Bounds on the Fading Number of Multiple-Input Single-Output Fading Channels with Memory

Final Report of NSC Project
"Capacity of Communication Channels"

Project-Number: NSC 94-2218-E-009-037
Project Duration: 1 December 2005–31 July 2006
Funded by: National Science Council, Taiwan
Author: Stefan M. Moser
Organization: Information Theory Laboratory
Department of Communication Engineering
National Chiao Tung University
Address: Engineering Building IV, Room 727
1001 Ta Hsueh Rd.
Hsinchu 300, Taiwan
E-mail: stefan.moser@ieee.org

Abstract

The demand of new wireless communication systems with much higher data rates that allow, *e.g.*, mobile wireless broadband Internet connections inspires a quick advance in wireless transmission technology. So far most systems rely on an approach where the channel state is measured with the help of regularly transmitted training sequences. The detection of the transmitted data is then done under the assumption of *perfect* knowledge of the channel state. This approach will not be sufficient anymore for very high data rate systems since the loss of bandwidth due to the training sequences is too large. Therefore, the research interest on joint estimation and detection schemes has been increased considerably.

Apart from potentially higher data rates a further advantage of such a system is that it allows for a fair analysis of the theoretical upper limit, the so-called *channel capacity*. "Fair" is used here in the sense that the capacity analysis does not ignore the estimation part of the system, *i.e.*, it takes into account the need of the receiver to gain some knowledge about the channel state without restricting it to assume some particular form (particularly, this approach does also include the approach with training sequences!). The capacity of such a joint estimation and detection scheme is often also known as *non-coherent capacity*.

Recent studies investigating the non-coherent capacity of fading channels have shown very unexpected results. In stark contrast to the capacity with perfect channel knowledge at the receiver, it has been shown that non-coherent fading channels become very power-inefficient at high signal-to-noise ratios (SNR) in the sense that increasing the transmission rate by an additional bit requires squaring the necessary SNR. Since transmission in such a regime will be highly inefficient, it is crucial to better understand this behavior and to be able to give an estimation as to where the inefficient regime starts. One parameter that provides a good approximation to such a border between the power-efficient low-SNR and the power-inefficient high-SNR regime is the so-called *fading number* which is defined as the second term in the high-SNR asymptotic expansion of channel capacity.

The results of this report concern this fading number. We restrict ourselves to fading channels with multiple antennas at the transmitter, but only one antenna at the receiver (a multiple-input single-output (MISO) situation), however, we do allow memory. Furthermore, the fading laws are not restricted to be Gaussian, but is assumed to be a general regular law with spatial and temporal memory. The main result of this report are a new upper bound and a new lower bound on the fading number of this MISO fading channel with memory. It can be seen as a further step towards the final goal of the fading number of general multiple-inputs multiple-outputs (MIMO) fading channels with memory.

In case of an isotropically distributed fading vector it is proven that the upper and lower bound coincide, *i.e.*, the general MISO fading number with memory is known precisely.

The upper and lower bounds show that a type of beam-forming is asymptotically optimal.

Keywords: Beam-forming, channel capacity, fading, fading number, flat fading channel, high SNR, joint estimation and detection, memory, MISO, multiple-antenna, non-coherent detection.

Contents

Acknowledgments

Helpful comments from Amos Lapidoth, Daniel Hösli, Tobias Koch, and Natalia Miliou are gratefully acknowledged. Specially, I would like to thank Amos who spent a whole day discussing these results with me and Tobias who gave the right hint to save a serious bug in an earlier version of the proof of the upper bound.

This work was supported by the National Science Council under NSC 94-2218-E-009-037.

These results have been presented at the *2006 IEEE International Symposium on Information Theory (ISIT 2006)*, Seattle, WA, USA. Some preliminary results to this work have been presented at the *2006 International Wireless Communications and Mobile Computing Conference (IWCMC 2006)*, Vancouver, Canada.

Chapter 1

Introduction

1.1 General Background

The importance of mobile communication system nowadays needs not to be emphasized. Worldwide millions of people rely daily on their mobile phone. While for the user a mobile phone looks very similar to a old-fashioned wired telephone, the engineering technique behind it is very much different. The reason for this is that in a wireless communication system several physical effects occur that change the behavior of the channel completely compared with wired communication:

- The signal may find many different paths from the sender to the receiver via various different reflections (buildings, trees, etc.). Therefore the receiver receives multiple copies of the same signal, however, since each path has different length and different attenuation, the various copies of the signal will arrive at different times and with different strength.

- Since the transmitter and/or the receiver might be in motion while transmitting, a physical effect called *Doppler effect* occurs: the frequency of the transmitted signal is shifted depending on the relative movement between receiver and transmitter.

- Since receiver and transmitter are moving and because the environment is permanently changing (*e.g.*, movements by wind, passing cars, people, etc.), the different signal paths are constantly changing.

The first two effects lead to a channel that not only adds noise to the transmitted signal (as this is the case for the traditional wired communication channel), but also changes the amplitude of the signal (so called *fading*) and in extreme cases introduces inter-symbol interference. Both effects can be combatted using appropriate transmissions schemes and coding.

The fact of the time variant nature of the channel is more difficult to deal with. Nowadays, usually a wireless communication system uses training sequences that are regularly transmitted between real data in order to measure the channel state, and then this knowledge is used to detect the data. This approach has the advantage that the system design can be split into two parts: one part dealing with estimating the channel and one part doing the detection under the assumption that the channel state is perfectly known.

The big disadvantage of the separate estimation and detection is that it is rather inefficient because bandwidth is lost for the transmission of the training sequences.

Particularly, if the channel is fast changing, the estimates will quickly become poor and the amount of needed training data will be exuberantly large.

A more promising approach is to design a system that uses the received data carrying the information at the same time for estimating the channel state. Such a *joint estimation and detection* approach will be particularly important for future systems where the required data rates are considerably larger than the rates provided by present systems (like, *e.g.*, GSM).

A further advantage of such joint estimation and detection systems is that they allow fair and realistic approximations to the physically feasible data rates. To elaborate more on this point, we need to briefly review some basic facts from Information Theory: in his famous landmark paper "A Mathematical Theory of Communication" [1] Claude E. Shannon proved that for every communication channel there exists a maximal rate—denoted *capacity*—above which one cannot transmit information reliably, *i.e.*, the probability of making decoding errors tends to one. On the other hand for every rate below the capacity it is theoretically possible to design a system such that the error probability is as small as one wishes. Of course, depending on the aimed probability of error, the system design will be rather complex and one will encounter possibly very long delays between the start of the transmission until the signal can be decoded. Particularly the latter is a large obstacle in real systems, because most communication systems cannot afford large delays. Nevertheless, the capacity shows the ultimate limit of communication rate of the available channel and is therefore fundamental for the understanding of the channel and also for the judgment of implemented systems regarding their efficiency.

So far the capacity analysis of above mentioned wireless communication channels were based on the assumption that the receiver has *perfect knowledge* of the channel state due to the training sequences. The capacity was then computed without taking into account the estimation scheme. Such an approach will definitely lead to an overly optimistic capacity, because

- even with 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 sequences is completely ignored.

The new approach of joint estimation and detection now allows to incorporate the estimation into the capacity analysis. As a matter of fact, we don't even need to make some assumption about how a particular estimation scheme might work, but can directly try to derive the ultimate data rate that the theoretically best system could achieve. The capacity of such a system is also known as the *non-coherent capacity of fading channels.*

Unfortunately, the evaluation of the non-coherent channel capacity involves an optimization that is very difficult—if not infeasible—to evaluate analytically or numerically.[1] Therefore, the question arises how one could get knowledge about the ultimate limit of reliable communication over fading channels without having to solve this infeasible expression.

A promising and interesting approach to this goal is the study of good upper and lower bounds to channel capacity. However, one needs to be aware that finding upper bounds to an expression that itself is a maximization might be rather challenging, too.

[1] As a matter of fact, this optimization is infeasible for most channels of interest.

In [2] and extracts thereof published before [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], large progress has been made in tackling this problem: a technique has been proposed for the derivation of upper bounds on channel capacity.[2] It is based on a dual expression for channel capacity where the maximization (of mutual information) over distributions on the channel input alphabet is replaced with a minimization (of average relative entropy) over distributions on the channel output alphabet. Every choice of an output distribution leads to an upper bound on mutual information. The chosen output distribution need not correspond to some distribution on the channel input. With a judicious choice of output distributions one can often derive tight upper bounds on channel capacity.

Furthermore, in [2] a technique has been proposed for the analysis of the asymptotic capacity of general cost-constrained channels. The technique is based on the observation that—under fairly mild conditions on the channel—every input distribution that achieves a mutual information with the same growth-rate in the cost constraint as the channel capacity must *escape to infinity*; *i.e.*, under such a distribution for some finite cost, the probability of the set of input symbols of lesser cost tends to zero as the cost constraint tends to infinity. For more details about this concept see Section 3.1.1.

Both techniques have been proven very successful: they have been successfully applied to various channel models:

- the free-space optical intensity channel [2], [6], [8];
- an optical intensity channel with input-dependent noise [2];
- the Poisson channel [2], [6], [8];
- multiple-antenna flat fading channels with memory where the fading process is assumed to be *regular* (*i.e.*, of finite entropy rate[3]) and where the realization of the fading process is unknown at the transmitter and unknown (or only partially known) at the receiver [2], [4], [7];
- multiple-antenna flat fading channels with memory where the fading process may be *irregular* (*i.e.*, of possibly infinite entropy rate) and where the realization of the fading process is unknown (or only partially known) at the receiver [14], [15], [16], [17], [18];
- fading channels with feedback [19], [2], [5];
- non-coherent fading networks [20], [21];
- a phase noise channel [22], [23].

