved by some input law under which one of transmitters is *always* sending a signal that is larger than all other signals. The application of the lemma will be restricted to the analysis of the mutual information *conditional on the ordering permutation being equal to* τ. Conditioned on this, Transmitter $\tau_1 (= [1])$ sends the largest signal with probability one. Even under this conditioning we cannot immediately apply the lemma, because there is no reason to believe that this strongest transmitter will be heard by all receivers. We shall only apply the lemma to study the mutual information between the inputs and *the subset of receivers that do indeed hear the strongest transmitter*.

Proof of Lemma 2: Let $\boldsymbol{Y} = \mathbb{H}\boldsymbol{X} + \boldsymbol{Z}$ and let

$$\mathcal{D} = \{\boldsymbol{x} \in \mathbb{C}^{n_\mathrm{T}} : \max_{1 \le t \le n_\mathrm{T}} |x(t)| = |x(t^*)|\} \tag{59}$$

so that (57) can be rewritten as $\Pr[\boldsymbol{X} \in \mathcal{D}] = 1$.

The proof of the lemma is very similar to the proof of [3, Theorem 4.2]. It too is based on the bound [3, eq. (333)]

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) \le \; & \log \pi^{n_\mathrm{R}} - \log \Gamma(n_\mathrm{R}) \\ & + \mathsf{E}_{\boldsymbol{X}}[n_\mathrm{R} \mathsf{E}[\log \|\boldsymbol{Y}\|^2 \,|\, \boldsymbol{X} = \boldsymbol{x}] - h(\boldsymbol{Y} \,|\, \boldsymbol{X} = \boldsymbol{x})] \\ & + \alpha(1 + \log \mathsf{E}[\|\boldsymbol{Y}\|^2] - \mathsf{E}[\log \|\boldsymbol{Y}\|^2]) \\ & + \log \Gamma(\alpha) - \alpha \log \alpha, \qquad \alpha > 0 \end{aligned} \tag{60}$$

where $\mathsf{E}_{\boldsymbol{X}}[\cdot]$ denotes expectation with respect to $\boldsymbol{X}$. From this inequality it follows that for inputs $\boldsymbol{X}$ satisfying (57) (i.e., satisfying $\Pr[\boldsymbol{X} \in \mathcal{D}] = 1$)

$$\begin{aligned} & I(\boldsymbol{X};\boldsymbol{Y}) \\ & \quad \le \log \pi^{n_\mathrm{R}} - \log \Gamma(n_\mathrm{R}) \\ & \qquad + \sup_{\boldsymbol{x} \in \mathcal{D}} \{n_\mathrm{R} \mathsf{E}[\log \|\boldsymbol{Y}\|^2 \,|\, \boldsymbol{X} = \boldsymbol{x}] - h(\boldsymbol{Y} \,|\, \boldsymbol{X} = \boldsymbol{x})\} \\ & \qquad + \alpha(1 + \log \mathsf{E}[\|\boldsymbol{Y}\|^2] - \mathsf{E}[\log \|\boldsymbol{Y}\|^2]) \\ & \qquad + \log \Gamma(\alpha) - \alpha \log \alpha, \quad \alpha > 0, \; \Pr[\boldsymbol{X} \in \mathcal{D}] = 1. \end{aligned} \tag{61}$$

We now proceed to analyze the various terms in (61). We begin with showing that the supremum, which does not depend on $\mathcal{E}$, is finite

$$\sup_{x \in \mathcal{D}} \{n_\mathrm{R} \mathsf{E}[\log \|\boldsymbol{Y}\|^2 \,|\, \boldsymbol{X} = \boldsymbol{x}] - h(\boldsymbol{Y} \,|\, \boldsymbol{X} = \boldsymbol{x})\} < \infty. \tag{62}$$

To this end, we use Jensen's inequality to obtain

$$\begin{aligned} & n_\mathrm{R} \mathsf{E}[\log \|\boldsymbol{Y}\|^2 \,|\, \boldsymbol{X} = \boldsymbol{x}] \\ & \quad \le n_\mathrm{R} \log \mathsf{E}[\|\boldsymbol{Y}\|^2 \,|\, \boldsymbol{X} = \boldsymbol{x}], \quad \boldsymbol{x} \in \mathbb{C}^{n_\mathrm{T}} \\ & \quad = n_\mathrm{R} \log(\mathsf{E}[\|\mathbb{H}\boldsymbol{X}\|^2 \,|\, \boldsymbol{X} = \boldsymbol{x}] + \mathsf{E}[\|\boldsymbol{Z}\|^2]), \quad \boldsymbol{x} \in \mathbb{C}^{n_\mathrm{T}} \\ & \quad \le n_\mathrm{R} \log\left(\mathsf{E}\left[\|\mathbb{H}\|_F^2\right] \cdot \|\boldsymbol{x}\|^2 + n_\mathrm{R}\right), \quad \boldsymbol{x} \in \mathbb{C}^{n_\mathrm{T}} \\ & \quad \le n_\mathrm{R} \log\left(\mathsf{E}\left[\|\mathbb{H}\|_F^2\right] \cdot n_\mathrm{T} \cdot |x(t^*)|^2 + n_\mathrm{R}\right), \quad \boldsymbol{x} \in \mathcal{D} \end{aligned} \tag{63}$$

where the second inequality follows from the Cauchy–Schwarz inequality with

$$\|\mathbb{H}\|_F^2 = \sum_{r,t} |H(r,t)|^2$$

denoting the squared Frobenius norm, and where the last inequality follows by restricting $\boldsymbol{x}$ to be in the set $\mathcal{D}$ where $\|\boldsymbol{x}\|^2 \le n_\mathrm{T} |x(t^*)|^2$.

