The Fading Number of Multiple-Input Multiple-Output Fading Channels with Memory

Stefan M. Moser, *Member, IEEE*

***Abstract*—The fading number of a general (not necessarily Gaussian) regular multiple-input–multiple-output (MIMO) fading channel with arbitrary temporal and spatial memory is derived. The channel is assumed to be noncoherent, *i.e.*, neither receiver nor transmitter have knowledge about the channel state, but they only know the probability law of the fading process. The fading number is the second term in the asymptotic expansion of channel capacity when the signal-to-noise ratio (SNR) tends to infinity. It is related to the border of the high-SNR region with double-logarithmic capacity growth.**

It is shown that the fading number can be achieved by an input that is the product of two independent processes: a stationary and circularly symmetric direction- (or unit-) vector process whose distribution is chosen such that the fading number is maximized, and a nonnegative magnitude process that is independent and identically distributed (IID) and escapes to infinity. Additionally, in the more general context of an arbitrary stationary channel model satisfying some weak conditions on the channel law, it is shown that there exists an optimal input distribution that is stationary apart from some edge effects.

***Index Terms*—Channel capacity, circular symmetry, escaping to infinity, fading number, flat fading, high signal-to-noise ratio (SNR), memory, multiple-input–multiple-output (MIMO), noncoherent detection, stationary input distribution.**

I. Introduction

A. General Background

FUTURE mobile communication systems will have to provide much higher data rates than what currently is available. To be able to design such systems we need to study wireless communication channels and try to understand how their behavior depends on various parameters like the number of antennas at the transmitter and receiver, the available power, feedback, or the implicitly available memory in the channel. An important parameter that is part of this theoretical understanding is the so-called *channel capacity*. It describes the ultimate physical limit on the maximum rate for which reliable information transmission still is possible. Note that this parameter is theoretical in the sense that we only assume limited power at the transmitter, but ignore other constraints of real systems like a maximum allowed transmission delay or limited computing resources.

Manuscript received April 06, 2007; revised May 02, 2008. Current version published May 20, 2009. This work was supported by the Industrial Technology Research Institute (ITRI), Zhudong, Taiwan, under Contracts G1-95003 and G1-96001 and by the National Science Council (NSC) under Grant NSC 96-2221-E-009-012-MY3. The material in this work was presented in part at the IEEE International Symposium on Information Theory (ISIT), Nice, France, June 2007.

The author is with the Department of Communication Engineering, National Chiao Tung University (NCTU), Hsinchu, Taiwan, (e-mail: stefan.moser@ieee.org).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2009.2018180

In the situation of wireless communication channels, the channel capacity is limited due to two main sources of transmission errors. Firstly, the receiver introduces thermal noise that can be well modeled by an additive random noise process. Secondly, because the signals are electromagnetic waves transmitted through air, the received signals suffer from random fluctuations in the magnitude and phase. This effect is known as *fading* and can be described by a multiplicative random noise process.

While the additive noise can be well approximated by an independent and identically distributed (IID) complex Gaussian process for almost all channels of interest, the detailed properties of the multiplicative noise depends on many parameters, system-internal and -external, and should therefore be kept as general as possible. Unfortunately, the analysis of the channel capacity in such generality is very difficult so that commonly the model is simplified in certain aspects.

One possible simplification is to assume that the receiver *perfectly* knows the fading realizations. This assumption is based on the idea that the transmitter will firstly transmit some known training symbols from which the receiver learns the current state of the multiplicative noise process. The capacity is then computed without taking into account the estimation scheme. It is common to call this *the coherent capacity of fading channels*. Such an approach will definitely lead to an overly optimistic capacity value because

- even with a large amount of training data the channel knowledge will never be perfect, but only an estimate; and because
- the data rate that is wasted for the training symbols is completely ignored.

In this paper we will not make this simplification, but stick with *noncoherent detection* where the receiver has no additional knowledge about the channel state. Note that the receiver is free to do anything in its power to gain knowledge about the fading based on the received signals.

Marzetta and Hochwald [2] simplify the noncoherent channel model by assuming that during blocks consisting of several symbol periods the fading remains constant, while the fading coefficients corresponding to different blocks are assumed to be independent. This model is generally known as *block fading model*. Note that it is pessimistic to assume that the blocks are independent of each other because memory provides additional information about the current fading level which in general will increase capacity. However, it is more problematic to conjecture that the fading coefficients are *perfectly* constant during one block. This means that for high enough signal-to-noise ratios (SNR) and for long enough blocks the receiver can get an (almost) perfect estimate of the fading value

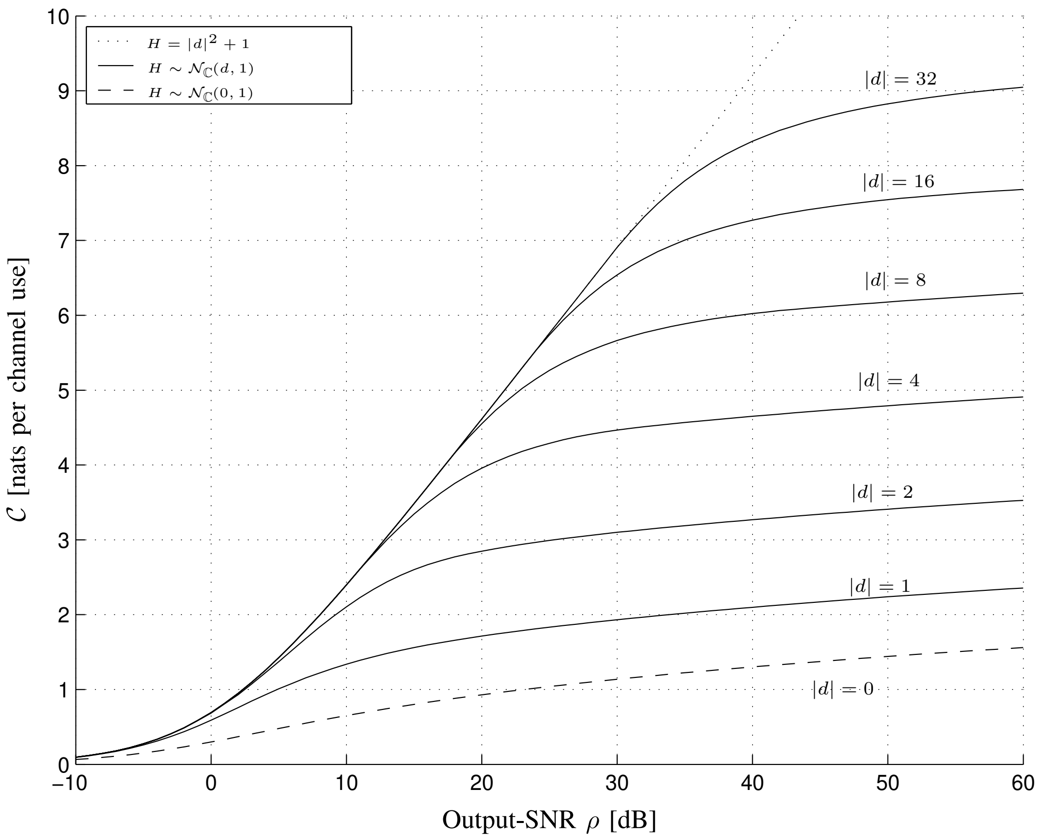

Fig. 1. An upper bound on the capacity of a Rician fading channel as a function of the output-SNR $\rho = (1 + |d|^2)\text{SNR}$ for different values of the specular component d. The dashed line corresponds to the situation of a Rayleigh fading channel with a zero line-of-sight component $d = 0$. The dotted line depicts the capacity of an additive Gaussian noise channel (without fading) of equal output-SNR ρ, namely $\log(1 + \rho)$.

within a block and use this knowledge to decode the received signal similarly to coherent detection. For larger SNR this seems to be overly optimistic. Indeed, as shown in [2] for single-input–single-output (SISO) Gaussian block fading and in [3] for multiple-input–multiple-output (MIMO) Gaussian block fading, the capacity of the block fading channel grows logarithmically in the SNR at high SNR, *i.e.*, the capacity has the same growth rate as the coherent capacity (and, as a matter of fact, as the capacity of an additive noise channel without fading, too).

In [4] Liang and Veeravalli generalize the SISO Gaussian block fading model by allowing some temporal correlation between the different fading coefficients within one block. They show that the rank of the block correlation matrix is crucial when determining the high-SNR channel capacity: if we have a rank-deficient correlation matrix, the effect of perfect predictability comes into play again similar to the situation of Marzetta and Hochwald [2]. This then again leads to a logarithmic growth of capacity. For a full-rank correlation matrix this is not true anymore. In this case the channel model reduces to a special case of the more general model described next.

The most general models only restrict the random noise processes to be stationary and ergodic, with additional variations in the exact fading law, the number of antennas, and the memory [5]–[13]. In [5] the authors investigate a memoryless SISO Rayleigh fading channel and derive some bounds. In [6] it is shown that the capacity-achieving input distribution for the memoryless SISO Rayleigh fading channel is discrete. In [7]–[9] the channel model is then generalized to MIMO and to general non-Gaussian fading distributions (possibly with memory) where the fading process is assumed to be *regular*, *i.e.*, its differential entropy rate is finite. The complementary situation of *nonregular* fading processes has been studied in [10]–[13].

It turns out that the capacity at high SNR is very sensitive to the exact assumptions of the channel model, in particular to the regularity assumption. If we assume a regular fading process, then the capacity grows only double-logarithmically in the SNR at high SNR [7, Theorem 4.2], [9, Theorem 6.10]. This means that at high power such a channel becomes extremely power-inefficient in the sense that whenever the capacity shall be increased by only one bit, the SNR needs to be squared or, on a dB-scale, the SNR needs to be doubled! So the high-SNR behavior is dramatically different from the optimistic models mentioned above.

For nonregular Gaussian fading the high-SNR behavior of capacity depends on the specific power spectral density and can be anything between the logarithmic and the double-logarithmic growth [11].

However, it is interesting to observe that for low SNR the difference between the different models is relatively small.

Indeed, the capacity of regular fading channels usually shows a very distinct turn at a certain SNR level where the growth rate changes from logarithmic to double-logarithmic. As an example Figure 1 shows the capacity of a noncoherent Rician fading channel with various values of the line-of-sight component. One clearly sees that the capacity curve, while growing logarithmically at lower SNR, suddenly has a sharp bend at a certain threshold where its growth becomes very slow. Moreover, one sees that this threshold depends strongly on the channel law, *i.e.*, on the line-of-sight component.

We conclude that at lower SNR the exact choice of the channel model has only a small impact on the capacity analysis, *i.e.*, the described simplifications (even the assumption of coherent detection) are useful in that regime. However, at high SNR many simplifications seem to lose their validity. Based on this observation we immediately ask ourselves whether we can say something about the separation between these two regimes. Particularly, in the situation of a regular fading model, we would like to know more about the threshold between the efficient low- to medium-SNR regime where the capacity grows logarithmically in the SNR and the highly inefficient high-SNR regime with a double-logarithmic growth. The dependence of this threshold on some system parameters like the number of antennas, the memory in the channel, or the availability of feedback might give valuable insight in good design criteria of wireless and mobile communication systems.

B. The Fading Number

In an attempt to more precisely quantify the mentioned threshold between the power-efficient and the power-inefficient regime, [7, Sec. IV.C] and [9, Sec. 6.5.2] define the *fading number* χ as the second term in the high-SNR asymptotic expansion of capacity, *i.e.*, at high SNR the channel capacity can be expressed as

$$\mathcal{C}(\text{SNR}) = \log(1+\log(1+\text{SNR})) + \chi + o(1). \tag{1}$$

Here, $o(1)$ denotes some terms that tend to zero as $\text{SNR} \uparrow \infty$.

Based on (1) we define the *high-SNR regime* to be the region where the $o(1)$-terms in (1) are negligible, *i.e.*, we say that a wireless communication system operates in the inefficient high-SNR regime if its capacity can be well approximated by

$$\mathcal{C}(\text{SNR}) \approx \log(1+\log(1+\text{SNR})) + \chi. \tag{2}$$

The important point to notice is that due to the extremely slow growth of $\log(1+\log(1+\text{SNR}))$ the fading number χ is usually the dominant term in the lower range of the high-SNR regime. In other words, $\log(1+\log(1+\text{SNR}))$ is only much larger than χ for extremely large values of SNR. An illustration of this behavior is given in Figure 2.

The fading number is therefore strongly connected to the point where the bend of the capacity curve occurs. As an example consider the following situation [13], [14]: assume for the moment that the threshold SNR_0 lies somewhere between 30 and 80 dB (it can be shown that this is a reasonable assumption for many channels that are encountered in practice).

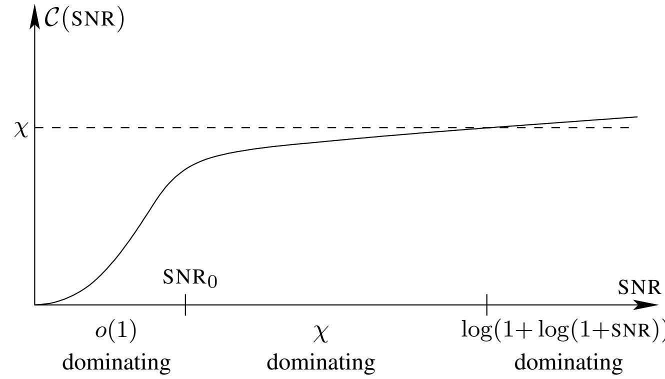

Fig. 2. Illustration of the different regimes of a typical regular fading channel. At low SNR the $o(1)$ terms are dominant, in the lower range of the high-SNR regime the fading number χ is dominant, and only at very high SNR the $\log(1+\log(1+\text{SNR}))$ term takes the lead.

In this case, the threshold capacity $\mathcal{C}_0 = \mathcal{C}(\text{SNR}_0)$ must be somewhere in the following interval:

$$\log\left(1+\log\left(1+30\text{ dB}\right)\right)+\chi \leq \mathcal{C}_0 \leq \log\left(1+\log\left(1+80\text{ dB}\right)\right)+\chi \tag{3}$$

i.e.,

$$\chi + 2.1\text{ nats} \leq \mathcal{C}_0 \leq \chi + 3\text{ nats}. \tag{4}$$

Hence, even though we have assumed a wide range from 30 to 80 dB, the capacity changes only very little (this is because the $\log\log$-term is growing extremely slowly). Hence, we get the following *rule of thumb*.

Conjecture 1 ([13], [14]): A communication system over a noncoherent regular fading channel[1] that operates at rates appreciably above $\chi + 2$ nats is in the high-SNR regime and therefore extremely power-inefficient.

The fading number can therefore be regarded as quality attribute of the channel: the larger the fading number is, the higher is the maximum rate at which the channel can be used without being extremely power-inefficient. It follows from this observation that a good system design will aim at achieving a large fading number.

The rest of this paper will concentrate on the analysis of the fading number of general MIMO fading channels with memory.

So far explicit expressions for the fading number are known in the special situation of general SISO fading channels with memory[2] [7, Theorem 4.41], [9, Theorem 6.41]:

$$\chi\big(\{H_k\}\big) = \log\pi + \mathsf{E}\left[\log|H_0|^2\right] - h\big(\{H_k\}\big) \tag{5}$$

and of general single-input–multiple-output (SIMO) fading channels with memory [8, Theorem 1], [9, Theorem 6.44]:

$$\chi\big(\{\mathbf{H}_k\}\big) = \chi_{\text{IID}}\left(\mathbf{H}_0 \,\middle|\, \mathbf{H}_{-\infty}^{-1}, \{\hat{\mathbf{H}}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\infty}\right). \tag{6}$$

[1]For more details about the exact assumptions made in this paper we refer to Section III.

[2]For an explanation of the notation used in this paper we refer to Section II.

Here $\chi_{\text{IID}}(\mathbf{H} \,|\, \mathbf{S})$ denotes the memoryless SIMO fading number with partial side-information $\mathbf{S}$ at the receiver [7, Note 4.31], [9, Eq. (6.194)]:

$$\chi_{\text{IID}}(\mathbf{H} \,|\, \mathbf{S}) = h_{\lambda}(\hat{\mathbf{H}} e^{i\Theta} \,|\, \mathbf{S}) + n_{\text{R}} \mathsf{E}\left[\log \|\mathbf{H}\|^2\right] - \log 2 - h(\mathbf{H} \,|\, \mathbf{S}). \quad (7)$$

The fading number of the multiple-input–single-output (MISO) fading channel has only been derived for the memoryless case [7, Theorem 4.27], [9, Theorem 6.27]:

$$\chi(\mathbf{H}^{\mathsf{T}}) = \sup_{\|\hat{\mathbf{x}}\|=1} \left\{ \log \pi + \mathsf{E}\left[\log |\mathbf{H}^{\mathsf{T}} \hat{\mathbf{x}}|^2\right] - h(\mathbf{H}^{\mathsf{T}} \hat{\mathbf{x}}) \right\}. \quad (8)$$

This fading number is achievable by inputs that can be expressed as the product of a constant unit vector in $\mathbb{C}^{n_{\text{T}}}$ and a circularly symmetric, scalar, complex random variable of the same law that achieves the memoryless SISO fading number [7]. Hence, the asymptotic capacity of a memoryless MISO fading channel is achieved by beam-forming where the beam-direction is chosen not to maximize the SNR, but the fading number.

For MISO fading with memory some bounds have been found [15]–[17]:

$$\chi(\{\mathbf{H}_k^{\mathsf{T}}\}) \le \sup_{\hat{\mathbf{x}}_{-\infty}^{0}} \Big\{ \log \pi + \mathsf{E}\left[\log |\mathbf{H}_0^{\mathsf{T}} \hat{\mathbf{x}}_0|^2\right] - h\big(\mathbf{H}_0^{\mathsf{T}} \hat{\mathbf{x}}_0 \,\big|\, \{\mathbf{H}_\ell^{\mathsf{T}} \hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big) \Big\} \quad (9)$$

and

$$\chi(\{\mathbf{H}_k^{\mathsf{T}}\}) \ge \sup_{\hat{\mathbf{x}}} \Big\{ \log \pi + \mathsf{E}\left[\log |\mathbf{H}_0^{\mathsf{T}} \hat{\mathbf{x}}|^2\right] - h\big(\mathbf{H}_0^{\mathsf{T}} \hat{\mathbf{x}} \,\big|\, \{\mathbf{H}_\ell^{\mathsf{T}} \hat{\mathbf{x}}\}_{\ell=-\infty}^{-1}\big) \Big\}. \quad (10)$$

The MIMO case has been solved recently in the memoryless situation [18]:

$$\chi(\mathbb{H}) = \sup_{\substack{Q_{\hat{\mathbf{X}}} \\ \text{circ. sym.}}} \left\{ h_{\lambda}\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) + n_{\text{R}} \mathsf{E}\left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2\right] - \log 2 - h(\mathbb{H}\hat{\mathbf{X}} \,|\, \hat{\mathbf{X}}) \right\}. \quad (11)$$

This paper generalizes these special cases to the most general situation of MIMO fading channels with memory and specifies the fading number exactly. We remark that the proofs are based on several new preliminary results that are interesting by themselves. In particular we prove a theorem which states that the optimal input to a stationary channel may be assumed to be stationary.

C. Outline

The rest of this paper is organized as follows: after a section about notation we will introduce the channel model in detail in Section III. In Section IV some preliminary results will be given. In particular we will present there a new theorem which states that—apart from edge effects and some weak conditions on the channel model—a stationary channel model has a capacity-achieving input distribution that is stationary. The corresponding proofs are found in the appendix.

Section V then presents the main result, *i.e.*, the fading number of a general MIMO fading channel with memory. We will give an outline of the proof there. The details can be found in Appendices D to I.

In Section VI we will consider some interesting special cases, in particular the fading number of MISO fading with memory which has been unknown so far. Note that in parallel to this paper a second publication [19] will treat the important special case of Gaussian MIMO fading channels in detail.

We conclude in Section VII.

II. Notation

As is by now fairly customary, we usually try to use upper-case letters for random quantities and lower-case letters for their realizations. This rule becomes awkward when dealing with matrices because matrices are usually written in upper case even if they are deterministic. To better differentiate between scalars, vectors, and matrices we have resorted to using different fonts for the different quantities. Upper-case letters such as X are used to denote scalar random variables taking value in the reals $\mathbb{R}$ or in the complex plane $\mathbb{C}$. Their realizations are typically written in lower-case, *e.g.*, x. Random vectors in the m-dimensional complex Euclidean space $\mathbb{C}^m$ are described by bold face capitals, *e.g.*, $\mathbf{X}$; for their realizations we use bold lower-case, *e.g.*, $\mathbf{x}$. Deterministic matrices are denoted by upper-case letters but of a special font, *e.g.*, H; and random matrices are denoted using another special upper-case font, *e.g.*, $\mathbb{H}$.

However, there will be a few exceptions to these rules. Since they are widely used in the literature, we will stick with the common customary shape of the entropy $H(\cdot)$ of a discrete random variable and of the mutual information functional $I(\cdot;\cdot)$. Moreover, we have decided to use the capital Q to denote the probability distribution of an input of a channel. In particular, Q_X and $Q_{\mathbf{X}}$ denote the probability distribution of a random variable X and random vector $\mathbf{X}$, respectively. Given an alphabet $\mathcal{A}$ we denote the set of all probability distributions over $\mathcal{A}$ by $\mathcal{P}(\mathcal{A})$.

The capacity is denoted by $\mathcal{C}$, the energy per symbol by $\mathcal{E}$, and the signal-to-noise ratio is denoted by SNR.

We use the shorthand H_a^b for $(H_a, H_{a+1}, \dots, H_b)$. For more complicated expressions, such as

$$(\mathbf{H}_a^{\mathsf{T}} \hat{\mathbf{x}}_a, \mathbf{H}_{a+1}^{\mathsf{T}} \hat{\mathbf{x}}_{a+1}, \dots, \mathbf{H}_b^{\mathsf{T}} \hat{\mathbf{x}}_b)$$

we use the dummy variable ℓ to clarify notation: $\{\mathbf{H}_\ell^{\mathsf{T}} \hat{\mathbf{x}}_\ell\}_{\ell=a}^{b}$.

The subscript k is reserved to denote discrete time. Curly brackets are used to distinguish between a random process and its manifestation at time k: $\{X_k\}$ is a discrete random process over time, while X_k is the random variable of this process at time k.

Hermitian conjugation is denoted by $(\cdot)^{\dagger}$, and $(\cdot)^{\mathsf{T}}$ stands for the transpose (without conjugation) of a matrix or vector. We use $\|\cdot\|$ to denote the Euclidean norm of vectors or the Euclidean operator norm of matrices. That is,

$$\|\mathbf{x}\| \triangleq \sqrt{\sum_{t=1}^{m} |x^{(t)}|^2}, \quad \mathbf{x} \in \mathbb{C}^m \quad (12)$$

$$\|\mathsf{A}\| \triangleq \max_{\|\hat{\mathbf{w}}\|=1} \|\mathsf{A}\hat{\mathbf{w}}\|. \tag{13}$$

Thus, $\|\mathsf{A}\|$ is the maximal singular value of the matrix A.

The Frobenius norm of matrices is denoted by $\|\cdot\|_{\mathrm{F}}$ and is given by the square root of the sum of the squared magnitudes of the elements of the matrix, *i.e.*,

$$\|\mathsf{A}\|_{\mathrm{F}} \triangleq \sqrt{\mathrm{tr}\,(\mathsf{A}^{\dagger}\mathsf{A})} \tag{14}$$

where $\mathrm{tr}\,(\cdot)$ denotes the trace of a matrix. Note that for every matrix A

$$\|\mathsf{A}\| \leq \|\mathsf{A}\|_{\mathrm{F}} \tag{15}$$

as can be verified by upper-bounding the squared magnitude of each of the components of $\mathsf{A}\hat{\mathbf{w}}$ using the Cauchy-Schwarz inequality.

We will often split a complex vector $\mathbf{v} \in \mathbb{C}^m$ up into its magnitude $\|\mathbf{v}\|$ and its *direction*

$$\hat{\mathbf{v}} \triangleq \frac{\mathbf{v}}{\|\mathbf{v}\|} \tag{16}$$

where we reserve this notation exclusively for unit vectors, *i.e.*, throughout the paper every vector carrying a hat, $\hat{\mathbf{v}}$ or $\hat{\mathbf{V}}$, denotes a (deterministic or random, respectively) vector of unit length

$$\|\hat{\mathbf{v}}\| = \|\hat{\mathbf{V}}\| = 1. \tag{17}$$

To be able to work with such *direction vectors* we shall need a differential entropy-like quantity for random vectors that take value on the unit sphere in $\mathbb{C}^m$. Note that with respect to a probability distribution over $\mathbb{C}^m$, the surface of the unit sphere in $\mathbb{C}^m$ has zero measure such that the corresponding differential entropy is undefined. We therefore introduce a new probability space that only lives on the surface of the unit sphere in $\mathbb{C}^m$ and denote its measure by λ. If a random vector $\hat{\mathbf{V}}$ takes value in the unit sphere and has the density $p^{\lambda}_{\hat{\mathbf{V}}}(\hat{\mathbf{v}})$ with respect to λ, then we shall let

$$h_{\lambda}(\hat{\mathbf{V}}) \triangleq -\mathsf{E}\left[\log p^{\lambda}_{\hat{\mathbf{V}}}(\hat{\mathbf{V}})\right] \tag{18}$$

if the expectation is defined.

We note that just as ordinary differential entropy is invariant under translation, so is $h_{\lambda}(\hat{\mathbf{V}})$ invariant under rotation. That is, if U is a deterministic unitary matrix, then

$$h_{\lambda}(\mathsf{U}\hat{\mathbf{V}}) = h_{\lambda}(\hat{\mathbf{V}}). \tag{19}$$

Also note that $h_{\lambda}(\hat{\mathbf{V}})$ is maximized if $\hat{\mathbf{V}}$ is uniformly distributed on the unit sphere, in which case

$$h_{\lambda}(\hat{\mathbf{V}}) = \log c_m \tag{20}$$

where c_m denotes the surface area of the unit sphere in $\mathbb{C}^m$

$$c_m = \frac{2\pi^m}{\Gamma(m)}. \tag{21}$$

The definition (18) can be easily extended to conditional entropies: if $\mathbf{W}$ is some random vector, and if conditional on $\mathbf{W} = \mathbf{w}$ the random vector $\hat{\mathbf{V}}$ has density $p^{\lambda}_{\hat{\mathbf{V}}|\mathbf{W}}(\hat{\mathbf{v}}|\mathbf{w})$, then we can define

$$h_{\lambda}\big(\hat{\mathbf{V}} \,\big|\, \mathbf{W} = \mathbf{w}\big) \triangleq -\mathsf{E}\left[\log p^{\lambda}_{\hat{\mathbf{V}}|\mathbf{W}}(\hat{\mathbf{V}}|\mathbf{W}) \,\middle|\, \mathbf{W} = \mathbf{w}\right] \tag{22}$$

and we can define $h_{\lambda}\big(\hat{\mathbf{V}} \,\big|\, \mathbf{W}\big)$ as the expectation (with respect to $\mathbf{W}$) of $h_{\lambda}\big(\hat{\mathbf{V}} \,\big|\, \mathbf{W} = \mathbf{w}\big)$.

Based on these definitions we have the following lemma.

Lemma 2: Let $\mathbf{V}$ be a complex random vector taking value in $\mathbb{C}^m$ and having differential entropy $h(\mathbf{V})$. Let $\|\mathbf{V}\|$ denote its norm and $\hat{\mathbf{V}}$ denotes its direction as in (16). Then

$$h(\mathbf{V}) = h(\|\mathbf{V}\|) + h_{\lambda}\big(\hat{\mathbf{V}} \,\big|\, \|\mathbf{V}\|\big) + (2m-1)\mathsf{E}[\log\|\mathbf{V}\|] \tag{23}$$

$$= h_{\lambda}(\hat{\mathbf{V}}) + h\big(\|\mathbf{V}\| \,\big|\, \hat{\mathbf{V}}\big) + (2m-1)\mathsf{E}[\log\|\mathbf{V}\|] \tag{24}$$

whenever all the quantities in (23) and (24), respectively, are defined. Here $h(\|\mathbf{V}\|)$ is the differential entropy of $\|\mathbf{V}\|$ when viewed as a real (scalar) random variable.

Proof: This lemma follows from a change of variables. Let $\mathbf{W}$ denote the real random vector in $\mathbb{R}^{2m}$ that consists of the real and imaginary part of $\mathbf{V}$ stacked on top of each other. Then we define

$$R \triangleq \|\mathbf{W}\| \qquad \text{and} \qquad \hat{\mathbf{W}} \triangleq \frac{\mathbf{W}}{\|\mathbf{W}\|} \tag{25}$$

and note that the infinitesimal volume $\mathrm{d}\mathbf{w}$ in the $2m$-dimensional Euclidean space corresponds to $\mathrm{d}r \cdot r^{2m-1}\,\mathrm{d}\hat{\mathbf{w}}$ where $\mathrm{d}\hat{\mathbf{w}}$ denotes an infinitesimal area on the unit sphere in $\mathbb{R}^{2m}$. Hence, the joint probability densities can be written as

$$p_R\big(\|\mathbf{v}\|\big)p^{\lambda}_{\hat{\mathbf{V}}|R}\big(\hat{\mathbf{v}} \,\big|\, \|\mathbf{v}\|\big) = p^{\lambda}_{\hat{\mathbf{V}}}(\hat{\mathbf{v}})p_{R|\hat{\mathbf{V}}}\big(\|\mathbf{v}\| \,\big|\, \hat{\mathbf{v}}\big) \tag{26}$$

$$= \|\mathbf{v}\|^{2m-1}p_{\mathbf{V}}(\mathbf{v}). \tag{27}$$

The result now follows from $h(\mathbf{V}) = -\mathsf{E}[\log p_{\mathbf{V}}(\mathbf{V})]$. ∎

We shall write $\mathbf{X} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \mathsf{K})$ if $\mathbf{X} - \boldsymbol{\mu}$ is a circularly symmetric, zero-mean, complex Gaussian random vector of covariance matrix $\mathsf{E}\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^{\dagger}\right] = \mathsf{K}$. By $X \sim \mathcal{U}([a,b])$ we denote a random variable that is uniformly distributed on the interval $[a,b]$.

Throughout the paper $\{e^{\mathrm{i}\Theta_k}\}$ denotes a complex random process that is IID according to a uniform distribution over the unit circle:

$$e^{\mathrm{i}\Theta_k} \text{ IID} \sim \text{Uniform on } \{z \in \mathbb{C} : |z| = 1\}. \tag{28}$$

When it appears in formulas with other random variables or processes, $\{e^{\mathrm{i}\Theta_k}\}$ is always to be understood as being independent of these other processes.

All rates specified in this paper are in nats per channel use, *i.e.*, $\log(\cdot)$ denotes the natural logarithmic function. The abbreviation *RHS* stands for *right-hand side* and *LHS* stands for *left-hand side*.

III. The Channel Model

We consider a channel with n_{T} transmit antennas and n_{R} receive antennas whose time-k output $\mathbf{Y}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ is given by

$$\mathbf{Y}_k = \mathbb{H}_k\mathbf{x}_k + \mathbf{Z}_k. \tag{29}$$

Here $\mathbf{x}_k \in \mathbb{C}^{n_{\mathrm{T}}}$ denotes the time-k channel input vector; the random matrix $\mathbb{H}_k \in \mathbb{C}^{n_{\mathrm{R}} \times n_{\mathrm{T}}}$ denotes the time-k fading matrix; and the random vector $\mathbf{Z}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ denotes additive noise.

We assume that the fading process $\{\mathbb{H}_k\}$ and the additive noise process $\{\mathbf{Z}_k\}$ are independent and of a joint law that does not depend on the channel input $\{\mathbf{x}_k\}$.

The random vector process $\{\mathbf{Z}_k\}$ is assumed to be a spatially and temporally white, zero-mean, circularly symmetric, complex Gaussian random process, *i.e.*, $\{\mathbf{Z}_k\}$ is temporally IID $\sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \sigma^2 \mathsf{I}_{n_\mathrm{R}}\right)$ for some $\sigma^2 > 0$. Here $\mathsf{I}_{n_\mathrm{R}}$ denotes the $n_\mathrm{R} \times n_\mathrm{R}$ identity matrix.

As for the multi-variate fading process $\{\mathbb{H}_k\}$, we shall only assume that it is stationary, ergodic, of finite second moment

$$\mathsf{E}\left[\|\mathbb{H}_k\|_\mathrm{F}^2\right] < \infty \tag{30}$$

and of finite differential entropy rate

$$h(\{\mathbb{H}_k\}) > -\infty \tag{31}$$

(the *regularity* assumption). Hence the components of $\mathbb{H}_k$ are in general correlated and depend on the past. Moreover, note that we do not necessarily assume that $\{\mathbb{H}_k\}$ is Gaussian, but allow any distribution that satisfies the above assumptions, *i.e.*, that is stationary, ergodic, regular and of finite second moment. The important special case of Gaussian fading is analyzed in more detail in a separate publication [19].

We would like to briefly comment about these assumptions. The assumption of stationarity reflects our lack of knowledge about the exact dependence of the fading law on time. Obviously we can not assume stationarity for all time as the fading law will change drastically if, *e.g.*, we move from an urban to a rural area. However, in a certain setting and for a reasonable time period, stationarity seems a natural choice. Note that the block fading model [2] is not stationary.

Ergodicity reflects our assumption that we are allowing very large blocklengths so that the channel "averages out." For systems with strong delay constraints this assumption will not be justified. Finally, by asking for a fading process that is regular we ensure that the fading process is "fully random" in the (engineering) sense that even if the past is perfectly known, the present values of the fading cannot be predicted error-free.[3] This assumption will be appropriate in certain situations and will not be in others. It seems therefore clear to us that both situations, regular and nonregular fading, should be investigated. We would like to emphasize once more that at high SNR this assumption has a dramatic effect on the capacity behavior [11].

As for the input, we consider two different constraints: a peak-power constraint or an average-power constraint. We use $\mathcal{E}$ to denote the maximal allowed instantaneous power in the former case, and to denote the allowed average power in the latter case. For both cases we set

$$\mathrm{SNR} \triangleq \frac{\mathcal{E}}{\sigma^2}. \tag{32}$$

The capacity $\mathcal{C}(\mathrm{SNR})$ of the channel (29) is given by

$$\mathcal{C}(\mathrm{SNR}) = \lim_{n\uparrow\infty} \frac{1}{n} \sup I\left(\mathbf{X}_1^n; \mathbf{Y}_1^n\right) \tag{33}$$

where the supremum is over the set of all probability distributions on $\mathbf{X}_1^n$ that satisfy the constraints, *i.e.*,

$$\|\mathbf{X}_k\|^2 \le \mathcal{E}, \quad \text{almost surely}, \quad k = 1, 2, \ldots, n \tag{34}$$

for a peak-power constraint, or

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[\|\mathbf{X}_k\|^2\right] \le \mathcal{E} \tag{35}$$

for an average-power constraint.

From [7, Theorem 4.2], [9, Theorem 6.10] we have

$$\overline{\lim_{\mathrm{SNR}\uparrow\infty}} \left\{\mathcal{C}(\mathrm{SNR}) - \log(1 + \log(1 + \mathrm{SNR}))\right\} < \infty. \tag{36}$$

Note that [7, Theorem 4.2], [9, Theorem 6.10] is stated under the assumption of an average-power constraint only. However, since a peak-power constraint is more stringent than an average-power constraint, (36) also holds in the situation of a peak-power constraint.

The fading number χ is now defined as in [7, Definition 4.6], [9, Definition 6.13] by

$$\chi(\{\mathbb{H}_k\}) \triangleq \overline{\lim_{\mathrm{SNR}\uparrow\infty}} \left\{\mathcal{C}(\mathrm{SNR}) - \log(1 + \log(1 + \mathrm{SNR}))\right\}. \tag{37}$$

Prima facie the fading number depends on whether a peak-power constraint (34) or an average-power constraint (35) is imposed on the input. However, it will turn out that the MIMO fading number with memory is identical for both cases.

Finally, we remark that for an arbitrary constant nonsingular $n_\mathrm{R} \times n_\mathrm{R}$ matrix A and an arbitrary constant nonsingular $n_\mathrm{T} \times n_\mathrm{T}$ matrix B

$$\chi\left(\{\mathsf{A}\mathbb{H}_k\mathsf{B}\}\right) = \chi\left(\{\mathbb{H}_k\}\right) \tag{38}$$

see [7, Lemma 4.7], [9, Lemma 6.14].