The bounds that have been derived in these contributions are often very tight. For various cases the asymptotic capacity in the limit when the available power (signal-to-noise ratio SNR) tends to infinity has been derived precisely. This is for example the case for the regular single-input multiple-output (SIMO) fading channel with memory and for the regular memoryless multiple-input single-output (MISO) fading channel. In other cases the *capacity pre-log* (*i.e.*, the ratio of channel capacity to the logarithm of the SNR in the limit when the SNR tends to infinity) could be quantified.

[2]The technique works for general channels, not fading channels only.

[3]*I.e.*, a process is called *regular* when the actual fading realization cannot be predicted even if the infinite past of the process is known.

Some of these results have been very unexpected. *E.g.*, it has been shown in [2] that regular fading processes have a capacity that grows only double-logarithmically in the SNR at high SNR. This means that at high power these channels become extremely power-inefficient in the sense that for every additional bit capacity the SNR needs to be squared or, respectively, on a dB-scale the SNR needs to be doubled! This behavior is independent of the particular law of the fading process, the law of the noise process, or the number of antennas at the transmitter or receiver. Moreover, the capacity-growth at high SNR is double-logarithmic irrespective whether there is memory in the fading process or not, and it even remains this slow when introducing *noiseless* feedback [19]! This is in stark contrast to the situation of additive noise channels and even to the so far known capacity results when assuming prefect knowledge of the channel state at the receiver: there the capacity grows logarithmically in the power and the mentioned factors (like, *e.g.*, number of antennas, memory, or feedback) have a strong (positive) impact on the capacity. For additive white Gaussian noise (AWGN) channels, *e.g.*, the number of receiver antennas multiplies the capacity and is therefore very beneficial!

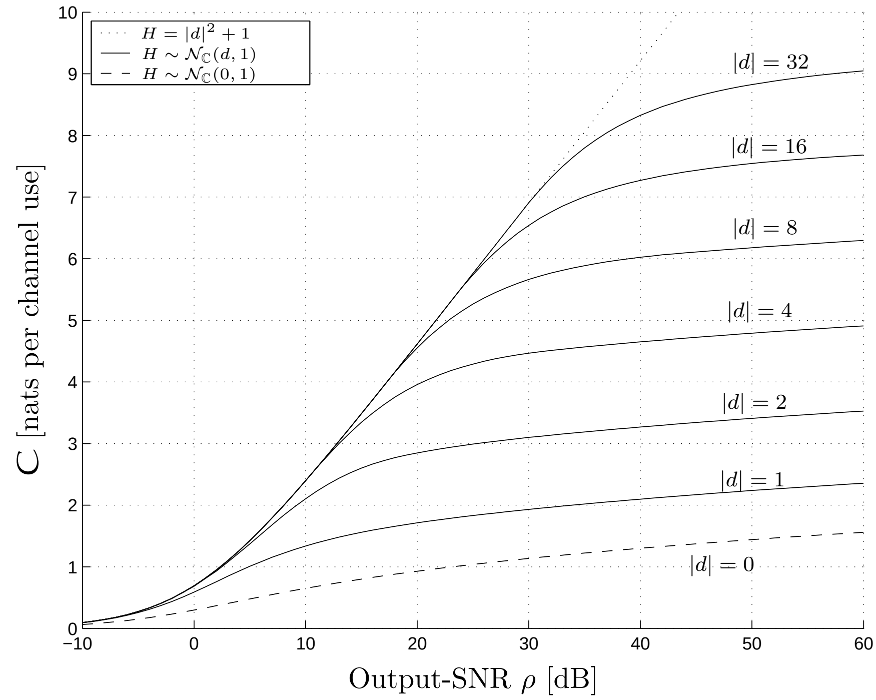

Figure 1.1: An upper bound on the capacity of a Rician fading channel for different values of the specular component d. The dotted line depicts the capacity of a Gaussian channel of equal output-SNR, namely $\log(1+\rho)$.

Therefore the question arises whether in the case of non-coherent fading channels multiple antennas or feedback is useful at all. It turns out that although the asymptotic growth rate of capacity is unchanged by these parameters, they still do have a large influence on the systems: the threshold above which the capacity growth changes from logarithmic to double-logarithmic is highly dependent on them! As an example Figure 1.1 shows the capacity of non-coherent Rayleigh fading channels with various numbers of receive antennas.

1.2 The Fading Number

In an attempt to quantify this threshold more precisely, the *fading number* has been introduced [7], [2]. The fading number is defined as the second term in the high-SNR capacity, *i.e.*, at high SNR the channel capacity can be expressed as

$$\mathsf{C}(\text{SNR}) = \log\log \text{SNR} + \chi + o(1). \tag{1.1}$$

Here, $o(1)$ denotes a term that tends to zero as the SNR tends to infinity; and χ is the fading number. For a mathematically more precise definition we refer to Chapter 2.2.

We would like now to motivate our claim that the fading number is related to the threshold between the efficient regime where capacity grows like $\log \text{SNR}$ and the inefficient regime where capacity only grows like $\log\log \text{SNR}$. To that goal we need to specify how to define this threshold. A very natural definition is as follows: we say that wireless communication system operates in the inefficient high-SNR regime, if its capacity can be well approximated by

$$\mathsf{C}(\text{SNR}) \approx \log\log \text{SNR} + \chi, \tag{1.2}$$

i.e., the $o(1)$-terms in (1.1) are small. Note that in the low- to medium-SNR regime these terms are dominating over the $\log\log \text{SNR}$-term.

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

$$\log\log(30 \text{ dB}) + \chi \le \mathsf{C}_0 \le \log\log(80 \text{ dB}) + \chi, \tag{1.3}$$

$$\implies \quad \chi + 2.1 \text{ nats} \le \mathsf{C}_0 \le \chi + 3 \text{ nats}. \tag{1.4}$$

From this immediately follows the following *rule of thumb*:

Conjecture 1. *A system that operates at rates appreciably above $\chi + 2$ nats is in the high-SNR regime and therefore extremely power-inefficient.*

Hence the fading number can 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.

Moreover, it follows from this observation that a system needs to be designed such as to have a large fading number. However, in order to understand how the fading number is influenced by the various design parameters like the number of antennas, feedback, etc., we need to know more about the exact value of χ. So far explicit expressions for the fading number were given for a number of fading models, *e.g.*, the fading number of single-input single-output (SISO) fading channels with memory was derived in [7], [2] and the single-input multiple-output (SIMO) case with memory was derived in [4], [3], [2].

However, there are still many interesting cases open and unsolved. For example, it is interesting to study the influence of multiple transmitter antennas on the fading number. The fading number of the multiple-input single-output (MISO) fading channel, for example, has only been derived in general for the memoryless case [7], [2]:

$$\chi(\mathbf{H}^\mathsf{T}) = \sup_{\|\hat{\mathbf{x}}\|=1} \left\{ \log\pi + \mathsf{E}\left[\log|\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}|^2\right] - h(\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}) \right\}. \tag{1.5}$$

This fading number is achievable by inputs that can be expressed as the product of a constant unit vector in $\mathbb{C}^{n_\mathrm{T}}$ 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 MISO fading channel is achieved by beam-forming where the beam-direction is chosen not to maximize the SNR, but the fading number.

In [15] and [16] Koch & Lapidoth investigate the fading number of MISO fading channels with memory where the fading is Gaussian. For the case of a mean-$\mathbf{d}$ Gaussian vector process with memory where $\{\mathbf{H}_k - \mathbf{d}\}$ is spatially independent and identically distributed (IID) and where each component is a zero-mean unit-variance circularly symmetric complex Gaussian process, the fading number is shown to be[4]

$$\chi_{\text{Gauss, spat. IID}}(\{\mathbf{H}_k^\mathsf{T}\}) = -1 + \log\|\mathbf{d}\|^2 - \mathrm{Ei}\bigl(-\|\mathbf{d}\|^2\bigr) + \log\frac{1}{\epsilon^2}, \tag{1.6}$$

where ϵ^2 denotes the prediction error when predicting one of the components of the fading vector based on the observation of its past.

Furthermore, Koch & Lapidoth derive an upper bound to the fading number for the general Gaussian case, *i.e.*, $\{\mathbf{H}_k - \mathbf{d}\}$ is a zero-mean circularly symmetric stationary ergodic complex Gaussian process with matrix-valued spectral distribution function $\mathsf{F}(\cdot)$ and with covariance matrix K. Assuming that the prediction error covariance matrix $\boldsymbol{\Sigma}$ is non-singular (regularity assumption) they show that

$$\chi_{\text{Gauss}}(\{\mathbf{H}_k^\mathsf{T}\}) \leq -1 + \log d_*^2 - \mathrm{Ei}\bigl(-d_*^2\bigr) + \log\frac{\|\mathsf{K}\|}{\lambda_{\min}}, \tag{1.7}$$

where

$$d_* = \max_{\|\hat{\mathbf{x}}\|=1} \frac{\bigl|\mathsf{E}[\mathbf{H}_k^\mathsf{T}]\,\hat{\mathbf{x}}\bigr|}{\sqrt{\mathsf{Var}\bigl(\mathbf{H}_k^\mathsf{T}\hat{\mathbf{x}}\bigr)}}; \tag{1.8}$$

$\lambda_{\min}$ denotes the smallest eigenvalue of $\boldsymbol{\Sigma}$; and where $\|\cdot\|$ denotes the Euclidean operator norm of matrices, *i.e.*, the largest singular value.

In this report we extend these results to general (not necessarily Gaussian) fading channels.

The remaining of this report is structured as follows: after some remarks about notation and a detailed mathematical definition of the channel model in the following chapter, we will present the main results, *i.e.*, a new upper and lower bound on the MISO fading number, in Chapter 3. There also some concepts are introduced that are important in the analysis of channel capacity, *e.g.*, the concept of distributions that *escape to infinity*, and relation between stationarity and capacity achieving input distributions.

We then specialize these results to the case of isotropically distributed fading processes in Chapter 4.1 and to Gaussian fading in Chapter 4.2. For isotropically distributed fading we will show that the upper and lower bound coincide. In the Gaussian case we shall derive the above mentioned results of Koch & Lapidoth as special cases of our bounds.