As to the differential entropy term in (62), we obtain two separate bounds. The first is useful when $\|\boldsymbol{x}\|^2$ is very small and is otherwise quite crude

$$\begin{aligned} h(\boldsymbol{Y} \,|\, \boldsymbol{X} = \boldsymbol{x}) & = h(\mathbb{H}\boldsymbol{x} + \boldsymbol{Z}) \\ & \ge h(\boldsymbol{Z}) \\ & = n_\mathrm{R} \log(\pi e), \qquad \boldsymbol{x} \in \mathbb{C}^{n_\mathrm{T}}. \end{aligned} \tag{64}$$

The second is

$$\begin{aligned} & h(\boldsymbol{Y} \,|\, \boldsymbol{X} = \boldsymbol{x}) \\ & \quad \ge h(\mathbb{H}\boldsymbol{x}) \\ & \quad \ge h\left(\mathbb{H}\boldsymbol{x} \;\middle|\; \{H(r,t)\}_{r \in \mathcal{R}, t \in \mathcal{T} \setminus \{t^*\}}\right) \\ & \quad = h\left(\{H(r',t^*) \cdot x(t^*)\}_{r' \in \mathcal{R}} \;\middle|\; \{H(r,t)\}_{r \in \mathcal{R}, t \in \mathcal{T} \setminus \{t^*\}}\right) \\ & \quad = n_\mathrm{R} \log |x(t^*)|^2 \\ & \qquad + h\left(\{H(r',t^*)\}_{r' \in \mathcal{R}} \;\middle|\; \{H(r,t)\}_{r \in \mathcal{R}, t \in \mathcal{T} \setminus \{t^*\}}\right) \\ & \quad = n_\mathrm{R} \log |x(t^*)|^2 \\ & \qquad + h\left(\{H(r',t^*)\}_{r' \in \mathcal{R}} \;\middle|\; \{H(r,t)\}_{t \in \mathcal{T} \setminus \{t^*\}, (r,t) \notin \mathcal{Z}}\right), \\ & \qquad\qquad |x(t^*)| > 0. \end{aligned} \tag{65}$$

Here the first inequality follows by ignoring the noise; the second inequality follows because conditioning cannot increase differential entropy; the subsequent equality by expressing $\mathbb{H}\boldsymbol{x}$ as

$$\begin{aligned} \mathbb{H}\boldsymbol{x} & = \begin{pmatrix} \sum_{t \in \mathcal{T}} H(1,t) x(t) \\ \vdots \\ \sum_{t \in \mathcal{T}} H(n_\mathrm{R},t) x(t) \end{pmatrix} \\ & = \begin{pmatrix} H(1,t^*) x(t^*) \\ \vdots \\ H(n_\mathrm{R},t^*) x(t^*) \end{pmatrix} + \begin{pmatrix} \sum_{t \in \mathcal{T} \setminus \{t^*\}} H(1,t) x(t) \\ \vdots \\ \sum_{t \in \mathcal{T} \setminus \{t^*\}} H(n_\mathrm{R},t) x(t) \end{pmatrix}, \end{aligned}$$

by noting that conditional on $\{H(r,t)\}_{r \in \mathcal{R}, t \in \mathcal{T} \setminus \{t^*\}}$ the second term on the right is deterministic, and by noting that the addition of a deterministic vector does not affect a vector's differential entropy; the next equality from the behavior of differential entropy under scaling; and the final equality because it is pointless to condition on deterministic random variables. Note that (55) guarantees that the right-hand side (RHS) of (65) is finite. Inequalities (63)–(65) combine to prove (62).

The analysis of the other terms in (61) and the choice of α in (61) as $\alpha(\mathcal{E}) = \alpha^*$, where α^* is given in (69), is identical to the analysis and choice in [3, Appendix II]

$$\begin{aligned} \log \mathsf{E}[\|\boldsymbol{Y}\|^2] & = \log(\mathsf{E}[\|\mathbb{H}\boldsymbol{X}\|^2] + \mathsf{E}[\|\boldsymbol{Z}\|^2]) \\ & \le \log(\mathsf{E}[\|\mathbb{H}\|^2] \mathsf{E}[\|\boldsymbol{X}\|^2] + \mathsf{E}[\|\boldsymbol{Z}\|^2]) \\ & \le \log\left(\mathsf{E}\left[\|\mathbb{H}\|_\mathrm{F}^2\right] \mathcal{E} + n_\mathrm{R}\right) \end{aligned} \tag{66}$$

$$\begin{aligned} \mathsf{E}[\log \|\boldsymbol{Y}\|^2] & = \mathsf{E}[\log \|\mathbb{H}\boldsymbol{X} + \boldsymbol{Z}\|^2] && (67) \\ & \ge \mathsf{E}[\log \|\boldsymbol{Z}\|^2] && (68) \\ \alpha^* & = (1 + \log \mathsf{E}[\|\boldsymbol{Y}\|^2] - \mathsf{E}[\log \|\boldsymbol{Y}\|^2])^{-1} && (69) \end{aligned}$$

where $\alpha^* \downarrow 0$ with the SNR. See [3, Appendix II] for details. □

Note 3: The Gaussianity of the noise in the above lemma is not crucial. As in [3 Appendix II], the result continues to hold whenever the additive noise $\boldsymbol{Z}$ is of finite second moment and of finite differential entropy.

With the aid of this lemma we can now prove (46). We shall upper-bound $I(\boldsymbol{X};\boldsymbol{Y} \,|\, \sigma(\boldsymbol{X}) = \tau)$ in κ phases. In the first phase, we shall upper-bound this mutual information by a double-logarithmic term, a constant, and another mutual information term. This latter mutual information term will be upper-bounded in the second phase by a double-logarithmic term, a constant, and yet another mutual information term, which is then upper-bounded in the third phase. In the final phase, Phase κ, we upper-bound the mutual information by a double-logarithmic term and a constant only, thus terminating the calculation. Since each phase contributes a double-logarithmic term (and a constant), the κ phases contribute together κ double-logarithmic terms (and κ constants, which are combined into one) as required.