IV. Preliminary Results

The proof of the main result relies on some observations that hold in a more general context and are therefore interesting by themselves. Some of these observations are known already and are repeated here without proof for the sake of completeness only, but some are new.

A. Capacity-Achieving Input Distributions and Stationarity

One of the main assumptions about our channel model is that the fading process and the additive noise are *stationary*. From an intuitive point of view it is clear that a stationary channel model should have a capacity-achieving input distribution that is also stationary. Unfortunately, we are not aware of a rigorous proof of this claim.

In [8, Lemma 5], [9, Lemma B.1] it is proven that—apart from edge effects—the optimum input distribution can be assumed to have equal marginals. Here we will extend this statement and prove that the capacity can be approached up to an $\epsilon > 0$ by a distribution that looks stationary apart from edge effects.

Theorem 3: Consider a channel model with input $\mathbf{x}_k \in \mathbb{C}^{n_\mathrm{T}}$ and output $\mathbf{Y}_k \in \mathbb{C}^{n_\mathrm{R}}$. Assume that the channel is both *stationary* and *unaffected by zero input vectors* $\mathbf{0}$ in the following sense: for every choice of $n \in \mathbb{N}$ and $t \in \mathbb{Z}$, for some integers $\underline{n} < -|t|$ and $\overline{n} > n + |t|$, and for every distribution $Q \in \mathcal{P}(\mathbb{C}^{n_\mathrm{T} \times n})$ we have

$$I\left(\mathbf{0}_{\underline{n}}^{0+t}, \mathbf{X}_{1+t}^{n+t}, \mathbf{0}_{n+1+t}^{\overline{n}}\,;\, \mathbf{Y}_{\underline{n}}^{\overline{n}}\right) = I\left(\mathbf{X}_1^n; \mathbf{Y}_1^n\right) \tag{39}$$

[3]Note that this is not a strictly mathematical explanation in general, but it is precise in the special case of a spatially independent Gaussian fading process.

whenever both $\mathbf{X}_{1+t}^{n+t}$ on the LHS and $\mathbf{X}_1^n$ on the RHS have the same distribution Q.

Now fix some nonnegative integer κ and some power $\mathcal{E} > 0$. Then for every $\epsilon > 0$ there corresponds some positive integer $\eta = \eta(\mathcal{E}, \epsilon)$ and some distribution $Q_{\mathcal{E},\epsilon}^{\kappa+1} \in \mathcal{P}\big(\mathbb{C}^{n_{\mathrm{T}} \times (\kappa+1)}\big)$ such that for a blocklength n sufficiently large there exists some input $\mathbf{X}_1^n$ satisfying the following:

1) The input $\mathbf{X}_1^n$ nearly achieves capacity in the sense that
$$\frac{1}{n} I\big(\mathbf{X}_1^n; \mathbf{Y}_1^n\big) \geq \mathcal{C}(\mathcal{E}) - \epsilon. \tag{40}$$
2) For every integer μ with $0 \leq \mu \leq \kappa$, every length-$(\mu+1)$ block of adjacent vectors
$$(\mathbf{X}_\ell, \ldots, \mathbf{X}_{\ell+\mu}) \tag{41}$$
taken from within the sequence
$$\mathbf{X}_\eta, \mathbf{X}_{\eta+1}, \ldots, \mathbf{X}_{n-2\eta+2} \tag{42}$$
has the same joint distribution $Q_{\mathcal{E},\epsilon}^{\mu+1}$, where this distribution $Q_{\mathcal{E},\epsilon}^{\mu+1}$ is given as the corresponding marginal distribution of $Q_{\mathcal{E},\epsilon}^{\kappa+1}$.
3) In particular, all vectors in (42) have the same marginal $Q_{\mathcal{E},\epsilon}^{1}$.
4) The marginal distribution $Q_{\mathcal{E},\epsilon}^{1}$ gives rise to a second moment $\mathcal{E}$:
$$\mathsf{E}\big[\|\mathbf{X}_k\|^2\big] = \mathcal{E}, \qquad k \in \{\eta, \ldots, n-2\eta+2\}. \tag{43}$$
5) The first $\eta - 1$ vectors and the last $2(\eta-1)$ vectors satisfy the power constraint possibly strictly:
$$\mathsf{E}\big[\|\mathbf{X}_k\|^2\big] \leq \mathcal{E}, \quad k \in \{1, \ldots, \eta-1\} \cup \{n-2\eta+3, \ldots, n\}. \tag{44}$$

Proof: The proof relies on a shift-and-mix argument based on the fact that when using deterministic zeros at the input, the corresponding output yields zero information. The details are given in Appendix A. ∎

Remark 4: Neglecting the edge-effects for the moment, Theorem 3 basically says that, for every $\mu \leq \kappa$, every block of $\mu + 1$ *adjacent* vectors has the same distribution independent of the time shift. From this it immediately follows that the distribution of every subset of (not necessarily adjacent) vectors of a $\mu+1$ block does not change when the vectors are shifted in time (simply marginalize those vectors out that are not member of the subset). Hence, Theorem 3 *almost* proves that the capacity-achieving input distribution is stationary: the only problem are the edge effects. Note that κ can be chosen freely, but has to remain fixed until n has been loosened to infinity. *I.e.*, to get rid of the edge effects one needs to firstly let n tend to infinity, before one can let κ grow.

Throughout the paper we will refer to $Q_{\mathcal{E},\epsilon}^{\kappa+1}$ and to a block of vectors $\mathbf{X}_0^\kappa \sim Q_{\mathcal{E},\epsilon}^{\kappa+1}$ as *quasi-stationary*.

B. Capacity-Achieving Input Distributions and Circular Symmetry

The next observation concerns circular symmetry. We say that a random vector $\mathbf{W}$ is *circularly symmetric* if
$$\mathbf{W} \stackrel{\mathscr{L}}{=} \mathbf{W} e^{\mathrm{i}\Theta} \tag{45}$$
where $\Theta \sim \mathcal{U}([0, 2\pi])$ is independent of $\mathbf{W}$ and where $\stackrel{\mathscr{L}}{=}$ stands for *equal in law*. Note that being circularly symmetric is not to be confused with *isotropically distributed*, which means that a vector has equal probability to point in every direction. Circular symmetry only concerns the phase of the components of a vector, not the vector's direction.

In case of a random process we make the following definition.

Definition 5: A vector random process $\{\mathbf{W}_k\}$ is said to be *circularly symmetric* if
$$\{\mathbf{W}_k\} \stackrel{\mathscr{L}}{=} \{\mathbf{W}_k e^{\mathrm{i}\Theta_k}\} \tag{46}$$
where the process $\{\Theta_k\}$ is IID $\sim \mathcal{U}([0, 2\pi])$ and independent of $\{\mathbf{W}_k\}$.

Remark 6: Note some subtleties of this definition: a random process being circularly symmetric does not only mean that for every time k the corresponding random vector $\mathbf{W}_k$ is circularly symmetric, but also that from past vectors $\mathbf{W}_{-\infty}^{k-1}$ one cannot gain any knowledge about the present phase, *i.e.*, the phase is IID. On the other hand, however, knowing the phase of one component of $\mathbf{W}_k$ in general does yield some knowledge about the phase of some other components at the *same* time k.

The following proposition says that for our channel model an optimal input can be assumed to be circularly symmetric.

Proposition 7: Assume a channel as given in (29). Then the capacity-achieving input process can be assumed to be circularly symmetric, *i.e.*, the input vectors $\{\mathbf{X}_k\}$ can be replaced by $\{\mathbf{X}_k e^{\mathrm{i}\Theta_k}\}$, where the random process $\{\Theta_k\}$ is IID $\sim \mathcal{U}([0, 2\pi])$ and independent of every other random quantity.

Proof: A proof is given in Appendix B. ∎

Remark 8: Note that the proof of Proposition 7 relies only on the fact that the additive noise is assumed to be circularly symmetric. Hence, for this proposition to hold the additive noise process does not need to be Gaussian distributed and may even have memory as long as it is circularly symmetric.

C. Capacity-Achieving Input Distributions and Escaping to Infinity

Next we give a brief review about the concept of input distributions that escape to infinity: a sequence of input distributions parametrized by the allowed cost (in our case of fading channels the cost is the available power or SNR) is said to *escape to infinity* if it assigns to every fixed compact set a probability that tends to zero as the allowed cost tends to infinity. In other words this means that in the limit—when the allowed cost tends to infinity—such a distribution does not use finite-cost symbols.

This notion is important because the asymptotic capacity of many channels of interest can only be achieved by input distributions that escape to infinity. As a matter of fact one can show that every input distribution that only achieves a mutual information of identical asymptotic *growth rate* as the capacity *must* escape to infinity. Loosely speaking, for many channels it is not favorable to use finite-cost input symbols whenever the cost constraint is loosened completely.

In the following we will only state this result specialized to the situation at hand. For a more general description and for all proofs we refer to [8, Sec. VII.C.3], [9, Sec. 2.6].

Definition 9: Let $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ be a family of input distributions for the memoryless version of the fading channel (29), *i.e.*, input distributions of the channel

$$\mathbf{Y} = \mathbb{H}\mathbf{x} + \mathbf{Z} \tag{47}$$

with input $\mathbf{x} \in \mathbb{C}^{n_{\text{T}}}$. Let this family be parametrized by the available average power $\mathcal{E}$ such that

$$\mathsf{E}_{Q_{\mathbf{X},\mathcal{E}}}\left[\|\mathbf{X}\|^2\right] \leq \mathcal{E}, \qquad \mathcal{E} \geq 0. \tag{48}$$

We say that the input distributions $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ *escape to infinity* if for every $\mathcal{E}_0 > 0$

$$\lim_{\mathcal{E}\uparrow\infty} Q_{\mathbf{X},\mathcal{E}}\left(\|\mathbf{X}\|^2 \leq \mathcal{E}_0\right) = 0. \tag{49}$$

We now have the following lemma.

Lemma 10: Let the memoryless MIMO fading channel be given as in (47) and let $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ be a family of distributions on the channel input that satisfy the power constraint (48). Let $I(Q_{\mathbf{X},\mathcal{E}})$ denote the mutual information between input and output of channel (47) when the input is distributed according to the law $Q_{\mathbf{X},\mathcal{E}}$. Assume that the family of input distributions $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ is such that the following condition is satisfied:

$$\lim_{\mathcal{E}\uparrow\infty} \frac{I(Q_{\mathbf{X},\mathcal{E}})}{\log\log\mathcal{E}} = 1. \tag{50}$$

Then $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ must escape to infinity.

Proof: A proof can be found in [8, Theorem 8, Remark 9], [9, Corollary 2.8]. ■

D. An Upper Bound on Channel Capacity

Since capacity is by definition a maximization of mutual information, it is implicitly difficult to find *upper* bounds on it. The following upper bound has been derived based on a dual expression of mutual information [7, Sec. V], [9, Sec. 2.3].

Lemma 11: Consider a memoryless channel with input $\mathbf{s} \in \mathbb{C}^{n_{\text{T}}}$ and output $T \in \mathbb{C}$. Then for an arbitrary distribution on the input $\mathbf{S}$ the mutual information between input and output of the channel is upper-bounded as follows:

$$\begin{aligned} I(\mathbf{S};T) \leq & -h(T|\mathbf{S}) + \log\pi + \alpha\log\beta + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) \\ & + (1-\alpha)\mathsf{E}\left[\log\left(|T|^2+\nu\right)\right] + \frac{1}{\beta}\mathsf{E}\left[|T|^2\right] \\ & + \frac{\nu}{\beta} \end{aligned} \tag{51}$$

where $\alpha, \beta > 0$ and $\nu \geq 0$ are parameters that can be chosen freely, but must not depend on the distribution of $\mathbf{S}$.

Proof: A proof can be found in [7, Sec. IV], [9, Sec. 2.4]. ■

E. Generalized Entropy Rates

The main result will be presented in various different forms, all containing some types of "entropy rates." The original definition of the *differential entropy rate* of a stochastic process $\{W_k\}$ is [20, Sec. 12.5]

$$h(\{W_k\}) \triangleq \lim_{n\uparrow\infty} \frac{1}{n} h(W_1, W_2, \ldots, W_n) \tag{52}$$

if the limit exists. Moreover, it can be shown that for stationary processes this limit always exists and is identical to

$$\lim_{k\uparrow\infty} h(W_k|W_{k-1},\ldots,W_1). \tag{53}$$

We will now extend this definition to conditional versions of entropy rates and prove that the limit exists as long as the involved processes are well-behaved (in particular, stationary). We will show this only for one example that, however, is representative for other forms, too.

Lemma 12: Let $\{\mathbb{H}_k\}$ be stationary, ergodic, of finite energy and regular. Let $\{\hat{\mathbf{X}}_k\}$ be a stationary unit-vector process. Then

1) the sequence $\frac{1}{n}h\left(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^n \,\middle|\, \hat{\mathbf{X}}_1^n\right)$ is nonincreasing in n;
2) the sequence $h\left(\mathbb{H}_n\hat{\mathbf{X}}_n \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n-1}, \hat{\mathbf{X}}_1^n\right)$ is nonincreasing in n;
3) for all $n \in \mathbb{N}$ we have
$$\begin{aligned} & h\left(\mathbb{H}_n\hat{\mathbf{X}}_n \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n-1}, \hat{\mathbf{X}}_1^n\right) \\ & \quad \leq \frac{1}{n}h\left(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^n \,\middle|\, \hat{\mathbf{X}}_1^n\right), \qquad \forall\, n \in \mathbb{N} \end{aligned} \tag{54}$$
and
4) the limits exist and are equal:
$$\begin{aligned} & h\left(\{\mathbb{H}_k\hat{\mathbf{X}}_k\} \,\middle|\, \{\hat{\mathbf{X}}_k\}\right) \\ & \quad \triangleq \lim_{n\uparrow\infty} \frac{1}{n}h\left(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^n \,\middle|\, \hat{\mathbf{X}}_1^n\right) && (55) \\ & \quad = \lim_{n\uparrow\infty} h\left(\mathbb{H}_n\hat{\mathbf{X}}_n \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n-1}, \hat{\mathbf{X}}_1^n\right). && (56) \end{aligned}$$

Proof: See Appendix C. ■

V. The Fading Number of MIMO Fading Channels with Memory

A. Main Result

We are now ready to state the main result.

Theorem 13: Consider a MIMO fading channel with memory (29) where the stationary and ergodic fading process $\{\mathbb{H}_k\}$ takes value in $\mathbb{C}^{n_{\text{R}}\times n_{\text{T}}}$ and satisfies $h(\{\mathbb{H}_k\}) > -\infty$ and $\mathsf{E}\left[\|\mathbb{H}_k\|_{\text{F}}^2\right] < \infty$. Then, irrespective of whether a peak-power constraint (34) or an average-power constraint (35) is imposed on the input, the fading number $\chi(\{\mathbb{H}_k\})$ is given by

$$\boxed{\begin{aligned}\chi(\{\mathbb{H}_k\}) &= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|} \right\}_{\ell=-\infty}^{-1} \right) \\ &\qquad + n_{\mathsf{R}}\mathsf{E}\left[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\right] - \log 2 \\ &\qquad - h\left(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\infty}^{0}\right) \Bigg\}.\end{aligned}} \tag{57}$$

Here the supremum is over all stochastic unit-vector processes $\{\hat{\mathbf{X}}_k\}$ that are stationary and circularly symmetric.

Moreover, the fading number is achievable by a stationary input that can be expressed as a product of two *independent* processes:

$$\mathbf{X}_k = R_k \cdot \hat{\mathbf{X}}_k \tag{58}$$

where $\{\hat{\mathbf{X}}_k\} \in \mathbb{C}^{n_{\mathsf{T}}}$ is a stationary and circularly symmetric unit-vector process with the distribution that achieves the maximum in (57), and $\{R_k\} \in \mathbb{R}_0^+$ is a scalar nonnegative IID random process such that

$$\log R_k^2 \sim \mathcal{U}([\log\log\mathcal{E}, \log\mathcal{E}]). \tag{59}$$

Note that this input satisfies the peak-power constraint (34) (and therefore also the average-power constraint (35)).

Proof: The proof is long and obscured by many technical details. We will therefore provide here an outline emphasizing the important key steps. For the details we refer to the Appendices D to I.

The proof consists of two parts: firstly we derive an upper bound on the fading number assuming an average-power constraint (35) on the channel input (see Appendix D). The key ingredients for this part are the four concepts introduced in Sections IV-A to IV-D.

Secondly we derive a lower bound on the fading number by assuming one particular input distribution on the channel that satisfies the peak-power constraint (34) (see Appendix G). We then show that the fading number that is achieved by this choice is identical to the upper bound on the fading number derived before. Since a peak-power constraint is more restrictive than the corresponding average-power constraint, the theorem follows.

a) Outline of Upper Bound: To derive the upper bound we consider the average-power constraint (35). Similarly to the proof of the SIMO fading number with memory [8, Sec. VII], [9, Sec. B.5.9] we use the chain rule to write

$$\frac{1}{n}I(\mathbf{X}_1^n;\mathbf{Y}_1^n) = \frac{1}{n}\sum_{k=1}^{n} I(\mathbf{X}_1^n;\mathbf{Y}_k \mid \mathbf{Y}_1^{k-1}) \tag{60}$$

and would like to separate each term on the RHS into terms that are memoryless and terms that take care of the memory. It is shown in (169)–(176) that

$$\begin{aligned} I(\mathbf{X}_1^n;\mathbf{Y}_k \mid \mathbf{Y}_1^{k-1}) &\le I(\mathbf{X}_k;\mathbf{Y}_k) - I(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}) + I(\mathbb{H}_k;\mathbb{H}_1^{k-1}) && (61)\\ &\le I(\mathbf{X}_k;\mathbf{Y}_k) + I(\mathbb{H}_k;\mathbb{H}_1^{k-1}) && (62)\end{aligned}$$

which would nicely do the trick. Unfortunately, (62) is not tight for two reasons. Firstly, note that in the situation of only one transmit antenna it is possible to get a good estimate of the fading realizations by simply dividing the received vector $\mathbf{Y}_k$ by the decoded value of X_k:

$$\frac{\mathbf{Y}_k}{X_k} = \frac{\mathbf{H}_k X_k + \mathbf{Z}_k}{X_k} = \mathbf{H}_k + \frac{\mathbf{Z}_k}{X_k} \to \mathbf{H}_k, \quad \text{for } |X_k| \uparrow \infty \tag{63}$$

as the SNR gets large. This is not possible anymore once we have multiple antennas at the transmitter as we cannot "divide by a vector." Instead we divide by the vector's norm:

$$\frac{\mathbf{Y}_k}{\|\mathbf{X}_k\|} = \frac{\mathbb{H}_k\mathbf{X}_k + \mathbf{Z}_k}{\|\mathbf{X}_k\|} = \mathbb{H}_k\hat{\mathbf{X}}_k + \frac{\mathbf{Z}_k}{\|\mathbf{X}_k\|} \to \mathbb{H}_k\hat{\mathbf{X}}_k, \quad \text{for } \|\mathbf{X}_k\| \uparrow \infty. \tag{64}$$

This estimation still depends on the direction of the input vector. Hence, we cannot gain the full knowledge $I(\mathbb{H}_k;\mathbb{H}_1^{k-1})$ but only $I(\mathbb{H}_k\hat{\mathbf{X}}_k;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1} \mid \hat{\mathbf{X}}_1^n)$.

The second reason why (62) is not tight is the term $I(\mathbf{Y}_k;\mathbf{Y}_1^{k-1})$ that (similar to the SIMO situation) we must not discard because it contains information about the past fading values even if we do not know the corresponding inputs. To see this note that from $\mathbf{Y}_k$ we can easily get

$$\begin{aligned}\frac{\mathbf{Y}_k}{\|\mathbf{Y}_k\|} &= \frac{\mathbb{H}_k\mathbf{X}_k + \mathbf{Z}_k}{\|\mathbb{H}_k\mathbf{X}_k + \mathbf{Z}_k\|} && (65)\\ &= \frac{\mathbb{H}_k\hat{\mathbf{X}}_k + \frac{\mathbf{Z}_k}{\|\mathbf{X}_k\|}}{\left\|\mathbb{H}_k\hat{\mathbf{X}}_k + \frac{\mathbf{Z}_k}{\|\mathbf{X}_k\|}\right\|} && (66)\\ &\to \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \qquad \text{for } \|\mathbf{X}_k\| \uparrow \infty && (67)\end{aligned}$$

which is an estimate for the "direction" of the fading. However, note that similarly to (64) and unlike to the SIMO case we cannot gain full knowledge about $\frac{\mathbb{H}_k}{\|\mathbb{H}_k\|}$ because the fading is a matrix-valued process.

So we get the following bound instead:

$$\begin{aligned} I(\mathbf{X}_1^n;\mathbf{Y}_k \mid \mathbf{Y}_1^{k-1}) &\le I(\mathbf{X}_k;\mathbf{Y}_k) - I\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{k-1}\right) \\ &\quad + I(\mathbb{H}_k\hat{\mathbf{X}}_k;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1} \mid \hat{\mathbf{X}}_1^n). \end{aligned} \tag{68}$$

Note that we have jumped over many details here, in particular, we need to rely on the observation of Lemma 10 that the capacity-achieving input distribution escapes to infinity ($\|\mathbf{X}_k\| \uparrow \infty$) in order to be able to discard the noise.

The first term on the RHS of (68) corresponds to memoryless MIMO fading. Hence, we might use the knowledge of the memoryless MIMO fading number (11) or we could use the bounding techniques known from [7, Sec. IV.A], [9, Sec. 2.4] to get an upper bound on this term. Unfortunately, both approaches fail, the former because the memoryless MIMO fading number contains a maximization that will loosen the bound when introduced at this stage. The latter approach turns out to lead to an even less tight bound.

Instead we split $I(\mathbf{X}_k; \mathbf{Y}_k)$ up into a magnitude term and a term that takes care of the direction

$$I(\mathbf{X}_k; \mathbf{Y}_k) = I(\mathbf{X}_k; \|\mathbf{Y}_k\|) + I\left(\mathbf{X}_k; \frac{\mathbf{Y}_k}{\|\mathbf{Y}_k\|} \,\middle|\, \|\mathbf{Y}_k\|\right) \tag{69}$$

and show that

$$I\left(\mathbf{X}_k; \frac{\mathbf{Y}_k}{\|\mathbf{Y}_k\|} \,\middle|\, \|\mathbf{Y}_k\|\right) \le I\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|, \hat{\mathbf{X}}_k; \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\right) \tag{70}$$

(where we again need to rely on the fact that the input distribution escapes to infinity).

The first term on the RHS of (69) almost looks like the mutual information between input and output of a memoryless MISO fading channel. We fix the problem that the output is nonnegative by multiplying $\|\mathbf{Y}_k\|$ by an independent circularly symmetric phase Θ_k. Because we assume that Θ_k is independent of all other random quantities, particularly of $\mathbf{X}_k$, this does not change the mutual information.

The bound (68) then looks as follows:

$$\begin{aligned} I&\left(\mathbf{X}_1^n; \mathbf{Y}_k \,\middle|\, \mathbf{Y}_1^{k-1}\right) \\ &\le I\left(\mathbf{X}_k; \|\mathbf{Y}_k\| e^{\mathrm{i}\Theta_k}\right) + I\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|, \hat{\mathbf{X}}_k; \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\right) \\ &\quad - I\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{k-1}\right) \\ &\quad + I\left(\mathbb{H}_k\hat{\mathbf{X}}_k; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1} \,\middle|\, \hat{\mathbf{X}}_1^n\right). \end{aligned} \tag{71}$$

This bound still depends on the unknown capacity-achieving input distribution. In order to eliminate this dependence we need to maximize it over all joint distributions on $\mathbf{X}_1, \ldots, \mathbf{X}_n$ that satisfy the average-power constraint. Unfortunately, when we only consider one fixed k, this maximization will loosen our bound. The reason lies in the third term on the RHS of (71) which can be loosely upper-bounded by zero. This loose upper bound can be achieved by the (obviously very bad) choice $\mathbf{X}_1 = \ldots = \mathbf{X}_{k-1} = \mathbf{0}$.

So it seems that we cannot consider each term of the sum in (60) separately. Fortunately, this is possible once we take Theorem 3 and Proposition 7 into account. They allow us to restrict ourselves to stationary and circularly symmetric input distributions. This excludes the mentioned bad choice and yields the following bound:

$$\begin{aligned} \frac{1}{n} I&\left(\mathbf{X}_1^n; \mathbf{Y}_1^n\right) \\ &\le \sup \Bigg\{ I\left(\mathbf{X}_0; \|\mathbf{Y}_0\| e^{\mathrm{i}\Theta_0}\right) + I\left(\|\mathbb{H}_0\hat{\mathbf{X}}_0\|, \hat{\mathbf{X}}_0; \frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\right) \\ &\quad - I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\infty}^{-1}\right) \\ &\quad + I\left(\mathbb{H}_0\hat{\mathbf{X}}_0; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1} \,\middle|\, \hat{\mathbf{X}}_{-\infty}^0\right) \Bigg\} \end{aligned} \tag{72}$$

where the supremum is over all stationary and circularly symmetric input processes.

Note that Theorem 3 has also allowed us to get rid of the dependence on k, *i.e.*, we can let n tend to infinity. Then the way is free to loosen the power constraint (*i.e.*, let $\mathcal{E} \uparrow \infty$) and to use the definition of the fading number (37).

We might now be tempted to use our knowledge about memoryless MISO fading. But again this approach fails due to the maximization in the expression of the memoryless MISO fading number (8). Instead we rely on Lemma 11 to get an upper bound on the first term on the RHS of (72). This bound will look very similar to the memoryless MISO fading number, however, does not involve a local beam-forming maximization. To make the expressions easier to read, we will use here the notation $\tilde{\chi}$ to refer to this part of the bound.

Hence, we get the following:

$$\begin{aligned} \chi(\{\mathbb{H}_k\}) \lessapprox \sup \Bigg\{ &\tilde{\chi}_{\text{MISO,IID}}\left(\|\mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0\| e^{\mathrm{i}\Theta_0}\right) \\ &+ I\left(\|\mathbb{H}_0\hat{\mathbf{X}}_0\|, \hat{\mathbf{X}}_0; \frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\right) \\ &- I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\infty}^{-1}\right) \\ &+ I\left(\mathbb{H}_0\hat{\mathbf{X}}_0; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1} \,\middle|\, \hat{\mathbf{X}}_{-\infty}^0\right) \Bigg\}. \end{aligned} \tag{73}$$

We see that the upper bound consists of a term that corresponds to the memoryless MISO fading number when the receiver only considers the magnitude of the received vector, a term that takes care of the contribution of the direction of the channel output, and two terms take care of the contribution of the memory in the channel.

Note that in the whole derivation we rely on the fact that the input distribution does not take on any value smaller than an arbitrary $\mathcal{E}_0$. However, Lemma 10 only guarantees this in the limit when the power tends to infinity. In order to solve that problem we need to introduce the event $\{\|\mathbf{X}\|^2 \ge \mathcal{E}_0\}$ and condition everything on this event.

For more details we refer to Appendix D.

b) Outline of Lower Bound: To derive a lower bound we choose a specific input distribution which naturally yields a lower bound to channel capacity and hence to the fading number. Let $\{\mathbf{X}_k\}$ be of the form

$$\mathbf{X}_k = R_k \cdot \hat{\mathbf{X}}_k. \tag{74}$$

Here $\{\hat{\mathbf{X}}_k\}$ is a sequence of random unit vectors forming a stochastic process that is stationary and circularly symmetric, but whose exact distribution will be specified later.

The stochastic process $\{R_k\}$ consists of random variables $R_k \in \mathbb{R}_0^+$ that are IID with

$$\log R_k^2 \sim \mathcal{U}\big([\log x_{\text{min}}^2, \log \mathcal{E}]\big) \tag{75}$$

where we choose x_{min}^2 as

$$x_{\text{min}}^2 \triangleq \log \mathcal{E}. \tag{76}$$

We assume that $\{R_k\} \perp\!\!\!\perp \{\hat{\mathbf{X}}_k\}$.

Note that this choice of $\{R_k\}$ satisfies the peak-power constraint (34) and therefore also the average-power constraint (35).

We then again start with the chain rule and write

$$\frac{1}{n} I\big(\mathbf{X}_1^n; \mathbf{Y}_1^n\big) = \frac{1}{n}\sum_{k=1}^{n} I\big(\mathbf{X}_k; \mathbf{Y}_1^n \,\big|\, \mathbf{X}_1^{k-1}\big) \tag{77}$$

where we would like to treat each term separately. Note that for the same reason why we were not allowed to discard the term $I\big(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}\big)$ in the derivation of the upper bound, we are not allowed to discard the future outputs $\mathbf{Y}_{k+1}^n$ on the RHS of (77).

After some algebraic changes we get the following lower bound:

$$\begin{aligned} &I\big(\mathbf{X}_k; \mathbf{Y}_1^n \,\big|\, \mathbf{X}_1^{k-1}\big) \\ &\quad\gtrapprox I\big(\hat{\mathbf{X}}_k; \hat{\mathbf{Y}}_{k+1}^n \,\big|\, \mathbf{Y}_1^{k-1}, \hat{\mathbf{X}}_1^{k-1}\big) \\ &\qquad + I\big(\mathbf{X}_k; \mathbf{Y}_k \,\big|\, \hat{\mathbf{Y}}_{k+1}^n, \mathbf{Y}_1^{k-1}, \hat{\mathbf{X}}_1^{k-1}\big). \end{aligned} \tag{78}$$

Note that the first term is bounded and that the second term corresponds to a memoryless MIMO fading channel with some side-information. To simplify notation let us denote this side-information by $\mathbf{S}_k$:

$$\mathbf{S}_k \triangleq \big(\hat{\mathbf{Y}}_{k+1}^n, \mathbf{Y}_1^{k-1}, \hat{\mathbf{X}}_1^{k-1}\big). \tag{79}$$

Contrary to the derivation of the upper bound that has been based on the memoryless MISO case, we will base the derivation of the lower bound on memoryless SIMO, *i.e.*, we split the second term on the RHS of (78) into the following two parts:

$$I\big(\mathbf{X}_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k\big) = I\big(\hat{\mathbf{X}}_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k\big) + I\big(R_k; \mathbf{Y}_k \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big). \tag{80}$$

Now we have the problem that the second mutual information term on the RHS of (80) does not correspond exactly to the SIMO situation since the input of the channel is real instead of complex. This is fixed by various arithmetic changes which at the end yield the following expression:

$$\begin{aligned} &I\big(\mathbf{X}_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k\big) \\ &\quad\approx I\big(R_k e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) \\ &\qquad + h_\lambda\big(\hat{\mathbf{Y}}_k \,\big|\, \mathbf{S}_k\big) - h_\lambda\big(\hat{\mathbf{Y}}_k e^{\mathrm{i}\Theta_k} \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) \qquad (81) \\ &\quad= I\big(\tilde{X}_k; \mathbb{H}_k \hat{\mathbf{X}}_k \tilde{X}_k + \mathbf{Z}_k \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) \\ &\qquad + h_\lambda\big(\hat{\mathbf{Y}}_k \,\big|\, \mathbf{S}_k\big) - h_\lambda\big(\hat{\mathbf{Y}}_k e^{\mathrm{i}\Theta_k} \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) \qquad (82) \end{aligned}$$

where we have introduced $\{\Theta_k\}$ to be IID uniformly distributed on $[0, 2\pi)$, independent of all other random quantities.

Note that our choice of R_k guarantees that $\tilde{X}_k \triangleq R_k e^{\mathrm{i}\Theta_k}$ achieves the fading number of memoryless SIMO fading with side-information [7, Proposition 4.30], [9, Proposition 6.30]. Hence, we get from (78) and (82)

$$\begin{aligned} \chi(\{\mathbb{H}_k\}) &\gtrapprox \chi_{\text{SIMO,IID}}\big(\mathbb{H}_k \hat{\mathbf{X}}_k \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) + h_\lambda\big(\hat{\mathbf{Y}}_k \,\big|\, \mathbf{S}_k\big) \\ &\quad - h_\lambda\big(\hat{\mathbf{Y}}_k e^{\mathrm{i}\Theta_k} \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) \\ &\quad + I\big(\hat{\mathbf{X}}_k; \hat{\mathbf{Y}}_{k+1}^n \,\big|\, \mathbf{Y}_1^{k-1}, \hat{\mathbf{X}}_1^{k-1}\big) \qquad (83) \\ &= h_\lambda\left(\frac{\mathbb{H}_k \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} e^{\mathrm{i}\Theta_k} \,\middle|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\right) \\ &\quad + n_{\mathrm{R}} \mathsf{E}\Big[\log \|\mathbb{H}_k \hat{\mathbf{X}}_k\|^2\Big] - \log 2 \\ &\quad - h\big(\mathbb{H}_k \hat{\mathbf{X}}_k \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) + h_\lambda\big(\hat{\mathbf{Y}}_k \,\big|\, \mathbf{S}_k\big) \\ &\quad - h_\lambda\big(\hat{\mathbf{Y}}_k e^{\mathrm{i}\Theta_k} \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) \\ &\quad + I\big(\hat{\mathbf{X}}_k; \hat{\mathbf{Y}}_{k+1}^n \,\big|\, \mathbf{Y}_1^{k-1}, \hat{\mathbf{X}}_1^{k-1}\big) \qquad (84) \end{aligned}$$

where we have used the expression of the fading number of memoryless SIMO fading with side-information (7).

Note that we have been cheating here since we have interchanged the order of the limits of $n \uparrow \infty$ and $\mathcal{E} \uparrow \infty$. To correct this we will need to introduce a parameter κ, get rid of n, and use the stationarity of our channel model and our choice of $\{\mathbf{X}_k\}$. Furthermore, we will have to discard the influence of the noise process in various places which is possible once we let $\mathcal{E} \uparrow \infty$ because $R_k \uparrow \infty$ with probability 1.

The result now follows by showing that (84) is equivalent to the upper bound. This will follow from some arithmetic changes, from stationarity, and from the fact that we choose the distribution of $\{\hat{\mathbf{X}}_k\}$ to achieve the supremum given in the upper bound.

For more details we refer to Appendix G. ■

B. Alternative Expressions and an Upper Bound

In the following we will state several equivalent expressions for the fading number given in (57). Depending on the context one particular form might be more convenient.

We start by defining a *constrained memoryless MIMO fading number* given a fixed distribution $Q_{\hat{\mathbf{X}}}$ on $\hat{\mathbf{X}}$:

$$\begin{aligned} \chi_{\hat{\mathbf{X}}}(\mathbb{H}) \triangleq{}& h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + n_{\mathrm{R}} \mathsf{E}\Big[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] \\ &- \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big). \end{aligned} \tag{85}$$

This corresponds to the situation where we additionally constrain the transmitter to use a fixed (possibly suboptimal) distribution on $\hat{\mathbf{X}}$, *i.e.*, the memoryless MIMO fading number is then given as (see (11))

$$\chi(\mathbb{H}) = \sup_{\substack{Q_{\hat{\mathbf{X}}} \\ \text{circ. sym.}}} \chi_{\hat{\mathbf{X}}}(\mathbb{H}). \tag{86}$$

Note the difference to the memoryless fading number with partial side-information $\hat{\mathbf{X}}$ at the receiver

$$\chi\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) = \mathsf{E}_{\hat{\mathbf{X}}}\Big[\chi\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}}\big)\Big]. \tag{87}$$

Here we assume that the realization of $\hat{\mathbf{X}}$ is known to the receiver which changes the problem from memoryless MIMO to memoryless SIMO with side-information.

From (85) we next define the following natural extension: the *constrained memoryless MIMO fading number with partial side-information* $\mathbf{S}$ given a fixed distribution $Q_{\hat{\mathbf{X}}}$ on $\hat{\mathbf{X}}$ is defined as follows:

$$\chi_{\hat{\mathbf{X}}}(\mathbb{H} \,|\, \mathbf{S}) \triangleq h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \mathbf{S}\right) + n_\mathsf{R}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\right] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}, \mathbf{S}\big). \tag{88}$$

Using these definitions we get the following alternative expressions.