The proof of the main result is found in Chapter 5; and we conclude in Chapter 6.

[4]Note that all results in this paper are in nats.

Chapter 2

Definitions and Notation

2.1 Notation

We try to use upper-case letters for random quantities and lower-case letters for their realizations. This rule, however, is broken when dealing with matrices and some constants. 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. For random vectors we use bold face capitals, *e.g.*, $\mathbf{X}$ and bold lower-case for their realizations, *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.*, $\mathbb{H}$. If scalars or deterministic scalar functions are not denoted using Greek or lower-case letters, we use yet another font, *e.g.*, C for capacity (in contrast to C) or $\mathsf{F}(\cdot)$ for the spectral density function (in contrast to $F(\cdot)$). The energy per symbol is denoted by $\mathcal{E}$ and the signal-to-noise ratio SNR is denoted by SNR.

We use the shorthand H_a^b for $(H_a, H_{a+1}, \ldots, H_b)$. For more complicated expressions, such as $(\mathbf{H}_a^{\mathsf{T}}\hat{\mathbf{x}}_a, \mathbf{H}_{a+1}^{\mathsf{T}}\hat{\mathbf{x}}_{a+1}, \ldots, \mathbf{H}_b^{\mathsf{T}}\hat{\mathbf{x}}_b)$, 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. The trace of a matrix is denoted by $\operatorname{tr}(\cdot)$.

We use $\|\cdot\|$ to denote the Euclidean norm of vectors or the Euclidean operator norm of matrices. That is,

$$\|\mathbf{x}\| \triangleq \sqrt{\sum_{t=1}^{m} |x^{(t)}|^2}, \quad \mathbf{x} \in \mathbb{C}^m \tag{2.1}$$

$$\|\mathsf{A}\| \triangleq \max_{\|\hat{\mathbf{w}}\|=1} \|\mathsf{A}\hat{\mathbf{w}}\|. \tag{2.2}$$

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

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

$$\hat{\mathbf{v}} \triangleq \frac{\mathbf{v}}{\|\mathbf{v}\|} \tag{2.3}$$

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

$$\|\hat{\mathbf{v}}\| = \|\hat{\mathbf{V}}\| = 1. \tag{2.4}$$

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 $\mathbb{C}^m$: let λ denote the area measure on the unit sphere in $\mathbb{C}^m$. 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

$$h_\lambda(\hat{\mathbf{V}}) \triangleq -\mathsf{E}\left[\log p^{\lambda}_{\hat{\mathbf{V}}}(\hat{\mathbf{V}})\right] \tag{2.5}$$

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

$$h_\lambda(\mathsf{U}\hat{\mathbf{V}}) = h_\lambda(\hat{\mathbf{V}}). \tag{2.6}$$

Also note that if $\hat{\mathbf{V}}$ is uniformly distributed on the unit sphere, then $h_\lambda(\hat{\mathbf{V}}) = \log c_m$, where c_m denotes the surface area of the unit sphere in $\mathbb{C}^m$

$$c_m = \frac{2\pi^m}{\Gamma(m)}. \tag{2.7}$$

The definition (2.5) 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

$$h_\lambda\big(\hat{\mathbf{V}} \,\big|\, \mathbf{W} = \mathbf{w}\big) \triangleq -\mathsf{E}\left[\log p^{\lambda}_{\hat{\mathbf{V}}|\mathbf{W}}(\hat{\mathbf{V}}|\mathbf{W}) \,\middle|\, \mathbf{W} = \mathbf{w}\right] \tag{2.8}$$

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 defined in* (2.3)*. Then*

$$h(\mathbf{V}) = h(\|\mathbf{V}\|) + h_\lambda\big(\hat{\mathbf{V}} \,\big|\, \|\mathbf{V}\|\big) + (2m-1)\mathsf{E}[\log\|\mathbf{V}\|] \tag{2.9}$$

$$= h_\lambda\big(\hat{\mathbf{V}}\big) + h\big(\|\mathbf{V}\| \,\big|\, \hat{\mathbf{V}}\big) + (2m-1)\mathsf{E}[\log\|\mathbf{V}\|] \tag{2.10}$$

whenever all the quantities in (2.9) *and* (2.10)*, respectively, are defined. Here $h(\|\mathbf{V}\|)$ is the differential entropy of $\|\mathbf{V}\|$ when viewed as a real (scalar) random variable.*

Proof. Omitted. □

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 Gaussian random vector of covariance matrix $\mathsf{E}\big[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\dagger\big] = \mathsf{K}$. By $X \sim \mathcal{U}([a,b])$ we denote a random variable that is uniformly distributed on the interval $[a,b]$.

All rates specified in this paper are in nats per channel use, *i.e.*, $\log(\cdot)$ denotes the natural logarithmic function.

2.2 The Channel Model

We consider a MISO fading channel whose time-k output $Y_k \in \mathbb{C}$ is given by

$$Y_k = \mathbf{H}_k^{\mathsf{T}}\mathbf{x}_k + Z_k \tag{2.11}$$

where $\mathbf{x}_k \in \mathbb{C}^{n_{\mathrm{T}}}$ denotes the time-k channel input vector; where the random vector $\mathbf{H}_k$ denotes the time-k fading vector; where $\mathbf{H}_k^{\mathsf{T}}$ denotes the transpose of the vector $\mathbf{H}_k$; and where Z_k denotes additive noise. Here $\mathbb{C}$ denotes the complex field, $\mathbb{C}^{n_{\mathrm{T}}}$ denotes the n_{T}-dimensional complex Euclidean space, and n_{T} is the number of transmit antennas. We assume that the additive noise is an IID zero-mean white Gaussian process of variance $\sigma^2 > 0$.

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

$$\mathsf{E}\left[\|\mathbf{H}_k\|^2\right] < \infty, \tag{2.12}$$

and of finite differential entropy rate

$$h(\{\mathbf{H}_k\}) > -\infty \tag{2.13}$$

(the *regularity assumption*).

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

As for the input, we consider two different constraints: a peak-power constraint and 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

$$\textsc{snr} \triangleq \frac{\mathcal{E}}{\sigma^2}. \tag{2.14}$$

The capacity $\mathsf{C}(\textsc{snr})$ of the channel (2.11) is given by

$$\mathsf{C}(\textsc{snr}) = \lim_{n\to\infty} \frac{1}{n} \sup I\left(\mathbf{X}_1^n; Y_1^n\right) \tag{2.15}$$

where we use $\mathbf{X}_j^k$ to denote $\mathbf{X}_j, \ldots, \mathbf{X}_k$ and where the supremum is over the set of all probability distributions on $\mathbf{X}_1^n$ satisfying the constraints, *i.e.*,

$$\|\mathbf{X}_k\|^2 \le \mathcal{E}, \quad \text{almost surely}, \quad k = 1, 2, \ldots, n \tag{2.16}$$

for a peak constraint, or

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[\|\mathbf{X}_k\|^2\right] \le \mathcal{E} \tag{2.17}$$

for an average constraint.

Specializing [7, Theorem 4.2] or [2, Theorem 6.10], respectively, to MISO fading, we have

$$\varlimsup_{\textsc{snr}\uparrow\infty} \Big\{\mathsf{C}(\textsc{snr}) - \log\log\textsc{snr}\Big\} < \infty. \tag{2.18}$$

The fading number χ is now defined as in [7, Definition 4.6] and in [2, Definition 6.13] by

$$\chi(\{\mathbf{H}_k^{\mathsf{T}}\}) \triangleq \varlimsup_{\textsc{snr}\uparrow\infty} \Big\{\mathsf{C}(\textsc{snr}) - \log\log\textsc{snr}\Big\}. \tag{2.19}$$

Prima facie the fading number depends on whether a peak-power constraint (2.16) or an average-power constraint (2.17) is imposed on the input. Since a peak-power constraint is more stringent than an average-power constraint, we will derive the upper bound using the average-power constraint and the lower bound using the peak-power constraint. In case of an isotropically distributed fading process we shall see that both constraints lead to identical fading numbers.

Chapter 3

Main Results

3.1 Preliminaries

Before we can state our new results, we need to give some preliminary results.

3.1.1 Escaping to Infinity

We start with a discussion about the concept of capacity achieving input distributions that escape to infinity.

A sequence of input distributions parameterized by the allowed cost (in our case of fading channels the cost is the available power or the SNR, respectively) 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. Loosely speaking, 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 of importance since 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 [7], [2].

Definition 3. *Let* $\{Q_{\mathcal{E}}\}_{\mathcal{E}\geq 0}$ *be a family of input distributions for the memoryless version of the fading channel* (2.11), i.e., *input distributions of the channel*

$$Y = \mathbf{H}^{\mathsf{T}}\mathbf{x} + Z \tag{3.1}$$

where $\mathbf{x} \in \mathbb{C}^{n_{\mathrm{T}}}$. *This family is parameterized by the available average power* $\mathcal{E}$ *such that*

$$\mathsf{E}_{Q_{\mathcal{E}}}\left[\|\mathbf{X}\|^2\right] \leq \mathcal{E}, \qquad \mathcal{E} \geq 0. \tag{3.2}$$

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

$$\lim_{\mathcal{E}\uparrow\infty} Q_{\mathcal{E}}\big(\|\mathbf{X}\|^2 \leq \mathcal{E}_0\big) = 0. \tag{3.3}$$

We now have the following:

Lemma 4. *Let the memoryless MISO fading channel be given as in* (3.1) *and let* $W(\cdot|\cdot)$ *denote the corresponding conditional channel law. Let* $\{Q_{\mathcal{E}}\}_{\mathcal{E}\geq 0}$ *be a family of input distributions satisfying the power constraint* (3.2) *and the condition*

$$\lim_{\mathcal{E}\uparrow\infty} \frac{I(Q_{\mathcal{E}}, W)}{\log\log\mathcal{E}} = 1. \tag{3.4}$$

Then $\{Q_{\mathcal{E}}\}_{\mathcal{E}\geq 0}$ *escapes to infinity.*

Proof. A proof can be found in [7], [2]. □

Hence, when computing bounds on the fading number (which is part of the capacity in the limit when $\mathcal{E}$ tends to infinity, see (2.19)) we may assume that

$$\Pr\left[\|\mathbf{X}\|^2 \leq \mathcal{E}_0\right] = 0. \tag{3.5}$$

3.1.2 An Upper Bound on Channel Capacity

In [7], [2] a new approach of finding upper bounds to channel capacity has been introduced. Since capacity is by definition a maximization of mutual information, it is implicitly difficult to find *upper* bounds on it. The new proposed technique bases on a dual expression of mutual information that leads to an expression of capacity as a minimization instead of a maximization. This way it becomes much easier to find upper bounds.