The details now follow. In Phase 1 we expand mutual information using the chain rule

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}\,|\,\sigma(\boldsymbol{X})=\tau) &= I(X(\mathcal{T});Y(\mathcal{R})\,|\,\sigma(\boldsymbol{X})=\tau)\\ &= I\left(X(\mathcal{T});Y(\mathcal{B}_1),Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right)\\ &= I(X(\mathcal{T});Y(\mathcal{B}_1)\,|\,\sigma(\boldsymbol{X})=\tau)\\ &\quad + I\left(X(\mathcal{T});Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,Y(\mathcal{B}_1),\sigma(\boldsymbol{X})=\tau\right). \end{aligned} \tag{70}$$

The first term on the RHS of (70) is easily treated using the lemma, because conditional on $\sigma(\boldsymbol{X})=\tau$, the component $X(t_1)$ (=$X([1])$) is of largest magnitude, and it is heard by all the receivers in $\mathcal{B}_1$. Consequently, we have by Lemma 2

$$I(X(\mathcal{T});Y(\mathcal{B}_1)\,|\,\sigma(\boldsymbol{X})=\tau)\le\log(1+\log(1+\mathcal{E}_\tau))+c_1 \tag{71}$$

where the constant c_1 is as in Lemma 2 independent of the SNR.

As for the second term on the RHS of (70), we use the chain rule once again to obtain

$$\begin{aligned} I\left(X(\mathcal{T});Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,Y(\mathcal{B}_1),\sigma(\boldsymbol{X})=\tau\right) &\le I\left(X(\mathcal{T}),Y(\mathcal{B}_1);Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right)\\ &= I\left(X(\mathcal{T});Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right)\\ &\quad + I\left(Y(\mathcal{B}_1);Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,X(\mathcal{T}),\sigma(\boldsymbol{X})=\tau\right)\\ &= I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right)\\ &\quad + I\left(Y(\mathcal{B}_1);Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,X(\mathcal{T}),\sigma(\boldsymbol{X})=\tau\right)\\ &\le I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right)\\ &\quad + I\left(\{H(r,t)\}_{\substack{t\in\mathcal{T}\\ r\in\mathcal{B}_1}};\{H(r,t)\}_{\substack{t\in\mathcal{T}\\ r\in\mathcal{B}_1^{\mathrm{c}}}}\right). \end{aligned} \tag{72}$$

Here the last inequality follows by the data processing inequality, and the preceding equality follows because $Y(\mathcal{B}_1^{\mathrm{c}})$ is conditionally independent of $X(\mathcal{A}_1)$ given $X(\mathcal{A}_1^{\mathrm{c}})$.

Thus, we have by (70) and (71) that the original mutual information term is upper-bounded by a double-logarithmic term, a constant term (which is finite by (35)), and another mutual information term

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}\,|\,\sigma(\boldsymbol{X})=\tau) &\le \log(1+\log(1+\mathcal{E}_\tau))+c_1\\ &\quad + I\left(\{H(r,t)\}_{\substack{t\in\mathcal{T}\\ r\in\mathcal{B}_1}};\{H(r,t)\}_{\substack{t\in\mathcal{T}\\ r\in\mathcal{B}_1^{\mathrm{c}}}}\right)\\ &\quad + I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right). \end{aligned} \tag{73}$$

The mutual information term

$$I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right)$$

on the RHS of the above is now upper-bounded in Phase 2. Notice that this term corresponds to a "smaller" fading channel where the inputs $\mathcal{A}_1$ are immaterial, as are the outputs $\mathcal{B}_1$. In Phase 2 we thus upper-bound this term as follows:

$$\begin{aligned} I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right) &= I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\cap\mathcal{B}_2\right),Y\left(\mathcal{B}_1^{\mathrm{c}}\cap\mathcal{B}_2^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right)\\ &= I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right);Y(\mathcal{B}_2)\,|\,\sigma(\boldsymbol{X})=\tau\right)\\ &\quad + I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\cap\mathcal{B}_2^{\mathrm{c}}\right)\,|\,Y(\mathcal{B}_2),\sigma(\boldsymbol{X})=\tau\right). \end{aligned} \tag{74}$$

The first term can be bounded using the lemma because $X(t_2)$ is the component of $X(\mathcal{A}_1^{\mathrm{c}})$ of largest magnitude, and it is heard by all receivers in $\mathcal{B}_2$

$$I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right);Y(\mathcal{B}_2)\,|\,\sigma(\boldsymbol{X})=\tau\right)\le\log(1+\log(1+\mathcal{E}_\tau))+c_2$$

for some constant c_2.

The second term in (74) can be expanded in analogy to (72) to yield

$$\begin{aligned} I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\cap\mathcal{B}_2^{\mathrm{c}}\right)\,|\,Y(\mathcal{B}_2),\sigma(\boldsymbol{X})=\tau\right) &\le I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right),Y(\mathcal{B}_2);Y\left(\mathcal{B}_1^{\mathrm{c}}\cap\mathcal{B}_2^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right)\\ &= I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\cap\mathcal{B}_2^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right)\\ &\quad + I\left(Y(\mathcal{B}_2);Y\left(\mathcal{B}_1^{\mathrm{c}}\cap\mathcal{B}_2^{\mathrm{c}}\right)\,|\,X\left(\mathcal{A}_1^{\mathrm{c}}\right),\sigma(\boldsymbol{X})=\tau\right)\\ &= I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\cap\mathcal{A}_2^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\cap\mathcal{B}_2^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right)\\ &\quad + I\left(\{H(r,t)\}_{\substack{t\in\mathcal{A}_1^{\mathrm{c}}\\ r\in\mathcal{B}_2}};\{H(r,t)\}_{\substack{t\in\mathcal{A}_1^{\mathrm{c}}\\ r\in\mathcal{B}_1^{\mathrm{c}}\cap\mathcal{B}_2^{\mathrm{c}}}}\right). \end{aligned}$$