Corollary 14: The MIMO fading number with memory can be written in the following five equivalent forms:

$$\begin{aligned}
\chi(\{\mathbb{H}_k\}) &= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\infty}^{-1}\right) + n_\mathsf{R}\mathsf{E}\left[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\right] - \log 2 \\
&\qquad - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\infty}^{0}\big)\Bigg\} &(89)\\
&= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ \chi_{\hat{\mathbf{X}}_0}(\mathbb{H}_0) - I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\infty}\right) + I\big(\mathbb{H}_0\hat{\mathbf{X}}_0; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1} \,\big|\, \hat{\mathbf{X}}_{-\infty}^{0}\big)\Bigg\} &(90)\\
&= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ \chi_{\hat{\mathbf{X}}_0}\left(\mathbb{H}_0 \,\middle|\, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\infty}, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\infty}^{-1}\right) \\
&\qquad + I\left(\hat{\mathbf{X}}_0; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\infty} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\infty}^{-1}\right)\Bigg\} &(91)\\
&= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ I\big(\{\hat{\mathbf{X}}_k\}; \{\mathbb{H}_k\hat{\mathbf{X}}_k\}\big) - \log 2 + \mathsf{E}\left[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|\right] - h\left(\{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|\} \,\middle|\, \left\{\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\right\}\right)\Bigg\} &(92)\\
&= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ I\left(\{\hat{\mathbf{X}}_k\}; \left\{\frac{\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k}\right\}_{r=2}^{n_\mathsf{R}}\right) - \log 2 + \mathsf{E}\left[\log\big|\mathbf{H}_0^{(1)\mathsf{T}}\hat{\mathbf{X}}_0\big|^2\right] \\
&\qquad + h\left(\{\Psi_k^{(1)}\} \,\middle|\, \{\Psi_k^{(r)} - \Psi_k^{(1)}\}_{r=2}^{n_\mathsf{R}}, \{|\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k|\}_{r=1}^{n_\mathsf{R}}\right) \\
&\qquad - h\left(\{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k\} \,\middle|\, \left\{\frac{\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k}\right\}_{r=2}^{n_\mathsf{R}}, \{\hat{\mathbf{X}}_k\}\right)\Bigg\}. &(93)
\end{aligned}$$

Here $\mathbf{H}_k^{(r)}$ denotes the r-th row of $\mathbb{H}_k$, and $\Psi_k^{(r)}$ denotes the phase of $\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k$. Moreover, in (92) we have defined

$$I\big(\{\hat{\mathbf{X}}_k\}; \{\mathbb{H}_k\hat{\mathbf{X}}_k\}\big) \triangleq \lim_{n\uparrow\infty}\frac{1}{n}I\big(\hat{\mathbf{X}}_1^n; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^n\big) \tag{94}$$

$$h\left(\{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|\} \,\middle|\, \left\{\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\right\}\right) \triangleq \lim_{n\uparrow\infty}\frac{1}{n}h\left(\{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^n\right) \tag{95}$$

and in (93)

$$I\left(\{\hat{\mathbf{X}}_k\}; \left\{\frac{\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k}\right\}_{r=2}^{n_\mathsf{R}}\right) \triangleq \lim_{n\uparrow\infty}\frac{1}{n}I\left(\hat{\mathbf{X}}_1^n; \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right) \tag{96}$$

$$h\left(\{\Psi_k^{(1)}\} \,\middle|\, \{\Psi_k^{(r)} - \Psi_k^{(1)}\}_{r=2}^{n_\mathsf{R}}, \{|\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k|\}_{r=1}^{n_\mathsf{R}}\right) \triangleq \lim_{n\uparrow\infty}\frac{1}{n}h\left(\{\Psi_\ell^{(1)}\}_{\ell=1}^n \,\middle|\, \{\Psi_\ell^{(r)} - \Psi_\ell^{(1)}\}_{\substack{r=2,\ldots,n_\mathsf{R},\\ \ell=1,\ldots,n}}, \{|\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell|\}_{\substack{r=1,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right) \tag{97}$$

$$h\left(\{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k\} \,\middle|\, \left\{\frac{\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k}\right\}_{r=2}^{n_\mathsf{R}}, \{\hat{\mathbf{X}}_k\}\right) \triangleq \lim_{n\uparrow\infty}\frac{1}{n}h\left(\{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}, \hat{\mathbf{X}}_1^n\right). \tag{98}$$

Proof: A proof can be found in Appendix J. ■

Here the expression (92) is interesting because it expresses the fading number without using the differential entropy-like quantity $h_\lambda(\cdot)$.

Note that the various forms of entropy rates used in (89)–(98) are all well-defined because the underlying processes are stationary. This has been proven in Lemma 12 for one particular case that is representative for all other cases.

Since the evaluation of (57) is in general rather difficult, we will next state two upper bounds to the MIMO fading number that are usually easier to compute.

Corollary 15: The fading number of a MIMO fading channel with memory as defined in Theorem 13 can be upper-bounded as follows:

$$\begin{aligned}\chi\big(\{\mathbb{H}_k\}\big) &\le n_{\mathrm{R}}\log\pi - \log\Gamma(n_{\mathrm{R}}) \\ &\quad + \inf_{\mathsf{A},\mathsf{B}}\sup_{\hat{\mathbf{x}}_{-\infty}^{0}}\Big\{n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}_0\|^2\big] \\ &\qquad - h\big(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}\,\big|\,\{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big)\Big\} \quad (99)\\ &\le n_{\mathrm{R}}\log\pi - \log\Gamma(n_{\mathrm{R}}) \\ &\quad + \inf_{\mathsf{A},\mathsf{B}}\sup_{\hat{\mathbf{x}}}\Big\{n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}\|^2\big] \\ &\qquad - h\big(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}\,\big|\,\mathbb{H}_{-\infty}^{-1}\big)\Big\} \quad (100)\end{aligned}$$

where the infimum is over all nonsingular $n_{\mathrm{R}}\times n_{\mathrm{R}}$ complex matrices A and nonsingular $n_{\mathrm{T}}\times n_{\mathrm{T}}$ complex matrices B.

Proof: Note that from conditioning that reduces entropy and from (20) and (21)

$$\begin{aligned} h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\,\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\infty}^{-1}\right) &\le h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\right) \le \log c_{n_{\mathrm{R}}} = \log\frac{2\pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})} \quad (101)\end{aligned}$$

where the latter upper bound is achieved with equality only if $\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}$ is uniformly distributed on the sphere, *i.e.*, isotropically distributed.

We now get from Theorem 13 and from (38):

$$\begin{aligned}\chi\big(\{\mathbb{H}_k\}\big) &= \inf_{\mathsf{A},\mathsf{B}}\chi\big(\{\mathsf{A}\mathbb{H}_k\mathsf{B}\}\big) \quad (102)\\ &\le \inf_{\mathsf{A},\mathsf{B}}\sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}}\Big\{n_{\mathrm{R}}\log\pi - \log\Gamma(n_{\mathrm{R}}) \\ &\qquad + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{X}}_0\|^2\Big] \\ &\qquad - h\big(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{X}}_0\,\big|\,\{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1},\hat{\mathbf{X}}_{-\infty}^{0}\big)\Big\} \quad (103)\\ &= \inf_{\mathsf{A},\mathsf{B}}\sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}}\mathsf{E}\Big[n_{\mathrm{R}}\log\pi - \log\Gamma(n_{\mathrm{R}}) \\ &\qquad + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{X}}_0\|^2\,\Big|\,\hat{\mathbf{X}}_0=\hat{\mathbf{x}}_0\Big] \\ &\qquad - h\big(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{X}}_0\,\big|\,\{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \\ &\qquad\qquad \{\hat{\mathbf{X}}_\ell=\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{0}\big)\Big] \quad (104)\\ &\le \inf_{\mathsf{A},\mathsf{B}}\sup_{\hat{\mathbf{x}}_{-\infty}^{0}}\Big\{n_{\mathrm{R}}\log\pi - \log\Gamma(n_{\mathrm{R}}) + n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}_0\|^2\big] \\ &\qquad - h\big(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}_0\,\big|\,\{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big)\Big\} \quad (105)\\ &\le \inf_{\mathsf{A},\mathsf{B}}\sup_{\hat{\mathbf{x}}_{-\infty}^{0}}\Big\{n_{\mathrm{R}}\log\pi - \log\Gamma(n_{\mathrm{R}}) + n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}_0\|^2\big] \\ &\qquad - h\big(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}_0\,\big|\,\{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1},\mathbb{H}_{-\infty}^{-1}\big)\Big\} \quad (106)\\ &= \inf_{\mathsf{A},\mathsf{B}}\sup_{\hat{\mathbf{x}}}\Big\{n_{\mathrm{R}}\log\pi - \log\Gamma(n_{\mathrm{R}}) + n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}\|^2\big] \\ &\qquad - h\big(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}\,\big|\,\mathbb{H}_{-\infty}^{-1}\big)\Big\}. \quad (107)\end{aligned}$$

Here, (102) follows from (38); the subsequent inequality (103) from (57) and (101); (104) holds due to linearity of expectation; then in (105) we upper-bound the expectation by the supremum over all possible values (this proves (99)); (106) is due to conditioning that reduces entropy; and the final equality (107) holds because conditional on $\mathbb{H}_{-\infty}^{-1}$, $\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}_0$ is independent of $\{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}$. ∎

VI. Some Special Cases

In this section we will now specialize the general result to some important special situations. While some of them have been known already, the case of MISO fading with memory has not been solved before.

A. Memoryless Fading

We start with the situation where the fading process has no temporal memory, *i.e.*, $\{\mathbb{H}_k\}$ is IID over time k. In this situation we will usually drop the time index and write $\mathbb{H}$.

The expression for the fading number of a memoryless MIMO fading channel (11) can be derived from (57) as follows: firstly note that only the first and the last term in (57) are influenced by memory. However, once we assume that there is no memory in the fading process $\{\mathbb{H}_k\}$, the past can only influence the present values via some memory in the input process $\{\hat{\mathbf{X}}_k\}$. Now note that the fourth term is conditioned on the input of the past *and of the present*. Hence, the past has no influence on this term either. Finally, note that the first term can be upper-bounded by dropping the conditioning, and note further that this upper-bound can actually be achieved if the input is chosen to be IID. Hence, an optimum choice of $Q_{\{\hat{\mathbf{X}}_k\}}$ will be memoryless, and (11) follows.

All other memoryless situations follow directly from this. In case of a SIMO fading channel there is only one possible choice for a circularly symmetric unit random variable $\hat{\mathbf{X}} = e^{\mathrm{i}\Theta}$, which is therefore implicitly the optimum one:

$$\chi(\mathbf{H}) = h_\lambda\big(\hat{\mathbf{H}}e^{\mathrm{i}\Theta}\big) + \mathsf{E}\big[\log\|\mathbf{H}\|^2\big] - \log 2 - h(\mathbf{H}). \quad (108)$$

For the MISO case note that independently of the distribution of $\mathbf{H}$ and $\hat{\mathbf{X}}$, the distribution of

$$\frac{\mathbf{H}^{\mathsf{T}}\hat{\mathbf{X}}}{|\mathbf{H}^{\mathsf{T}}\hat{\mathbf{X}}|}e^{\mathrm{i}\Theta}$$

is identical to the distribution of $e^{\mathrm{i}\Theta}$. Hence,

$$h_\lambda\left(\frac{\mathbf{H}^{\mathsf{T}}\hat{\mathbf{X}}}{|\mathbf{H}^{\mathsf{T}}\hat{\mathbf{X}}|}e^{\mathrm{i}\Theta}\right) = h_\lambda\big(e^{\mathrm{i}\Theta}\big) = \log 2\pi \quad (109)$$

and the fading number becomes

$$\begin{aligned}\chi(\mathbf{H}^{\mathsf{T}}) &= \sup_{Q_{\hat{\mathbf{X}}}}\Big\{\log 2\pi + \mathsf{E}\Big[\log|\mathbf{H}^{\mathsf{T}}\hat{\mathbf{X}}|^2\Big] - \log 2 \\ &\qquad - h\big(\mathbf{H}^{\mathsf{T}}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}}\big)\Big\} \quad (110)\\ &= \sup_{Q_{\hat{\mathbf{X}}}}\mathsf{E}_{\hat{\mathbf{X}}}\Big[\log\pi + \mathsf{E}\Big[\log|\mathbf{H}^{\mathsf{T}}\hat{\mathbf{X}}|^2\,\Big|\,\hat{\mathbf{X}}=\hat{\mathbf{x}}\Big]\end{aligned}$$

$$- h\big(\mathbf{H}^{\mathsf{T}}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}}\big)\bigg] \tag{111}$$

$$\le \sup_{\hat{\mathbf{x}}} \Big\{\log \pi + \mathsf{E}\big[\log |\mathbf{H}^{\mathsf{T}}\hat{\mathbf{x}}|^2\big] - h\big(\mathbf{H}^{\mathsf{T}}\hat{\mathbf{x}}\big)\Big\} \tag{112}$$

which can be achieved for a distribution of $Q_{\hat{\mathbf{X}}}$ that with probability 1 takes on the value $\hat{\mathbf{x}}$ that achieves the fading number in (112) (beam-forming).

Finally, the SISO case is a combination of the arguments of the SIMO and MISO case, *i.e.*,

$$h_\lambda\big(e^{\mathrm{i}\Theta}\big) = \log 2\pi \tag{113}$$

which yields

$$\chi(H) = \log \pi + \mathsf{E}\big[\log |H|^2\big] - h(H). \tag{114}$$

B. MISO Fading With Memory

Next we are going to study the special case of MISO fading with memory for which the fading number has been unknown so far. If we specialize Theorem 13 to the situation of only one antenna at the receiver we get the following corollary.

Corollary 16: Consider a MISO fading channel with memory where the stationary and ergodic fading process $\{\mathbf{H}_k\}$ takes value in $\mathbb{C}^{n_{\mathrm{T}}}$ and satisfies $h(\{\mathbf{H}_k\}) > -\infty$ and $\mathsf{E}\big[\|\mathbf{H}_k\|^2\big] < \infty$. Then, irrespective of whether a peak-power constraint (34) or an average-power constraint (35) is imposed on the input, the fading number $\chi\big(\{\mathbf{H}_k^{\mathsf{T}}\}\big)$ is given by

$$\chi\big(\{\mathbf{H}_k^{\mathsf{T}}\}\big) = \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}}} \bigg\{ \log \pi + \mathsf{E}\Big[\log |\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{X}}_0|^2\Big] - h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{X}}_0 \,\big|\, \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\infty}^{0}\big)\bigg\} \tag{115}$$

where $\hat{\mathbf{X}}_k \triangleq \frac{\mathbf{X}_k}{\|\mathbf{X}_k\|}$ denote vectors of unit length, and where the maximization is over all stochastic processes $\{\hat{\mathbf{X}}_k\}$ that are stationary.

Moreover, the fading number is achievable by an input that can be expressed as a product of three *independent* processes:

$$\mathbf{X}_k = R_k \cdot \hat{\mathbf{X}}_k \cdot e^{\mathrm{i}\Theta_k}. \tag{116}$$

Here $\{\hat{\mathbf{X}}_k\} \in \mathbb{C}^{n_{\mathrm{T}}}$ is a stationary unit-vector process with the distribution that maximizes (115); $\{R_k\} \in \mathbb{R}_0^+$ is a scalar nonnegative IID random process satisfying (59); and $\{\Theta_k\}$ is IID $\sim \mathcal{U}([0, 2\pi))$ as defined in Definition 5.

Proof: This follows directly from Theorem 13 by the observation that independently of the distribution of $\{\mathbf{H}_k\}$ and $\{\hat{\mathbf{X}}_k\}$, in the MISO case the distribution of

$$\left\{\frac{\mathbf{H}_k^{\mathsf{T}}\hat{\mathbf{X}}_k}{|\mathbf{H}_k^{\mathsf{T}}\hat{\mathbf{X}}_k|} e^{\mathrm{i}\Theta_k}\right\}$$

is identical to the distribution of $\{e^{\mathrm{i}\Theta_k}\}$ and therefore

$$h_\lambda\left(\frac{\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{X}}_0}{\|\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{X}}_0\|} e^{\mathrm{i}\Theta_0} \,\middle|\, \left\{\frac{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{X}}_\ell}{\|\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{X}}_\ell\|} e^{\mathrm{i}\Theta_\ell}\right\}_{\ell=-\infty}^{-1}\right) = \log 2\pi. \tag{117}$$

Note that the remaining terms do not depend on the phase of $\{\mathbf{X}_k\}$. ■

Remark 17: We would like to point out that in the case of a MISO fading process without memory the optimal input uses beam-forming with a deterministic direction that maximizes the fading number (see (8)). Once the fading process has memory this is not the case anymore. However, it is straightforward to derive the upper bound (9) and the lower bound (10) that are of beam-forming type [16]: the upper bound follows by upper-bounding the expectation over $\hat{\mathbf{X}}_{-\infty}^0$ by the supremum over $\hat{\mathbf{x}}_{-\infty}^0$. For the lower bound we choose the following stationary and circularly symmetric distribution on $\{\hat{\mathbf{X}}_k\}$:

$$\hat{\mathbf{X}}_k = \hat{\mathbf{x}} e^{\mathrm{i}\Theta_k} \tag{118}$$

where $\hat{\mathbf{x}}$ is the deterministic direction that achieves the maximum in (10).

C. SIMO and SISO Fading With Memory

In the situation with only one antenna at the transmitter the input vector $\mathbf{X}_k$ is reduced to a random variable X_k and therefore the input direction $\hat{\mathbf{X}}_k$ to a phase $e^{\mathrm{i}\Phi_k}$. Hence the expression (57) gets simplified considerably by the fact that there is only one choice of a circularly symmetric distribution of $e^{\mathrm{i}\Phi_k}$:

$$\hat{\mathbf{X}}_k = e^{\mathrm{i}\Phi_k} = e^{\mathrm{i}\Theta_k}, \quad \forall\, k \tag{119}$$

i.e., the supremum disappears.

The fading number of a SIMO fading channel with memory follows then from (57) in a straightforward way from the fact that

$$h\big(\mathbf{H}_0 e^{\mathrm{i}\Theta_0} \,\big|\, \{\mathbf{H}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\infty}^{-1}, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\infty}^{0}\big) = h\big(\mathbf{H}_0 \,\big|\, \mathbf{H}_{-\infty}^{-1}\big). \tag{120}$$

The expression (6) can be derived from the alternative form (91).

For the fading number of a SISO fading channel with memory (5) we use

$$h_\lambda\big(e^{\mathrm{i}\Theta_0} \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\infty}^{-1}\big) = \log 2\pi. \tag{121}$$

VII. CONCLUSIONS

We have derived the fading number of a general MIMO fading channel with memory where the distribution of the fading process is not restricted to be Gaussian, but may have any stationary, ergodic, and regular distribution of finite energy. In particular we allow both temporal and spatial memory. The channel is assumed to be noncoherent, *i.e.*, neither receiver nor transmitter know the realization of the fading process, but they only know its probability distribution.

We have shown that the MIMO fading number is achievable by an input that can be written as a product of two independent processes: an IID nonnegative "magnitude" process and a stationary and circularly symmetric "direction" process. The former has the standard logarithmically uniform distribution (59) that has been used in previous publications about the fading number [7]–[9]. It escapes to infinity as guaranteed by Lemma 10. The "direction" process depends on the particular law of the fading process, *i.e.*, it needs to be chosen such as to maximize the fading number.

Note that the fading number is not given in a completely closed form but as an expression that still contains a maximization. This is to be expected due to the generality of the result. Recall that the stationary and ergodic matrix-valued fading process $\{\mathbb{H}_k\}$ is only constrained[4] to be regular (see (31)). It can contain various types of temporal and spatial memory and have various different distributions. In particular, we do not restrict it to be Gaussian. We believe that it will be hardly possible to further simply (57) without making more detailed assumptions about $\{\mathbb{H}_k\}$.

Unfortunately, the evaluation of (57) is difficult even if we specify the fading process in more detail. There are several reasons for this. Firstly, the fading number is not determined by the fading process directly, but by the projection of the fading process into $\mathbb{C}^{n_\mathrm{R}}$ using an optimal choice of the input "direction" process $\{\hat{\mathbf{X}}_k\}$. Note that $\{\hat{\mathbf{X}}_k\}$ not only determines this projection, but simultaneously also conveys information in itself.[5] Hence we should choose $\{\hat{\mathbf{X}}_k\}$ such as to find a good projection and also to maximize the amount of information it can convey. The optimal choice of the input $\{\hat{\mathbf{X}}_k\}$ therefore is a trade-off between these two—in general contradicting—objectives.

Secondly, even though straightforward in theory, the evaluation of the entropy h_λ with respect to the surface of the unit sphere in $\mathbb{C}^{n_\mathrm{R}}$ might be cumbersome.

In spite of these difficulties the fading number offers very interesting insights about the general behavior of capacity in the high-SNR regime. We refer to the discussion about the fading number in Section I-B, and also, as an example, to the special situation of Gaussian fading processes with memory where (57) allows general statements about the dependence of the high-SNR capacity on the memory and the number of transmit and receive antennas. For more details about this important special case of Gaussian fading we refer to a separate publication [19].

The proof of the main result is strongly based on a new theorem showing that the capacity-achieving input distribution of a stationary channel model can (almost) be assumed to be stationary (see Theorem 3). Even though this result is very intuitive, we are not aware of any proof in the literature. We believe this preliminary result to be of importance in many other situations, too.

We also have derived the MISO fading number with memory and the already known SIMO and SISO fading numbers as special cases from this general result. In the case of MISO fading with memory it is interesting to note that in contrast to the memoryless situation the fading number is in general *not* achieved by beam-forming.

[4]Note that the restriction of having finite energy is obvious as otherwise the capacity is unbounded.

[5]This is in stark contrast to the special case of memoryless MISO fading where the $\hat{\mathbf{x}}$ is only used for beam-forming without conveying information.

Appendix A
Proof of Theorem 3

The proof follows the same lines as the proofs of [8, Lemma 5] and [9, Lemma B.1]. It is based on a shift-and-mix argument.

Fix some arbitrary $\epsilon > 0$, $\mathcal{E} > 0$, and an integer $\kappa > 0$. Recalling that

$$\mathcal{C}(\mathcal{E}) = \lim_{n\uparrow\infty} \frac{1}{n} \sup I(\mathbf{X}_1,\ldots,\mathbf{X}_n;\mathbf{Y}_1,\ldots,\mathbf{Y}_n) \tag{122}$$

where the supremum is over all joint distributions on $(\mathbf{X}_1,\ldots,\mathbf{X}_n) \in \mathbb{C}^{n_\mathrm{T}\times n}$ under which $\sum_{k=1}^n \mathsf{E}\left[\|\mathbf{X}_k\|^2\right] = n\mathcal{E}$, we conclude that there must exist some integer $\eta \geq 1$ and some joint distribution $Q^* \in \mathcal{P}(\mathbb{C}^{n_\mathrm{T}\times\eta})$ such that if $(\mathbf{X}_1,\ldots,\mathbf{X}_\eta) \sim Q^*$ then

$$\frac{1}{\eta}\sum_{k=1}^{\eta} \mathsf{E}\left[\|\mathbf{X}_k\|^2\right] = \mathcal{E} \tag{123}$$

and

$$\frac{1}{\eta} I(\mathbf{X}_1,\ldots,\mathbf{X}_\eta;\mathbf{Y}_1,\ldots,\mathbf{Y}_\eta) > \mathcal{C}(\mathcal{E}) - \frac{\epsilon}{2}. \tag{124}$$

Let $\mathbb{W}$ be a $n_\mathrm{T} \times \left(\eta\cdot\left\lceil\frac{\kappa}{\eta}+1\right\rceil\right)$ random matrix whose distribution consists of $\left\lceil\frac{\kappa}{\eta}+1\right\rceil$ independent $n_\mathrm{T}\times\eta$ blocks that are distributed according to Q^*. The distribution of $\mathbb{W}$ can then be written as the product of $\left\lceil\frac{\kappa}{\eta}+1\right\rceil$ distributions Q^*:

$$\begin{aligned} &Q_\mathbb{W}\left(\mathbf{w}_1,\ldots,\mathbf{w}_{\eta\left\lceil\frac{\kappa}{\eta}+1\right\rceil}\right) \\ &\quad= \prod_{i=0}^{\left\lceil\frac{\kappa}{\eta}+1\right\rceil-1} Q^*\left(\mathbf{w}_{i\cdot\eta+1},\ldots,\mathbf{w}_{i\cdot\eta+\eta}\right), \qquad \mathbf{w}_\ell \in \mathbb{C}^{n_\mathrm{T}}. \end{aligned} \tag{125}$$

Let us next compute the marginal distribution of $Q_\mathbb{W}$ for a certain block of length $\kappa+1$, $(\mathbf{W}_\ell,\ldots,\mathbf{W}_{\ell+\kappa})$. This marginal distribution depends on the particular choice of the starting point ℓ of the block, however, note that different choices of ℓ will result in at most η different marginal distributions. This follows from the definition of $Q_\mathbb{W}$ in (125). Let $Q_{\mathcal{E},\epsilon}^{\kappa+1}$ be the probability law on $\mathbb{C}^{n_\mathrm{T}\times(\kappa+1)}$ that is a mixture of these η different block-marginals of $Q_\mathbb{W}$, *i.e.*, for every Borel set $\mathcal{B} \subseteq \mathbb{C}^{n_\mathrm{T}\times(\kappa+1)}$

$$Q_{\mathcal{E},\epsilon}^{\kappa+1}(\mathcal{B}) \triangleq \frac{1}{\eta}\sum_{\ell=1}^{\eta} Q_\mathbb{W}\left[(\mathbf{W}_\ell,\ldots,\mathbf{W}_{\ell+\kappa}) \in \mathcal{B}\right]. \tag{126}$$

Note that in the situation when $\kappa < \eta$, $Q_{\mathcal{E},\epsilon}^{\kappa+1}$ can alternatively be written as

$$\begin{aligned} &Q_{\mathcal{E},\epsilon}^{\kappa+1}(\mathcal{B}) \\ &\quad\triangleq \frac{1}{\eta}\sum_{\ell=1}^{\eta-\kappa} Q^*\left[(\mathbf{X}_\ell,\mathbf{X}_{\ell+1},\ldots,\mathbf{X}_{\ell+\kappa}) \in \mathcal{B}\right] \\ &\qquad + \frac{1}{\eta}\sum_{\ell=\eta-\kappa+1}^{\eta} Q^*\left[(\mathbf{X}_\ell,\ldots,\mathbf{X}_\eta) \in \mathcal{B}_{1,\ldots,\eta-\ell+1}\right] \\ &\qquad\qquad \cdot Q^*\left[(\mathbf{X}_1,\ldots,\mathbf{X}_{\ell+\kappa-\eta}) \in \mathcal{B}_{\eta-\ell+2,\ldots,\kappa+1}\right] \end{aligned} \tag{127}$$

where we used $\mathcal{B}_{i,\dots,j}$ to denote the set of all corresponding $n_\mathrm{T} \times (j-i+1)$ submatrices of $\mathcal{B}$ that are created by taking only columns i to j of each matrix in $\mathcal{B}$.

Note further that from our definition it follows that $Q^{\kappa+1}_{\mathcal{E},\epsilon}$ is *quasi-stationary* in the sense that if $\mathbf{X}_0^\kappa \sim Q^{\kappa+1}_{\mathcal{E},\epsilon}$ then every length-$(\mu+1)$ subblock $\mathbf{X}_\ell^{\ell+\mu}$ has the same distribution $Q^{\mu+1}_{\mathcal{E},\epsilon}$ for all $\ell = 0,\dots,\kappa-\mu$. The distribution $Q^{\mu+1}_{\mathcal{E},\epsilon}$ can be computed from $Q^{\kappa+1}_{\mathcal{E},\epsilon}$ as marginal distribution, $0 \le \mu \le \kappa$.

In the theorem we have assumed that n is given and sufficiently large. In particular we will assume that $n \gg \max\{\eta,\kappa\}$. We shall next describe the required input distribution as follows: let

$$\nu \triangleq \left\lfloor \frac{n-\eta+1}{\eta} \right\rfloor \tag{128}$$

and let the length-$(n+\eta-1)$ sequence $\tilde{\mathbb{X}}$ of random n_T-vectors be defined by

$$\tilde{\mathbb{X}} \triangleq \Big(\underbrace{\mathbf{0},\dots,\mathbf{0}}_{\eta-1},\underbrace{\mathbf{\Xi}_1^{(1)},\dots,\mathbf{\Xi}_\eta^{(1)}}_{\eta},\dots, \dots,\underbrace{\mathbf{\Xi}_1^{(\nu)},\dots,\mathbf{\Xi}_\eta^{(\nu)}}_{\eta},\mathbf{0},\dots,\mathbf{0}\Big) \tag{129}$$

so that

$$\tilde{\mathbf{X}}_k = \begin{cases} \mathbf{0}, & \text{if } 1 \le k \le \eta-1 \\ \mathbf{\Xi}^{\lfloor k/\eta \rfloor}_{(k \bmod \eta)+1}, & \text{if } \eta \le k \le (\nu+1)\eta - 1 \\ \mathbf{0}, & \text{if } (\nu+1)\eta \le k \le n+\eta-1 \end{cases} \tag{130}$$

where $\mathbf{0}$ is the zero n_T-vector and where

$$\left\{\left(\mathbf{\Xi}_1^{(j)},\dots,\mathbf{\Xi}_\eta^{(j)}\right)\right\}_{j=1}^{\nu} \text{ are IID} \sim Q^*. \tag{131}$$

If we choose $\tilde{\mathbb{X}}$ as input for our channel, then it follows from the fact that zeros have no effect and from (124) that

$$\frac{1}{n} I\big(\tilde{\mathbf{X}}_1^n;\mathbf{Y}_1^n\big) > \frac{1}{n}\cdot \nu\eta \left(\mathcal{C}(\mathcal{E}) - \frac{\epsilon}{2}\right). \tag{132}$$

Again, since the lead-in and trailing zeros are of no consequence and since shifting does not change mutual information, this same mutual information results if we shift $\tilde{\mathbf{X}}_k$ by t (provided that $0 \le t \le \eta-1$ and n is large enough so that we do not lose any nonzero input vector):

$$\frac{1}{n} I\big(\tilde{\mathbf{X}}_{1+t}^{n+t};\mathbf{Y}_1^n\big) > \frac{1}{n}\cdot \nu\eta \left(\mathcal{C}(\mathcal{E}) - \frac{\epsilon}{2}\right). \tag{133}$$

Consequently, if we define $\mathbf{X}_1,\dots,\mathbf{X}_n$ by the mixture of the time shift of $\tilde{\mathbb{X}}$, *i.e.*,

$$\mathbf{X}_k \triangleq \tilde{\mathbf{X}}_{k+T}, \qquad 1 \le k \le n \tag{134}$$

where

$$T \sim \mathcal{U}(\{0,\dots,\eta-1\}) \tag{135}$$

is independent of $\tilde{\mathbb{X}}$, then by the concavity of mutual information in the input distribution we obtain that

$$\begin{aligned} \frac{1}{n} I\left(\mathbf{X}_1^n;\mathbf{Y}_1^n\right) &> \frac{\eta\nu}{n}\left(\mathcal{C}(\mathcal{E}) - \frac{\epsilon}{2}\right) && (136)\\ &= \frac{\eta\left\lfloor\frac{n-\eta+1}{\eta}\right\rfloor}{n}\left(\mathcal{C}(\mathcal{E}) - \frac{\epsilon}{2}\right) && (137) \end{aligned}$$

which exceeds $\mathcal{C}(\mathcal{E}) - \epsilon$ for sufficiently large n.

Except at the edges, the above mixture guarantees that every block of $\mu+1$ vectors has the same distribution

$$Q^{\mu+1}_{\mathcal{E},\epsilon}(\mathcal{B}) \triangleq \frac{1}{\eta}\sum_{\ell=1}^{\eta} Q_{\mathbb{W}}\big[(\mathbf{W}_\ell,\dots,\mathbf{W}_{\ell+\mu}) \in \mathcal{B}\big] \tag{138}$$

for every μ, $0 \le \mu \le \kappa$ and every Borel set $\mathcal{B} \subseteq \mathbb{C}^{n_\mathrm{T}\times(\mu+1)}$, *i.e.*, $\mathbf{X}_1^n$ is (apart from the edges) quasi-stationary.

Note further that by (123) we have for $\mu = 0$

$$\int_{\mathbb{C}^{n_\mathrm{T}}} \|\mathbf{x}\|^2 \,\mathrm{d}Q^1_{\mathcal{E},\epsilon}(\mathbf{x}) = \mathcal{E}. \tag{139}$$

The power in the edges can be smaller than $\mathcal{E}$ because of the mixture with deterministic zero vectors.

Appendix B
Proof of Proposition 7

Assume that $\{\Theta_k\}$ are IID $\sim \mathcal{U}([0,2\pi])$, independent of every other random quantity. Then

$$\begin{aligned} &\frac{1}{n} I\big(\mathbf{X}_1^n;\mathbf{Y}_1^n\big)\\ &= \frac{1}{n} I\big(\mathbf{X}_1^n;\mathbf{Y}_1^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) && (140)\\ &= \frac{1}{n} I\big(\{\mathbf{X}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n;\{\mathbf{Y}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) && (141)\\ &= \frac{1}{n} I\big(\{\mathbf{X}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n;\{\mathbb{H}_\ell\mathbf{X}_\ell e^{\mathrm{i}\Theta_\ell} + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) && (142)\\ &= \frac{1}{n} I\big(\tilde{\mathbf{X}}_1^n;\{\mathbb{H}_\ell\tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) && (143)\\ &= \frac{1}{n} h\big(\{\mathbb{H}_\ell\tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big)\\ &\quad - \frac{1}{n} h\big(\{\mathbb{H}_\ell\tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \tilde{\mathbf{X}}_1^n, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) && (144)\\ &= \frac{1}{n} h\big(\{\mathbb{H}_\ell\tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big)\\ &\quad - \frac{1}{n} h\big(\{\mathbb{H}_\ell\tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \tilde{\mathbf{X}}_1^n\big) && (145)\\ &\le \frac{1}{n} h\big(\{\mathbb{H}_\ell\tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n\big)\\ &\quad - \frac{1}{n} h\big(\{\mathbb{H}_\ell\tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \tilde{\mathbf{X}}_1^n\big) && (146)\\ &= \frac{1}{n} I\big(\tilde{\mathbf{X}}_1^n;\{\mathbb{H}_\ell\tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n\big). && (147) \end{aligned}$$

Here (140) follows because $\{\Theta_k\}$ is independent of every other random quantity; (142) follows because $\{\mathbf{Z}_k\}$ is circularly symmetric; in (143) we define the new input $\tilde{\mathbf{X}}_\ell \triangleq \mathbf{X}_\ell e^{\mathrm{i}\Theta_\ell}$; and (146) follows since conditioning reduces entropy.

Hence, a circularly symmetric input achieves a mutual information that is at least as big as the original mutual information.

APPENDIX C
PROOF OF LEMMA 12

We start with the proof of the second statement which follows directly from the fact that conditioning cannot increase differential entropy:

$$\begin{aligned} &h\big(\mathbb{H}_{n+1}\hat{\mathbf{X}}_{n+1} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}, \hat{\mathbf{X}}_1^{n+1}\big) \\ &\quad\le h\big(\mathbb{H}_{n+1}\hat{\mathbf{X}}_{n+1} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=2}^{n}, \hat{\mathbf{X}}_2^{n+1}\big) \quad (148)\\ &\quad= h\big(\mathbb{H}_{n}\hat{\mathbf{X}}_{n} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n-1}, \hat{\mathbf{X}}_1^{n}\big) \quad (149)\end{aligned}$$

where the last equality follows from stationarity.

We next use the second statement to prove the third:

$$\begin{aligned} &h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n}\big) \\ &\quad= \sum_{k=1}^{n} h\big(\mathbb{H}_{k}\hat{\mathbf{X}}_{k} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1}, \hat{\mathbf{X}}_1^{n}\big) \quad (150)\\ &\quad= \sum_{k=1}^{n} h\big(\mathbb{H}_{k}\hat{\mathbf{X}}_{k} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1}, \hat{\mathbf{X}}_1^{k}\big) \quad (151)\\ &\quad\ge \sum_{k=1}^{n} h\big(\mathbb{H}_{n}\hat{\mathbf{X}}_{n} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n-1}, \hat{\mathbf{X}}_1^{n}\big) \quad (152)\\ &\quad= nh\big(\mathbb{H}_{n}\hat{\mathbf{X}}_{n} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n-1}, \hat{\mathbf{X}}_1^{n}\big) \quad (153)\end{aligned}$$

where (150) follows from the chain rule; (151) from the fact that conditional on $\hat{\mathbf{X}}_1^k$, $\hat{\mathbf{X}}_{k+1}^n$ is independent of all other random variables in the expression; and (152) follows from the second statement.