Again, here we only state the upper bound in a form needed in the derivation of Theorem 7, for a more general form, for more mathematical details, and for all proofs we refer to [7], [2].

Lemma 5. *Consider a memoryless channel*[1] *with input alphabet* $\mathbb{C}^{n_{\mathrm{R}}}$ *and output alphabet* $\mathbb{C}$ *as given in* (3.1). *Then the mutual information between input and output of the channel is upper-bounded as follows:*

$$\begin{aligned} I(\mathbf{X};Y) = &-h(Y|\mathbf{X}) + \log\pi + \alpha\log\beta + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) \\ &+ (1-\alpha)\mathsf{E}\left[\log\left(|Y|^2+\nu\right)\right] + \frac{1}{\beta}\mathsf{E}\left[|Y|^2\right] + \frac{\nu}{\beta} \end{aligned} \tag{3.6}$$

where $\alpha, \beta > 0$ *and* $\nu \geq 0$ *are parameters that can be chosen freely.*

Proof. A proof can be found in [7], [2]. □

3.1.3 Capacity Achieving Input Distributions and Stationarity

One of the main assumption about our channel model is that the fading process and the additive noise are *stationary*. This assumption is crucial both for the results as well as the derivation, *i.e.*, we don't believe the results to still be valid in a non-stationary setting.

From an intuitive point of view a stationary channel model should have a capacity achieving input distribution that is stationary. Unfortunately, we are not aware of a rigorous proof of this claim. However, we are able to prove a less strong statement which is basically saying that for our channel model we may limit ourselves to joint input distributions under which the input vectors have the same law for (almost) all time k:

[1] Actually, the lemma requires some mathematical conditions on the alphabets and the channel law to be satisfied. However, all these conditions are satisfied in our context. For more detail see [7], [2].

Lemma 6. *Fix some power $\mathcal{E}$ with corresponding* $\text{SNR} \triangleq \mathcal{E}/\sigma^2$*. Let $\mathsf{C}(\mathcal{E})$ denote the corresponding channel capacity. Then for every fixed $\epsilon > 0$ there corresponds some positive integer $\eta = \eta(\mathcal{E}, \epsilon)$ and some distribution $Q_{\mathcal{E},\epsilon} = Q(\mathcal{E}, \epsilon)$ on $\mathbb{C}^{n_\text{T}}$ such that for every 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*

$$\frac{1}{n} I\bigl(\mathbf{X}_1^n; Y_1^n\bigr) \geq \mathsf{C}(\mathcal{E}) - \epsilon. \tag{3.7}$$

2. *Except for the first $\eta - 1$ vectors $\mathbf{X}_1^{\eta-1}$ and for at most the last $2(\eta - 1)$ vectors $\mathbf{X}_{n-2\eta+3}^n$ the vectors*

$$\mathbf{X}_\eta, \mathbf{X}_{\eta+1}, \ldots, \mathbf{X}_{n-2\eta+2} \tag{3.8}$$

   all have the same distribution $Q_{\mathcal{E},\epsilon}$.

3. *This marginal distribution $Q_{\mathcal{E},\epsilon}$ gives rise to a second moment $\mathcal{E}$:*

$$\mathsf{E}\bigl[\|\mathbf{X}_\ell\|^2\bigr] = \mathcal{E}, \qquad \ell = \eta, \ldots, n - 2\eta + 2. \tag{3.9}$$

4. *The first $\eta - 1$ symbols and the last $2(\eta - 1)$ vectors satisfy the power constraint possibly strictly*

$$\mathsf{E}\bigl[\|\mathbf{X}_\ell\|^2\bigr] \leq \mathcal{E}, \quad \ell \in \{1, \ldots, \eta - 1\} \cup \{n - 2\eta + 3, \ldots, n\}. \tag{3.10}$$

Proof. See Appendix A. □

Note that this lemma and its proof are analogous to a very similar lemma needed in the derivation of the fading number of SIMO fading channels with memory [7], [2].

3.2 Main Results

We are now ready to state the new bounds on the fading number of a MISO fading channel with memory. We start with an upper bound:

Theorem 7. *Consider a MISO fading channel with memory* (2.11) *where the stationary and ergodic fading process $\{\mathbf{H}_k\}$ takes value in $\mathbb{C}^{n_\text{T}}$ and satisfies $h(\{\mathbf{H}_k\}) > -\infty$ and $\mathsf{E}\bigl[\|\mathbf{H}_k\|^2\bigr] < \infty$. Then, irrespective of whether a peak-power constraint* (2.16) *or an average-power constraint* (2.17) *is imposed on the input, the fading number $\chi\bigl(\{\mathbf{H}_k^\mathsf{T}\}\bigr)$ is upper-bounded by*

$$\chi\bigl(\{\mathbf{H}_k^\mathsf{T}\}\bigr) \leq \sup_{\hat{\mathbf{x}}_{-\infty}^0} \Bigl\{ \log \pi + \mathsf{E}\bigl[\log |\mathbf{H}_0^\mathsf{T} \hat{\mathbf{x}}_0|^2\bigr] - h\Bigl(\mathbf{H}_0^\mathsf{T} \hat{\mathbf{x}}_0 \,\Big|\, \{\mathbf{H}_\ell^\mathsf{T} \hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\Bigr) \Bigr\} \tag{3.11}$$

where $\hat{\mathbf{x}}_\ell \triangleq \frac{\mathbf{x}_\ell}{\|\mathbf{x}_\ell\|}$ denotes a vector of unit length.

Proof. We give here only an outline of the proof. The details can be found in Chapter 5.1.

The basic idea of the proof is to split the mutual information into a term that does not take into account the memory of the fading process and a term that takes care of the memory:

$$\frac{1}{n} I\bigl(\mathbf{X}_1^n; Y_1^n\bigr) \approx I(\mathbf{X}_0; Y_0) + \frac{1}{n} \sum_{k=1}^{n} I\bigl(\mathbf{H}_k^\mathsf{T} \hat{\mathbf{X}}_k; \{\mathbf{H}_\ell^\mathsf{T} \hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1} \,\big|\, \hat{\mathbf{X}}_1^n\bigr). \tag{3.12}$$

The derivation of this expression is complicated by the fact that Lemma 6 only guarantees equal marginals away from the edges, *i.e.*, we need to take care of the edge effects.

In a next step we now upper-bound the first term on the RHS using Lemma 5. This leads to a rather complicated looking expression with various terms that depend on the blocklength n, the power $\mathcal{E}$, the free parameters α, β, and ν, and, of course, on the input distribution. However, interestingly, there are no terms that depend on the input direction $\hat{\mathbf{X}}_k$ and the input amplitude $\|\mathbf{X}_k\|$ at the same time. We can therefore separate the expression in a group of terms that depend on $\mathcal{E}$ and another group of terms that do not depend on $\mathcal{E}$.

By an appropriate choice of the free parameters, and by letting n and $\mathcal{E}$ (in this order) go to infinity, we end up with the following bound:

$$\chi(\{\mathbf{H}_k^\mathsf{T}\}) \leq \sup_{Q_{\hat{\mathbf{X}}_{-\infty}^0}} \left\{ \log \pi + \mathsf{E}\Big[\log |\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0|^2\Big] - h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\kappa}^0\big)\right\} \tag{3.13}$$

$$\leq \sup_{\hat{\mathbf{x}}_{-\infty}^0} \left\{ \log \pi + \mathsf{E}\big[\log |\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}}_0|^2\big] - h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}}_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}_\ell\}_{\ell=-\kappa}^{-1}\big)\right\}. \tag{3.14}$$

□

Next we state a lower bound to the fading number of a MISO fading channel:

Theorem 8. *Consider a MISO fading channel with memory* (2.11) *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 the fading number* $\chi\big(\{\mathbf{H}_k^\mathsf{T}\}\big)$ *is lower-bounded by*

$$\chi\big(\{\mathbf{H}_k^\mathsf{T}\}\big) \geq \sup_{\hat{\mathbf{x}}} \left\{ \log \pi + \mathsf{E}\big[\log |\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}}|^2\big] - h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}} \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=-\infty}^{-1}\big)\right\} \tag{3.15}$$

where $\hat{\mathbf{x}} \triangleq \frac{\mathbf{x}}{\|\mathbf{x}\|}$ *denotes a vector of unit length.*

Moreover, this lower bound is achievable by IID inputs that can be expressed as the product of a constant unit vector $\hat{\mathbf{x}} \in \mathbb{C}^{n_\mathrm{T}}$ *and a circularly symmetric, scalar, complex IID random process* $\{X_k\}$ *such that*

$$\log |X_k|^2 \sim \mathcal{U}\left([\log\log \mathcal{E}, \log \mathcal{E}]\right). \tag{3.16}$$

Note that this input satisfies the peak-power constraint (2.16) *(and therefore also the average-power constraint* (2.17)*).*

Proof. We give here only an outline of the proof. The details can be found in Chapter 5.2.