The mutual information term

$$I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\cap\mathcal{A}_2^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\cap\mathcal{B}_2^{\mathrm{c}}\right)\,|\,\sigma(\boldsymbol{X})=\tau\right)$$

is now upper-bounded in Phase 3. This process is continued until the final phase, Phase κ, when the term

$$I\left(X\left(\mathcal{A}_1^{\mathrm{c}}\cap\cdots\cap\mathcal{A}_{\kappa-1}^{\mathrm{c}}\right);Y\left(\mathcal{B}_1^{\mathrm{c}}\cap\cdots\cap\mathcal{B}_{\kappa-1}^{\mathrm{c}}\right)\right) = I(X(\mathcal{A}_\kappa);Y(\mathcal{B}_\kappa))$$

is upper-bounded using the lemma by a double-logarithmic term and a constant without an additional mutual information term. Indeed, the component $X(t_\kappa)$ is of largest magnitude among the terms in $X(\mathcal{A}_\kappa)$ and it influences all the receivers in $\mathcal{B}_\kappa$.

It is thus seen that performing a total of κ phases yields the bound (46) and hence, by (51), also (47). The converse now follows from (47) and (44) using Jensen's inequality because the double-logarithmic function is concave and because, in view of (45)

$$\sum_{\tau:\Pr[\sigma(\boldsymbol{X})=\tau]>0}\Pr[\sigma(\boldsymbol{X})=\tau]\cdot\mathcal{E}_\tau=\mathsf{E}[\|\boldsymbol{X}\|^2]. \tag{75}$$

B. The Direct Part

To prove the direct part we shall demonstrate that if $(t_1,\ldots,t_\kappa)\in\mathcal{T}^\kappa$ is any power chain with respect to $\mathcal{Z}$ then we can find a distribution on $\boldsymbol{X}$ under which its components are independent (thus guaranteeing the achievability of $I(\boldsymbol{X};\boldsymbol{Y})$ under multiple-access conditions) and such that

$$\overline{\lim_{\mathcal{E}\to\infty}}\{\kappa\cdot\log\log\mathcal{E}-I(\boldsymbol{X};\boldsymbol{Y})\}<\infty. \tag{76}$$

We shall also demonstrate the existence of a κ-tuple $(r_1,\ldots,r_\kappa)\in\mathcal{R}^\kappa$ such that for the above input vector $\boldsymbol{X}$

$$\overline{\lim_{\mathcal{E}\to\infty}}\left\{\kappa\cdot\log\log\mathcal{E}-\sum_{\nu=1}^{\kappa}I(X(t_\nu);Y(r_\nu))\right\}<\infty$$

thus demonstrating the achievability of $\kappa\cdot\log\log\mathcal{E}$ with κ single-user detectors.

Let the power chain $(t_1,\ldots,t_\kappa) \in \mathcal{T}^\kappa$ be fixed. Consider a distribution for $\boldsymbol{X}$ under which the components of $\boldsymbol{X}$ are independent with laws that can be described as follows. If some $t \in \mathcal{T}$ is not in $\{t_1,\ldots,t_\kappa\}$, we set $X(t)$ to be deterministically zero

$$t \notin \{t_1,\ldots,t_\kappa\} \Rightarrow (X(t) = 0 \text{ a.s.}). \tag{77}$$

As to the other components of $\boldsymbol{X}$, we choose them to be circularly symmetric with squared magnitudes whose logarithms are uniformly distributed on intervals

$$\log|X(t_\nu)|^2 \sim \text{Uniform}\left(\log x^2_{\min,\nu}, \log x^2_{\max,\nu}\right), \qquad \nu = 1,\ldots,\kappa \tag{78}$$

where the endpoints satisfy

$$0 < x^2_{\min,\nu} < x^2_{\max,\nu} \le \mathcal{E}, \qquad \nu = 1,\ldots,\kappa \tag{79}$$

and will be specified later. (See (102) and (103).) Note that with this choice of the laws

$$h(\log|X(t_\nu)|^2) = \log\log\frac{x^2_{\max,\nu}}{x^2_{\min,\nu}}, \qquad \nu = 1,\ldots,\kappa. \tag{80}$$

Since $(t_1,\ldots,t_\kappa)$ is a power chain, it follows that for every $1 \le \nu \le \kappa$ we can find a receiver $r_\nu \in \mathcal{R}$ such that

$$r_\nu \in \mathcal{R}_{t_\nu} \Bigg\backslash \bigcup_{\eta=1}^{\nu-1} \mathcal{R}_{t_\eta}. \tag{81}$$

Thus, Receiver r_ν can hear Transmitter t_ν

$$(r_\nu, t_\nu) \notin \mathcal{Z} \tag{82}$$

but it is uninfluenced by the transmitters $t_1,\ldots,t_{\nu-1}$

$$(r_\nu, t_\eta) \in \mathcal{Z}, \qquad \eta = 1,\ldots,\nu-1. \tag{83}$$

(Receiver r_ν may be affected by transmitters $t_{\nu+1},\ldots,t_\kappa$, but those, as we shall see, will be chosen to have powers that are much smaller than the power assigned to transmitter t_ν.) Let $r_1,\ldots,r_\kappa$ thus satisfy (82) and (83).