Next, we prove the first statement:

$$\begin{aligned} &h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n+1} \,\big|\, \hat{\mathbf{X}}_1^{n+1}\big) \\ &\quad= h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n+1}\big) \\ &\qquad+ h\big(\mathbb{H}_{n+1}\hat{\mathbf{X}}_{n+1} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}, \hat{\mathbf{X}}_1^{n+1}\big) \quad (154)\\ &\quad\le h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n}\big) \\ &\qquad+ h\big(\mathbb{H}_{n}\hat{\mathbf{X}}_{n} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n-1}, \hat{\mathbf{X}}_1^{n}\big) \quad (155)\\ &\quad\le h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n}\big) + \frac{1}{n} h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n}\big) \quad (156)\\ &\quad= \frac{n+1}{n} h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n}\big) \quad (157)\end{aligned}$$

where (154) follows from the chain rule; where (155) follows from the second statement and from the fact that given $\hat{\mathbf{X}}_1^n$ the random vector $\hat{\mathbf{X}}_{n+1}$ is independent of $\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}$, and where (156) follows from the third statement.

Finally, to prove the fourth statement we note that for an arbitrary $m \in \mathbb{N}$

$$\begin{aligned} &h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n+m} \,\big|\, \hat{\mathbf{X}}_1^{n+m}\big) \\ &\quad= h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n+m}\big) \\ &\qquad+ h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=n+1}^{n+m} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}, \hat{\mathbf{X}}_1^{n+m}\big) \quad (158)\\ &\quad= h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n}\big) \\ &\qquad+ \sum_{k=n+1}^{n+m} h\big(\mathbb{H}_{k}\hat{\mathbf{X}}_{k} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1}, \hat{\mathbf{X}}_1^{n+m}\big) \quad (159)\\ &\quad= h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n}\big) \\ &\qquad+ \sum_{k=n+1}^{n+m} h\big(\mathbb{H}_{k}\hat{\mathbf{X}}_{k} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1}, \hat{\mathbf{X}}_1^{k}\big) \quad (160)\\ &\quad\le h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n}\big) \\ &\qquad+ \sum_{k=n+1}^{n+m} h\big(\mathbb{H}_{n+1}\hat{\mathbf{X}}_{n+1} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}, \hat{\mathbf{X}}_1^{n+1}\big) \quad (161)\\ &\quad= h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n}\big) \\ &\qquad+ mh\big(\mathbb{H}_{n+1}\hat{\mathbf{X}}_{n+1} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}, \hat{\mathbf{X}}_1^{n+1}\big). \quad (162)\end{aligned}$$

Here in (158) and (159) we use the chain rule; (160) follows from the fact that conditional on $\hat{\mathbf{X}}_1^k$, $\hat{\mathbf{X}}_{k+1}^{n+m}$ are independent of all other random variables in the expression; and (161) follows from the second statement. Hence,

$$\begin{aligned} &\underbrace{\frac{1}{n+m} h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n+m} \,\big|\, \hat{\mathbf{X}}_1^{n+m}\big)}_{\to \lim_{m\uparrow\infty} \frac{1}{m} h\left(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{m} \,\middle|\, \hat{\mathbf{X}}_1^{m}\right) \text{ as } m\uparrow\infty} \\ &\quad\le \underbrace{\frac{1}{n+m} h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\big|\, \hat{\mathbf{X}}_1^{n}\big)}_{\to 0 \text{ as } m\uparrow\infty} \\ &\qquad+ \underbrace{\frac{m}{n+m}}_{\to 1 \text{ as } m\uparrow\infty} h\big(\mathbb{H}_{n+1}\hat{\mathbf{X}}_{n+1} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}, \hat{\mathbf{X}}_1^{n+1}\big). \quad (163)\end{aligned}$$

Letting m tend to infinity we then get

$$\begin{aligned} &\lim_{m\uparrow\infty} \frac{1}{m} h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{m} \,\big|\, \hat{\mathbf{X}}_1^{m}\big) \\ &\qquad\le h\big(\mathbb{H}_{n+1}\hat{\mathbf{X}}_{n+1} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}, \hat{\mathbf{X}}_1^{n+1}\big) \quad (164)\end{aligned}$$

which combined with the third statement proves the fourth statement.

APPENDIX D
DERIVATION OF AN UPPER BOUND FOR THEOREM 13

Fix some power $\mathcal{E} > 0$, and let $\epsilon > 0$ be arbitrary. Let $\eta = \eta(\mathcal{E}, \epsilon)$ be a positive integer whose existence is guaranteed in Theorem 3. Fix a nonnegative integer κ, and let $Q_{\mathcal{E},\epsilon}^{\kappa+1} \in \mathcal{P}(\mathbb{C}^{n_\mathrm{T}\times(\kappa+1)})$ be the block-input distribution whose existence is guaranteed in Theorem 3. Let blocklength n and input $\mathbf{X}_1^n$ satisfy (40)–(44) so that

$$\begin{aligned} \mathcal{C}(\mathcal{E}) &\le \frac{1}{n} I\big(\mathbf{X}_1^n; \mathbf{Y}_1^n\big) + \epsilon \quad (165)\\ &= \frac{1}{n} \sum_{k=1}^{n} I\big(\mathbf{X}_1^n; \mathbf{Y}_k \,\big|\, \mathbf{Y}_1^{k-1}\big) + \epsilon. \quad (166)\end{aligned}$$

For $1 \le k \le \eta + \kappa - 1$ and for $n - 2\eta + 3 \le k \le n$ we use the crude bound

$$\begin{aligned} I\big(\mathbf{X}_1^n; \mathbf{Y}_k \,\big|\, \mathbf{Y}_1^{k-1}\big) &\le I\big(\mathbf{X}_k; \mathbf{Y}_k\big) + I\big(\mathbb{H}_0; \mathbb{H}_{-\infty}^{-1}\big) \quad (167)\\ &\le \mathcal{C}_{\mathrm{IID}}(\mathcal{E}) + I\big(\mathbb{H}_0; \mathbb{H}_{-\infty}^{-1}\big) \quad (168)\end{aligned}$$

where $\mathcal{C}_{\mathrm{IID}}(\mathcal{E})$ denotes the capacity of the memoryless MIMO fading channel as given in (47) with an available average power of at most $\mathcal{E}$ as guaranteed in (43) and (44). The first inequality can be derived as follows:

$$\begin{aligned}
I\big(\mathbf{X}_1^n;\mathbf{Y}_k\,\big|\,\mathbf{Y}_1^{k-1}\big)
&= I\big(\mathbf{X}_1^n,\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big) - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (169)\\
&= I\big(\mathbf{X}_1^{k-1},\mathbf{Y}_1^{k-1},\mathbf{X}_k;\mathbf{Y}_k\big) - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (170)\\
&\le I\big(\mathbf{X}_1^{k-1},\mathbf{Y}_1^{k-1},\mathbb{H}_1^{k-1},\mathbf{X}_k;\mathbf{Y}_k\big) - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (171)\\
&= I\big(\mathbb{H}_1^{k-1},\mathbf{X}_k;\mathbf{Y}_k\big) - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (172)\\
&= I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\mathbb{H}_1^{k-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_k\big) - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (173)\\
&= I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\mathbb{H}_1^{k-1};\mathbf{X}_k,\mathbf{Y}_k\big) - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (174)\\
&\le I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\mathbb{H}_1^{k-1};\mathbb{H}_k,\mathbf{X}_k,\mathbf{Y}_k\big)\\
&\quad - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (175)\\
&= I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\mathbb{H}_1^{k-1};\mathbb{H}_k\big) - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (176)\\
&\le I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\mathbb{H}_0;\mathbb{H}_{-\infty}^{-1}\big) - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (177)\\
&\le I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\mathbb{H}_0;\mathbb{H}_{-\infty}^{-1}\big). && (178)
\end{aligned}$$

Here (169) follows from the chain rule; (170) follows because we prohibit feedback; (171) from the inclusion of the additional random vectors $\mathbb{H}_1^{k-1}$ in the mutual information term; (172) follows because, conditional on the past fading and the present input, the past inputs and outputs are independent of the present output; (173) follows from the chain rule; in (174) we use the chain rule and note that $I\big(\mathbb{H}_1^{k-1};\mathbf{X}_k\big)=0$; the following two steps (175)–(176) are analogous to (171)–(172); (177) follows once more from the inclusion of additional random vectors in the mutual information and from stationarity; and the final inequality (178) from the nonnegativity of mutual information.

Note that (168) is uniformly bounded in n. So we conclude that

$$\begin{aligned}
\mathcal{C}(\mathcal{E}) &\le \varliminf_{n\uparrow\infty}\frac{1}{n}I\big(\mathbf{X}_1^n;\mathbf{Y}_1^n\big)+\epsilon && (179)\\
&= \varliminf_{n\uparrow\infty}\frac{1}{n-\kappa-3(\eta-1)}\sum_{k=\eta+\kappa}^{n-2(\eta-1)} I\big(\mathbf{X}_1^n;\mathbf{Y}_k\,\big|\,\mathbf{Y}_1^{k-1}\big)\\
&\quad+\epsilon. && (180)
\end{aligned}$$

This allows us to focus on $\eta+\kappa\le k\le n-2(\eta-1)$ for which Theorem 3 guarantees that every $(\kappa+1)$-block $(\mathbf{X}_{k-\kappa},\ldots,\mathbf{X}_k)$ has the same distribution $Q_{\mathcal{E},\epsilon}^{\kappa+1}$.

We now continue by further upper-bounding $I\big(\mathbf{X}_1^n;\mathbf{Y}_k\,\big|\,\mathbf{Y}_1^{k-1}\big)$ for such k:

$$\begin{aligned}
I\big(\mathbf{X}_1^n;\mathbf{Y}_k\,\big|\,\mathbf{Y}_1^{k-1}\big)
&= I\big(\mathbf{X}_1^n,\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big) - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (181)\\
&= I\big(\mathbf{X}_1^{k-1},\mathbf{Y}_1^{k-1},\mathbf{X}_k;\mathbf{Y}_k\big) - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (182)\\
&\le I\big(\mathbf{X}_1^{k-1},\mathbf{Y}_1^{k-1},\mathbb{H}_1^{k-\kappa-1},\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1},\mathbf{X}_k;\mathbf{Y}_k\big)\\
&\quad - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (183)\\
&= I\big(\mathbf{X}_{k-\kappa}^{k-1},\mathbb{H}_1^{k-\kappa-1},\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1},\mathbf{X}_k;\mathbf{Y}_k\big)\\
&\quad - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (184)\\
&= I(\mathbf{X}_k;\mathbf{Y}_k) + \underbrace{I\big(\mathbf{X}_{k-\kappa}^{k-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_k\big)}_{=0}\\
&\quad + I\big(\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_{k-\kappa}^{k}\big)\\
&\quad + I\big(\mathbb{H}_1^{k-\kappa-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_{k-\kappa}^{k},\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big)\\
&\quad - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (185)\\
&= I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_{k-\kappa}^{k}\big)\\
&\quad + \underbrace{I\big(\mathbb{H}_1^{k-\kappa-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_{k-\kappa}^{k},\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big)}_{\le\delta_1(\kappa)}\\
&\quad - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big) && (186)\\
&\le I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_{k-\kappa}^{k}\big)\\
&\quad + \delta_1(\kappa) - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big). && (187)
\end{aligned}$$

Here the first two steps (181)–(182) are identical to (169)–(170); (183) follows from the inclusion of the additional random matrices $\mathbb{H}_1^{k-\kappa-1}$ and the random vectors $\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}$ in the mutual information term; (184) follows because, conditional on $(\mathbb{H}_1^{k-\kappa-1},\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1},\mathbf{X}_{k-\kappa}^{k})$, the past outputs $\mathbf{Y}_1^{k-1}$ and the past inputs $\mathbf{X}_1^{k-\kappa-1}$ are independent of the present output $\mathbf{Y}_k$; then (185) follows from the chain rule; (186) from the independence of the past inputs and the present output when conditioning on the present input; and in (187) we use the following lemma.

Lemma 18: Let $\mathbf{X}_{k-\kappa}^{k}$ be quasi-stationary as defined in Theorem 3 and Remark 4. Then

$$I\big(\mathbb{H}_1^{k-\kappa-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_{k-\kappa}^{k},\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big)\le\delta_1(\kappa) \qquad (188)$$

where $\delta_1(\kappa)$ does neither depend on k nor on the input $\{\mathbf{X}_k\}$ and tends to zero as κ tends to infinity.

Proof: See Appendix E. ■

We continue as follows:

$$\begin{aligned}
&I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_{k-\kappa}^{k}\big)\\
&- I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big)+\delta_1(\kappa)\\
&\quad\le I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1};\mathbf{Y}_k,\mathbb{H}_k\mathbf{X}_k\,\big|\,\mathbf{X}_{k-\kappa}^{k}\big)\\
&\qquad - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big)+\delta_1(\kappa) && (189)\\
&\quad= I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1};\mathbb{H}_k\mathbf{X}_k\,\big|\,\mathbf{X}_{k-\kappa}^{k}\big)\\
&\qquad + \underbrace{I\big(\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_{k-\kappa}^{k},\mathbb{H}_k\mathbf{X}_k\big)}_{=0}\\
&\qquad - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big)+\delta_1(\kappa) && (190)\\
&\quad= I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1};\mathbb{H}_k\mathbf{X}_k\,\big|\,\mathbf{X}_{k-\kappa}^{k}\big)\\
&\qquad - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big)+\delta_1(\kappa) && (191)\\
&\quad= I(\mathbf{X}_k;\mathbf{Y}_k)\\
&\qquad + I\big(\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1};\mathbb{H}_k\mathbf{X}_k\,\big|\,\{\|\mathbf{X}_\ell\|\}_{\ell=k-\kappa}^{k},\hat{\mathbf{X}}_{k-\kappa}^{k}\big)\\
&\qquad - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big)+\delta_1(\kappa) && (192)\\
&\quad= I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1};\mathbb{H}_k\hat{\mathbf{X}}_k\,\big|\,\hat{\mathbf{X}}_{k-\kappa}^{k}\big)\\
&\qquad - I\big(\mathbf{Y}_k;\mathbf{Y}_1^{k-1}\big)+\delta_1(\kappa) && (193)\\
&\quad\le I(\mathbf{X}_k;\mathbf{Y}_k) + I\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1};\mathbb{H}_k\hat{\mathbf{X}}_k\,\big|\,\hat{\mathbf{X}}_{k-\kappa}^{k}\big)\\
&\qquad - I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big)+\delta_1(\kappa). && (194)
\end{aligned}$$

Here (189) follows from the inclusion of the random vector $\mathbb{H}_k\mathbf{X}_k$ in the mutual information term; (190) from chain rule; (191) follows because the additive noise $\mathbf{Z}_k$ is independent of the fading $\mathbb{H}_{k-\kappa}^{k-1}$; in (192) we introduce $\hat{\mathbf{X}}_\ell\triangleq\mathbf{X}_\ell/\|\mathbf{X}_\ell\|$; (193) follows from dividing each term by the magnitude of the input vectors and from the fact that given $\hat{\mathbf{X}}_{k-\kappa}^{k}$, $\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k}$ is independent of $\{\|\mathbf{X}_\ell\|\}_{\ell=k-\kappa}^{k}$; and in the final inequality (194) we drop some arguments in the negative mutual information term.

Note that (194) only depends on $\mathbf{X}_{k-\kappa}^{k}$ which, according to Theorem 3, has a distribution $Q_{\mathcal{E},\epsilon}^{\kappa+1}$. Hence, using stationarity and combining (194) with (180) we find

$$\begin{aligned}
&\mathcal{C}(\mathcal{E}) \\
&\quad\le \varlimsup_{n\uparrow\infty} \frac{1}{n-\kappa-3(\eta-1)} \\
&\qquad\cdot \sum_{k=\eta+\kappa}^{n-2(\eta-1)} \Big(I(\mathbf{X}_k;\mathbf{Y}_k) - I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big) \\
&\qquad\qquad + I\big(\mathbb{H}_k\hat{\mathbf{X}}_k;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}\,\big|\,\hat{\mathbf{X}}_{k-\kappa}^{k}\big)\Big) \\
&\qquad + \delta_1(\kappa)+\epsilon && (195)\\
&\quad= \varlimsup_{n\uparrow\infty} \frac{1}{n-\kappa-3(\eta-1)} \\
&\qquad\cdot \sum_{k=\eta+\kappa}^{n-2(\eta-1)} \Big(I(\mathbf{X}_k;\mathbb{H}_k\mathbf{X}_k+\mathbf{Z}_k) \\
&\qquad\qquad - I\big(\mathbb{H}_k\mathbf{X}_k+\mathbf{Z}_k;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=k-\kappa}^{k-1}\big) \\
&\qquad\qquad + I\big(\mathbb{H}_k\hat{\mathbf{X}}_k;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}\,\big|\,\hat{\mathbf{X}}_{k-\kappa}^{k}\big)\Big) \\
&\qquad + \delta_1(\kappa)+\epsilon && (196)\\
&\quad= \varlimsup_{n\uparrow\infty} \frac{1}{n-\kappa-3(\eta-1)} \\
&\qquad\cdot \sum_{k=\eta+\kappa}^{n-2(\eta-1)} \Big(I(\mathbf{X}_{\eta+\kappa};\mathbb{H}_{\eta+\kappa}\mathbf{X}_{\eta+\kappa}+\mathbf{Z}_{\eta+\kappa}) \\
&\qquad\qquad - I\big(\mathbb{H}_{\eta+\kappa}\mathbf{X}_{\eta+\kappa}+\mathbf{Z}_{\eta+\kappa};\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=\eta}^{\eta+\kappa-1}\big) \\
&\qquad\qquad + I\big(\mathbb{H}_{\eta+\kappa}\hat{\mathbf{X}}_{\eta+\kappa};\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=\eta}^{\eta+\kappa-1}\,\big|\,\hat{\mathbf{X}}_{\eta}^{\eta+\kappa}\big)\Big) \\
&\qquad + \delta_1(\kappa)+\epsilon && (197)\\
&\quad= I(\mathbf{X}_{\eta+\kappa};\mathbb{H}_{\eta+\kappa}\mathbf{X}_{\eta+\kappa}+\mathbf{Z}_{\eta+\kappa}) \\
&\qquad - I\big(\mathbb{H}_{\eta+\kappa}\mathbf{X}_{\eta+\kappa}+\mathbf{Z}_{\eta+\kappa};\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=\eta}^{\eta+\kappa-1}\big) \\
&\qquad + I\big(\mathbb{H}_{\eta+\kappa}\hat{\mathbf{X}}_{\eta+\kappa};\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=\eta}^{\eta+\kappa-1}\,\big|\,\hat{\mathbf{X}}_{\eta}^{\eta+\kappa}\big)\Big) \\
&\qquad + \delta_1(\kappa)+\epsilon. && (198)
\end{aligned}$$

Here in (197) we shift all k to $\eta+\kappa$ which is possible due to the stationarity of $\{\mathbb{H}_k,\mathbf{Z}_k\}$ and due to the fact that for all $k\in\{\eta+\kappa,\ldots,n-2(\eta-1)\}$ the distribution of $\mathbf{X}_{k-\kappa}^{k}$ is $Q_{\mathcal{E},\epsilon}^{\kappa+1}$ given in Theorem 3.

To ease readability and to prevent the need of carrying η throughout the rest of the proof, we will introduce here a slight misuse in notation: for pure notational convenience we will assume from now on that $\mathbf{X}_{-\kappa}^{0}\sim Q_{\mathcal{E},\epsilon}^{\kappa+1}$, *i.e.*, that from now on $\mathbf{X}_{-\kappa}^{0}$ is quasi-stationary. Note that this is a notational choice and therefore does not contradict the edge-effects of Theorem 3! Hence, we rewrite (198) as follows:

$$\begin{aligned}
\mathcal{C}(\mathcal{E}) \le{}& I(\mathbf{X}_0;\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0) \\
&- I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\big) \\
&+ I\big(\mathbb{H}_0\hat{\mathbf{X}}_0;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,\hat{\mathbf{X}}_{-\kappa}^{0}\big)+\delta_1(\kappa)+\epsilon. \qquad (199)
\end{aligned}$$

Before we continue bounding (199) note that Lemma 10 guarantees that $Q_{\mathcal{E},\epsilon}^{1}$ must *escape to infinity* as defined in Definition 9, *i.e.*, for an arbitrary $\mathcal{E}_0\ge 0$

$$\lim_{\mathcal{E}\uparrow\infty}\Pr\big[\|\mathbf{X}_k\|^2\le\mathcal{E}_0\big]=0,\quad \forall\,k\in\{-\kappa,\ldots,0\}. \qquad (200)$$

So fix $\mathcal{E}_0$ and define an indicator random variable E as follows:

$$E\triangleq\begin{cases}1, & \text{if } \|\mathbf{X}_k\|^2\ge\mathcal{E}_0, \quad \forall\,k\in\{-\kappa,\ldots,0\}\\ 0, & \text{otherwise.}\end{cases} \qquad (201)$$

Let

$$p\triangleq\Pr[E=1] \qquad (202)$$

and note that by the union bound

$$\begin{aligned}
1-p &= \Pr[E=0] && (203)\\
&\le \sum_{k=-\kappa}^{0}\Pr\big[\|\mathbf{X}_k\|^2<\mathcal{E}_0\big] && (204)\\
&= (\kappa+1)Q_{\mathcal{E},\epsilon}^{1}\big(\|\mathbf{X}_0\|^2<\mathcal{E}_0\big) && (205)
\end{aligned}$$

i.e.,

$$p\ge 1-(\kappa+1)Q_{\mathcal{E},\epsilon}^{1}\big(\|\mathbf{X}_0\|^2<\mathcal{E}_0\big). \qquad (206)$$

Hence from Lemma 10 then follows that

$$\lim_{\mathcal{E}\uparrow\infty}p=1. \qquad (207)$$

Finally, in order to simplify notation, we will write R_k for $\|\mathbf{X}_k\|$, *i.e.*, we have $\mathbf{X}_k=R_k\cdot\hat{\mathbf{X}}_k$.

We start with the second term on the RHS of (199) and derive the following lower bound:

$$\begin{aligned}
&I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\big) \\
&\quad= I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1},E\big) \\
&\qquad - I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;E\,\big|\,\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\big) && (208)\\
&\quad= I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;E\big) \\
&\qquad + I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E\big) \\
&\qquad - I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;E\,\big|\,\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\big) && (209)\\
&\quad\ge I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E\big) \\
&\qquad - I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;E\,\big|\,\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\big) && (210)\\
&\quad= I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E\big) \\
&\qquad - H\big(E\,\big|\,\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\big) \\
&\qquad + H\big(E\,\big|\,\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{0}\big) && (211)\\
&\quad\ge I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E\big) \\
&\qquad - H\big(E\,\big|\,\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\big) && (212)\\
&\quad\ge I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E\big)-H(E) && (213)\\
&\quad= I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E\big)-H_{\mathrm{b}}(p) && (214)\\
&\quad= pI\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E=1\big) \\
&\qquad + (1-p)I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E=0\big) \\
&\qquad - H_{\mathrm{b}}(p) && (215)\\
&\quad\ge pI\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E=1\big) \\
&\qquad - H_{\mathrm{b}}(p) && (216)
\end{aligned}$$

where

$$H_{\mathrm{b}}(\xi)\triangleq-\xi\log\xi-(1-\xi)\log(1-\xi) \qquad (217)$$

denotes the binary entropy function. Here the first two equalities (208) and (209) follow from the chain rule; in (210) we drop a nonnegative mutual information term; (211) and (212) follow from the definition of mutual information and the fact that entropy is nonnegative (note that since E is a binary random variable we deal with discrete entropies and not differential entropies at this point); in (213) we then drop some conditioning which increases entropy; and in (216) we again drop some nonnegative mutual information terms.

We now proceed as follows:

$$\begin{aligned} &pI\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E=1\big)-H_{\mathrm{b}}(p)\\ &\quad= pI\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1},\mathbf{Z}_{-\kappa}^{-1}\,\big|\,E=1\big)\\ &\qquad - p\,\underbrace{I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\mathbf{Z}_{-\kappa}^{-1}\,\big|\,\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1},E=1\big)}_{\le\delta_2(\mathcal{E}_0,\kappa)}\\ &\qquad - H_{\mathrm{b}}(p) &(218)\\ &\quad\ge pI\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1},\mathbf{Z}_{-\kappa}^{-1}\,\big|\,E=1\big)\\ &\qquad - p\delta_2(\mathcal{E}_0,\kappa)-H_{\mathrm{b}}(p) &(219)\\ &\quad= pI\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E=1\big)\\ &\qquad - p\delta_2(\mathcal{E}_0,\kappa)-H_{\mathrm{b}}(p) &(220)\\ &\quad= pI\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0,\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E=1\big)\\ &\qquad - p\,\underbrace{I\big(\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0,E=1\big)}_{\le\delta_3(\mathcal{E}_0,\kappa)}\\ &\qquad - p\delta_2(\mathcal{E}_0,\kappa)-H_{\mathrm{b}}(p) &(221)\\ &\quad\ge pI\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0,\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E=1\big)\\ &\qquad - p\delta_2(\mathcal{E}_0,\kappa)-p\delta_3(\mathcal{E}_0,\kappa)-H_{\mathrm{b}}(p) &(222)\\ &\quad= pI\big(\mathbb{H}_0\mathbf{X}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E=1\big)\\ &\qquad - \underbrace{p\delta_2(\mathcal{E}_0,\kappa)}_{\le\delta_2(\mathcal{E}_0,\kappa)}-\underbrace{p\delta_3(\mathcal{E}_0,\kappa)}_{\le\delta_3(\mathcal{E}_0,\kappa)}-H_{\mathrm{b}}(p) &(223)\\ &\quad\ge pI\big(\mathbb{H}_0\mathbf{X}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E=1\big)\\ &\qquad - \delta_2(\mathcal{E}_0,\kappa)-\delta_3(\mathcal{E}_0,\kappa)-H_{\mathrm{b}}(p) &(224)\end{aligned}$$

where we have used the following lemma.

Lemma 19: Let $\{E=1\}$ be the event that $\|\mathbf{X}_k\|^2\ge\mathcal{E}_0$ for all $k=-\kappa,\ldots,0$. Then

$$I\big(\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0;\mathbf{Z}_{-\kappa}^{-1}\,\big|\,\{\mathbb{H}_\ell\mathbf{X}_\ell+\mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1},E=1\big)\le\delta_2(\mathcal{E}_0,\kappa) \tag{225}$$

$$I\big(\mathbf{Z}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,\mathbb{H}_0\mathbf{X}_0+\mathbf{Z}_0,E=1\big)\le\delta_3(\mathcal{E}_0,\kappa) \tag{226}$$

where $\delta_2(\mathcal{E}_0,\kappa)$ and $\delta_3(\mathcal{E}_0,\kappa)$ are independent of $\{\mathbf{X}_k\}$ and depend on $\mathcal{E}_0$ such that

$$\lim_{\mathcal{E}_0\uparrow\infty}\delta_2(\mathcal{E}_0,\kappa)=0 \tag{227}$$

$$\lim_{\mathcal{E}_0\uparrow\infty}\delta_3(\mathcal{E}_0,\kappa)=0. \tag{228}$$

Proof: See Appendix F. ■

We continue by splitting $\mathbb{H}_k\mathbf{X}_k$ into magnitude and direction:

$$\begin{aligned}\mathbb{H}_k\mathbf{X}_k &= \frac{\mathbb{H}_k\mathbf{X}_k}{\|\mathbb{H}_k\mathbf{X}_k\|}\cdot\|\mathbb{H}_k\mathbf{X}_k\| &(229)\\ &= \frac{\mathbb{H}_k\hat{\mathbf{X}}_k\cdot R_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\cdot R_k\|}\cdot\|\mathbb{H}_k\hat{\mathbf{X}}_k\cdot R_k\| &(230)\\ &= \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\cdot\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k. &(231)\end{aligned}$$

Hence, we get

$$\begin{aligned} &pI\big(\mathbb{H}_0\mathbf{X}_0;\{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,E=1\big)\\ &\quad= pI\left(\|\mathbb{H}_0\hat{\mathbf{X}}_0\|R_0,\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\right.\\ &\qquad\qquad\left.\big\{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|R_\ell\big\}_{\ell=-\kappa}^{-1},\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\,\middle|\,E=1\right) &(232)\\ &\quad\ge pI\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\,\middle|\,E=1\right) &(233)\\ &\quad= I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\,\middle|\,E\right)\\ &\qquad-(1-p)I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\,\middle|\,E=0\right) &(234)\end{aligned}$$

where the inequality (233) follows from dropping some arguments of mutual information. Note that even though $\|\hat{\mathbf{X}}_k\|=1$ by definition, it still might depend on E because we are not allowed to assume that $\hat{\mathbf{X}}_k$ and $R_k\triangleq\|\mathbf{X}_k\|$ are independent. At this stage of the proof we only know that $\mathbf{X}_{-\kappa}^0\sim Q_{\mathcal{E},\epsilon}^{\kappa+1}$ where $Q_{\mathcal{E},\epsilon}^{\kappa+1}$ is defined in Theorem 3 and *does* depend on $\mathcal{E}$.

We lower-bound the first term in (234) by

$$\begin{aligned} &I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\,\middle|\,E\right)\\ &\quad= I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1},E\right)\\ &\qquad - I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};E\right) &(235)\\ &\quad= I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1},E\right)-H(E)\\ &\qquad + \underbrace{H\left(E\,\middle|\,\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\right)}_{\ge 0} &(236)\\ &\quad\ge I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\right)-H_{\mathrm{b}}(p) &(237)\end{aligned}$$

and upper-bound the second term in (234) by

$$\begin{aligned} &I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\,\middle|\,E=0\right)\\ &\quad\le I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|},\mathbb{H}_0\hat{\mathbf{X}}_0;\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1},\right.\end{aligned}$$

$$\left.\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1} \,\middle|\, E=0\right) \quad (238)$$

$$= I\big(\mathbb{H}_0\hat{\mathbf{X}}_0; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1} \,\big|\, E=0\big) \quad (239)$$

$$= h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, E=0\big) - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, E=0\big) \quad (240)$$

$$\leq h\big(\mathbb{H}_0\hat{\mathbf{X}}_0\big) - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_0, \mathbb{H}_{-\kappa}^{-1}, E=0\big) \quad (241)$$

$$= h\big(\mathbb{H}_0\hat{\mathbf{X}}_0\big) - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \hat{\mathbf{X}}_0, \mathbb{H}_{-\kappa}^{-1}, E=0\big) \quad (242)$$

$$\leq n_{\mathrm{R}} \log\left(\pi e \frac{\mathsf{E}\left[\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\right]}{n_{\mathrm{R}}}\right) - \inf_{\hat{\mathbf{x}}} h\big(\mathbb{H}_0\hat{\mathbf{x}} \,\big|\, \mathbb{H}_{-\kappa}^{-1}\big) \quad (243)$$

$$\leq n_{\mathrm{R}} \log\left(\pi e \frac{\mathsf{E}\left[\|\mathbb{H}_0\|^2 \cdot \|\hat{\mathbf{X}}_0\|^2\right]}{n_{\mathrm{R}}}\right) - \inf_{\hat{\mathbf{x}}} h\big(\mathbb{H}_0\hat{\mathbf{x}} \,\big|\, \mathbb{H}_{-\kappa}^{-1}\big) \quad (244)$$

$$= n_{\mathrm{R}} \log\left(\pi e \frac{\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]}{n_{\mathrm{R}}}\right) - \inf_{\hat{\mathbf{x}}} h\big(\mathbb{H}_0\hat{\mathbf{x}} \,\big|\, \mathbb{H}_{-\kappa}^{-1}\big) \quad (245)$$

$$\triangleq \xi_1. \quad (246)$$

Here (238) follows from adding additional arguments to mutual information; (239) holds because conditional on $\mathbb{H}_\ell\hat{\mathbf{X}}_\ell$, $\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}$ is deterministic; in (240) we use the definition of mutual information; (241) follows from the fact that conditioning reduces entropy; in (242) we drop $\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}$ because given $\hat{\mathbf{X}}_0$ and $\mathbb{H}_{-\kappa}^{-1}$ it is independent of $\mathbb{H}_0\hat{\mathbf{X}}_0$. In (243) we rely on the fact that of all random vectors of given marginals differential entropy is maximized by the one whose components are independent [20], on the fact that of all complex random variables of a given second moment differential entropy is maximized by the circularly symmetric Gaussian distribution, and by the concavity (in the variance) of the differential entropy of a circularly symmetric complex Gaussian. The inequality (244) follows from Cauchy-Schwarz; and (245) from the fact that $\hat{\mathbf{X}}_0$ is a unit vector. Note that

$$0 \leq \xi_1 < \infty \quad (247)$$

because our assumption $\mathsf{E}\left[\|\mathbb{H}_0\|_{\mathrm{F}}^2\right] < \infty$ implies by (15) that $\mathsf{E}\left[\|\mathbb{H}_0\|^2\right] < \infty$ and because our assumption $h(\{\mathbb{H}_k\}) > -\infty$ implies by a conditional version of [7, Lemma 6.6], [9, Lemma A.14] that

$$\inf_{\hat{\mathbf{x}}} h\big(\mathbb{H}_0\hat{\mathbf{x}} \,\big|\, \mathbb{H}_{-\kappa}^{-1}\big) > -\infty. \quad (248)$$

Plugging (246) and (237) into (234), (224), and then into (199) we get

$$\mathcal{C}(\mathcal{E}) \leq I(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0) - I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\right) + 2H_{\mathrm{b}}(p) + (1-p)\xi_1 + \delta_2(\mathcal{E}_0,\kappa) + \delta_3(\mathcal{E}_0,\kappa) + I\big(\mathbb{H}_0\hat{\mathbf{X}}_0; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1} \,\big|\, \hat{\mathbf{X}}_{-\kappa}^0\big) + \delta_1(\kappa) + \epsilon. \quad (249)$$

Before we continue to bound the first term on the RHS of (249), we define another indicator random variable

$$E_0 \triangleq \begin{cases} 1, & \text{if } \|\mathbf{X}_0\|^2 \geq \mathcal{E}_0 \\ 0, & \text{otherwise} \end{cases} \quad (250)$$

and let

$$p_0 \triangleq \Pr[E_0 = 1] = \Pr\big[\|\mathbf{X}_0\|^2 \geq \mathcal{E}_0\big]. \quad (251)$$

From Lemma 10 then follows that

$$\lim_{\mathcal{E}\uparrow\infty} p_0 = 1. \quad (252)$$

Similarly to (208)–(216) we get

$$I(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0) \leq I(\mathbf{X}_0, E_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0) \quad (253)$$

$$= I(E_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0) + I\big(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0 \,\big|\, E_0\big) \quad (254)$$

$$= H(E_0) - H\big(E_0 \,\big|\, \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0\big) + I\big(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0 \,\big|\, E_0\big) \quad (255)$$

$$\leq H(E_0) + I\big(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0 \,\big|\, E_0\big) \quad (256)$$

$$= H_{\mathrm{b}}(p_0) + p_0 I\big(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0 \,\big|\, E_0 = 1\big) + (1-p_0) I\big(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0 \,\big|\, E_0 = 0\big) \quad (257)$$

$$\leq H_{\mathrm{b}}(p_0) + I\big(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0 \,\big|\, E_0 = 1\big) + (1-p_0)\mathcal{C}_{\mathrm{IID}}(\mathcal{E}_0). \quad (258)$$

Here, in (253) we add E_0 to the argument of mutual information; (254) and (255) follow from the chain rule and the definition of mutual information (note again that E_0 is a binary random variable, *i.e.*, we need entropy instead of differential entropy); in (256) we lower-bound an entropy term by zero; and in the last inequality (258) we upper-bound $p_0 \leq 1$ and interpret the second mutual information term as mutual information between input and output of a memoryless MIMO fading channel (see (47)) under the constraint that the maximum available average power is $\mathcal{E}_0$. Hence, we can upper-bound this term by the memoryless capacity $\mathcal{C}_{\mathrm{IID}}(\mathcal{E}_0)$.