The lower bound is based on the assumption of a specific input distribution which is chosen to be of the form

$$\mathbf{X}_k = X_k \cdot \hat{\mathbf{x}} \tag{3.17}$$

where $\hat{\mathbf{x}}$ is a deterministic unit vector (the beam-direction) and where $\{X_k\}$ is IID circularly symmetric with

$$\log |X_k|^2 \sim \mathcal{U}\left([\log\log \mathcal{E}, \log \mathcal{E}]\right). \tag{3.18}$$

Note that this choice for $\{X_k\}$ achieves the fading number for the SISO fading channel

$$Y_k = (\mathbf{H}_k^\mathsf{T}\hat{\mathbf{x}}) \cdot X_k + Z_k \tag{3.19}$$

with fading process $\{H_k\} = \{\mathbf{H}_k^\mathsf{T}\hat{\mathbf{x}}\}$. The lower bound is then derived by proving

$$\frac{1}{n} I(\mathbf{X}_1^n; Y_1^n) \approx \frac{1}{n} \sum_{k=1}^{n} I\big(X_k; Y_k \big| \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=1}^{k-1}\big) \tag{3.20}$$

and using the results of memoryless SISO fading channels with side-information [7], [2]. $\square$

Chapter 4

Special Cases

Before we discuss the proofs more in detail, we would like to add some insight by considering some special cases and specializing Theorem 7 and Theorem 8 to these situations. Firstly, we will analyze a fading process that is isotropically distributed, and secondly we investigate the in practice very important special case of Gaussian fading.

4.1 Isotropically Distributed Fading

Let's consider the special case of isotropically distributed fading processes, *i.e.*, for every deterministic unitary $n_\mathrm{T} \times n_\mathrm{T}$ matrix U

$$\mathbf{H}_k \overset{\mathscr{L}}{=} \mathsf{U}\mathbf{H}_k, \tag{4.1}$$

where we use "$\overset{\mathscr{L}}{=}$" to denote *equal in law.*

In this case we have the following corollary:

Corollary 9. *Consider a MISO fading channel with memory* (2.11) *where the stationary and ergodic fading process* $\{\mathbf{H}_k\}$ *takes value in* $\mathbb{C}^{n_\mathrm{T}}$*, satisfies* $h(\{\mathbf{H}_k\}) > -\infty$ *and* $\mathsf{E}\left[\|\mathbf{H}_k\|^2\right] < \infty$*, and is* isotropically distributed*. Then the upper bound* (3.11) *and the lower bound* (3.15) *coincide and the fading number* $\chi_{\mathrm{iso}}\big(\{\mathbf{H}_k^\mathsf{T}\}\big)$ *is given by*

$$\chi_{\mathrm{iso}}\big(\{\mathbf{H}_k^\mathsf{T}\}\big) = \log \pi + \mathsf{E}\left[\log |\mathbf{H}_0^\mathsf{T}\hat{\mathbf{e}}|^2\right] - h\Big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{e}} \,\Big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{e}}\}_{\ell=-\infty}^{-1}\Big) \tag{4.2}$$

where $\hat{\mathbf{e}}$ *is some deterministic unit vector.*

Proof. This corollary follows immediately from Theorem 7 and 8 by noting that for every $\hat{\mathbf{e}}$

$$\mathbf{H}_k^\mathsf{T}\hat{\mathbf{e}} \overset{\mathscr{L}}{=} \mathbf{H}_k^\mathsf{T}\mathsf{U}^\mathsf{T}\hat{\mathbf{e}} = \mathbf{H}_k^\mathsf{T}\hat{\mathbf{e}}' \tag{4.3}$$

where the first equality in law follows from (4.1) and the second equality by defining a new unit vector $\hat{\mathbf{e}}' \triangleq \mathsf{U}^\mathsf{T}\hat{\mathbf{e}}$. Note that for the MISO case *isotropically distributed* is equivalent to *rotation commutative in the generalized sense* as defined in [7, Definition 4.37] or [2, Definition 6.37]. □

4.2 Gaussian Fading

In this section we assume that the fading process $\{\mathbf{H}_k\}$ is a mean-$\mathbf{d}$ Gaussian process such that $\{\tilde{\mathbf{H}}_k\} = \{\mathbf{H}_k - \mathbf{d}\}$ is a zero-mean, circularly symmetric, stationary, ergodic, complex Gaussian process with matrix-valued spectral distribution function $\mathsf{F}(\cdot)$, and with covariance matrix K. Furthermore, we assume that the prediction error covariance matrix $\mathsf{\Sigma}$ is non-singular (regularity assumption).

4.2.1 Upper Bound for Gaussian Fading

We start with a new derivation of the upper bound (1.7) based on Theorem 7. We will see that (1.7) is in general less tight than (3.11).

We start by loosening the upper bound (3.11) as follows:

$$\chi\big(\{\mathbf{H}_k^{\mathsf{T}}\}\big) \le \sup_{\hat{\mathbf{x}}_{-\infty}^{0}} \Big\{ \log\pi + \mathsf{E}\big[\log|\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0|^2\big] - h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0 \,\big|\, \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big)\Big\} \tag{4.4}$$

$$= \sup_{\hat{\mathbf{x}}_{-\infty}^{0}} \Big\{ \log\pi + \mathsf{E}\big[\log|\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0|^2\big] - h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0\big) + h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0\big) - h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0 \,\big|\, \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big)\Big\} \tag{4.5}$$

$$\le \sup_{\hat{\mathbf{x}}_0} \Big\{ \log\pi + \mathsf{E}\big[\log|\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0|^2\big] - h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0\big)\Big\} + \sup_{\hat{\mathbf{x}}_{-\infty}^{0}} \Big\{ h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0\big) - h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0 \,\big|\, \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big)\Big\} \tag{4.6}$$

$$= \sup_{\hat{\mathbf{x}}_0} \Big\{ \log\pi + \mathsf{E}\big[\log|\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0|^2\big] - h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0\big)\Big\} + \sup_{\hat{\mathbf{x}}_{-\infty}^{0}} I\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0; \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big), \tag{4.7}$$

where (4.6) follows from

$$\sup_x \{f(x) + g(x)\} \le \sup_x f(x) + \sup_x g(x). \tag{4.8}$$

In [7, Corollary 4.28], [2, Corollary 6.28] it has been shown that the IID MISO fading number (1.5) for Gaussian fading is given by

$$\chi(\mathbf{H}^{\mathsf{T}}) = \sup_{\|\hat{\mathbf{x}}\|=1} \big\{\log\pi + \mathsf{E}\big[\log|\mathbf{H}^{\mathsf{T}}\hat{\mathbf{x}}|^2\big] - h\big(\mathbf{H}^{\mathsf{T}}\hat{\mathbf{x}}\big)\big\} \tag{4.9}$$

$$= -1 + \log d_*^2 - \mathrm{Ei}\big(-d_*^2\big) \tag{4.10}$$

where d_* is given in (1.8). This proves the equivalence of the first supremum in (4.7) with the first three terms of (1.7). It therefore only remains to prove that

$$\sup_{\hat{\mathbf{x}}_{-\infty}^{0}} I\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0; \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big) \le \log \frac{\|\mathsf{K}\|}{\lambda_{\min}}. \tag{4.11}$$

To this goal note that

$$\sup_{\hat{\mathbf{x}}_{-\infty}^{0}} I\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0; \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big) \le \sup_{\hat{\mathbf{x}}_{-\infty}^{0}} I\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0; \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}, \mathbf{H}_{-\infty}^{-1}\big) \tag{4.12}$$

$$= \sup_{\hat{\mathbf{x}}_0} I\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0; \mathbf{H}_{-\infty}^{-1}\big) \tag{4.13}$$

$$= \sup_{\hat{\mathbf{x}}_0} \Big\{ h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0\big) - h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0 \,\big|\, \mathbf{H}_{-\infty}^{-1}\big)\Big\} \tag{4.14}$$

$$= \sup_{\hat{\mathbf{x}}_0} \Big\{ \log\Big(\pi e \hat{\mathbf{x}}_0^{\dagger}\mathsf{K}\hat{\mathbf{x}}_0\Big) - h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0 \,\big|\, \mathbf{H}_{-\infty}^{-1}\big)\Big\}. \tag{4.15}$$

Here, the first inequality follows from the inclusion of additional random variables in the mutual information; the subsequent equality from the fact that given the past realization of the fading, $\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0$ is independent of $\{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}$; and in the last equality we have used the expression for the differential entropy of a Gaussian random variable with K denoting the covariance matrix of $\{\mathbf{H}_k\}$.

Note that the first inequality in general is not tight, *i.e.*, (1.7) is in general looser than (4.7) which in turn is in general looser than (3.11).

To compute the second term on the RHS of (4.15), we express the fading $\mathbf{H}_0$ as

$$\mathbf{H}_0 = \bar{\mathbf{H}}_0 + \tilde{\mathbf{H}}_0 \tag{4.16}$$

with $\bar{\mathbf{H}}_0$ being the best estimate of $\mathbf{H}_0$ based on the past realizations. We note that $\tilde{\mathbf{H}}_0 \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ denotes the prediction error covariance matrix. Hence

$$h\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0 \,\big|\, \mathbf{H}_{-\infty}^{-1}\big) = \log\Big(\pi e \hat{\mathbf{x}}_0^{\dagger}\boldsymbol{\Sigma}\hat{\mathbf{x}}_0\Big). \tag{4.17}$$

The bound (4.11) now follows by the Rayleigh-Ritz Theorem [24, Theorem 4.2.2], [2, Theorem A.9]

$$\lambda_{\min} = \min_{\hat{\mathbf{x}}} \hat{\mathbf{x}}^{\dagger}\boldsymbol{\Sigma}\hat{\mathbf{x}}, \tag{4.18}$$

the definition of the Euclidean norm of matrices, and the properties of positive semi-definite matrices:

$$\max_{\hat{\mathbf{x}}} \hat{\mathbf{x}}^{\dagger}\mathsf{K}\hat{\mathbf{x}} = \max_{\hat{\mathbf{x}}} \hat{\mathbf{x}}^{\dagger}\mathsf{S}^{\mathsf{T}}\mathsf{S}\hat{\mathbf{x}} = \max_{\hat{\mathbf{x}}} \|\mathsf{S}\hat{\mathbf{x}}\|^2 = \|\mathsf{S}\|^2 = \|\mathsf{K}\|. \tag{4.19}$$

4.2.2 Spatially IID Gaussian Fading

We next specialize the assumptions to the case where $\{\tilde{\mathbf{H}}_k\} = \{\mathbf{H}_k - \mathbf{d}\}$ is a spatially IID process where each component is a zero-mean unit-variance circularly symmetric complex Gaussian process of spectral distribution function $\mathsf{F}(\cdot)$. For this case we will now present a new derivation of the result (1.6) based on our new bounds.