The mutual information $I(\boldsymbol{X};\boldsymbol{Y})$ can be now lower bounded as follows:

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &= I\left(\{X(t_\nu)\}_{\nu=1}^\kappa;\boldsymbol{Y}\right) \\ &= \sum_{\nu=1}^\kappa I\left(X(t_\nu);\boldsymbol{Y} \,\middle|\, \{X(t_\eta)\}_{\eta=\nu+1}^\kappa\right) \\ &\ge \sum_{\nu=1}^\kappa I\left(X(t_\nu);Y(r_\nu) \,\middle|\, \{X(t_\eta)\}_{\eta=\nu+1}^\kappa\right) \\ &\ge \sum_{\nu=1}^\kappa I(X(t_\nu);Y(r_\nu)). \end{aligned} \tag{84}$$

Here, the first equality follows by (77); the second by the chain rule; the subsequent inequality by restricting the set of observables in each of the terms; and the final inequality because the components of $\boldsymbol{X}$ are independent.

We shall next show that with a judicious choice of the constants

$$\{x_{\min,\nu}\}, \{x_{\max,\nu}\}, \qquad \nu = 1,\ldots,\kappa$$

in (78) we can guarantee that each of the κ terms in (84) grow double-logarithmically in the SNR.

Consider the expression

$$I(X(t_\nu);Y(r_\nu)) \tag{85}$$

for some $1 \le \nu \le \kappa$. By (77) and (83) it follows that we can express $Y(r_\nu)$ as

$$Y(r_\nu) = H(r_\nu,t_\nu)X(t_\nu) + W(r_\nu) \tag{86}$$

where

$$W(r_\nu) = \sum_{\eta=\nu+1}^\kappa H(r_\nu,t_\eta)X(t_\eta) + Z(r_\nu), \quad 1 \le \nu \le \kappa. \tag{87}$$

Note that the term $W(r_\nu)$ cannot be treated as independent additive noise because $W(r_\nu)$ and $H(r_\nu,t_\nu)$ may be dependent. (This dependence comes about because $H(r_\nu,t_\nu)$ may depend on $H(r_\nu,t_\eta)$ for $\eta = \nu+1,\ldots,\kappa$.) However, conditional on $H(r_\nu,t_\nu)$, the random variables $X(t_\nu)$ and $W(r_\nu)$ are independent, i.e.,

$$X(t_\nu) \multimap\!\!\!- H(r_\nu,t_\nu) \multimap\!\!\!- W(r_\nu) \tag{88}$$

forms a Markov chain. To analyze $I(X(t_\nu);Y(r_\nu))$ we shall use the following lemma.

Lemma 4: 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 $X \multimap\!\!\!- H \multimap\!\!\!- W$ forms a Markov chain.[2] Then

$$I(X;HX+W) \ge h(X) - \mathsf{E}[\log|X|^2] + \mathsf{E}[\log|H|^2] - \mathsf{E}\left[\log\left(\pi e\left(\sigma_H + \frac{\sigma_W}{|X|}\right)^2\right)\right] \tag{89}$$

where $\sigma_W^2 \ge 0$ and $\sigma_H^2 > 0$ are the variances of W and H, respectively. Consequently, if $|X| \ge x_{\min}$ with probability one for some positive constant $x_{\min} > 0$, then

$$I(X;HX+W) \ge h(X) - \mathsf{E}[\log|X|^2] + \mathsf{E}[\log|H|^2] - \log\left(\pi e\left(\sigma_H + \frac{\sigma_W}{x_{\min}}\right)^2\right), \qquad |X| \ge x_{\min} \quad \text{a.s.} \tag{90}$$

If, additionally, X is circularly symmetric, then

$$I(X;HX+W) \ge h(\log|X|^2) + \log\pi + \mathsf{E}[\log|H|^2] - \log\left(\pi e\left(\sigma_H + \frac{\sigma_W}{x_{\min}}\right)^2\right), \qquad |X| \ge x_{\min} \quad \text{circ. sym.} \tag{91}$$

Proof of Lemma 4: First note that the assumptions that X has a finite second moment and finite differential entropy guarantee that the logarithm of its magnitude is of finite expectation [3, Lemma 7.7] so that the lemma's claim is meaningful.

[2]This is equivalent to X being independent of (H,W).

The proof proceeds by expressing $I(X; HX+W)$ as

$$I(X; HX+W) = h(HX+W) - h(HX+W \mid X) \quad (92)$$

and by then bounding the terms on the RHS. We begin with the first

$$\begin{aligned} h(HX+W) &\geq h(HX+W \mid H) \\ &\geq h(HX \mid H) \\ &= h(X) + \mathsf{E}[\log |H|^2] \end{aligned} \quad (93)$$

where the first inequality follows because conditioning (on H) cannot increase differential entropy; the second because conditional on H, the random variables X and W are independent; and the subsequent equality from the behavior of differential entropy of complex random variables under deterministic scaling and from the independence of X and H.

As to the other term in (92), we note that conditional on $X = x$, the differential entropy of the random variable $HX + W$ is upper-bounded by that of a circularly symmetric Gaussian of equal variance. Hence,

$$\begin{aligned} h(HX+W \mid X) &= \mathsf{E}[\log |X|^2] + h\left(H + \frac{W}{X} \,\middle|\, X\right) \\ &\leq \mathsf{E}[\log |X|^2] + \mathsf{E}\left[\log \pi e \cdot \mathsf{Var}\left(H + \frac{W}{X} \,\middle|\, X\right)\right] \\ &\leq \mathsf{E}[\log |X|^2] + \mathsf{E}\left[\log \pi e \left(\sigma_H + \frac{\sigma_W}{|X|}\right)^2\right] \end{aligned} \quad (94)$$

where σ_H^2 and σ_W^2 are the respective variances of H and W. Here the last inequality follows from the Cauchy–Schwarz inequality, from the independence of H and X, and from the independence of W and X.