We remark that we do not know an analytic expression for $\mathcal{C}_{\mathrm{IID}}(\mathcal{E}_0)$. However, we know that it is finite (bounded) and independent of $\mathcal{E}$ so that from (252) we have

$$\lim_{\mathcal{E}\uparrow\infty} \big\{H_{\mathrm{b}}(p_0) + (1-p_0)\mathcal{C}_{\mathrm{IID}}(\mathcal{E}_0)\big\} = 0. \quad (259)$$

Using (231) we continue with the mutual information term in (258) as follows:

$$I\big(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0 \,\big|\, E_0 = 1\big) \leq I\big(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0, \mathbf{Z}_0 \,\big|\, E_0 = 1\big) \quad (260)$$

$$= I\big(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 \,\big|\, E_0 = 1\big) \quad (261)$$

$$= I\left(\mathbf{X}_0; \|\mathbb{H}_0\mathbf{X}_0\|, \frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, E_0 = 1\right) \quad (262)$$

$$= I\big(\mathbf{X}_0; \|\mathbb{H}_0\mathbf{X}_0\| \,\big|\, E_0 = 1\big) + I\left(\mathbf{X}_0; \frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \|\mathbb{H}_0\mathbf{X}_0\|, E_0 = 1\right) \quad (263)$$

$$= I\big(\mathbf{X}_0; \|\mathbb{H}_0\mathbf{X}_0\|, e^{\mathrm{i}\Theta_0} \,\big|\, E_0 = 1\big) - I\big(\mathbf{X}_0; e^{\mathrm{i}\Theta_0} \,\big|\, \|\mathbb{H}_0\mathbf{X}_0\|, E_0 = 1\big)$$

$$+ I\left(\mathbf{X}_0; \frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \|\mathbb{H}_0\mathbf{X}_0\|, E_0 = 1\right) \tag{264}$$

$$= I\big(\mathbf{X}_0; \|\mathbb{H}_0\mathbf{X}_0\| e^{\mathrm{i}\Theta_0} \,\big|\, E_0 = 1\big) + I\left(\mathbf{X}_0; \frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \|\mathbb{H}_0\mathbf{X}_0\|, E_0 = 1\right). \tag{265}$$

Here (260) follows from adding an additional random vector $\mathbf{Z}_0$ to the argument of the mutual information; (261) from subtracting the known vector $\mathbf{Z}_0$ from $\mathbf{Y}_0$ and from the independence between the noise and all other random quantities; then in (262) we split $\mathbb{H}_0\mathbf{X}_0$ into magnitude and direction vector (see (231)); (263) follows from the chain rule again; in (264) we use the chain rule and introduce $e^{\mathrm{i}\Theta_0}$ that is independent of all the other random quantities and that is uniformly distributed on the complex unit circle; and the last equality (265) follows from the independence of $e^{\mathrm{i}\Theta_0}$ from all other random quantities.

Next we apply Lemma 11 to the first term in (265), *i.e.*, we choose $\mathbf{S} = \mathbf{X}_0$ and $T = \|\mathbb{H}_0\mathbf{X}_0\| e^{\mathrm{i}\Theta_0}$. Note that we need to condition everything on the event $E_0 = 1$:

$$\begin{aligned} &I\big(\mathbf{X}_0; \|\mathbb{H}_0\mathbf{X}_0\| e^{\mathrm{i}\Theta_0} \,\big|\, E_0 = 1\big) \\ &\quad \le -h\big(\|\mathbb{H}_0\mathbf{X}_0\| e^{\mathrm{i}\Theta_0} \,\big|\, \mathbf{X}_0, E_0 = 1\big) + \log \pi + \alpha \log \beta \\ &\qquad + \log \Gamma\left(\alpha, \frac{\nu}{\beta}\right) \\ &\qquad + (1-\alpha)\mathsf{E}\left[\log\left(\|\mathbb{H}_0\mathbf{X}_0\|^2 + \nu\right) \,\middle|\, E_0 = 1\right] \\ &\qquad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\mathbf{X}_0\|^2 \,\middle|\, E_0 = 1\right] + \frac{\nu}{\beta} \end{aligned} \tag{266}$$

where $\alpha, \beta > 0$, and $\nu \ge 0$ can be chosen freely, but must not depend on $Q_{\mathcal{E},\epsilon}^{\kappa+1}$.

Note that from a conditional version of Lemma 2 with $m = 1$ follows that

$$\begin{aligned} &h\big(\|\mathbb{H}_0\mathbf{X}_0\| e^{\mathrm{i}\Theta_0} \,\big|\, \mathbf{X}_0 = \mathbf{x}_0, E_0 = 1\big) \\ &\quad = h_\lambda\big(e^{\mathrm{i}\Theta_0} \,\big|\, \mathbf{X}_0 = \mathbf{x}_0, E_0 = 1\big) \\ &\qquad + h\big(\|\mathbb{H}_0\mathbf{X}_0\| \,\big|\, e^{\mathrm{i}\Theta_0}, \mathbf{X}_0 = \mathbf{x}_0, E_0 = 1\big) \\ &\qquad + \mathsf{E}\big[\log \|\mathbb{H}_0\mathbf{X}_0\| \,\big|\, \mathbf{X}_0 = \mathbf{x}_0, E_0 = 1\big] \end{aligned} \tag{267}$$

$$\begin{aligned} &\quad = \log 2\pi + h\big(\|\mathbb{H}_0\mathbf{X}_0\| \,\big|\, \mathbf{X}_0 = \mathbf{x}_0, E_0 = 1\big) \\ &\qquad + \mathsf{E}\big[\log \|\mathbb{H}_0\mathbf{X}_0\| \,\big|\, \mathbf{X}_0 = \mathbf{x}_0, E_0 = 1\big] \end{aligned} \tag{268}$$

where we have used that $e^{\mathrm{i}\Theta_0}$ is independent of all other random quantities and uniformly distributed on the unit circle. Taking the expectation over $\mathbf{X}_0$ conditional on $E_0 = 1$ and noting that by the law of total expectation

$$\begin{aligned} &\mathsf{E}_{\mathbf{X}_0}\Big[\mathsf{E}\big[\log \|\mathbb{H}_0\mathbf{X}_0\| \,\big|\, \mathbf{X}_0 = \mathbf{x}_0, E_0 = 1\big] \,\Big|\, E_0 = 1\Big] \\ &\qquad\qquad = \mathsf{E}\big[\log \|\mathbb{H}_0\mathbf{X}_0\| \,\big|\, E_0 = 1\big] \end{aligned} \tag{269}$$

we then get

$$\begin{aligned} &h\big(\|\mathbb{H}_0\mathbf{X}_0\| e^{\mathrm{i}\Theta_0} \,\big|\, \mathbf{X}_0, E_0 = 1\big) \\ &\quad = \log 2\pi + h\big(\|\mathbb{H}_0\mathbf{X}_0\| \,\big|\, \mathbf{X}_0, E_0 = 1\big) \\ &\qquad + \mathsf{E}\big[\log \|\mathbb{H}_0\mathbf{X}_0\| \,\big|\, E_0 = 1\big] \end{aligned} \tag{270}$$

$$\begin{aligned} &\quad = \log 2\pi + h\big(\|\mathbb{H}_0\hat{\mathbf{X}}_0\| \cdot R_0 \,\big|\, \hat{\mathbf{X}}_0, R_0, E_0 = 1\big) \\ &\qquad + \mathsf{E}\Big[\log \|\mathbb{H}_0\hat{\mathbf{X}}_0\| \cdot R_0 \,\Big|\, E_0 = 1\Big] \end{aligned} \tag{271}$$

$$\begin{aligned} &\quad = \log 2\pi + h\big(\|\mathbb{H}_0\hat{\mathbf{X}}_0\| \,\big|\, \hat{\mathbf{X}}_0, R_0, E_0 = 1\big) \\ &\qquad + \mathsf{E}\big[\log R_0 \,\big|\, E_0 = 1\big] + \mathsf{E}\Big[\log \|\mathbb{H}_0\hat{\mathbf{X}}_0\| \,\Big|\, E_0 = 1\Big] \\ &\qquad + \mathsf{E}\big[\log R_0 \,\big|\, E_0 = 1\big] \end{aligned} \tag{272}$$

$$\begin{aligned} &\quad = \log 2\pi + h\big(\|\mathbb{H}_0\hat{\mathbf{X}}_0\| \,\big|\, \hat{\mathbf{X}}_0, E_0 = 1\big) \\ &\qquad + 2\mathsf{E}\big[\log R_0 \,\big|\, E_0 = 1\big] \\ &\qquad + \mathsf{E}\Big[\log \|\mathbb{H}_0\hat{\mathbf{X}}_0\| \,\Big|\, E_0 = 1\Big] \end{aligned} \tag{273}$$

where (271) follows from the definition of $R_0 \triangleq \|\mathbf{X}_0\|$; where (272) follows from the scaling property of entropy with a *real* argument; and where (273) follows because given $\hat{\mathbf{X}}_0$, $\|\mathbb{H}_0\hat{\mathbf{X}}_0\|$ is independent of R_0.

We assume $0 < \alpha < 1$ such that $1 - \alpha > 0$. Then we define

$$\epsilon_\nu \triangleq \sup_{\|\mathbf{x}\|^2 \ge \mathcal{E}_0} \Big\{ \mathsf{E}\left[\log\left(\|\mathbb{H}_0\mathbf{x}\|^2 + \nu\right)\right] - \mathsf{E}\left[\log \|\mathbb{H}_0\mathbf{x}\|^2\right] \Big\} \tag{274}$$

such that

$$\begin{aligned} &(1-\alpha)\mathsf{E}\left[\log\left(\|\mathbb{H}_0\mathbf{X}_0\|^2 + \nu\right) \,\middle|\, E_0 = 1\right] \\ &\quad = (1-\alpha)\mathsf{E}\left[\log \|\mathbb{H}_0\mathbf{X}_0\|^2 \,\middle|\, E_0 = 1\right] \\ &\qquad + (1-\alpha)\Big(\mathsf{E}\left[\log\left(\|\mathbb{H}_0\mathbf{X}_0\|^2 + \nu\right) \,\middle|\, E_0 = 1\right] \\ &\qquad\qquad - \mathsf{E}\left[\log \|\mathbb{H}_0\mathbf{X}_0\|^2 \,\middle|\, E_0 = 1\right]\Big) \end{aligned} \tag{275}$$

$$\begin{aligned} &\quad \le (1-\alpha)\mathsf{E}\left[\log \|\mathbb{H}_0\mathbf{X}_0\|^2 \,\middle|\, E_0 = 1\right] \\ &\qquad + (1-\alpha)\sup_{\|\mathbf{x}\|^2 \ge \mathcal{E}_0} \Big\{\mathsf{E}\left[\log\left(\|\mathbb{H}_0\mathbf{x}\|^2 + \nu\right)\right] \\ &\qquad\qquad - \mathsf{E}\left[\log \|\mathbb{H}_0\mathbf{x}\|^2\right]\Big\} \end{aligned} \tag{276}$$

$$= (1-\alpha)\mathsf{E}\left[\log \|\mathbb{H}_0\mathbf{X}_0\|^2 \,\middle|\, E_0 = 1\right] + (1-\alpha)\epsilon_\nu \tag{277}$$

$$\le (1-\alpha)\mathsf{E}\left[\log \|\mathbb{H}_0\mathbf{X}_0\|^2 \,\middle|\, E_0 = 1\right] + \epsilon_\nu. \tag{278}$$

Note that in (276) we use our knowledge $E_0 = 1$, *i.e.*, $R_0^2 \ge \mathcal{E}_0$.

Finally, we bound

$$\begin{aligned} &\frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\mathbf{X}_0\|^2 \,\middle|\, E_0 = 1\right] \\ &\qquad \le \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2 \cdot \|\mathbf{X}_0\|^2 \,\middle|\, E_0 = 1\right] \end{aligned} \tag{279}$$

$$= \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right] \cdot \mathsf{E}\left[R_0^2 \,\middle|\, E_0 = 1\right] \tag{280}$$

$$\le \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right] \cdot \frac{\mathcal{E}}{p_0} \tag{281}$$

where we use Cauchy-Schwarz in (279), then in (280) split the expectation of a product into a product of expectations because R_0 and $\mathbb{H}_0$ are independent, and finally in (281) use the fact that R_0 needs to satisfy the average-power constraint (35) to get the following bound:

$$\mathcal{E} \ge \mathsf{E}\left[R_0^2\right] \tag{282}$$

$$= p_0\mathsf{E}\left[R_0^2 \,\middle|\, E_0 = 1\right] + (1-p_0)\mathsf{E}\left[R_0^2 \,\middle|\, E_0 = 0\right] \tag{283}$$

$$\ge p_0\mathsf{E}\left[R_0^2 \,\middle|\, E_0 = 1\right]. \tag{284}$$

Plugging (281), (278), and (273) into (266) yields

$$\begin{aligned}
&I\big(\mathbf{X}_0; \|\mathbb{H}_0\mathbf{X}_0\|e^{\mathrm{i}\Theta_0} \,\big|\, E_0=1\big)\\
&\quad\le -\log 2 - h\big(\|\mathbb{H}_0\hat{\mathbf{X}}_0\| \,\big|\, \hat{\mathbf{X}}_0, E_0=1\big)\\
&\qquad - 2\mathsf{E}[\log R_0 \mid E_0=1] - \mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\| \,\Big|\, E_0=1\Big]\\
&\qquad + \alpha\log\beta + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right)\\
&\qquad + (1-\alpha)\mathsf{E}\left[\log\|\mathbb{H}_0\mathbf{X}_0\|^2 \,\big|\, E_0=1\right] + \epsilon_\nu\\
&\qquad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\cdot\frac{\mathcal{E}}{p_0} + \frac{\nu}{\beta}.
\end{aligned} \tag{285}$$

Next we continue with the second term in (265):

$$\begin{aligned}
&I\left(\mathbf{X}_0; \frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \|\mathbb{H}_0\mathbf{X}_0\|, E_0=1\right)\\
&\quad = h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \|\mathbb{H}_0\hat{\mathbf{X}}_0\|\cdot R_0, E_0=1\right)\\
&\qquad - h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \|\mathbb{H}_0\hat{\mathbf{X}}_0\|\cdot R_0, \hat{\mathbf{X}}_0, R_0, E_0=1\right) && (286)\\
&\quad = h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \|\mathbb{H}_0\hat{\mathbf{X}}_0\|\cdot R_0, E_0=1\right)\\
&\qquad - h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \|\mathbb{H}_0\hat{\mathbf{X}}_0\|, \hat{\mathbf{X}}_0, E_0=1\right) && (287)\\
&\quad \le h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, E_0=1\right)\\
&\qquad - h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \|\mathbb{H}_0\hat{\mathbf{X}}_0\|, \hat{\mathbf{X}}_0, E_0=1\right). && (288)
\end{aligned}$$

Here, (287) follows because given $\hat{\mathbf{X}}_0$ and $\|\mathbb{H}_0\hat{\mathbf{X}}_0\|$, the term $\mathbb{H}_0\hat{\mathbf{X}}_0/\|\mathbb{H}_0\hat{\mathbf{X}}_0\|$ does not depend on R_0; and (288) follows because conditioning cannot increase entropy.

Hence, using (288) and (285) in (265) we get

$$\begin{aligned}
&I\big(\mathbf{X}_0; \mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0 \,\big|\, E_0=1\big)\\
&\quad\le -\log 2 - h\big(\|\mathbb{H}_0\hat{\mathbf{X}}_0\| \,\big|\, \hat{\mathbf{X}}_0, E_0=1\big)\\
&\qquad - 2\mathsf{E}[\log R_0 \mid E_0=1] - \mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\| \,\Big|\, E_0=1\Big]\\
&\qquad + \alpha\log\beta + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right)\\
&\qquad + (1-\alpha)\mathsf{E}\left[\log\|\mathbb{H}_0\mathbf{X}_0\|^2 \,\big|\, E_0=1\right] + \epsilon_\nu + \frac{\nu}{\beta}\\
&\qquad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\frac{\mathcal{E}}{p_0} + h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, E_0=1\right)\\
&\qquad - h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \|\mathbb{H}_0\hat{\mathbf{X}}_0\|, \hat{\mathbf{X}}_0, E_0=1\right) && (289)\\
&\quad = -\log 2 - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \hat{\mathbf{X}}_0, E_0=1\big)\\
&\qquad + (2n_\mathsf{R}-1)\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\| \,\Big|\, E_0=1\Big]\\
&\qquad - 2\mathsf{E}[\log R_0 \mid E_0=1] - \mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\| \,\Big|\, E_0=1\Big]\\
&\qquad + \alpha\log\beta + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right)\\
&\qquad + 2\mathsf{E}[\log\|\mathbb{H}_0\mathbf{X}_0\| \mid E_0=1]\\
&\qquad - \alpha\mathsf{E}\left[\log\|\mathbb{H}_0\mathbf{X}_0\|^2 \,\big|\, E_0=1\right] + \epsilon_\nu + \frac{\nu}{\beta}\\
&\qquad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\frac{\mathcal{E}}{p_0} + h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, E_0=1\right) && (290)\\
&\quad = h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, E_0=1\right) - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \hat{\mathbf{X}}_0, E_0=1\big)\\
&\qquad + n_\mathsf{R}\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2 \,\Big|\, E_0=1\Big] - \log 2\\
&\qquad + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) + \epsilon_\nu\\
&\qquad + \alpha\Big(\log\beta - \mathsf{E}\left[\log\|\mathbb{H}_0\mathbf{X}_0\|^2 \,\big|\, E_0=1\right]\Big)\\
&\qquad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\frac{\mathcal{E}}{p_0} + \frac{\nu}{\beta} && (291)\\
&\quad \le h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, E_0=1\right) - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \hat{\mathbf{X}}_0, E_0=1\big)\\
&\qquad + n_\mathsf{R}\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2 \,\Big|\, E_0=1\Big] - \log 2\\
&\qquad + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) + \epsilon_\nu + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi_2\Big)\\
&\qquad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\frac{\mathcal{E}}{p_0} + \frac{\nu}{\beta}. && (292)
\end{aligned}$$

Here, (290) follows from a conditional version of Lemma 2 similar to (267)–(273) which allows us to combine the second and the last term in (289); in (291) we arithmetically rearrange the terms; and (292) follows from the following bound:

$$\begin{aligned}
&\mathsf{E}\left[\log\|\mathbb{H}_0\mathbf{X}_0\|^2 \,\big|\, E_0=1\right]\\
&\quad \ge \inf_{\|\mathbf{x}\|^2\ge\mathcal{E}_0}\mathsf{E}\left[\log\|\mathbb{H}_0\mathbf{x}\|^2\right] && (293)\\
&\quad = \log\mathcal{E}_0 + \inf_{\hat{\mathbf{x}}}\mathsf{E}\left[\log\|\mathbb{H}_0\hat{\mathbf{x}}\|^2\right] && (294)\\
&\quad \triangleq \log\mathcal{E}_0 + \xi_2 && (295)
\end{aligned}$$

where the last line should be read as definition for ξ_2. Notice that

$$-\infty < \xi_2 < \infty \tag{296}$$

as can be argued as follows: the lower bound on ξ_2 follows from [7, Lemma 6.7f)], [9, Lemma A.15f)] because $h(\mathbb{H}_0) > -\infty$ and $\mathsf{E}\left[\|\mathbb{H}_0\|_\mathrm{F}^2\right] < \infty$. The upper bound on ξ_2 can be verified using the concavity of the logarithm function and Jensen's inequality.

Hence, plugging (292) and (258) into (249) we get the following bound on capacity:

$$\begin{aligned}
\mathcal{C}(\mathcal{E}) \le{}& H_\mathrm{b}(p_0) + h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, E_0=1\right)\\
&- h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \hat{\mathbf{X}}_0, E_0=1\big)\\
&+ n_\mathsf{R}\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2 \,\Big|\, E_0=1\Big] - \log 2\\
&+ \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) + \epsilon_\nu + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi_2\Big)
\end{aligned}$$

$$
\begin{aligned}
&\quad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\frac{\mathcal{E}}{p_0} + \frac{\nu}{\beta} + (1-p_0)\mathcal{C}_{\text{IID}}(\mathcal{E}_0) \\
&\quad - I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\right) \\
&\quad + I\big(\mathbb{H}_0\hat{\mathbf{X}}_0; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1} \,\big|\, \hat{\mathbf{X}}_{-\kappa}^0\big) \\
&\quad + (1-p)\xi_1 + 2H_{\text{b}}(p) + \delta_1(\kappa) + \delta_2(\mathcal{E}_0,\kappa) \\
&\quad + \delta_3(\mathcal{E}_0,\kappa) + \epsilon && (297) \\
&= h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\, E_0 = 1\right) - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \hat{\mathbf{X}}_0, E_0 = 1\big) \\
&\quad + n_{\text{R}}\mathsf{E}\left[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2 \,\middle|\, E_0 = 1\right] - \log 2 \\
&\quad + I\big(\mathbb{H}_0\hat{\mathbf{X}}_0; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1} \,\big|\, \hat{\mathbf{X}}_{-\kappa}^0\big) \\
&\quad - I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\right) \\
&\quad + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) + \epsilon_\nu + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi_2\Big) \\
&\quad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\frac{\mathcal{E}}{p_0} + \frac{\nu}{\beta} + (1-p_0)\mathcal{C}_{\text{IID}}(\mathcal{E}_0) \\
&\quad + H_{\text{b}}(p_0) + 2H_{\text{b}}(p) + (1-p)\xi_1 + \delta_1(\kappa) \\
&\quad + \delta_2(\mathcal{E}_0,\kappa) + \delta_3(\mathcal{E}_0,\kappa) + \epsilon. && (298)
\end{aligned}
$$

This bound still depends on the distribution $Q_{\mathcal{E},\epsilon}^{\kappa+1}$ which is guaranteed to exist by Theorem 3, but whose exact form is not known. In order to get around this problem we will now further upper-bound this expression by maximizing it over all possible choices of $Q \in \mathcal{P}\big(\mathbb{C}^{n_{\text{T}}\times(\kappa+1)}\big)$ that are quasi-stationary and circularly symmetric:[6]

$$
\begin{aligned}
\mathcal{C}(\mathcal{E}) &\leq \sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^0} \\ \text{quasi-stat.} \\ \text{circ. sym.}}} \Bigg\{ h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\right) - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \hat{\mathbf{X}}_0\big) \\
&\qquad + n_{\text{R}}\mathsf{E}\left[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\right] - \log 2 \\
&\qquad + I\big(\mathbb{H}_0\hat{\mathbf{X}}_0; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1} \,\big|\, \hat{\mathbf{X}}_{-\kappa}^0\big) \\
&\qquad - I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\right)\Bigg\} \\
&\quad + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) + \epsilon_\nu + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi_2\Big) \\
&\quad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\frac{\mathcal{E}}{p_0} + \frac{\nu}{\beta} + (1-p_0)\mathcal{C}_{\text{IID}}(\mathcal{E}_0) \\
&\quad + H_{\text{b}}(p_0) + 2H_{\text{b}}(p) + (1-p)\xi_1 + \delta_1(\kappa) \\
&\quad + \delta_2(\mathcal{E}_0,\kappa) + \delta_3(\mathcal{E}_0,\kappa) + \epsilon && (299) \\
&= \sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^0} \\ \text{quasi-stat.} \\ \text{circ. sym.}}} \Bigg\{ h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\right) \\
&\qquad + n_{\text{R}}\mathsf{E}\left[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\right] - \log 2 \\
&\qquad - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^0\big)\Bigg\} \\
&\quad + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) + \epsilon_\nu + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi_2\Big) \\
&\quad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\frac{\mathcal{E}}{p_0} + \frac{\nu}{\beta} + (1-p_0)\mathcal{C}_{\text{IID}}(\mathcal{E}_0) \\
&\quad + H_{\text{b}}(p_0) + 2H_{\text{b}}(p) + (1-p)\xi_1 + \delta_1(\kappa) \\
&\quad + \delta_2(\mathcal{E}_0,\kappa) + \delta_3(\mathcal{E}_0,\kappa) + \epsilon && (300)
\end{aligned}
$$

[6]Note that by Proposition 7 we know that $Q_{\mathcal{E},\epsilon}^{\kappa+1}$ is circularly symmetric.

where (300) follows from the definition of mutual information. As mentioned, the supremum is over all distributions of blocks of unit vectors

$$
Q_{\hat{\mathbf{X}}_{-\kappa}^0} \in \mathcal{P}\big(\mathbb{C}^{n_{\text{T}}\times(\kappa+1)}\big) \tag{301}
$$

that are quasi-stationary in the sense of Theorem 3 and that are circularly symmetric as guaranteed by Proposition 7. *I.e.*, for every integer μ with $0 \leq \mu \leq \kappa$, every block of $\mu+1$ vectors $\big(\hat{\mathbf{X}}_k, \ldots, \hat{\mathbf{X}}_{k+\mu}\big)$ has the same distribution, and the vectors $\hat{\mathbf{X}}_k$ are circularly symmetric according to Definition 5.

Note that from (299) the alternative expression (90) can be derived.

We next use this upper bound on capacity to get an upper bound on the fading number (37):

$$
\begin{aligned}
&\chi(\{\mathbb{H}_k\}) \\
&\quad = \varlimsup_{\mathcal{E}\uparrow\infty}\left\{\mathcal{C}(\mathcal{E}) - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} && (302) \\
&\quad \leq \varlimsup_{\mathcal{E}\uparrow\infty}\Bigg\{ \sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^0} \\ \text{quasi-stat.} \\ \text{circ. sym.}}} \Bigg\{ h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\right) \\
&\qquad\qquad + n_{\text{R}}\mathsf{E}\left[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\right] - \log 2 \\
&\qquad\qquad - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^0\big)\Bigg\} \\
&\qquad + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) + \epsilon_\nu + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi_2\Big) \\
&\qquad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\frac{\mathcal{E}}{p_0} + \frac{\nu}{\beta} + (1-p_0)\mathcal{C}_{\text{IID}}(\mathcal{E}_0) \\
&\qquad + H_{\text{b}}(p_0) + 2H_{\text{b}}(p) + (1-p)\xi_1 + \delta_1(\kappa) \\
&\qquad + \delta_2(\mathcal{E}_0,\kappa) + \delta_3(\mathcal{E}_0,\kappa) + \epsilon \\
&\qquad - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right)\Bigg\} && (303) \\
&\quad = \sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^0} \\ \text{quasi-stat.} \\ \text{circ. sym.}}} \Bigg\{ h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\right) \\
&\qquad\qquad + n_{\text{R}}\mathsf{E}\left[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\right] - \log 2
\end{aligned}
$$

$$
\begin{aligned}
&\qquad - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\big) \Bigg\} \\
&+ \varlimsup_{\mathcal{E}\uparrow\infty} \Bigg\{ \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) - \log\frac{1}{\alpha} + \epsilon_\nu \\
&\qquad + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi_2\Big) + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\frac{\mathcal{E}}{p_0} \\
&\qquad + \frac{\nu}{\beta} + \log\frac{1}{\alpha} - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right) \Bigg\} \\
&+ \varlimsup_{\mathcal{E}\uparrow\infty} \Big\{ (1-p_0)\mathcal{C}_{\mathrm{IID}}(\mathcal{E}_0) + H_{\mathrm{b}}(p_0) + 2H_{\mathrm{b}}(p) \\
&\qquad + (1-p)\xi_1 \Big\} \\
&+ \delta_1(\kappa) + \delta_2(\mathcal{E}_0,\kappa) + \delta_3(\mathcal{E}_0,\kappa) + \epsilon \qquad (304)\\
&= \sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^{0}} \\ \text{quasi-stat.} \\ \text{circ. sym.}}} \Bigg\{ h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|} \right\}_{\ell=-\kappa}^{-1} \right) \\
&\qquad + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\Big] - \log 2 \\
&\qquad - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\big) \Bigg\} \\
&+ \log\left(1 - e^{-\nu}\right) + \nu - \log\nu + \epsilon_\nu \\
&+ \delta_1(\kappa) + \delta_2(\mathcal{E}_0,\kappa) + \delta_3(\mathcal{E}_0,\kappa) + \epsilon. \qquad (305)
\end{aligned}
$$

Here, (304) follows from the fact that $\delta_1(\kappa)$, $\delta_2(\mathcal{E}_0,\kappa)$, and $\delta_3(\mathcal{E}_0,\kappa)$ are independent of $Q_{\mathcal{E},\epsilon}^{\kappa+1}$ and therefore independent of $\mathcal{E}$ (this is shown in Appendix E and F), and from the fact that the four terms in the supremum do not depend on $\mathcal{E}$ either. Moreover we have added and subtracted a term $\log\frac{1}{\alpha}$. In (305) we have used that p and p_0 both tend to 1 as $\mathcal{E}$ tends to infinity (see (207) and (259)) and we have made the following choices on the free parameters α and β:

$$\alpha \triangleq \alpha(\mathcal{E}) = \frac{\nu}{\log\mathcal{E} + \log\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]} \qquad (306)$$

$$\beta \triangleq \beta(\mathcal{E}) = \frac{1}{\alpha(\mathcal{E})}e^{\nu/\alpha}. \qquad (307)$$

Note that for this choice

$$\varlimsup_{\mathcal{E}\uparrow\infty}\left\{\log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) - \log\frac{1}{\alpha}\right\} = \log\left(1 - e^{-\nu}\right) \qquad (308)$$

$$\varlimsup_{\mathcal{E}\uparrow\infty}\alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi_2\Big) = \nu \qquad (309)$$

$$\varlimsup_{\mathcal{E}\uparrow\infty}\left\{\frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}_0\|^2\right]\frac{\mathcal{E}}{p_0} + \frac{\nu}{\beta}\right\} = 0 \qquad (310)$$

$$\varlimsup_{\mathcal{E}\uparrow\infty}\left\{\log\frac{1}{\alpha} - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} = -\log\nu. \qquad (311)$$

(Compare with [7, App. VII], [9, Sec. B.5.9].)

In a next step we let ν go to zero. Note that $\epsilon_\nu \downarrow 0$ as $\nu \downarrow 0$ as can be seen from (274). Note further that

$$\lim_{\nu\downarrow 0}\left\{\log\left(1 - e^{-\nu}\right) - \log\nu\right\} = 0. \qquad (312)$$

Therefore, we get

$$
\begin{aligned}
\chi(\{\mathbb{H}_k\}) \le \sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^{0}} \\ \text{quasi-stat.} \\ \text{circ. sym.}}} &\Bigg\{ h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|} \right\}_{\ell=-\kappa}^{-1} \right) \\
&+ n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\Big] - \log 2 \\
&- h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\big) \Bigg\} \\
&+ \delta_1(\kappa) + \delta_2(\mathcal{E}_0,\kappa) + \delta_3(\mathcal{E}_0,\kappa) + \epsilon. \qquad (313)
\end{aligned}
$$

Next let $\mathcal{E}_0$ tend to infinity. Then it is shown in Appendix F that $\delta_2(\mathcal{E}_0,\kappa) \downarrow 0$ and $\delta_3(\mathcal{E}_0,\kappa) \downarrow 0$.

Finally, we let κ tend to infinity and recall that by Lemma 18 $\delta_1(\kappa) \downarrow 0$ when $\kappa \uparrow \infty$. Moreover, below we will show uniform convergence which guarantees that we are allowed to swap the supremum over $Q_{\hat{\mathbf{X}}_{-\kappa}^{0}}$ and the limit of κ tending to infinity. The result then follows because ϵ is arbitrary.

So it only remains to prove uniform convergence of the expression inside the supremum in (313) for $\kappa \uparrow \infty$. To that goal we define

$$
\begin{aligned}
\chi_\kappa(Q) \triangleq\ & h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|} \right\}_{\ell=-\kappa}^{-1} \right) \\
&+ n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\Big] - \log 2 \\
&- h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\big) \qquad (314)\\
\chi(Q) \triangleq\ & h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|} \right\}_{\ell=-\infty}^{-1} \right) \\
&+ n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\Big] - \log 2 \\
&- h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\infty}^{0}\big) \qquad (315)
\end{aligned}
$$

where Q stands for the stationary and circularly symmetric distribution of $\{\hat{\mathbf{X}}_k\}$. From the lower bound derived in Appendix G and from Theorem 3 we know that we can restrict Q to the set $\mathcal{Q}_\epsilon$ of distributions that will achieve the fading number $\chi(\{\mathbb{H}_k\})$ up to an $\epsilon > 0$:

$$\chi(Q) \ge \chi(\{\mathbb{H}_k\}) - \epsilon. \qquad (316)$$

Moreover, below we will show that $\sup_Q \chi_\kappa(Q)$ is a lower bound to the restricted fading number $\chi^{(\kappa)}(\{\mathbb{H}_k\})$ when the receiver only takes into account the memory up to a length of κ, but ignores all further past:

$$\sup_Q \chi_\kappa(Q) \le \chi^{(\kappa)}(\{\mathbb{H}_k\}). \qquad (317)$$

We hence have

$$
\begin{aligned}
\chi_\kappa(Q) - \chi(Q) &\le \sup_Q \chi_\kappa(Q) - \chi(\{\mathbb{H}_k\}) + \epsilon \qquad (318)\\
&\le \chi^{(\kappa)}(\{\mathbb{H}_k\}) - \chi(\{\mathbb{H}_k\}) + \epsilon \qquad (319)\\
&\le \epsilon. \qquad (320)
\end{aligned}
$$

Here, in (318) we maximize $\chi_\kappa(Q)$ over Q and use (316); in (319) we use (317); and in the last inequality (320) we rely on the fact that restricting the memory in the receiver cannot increase the fading number.

On the other hand, using the definitions (314) and (315) we have

$$\chi(Q) - \chi_\kappa(Q)$$
$$= I\big(\mathbb{H}_0\hat{\mathbf{X}}_0; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-\kappa-1} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\infty}^{0}\big)$$
$$\underbrace{- I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\infty}^{-\kappa-1} \,\middle|\, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\right)}_{\geq 0} \tag{321}$$
$$\leq I\big(\mathbb{H}_0\hat{\mathbf{X}}_0; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-\kappa-1}, \mathbb{H}_{-\infty}^{-\kappa-1} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\infty}^{0}\big) \tag{322}$$
$$= I\big(\mathbb{H}_0\hat{\mathbf{X}}_0; \mathbb{H}_{-\infty}^{-\kappa-1} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\infty}^{0}\big) \tag{323}$$
$$\leq \delta_1(\kappa) \tag{324}$$

where (322) follows because mutual information is nonnegative and because we add $\mathbb{H}_{-\infty}^{-\kappa-1}$ to the arguments of the first mutual information; where in (323) we drop $\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-\kappa-1}$ since it is a deterministic function of $\mathbb{H}_{-\infty}^{-\kappa-1}$ and $\hat{\mathbf{X}}_{-\infty}^{-\kappa-1}$; and where (324) follows from the derivation (343)–(365) given in Appendix E. Note that $\delta_1(\kappa)$ does not depend on Q and tends monotonically to 0 as κ tends to infinity.