Note that we cannot apply Corollary 9 here: even though $\{\tilde{\mathbf{H}}_k\}$ is isotropically distributed, $\{\mathbf{H}_k\}$ is not due to its mean vector $\mathbf{d}$.

However, the term $I\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0; \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big)$ does not depend on the particular choice of $\hat{\mathbf{x}}_\ell$:

$$\begin{aligned} I\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0; \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big) &= I\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0 - \mathbf{d}^{\mathsf{T}}\hat{\mathbf{x}}_0; \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell - \mathbf{d}^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big) && (4.20)\\ &= I\big(\tilde{\mathbf{H}}_0^{\mathsf{T}}\hat{\mathbf{x}}_0; \{\tilde{\mathbf{H}}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big) && (4.21)\\ &= I\big(\tilde{\mathbf{H}}_0^{\mathsf{T}}\hat{\mathbf{e}}; \{\tilde{\mathbf{H}}_\ell^{\mathsf{T}}\hat{\mathbf{e}}\}_{\ell=-\infty}^{-1}\big) && (4.22)\\ &= I\big(H_0^{(1)}; \{H_\ell^{(1)}\}_{\ell=-\infty}^{-1}\big) && (4.23)\\ &= \log\frac{1}{\epsilon^2}. && (4.24) \end{aligned}$$

Equation (1.6) now follows from (4.10), Theorem 7, and Theorem 8 by noting that

$$\max_{\|\hat{\mathbf{x}}\|=1} \frac{|\mathsf{E}[\mathbf{H}_k^{\mathsf{T}}]\,\hat{\mathbf{x}}|}{\sqrt{\mathsf{Var}\big(\mathbf{H}_k^{\mathsf{T}}\hat{\mathbf{x}}\big)}} = \max_{\|\hat{\mathbf{x}}\|=1} |\mathbf{d}^{\mathsf{T}}\hat{\mathbf{x}}| = \|\mathbf{d}\|, \tag{4.25}$$

where the maximum is achieved for $\hat{\mathbf{x}} = \mathbf{d}/\|\mathbf{d}\|$.

Chapter 5

Proof of Main Results

Note that the proofs are in part pretty technical. We therefore recommend the reader to firstly have a look at the overview given in Chapter 3.

5.1 Derivation of the Upper Bound of Theorem 7

Fix $\mathcal{E} > 0$, let the positive integer κ be arbitrary, and fix $\epsilon > 0$. Let $\eta = \eta(\mathcal{E}, \epsilon) \in \mathbb{Z}^+$ and $Q_{\mathcal{E},\epsilon} = Q(\mathcal{E}, \epsilon) \in \mathcal{P}(\mathbb{C}^{n_\mathrm{T}})$ be the integer and the input distribution[1] on $\mathbb{C}^{n_\mathrm{T}}$ whose existence is guaranteed in Lemma 6. Let blocklength n and input $\mathbf{X}_1^n$ satisfy (3.7)–(3.10) so that, in particular,

$$\begin{aligned} \mathsf{C}(\mathcal{E}) &\leq \frac{1}{n} I\big(\mathbf{X}_1^n; Y_1^n\big) + \epsilon && (5.1)\\ &= \frac{1}{n} \sum_{k=1}^{n} I\big(\mathbf{X}_1^n; Y_k \,\big|\, Y_1^{k-1}\big) + \epsilon. && (5.2) \end{aligned}$$

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

$$\begin{aligned} I\big(\mathbf{X}_1^n; Y_k \,\big|\, Y_1^{k-1}\big) &\leq I\big(\mathbf{X}_k; Y_k\big) + I\big(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}\big) && (5.3)\\ &\leq \mathsf{C}_{\mathrm{IID}}(\mathcal{E}) + I\big(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}\big), && (5.4) \end{aligned}$$

which is uniformly bounded in n. Here the second inequality follows from (3.9) and (3.10). The first inequality can be derived as follows:

$$\begin{aligned} I\big(\mathbf{X}_1^n; Y_k \,\big|\, Y_1^{k-1}\big) &= I\big(\mathbf{X}_1^n, Y_1^{k-1}; Y_k\big) - I\big(Y_k; Y_1^{k-1}\big) && (5.5)\\ &\leq I\big(\mathbf{X}_1^n, Y_1^{k-1}; Y_k\big) && (5.6)\\ &= I\big(\mathbf{X}_1^{k-1}, Y_1^{k-1}, \mathbf{X}_k; Y_k\big) && (5.7)\\ &\leq I\big(\mathbf{X}_1^{k-1}, Y_1^{k-1}, \mathbf{H}_1^{k-1}, \mathbf{X}_k; Y_k\big) && (5.8)\\ &= I\big(\mathbf{H}_1^{k-1}, \mathbf{X}_k; Y_k\big) && (5.9)\\ &= I(\mathbf{X}_k; Y_k) + I\big(\mathbf{H}_1^{k-1}; Y_k \,\big|\, \mathbf{X}_k\big) && (5.10)\\ &= I(\mathbf{X}_k; Y_k) + I\big(\mathbf{H}_1^{k-1}; \mathbf{X}_k, Y_k\big) && (5.11)\\ &\leq I(\mathbf{X}_k; Y_k) + I\big(\mathbf{H}_1^{k-1}; \mathbf{H}_k, \mathbf{X}_k, Y_k\big) && (5.12)\\ &= I(\mathbf{X}_k; Y_k) + I\big(\mathbf{H}_1^{k-1}; \mathbf{H}_k\big) && (5.13)\\ &\leq I(\mathbf{X}_k; Y_k) + I\big(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}\big). && (5.14) \end{aligned}$$

[1] Given an alphabet $\mathcal{A}$ we denote the set of all distributions over $\mathcal{A}$ by $\mathcal{P}(\mathcal{A})$.

Here the first equality follows from the chain rule; the subsequent inequality from the non-negativity of mutual information; the subsequent equality follows because we prohibit feedback; the subsequent inequality from the inclusion of the additional random vectors $\mathbf{H}_1^{k-1}$ in the mutual information term; (5.9) follows because, conditional on the past fading and the present input, the past inputs and outputs are independent of the present output Y_k; the subsequent equality follows from the chain rule; the following three steps are analogous to the first steps; and the last inequality follows once more from the inclusion of additional random vectors in the mutual information.

We conclude that

$$
\begin{aligned}
\mathsf{C}(\mathcal{E}) &\leq \varliminf_{n\uparrow\infty} \frac{1}{n} I\big(\mathbf{X}_1^n; Y_1^n\big) + \epsilon && (5.15)\\
&= \varliminf_{n\uparrow\infty} \frac{1}{n-\kappa-3(\eta-1)} \sum_{k=\eta+\kappa}^{n-2\eta+2} I\big(\mathbf{X}_1^n; Y_k \,\big|\, Y_1^{k-1}\big) + \epsilon. && (5.16)
\end{aligned}
$$

This allows us to focus on $\eta + \kappa \leq k \leq n - 2(\eta - 1)$ which guarantees that $\mathbf{X}_{k-\kappa}, \ldots, \mathbf{X}_k$ are each distributed according to $Q_{\mathcal{E},\epsilon}$.

We now continue by further upper-bounding $I\big(\mathbf{X}_1^n; Y_k \,\big|\, Y_1^{k-1}\big)$ for such k:

$$
\begin{aligned}
&I\big(\mathbf{X}_1^n; Y_k \,\big|\, Y_1^{k-1}\big)\\
&\quad= I\big(\mathbf{X}_1^n, Y_1^{k-1}; Y_k\big) - I\big(Y_k; Y_1^{k-1}\big) && (5.17)\\
&\quad\leq I\big(\mathbf{X}_1^n, Y_1^{k-1}; Y_k\big) && (5.18)\\
&\quad= I\big(\mathbf{X}_1^{k-1}, Y_1^{k-1}, \mathbf{X}_k; Y_k\big) && (5.19)\\
&\quad\leq I\big(\mathbf{X}_1^{k-1}, Y_1^{k-1}, \mathbf{H}_1^{k-\kappa-1}, \{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}, \mathbf{X}_k; Y_k\big) && (5.20)\\
&\quad= I\big(\mathbf{X}_{k-\kappa}^{k-1}, \mathbf{H}_1^{k-\kappa-1}, \{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}, \mathbf{X}_k; Y_k\big) && (5.21)\\
&\quad= I(\mathbf{X}_k; Y_k) + \underbrace{I\big(\mathbf{X}_{k-\kappa}^{k-1}; Y_k \,\big|\, \mathbf{X}_k\big)}_{=0} + I\big(\{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}; Y_k \,\big|\, \mathbf{X}_{k-\kappa}^k\big)\\
&\qquad + I\big(\mathbf{H}_1^{k-\kappa-1}; Y_k \,\big|\, \mathbf{X}_{k-\kappa}^k, \{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big) && (5.22)\\
&\quad= I(\mathbf{X}_k; Y_k) + I\big(\{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}; Y_k \,\big|\, \mathbf{X}_{k-\kappa}^k\big)\\
&\qquad + I\big(\mathbf{H}_1^{k-\kappa-1}; Y_k \,\big|\, \mathbf{X}_{k-\kappa}^k, \{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big), && (5.23)
\end{aligned}
$$

where the first three steps are identical to (5.5)–(5.7); (5.20) follows from the inclusion of the additional random vectors $\mathbf{H}_1^{k-\kappa-1}$ and the additional random variables $\{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}$ in the mutual information term; the subsequent equality follows because, conditional on the past terms $\mathbf{H}_1^{k-\kappa-1}, \{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}$, and on the present and past inputs $\mathbf{X}_{k-\kappa}^k$, the past outputs Y_1^{k-1} and the past inputs $\mathbf{X}_1^{k-\kappa-1}$ are independent of the present output Y_k; the subsequent equality follows from the chain rule; and the last equality from the independence of the past inputs and the present output when conditioning on the present input.