Combining (93) and (94) with (92) yields (89), which combines with the monotonicity of the logarithm function to imply (90). Finally, to obtain (91) we note that if X is circularly symmetric then

$$h(X) - \mathsf{E}[\log |X|^2] = h(\log |X|^2) + \log \pi \quad (95)$$

which follows, for example, from [3, eqs. (320) and (316)]. □

We continue the proof of the direct part of Theorem 1 by applying Lemma 4 to the analysis of (85)–(87) with $X(t_\nu)$, $H(r_\nu, t_\nu)$, and $W(r_\nu)$ replacing X, H, and W, respectively. To proceed we need an upper bound on the variance of $W(r_\nu)$. Such a bound can be derived using the Cauchy–Schwarz inequality. From (87) we have

$$\begin{aligned} \mathsf{Var}(W(r_\nu)) &\leq \mathsf{E}[|W(r_\nu)|^2] \\ &= \mathsf{E}[|Z(r_\nu)|^2] + \mathsf{E}\left[\left|\sum_{\eta=\nu+1}^{\kappa} H(r_\nu, t_\eta) X(t_\eta)\right|^2\right] \\ &\leq 1 + \sum_{\eta=\nu+1}^{\kappa} \mathsf{E}[|H(r_\nu, t_\eta)|^2] \cdot \sum_{\eta=\nu+1}^{\kappa} \mathsf{E}[|X(t_\eta)|^2] \\ &\leq 1 + \mathsf{E}\left[\|\mathbb{H}\|_F^2\right] \cdot (\kappa - \nu) \max_{\nu < \eta \leq \kappa} x_{\max,\eta}^2 \end{aligned} \quad (96)$$

where the second inequality follows from the Cauchy–Schwarz inequality

$$\left|\sum_{\eta=\nu+1}^{\kappa} H(r_\nu, t_\eta) X(t_\eta)\right|^2 \leq \left(\sum_{\eta=\nu+1}^{\kappa} |H(r_\nu, t_\eta)|^2\right) \cdot \left(\sum_{\eta=\nu+1}^{\kappa} |X(t_\eta)|^2\right) \quad (97)$$

and from the independence of $\mathbb{H}$ and $\boldsymbol{X}$.

It thus follows from Lemma 4 and from (96) that the mutual information in (85) will satisfy

$$\overline{\lim_{\mathcal{E}\to\infty}} \{\log\log \mathcal{E} - I(X(t_\nu); Y(r_\nu))\} < \infty \quad (98)$$

for $\{X(t_\eta)\}_{\eta=\nu+1}^{\kappa}$ satisfying (with probability one)

$$x_{\min,\eta} \leq |X(t_\eta)| \leq x_{\max,\eta}, \qquad \eta = \nu+1, \ldots, \kappa \quad (99)$$

whenever both

$$\lim_{\mathcal{E}\to\infty} \frac{x_{\min,\nu}^2}{1 + \mathsf{E}\left[\|\mathbb{H}\|_F^2\right] \cdot (\kappa - \nu) \max_{\nu<\eta\leq\kappa} x_{\max,\eta}^2} = \infty \quad (100)$$

—so that the last term on the RHS of (91) tends to a constant, i.e., to the constant $\log(\pi e \cdot \mathsf{Var}(H(r_\nu, t_\nu)))$—and

$$\overline{\lim_{\mathcal{E}\to\infty}} \left\{\log\log \mathcal{E} - \log\log \frac{x_{\max,\nu}^2}{x_{\min,\nu}^2}\right\} < \infty \quad (101)$$

—so that by (80) the first term on the RHS of (91) has the right asymptotic growth.

To conclude the proof it is thus only required to find choices for $\{x_{\min,\eta}, x_{\max,\eta}\}_{\eta=1}^{\kappa}$ that will guarantee that both (100) and (101) hold. An example of such a choice is (for large enough $\mathcal{E}$)

$$x_{\max,\eta}^2 = \mathcal{E}^{1-(\eta-1)/\kappa}, \qquad \eta = 1, \ldots, \kappa \quad (102)$$

$$x_{\min,\eta}^2 = \mathcal{E}^{1-\eta/\kappa} \log \mathcal{E}, \qquad \eta = 1, \ldots, \kappa. \quad (103)$$

IV. Discussion and Summary

In this paper, we considered noncoherent fading networks with vector-valued additive and matrix-valued multiplicative noises. We have shown that, at high SNR, the capacity of the network grows like an integer multiple of log log SNR. This integer multiple is determined by the location of the deterministic zeros of the fading matrix. Loosely speaking, this integer can be viewed as the effective number of parallel channels that can be supported by the network, i.e., as the maximal number of point-to-point single-user scalar channels that can be supported by the network in a manner that will allow, with proper power allocation, negligible cross-interference.

It is felt that this integer is an important parameter of the network, but that more parameters are needed to obtain more precise approximations of the system's throughput. For example, to assess the rates above which every increase in throughput of one bit per channel-use requires squaring the SNR one can consider the network's fading number χ. This can be defined, analogously to the MIMO fading number [3], as

$$\chi = \overline{\lim_{\mathrm{SNR}\to\infty}} \{C(\mathrm{SNR}) - \kappa^* \cdot \log\log \mathrm{SNR}\} \quad (104)$$

where κ^* is the length of the longest power chain.

Not surprisingly, the evaluation of χ is much more elaborate than that of κ^*. That χ is not just a function of κ^* is readily seen from the single-user MIMO case ($\mathcal{Z} = \emptyset$) where $\kappa^* = 1$ but where the fading number is highly dependent on the number of transmit and receive antennas [7]. Needless to say, as in the SISO case [3], the memory in the process also plays a key role in determining the fading number.