Hence, for every $\epsilon > 0$ we can find a κ_ϵ such that for all $\kappa > \kappa_\epsilon$ we have

$$|\chi_\kappa(Q) - \chi(Q)| \leq \epsilon \tag{325}$$

for all $Q \in \mathcal{Q}_\epsilon$. This proves uniform convergence, *i.e.*,

$$\lim_{\kappa\uparrow\infty} \sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^0} \\ \text{quasi-stat.} \\ \text{circ. sym.}}} \chi_\kappa\left(Q_{\hat{\mathbf{X}}_{-\kappa}^0}\right) = \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}} \\ \text{stationary} \\ \text{circ. sym.}}} \lim_{\kappa\uparrow\infty} \chi_\kappa\left(Q_{\hat{\mathbf{X}}_{-\kappa}^0}\right). \tag{326}$$

We are left to prove (317). Assuming that the receiver only takes into account the past κ terms of the received signal, and using an input distribution as used in Appendix G we get

$$\mathcal{C}^{(\kappa)}(\mathcal{E})$$
$$\geq \varlimsup_{n\uparrow\infty} \frac{1}{n} \sum_{k=\kappa+1}^{n-\kappa} I\big(\mathbf{X}_1^n; \mathbf{Y}_k \,\big|\, \mathbf{Y}_1^{k-1}\big) \tag{327}$$
$$\geq \varlimsup_{n\uparrow\infty} \frac{1}{n} \sum_{k=\kappa+1}^{n-\kappa} I\big(\mathbf{X}_{k-\kappa}^k; \mathbf{Y}_k \,\big|\, \mathbf{Y}_{k-\kappa}^{k-1}\big) \tag{328}$$
$$= I\big(\mathbf{X}_{-\kappa}^0; \mathbf{Y}_0 \,\big|\, \mathbf{Y}_{-\kappa}^{-1}\big) \tag{329}$$
$$= I\big(\mathbf{X}_{-\kappa}^0, \mathbf{Y}_{-\kappa}^{-1}; \mathbf{Y}_0\big) - I\big(\mathbf{Y}_{-\kappa}^{-1}; \mathbf{Y}_0\big) \tag{330}$$
$$= I(\mathbf{X}_0; \mathbf{Y}_0) + I\big(\mathbf{X}_{-\kappa}^{-1}, \mathbf{Y}_{-\kappa}^{-1}; \mathbf{Y}_0 \,\big|\, \mathbf{X}_0\big) - I\big(\mathbf{Y}_{-\kappa}^{-1}; \mathbf{Y}_0\big) \tag{331}$$
$$\geq I(\mathbf{X}_0; \mathbf{Y}_0) + I\big(\mathbf{X}_{-\kappa}^{-1}, \mathbf{Y}_{-\kappa}^{-1}, \mathbf{Z}_{-\kappa}^{-1}; \mathbf{Y}_0 \,\big|\, \mathbf{X}_0\big) - \delta_2(\mathcal{E}_0, \kappa) - I\big(\mathbf{Y}_{-\kappa}^{-1}; \mathbf{Y}_0\big) \tag{332}$$
$$= I(\mathbf{X}_0; \mathbf{Y}_0) + I\big(\mathbf{X}_{-\kappa}^{-1}, \{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}; \mathbf{Y}_0 \,\big|\, \mathbf{X}_0\big) - I\big(\mathbf{Y}_{-\kappa}^{-1}; \mathbf{Y}_0\big) - \delta_2(\mathcal{E}_0, \kappa) \tag{333}$$
$$\geq I(\mathbf{X}_0; \mathbf{Y}_0) + I\big(\mathbf{X}_{-\kappa}^{-1}, \{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}; \mathbf{Y}_0, \mathbf{Z}_0 \,\big|\, \mathbf{X}_0\big) - I\big(\mathbf{Y}_{-\kappa}^{-1}; \mathbf{Y}_0\big) - \delta_2(\mathcal{E}_0, \kappa) - \delta_3(\mathcal{E}_0, \kappa) \tag{334}$$
$$= I(\mathbf{X}_0; \mathbf{Y}_0) + I\big(\mathbf{X}_{-\kappa}^{-1}, \{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}; \mathbb{H}_0\mathbf{X}_0 \,\big|\, \mathbf{X}_0\big) - I\big(\mathbf{Y}_{-\kappa}^{-1}; \mathbf{Y}_0\big) - \delta_2(\mathcal{E}_0, \kappa) - \delta_3(\mathcal{E}_0, \kappa) \tag{335}$$
$$\geq I(\mathbf{X}_0; \mathbf{Y}_0) + I\big(\hat{\mathbf{X}}_{-\kappa}^{-1}, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}; \mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \hat{\mathbf{X}}_0\big) - I\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}; \frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\right) - \delta_2(\mathcal{E}_0, \kappa) - \delta_3(\mathcal{E}_0, \kappa). \tag{336}$$

Here, (327) follows by dropping the supremum (we assume as input distribution the stationary distribution given in Appendix G) and some nonnegative terms in the sum; in (328) we drop some more terms and make use of the assumption that the receiver ignores the memory beyond $-\kappa$; (329) follows from stationarity; in (332) we use a derivation very similar to (366)–(378) and in (334) a derivation according to (380)–(392) in Appendix F, with the definition of δ_2 and δ_3 given there.

The bound (317) now follows from [18, Eq. (192)] and by letting $\mathcal{E}_0$ tend to infinity.

Appendix E
Proof of Lemma 18

We assume that $\mathbf{X}_{k-\kappa}^k$ is distributed according to the quasi-stationary distribution $Q_{\mathcal{E},\epsilon}^{\kappa+1}$. We bound as follows:

$$I\big(\mathbb{H}_1^{k-\kappa-1}; \mathbf{Y}_k \,\big|\, \mathbf{X}_{k-\kappa}^k, \{\mathbb{H}_\ell\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big)$$
$$= I\big(\mathbb{H}_1^{k-\kappa-1}; \mathbf{Y}_k \,\big|\, \hat{\mathbf{X}}_{k-\kappa}^k, \{\|\mathbf{X}_\ell\|\}_{\ell=k-\kappa}^k, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}\big) \tag{337}$$
$$\leq I\big(\mathbb{H}_1^{k-\kappa-1}; \mathbf{Y}_k, \mathbb{H}_k\hat{\mathbf{X}}_k \,\big|\, \hat{\mathbf{X}}_{k-\kappa}^k, \{\|\mathbf{X}_\ell\|\}_{\ell=k-\kappa}^k, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}\big) \tag{338}$$
$$= I\big(\mathbb{H}_1^{k-\kappa-1}; \mathbb{H}_k\hat{\mathbf{X}}_k \,\big|\, \hat{\mathbf{X}}_{k-\kappa}^k, \{\|\mathbf{X}_\ell\|\}_{\ell=k-\kappa}^k, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}\big) \tag{339}$$
$$= I\big(\mathbb{H}_1^{k-\kappa-1}; \mathbb{H}_k\hat{\mathbf{X}}_k \,\big|\, \hat{\mathbf{X}}_{k-\kappa}^k, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}\big) \tag{340}$$
$$= I\big(\mathbb{H}_{-k+1}^{-\kappa-1}; \mathbb{H}_0\hat{\mathbf{X}}_k \,\big|\, \hat{\mathbf{X}}_{k-\kappa}^k, \{\mathbb{H}_\ell\hat{\mathbf{X}}_{k+\ell}\}_{\ell=-\kappa}^{-1}\big). \tag{341}$$

Here, in (337) we split the vectors $\mathbf{X}_\ell$ up into magnitude and direction; in (338) we add the additional term $\mathbb{H}_k\hat{\mathbf{X}}_k$ to the argument of mutual information; in (339) we drop $\mathbf{Y}_k$ because given $\mathbb{H}_k\hat{\mathbf{X}}_k$ it is independent of the other random quantities; then in (340) we remove the conditioning on $\|\mathbf{X}_\ell\|$ because it does not provide any useful information; and in the last step (341) we made use of the stationarity of $\{\mathbb{H}_k\}$.

Similar to the derivation of the upper bound, in the following we will again introduce a shorthand and rename $\hat{\mathbf{X}}_{k-\kappa}^k$ by $\hat{\mathbf{X}}_{-\kappa}^0$. Note that since the upper bound that is derived in this appendix will not depend on $\{\hat{\mathbf{X}}_k\}$, we lose the dependence on k in the end anyway.

Hence, letting $\hat{\mathbf{X}}_{-\kappa}^0 \sim Q_{\mathcal{E},\epsilon}^{\kappa+1}$, we rewrite (341) as follows:

$$I\big(\mathbb{H}_{-k+1}^{-\kappa-1}; \mathbb{H}_0\hat{\mathbf{X}}_k \,\big|\, \hat{\mathbf{X}}_{k-\kappa}^k, \{\mathbb{H}_\ell\hat{\mathbf{X}}_{k+\ell}\}_{\ell=-\kappa}^{-1}\big)$$
$$= I\big(\mathbb{H}_{-k+1}^{-\kappa-1}; \mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \hat{\mathbf{X}}_{-\kappa}^0, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\big) \tag{342}$$
$$\leq I\big(\mathbb{H}_{-\infty}^{-\kappa-1}; \mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \hat{\mathbf{X}}_{-\kappa}^0, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\big) \tag{343}$$
$$= h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^0\big) - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^0, \mathbb{H}_{-\infty}^{-\kappa-1}\big) \tag{344}$$

$$\leq \frac{1}{\kappa+1} h\big(\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0} \,\big|\, \hat{\mathbf{X}}_{-\kappa}^{0}\big) - h\big(\mathbb{H}_0 \hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}, \mathbb{H}_{-\infty}^{-\kappa-1}\big). \tag{345}$$

Here in (343) we add more terms to the argument of mutual information; and (345) follows from the third statement of Lemma 12.

Now note that for $\hat{\mathbf{X}}_{-\kappa}^{0}$ being quasi-stationary and for all $i \in \{-\kappa, \ldots, -1\}$ we have

$$h\big(\mathbb{H}_i \hat{\mathbf{X}}_i \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{i-1}, \hat{\mathbf{X}}_{-\kappa}^{i}, \mathbb{H}_{-\infty}^{-\kappa-1}\big)$$
$$= h\big(\mathbb{H}_{i+1} \hat{\mathbf{X}}_{i+1} \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa+1}^{i}, \hat{\mathbf{X}}_{-\kappa+1}^{i+1}, \mathbb{H}_{-\infty}^{-\kappa}\big) \tag{346}$$
$$= h\big(\mathbb{H}_{i+1} \hat{\mathbf{X}}_{i+1} \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa+1}^{i}, \hat{\mathbf{X}}_{-\kappa+1}^{i+1}, \hat{\mathbf{X}}_{-\kappa}, \mathbb{H}_{-\infty}^{-\kappa}\big) \tag{347}$$
$$= h\big(\mathbb{H}_{i+1} \hat{\mathbf{X}}_{i+1} \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa+1}^{i}, \hat{\mathbf{X}}_{-\kappa}^{i+1}, \mathbb{H}_{-\infty}^{-\kappa}, \mathbb{H}_{-\kappa}\hat{\mathbf{X}}_{-\kappa}\big) \tag{348}$$
$$\leq h\big(\mathbb{H}_{i+1} \hat{\mathbf{X}}_{i+1} \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa+1}^{i}, \hat{\mathbf{X}}_{-\kappa}^{i+1}, \mathbb{H}_{-\infty}^{-\kappa-1}, \mathbb{H}_{-\kappa}\hat{\mathbf{X}}_{-\kappa}\big) \tag{349}$$
$$= h\big(\mathbb{H}_{i+1} \hat{\mathbf{X}}_{i+1} \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{i}, \hat{\mathbf{X}}_{-\kappa}^{i+1}, \mathbb{H}_{-\infty}^{-\kappa-1}\big) \tag{350}$$

where (346) follows from the stationarity of $\{\mathbb{H}_k\}$ and the quasi-stationarity of $\hat{\mathbf{X}}_{-\kappa}^{0}$ (note that $i < 0$ so that $i+1 \leq 0$); in (347) we add $\hat{\mathbf{X}}_{-\kappa}$ which, conditional on $\hat{\mathbf{X}}_{-\kappa+1}^{i+1}$, is independent of the other random quantities; then in (348) we add $\mathbb{H}_{-\kappa}\hat{\mathbf{X}}_{-\kappa}$ to the conditioning which does not change anything as it is a function of the given terms $\mathbb{H}_{-\kappa}$ and $\hat{\mathbf{X}}_{-\kappa}$; and the inequality (349) then follows by dropping $\mathbb{H}_{-\kappa}$ which cannot reduce entropy.

Therefore,

$$h\big(\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0} \,\big|\, \hat{\mathbf{X}}_{-\kappa}^{0}, \mathbb{H}_{-\infty}^{-\kappa-1}\big)$$
$$= \sum_{i=-\kappa}^{0} h\big(\mathbb{H}_i \hat{\mathbf{X}}_i \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{i-1}, \hat{\mathbf{X}}_{-\kappa}^{0}, \mathbb{H}_{-\infty}^{-\kappa-1}\big) \tag{351}$$
$$= \sum_{i=-\kappa}^{0} h\big(\mathbb{H}_i \hat{\mathbf{X}}_i \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{i-1}, \hat{\mathbf{X}}_{-\kappa}^{i}, \mathbb{H}_{-\infty}^{-\kappa-1}\big) \tag{352}$$
$$\leq \sum_{i=-\kappa}^{0} h\big(\mathbb{H}_0 \hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}, \mathbb{H}_{-\infty}^{-\kappa-1}\big) \tag{353}$$
$$= (\kappa+1) h\big(\mathbb{H}_0 \hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}, \mathbb{H}_{-\infty}^{-\kappa-1}\big). \tag{354}$$

Here, (351) follows from the chain rule; in (352) we drop $\hat{\mathbf{X}}_{i+1}^{0}$ because conditioned on $\hat{\mathbf{X}}_{-\kappa}^{i}$ they are independent of the other random quantities; and in (353) we apply (350) several times to each term of the sum.

Hence we have

$$h\big(\mathbb{H}_0 \hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}, \mathbb{H}_{-\infty}^{-\kappa-1}\big) \geq \frac{1}{\kappa+1} h\big(\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0} \,\big|\, \hat{\mathbf{X}}_{-\kappa}^{0}, \mathbb{H}_{-\infty}^{-\kappa-1}\big). \tag{355}$$

Using this in (345) and (341) we finally get

$$I\big(\mathbb{H}_1^{k-\kappa-1}; \mathbf{Y}_k \,\big|\, \mathbf{X}_{k-\kappa}^{k}, \{\mathbb{H}_\ell \mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big)$$
$$\leq \frac{1}{\kappa+1} h\big(\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0} \,\big|\, \hat{\mathbf{X}}_{-\kappa}^{0}\big) - \frac{1}{\kappa+1} h\big(\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0} \,\big|\, \hat{\mathbf{X}}_{-\kappa}^{0}, \mathbb{H}_{-\infty}^{-\kappa-1}\big) \tag{356}$$
$$= \frac{1}{\kappa+1} I\big(\mathbb{H}_{-\infty}^{-\kappa-1}; \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0} \,\big|\, \hat{\mathbf{X}}_{-\kappa}^{0}\big) \tag{357}$$
$$= \frac{1}{\kappa+1} I\big(\mathbb{H}_{-\infty}^{-\kappa-1}; \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0}, \hat{\mathbf{X}}_{-\kappa}^{0}\big) \tag{358}$$
$$\leq \frac{1}{\kappa+1} I\big(\mathbb{H}_{-\infty}^{-\kappa-1}; \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0}, \hat{\mathbf{X}}_{-\kappa}^{0}, \mathbb{H}_{-\kappa}^{0}\big) \tag{359}$$
$$= \frac{1}{\kappa+1} I\big(\mathbb{H}_{-\infty}^{-\kappa-1}; \mathbb{H}_{-\kappa}^{0}\big) \tag{360}$$
$$= \frac{1}{\kappa+1} h(\mathbb{H}_{-\kappa}^{0}) - \frac{1}{\kappa+1} h(\mathbb{H}_{-\kappa}^{0} | \mathbb{H}_{-\infty}^{-\kappa-1}) \tag{361}$$
$$= \frac{1}{\kappa+1} h(\mathbb{H}_{-\kappa}^{0}) - \frac{1}{\kappa+1} \sum_{i=-\kappa}^{0} h(\mathbb{H}_i | \mathbb{H}_{-\infty}^{-i-1}) \tag{362}$$
$$= \frac{1}{\kappa+1} h(\mathbb{H}_{-\kappa}^{0}) - \frac{1}{\kappa+1} \sum_{i=-\kappa}^{0} h(\mathbb{H}_0 | \mathbb{H}_{-\infty}^{-1}) \tag{363}$$
$$= \frac{1}{\kappa+1} h(\mathbb{H}_{-\kappa}^{0}) - h(\mathbb{H}_0 | \mathbb{H}_{-\infty}^{-1}) \tag{364}$$
$$\triangleq \delta_1(\kappa). \tag{365}$$

Here, (358) follows because $\hat{\mathbf{X}}_{-\kappa}^{0}$ is independent of $\mathbb{H}_{-\infty}^{-\kappa-1}$; in (359) we add $\mathbb{H}_{-\kappa}^{0}$ to the argument of mutual information; in (360) we drop $(\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0}, \hat{\mathbf{X}}_{-\kappa}^{0})$ because conditional on $\mathbb{H}_{-\kappa}^{0}$ it is independent of $\mathbb{H}_{-\infty}^{-\kappa-1}$; and (363) follows from stationarity.

Note that $\delta_1(\kappa)$ does neither depend on k nor on the input $\{\mathbf{X}_k\}$ and that—due to the stationarity of $\{\mathbb{H}_k\}$—it monotonically tends to zero as κ tends to infinity.

Appendix F
Proof of Lemma 19

We start with the first bound:

$$I\big(\mathbb{H}_0 \mathbf{X}_0 + \mathbf{Z}_0; \mathbf{Z}_{-\kappa}^{-1} \,\big|\, \{\mathbb{H}_\ell \mathbf{X}_\ell + \mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}, E=1\big)$$
$$= h\big(\mathbf{Z}_{-\kappa}^{-1} \,\big|\, \{\mathbb{H}_\ell \mathbf{X}_\ell + \mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}, E=1\big) - h\big(\mathbf{Z}_{-\kappa}^{-1} \,\big|\, \{\mathbb{H}_\ell \mathbf{X}_\ell + \mathbf{Z}_\ell\}_{\ell=-\kappa}^{0}, E=1\big) \tag{366}$$
$$\leq h\big(\mathbf{Z}_{-\kappa}^{-1}\big) - h\big(\mathbf{Z}_{-\kappa}^{-1} \,\big|\, \{\mathbb{H}_\ell \mathbf{X}_\ell + \mathbf{Z}_\ell\}_{\ell=-\kappa}^{0}, \mathbf{X}_{-\kappa}^{-1}, \mathbb{H}_0, E=1\big) \tag{367}$$
$$= h\big(\mathbf{Z}_{-\kappa}^{-1}\big) - h\big(\mathbf{Z}_{-\kappa}^{-1} \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell R_\ell + \mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}, R_{-\kappa}^{-1}, \mathbb{H}_0, E=1\big) \tag{368}$$
$$\leq h\big(\mathbf{Z}_{-\kappa}^{-1}\big) - \inf_{\substack{r_{-\kappa}^2 \geq \mathcal{E}_0, \ldots, \\ r_{-1}^2 \geq \mathcal{E}_0}} h\big(\mathbf{Z}_{-\kappa}^{-1} \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell r_\ell + \mathbf{Z}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}, \mathbb{H}_0, E=1\big) \tag{369}$$
$$= h\big(\mathbf{Z}_{-\kappa}^{-1} \,\big|\, \hat{\mathbf{X}}_{-\kappa}^{-1}, \mathbb{H}_0, E=1\big) - h\left(\mathbf{Z}_{-\kappa}^{-1} \,\middle|\, \left\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell \sqrt{\mathcal{E}_0} + \mathbf{Z}_\ell\right\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}, \mathbb{H}_0, E=1\right) \tag{370}$$
$$= I\left(\mathbf{Z}_{-\kappa}^{-1}; \left\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell \sqrt{\mathcal{E}_0} + \mathbf{Z}_\ell\right\}_{\ell=-\kappa}^{-1} \,\middle|\, \hat{\mathbf{X}}_{-\kappa}^{-1}, \mathbb{H}_0, E=1\right) \tag{371}$$
$$= I\left(\left\{\frac{\mathbf{Z}_\ell}{\sqrt{\mathcal{E}_0}}\right\}_{\ell=-\kappa}^{-1}; \left\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell + \frac{\mathbf{Z}_\ell}{\sqrt{\mathcal{E}_0}}\right\}_{\ell=-\kappa}^{-1} \,\middle|\, \hat{\mathbf{X}}_{-\kappa}^{-1},\right.$$

$$\left. \mathbb{H}_0, E=1 \right) \quad (372)$$

$$\leq \sup_{\substack{\hat{\mathbf{x}}_{-\kappa},\ldots,\\ \hat{\mathbf{x}}_{-1}}} I\left(\left\{ \frac{\mathbf{Z}_\ell}{\sqrt{\mathcal{E}_0}} \right\}_{\ell=-\kappa}^{-1} ; \left\{ \mathbb{H}_\ell \hat{\mathbf{x}}_\ell + \frac{\mathbf{Z}_\ell}{\sqrt{\mathcal{E}_0}} \right\}_{\ell=-\kappa}^{-1} \,\middle|\, \mathbb{H}_0 \right) \quad (373)$$

$$= \sup_{\substack{\hat{\mathbf{x}}_{-\kappa},\ldots,\\ \hat{\mathbf{x}}_{-1}}} I\left(\left\{ \frac{\sigma \mathbb{Z}_\ell \hat{\mathbf{x}}_\ell}{\sqrt{\mathcal{E}_0}} \right\}_{\ell=-\kappa}^{-1} ; \left\{ \mathbb{H}_\ell \hat{\mathbf{x}}_\ell + \frac{\sigma \mathbb{Z}_\ell \hat{\mathbf{x}}_\ell}{\sqrt{\mathcal{E}_0}} \right\}_{\ell=-\kappa}^{-1} \,\middle|\, \mathbb{H}_0 \right) \quad (374)$$

$$= \sup_{\substack{\hat{\mathbf{x}}_{-\kappa},\ldots,\\ \hat{\mathbf{x}}_{-1}}} I\left(\left\{ \frac{\sigma \mathbb{Z}_\ell \hat{\mathbf{x}}_\ell}{\sqrt{\mathcal{E}_0}} \right\}_{\ell=-\kappa}^{-1} ; \left\{ \left(\mathbb{H}_\ell + \frac{\sigma \mathbb{Z}_\ell}{\sqrt{\mathcal{E}_0}} \right) \hat{\mathbf{x}}_\ell \right\}_{\ell=-\kappa}^{-1} \,\middle|\, \mathbb{H}_0 \right) \quad (375)$$

$$\leq I\left(\left\{ \frac{\sigma \mathbb{Z}_\ell}{\sqrt{\mathcal{E}_0}} \right\}_{\ell=-\kappa}^{-1} ; \left\{ \mathbb{H}_\ell + \frac{\sigma \mathbb{Z}_\ell}{\sqrt{\mathcal{E}_0}} \right\}_{\ell=-\kappa}^{-1} \,\middle|\, \mathbb{H}_0 \right) \quad (376)$$

$$= h\left(\left\{ \mathbb{H}_\ell + \frac{\sigma \mathbb{Z}_\ell}{\sqrt{\mathcal{E}_0}} \right\}_{\ell=-\kappa}^{-1} \,\middle|\, \mathbb{H}_0 \right) - h\big(\mathbb{H}_{-\kappa}^{-1} \,\big|\, \mathbb{H}_0\big) \quad (377)$$

$$\triangleq \delta_2(\mathcal{E}_0, \kappa). \quad (378)$$

Here (367) follows from conditioning that reduces entropy; in (368) we drop $\mathbf{Y}_0$ since given $\mathbb{H}_0$ it is independent of the other random variables; in (369) and (370) we upper-bound the terms by choosing the smallest value for R_ℓ which is $\sqrt{\mathcal{E}_0}$ and we condition the first entropy on random variables that are independent of the noise; in (373) we upper-bound the expectation by a supremum; in (374) we introduce a random $n_\mathrm{R} \times n_\mathrm{T}$ matrix process $\{\mathbb{Z}_k\}$ that is IID with components that are independent $\mathcal{N}_\mathbb{C}(0,1)$ so that $\sigma \mathbb{Z}_\ell \hat{\mathbf{x}}_\ell \sim \mathcal{N}_\mathbb{C}\big(0, \sigma^2 \mathsf{I}_{n_\mathrm{R}}\big)$; and (376) follows from the data processing inequality.

From [7, Lemma 6.11], [9, Lemma A.19] we conclude that for every realization of $\mathbb{H}_0$ and for any κ the expression

$$h\left(\left\{ \mathbb{H}_\ell + \frac{\sigma \mathbb{Z}_\ell}{\sqrt{\mathcal{E}_0}} \right\}_{\ell=-\kappa}^{-1} \,\middle|\, \mathbb{H}_0 = \mathsf{H}_0 \right) \quad (379)$$

converges monotonically in $\mathcal{E}_0$ to $h\big(\mathbb{H}_{-\kappa}^{-1} \,\big|\, \mathbb{H}_0 = \mathsf{H}_0\big)$. By the Monotone Convergence Theorem (MCT) [21] this is also true when we average over $\mathbb{H}_0$.

Similarly, we find:

$$\begin{aligned} I\big(&\mathbf{Z}_0; \{\mathbb{H}_\ell \mathbf{X}_\ell\}_{\ell=-\kappa}^{-1} \,\big|\, \mathbb{H}_0 \mathbf{X}_0 + \mathbf{Z}_0, E=1\big) \\ &= h\big(\mathbf{Z}_0 \,\big|\, \mathbb{H}_0 \mathbf{X}_0 + \mathbf{Z}_0, E=1\big) \\ &\quad - h\big(\mathbf{Z}_0 \,\big|\, \{\mathbb{H}_\ell \mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}, \mathbb{H}_0 \mathbf{X}_0 + \mathbf{Z}_0, E=1\big) && (380) \\ &\leq h\big(\mathbf{Z}_0\big) \\ &\quad - h\big(\mathbf{Z}_0 \,\big|\, \{\mathbb{H}_\ell \mathbf{X}_\ell\}_{\ell=-\kappa}^{-1}, \mathbb{H}_0 \mathbf{X}_0 + \mathbf{Z}_0, \mathbf{X}_0, \mathbb{H}_{-\kappa}^{-1}, E=1\big) && (381) \\ &= h\big(\mathbf{Z}_0\big) - h\big(\mathbf{Z}_0 \,\big|\, \mathbb{H}_0 \hat{\mathbf{X}}_0 R_0 + \mathbf{Z}_0, \hat{\mathbf{X}}_0, R_0, \mathbb{H}_{-\kappa}^{-1}, E=1\big) && (382) \\ &\leq h\big(\mathbf{Z}_0\big) - \inf_{r_0^2 \geq \mathcal{E}_0} h\big(\mathbf{Z}_0 \,\big|\, \mathbb{H}_0 \hat{\mathbf{X}}_0 r_0 + \mathbf{Z}_0, \hat{\mathbf{X}}_0, \mathbb{H}_{-\kappa}^{-1}, E=1\big) && (383) \\ &= h\big(\mathbf{Z}_0 \,\big|\, \hat{\mathbf{X}}_0, \mathbb{H}_{-\kappa}^{-1}, E=1\big) \\ &\quad - h\left(\mathbf{Z}_0 \,\middle|\, \mathbb{H}_0 \hat{\mathbf{X}}_0 \sqrt{\mathcal{E}_0} + \mathbf{Z}_0, \hat{\mathbf{X}}_0, \mathbb{H}_{-\kappa}^{-1}, E=1\right) && (384) \\ &= I\left(\mathbf{Z}_0; \mathbb{H}_0 \hat{\mathbf{X}}_0 \sqrt{\mathcal{E}_0} + \mathbf{Z}_0 \,\middle|\, \hat{\mathbf{X}}_0, \mathbb{H}_{-\kappa}^{-1}, E=1\right) && (385) \\ &= I\left(\frac{\mathbf{Z}_0}{\sqrt{\mathcal{E}_0}}; \mathbb{H}_0 \hat{\mathbf{X}}_0 + \frac{\mathbf{Z}_0}{\sqrt{\mathcal{E}_0}} \,\middle|\, \hat{\mathbf{X}}_0, \mathbb{H}_{-\kappa}^{-1}, E=1\right) && (386) \\ &\leq \sup_{\hat{\mathbf{x}}_0} I\left(\frac{\mathbf{Z}_0}{\sqrt{\mathcal{E}_0}}; \mathbb{H}_0 \hat{\mathbf{x}}_0 + \frac{\mathbf{Z}_0}{\sqrt{\mathcal{E}_0}} \,\middle|\, \mathbb{H}_{-\kappa}^{-1}\right) && (387) \\ &= \sup_{\hat{\mathbf{x}}_0} I\left(\frac{\sigma \mathbb{Z}_0 \hat{\mathbf{x}}_0}{\sqrt{\mathcal{E}_0}}; \mathbb{H}_0 \hat{\mathbf{x}}_0 + \frac{\sigma \mathbb{Z}_0 \hat{\mathbf{x}}_0}{\sqrt{\mathcal{E}_0}} \,\middle|\, \mathbb{H}_{-\kappa}^{-1}\right) && (388) \\ &= \sup_{\hat{\mathbf{x}}_0} I\left(\frac{\sigma \mathbb{Z}_0 \hat{\mathbf{x}}_0}{\sqrt{\mathcal{E}_0}}; \left(\mathbb{H}_0 + \frac{\sigma \mathbb{Z}_0}{\sqrt{\mathcal{E}_0}}\right) \hat{\mathbf{x}}_0 \,\middle|\, \mathbb{H}_{-\kappa}^{-1}\right) && (389) \\ &\leq I\left(\frac{\sigma \mathbb{Z}_0}{\sqrt{\mathcal{E}_0}}; \mathbb{H}_0 + \frac{\sigma \mathbb{Z}_0}{\sqrt{\mathcal{E}_0}} \,\middle|\, \mathbb{H}_{-\kappa}^{-1}\right) && (390) \\ &= h\left(\mathbb{H}_0 + \frac{\sigma \mathbb{Z}_0}{\sqrt{\mathcal{E}_0}} \,\middle|\, \mathbb{H}_{-\kappa}^{-1}\right) - h\big(\mathbb{H}_0 \,\big|\, \mathbb{H}_{-\kappa}^{-1}\big) && (391) \\ &\triangleq \delta_3(\mathcal{E}_0, \kappa) && (392) \end{aligned}$$

where (390) again follows from the data processing inequality.

From [7, Lemma 6.11], [9, Lemma A.19] we conclude that for every realization of $\mathbb{H}_{-\kappa}^{-1}$ the expression

$$h\left(\mathbb{H}_0 + \frac{\sigma \mathbb{Z}_0}{\sqrt{\mathcal{E}_0}} \,\middle|\, \{\mathbb{H}_\ell = \mathsf{H}_\ell\}_{\ell=-\kappa}^{-1}\right) \quad (393)$$

converges monotonically in $\mathcal{E}_0$ to $h\big(\mathbb{H}_0 \,\big|\, \{\mathbb{H}_\ell = \mathsf{H}_\ell\}_{\ell=-\kappa}^{-1}\big)$. By the Monotone Convergence Theorem (MCT) [21] this is also true when we average over $\mathbb{H}_{-\kappa}^{-1}$.

APPENDIX G
DERIVATION OF A LOWER BOUND FOR THEOREM 13

To derive a lower bound we choose a specific input distribution which naturally yields a lower bound to channel capacity. Let $\{\mathbf{X}_k\}$ be of the form

$$\mathbf{X}_k \triangleq R_k \cdot \hat{\mathbf{X}}_k. \quad (394)$$

Here $\{\hat{\mathbf{X}}_k\}$ is a sequence of random unit vectors forming a stochastic process that is stationary and circularly symmetric, but whose exact distribution will be specified later. The stochastic process $\{R_k\}$ is chosen to be independent of $\{\hat{\mathbf{X}}_k\}$ and to consist of random variables $R_k \in \mathbb{R}_0^+$ that are IID with

$$\log R_k^2 \sim \mathcal{U}\big([\log x_{\min}^2, \log \mathcal{E}]\big) \quad (395)$$

where we choose $x_{\min}^2$ as

$$x_{\min}^2 \triangleq \log \mathcal{E}. \quad (396)$$

Note that this choice of $\{R_k\}$ satisfies the peak-power constraint (34) and therefore also the average-power constraint (35).

Fix some (large) positive integer κ and use the chain rule and the nonnegativity of mutual information to obtain:

$$\begin{aligned} \frac{1}{n} I\big(\mathbf{X}_1^n; \mathbf{Y}_1^n\big) &= \frac{1}{n} \sum_{k=1}^{n} I\big(\mathbf{X}_k; \mathbf{Y}_1^n \,\big|\, \mathbf{X}_1^{k-1}\big) && (397) \\ &\geq \frac{1}{n} \sum_{k=\kappa+1}^{n-\kappa} I\big(\mathbf{X}_k; \mathbf{Y}_1^n \,\big|\, \mathbf{X}_1^{k-1}\big). && (398) \end{aligned}$$

Then for every $\kappa + 1 \le k \le n - \kappa$, we can use the fact that $\{R_k\}$ is IID and that $\{\hat{\mathbf{X}}_k\}$ is stationary and circularly symmetric to lower-bound $I(\mathbf{X}_k; \mathbf{Y}_1^n \mid \mathbf{X}_1^{k-1})$ as follows:

$$
\begin{aligned}
&I(\mathbf{X}_k; \mathbf{Y}_1^n \mid \mathbf{X}_1^{k-1}) \\
&= I(\mathbf{X}_k; \mathbf{Y}_1^n, \mathbf{X}_1^{k-1}) - I(\mathbf{X}_k; \mathbf{X}_1^{k-1}) && (399)\\
&= I(\mathbf{X}_k; \mathbf{Y}_1^n, \mathbf{X}_1^{k-1}) - I(\mathbf{X}_k; \mathbf{X}_{k-\kappa}^{k-1}) \\
&\quad - \underbrace{I(\mathbf{X}_k; \mathbf{X}_1^{k-\kappa-1} \mid \mathbf{X}_{k-\kappa}^{k-1})}_{\le \delta_4(\kappa)} && (400)\\
&\ge I(\mathbf{X}_k; \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{X}_{k-\kappa}^{k-1}) - I(\mathbf{X}_k; \mathbf{X}_{k-\kappa}^{k-1}) - \delta_4(\kappa) && (401)\\
&= I(\mathbf{X}_k; \mathbf{Y}_{k-\kappa}^{k+\kappa} \mid \mathbf{X}_{k-\kappa}^{k-1}) - \delta_4(\kappa) && (402)\\
&= I(\mathbf{X}_k; \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{Z}_{k-\kappa}^{k-1}, \mathbf{Z}_{k+1}^{k+\kappa} \mid \mathbf{X}_{k-\kappa}^{k-1}) \\
&\quad - \underbrace{I(\mathbf{X}_k; \mathbf{Z}_{k-\kappa}^{k-1}, \mathbf{Z}_{k+1}^{k+\kappa} \mid \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{X}_{k-\kappa}^{k-1})}_{\le \delta_5(x_{\min}, \kappa)} - \delta_4(\kappa) && (403)\\
&\ge I(\mathbf{X}_k; \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{Z}_{k-\kappa}^{k-1}, \mathbf{Z}_{k+1}^{k+\kappa} \mid \mathbf{X}_{k-\kappa}^{k-1}) - \delta_5(x_{\min}, \kappa) \\
&\quad - \delta_4(\kappa) && (404)\\
&= I\big(\mathbf{X}_k; \mathbf{Y}_k, \{\mathbb{H}_\ell \mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}, \{\mathbb{H}_\ell \mathbf{X}_\ell\}_{\ell=k+1}^{k+\kappa} \mid \mathbf{X}_{k-\kappa}^{k-1}\big) \\
&\quad - \delta_4(\kappa) - \delta_5(x_{\min}, \kappa) && (405)\\
&= I\left(\mathbf{X}_k; \mathbf{Y}_k, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \big\{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| R_\ell\big\}_{\ell=k+1}^{k+\kappa}, \right. \\
&\qquad \left. \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa} \middle| R_{k-\kappa}^{k-1}, \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) \\
&\quad - \delta_4(\kappa) - \delta_5(x_{\min}, \kappa) && (406)\\
&\ge I\left(\mathbf{X}_k; \mathbf{Y}_k, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa} \middle| R_{k-\kappa}^{k-1}, \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) \\
&\quad - \delta_4(\kappa) - \delta_5(x_{\min}, \kappa) && (407)\\
&= I\left(\mathbf{X}_k; \mathbf{Y}_k, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa} \middle| \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) \\
&\quad - \delta_4(\kappa) - \delta_5(x_{\min}, \kappa). && (408)
\end{aligned}
$$

Here (399) and (400) follow from the chain rule; (401) follows from dropping some random quantities in the argument of the first mutual information term and from the following bound:

$$
\begin{aligned}
&I(\mathbf{X}_k; \mathbf{X}_1^{k-\kappa-1} \mid \mathbf{X}_{k-\kappa}^{k-1}) \\
&\quad = I(\mathbf{X}_0; \mathbf{X}_{-k+1}^{-\kappa-1} \mid \mathbf{X}_{-\kappa}^{-1}) && (409)\\
&\quad = I(R_0, \hat{\mathbf{X}}_0; R_{-k+1}^{-\kappa-1}, \hat{\mathbf{X}}_{-k+1}^{-\kappa-1} \mid R_{-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}) && (410)\\
&\quad = I(R_0; R_{-k+1}^{-\kappa-1} \mid R_{-\kappa}^{-1}) + I(\hat{\mathbf{X}}_0; \hat{\mathbf{X}}_{-k+1}^{-\kappa-1} \mid \hat{\mathbf{X}}_{-\kappa}^{-1}) && (411)\\
&\quad = I(\hat{\mathbf{X}}_0; \hat{\mathbf{X}}_{-k+1}^{-\kappa-1} \mid \hat{\mathbf{X}}_{-\kappa}^{-1}) && (412)\\
&\quad \le I(\hat{\mathbf{X}}_0; \hat{\mathbf{X}}_{-\infty}^{-\kappa-1} \mid \hat{\mathbf{X}}_{-\kappa}^{-1}) && (413)\\
&\quad = h_\lambda(\hat{\mathbf{X}}_0 \mid \hat{\mathbf{X}}_{-\kappa}^{-1}) - h_\lambda(\hat{\mathbf{X}}_0 \mid \hat{\mathbf{X}}_{-\infty}^{-1}) && (414)\\
&\quad = h_\lambda(\hat{\mathbf{X}}_0 \mid \hat{\mathbf{X}}_{-\kappa}^{-1}) - h_\lambda(\{\hat{\mathbf{X}}_k\}) && (415)\\
&\quad \triangleq \delta_4(\kappa) && (416)
\end{aligned}
$$

where the last equality should be read as definition of $\delta_4(\kappa)$. Here (409) follows from stationarity of $\{\mathbf{X}_k\}$; in (410) we split $\mathbf{X}_k$ into magnitude and direction; (411) follows because $\{R_k\}$ and $\{\hat{\mathbf{X}}_k\}$ are independent; in (412) we use the fact that $\{R_k\}$ is IID; (413) follows from adding terms to mutual information; and (415) from the definition of entropy rate and from stationarity. Note that while $\delta_4(\kappa)$ does depend on the distribution of $\{\hat{\mathbf{X}}_k\}$, it is independent of k and $\mathcal{E}$ and tends to zero as κ tends to infinity.