We continue by bounding the last term in (5.23):

$$
\begin{aligned}
&I\big(\mathbf{H}_1^{k-\kappa-1}; Y_k \,\big|\, \mathbf{X}_{k-\kappa}^k, \{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big)\\
&\quad= I\big(\mathbf{H}_1^{k-\kappa-1}; Y_k, \mathbf{X}_k \,\big|\, \mathbf{X}_{k-\kappa}^{k-1}, \{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big)\\
&\qquad - I\big(\mathbf{H}_1^{k-\kappa-1}; \mathbf{X}_k \,\big|\, \mathbf{X}_{k-\kappa}^{k-1}, \{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big) && (5.24)\\
&\quad\leq I\big(\mathbf{H}_1^{k-\kappa-1}; Y_k, \mathbf{X}_k \,\big|\, \mathbf{X}_{k-\kappa}^{k-1}, \{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big) && (5.25)\\
&\quad\leq I\big(\mathbf{H}_1^{k-\kappa-1}; Y_k, \mathbf{X}_k, \mathbf{H}_k \,\big|\, \mathbf{X}_{k-\kappa}^{k-1}, \{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big) && (5.26)
\end{aligned}
$$

$$= I\big(\mathbf{H}_1^{k-\kappa-1}; \mathbf{H}_k \,\big|\, \mathbf{X}_{k-\kappa}^{k-1}, \{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}\big) \tag{5.27}$$
$$= I\big(\mathbf{H}_1^{k-\kappa-1}; \mathbf{H}_k \,\big|\, \hat{\mathbf{X}}_{k-\kappa}^{k-1}, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}\big) \tag{5.28}$$
$$= \mathsf{E}\Big[I\big(\mathbf{H}_1^{k-\kappa-1}; \mathbf{H}_k \,\big|\, \{\hat{\mathbf{X}}_\ell = \hat{\mathbf{x}}_\ell\}_{\ell=k-\kappa}^{k-1}, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}_\ell\}_{\ell=k-\kappa}^{k-1}\big)\Big] \tag{5.29}$$
$$\leq \sup_{\hat{\mathbf{x}}_{k-\kappa}^{k-1}} I\big(\mathbf{H}_1^{k-\kappa-1}; \mathbf{H}_k \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}_\ell\}_{\ell=k-\kappa}^{k-1}\big) \tag{5.30}$$
$$= \sup_{\hat{\mathbf{e}}_{-\kappa}^{-1}} I\big(\mathbf{H}_{-k+1}^{-\kappa-1}; \mathbf{H}_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{e}}_\ell\}_{\ell=-\kappa}^{-1}\big) \tag{5.31}$$
$$\leq \sup_{\hat{\mathbf{e}}_{-\kappa}^{-1}} I\big(\mathbf{H}_{-\infty}^{-\kappa-1}; \mathbf{H}_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{e}}_\ell\}_{\ell=-\kappa}^{-1}\big) \tag{5.32}$$
$$\triangleq \delta(\kappa). \tag{5.33}$$

Here, the first equality follows from the chain rule; the subsequent inequality from the non-negativity of mutual information; the subsequent from inclusion of additional random vectors in the mutual information; the subsequent equality from the independence of the present input and output on the past fading when conditioned on the present fading; in the subsequent equality we introduce $\hat{\mathbf{X}} = \mathbf{X}/\|\mathbf{X}\|$; (5.31) follows from stationarity; and the subsequent inequality again from inclusion of additional terms into the mutual information.

Note that $\delta(\kappa)$ does not depend on k anymore and tends to zero as κ tends to infinity.

Hence, we continue with (5.23) as follows:

$$I\big(\mathbf{X}_1^n; Y_k \,\big|\, Y_1^{k-1}\big)$$
$$\leq I(\mathbf{X}_k; Y_k) + I\big(\{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}; Y_k \,\big|\, \mathbf{X}_{k-\kappa}^k\big) + \delta(\kappa) \tag{5.34}$$
$$\leq I(\mathbf{X}_k; Y_k) + I\big(\{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}; Y_k, \mathbf{H}_k^\mathsf{T}\mathbf{X}_k \,\big|\, \mathbf{X}_{k-\kappa}^k\big) + \delta(\kappa) \tag{5.35}$$
$$= I(\mathbf{X}_k; Y_k) + I\big(\{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}; \mathbf{H}_k^\mathsf{T}\mathbf{X}_k \,\big|\, \mathbf{X}_{k-\kappa}^k\big) + \underbrace{I\big(\{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}; Y_k \,\big|\, \mathbf{X}_{k-\kappa}^k, \mathbf{H}_k^\mathsf{T}\mathbf{X}_k\big)}_{=0} + \delta(\kappa) \tag{5.36}$$
$$= I(\mathbf{X}_k; Y_k) + I\big(\{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}; \mathbf{H}_k^\mathsf{T}\mathbf{X}_k \,\big|\, \mathbf{X}_{k-\kappa}^k\big) + \delta(\kappa) \tag{5.37}$$
$$= I(\mathbf{X}_k; Y_k) + I\big(\{\mathbf{H}_\ell^\mathsf{T}\mathbf{X}_\ell\}_{\ell=k-\kappa}^{k-1}; \mathbf{H}_k^\mathsf{T}\mathbf{X}_k \,\big|\, \{\|\mathbf{X}_\ell\|\}_{\ell=k-\kappa}^k, \hat{\mathbf{X}}_{k-\kappa}^k\big) + \delta(\kappa) \tag{5.38}$$
$$= I(\mathbf{X}_k; Y_k) + I\big(\{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1}; \mathbf{H}_k^\mathsf{T}\hat{\mathbf{X}}_k \,\big|\, \hat{\mathbf{X}}_{k-\kappa}^k\big) + \delta(\kappa). \tag{5.39}$$

Here (5.35) follows from the inclusion of the random variable $\mathbf{H}_k^\mathsf{T}\mathbf{X}_k$ in the mutual information term; the subsequent equality from chain rule; (5.37) follows because the additive noise Z_k is independent of the fading $\mathbf{H}_{k-\kappa}^{k-1}$; in the subsequent equality we introduce $\hat{\mathbf{X}}_\ell \triangleq \mathbf{X}_\ell/\|\mathbf{X}_\ell\|$; and the final equality follows from dividing each term by the magnitude of the input vectors.

Combined with (5.16) this yields

$$\mathsf{C}(\mathcal{E}) \leq \varliminf_{n\uparrow\infty} \frac{1}{n-\kappa-3(\eta-1)} \sum_{k=\eta+\kappa}^{n-2\eta+2} \bigg(I(\mathbf{X}_k; Y_k) + I\big(\mathbf{H}_k^\mathsf{T}\hat{\mathbf{X}}_k; \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=k-\kappa}^{k-1} \,\big|\, \hat{\mathbf{X}}_{k-\kappa}^k\big)\bigg) + \delta(\kappa) + \epsilon \tag{5.40}$$
$$= \varliminf_{n\uparrow\infty} \frac{1}{n-\kappa-3(\eta-1)} \sum_{k=\eta+\kappa}^{n-2\eta+2} \bigg(I(\mathbf{X}_k; \mathbf{H}_0^\mathsf{T}\mathbf{X}_k + Z_0)$$

$$+ I\Big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_k; \mathbf{H}_{-1}^\mathsf{T}\hat{\mathbf{X}}_{k-1}, \ldots, \mathbf{H}_{-\kappa}^\mathsf{T}\hat{\mathbf{X}}_{k-\kappa} \,\Big|\, \hat{\mathbf{X}}_{k-\kappa}^{k}\Big)\Bigg) + \delta(\kappa) + \epsilon \quad (5.41)$$

$$= I(\mathbf{X}_0; \mathbf{H}_0^\mathsf{T}\mathbf{X}_0 + Z_0) + \delta(\kappa) + \epsilon$$
$$+ \varliminf_{n\uparrow\infty} \frac{1}{n-\kappa-3(\eta-1)} \sum_{k=\eta+\kappa}^{n-2\eta+2} I\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_k; \mathbf{H}_{-1}^\mathsf{T}\hat{\mathbf{X}}_{k-1}, \ldots, \mathbf{H}_{-\kappa}^\mathsf{T}\hat{\mathbf{X}}_{k-\kappa} \,\big|\, \hat{\mathbf{X}}_{k-\kappa}^{k}\big). \quad (5.42)$$

Here the first equality follows from the stationarity of $\{\mathbf{H}_k, Z_k\}$; and the subsequent equality follows from the fact that for all $k \in \{\eta, \ldots, n-2\eta+2\}$ the distribution of $\mathbf{X}_k$ is $Q_{\mathcal{E},\epsilon}$ given in Lemma 6. Note that we have changed notation here: for notational convenience we will assume from now on that also $\mathbf{X}_0 \sim Q_{\mathcal{E},\epsilon}$.