Consider, for example, the case where the $n_{\mathrm{R}} \cdot n_{\mathrm{T}}$ components of $\{\mathbb{H}_k\}$ are i.i.d., each being a zero-mean, unit-variance, circularly symmetric, stationary, Gaussian process of spectral distribution function $F(\,\cdot\,)$ with corresponding positive prediction error

$$\epsilon^2 = \exp\left\{ \int_{-1/2}^{1/2} \log F'(\lambda)\,\mathrm{d}\lambda \right\}. \tag{105}$$

In this case, one can show using techniques very similar to those in [8] that

$$\chi_{\mathrm{SU,FB}} \leq \chi_{\mathrm{SU,i.i.d.}} + n_{\mathrm{R}} \log \frac{1}{\epsilon^2} \tag{106}$$

where the term $\chi_{\mathrm{SU,\,i.i.d.}}$ corresponds to the fading number of the network under full cooperation conditions but with memoryless fading of the same marginal law as that of the original network, i.e., i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ components. On the other hand, by analyzing more precisely the power allocation scheme (102), (103) we obtain that in the absence of feedback and under multiple-access conditions

$$\chi_{\mathrm{MAC}} \geq \kappa^* \left(-1 - \gamma + \log \frac{1}{\epsilon^2} \right) - \kappa^* \log \kappa^* \tag{107}$$

where γ denotes Euler's constant and where we are relying on the asymptotic results on the capacity of SISO Rayleigh-fading channels with memory [3, Corollary 4.42]. While the bounds (106) and (107) can be quite loose in some scenarios[3] there are cases where they give reasonable estimates, e.g., for Wyner's cellular model (33) with the number of cells M being large.

We have already noted that Theorem 1 holds under very general conditions. It does not, for example, require that the fading be Gaussian. It suffices that the fading be of finite second moment and that the nonzero components have finite joint differential entropy rate. In fact, the additive noise need not be Gaussian either. It suffices that it be stationary and ergodic with a finite second moment and finite differential entropy rate. It is thus natural to ask whether we can also relax the conditions on the zeros. Can the result be extended to networks where some components of the fading matrix are deterministic but nonzero or deterministically dependent on others? This requires some care and is beyond the scope of this paper. To see some of the difficulties, first note that in the $n_{\mathrm{R}} = n_{\mathrm{T}} = 1$ case there is a dramatic difference between the sole component of the fading matrix being zero or some deterministic nonzero. The former leads to zero capacity whereas the latter to a nonfading Gaussian channel, and hence to a logarithmic growth of capacity. Yet another example is the 2×2 MIMO case. If the two rows are identical with each being of finite differential entropy then capacity grows double-logarithmically. If the matrix is diagonal with the diagonal elements being identical (as in the SISO block-constant fading model), capacity grows like $1/2 \cdot \log \mathrm{SNR}$ [10].

Theorem 1 can, however, be somewhat generalized using [3, Lemma 4.7] which demonstrates that the fading number (and hence also the pre-loglog κ^*) is invariant under deterministic nonsingular matrix multiplication

$$\chi_{\mathrm{SU,FB}}(\{\mathbb{H}_k\}) = \chi_{\mathrm{SU,FB}}(\{\mathsf{G}\mathbb{H}_k\mathsf{F}\}), \quad \det(\mathsf{G}), \det(\mathsf{F}) \neq 0 \tag{108}$$

and

$$\chi_{\mathrm{MAC}}(\{\mathbb{H}_k\}) = \chi_{\mathrm{MAC}}(\{\mathsf{G}\mathbb{H}_k\}), \quad \det(\mathsf{G}) \neq 0. \tag{109}$$

It has been pointed out to me by Shlomo Shamai that in some broadcast scenarios the fading levels experienced by the different users may be highly correlated so that the assumption that the nonzero components of the fading matrix are of finite joint differential entropy may be violated. Such scenarios can be sometimes addressed using our results by noting that in broadcast scenarios the achievable rates are determined by the marginals of the network law [11]. Thus, in some such scenarios one can replace the fading matrix with a fading matrix whose rows are independent, but such that each row is of the same law as that in the original matrix.

Appendix

In this appendix we show that if the stationary and ergodic fading process $\{\mathbb{H}_k\}$ satisfies (9) and (14) then the difference between the channel's feedback capacity and the capacity of the memoryless channel of i.i.d. fading $\{\tilde{\mathbb{H}}_k\}$, where the law of $\tilde{\mathbb{H}}_k$ is identical to the law of $\mathbb{H}_1$, is bounded in the SNR.

This proof is taken almost verbatim from [6] and is included here only for the sake of completeness. We denote the message to be transmitted by M, and we assume that the time-k transmitted input $\boldsymbol{X}_k$ is now a function of M and of the previous outputs $\boldsymbol{Y}_1^{k-1}$. The proof, as in, for example, [11, Sec. 8.12], is based on Fano's inequality and on an upper bound on $n^{-1} \cdot I(M; \boldsymbol{Y}_1^n)$.

$$\begin{aligned}
&\frac{1}{n} I\left(M; \boldsymbol{Y}_1^n\right) \\
&\quad = \frac{1}{n} \sum_{k=1}^{n} I\left(M; \boldsymbol{Y}_k \,\middle|\, \boldsymbol{Y}_1^{k-1}\right) && (110)\\
&\quad = \frac{1}{n} \sum_{k=1}^{n} \left(I\left(M, \boldsymbol{Y}_1^{k-1}; \boldsymbol{Y}_k\right) - I\left(\boldsymbol{Y}_1^{k-1}; \boldsymbol{Y}_k\right)\right) && (111)\\
&\quad \leq \frac{1}{n} \sum_{k=1}^{n} I\left(M, \boldsymbol{Y}_1^{k-1}; \boldsymbol{Y}_k\right) && (112)\\
&\quad \leq \frac{1}{n} \sum_{k=1}^{n} I\left(M, \boldsymbol{Y}_1^{k-1}, \mathbb{H}_1^{k-1}; \boldsymbol{Y}_k\right) && (113)\\
&\quad = \frac{1}{n} \sum_{k=1}^{n} I\left(M, \boldsymbol{Y}_1^{k-1}, \mathbb{H}_1^{k-1}, \boldsymbol{X}_k; \boldsymbol{Y}_k\right) && (114)\\
&\quad = \frac{1}{n} \sum_{k=1}^{n} \Big(I\left(\mathbb{H}_1^{k-1}, \boldsymbol{X}_k; \boldsymbol{Y}_k\right) \\
&\qquad + I\left(M, \boldsymbol{Y}_1^{k-1}; \boldsymbol{Y}_k \,\middle|\, \mathbb{H}_1^{k-1}, \boldsymbol{X}_k\right)\Big) && (115)
\end{aligned}$$