Then, (402) and (403) follow again from the chain rule, and (404) follows from the following lemma.

Lemma 20: Let $\{\mathbf{X}_k\}$ be as specified in (394)–(396). Then

$$
I(\mathbf{X}_k; \mathbf{Z}_{k-\kappa}^{k-1}, \mathbf{Z}_{k+1}^{k+\kappa} \mid \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{X}_{k-\kappa}^{k-1}) \le \delta_5(x_{\min}, \kappa) \tag{417}
$$

where $\delta_5(x_{\min}, \kappa)$ is independent of k and the distribution of $\{\mathbf{X}_k\}$, and tends to zero as $x_{\min} \uparrow \infty$.

Proof: See Appendix H. ■

In (405) we firstly extract $\{\mathbb{H}_\ell \mathbf{X}_\ell\}$ from $\{\mathbf{Y}_\ell\}$ using the noise $\{\mathbf{Z}_\ell\}$ and then drop $\{\mathbf{Y}_\ell, \mathbf{Z}_\ell\}$ because it is independent of $(\mathbf{X}_k, \mathbf{Y}_k)$ when conditioned on $\{\mathbb{H}_\ell \mathbf{X}_\ell\}$; in (406) we split $\mathbb{H}_k \mathbf{X}_k$ into magnitude and direction as in (231) and divide $\mathbb{H}_\ell \mathbf{X}_\ell$ by R_ℓ for $\ell = k-\kappa, \ldots, k-1$; in (407) we drop the arguments $\big\{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| R_\ell\big\}_{\ell=k+1}^{k+\kappa}$; and (408) follows from the fact that $\{R_k\}$ is IID and independent of $\{\hat{\mathbf{X}}_k\}$.

We continue with the mutual information term in (408) as follows:

$$
\begin{aligned}
&I\left(\mathbf{X}_k; \mathbf{Y}_k, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa} \middle| \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) \\
&\ = I\left(\mathbf{X}_k; \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa} \middle| \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) \\
&\quad + I\left(\mathbf{X}_k; \mathbf{Y}_k \middle| \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa}, \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) && (418)\\
&\ = I\left(\hat{\mathbf{X}}_k; \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa} \middle| \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) \\
&\quad + I\left(\hat{\mathbf{X}}_k, R_k; \mathbf{Y}_k \middle| \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa}, \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) && (419)\\
&\ = I\left(\hat{\mathbf{X}}_k; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa} \middle| \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) \\
&\quad + I\left(\hat{\mathbf{X}}_k, R_k; \mathbf{Y}_k \middle| \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa}, \right.
\end{aligned}
$$

$$\left.\hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) \quad (420)$$

$$= I\left(\hat{\mathbf{X}}_k; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) + I\left(\hat{\mathbf{X}}_k; \mathbf{Y}_k \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa}, \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) + I\left(R_k; \mathbf{Y}_k \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa}, \hat{\mathbf{X}}_{k-\kappa}^{k}\right). \quad (421)$$

Here (418) follows from the chain rule; in (419) we split $\mathbf{X}_k$ into $\hat{\mathbf{X}}_k$ and R_k and drop R_k in the first argument of the first mutual information term because it is independent of everything else; (420) follows because conditional on $\hat{\mathbf{X}}_{k-\kappa}^{k-1}$, $\hat{\mathbf{X}}_k$ and $\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}$ are independent; and (421) follows again from the chain rule.

To make our life easier we introduce the following shorthand

$$\mathbf{S}_k \triangleq \left(\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa}, \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right). \quad (422)$$

Note that $\mathbf{S}_k$ does not contain any term at time k, but only past and future terms. Then we get from (421) and (408):

$$\begin{aligned} I\big(\mathbf{X}_k; \mathbf{Y}_1^n \,\big|\, \mathbf{X}_1^{k-1}\big) &\geq I\left(\hat{\mathbf{X}}_k; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) \\ &\quad + I\big(\hat{\mathbf{X}}_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k\big) + I\big(R_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) - \delta_4(\kappa) \\ &\quad - \delta_5(x_{\min}, \kappa). \end{aligned} \quad (423)$$

To continue we introduce a new stochastic process $\{\Theta_k\}$ which is assumed to be independent of every other random quantity and IID $\sim \mathcal{U}([0, 2\pi])$. The third mutual information term in (423) can then be written as

$$\begin{aligned} I\big(R_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) &= I\big(R_k; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\ &\quad - I\big(R_k; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) + I\big(R_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \quad (424) \\ &= I\big(R_k, e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\ &\quad - I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k, R_k\big) \\ &\quad - I\big(R_k; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) + I\big(R_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \quad (425) \end{aligned}$$

where (425) follows from the chain rule. The last two terms in (425) can be rearranged as follows:

$$\begin{aligned} &-I\big(R_k; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) + I\big(R_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\ &= -h\big(\mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) + h\big(\mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k, R_k\big) \\ &\quad + h\big(\mathbf{Y}_k \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) - h\big(\mathbf{Y}_k \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k, R_k\big) \quad (426) \\ &= -h\big(\mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) + h\big(\mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k, R_k\big) \\ &\quad + h\big(\mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k, e^{\mathrm{i}\Theta_k}\big) \\ &\quad - h\big(\mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k, R_k, e^{\mathrm{i}\Theta_k}\big) \quad (427) \\ &= -I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\ &\quad + I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k, R_k\big). \quad (428) \end{aligned}$$

Here (427) follows because $e^{\mathrm{i}\Theta_k}$ is independent of everything else so that we can add it to the conditioning part of the entropy without changing its values, and because differential entropy remains unchanged if its argument is multiplied by a constant complex number of magnitude 1.

Putting this into (425) yields

$$\begin{aligned} I\big(R_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) &= I\big(R_k, e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\ &\quad - I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k, R_k\big) \\ &\quad - I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\ &\quad + I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k, R_k\big) \quad (429) \\ &= I\big(R_k, e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\ &\quad - I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \quad (430) \\ &= I\big(R_k e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\ &\quad - I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \quad (431) \end{aligned}$$

where the last equality holds because from $R_k e^{\mathrm{i}\Theta_k}$ the random variables R_k and $e^{\mathrm{i}\Theta_k}$ can be gained back.

Hence, plugging this into (423) we get the following bound:

$$\begin{aligned} I\big(\mathbf{X}_k; \mathbf{Y}_1^n \,\big|\, \mathbf{X}_1^{k-1}\big) &\geq I\left(\hat{\mathbf{X}}_k; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) \\ &\quad + I\big(\hat{\mathbf{X}}_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k\big) + I\big(R_k e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\ &\quad - I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) - \delta_4(\kappa) - \delta_5(x_{\min}, \kappa). \end{aligned} \quad (432)$$

We continue with bounding the second term on the RHS of (432):

$$\begin{aligned} I\big(\hat{\mathbf{X}}_k; \mathbf{Y}_k \,\big|\, \mathbf{S}_k\big) &= I\big(\hat{\mathbf{X}}_k; \mathbf{Y}_k, \mathbf{Z}_k \,\big|\, \mathbf{S}_k\big) - \underbrace{I\big(\hat{\mathbf{X}}_k; \mathbf{Z}_k \,\big|\, \mathbf{Y}_k, \mathbf{S}_k\big)}_{\leq \delta_6(x_{\min}, \kappa)} \quad (433) \\ &\geq I\big(\hat{\mathbf{X}}_k; \mathbf{Y}_k, \mathbf{Z}_k \,\big|\, \mathbf{S}_k\big) - \delta_6(x_{\min}, \kappa) \quad (434) \\ &= I\big(\hat{\mathbf{X}}_k; \mathbb{H}_k \mathbf{X}_k \,\big|\, \mathbf{S}_k\big) - \delta_6(x_{\min}, \kappa) \quad (435) \\ &= I\left(\hat{\mathbf{X}}_k; \frac{\mathbb{H}_k \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}, \|\mathbb{H}_k \hat{\mathbf{X}}_k\| R_k \,\middle|\, \mathbf{S}_k\right) \\ &\quad - \delta_6(x_{\min}, \kappa) \quad (436) \\ &= I\left(\hat{\mathbf{X}}_k; \frac{\mathbb{H}_k \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} \,\middle|\, \mathbf{S}_k\right) \\ &\quad + I\left(\hat{\mathbf{X}}_k; \|\mathbb{H}_k \hat{\mathbf{X}}_k\| R_k \,\middle|\, \frac{\mathbb{H} \hat{\mathbf{X}}_k}{\|\mathbb{H} \hat{\mathbf{X}}_k\|}, \mathbf{S}_k\right) \\ &\quad - \delta_6(x_{\min}, \kappa). \quad (437) \end{aligned}$$

Here (433) follows from the chain rule and (434) from the following lemma.

Lemma 21: Let $\{\mathbf{X}_k\}$ be as specified in (394)–(396). Then

$$I\big(\hat{\mathbf{X}}_k; \mathbf{Z}_k \,\big|\, \mathbf{Y}_k, \mathbf{S}_k\big) \leq \delta_6(x_{\min}, \kappa) \tag{438}$$

where $\delta_6(x_{\min}, \kappa)$ is independent of k and $\{\mathbf{X}_k\}$, and tends to zero as $x_{\min} \uparrow \infty$.

Proof: See Appendix I. ∎

In (435) we use $\mathbf{Z}_k$ in order to extract $\mathbb{H}\mathbf{X}_k$ from $\mathbf{Y}_k$ and then drop $\mathbf{Y}_k$ and $\mathbf{Z}_k$ since given $\mathbb{H}\mathbf{X}_k$ they are independent of the other random variables; in (436) we split $\mathbb{H}\mathbf{X}_k$ up into magnitude and direction as in (231); and (437) follows again from the chain rule.

Next we bound the fourth term on the RHS of (432):

$$\begin{aligned}
&I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\
&\quad \leq I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k}, \mathbf{Z}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) && (439)\\
&\quad = I\big(e^{\mathrm{i}\Theta_k}; \mathbb{H}_k\mathbf{X}_k e^{\mathrm{i}\Theta_k}, \mathbf{Z}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) && (440)\\
&\quad = I\big(e^{\mathrm{i}\Theta_k}; \mathbb{H}_k\mathbf{X}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\
&\qquad + I\big(e^{\mathrm{i}\Theta_k}; \mathbf{Z}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbb{H}_k\mathbf{X}_k e^{\mathrm{i}\Theta_k}, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) && (441)\\
&\quad = I\big(e^{\mathrm{i}\Theta_k}; \mathbb{H}_k\mathbf{X}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) && (442)\\
&\quad = I\left(e^{\mathrm{i}\Theta_k}; \|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k} \,\middle|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\right) && (443)\\
&\quad = I\left(e^{\mathrm{i}\Theta_k}; \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k} \,\middle|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\right) \\
&\qquad + I\left(e^{\mathrm{i}\Theta_k}; \|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}, \mathbf{S}_k, \hat{\mathbf{X}}_k\right). && (444)
\end{aligned}$$

Here, (439) follows from adding a term to mutual information; in (440) we extract $\mathbb{H}_k\mathbf{X}_k e^{\mathrm{i}\Theta_k}$ from $\mathbf{Y}_k e^{\mathrm{i}\Theta_k}$; then in (441) we use the chain rule; (442) holds because the noise $\mathbf{Z}_k$ is circularly symmetric and independent of all other random quantities so that the phase $e^{\mathrm{i}\Theta_k}$ is destroyed; in (443) we split $\mathbb{H}_k\mathbf{X}_k e^{\mathrm{i}\Theta_k}$ up into magnitude and direction similar to (231); and (444) follows again from the chain rule.

We plug (444) and (437) into (432) and get:

$$\begin{aligned}
&I\big(\mathbf{X}_k; \mathbf{Y}_1^n \,\big|\, \mathbf{X}_1^{k-1}\big) \\
&\quad \geq I\left(\hat{\mathbf{X}}_k; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}, \hat{\mathbf{X}}_{k-\kappa}^{k-1}\right) \\
&\qquad + I\left(\hat{\mathbf{X}}_k; \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|} \,\middle|\, \mathbf{S}_k\right) \\
&\qquad + I\left(\hat{\mathbf{X}}_k; \|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}\hat{\mathbf{X}}_k}{\|\mathbb{H}\hat{\mathbf{X}}_k\|}, \mathbf{S}_k\right) \\
&\qquad + I\big(R_k e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \,\big|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\big) \\
&\qquad - I\left(e^{\mathrm{i}\Theta_k}; \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k} \,\middle|\, \mathbf{S}_k, \hat{\mathbf{X}}_k\right) \\
&\qquad - I\left(e^{\mathrm{i}\Theta_k}; \|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}, \mathbf{S}_k, \hat{\mathbf{X}}_k\right) \\
&\qquad - \delta_4(\kappa) - \delta_5(x_{\min}, \kappa) - \delta_6(x_{\min}, \kappa). && (445)
\end{aligned}$$

We continue by bounding the third and the sixth term on the RHS of (445):

$$\begin{aligned}
&I\left(\hat{\mathbf{X}}_k; \|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \mathbf{S}_k\right) \\
&- I\left(e^{\mathrm{i}\Theta_k}; \|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}, \mathbf{S}_k, \hat{\mathbf{X}}_k\right) \\
&\quad = h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \mathbf{S}_k\right) \\
&\qquad - h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \mathbf{S}_k, \hat{\mathbf{X}}_k\right) \\
&\qquad - h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}, \mathbf{S}_k, \hat{\mathbf{X}}_k\right) \\
&\qquad + h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}, \mathbf{S}_k, \hat{\mathbf{X}}_k, e^{\mathrm{i}\Theta_k}\right) && (446)\\
&\quad = h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \mathbf{S}_k\right) \\
&\qquad - h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \mathbf{S}_k, \hat{\mathbf{X}}_k\right) \\
&\qquad - h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}, \mathbf{S}_k, \hat{\mathbf{X}}_k\right) \\
&\qquad + h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \mathbf{S}_k, \hat{\mathbf{X}}_k\right) && (447)\\
&\quad = h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \mathbf{S}_k\right) \\
&\qquad - h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}, \mathbf{S}_k, \hat{\mathbf{X}}_k\right) && (448)\\
&\quad \geq h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \mathbf{S}_k\right) \\
&\qquad - h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}, \mathbf{S}_k\right) && (449)\\
&\quad = h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \mathbf{S}_k\right) \\
&\qquad - h\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|R_k \,\middle|\, \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \mathbf{S}_k\right) && (450)\\
&\quad = 0. && (451)
\end{aligned}$$

Here, (447) follows because $e^{\mathrm{i}\Theta_k}$ is independent of all other random quantities; (449) follows from conditioning that reduces entropy; and (450) holds because we have assumed $\hat{\mathbf{X}}_k$ to be circularly symmetric. Note that since we have dropped the conditioning on $\hat{\mathbf{X}}_k$ in the second differential entropy term, we cannot recover the phase of $\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}$ and that therefore the uniform phase of $\hat{\mathbf{X}}_k$ "destroys" the phase of $e^{\mathrm{i}\Theta_k}$; see also Definition 5 and Remark 6.

Next we bound the second and fifth term on the RHS of (445):

$$\begin{aligned}
I&\left(\hat{\mathbf{X}}_k; \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\,\middle|\,\mathbf{S}_k\right) - I\left(e^{\mathrm{i}\Theta_k}; \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}\,\middle|\,\mathbf{S}_k,\hat{\mathbf{X}}_k\right)\\
&= h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\,\middle|\,\mathbf{S}_k\right) - h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\,\middle|\,\mathbf{S}_k,\hat{\mathbf{X}}_k\right)\\
&\quad - h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}\,\middle|\,\mathbf{S}_k,\hat{\mathbf{X}}_k\right)\\
&\quad + h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}\,\middle|\,\mathbf{S}_k,\hat{\mathbf{X}}_k,e^{\mathrm{i}\Theta_k}\right) \qquad (452)\\
&= h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\,\middle|\,\mathbf{S}_k\right) - h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\,\middle|\,\mathbf{S}_k,\hat{\mathbf{X}}_k\right)\\
&\quad - h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}\,\middle|\,\mathbf{S}_k,\hat{\mathbf{X}}_k\right)\\
&\quad + h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\,\middle|\,\mathbf{S}_k,\hat{\mathbf{X}}_k\right) \qquad (453)\\
&= h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\,\middle|\,\mathbf{S}_k\right) - h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}\,\middle|\,\mathbf{S}_k,\hat{\mathbf{X}}_k\right) \qquad (454)
\end{aligned}$$

where (453) follows from (19) with a choice $\mathsf{U} = e^{-\mathrm{i}\theta_k}\mathsf{I}_{n_\mathrm{R}}$ and from the fact that $e^{\mathrm{i}\Theta_k}$ is independent of all other random quantities.

Plugging (454) and (451) into (445) leaves us with the following bound:

$$\begin{aligned}
I&\big(\mathbf{X}_k;\mathbf{Y}_1^n\,\big|\,\mathbf{X}_1^{k-1}\big)\\
&\ge I\left(\hat{\mathbf{X}}_k;\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa}\,\middle|\,\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1},\hat{\mathbf{X}}_{k-\kappa}^{k-1}\right)\\
&\quad + I\big(R_k e^{\mathrm{i}\Theta_k};\mathbf{Y}_k e^{\mathrm{i}\Theta_k}\,\big|\,\mathbf{S}_k,\hat{\mathbf{X}}_k\big)\\
&\quad + h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\,\middle|\,\mathbf{S}_k\right) - h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}\,\middle|\,\mathbf{S}_k,\hat{\mathbf{X}}_k\right)\\
&\quad - \delta_4(\kappa) - \delta_5(x_{\min},\kappa) - \delta_6(x_{\min},\kappa). \qquad (455)
\end{aligned}$$

Hence, using this in (398) we get

$$\begin{aligned}
&\frac{1}{n}I\big(\mathbf{X}_1^n;\mathbf{Y}_1^n\big)\\
&\ge \frac{1}{n}\sum_{k=\kappa+1}^{n-\kappa}\Bigg[I\big(R_k e^{\mathrm{i}\Theta_k};\mathbb{H}_k\hat{\mathbf{X}}_k R_k e^{\mathrm{i}\Theta_k} + \mathbf{Z}_k\,\big|\,\mathbf{S}_k,\hat{\mathbf{X}}_k\big)\\
&\qquad + I\left(\hat{\mathbf{X}}_k;\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=k+1}^{k+\kappa}\,\middle|\,\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1},\hat{\mathbf{X}}_{k-\kappa}^{k-1}\right)\\
&\qquad + h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\,\middle|\,\mathbf{S}_k\right)\\
&\qquad - h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}e^{\mathrm{i}\Theta_k}\,\middle|\,\mathbf{S}_k,\hat{\mathbf{X}}_k\right)\\
&\qquad - \delta_4(\kappa) - \delta_5(x_{\min},\kappa) - \delta_6(x_{\min},\kappa)\Bigg] \qquad (456)\\
&= \frac{1}{n}\sum_{k=\kappa+1}^{n-\kappa}\Bigg[I\big(R_0 e^{\mathrm{i}\Theta_0};\mathbb{H}_0\hat{\mathbf{X}}_0 R_0 e^{\mathrm{i}\Theta_0} + \mathbf{Z}_0\,\big|\,\mathbf{S}_0,\hat{\mathbf{X}}_0\big)\\
&\qquad + I\left(\hat{\mathbf{X}}_0;\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\,\middle|\,\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1},\hat{\mathbf{X}}_{-\kappa}^{-1}\right)\\
&\qquad + h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\,\mathbf{S}_0\right)\\
&\qquad - h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}e^{\mathrm{i}\Theta_0}\,\middle|\,\mathbf{S}_0,\hat{\mathbf{X}}_0\right)\\
&\qquad - \delta_4(\kappa) - \delta_5(x_{\min},\kappa) - \delta_6(x_{\min},\kappa)\Bigg] \qquad (457)\\
&= \frac{n-2\kappa}{n}\Bigg[I\big(\tilde{X}_0;\mathbb{H}_0\hat{\mathbf{X}}_0\tilde{X}_0 + \mathbf{Z}_0\,\big|\,\mathbf{S}_0,\hat{\mathbf{X}}_0\big)\\
&\qquad + I\left(\hat{\mathbf{X}}_0;\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\,\middle|\,\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1},\hat{\mathbf{X}}_{-\kappa}^{-1}\right)\\
&\qquad + h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\,\mathbf{S}_0\right)\\
&\qquad - h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}e^{\mathrm{i}\Theta_0}\,\middle|\,\mathbf{S}_0,\hat{\mathbf{X}}_0\right)\\
&\qquad - \delta_4(\kappa) - \delta_5(x_{\min},\kappa) - \delta_6(x_{\min},\kappa)\Bigg] \qquad (458)
\end{aligned}$$

where (457) follows from stationarity, and (458) from defining a new complex random variable $\tilde{X}_0 \triangleq R_0 e^{\mathrm{i}\Theta_0}$. Note that we have changed our shorthand (422) accordingly to

$$\mathbf{S}_0 \triangleq \left(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right). \qquad (459)$$

Letting n tend to infinity we obtain

$$\begin{aligned}
\mathcal{C}(\mathcal{E}) &\ge I\left(\hat{\mathbf{X}}_0;\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\,\middle|\,\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1},\hat{\mathbf{X}}_{-\kappa}^{-1}\right)\\
&\quad + I\big(\tilde{X}_0;\mathbb{H}_0\hat{\mathbf{X}}_0\tilde{X}_0 + \mathbf{Z}_0\,\big|\,\mathbf{S}_0,\hat{\mathbf{X}}_0\big)\\
&\quad + h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\,\mathbf{S}_0\right)\\
&\quad - h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}e^{\mathrm{i}\Theta_0}\,\middle|\,\mathbf{S}_0,\hat{\mathbf{X}}_0\right)\\
&\quad - \delta_4(\kappa) - \delta_5(x_{\min},\kappa) - \delta_6(x_{\min},\kappa) \qquad (460)
\end{aligned}$$

where the second term on the RHS can be viewed as mutual information across a memoryless SIMO fading channel with fading vector $\mathbf{H} = \mathbb{H}_0\hat{\mathbf{X}}_0$ in the presence of the side-information $(\mathbf{S}_0,\hat{\mathbf{X}}_0)$.

We next let the power grow to infinity $\mathcal{E}\uparrow\infty$ and use the definition of the fading number. Note that the distribution of $\tilde{X}_0$ (the product of (395) with the circularly symmetric law from $e^{\mathrm{i}\Theta_0}$) achieves the fading number of IID SIMO fading with side-information [7, Proposition 4.23], [9, Proposition 6.23]. Moreover, our choice (396) guarantees that $\delta_5(x_{\min},\kappa)$ and $\delta_6(x_{\min},\kappa)$ tend to zero as $\mathcal{E}\uparrow\infty$ (see

Appendices H and I). Therefore, we obtain the following bound:

$$
\begin{aligned}
&\chi(\{\mathbb{H}_k\}) \\
&= \varlimsup_{\mathcal{E}\uparrow\infty} \left\{ \mathcal{C}(\mathcal{E}) - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} &(461)\\
&\geq I\left(\hat{\mathbf{X}}_0; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right) \\
&\quad + \varlimsup_{\mathcal{E}\uparrow\infty} \left\{ I\big(\tilde{X}_0; \mathbb{H}_0 \hat{\mathbf{X}}_0 \tilde{X}_0 + \mathbf{Z}_0 \,\big|\, \mathbf{S}_0, \hat{\mathbf{X}}_0\big) \right. \\
&\qquad\qquad \left. -\ \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} \\
&\quad + h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \,\middle|\, \mathbf{S}_0\right) - h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} e^{\mathrm{i}\Theta_0} \,\middle|\, \mathbf{S}_0, \hat{\mathbf{X}}_0\right) \\
&\quad - \delta_4(\kappa) &(462)\\
&= I\left(\hat{\mathbf{X}}_0; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right) \\
&\quad + \chi_{\mathrm{IID}}\big(\mathbb{H}_0 \hat{\mathbf{X}}_0 \,\big|\, \mathbf{S}_0, \hat{\mathbf{X}}_0\big) + h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \,\middle|\, \mathbf{S}_0\right) \\
&\quad - h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} e^{\mathrm{i}\Theta_0} \,\middle|\, \mathbf{S}_0, \hat{\mathbf{X}}_0\right) - \delta_4(\kappa) &(463)\\
&= I\left(\hat{\mathbf{X}}_0; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right) \\
&\quad + h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} e^{\mathrm{i}\Theta_0} \,\middle|\, \mathbf{S}_0, \hat{\mathbf{X}}_0\right) + n_{\mathsf{R}} \mathsf{E}\left[\log \|\mathbb{H}_0 \hat{\mathbf{X}}_0\|^2\right] \\
&\quad -\ \log 2 - h\big(\mathbb{H}_0 \hat{\mathbf{X}}_0 \,\big|\, \mathbf{S}_0, \hat{\mathbf{X}}_0\big) + h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \,\middle|\, \mathbf{S}_0\right) \\
&\quad - h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} e^{\mathrm{i}\Theta_0} \,\middle|\, \mathbf{S}_0, \hat{\mathbf{X}}_0\right) - \delta_4(\kappa) &(464)\\
&= I\left(\hat{\mathbf{X}}_0; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right) \\
&\quad + h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \,\middle|\, \mathbf{S}_0\right) + n_{\mathsf{R}} \mathsf{E}\left[\log \|\mathbb{H}_0 \hat{\mathbf{X}}_0\|^2\right] \\
&\quad -\ \log 2 - h\big(\mathbb{H}_0 \hat{\mathbf{X}}_0 \,\big|\, \mathbf{S}_0, \hat{\mathbf{X}}_0\big) - \delta_4(\kappa) &(465)\\
&= I\left(\hat{\mathbf{X}}_0; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right) \\
&\quad + h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right) \\
&\quad + n_{\mathsf{R}} \mathsf{E}\left[\log \|\mathbb{H}_0 \hat{\mathbf{X}}_0\|^2\right] - \log 2 \\
&\quad - h\left(\mathbb{H}_0 \hat{\mathbf{X}}_0 \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}, \hat{\mathbf{X}}_{-\kappa}^{0}\right) \\
&\quad - \delta_4(\kappa). &(466)
\end{aligned}
$$

Here, (461) follows by definition (see (37)); in (462) we use (460) and note that apart from two terms all other terms do not depend on $\mathcal{E}$; (463) follows because $\tilde{X}_0$ achieves the memoryless SIMO fading number with side information; in (464) we then replace this fading number by its expression (7); and in (466) we have expanded the shorthand $\mathbf{S}_0$.

Notice that from (466) the alternative expression (91) can be derived.

Finally, we need to show that (466) is identical to the upper bound that we have derived above. To that goal note that

$$
\begin{aligned}
&h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right) \\
&+ I\left(\hat{\mathbf{X}}_0; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right) \\
&= h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right) \\
&\quad - I\left(\hat{\mathbf{X}}_{-\kappa}^{-1}; \frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right) \\
&\quad + I\left(\hat{\mathbf{X}}_0; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right) &(467)\\
&= h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right) \\
&\quad - h_\lambda\left(\hat{\mathbf{X}}_{-\kappa}^{-1} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right) \\
&\quad + h_\lambda\left(\hat{\mathbf{X}}_{-\kappa}^{-1} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=0}^{\kappa}\right) \\
&\quad + h_\lambda\big(\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}\big) \\
&\quad - h_\lambda\left(\hat{\mathbf{X}}_0 \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right) &(468)\\
&= h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right) \\
&\quad - h_\lambda\left(\hat{\mathbf{X}}_{-\kappa}^{0} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right) \\
&\quad + h_\lambda\left(\hat{\mathbf{X}}_{-\kappa}^{-1} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=0}^{\kappa}\right) \\
&\quad + h_\lambda\big(\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}\big) &(469)\\
&= h_\lambda\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right) \\
&\quad + I\left(\hat{\mathbf{X}}_{-\kappa}^{0}; \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\right) \\
&\quad - h_\lambda\big(\hat{\mathbf{X}}_{-\kappa}^{0} \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\big) \\
&\quad + h_\lambda\left(\hat{\mathbf{X}}_{-\kappa}^{-1} \,\middle|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|}\right\}_{\ell=0}^{\kappa}\right) \\
&\quad + h_\lambda\big(\hat{\mathbf{X}}_{-\kappa}^{0} \,\big|\, \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\big)
\end{aligned}
$$

$$
\begin{aligned}
&- h_\lambda\big(\hat{\mathbf{X}}_{-\kappa}^{-1} \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\big) &(470)\\
&= h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right)\\
&\quad + I\left(\hat{\mathbf{X}}_{-\kappa}^{0}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\right)\\
&\quad - I\left(\hat{\mathbf{X}}_{-\kappa}^{-1}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=0}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\right) &(471)\\
&= h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right)\\
&\quad + h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\right)\\
&\quad - h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\right)\\
&\quad - h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=0}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\right)\\
&\quad + h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=0}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right) &(472)\\
&= -h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\right)\\
&\quad + h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=0}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}\right) &(473)\\
&= -h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right) + h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=0}^{\kappa}\right)\\
&\quad + I\left(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right)\\
&\quad - I\left(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=0}^{\kappa}\right) &(474)\\
&= h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\right)\\
&\quad + I\left(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right)\\
&\quad - I\left(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=0}^{\kappa}\right). &(475)
\end{aligned}
$$

Here, (467) and (468) follow from the definition of mutual information; for (469) we combine the second and fifth term of (468) using the chain rule; (470) follows from the definition of mutual information applied to the second term in (469) and the chain rule applied to the last term; then in (471) and (472) we use the definition of mutual information again; (473) follows from the chain rule which shows that the first, the second, and the fourth term in (472) cancel each other; in (474) we use the definition of mutual information twice; and (475) follows from stationarity and from the chain rule.

Next, using the definition of mutual information several times more, we get the following chain of equalities:

$$
\begin{aligned}
&-h\left(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}, \hat{\mathbf{X}}_{-\kappa}^{0}\right)\\
&\quad = -h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\big)\\
&\qquad + I\left(\mathbb{H}_0\hat{\mathbf{X}}_0; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\right) &(476)\\
&\quad = -h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\big)\\
&\qquad + h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\right)\\
&\qquad - h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0}, \hat{\mathbf{X}}_{-\kappa}^{0}\right) &(477)\\
&\quad = -h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\big)\\
&\qquad + h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}\right)\\
&\qquad - I\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\right)\\
&\qquad - h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa} \,\middle|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0}, \hat{\mathbf{X}}_{-\kappa}^{0}\right) &(478)\\
&\quad = -h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\big)\\
&\qquad + I\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0}, \hat{\mathbf{X}}_{-\kappa}^{0}\right)\\
&\qquad - I\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\right). &(479)
\end{aligned}
$$

Hence, plugging (479) and (475) into (466) we have

$$
\begin{aligned}
\chi(\{\mathbb{H}_k\}) \geq{}& h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|} \,\middle|\, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\kappa}^{-1}\right)\\
&+ n_{\mathsf{R}}\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\Big] - \log 2\\
&- h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{0}\big)\\
&+ I\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\kappa}; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{0}, \hat{\mathbf{X}}_{-\kappa}^{0}\right)\\
&- I\left(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}_{-\kappa}^{-1}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=0}^{\kappa}\right)\\
&- \delta_4(\kappa). \qquad (480)
\end{aligned}
$$

Finally, we let κ go to infinity. The result now follows from the fact that $\delta_4(\kappa)$ tends to zero as $\kappa \uparrow \infty$ (see (416)), from stationarity of the input and the fading process which makes sure that the two mutual information terms cancel, and by choosing the distribution of $\{\hat{\mathbf{X}}_k\}$ such as to maximize this lower bound under the constraints that the distribution needs to be stationary and circularly symmetric.[7]

APPENDIX H
PROOF OF LEMMA 20

We derive the following bound:

$$\begin{aligned}
&I\big(\mathbf{X}_k;\mathbf{Z}_{k-\kappa}^{k-1},\mathbf{Z}_{k+1}^{k+\kappa}\,\big|\,\mathbf{Y}_{k-\kappa}^{k+\kappa},\mathbf{X}_{k-\kappa}^{k-1}\big)\\
&\quad\le I\big(\mathbf{X}_k;\mathbf{Z}_{k-\kappa}^{k+\kappa}\,\big|\,\mathbf{Y}_{k-\kappa}^{k+\kappa},\mathbf{X}_{k-\kappa}^{k-1}\big) && (481)\\
&\quad= h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\,\big|\,\mathbf{Y}_{k-\kappa}^{k+\kappa},\mathbf{X}_{k-\kappa}^{k-1}\big)\\
&\qquad - h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\,\big|\,\mathbf{Y}_{k-\kappa}^{k+\kappa},\mathbf{X}_{k-\kappa}^{k}\big) && (482)\\
&\quad\le h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\big) - h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\,\big|\,\mathbf{Y}_{k-\kappa}^{k+\kappa},\mathbf{X}_{k-\kappa}^{k+\kappa}\big) && (483)\\
&\quad= h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\big) - h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\,\big|\,\mathbf{Y}_{k-\kappa}^{k+\kappa},\hat{\mathbf{X}}_{k-\kappa}^{k+\kappa},R_{k-\kappa}^{k+\kappa}\big) && (484)\\
&\quad\le h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\big)\\
&\qquad - \inf_{\substack{r_{k-\kappa}\ge x_{\min},\ldots,\\ r_{k+\kappa}\ge x_{\min}}} h\Big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\,\Big|\,\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\cdot r_\ell+\mathbf{Z}_\ell\}_{\ell=k-\kappa}^{k+\kappa},\hat{\mathbf{X}}_{k-\kappa}^{k+\kappa}\Big) && (485)\\
&\quad= h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\big)\\
&\qquad - h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\,\big|\,\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\cdot x_{\min}+\mathbf{Z}_\ell\}_{\ell=k-\kappa}^{k+\kappa},\hat{\mathbf{X}}_{k-\kappa}^{k+\kappa}\big) && (486)\\
&\quad= I\big(\mathbf{Z}_{k-\kappa}^{k+\kappa};\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\cdot x_{\min}+\mathbf{Z}_\ell\}_{\ell=k-\kappa}^{k+\kappa},\hat{\mathbf{X}}_{k-\kappa}^{k+\kappa}\big) && (487)\\
&\quad= I\left(\left\{\frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=k-\kappa}^{k+\kappa};\left\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell+\frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=k-\kappa}^{k+\kappa}\,\middle|\,\hat{\mathbf{X}}_{k-\kappa}^{k+\kappa}\right) && (488)\\
&\quad\le \sup_{\substack{\hat{\mathbf{x}}_{k-\kappa},\ldots,\\ \hat{\mathbf{x}}_{k+\kappa}}} I\left(\left\{\frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=k-\kappa}^{k+\kappa};\left\{\mathbb{H}_\ell\hat{\mathbf{x}}_\ell+\frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=k-\kappa}^{k+\kappa}\right) && (489)\\
&\quad= \sup_{\substack{\hat{\mathbf{e}}_{-\kappa},\ldots,\\ \hat{\mathbf{e}}_{\kappa}}} I\left(\left\{\frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa};\left\{\mathbb{H}_\ell\hat{\mathbf{e}}_\ell+\frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa}\right) && (490)\\
&\quad= \sup_{\substack{\hat{\mathbf{e}}_{-\kappa},\ldots,\\ \hat{\mathbf{e}}_{\kappa}}} I\left(\left\{\frac{\sigma\mathbb{Z}_\ell\hat{\mathbf{e}}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa};\left\{\mathbb{H}_\ell\hat{\mathbf{e}}_\ell+\frac{\sigma\mathbb{Z}_\ell\hat{\mathbf{e}}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa}\right) && (491)\\
&\quad= \sup_{\substack{\hat{\mathbf{e}}_{-\kappa},\ldots,\\ \hat{\mathbf{e}}_{\kappa}}} I\left(\left\{\frac{\sigma\mathbb{Z}_\ell\hat{\mathbf{e}}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa};\left\{\left(\mathbb{H}_\ell+\frac{\sigma\mathbb{Z}_\ell}{x_{\min}}\right)\hat{\mathbf{e}}_\ell\right\}_{\ell=-\kappa}^{\kappa}\right) && (492)\\
&\quad\le I\left(\left\{\frac{\sigma\mathbb{Z}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa};\left\{\mathbb{H}_\ell+\frac{\sigma\mathbb{Z}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa}\right) && (493)\\
&\quad= h\left(\left\{\mathbb{H}_\ell+\frac{\sigma\mathbb{Z}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa}\right) - h\big(\mathbb{H}_{-\kappa}^{\kappa}\big) && (494)\\
&\quad\triangleq \delta_5(x_{\min},\kappa). && (495)
\end{aligned}$$

Here (481) follows from adding $\mathbf{Z}_k$ to the arguments of the mutual information; (483) follows from conditioning that reduces entropy; in (484) we split $\mathbf{X}_\ell$ into direction and magnitude; in (485) we replace the expectation over $R_{k-\kappa}^{k+\kappa}$ by a corresponding minimization; for (486) we note that the minimum is achieved for the smallest value of R_ℓ which is $x_{\min}$; (488) follows because $\{\mathbf{Z}_k\}$ is independent of $\{\hat{\mathbf{X}}_k\}$; in (489) we again replace an expectation by an optimization; (490) follows from stationarity; in (491) we introduce a random $n_{\mathrm{R}}\times n_{\mathrm{T}}$ matrix process $\{\mathbb{Z}_k\}$ that is IID with components that are independent $\mathcal{N}_{\mathbb{C}}(0,1)$ so that $\sigma\mathbb{Z}_\ell\hat{\mathbf{e}}_\ell\sim\mathcal{N}_{\mathbb{C}}\big(0,\sigma^2\mathsf{I}_{n_{\mathrm{R}}}\big)$; and (493) follows from the data processing inequality.