[3]The upper bound is, for example, loose in single-input multiple-output (SIMO) scenarios [9], and the lower bound can be quite loose in MIMO channels [7].

$$= \frac{1}{n}\sum_{k=1}^{n} I\left(\mathbb{H}_1^{k-1}, \boldsymbol{X}_k; \boldsymbol{Y}_k\right) \tag{116}$$

$$= \frac{1}{n}\sum_{k=1}^{n} \left(I(\boldsymbol{X}_k; \boldsymbol{Y}_k) + I\left(\mathbb{H}_1^{k-1}; \boldsymbol{Y}_k \,\middle|\, \boldsymbol{X}_k\right)\right). \tag{117}$$

Here the first two equalities follow from the chain rule; the subsequent inequality from the nonnegativity of mutual information; the following inequality from adding random matrices; the subsequent equality follows since $\boldsymbol{X}_k$ is a deterministic function of M and $\boldsymbol{Y}_1^{k-1}$; then we have used the chain rule again; (116) follows since

$$I\left(M, \boldsymbol{Y}_1^{k-1}; \boldsymbol{Y}_k \,\middle|\, \mathbb{H}_1^{k-1}, \boldsymbol{X}_k\right) = 0; \tag{118}$$

and finally we have used the chain rule once more.

The term $I(\boldsymbol{X}_k; \boldsymbol{Y}_k)$ does not depend on the memory in the fading process and is thus identical for $\{\mathbb{H}_k\}$ and for $\{\tilde{\mathbb{H}}_k\}$. As for the other term, we upper-bound it as follows:

$$\begin{aligned}
I&\left(\mathbb{H}_1^{k-1}; \boldsymbol{Y}_k \,\middle|\, \boldsymbol{X}_k\right) \\
&\leq I\left(\mathbb{H}_1^{k-1}; \boldsymbol{Y}_k, \mathbb{H}_k \,\middle|\, \boldsymbol{X}_k\right) &(119)\\
&= I\left(\mathbb{H}_1^{k-1}; \mathbb{H}_k \,\middle|\, \boldsymbol{X}_k\right) + I\left(\mathbb{H}_1^{k-1}; \boldsymbol{Y}_k \,\middle|\, \boldsymbol{X}_k, \mathbb{H}_k\right) &(120)\\
&= I\left(\mathbb{H}_1^{k-1}; \mathbb{H}_k \,\middle|\, \boldsymbol{X}_k\right) &(121)\\
&\leq I\left(\mathbb{H}_1^{k-1}, \boldsymbol{X}_k; \mathbb{H}_k\right) &(122)\\
&= I\left(\mathbb{H}_k; \mathbb{H}_1^{k-1}\right) + I\left(\mathbb{H}_k; \boldsymbol{X}_k \mid \mathbb{H}_1^{k-1}\right) &(123)\\
&= I\left(\mathbb{H}_k; \mathbb{H}_1^{k-1}\right). &(124)
\end{aligned}$$

The feedback capacity of the channel with fading $\{\mathbb{H}_k\}$ can thus exceed the capacity of the memoryless fading channel with equal-marginal fading $\{\tilde{\mathbb{H}}_k\}$ by at most $I(\mathbb{H}_k; \mathbb{H}_1^{k-1})$, which does not depend on the SNR and which, by assumption (14), is finite.

ACKNOWLEDGMENT

Discussions with Stefan M. Moser and Jatin Thukral are gratefully acknowledged. The comments of the Associate Editor Yossef Steinberg and the anonymous referees are also much appreciated.

REFERENCES

[1] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1713–1727, Nov. 1994.

[2] S. Shamai (Shitz) and A. D. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels—Parts I and II," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1877–1911, Nov. 1997.

[3] 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.

[4] S. A. Jafar, "Too much mobility limits the capacity of wireless ad-hoc networks," *IEEE Trans. Inf. Theory*, to be published.

[5] S. Hanly and P. Whiting, "Information-theoretic capacity of multi-receiver networks," *Telecommun. Syst.*, vol. 1, no. 1–42, 1993.

[6] S. M. Moser, "Duality-based bounds on channel capacity," Ph.D. dissertation, Swiss Federal Inst. Technol. (ETH), Zurich, Switzerland, 2004.

[7] T. Koch and A. Lapidoth, "Degrees of freedom in noncoherent stationary MIMO fading channels," in *Proc. Winter School on Coding and Information Theory*, Bratislava, Slovakia, Feb. 2005, pp. 91–97.

[8] A. Lapidoth and S. M. Moser, "On noncoherent fading channels with feedback," in *Proc. Winterschool on Coding and Information Theory*, Bratislava, Slovakia, Feb. 2005, pp. 113–118.

[9] ——, "The fading number of SIMO fading channels with memory," in *Proc. 2004 Int. Symp. Information Theory and Its Applications (ISITA'04)*, Parma, Italy, Oct. 2004.

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

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.