From [7, Lemma 6.11], [9, Lemma A.19] we conclude that the expression

$$h\left(\left\{\mathbb{H}_\ell+\frac{\sigma\mathbb{Z}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa}\right) \tag{496}$$

converges monotonically in $x_{\min}$ to $h\big(\mathbb{H}_{-\kappa}^{\kappa}\big)$.

APPENDIX I
PROOF OF LEMMA 21

This bound is derived similarly to the derivation of Lemma 20 in the previous appendix.

$$\begin{aligned}
&I\big(\hat{\mathbf{X}}_k;\mathbf{Z}_k\,\big|\,\mathbf{Y}_k,\mathbf{S}_k\big)\\
&\quad= h\big(\mathbf{Z}_k\,\big|\,\mathbf{Y}_k,\mathbf{S}_k\big) - h\big(\mathbf{Z}_k\,\big|\,\mathbf{Y}_k,\mathbf{S}_k,\hat{\mathbf{X}}_k\big) && (497)\\
&\quad\le h(\mathbf{Z}_k) - h\big(\mathbf{Z}_k\,\big|\,\mathbf{Y}_k,\mathbf{S}_k,\mathbb{H}_{k-\kappa}^{k-1},\mathbb{H}_{k+1}^{k+\kappa},R_k,\hat{\mathbf{X}}_k\big) && (498)\\
&\quad= h(\mathbf{Z}_k)\\
&\qquad - h\big(\mathbf{Z}_k\,\big|\,\mathbb{H}_k\hat{\mathbf{X}}_kR_k+\mathbf{Z}_k,R_k,\hat{\mathbf{X}}_k,\mathbb{H}_{k-\kappa}^{k-1},\mathbb{H}_{k+1}^{k+\kappa}\big) && (499)\\
&\quad\le h(\mathbf{Z}_k)\\
&\qquad - \inf_{r_k\ge x_{\min}} h\big(\mathbf{Z}_k\,\big|\,\mathbb{H}_k\hat{\mathbf{X}}_kr_k+\mathbf{Z}_k,\hat{\mathbf{X}}_k,\mathbb{H}_{k-\kappa}^{k-1},\mathbb{H}_{k+1}^{k+\kappa}\big) && (500)\\
&\quad= h(\mathbf{Z}_k)\\
&\qquad - h\big(\mathbf{Z}_k\,\big|\,\mathbb{H}_k\hat{\mathbf{X}}_kx_{\min}+\mathbf{Z}_k,\hat{\mathbf{X}}_k,\mathbb{H}_{k-\kappa}^{k-1},\mathbb{H}_{k+1}^{k+\kappa}\big) && (501)\\
&\quad= I\big(\mathbf{Z}_k;\mathbb{H}_k\hat{\mathbf{X}}_kx_{\min}+\mathbf{Z}_k,\hat{\mathbf{X}}_k,\mathbb{H}_{k-\kappa}^{k-1},\mathbb{H}_{k+1}^{k+\kappa}\big) && (502)\\
&\quad= I\left(\frac{\mathbf{Z}_k}{x_{\min}};\mathbb{H}_k\hat{\mathbf{X}}_k+\frac{\mathbf{Z}_k}{x_{\min}}\,\middle|\,\hat{\mathbf{X}}_k,\mathbb{H}_{k-\kappa}^{k-1},\mathbb{H}_{k+1}^{k+\kappa}\right) && (503)\\
&\quad\le \sup_{\hat{\mathbf{x}}_k} I\left(\frac{\mathbf{Z}_k}{x_{\min}};\mathbb{H}_k\hat{\mathbf{x}}_k+\frac{\mathbf{Z}_k}{x_{\min}}\,\middle|\,\mathbb{H}_{k-\kappa}^{k-1},\mathbb{H}_{k+1}^{k+\kappa}\right) && (504)\\
&\quad= \sup_{\hat{\mathbf{e}}} I\left(\frac{\mathbf{Z}_0}{x_{\min}};\mathbb{H}_0\hat{\mathbf{e}}+\frac{\mathbf{Z}_0}{x_{\min}}\,\middle|\,\mathbb{H}_{-\kappa}^{-1},\mathbb{H}_1^{\kappa}\right) && (505)\\
&\quad= \sup_{\hat{\mathbf{e}}} I\left(\frac{\sigma\mathbb{Z}_0\hat{\mathbf{e}}}{x_{\min}};\mathbb{H}_0\hat{\mathbf{e}}+\frac{\sigma\mathbb{Z}_0\hat{\mathbf{e}}}{x_{\min}}\,\middle|\,\mathbb{H}_{-\kappa}^{-1},\mathbb{H}_1^{\kappa}\right) && (506)\\
&\quad= \sup_{\hat{\mathbf{e}}} I\left(\frac{\sigma\mathbb{Z}_0\hat{\mathbf{e}}}{x_{\min}};\left(\mathbb{H}_0+\frac{\sigma\mathbb{Z}_0}{x_{\min}}\right)\hat{\mathbf{e}}\,\middle|\,\mathbb{H}_{-\kappa}^{-1},\mathbb{H}_1^{\kappa}\right) && (507)\\
&\quad\le I\left(\frac{\sigma\mathbb{Z}_0}{x_{\min}};\mathbb{H}_0+\frac{\sigma\mathbb{Z}_0}{x_{\min}}\,\middle|\,\mathbb{H}_{-\kappa}^{-1},\mathbb{H}_1^{\kappa}\right) && (508)\\
&\quad= h\left(\mathbb{H}_0+\frac{\sigma\mathbb{Z}_0}{x_{\min}}\,\middle|\,\mathbb{H}_{-\kappa}^{-1},\mathbb{H}_1^{\kappa}\right) - h\big(\mathbb{H}_0\,\big|\,\mathbb{H}_{-\kappa}^{-1},\mathbb{H}_1^{\kappa}\big) && (509)\\
&\quad\triangleq \delta_6(x_{\min},\kappa). && (510)
\end{aligned}$$

Here, (498) follows again from conditioning that reduces entropy; (499) from the fact that conditional on $(\mathbb{H}_{k-\kappa}^{k-1},\mathbb{H}_{k+1}^{k+\kappa},\mathbf{X}_k)$ the pair $(\mathbf{Z}_k,\mathbf{Y}_k)$ is independent of $\mathbf{S}_k$; and the remaining steps (500)–(510) are identical to (485)–(495). Note that we introduce a random $n_{\mathrm{R}}\times n_{\mathrm{T}}$ matrix

[7]Note that originally we have *chosen* these constraints, *i.e.*, we do not necessarily have to use such inputs. However, during the derivation we have relied on this choice several times, so that at this stage we cannot change it anymore.

$\mathbb{Z}_0$ with components that are independent $\mathcal{N}_{\mathbb{C}}(0,1)$ so that $\sigma\mathbb{Z}_0\hat{\mathbf{e}} \sim \mathcal{N}_{\mathbb{C}}\big(0,\sigma^2\mathsf{I}_{n_{\mathrm{R}}}\big)$.

From [7, Lemma 6.11], [9, Lemma A.19] we conclude that for every realization of $(\mathbb{H}_{-\kappa}^{-1},\mathbb{H}_1^{\kappa})$ the expression

$$h\left(\mathbb{H}_0+\frac{\sigma\mathbb{Z}_0}{x_{\min}}\,\middle|\,\{\mathbb{H}_\ell=\mathsf{H}_\ell\}_{\ell=-\kappa}^{-1},\{\mathbb{H}_\ell=\mathsf{H}_\ell\}_{\ell=1}^{\kappa}\right) \tag{511}$$

converges monotonically in $x_{\min}$ to $h\big(\mathbb{H}_0\,\big|\,\{\mathbb{H}_\ell=\mathsf{H}_\ell\}_{\ell=-\kappa}^{-1},\{\mathbb{H}_\ell=\mathsf{H}_\ell\}_{\ell=1}^{\kappa}\big)$. By the Monotone Convergence Theorem (MCT) [21] this is also true when we average over $(\mathbb{H}_{-\kappa}^{-1},\mathbb{H}_1^{\kappa})$.

Appendix J
Proof of Corollary 14

We first need an extension of Lemma 2.

Lemma 22: Let $\mathbf{V}_1,\ldots,\mathbf{V}_\nu$ be ν complex random vectors taking value in $\mathbb{C}^m$ and having differential entropy $h(\mathbf{V}_1,\ldots,\mathbf{V}_\nu)$. Let $\|\mathbf{V}_k\|$ denote the norm of $\mathbf{V}_k$, and $\hat{\mathbf{V}}_k$ its direction

$$\hat{\mathbf{V}}_k \triangleq \frac{\mathbf{V}_k}{\|\mathbf{V}_k\|}. \tag{512}$$

Then

$$\begin{aligned}
h\big(\mathbf{V}_1^\nu\big) &= h_\lambda\big(\hat{\mathbf{V}}_1^\nu\big)+h\Big(\big\{\|\mathbf{V}_\ell\|\big\}_{\ell=1}^{\nu}\,\Big|\,\hat{\mathbf{V}}_1^\nu\Big)\\
&\quad+(2m-1)\sum_{k=1}^{\nu}\mathsf{E}[\log\|\mathbf{V}_k\|] &(513)\\
&= h\Big(\big\{\|\mathbf{V}_\ell\|\big\}_{\ell=1}^{\nu}\Big)+h_\lambda\Big(\hat{\mathbf{V}}_1^\nu\,\Big|\,\big\{\|\mathbf{V}_\ell\|\big\}_{\ell=1}^{\nu}\Big)\\
&\quad+(2m-1)\sum_{k=1}^{\nu}\mathsf{E}[\log\|\mathbf{V}_k\|] &(514)
\end{aligned}$$

whenever all the quantities in (513) and (514), respectively, are defined.

Proof: The proof is based on a repetitive use of the chain rule and a conditional version of Lemma 2. The details are omitted. ■

The equivalence of (89) and (90) can be proven based on (299). Similarly, we rely on (466) to prove that (89) and (91) are equivalent. The details are omitted.

In the following we will prove the equivalence of (90), (92), and (93).

A. *Proof of Equivalence of* (90) *and* (92)

We start with the equivalence of (90) and (92).

Fix the distribution of $\{\hat{\mathbf{X}}_k\}$ for the moment and write:

$$\begin{aligned}
&\chi_{\hat{\mathbf{X}}_0}(\mathbb{H}_0)+I\big(\mathbb{H}_0\hat{\mathbf{X}}_0;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}\,\big|\,\hat{\mathbf{X}}_{-\infty}^0\big)\\
&-I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\infty}\right)\\
&\quad=\lim_{n\uparrow\infty}\Bigg\{\chi_{\hat{\mathbf{X}}_0}(\mathbb{H}_0)+h\big(\mathbb{H}_0\hat{\mathbf{X}}_0\,\big|\,\hat{\mathbf{X}}_0\big)\\
&\qquad-\frac{1}{n}h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}\,\big|\,\hat{\mathbf{X}}_1^n\big)-h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\right)\\
&\qquad+\frac{1}{n}h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{n}\right)\Bigg\} &(515)
\end{aligned}$$

where we have used stationarity of $\{\mathbb{H}_k\}$ and $\{\hat{\mathbf{X}}_k\}$ and Lemma 12 to write

$$\begin{aligned}
&\lim_{n\uparrow\infty}h\big(\mathbb{H}_n\hat{\mathbf{X}}_n\,\big|\,\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n-1},\hat{\mathbf{X}}_1^n\big)\\
&\qquad=\lim_{n\uparrow\infty}\frac{1}{n}h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}\,\big|\,\hat{\mathbf{X}}_1^n\big) &(516)
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{n\uparrow\infty}h_\lambda\left(\frac{\mathbb{H}_n\hat{\mathbf{X}}_n}{\|\mathbb{H}_n\hat{\mathbf{X}}_n\|}\,\middle|\,\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{n-1}\right)\\
&\qquad=\lim_{n\uparrow\infty}\frac{1}{n}h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{n}\right). &(517)
\end{aligned}$$

Next we rely on Lemma 22 with $\nu=n$ and $\mathbf{V}_k=\mathbb{H}_k\hat{\mathbf{X}}_k$ (*i.e.*, $m=n_{\mathrm{R}}$) to write

$$\begin{aligned}
&\frac{1}{n}h_\lambda\left(\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{n}\right)\\
&\quad=\frac{1}{n}h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}\big)\\
&\qquad-\frac{1}{n}h\left(\big\{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|\big\}_{\ell=1}^{n}\,\middle|\,\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{n}\right)\\
&\qquad-\frac{2n_{\mathrm{R}}-1}{n}\sum_{k=1}^{n}\mathsf{E}\Big[\log\|\mathbb{H}_k\hat{\mathbf{X}}_k\|\Big] &(518)\\
&\quad=\frac{1}{n}h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}\big)\\
&\qquad-\frac{1}{n}h\left(\big\{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|\big\}_{\ell=1}^{n}\,\middle|\,\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{n}\right)\\
&\qquad-(2n_{\mathrm{R}}-1)\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|\Big] &(519)
\end{aligned}$$

where (519) follows from stationarity. Together with (85) and (515) this yields

$$\begin{aligned}
&\chi_{\hat{\mathbf{X}}_0}(\mathbb{H}_0)+I\big(\mathbb{H}_0\hat{\mathbf{X}}_0;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}\,\big|\,\hat{\mathbf{X}}_{-\infty}^0\big)\\
&-I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\infty}\right)\\
&\quad=\lim_{n\uparrow\infty}\Bigg\{h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\right)+n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\Big]-\log 2\\
&\qquad-h\big(\mathbb{H}_0\hat{\mathbf{X}}_0\,\big|\,\hat{\mathbf{X}}_0\big)+h\big(\mathbb{H}_0\hat{\mathbf{X}}_0\,\big|\,\hat{\mathbf{X}}_0\big)\\
&\qquad-\frac{1}{n}h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}\,\big|\,\hat{\mathbf{X}}_1^n\big)-h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\right)\\
&\qquad+\frac{1}{n}h\big(\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}\big)\\
&\qquad-\frac{1}{n}h\left(\big\{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|\big\}_{\ell=1}^{n}\,\middle|\,\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{n}\right)\\
&\qquad-(2n_{\mathrm{R}}-1)\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|\Big]\Bigg\} &(520)\\
&\quad=\lim_{n\uparrow\infty}\Bigg\{\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|\Big]-\log 2+\frac{1}{n}I\big(\hat{\mathbf{X}}_1^n;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}\big)
\end{aligned}$$

$$- \frac{1}{n} h \left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^n \,\middle|\, \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} \right\}_{\ell=1}^n \right) \Bigg\} \tag{521}$$

which proves the equivalence of (90) and (92).

B. Proof of Equivalence of (92) and (93)

In the following $\mathbf{H}_k^{(r)}$ denotes the r-th *row* of $\mathbb{H}_k$ (written as column vector), and $\Psi_k^{(r)}$ denotes the phase of $\mathbf{H}_k^{(r)\mathsf{T}} \hat{\mathbf{X}}_k$. We start with an observation: if every component of the unit vector $\frac{\mathbb{H}_k \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}$ is divided by the first component, one gets the following vector:

$$\left(1, \frac{\mathbf{H}_k^{(2)\mathsf{T}} \hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k}, \ldots, \frac{\mathbf{H}_k^{(n_\mathrm{R})\mathsf{T}} \hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k} \right)^{\mathsf{T}}. \tag{522}$$

Hence, using the fact that conditioning cannot increase entropy we get the following inequality:

$$\begin{aligned} &\frac{1}{n} h \left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^n \,\middle|\, \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} \right\}_{\ell=1}^n \right) \\ &= \frac{1}{n} h \left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^n \,\middle|\, \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} \right\}_{\ell=1}^n, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathrm{R} \\ \ell=1,\ldots,n}} \right) &(523) \\ &\leq \frac{1}{n} h \left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^n \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathrm{R} \\ \ell=1,\ldots,n}} \right). &(524) \end{aligned}$$

On the other hand, we can prove that the knowledge of the following expressions is equivalent:

$$\begin{aligned} &\frac{\mathbf{H}_k^{(2)\mathsf{T}} \hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k}, \ldots, \frac{\mathbf{H}_k^{(n_\mathrm{R})\mathsf{T}} \hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k} \\ &\iff \frac{\mathbf{H}_k^{(2)\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} \cdot \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k}, \ldots, \frac{\mathbf{H}_k^{(n_\mathrm{R})\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} \cdot \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k} &(525) \\ &\iff \frac{|\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}, \frac{\mathbf{H}_k^{(2)\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} \cdot \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k}, \ldots \\ &\qquad \ldots, \frac{\mathbf{H}_k^{(n_\mathrm{R})\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} \cdot \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k} &(526) \\ &\iff \frac{|\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}, \frac{\mathbf{H}_k^{(2)\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} e^{-\mathrm{i}\Psi_k^{(1)}}, \ldots \\ &\qquad \ldots, \frac{\mathbf{H}_k^{(n_\mathrm{R})\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} e^{-\mathrm{i}\Psi_k^{(1)}} &(527) \\ &\iff \frac{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} e^{-\mathrm{i}\Psi_k^{(1)}}, \frac{\mathbf{H}_k^{(2)\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} e^{-\mathrm{i}\Psi_k^{(1)}}, \ldots \\ &\qquad \ldots, \frac{\mathbf{H}_k^{(n_\mathrm{R})\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} e^{-\mathrm{i}\Psi_k^{(1)}} &(528) \\ &\iff \frac{\mathbb{H}_k \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} e^{-\mathrm{i}\Psi_k^{(1)}}. &(529) \end{aligned}$$

Here, the first equivalence (525) is obvious as we have not changed anything. To derive the equivalence (526) note that

$$\mathbb{H}_k \hat{\mathbf{X}}_k = \begin{pmatrix} \mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k \\ \vdots \\ \mathbf{H}_k^{(n_\mathrm{R})\mathsf{T}} \hat{\mathbf{X}}_k \end{pmatrix} \tag{530}$$

and that

$$\|\mathbb{H}_k \hat{\mathbf{X}}\|^2 = |\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|^2 + \cdots + |\mathbf{H}_k^{(n_\mathrm{R})\mathsf{T}} \hat{\mathbf{X}}_k|^2 \tag{531}$$

such that

$$\begin{aligned} &1 + \left| \frac{\mathbf{H}_k^{(2)\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} \cdot \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k} \right|^2 + \cdots + \left| \frac{\mathbf{H}_k^{(n_\mathrm{R})\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} \cdot \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k} \right|^2 \\ &= 1 + \left(\left| \frac{\mathbf{H}_k^{(2)\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} \right|^2 + \cdots + \left| \frac{\mathbf{H}_k^{(n_\mathrm{R})\mathsf{T}} \hat{\mathbf{X}}_k}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|} \right|^2 \right) \cdot \\ &\qquad \cdot \left| \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k} \right|^2 &(532) \\ &= 1 + \left(\frac{|\mathbf{H}_k^{(2)\mathsf{T}} \hat{\mathbf{X}}_k|^2 + \cdots + |\mathbf{H}_k^{(n_\mathrm{R})\mathsf{T}} \hat{\mathbf{X}}_k|^2}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|^2} \right) \cdot \\ &\qquad \cdot \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|^2}{|\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|^2} &(533) \\ &= 1 + \left(\frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|^2 - |\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|^2}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|^2} \right) \cdot \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|^2}{|\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|^2} &(534) \\ &= 1 + \left(1 - \frac{|\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|^2}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|^2} \right) \cdot \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|^2}{|\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|^2} &(535) \\ &= 1 + \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|^2}{|\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|^2} - 1 &(536) \\ &= \frac{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|^2}{|\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|^2} &(537) \end{aligned}$$

where (532) and (533) follow from algebraic rearrangements; (534) from (531); and the rest again from rearranging the terms. Hence, from (525) we can compute $\frac{|\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}$.

In (527) we multiply all but the first components by $\frac{|\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k|}{\|\mathbb{H}_k \hat{\mathbf{X}}_k\|}$ and recall that $\Psi_k^{(1)}$ denotes the phase of $\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k$. And (528) and (529) are again trivial rearrangements.

Hence, we get

$$\begin{aligned} &\frac{1}{n} h \left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^n \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathrm{R} \\ \ell=1,\ldots,n}} \right) \\ &= \frac{1}{n} h \left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^n \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathrm{R} \\ \ell=1,\ldots,n}}, \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} e^{-\mathrm{i}\Psi_\ell^{(1)}} \right\}_{\ell=1}^n \right) &(538) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}, \right. \\
&\qquad \left. \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} e^{-\mathrm{i}(\Psi_\ell^{(1)}+\Theta_\ell)} \right\}_{\ell=1}^{n}, \left\{ e^{\mathrm{i}\Theta_\ell} \right\}_{\ell=1}^{n} \right) && (539)\\
&\leq \frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} e^{-\mathrm{i}(\Psi_\ell^{(1)}+\Theta_\ell)} \right\}_{\ell=1}^{n} \right) && (540)\\
&= \frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} e^{\mathrm{i}\tilde{\Theta}_\ell} \right\}_{\ell=1}^{n} \right) && (541)\\
&= \frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} \right\}_{\ell=1}^{n} \right). && (542)
\end{aligned}
$$

Here, (538) follows from (529); in (539) we add $\{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{n}$ where we assume that $\{\Theta_k\}$ is IID uniformly distributed on $[0, 2\pi)$ and independent of any other random quantity in the expression; (540) relies on the fact that conditioning cannot increase entropy; then in (541) we introduce $\tilde{\Theta}_\ell \triangleq \Psi_\ell^{(1)} + \Theta_\ell$ which is still independent and IID $\sim \mathcal{U}([0, 2\pi))$; and (542) follows since we know that $\{\hat{\mathbf{X}}_k\}$ is circularly symmetric, *i.e.*, the law of $\{\hat{\mathbf{X}} e^{\mathrm{i}\tilde{\Theta}_k}\}$ is identical to the distribution of $\{\hat{\mathbf{X}}_k\}$.

So, since by (524)

$$
\begin{aligned}
&\frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} \right\}_{\ell=1}^{n} \right)\\
&\qquad \leq \frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right) && (543)
\end{aligned}
$$

and by (542)

$$
\begin{aligned}
&\frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} \right\}_{\ell=1}^{n} \right)\\
&\qquad \geq \frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right) && (544)
\end{aligned}
$$

it follows that

$$
\begin{aligned}
&\frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} \right\}_{\ell=1}^{n} \right)\\
&\qquad = \frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right). && (545)
\end{aligned}
$$

Using this observation we now continue with some arithmetic changes:

$$
\begin{aligned}
&\frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} \right\}_{\ell=1}^{n} \right)\\
&\quad = \frac{1}{n} h\left(\left\{ \|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right) && (546)\\
&\quad = \frac{1}{n} h\left(\left\{ \sqrt{1 + \left| \frac{\mathbf{H}_\ell^{(2)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right|^2 + \cdots + \left| \frac{\mathbf{H}_\ell^{(2)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right|^2} \cdot \left| \mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell \right| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right) && (547)\\
&\quad = \frac{1}{n} h\left(\left\{ \left| \mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell \right| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right)\\
&\qquad + \frac{1}{n} \sum_{k=1}^{n} \mathsf{E}\left[\log \sqrt{1 + \left| \frac{\mathbf{H}_k^{(2)\mathsf{T}} \hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k} \right|^2 + \cdots + \left| \frac{\mathbf{H}_k^{(2)\mathsf{T}} \hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}} \hat{\mathbf{X}}_k} \right|^2} \right] && (548)\\
&\quad = \frac{1}{n} h\left(\left\{ \left| \mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell \right| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right)\\
&\qquad + \frac{1}{2} \mathsf{E}\left[\log \left(1 + \left| \frac{\mathbf{H}_0^{(2)\mathsf{T}} \hat{\mathbf{X}}_0}{\mathbf{H}_0^{(1)\mathsf{T}} \hat{\mathbf{X}}_0} \right|^2 + \cdots + \left| \frac{\mathbf{H}_0^{(2)\mathsf{T}} \hat{\mathbf{X}}_0}{\mathbf{H}_0^{(1)\mathsf{T}} \hat{\mathbf{X}}_0} \right|^2 \right) \right] && (549)\\
&\quad = \frac{1}{n} h\left(\left\{ \left| \mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell \right| \right\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right)\\
&\qquad + \mathsf{E}\left[\log \|\mathbb{H}_0 \hat{\mathbf{X}}_0\| \right] - \mathsf{E}\left[\log \left| \mathbf{H}_0^{(1)\mathsf{T}} \hat{\mathbf{X}}_0 \right| \right]. && (550)
\end{aligned}
$$

Here (546) follows from (545); in (547) we use (531); (548) follows from the scaling property of *scalar* differential entropy; (549) follows from stationarity; and (550) again from (531).

Next we have a look at the following mutual information term:

$$
\begin{aligned}
&\frac{1}{n} I\left(\hat{\mathbf{X}}_1^n; \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \right)\\
&\quad = \frac{1}{n} I\left(\hat{\mathbf{X}}_1^n; \{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell\}_{\ell=1}^{n}, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right) && (551)\\
&\quad = \frac{1}{n} I\left(\hat{\mathbf{X}}_1^n; \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right)\\
&\qquad + \frac{1}{n} I\left(\hat{\mathbf{X}}_1^n; \{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right) && (552)\\
&\quad = \frac{1}{n} I\left(\hat{\mathbf{X}}_1^n; \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right)\\
&\qquad + \frac{1}{n} h\left(\{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell\}_{\ell=1}^{n} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^{(r)\mathsf{T}} \hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}} \hat{\mathbf{X}}_\ell} \right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}} \right)
\end{aligned}
$$

$$-\frac{1}{n}h\left(\left\{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\right\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}, \hat{\mathbf{X}}_1^n\right) \tag{553}$$

where (551) follows from writing the vectors component-wise and divide the components by the first component and where (552) follows from the chain rule.

We continue with the second term in (553) and use a conditional version of Lemma 22 with $\nu = n$ and $m = 1$ to write

$$\begin{aligned}
&\frac{1}{n}h\left(\left\{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\right\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right)\\
&\quad= \frac{1}{n}h\left(\left\{\left|\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\right|\right\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right)\\
&\qquad+ \frac{1}{n}h\left(\left\{\Psi_\ell^{(1)}\right\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}, \left\{\left|\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\right|\right\}_{\ell=1}^n\right)\\
&\qquad+ \frac{1}{n}\sum_{k=1}^n \mathsf{E}\left[\log\left|\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k\right|\right] \qquad (554)\\
&\quad= \frac{1}{n}h\left(\left\{\left|\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\right|\right\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right)\\
&\qquad+ \frac{1}{n}h\left(\left\{\Psi_\ell^{(1)}\right\}_{\ell=1}^n \,\middle|\, \left\{\Psi_\ell^{(r)}-\Psi_\ell^{(1)}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}, \left\{\left|\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell\right|\right\}_{\substack{r=1,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right)\\
&\qquad+ \mathsf{E}\left[\log\left|\mathbf{H}_0^{(1)\mathsf{T}}\hat{\mathbf{X}}_0\right|\right] \qquad (555)
\end{aligned}$$

where the last equality follows from stationarity and by noting that

$$\left(\left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}, \left\{\left|\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\right|\right\}_{\ell=1}^n\right)$$

is equivalent to

$$\left(\left\{\Psi_\ell^{(r)}-\Psi_\ell^{(1)}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}, \left\{\left|\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell\right|\right\}_{\substack{r=1,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right).$$

We are now able to show the equivalence of (92) and (93). We start with (92) and incorporate the expressions (550), (553), and (555):

$$\begin{aligned}
&\frac{1}{n}I\left(\hat{\mathbf{X}}_1^n; \left\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\right\}_{\ell=1}^n\right) - \log 2 + \mathsf{E}\left[\log\left\|\mathbb{H}_0\hat{\mathbf{X}}_0\right\|\right]\\
&- \frac{1}{n}h\left(\left\{\left\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\right\|\right\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\left\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\right\|}\right\}_{\ell=1}^n\right)\\
&= \frac{1}{n}I\left(\hat{\mathbf{X}}_1^n; \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right)\\
&\quad+ \frac{1}{n}h\left(\left\{\left|\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\right|\right\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right)\\
&\quad+ \frac{1}{n}h\left(\left\{\Psi_\ell^{(1)}\right\}_{\ell=1}^n \,\middle|\, \left\{\Psi_\ell^{(r)}-\Psi_\ell^{(1)}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}, \left\{\left|\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell\right|\right\}_{\substack{r=1,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right)\\
&\quad+ \mathsf{E}\left[\log\left|\mathbf{H}_0^{(1)\mathsf{T}}\hat{\mathbf{X}}_0\right|\right]\\
&\quad- \frac{1}{n}h\left(\left\{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\right\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}, \hat{\mathbf{X}}_1^n\right)\\
&\quad- \log 2 + \mathsf{E}\left[\log\left\|\mathbb{H}_0\hat{\mathbf{X}}_0\right\|\right]\\
&\quad- \frac{1}{n}h\left(\left\{\left|\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\right|\right\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right)\\
&\quad- \mathsf{E}\left[\log\left\|\mathbb{H}_0\hat{\mathbf{X}}_0\right\|\right] + \mathsf{E}\left[\log\left|\mathbf{H}_0^{(1)\mathsf{T}}\hat{\mathbf{X}}_0\right|\right] \qquad (556)\\
&= \frac{1}{n}I\left(\hat{\mathbf{X}}_1^n; \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right)\\
&\quad- \log 2 + \mathsf{E}\left[\log\left|\mathbf{H}_0^{(1)\mathsf{T}}\hat{\mathbf{X}}_0\right|^2\right]\\
&\quad+ \frac{1}{n}h\left(\left\{\Psi_\ell^{(1)}\right\}_{\ell=1}^n \,\middle|\, \left\{\Psi_\ell^{(r)}-\Psi_\ell^{(1)}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}, \left\{\left|\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell\right|\right\}_{\substack{r=1,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}\right)\\
&\quad- \frac{1}{n}h\left(\left\{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\right\}_{\ell=1}^n \,\middle|\, \left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathsf{R}\\ \ell=1,\ldots,n}}, \hat{\mathbf{X}}_1^n\right) \qquad (557)
\end{aligned}$$

which corresponds to (93).

ACKNOWLEDGMENTS

The author is in debt with many persons who contributed to this paper. Most thanks go to Tobias Koch and Amos Lapidoth. Without their efforts, contributions, and encouragements this paper would not have been possible. Apart from many other suggestions and improvements they have found several flaws in earlier versions of the proof of the upper bound and helped to fix them, and they have provided the fundament of the proof in Appendix J-B.

Further I would like to offer many thanks to Su Yu Ted, Chen Po-Ning, Li Ying, Rudolf Blazek, and Chen Guan-Yu for long and fruitful discussions.

Finally, I thank the reviewers for their valuable comments.

This work was supported by the Industrial Technology Research Institute (ITRI), Zhudong, Taiwan, under contracts G1-95003 and G1-96001 and by the National Science Council under NSC 96-2221-E-009-012-MY3.

References

[1] S. M. Moser, "The fading number of multiple-input multiple-output fading channels with memory," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 24–30, 2007, pp. 521–525.

[2] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, January 1999.

[3] L. Zheng and D. N. C. Tse, "Communicating on the Grassmann manifold: a geometric approach to the non coherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, February 2002.

[4] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3095–3110, December 2004.

[5] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *Electronics Letters*, vol. 33, no. 16, pp. 1368–1370, July 31, 1997.

[6] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh fading channels," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

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

[8] ——, "The fading number of single-input multiple-output fading channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 437–453, February 2006.

[9] S. M. Moser, "Duality-based bounds on channel capacity," Ph.D. dissertation, Swiss Federal Institute of Technology (ETH), Zurich, October 2004, Diss. ETH No. 15769. [Online]. Available: http://moser.cm.nctu.edu.tw/

[10] T. Koch, "On the asymptotic capacity of multiple-input single-output fading channels with memory," Master's thesis, Signal and Information Processing Laboratory, ETH Zurich, Switzerland, April 2004, supervised by Prof. Dr. Amos Lapidoth.

[11] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 437–446, February 2005.

[12] T. Koch and A. Lapidoth, "Degrees of freedom in non-coherent stationary MIMO fading channels," in *Proceedings Winter School on Coding and Information Theory*, Bratislava, Slovakia, February 20–25, 2005, pp. 91–97.

[13] ——, "The fading number and degrees of freedom in non-coherent MIMO fading channels: a peace pipe," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, September 4–9, 2005, pp. 661–665.

[14] A. Lapidoth, "The fading number and degrees of freedom: A peace pipe," presentation, Shushan Purim, Israel, March 27, 2005.

[15] S. M. Moser, "Bounds on the fading number of multiple-input single-output fading channels with memory," in *Proceedings International Wireless Communications and Mobile Computing Conference (IWCMC)*, Vancouver, Canada, July 3–6, 2006, pp. 3–8.

[16] ——, "On the fading number of multiple-input single-output fading channels with memory," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Seattle, WA, USA, July 9–14, 2006, pp. 1056–1060.

[17] ——, "Bounds on the fading number of multiple-input single-output fading channels with memory," Department of Communication Engineering, National Chiao Tung University (NCTU), final report of National Science Council (NSC) project NSC 94-2218-E-009-037, from December 1, 2005, to July 31, 2006.

[18] ——, "The fading number of memoryless multiple-input multiple-output fading channels," *IEEE Transactions on Information Theory*, vol. 53, no. 7, pp. 2652–2666, July 2007.

[19] T. Koch, A. Lapidoth, and S. M. Moser, "Fading number and degrees of freedom," 2008, in preparation.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.

[21] H. A. Priestley, *Introduction to Integration*. Oxford University Press, 1997.

Stefan M. Moser (S'01–M'05) was born in Switzerland. He received the M.Sc. degree in electrical engineering (with distinction) in 1999, the M.Sc. degree in industrial management (M.B.A.) in 2003, and the Ph.D. degree in the field of information theory in 2004, all from the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland.

During 1999 to 2003, he was a Research and Teaching Assistant, and from 2004 to 2005, he was a Senior Research Assistant with the Signal and Information Processing Laboratory, ETH Zurich. Since August 2005, he has been an Assistant Professor, and since August 2008, an Associate Professor with the Department of Communication Engineering, National Chiao Tung University (NCTU), Hsinchu, Taiwan. His research interests are in information theory and digital communications.

Dr. Moser received the National Chiao Tung University Outstanding Researchers Award in 2007, 2008, and 2009, the National Chiao Tung University Excellent Teaching Award and the National Chiao Tung University Outstanding Mentoring Award both in 2007, the Willi Studer Award of ETH in 1999, and the ETH Silver Medal for an excellent Master's thesis in 1999.