We continue to upper-bound the first term $I(\mathbf{X}_0; Y_0)$ under the constraint that $\mathbf{X}_0 \sim Q_{\mathcal{E},\epsilon}$. Note that from Lemma 4 we now that $Q_{\mathcal{E},\epsilon}$ *escapes to infinity*, *i.e.*, $\Pr\big[\|\mathbf{X}_0\|^2 \le \mathcal{E}_{\text{low}}\big] = 0$ for some $\mathcal{E}_{\text{low}} \ge 0$.

$$\begin{aligned}
I(\mathbf{X}_0; Y_0) &\le I(\mathbf{X}_0; \mathbf{H}_0^\mathsf{T}\mathbf{X}_0 + Z_0, Z_0) &(5.43)\\
&= I(\mathbf{X}_0; Z_0) + I(\mathbf{X}_0; \mathbf{H}_0^\mathsf{T}\mathbf{X}_0 + Z_0 | Z_0) &(5.44)\\
&= I(\mathbf{X}_0; \mathbf{H}_0^\mathsf{T}\mathbf{X}_0 | Z_0) &(5.45)\\
&= I(\mathbf{X}_0; \mathbf{H}_0^\mathsf{T}\mathbf{X}_0). &(5.46)
\end{aligned}$$

We will now upper-bound this term by the bound given in Lemma 5:

$$\begin{aligned}
I(\mathbf{X}_0; \mathbf{H}_0^\mathsf{T}\mathbf{X}_0) \le{}& -h_{Q_{\mathcal{E},\epsilon}}\big(\mathbf{H}_0^\mathsf{T}\mathbf{X}_0 \big| \mathbf{X}_0\big) + \log\pi + \alpha\log\beta + \log\Gamma(\alpha, \nu/\beta)\\
&+ (1-\alpha)\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\log\big(|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2 + \nu\big)\big] + \frac{1}{\beta}\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2\big] + \frac{\nu}{\beta}, \quad (5.47)
\end{aligned}$$

where $\alpha, \beta > 0$, and $\nu \ge 0$ can be chosen freely. We fix these parameters and assume $0 < \alpha < 1$ such that $1 - \alpha > 0$. Then define

$$\epsilon_\nu \triangleq \sup_{\|\mathbf{x}\|^2 \ge \mathcal{E}_{\text{low}}} \big\{\mathsf{E}\big[\log\big(|\mathbf{H}_0^\mathsf{T}\mathbf{x}|^2 + \nu\big)\big] - \mathsf{E}\big[\log|\mathbf{H}_0^\mathsf{T}\mathbf{x}|^2\big]\big\}. \quad (5.48)$$

Then

$$\begin{aligned}
&(1-\alpha)\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\log\big(|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2 + \nu\big)\big]\\
&\quad= (1-\alpha)\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\log|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2\big]\\
&\qquad+ (1-\alpha)\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\log\big(|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2 + \nu\big)\big] - (1-\alpha)\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\log|\mathbf{H}_0\mathbf{X}_0|^2\big] &(5.49)\\
&\quad\le (1-\alpha)\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\log|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2\big]\\
&\qquad+ (1-\alpha)\sup_{\|\mathbf{x}\|^2 \ge \mathcal{E}_{\text{low}}} \big\{\mathsf{E}\big[\log\big(|\mathbf{H}_0^\mathsf{T}\mathbf{x}|^2 + \nu\big)\big] - \mathsf{E}\big[\log|\mathbf{H}_0\mathbf{x}|^2\big]\big\} &(5.50)\\
&\quad= (1-\alpha)\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\log|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2\big] + (1-\alpha)\epsilon_\nu &(5.51)\\
&\quad\le (1-\alpha)\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\log|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2\big] + \epsilon_\nu. &(5.52)
\end{aligned}$$

Plugging this into (5.47) yields

$$\begin{aligned}
&I(\mathbf{X}_0; \mathbf{H}_0^\mathsf{T}\mathbf{X}_0)\\
&\quad\le \mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\mathsf{E}\big[\log|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2 \,\big|\, \mathbf{X}_0 = \mathbf{x}_0\big] - h\big(\mathbf{H}_0^\mathsf{T}\mathbf{X}_0 \big| \mathbf{X}_0 = \mathbf{x}_0\big)\big]\\
&\qquad+ \log\pi + \log\Gamma(\alpha, \nu/\beta) + \epsilon_\nu\\
&\qquad+ \alpha\Big(\log\beta - \mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\log|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2\big]\Big) + \frac{1}{\beta}\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2\big] + \frac{\nu}{\beta} &(5.53)
\end{aligned}$$

$$
\begin{aligned}
&= \mathsf{E}_{Q_{\mathcal{E},\epsilon}}\Big[\mathsf{E}\Big[\log|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0|^2 \Big| \hat{\mathbf{X}}_0 = \hat{\mathbf{x}}_0\Big] - h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0\big|\hat{\mathbf{X}}_0 = \hat{\mathbf{x}}_0\big)\Big] \\
&\quad + \log\pi + \log\Gamma(\alpha,\nu/\beta) + \epsilon_\nu \\
&\quad + \alpha\Big(\log\beta - \mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\log|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2\big]\Big) + \frac{1}{\beta}\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\Big[|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0|^2\cdot\|\mathbf{X}_0\|^2\Big] + \frac{\nu}{\beta} && (5.54)\\
&\le \mathsf{E}_{Q_{\mathcal{E},\epsilon}}\Big[\log|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0|^2\Big] - h_{Q_{\mathcal{E},\epsilon}}\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0\big|\hat{\mathbf{X}}_0\big) + \log\pi + \log\Gamma(\alpha,\nu/\beta) + \epsilon_\nu \\
&\quad + \alpha\Big(\log\beta - \inf_{\|\mathbf{x}\|^2\ge\mathcal{E}_{\text{low}}}\mathsf{E}\big[\log|\mathbf{H}_0^\mathsf{T}\mathbf{x}|^2\big]\Big) + \frac{1}{\beta}\sup_{\|\hat{\mathbf{x}}\|=1}\mathsf{E}\big[|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}}|^2\big]\cdot\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\|\mathbf{X}_0\|^2\big] \\
&\quad + \frac{\nu}{\beta} && (5.55)\\
&= \mathsf{E}_{Q_{\mathcal{E},\epsilon}}\Big[\log|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0|^2\Big] - h_{Q_{\mathcal{E},\epsilon}}\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0\big|\hat{\mathbf{X}}_0\big) + \log\pi + \log\Gamma(\alpha,\nu/\beta) + \epsilon_\nu \\
&\quad + \alpha\Big(\log\beta - \log\mathcal{E}_{\text{low}} - \xi\Big) + \frac{1}{\beta}\sup_{\|\hat{\mathbf{x}}\|=1}\mathsf{E}\big[|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}}|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta}. && (5.56)
\end{aligned}
$$

Here, the first inequality follows from (5.47) using bound (5.52); in the subsequent equality we split $\mathbf{X}_0$ up into its magnitude $\|\mathbf{X}_0\|$ and direction $\hat{\mathbf{X}}_0$ and used the scaling property of differential entropy of complex random variables; and the final equality follows from the following definition:

$$
\begin{aligned}
\mathsf{E}_{Q_{\mathcal{E},\epsilon}}\big[\log|\mathbf{H}_0^\mathsf{T}\mathbf{X}_0|^2\big] &\ge \inf_{\|\mathbf{x}\|^2\ge\mathcal{E}_{\text{low}}}\mathsf{E}\big[\log|\mathbf{H}_0^\mathsf{T}\mathbf{x}|^2\big] && (5.57)\\
&= \log\mathcal{E}_{\text{low}} + \inf_{\|\hat{\mathbf{x}}\|=1}\mathsf{E}\big[\log|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}}|^2\big] && (5.58)\\
&\triangleq \log\mathcal{E}_{\text{low}} + \xi. && (5.59)
\end{aligned}
$$

Here the last line should be taken as a definition for ξ. Notice that

$$
-\infty < \xi < \infty \tag{5.60}
$$

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

Plugging (5.56) into (5.42) yields the following bound on capacity:

$$
\begin{aligned}
\mathsf{C}(\mathcal{E}) &\le \mathsf{E}\Big[\log|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0|^2\Big] - h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0\big|\hat{\mathbf{X}}_0\big) + \log\pi + \log\Gamma(\alpha,\nu/\beta) + \epsilon_\nu \\
&\quad + \alpha\Big(\log\beta - \log\mathcal{E}_{\text{low}} - \xi\Big) + \frac{1}{\beta}\sup_{\|\hat{\mathbf{x}}\|=1}\mathsf{E}\big[|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}}|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta} + \delta(\kappa) + \epsilon \\
&\quad + \varliminf_{n\uparrow\infty}\frac{1}{n-\kappa-3(\eta-1)}\sum_{k=\eta+\kappa}^{n-2\eta+2} I\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_k;\mathbf{H}_{-1}^\mathsf{T}\hat{\mathbf{X}}_{k-1},\ldots,\mathbf{H}_{-\kappa}^\mathsf{T}\hat{\mathbf{X}}_{k-\kappa}\,\big|\,\hat{\mathbf{X}}_{k-\kappa}^{k}\big).
\end{aligned}
\tag{5.61}
$$

Note that in this upper bound there are no terms that depend simultaneously on $\hat{\mathbf{X}}_k$ and $\|\mathbf{X}_k\|$. Hence, even if according to the (unknown) capacity-achieving input distribution there is a dependence between direction and magnitude, this dependence has no influence on this upper bound. We may therefore, without loss of generality, assume that $\{\hat{\mathbf{X}}_k\} \perp\!\!\!\perp \{\|\mathbf{X}_k\|\}$. Therefore we may assume that $\{\hat{\mathbf{X}}_k\}$ does not depend on the value of $\mathcal{E}$.

Next we use this bound in order to get an upper bound on the fading number of

MISO fading channels with memory:

$$\begin{aligned}
&\chi(\{\mathbf{H}_k^{\mathsf{T}}\}) \\
&\quad = \varlimsup_{\mathcal{E}\uparrow\infty} \left\{ \mathsf{C}(\mathcal{E}) - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} && (5.62)\\
&\quad \le \varlimsup_{\mathcal{E}\uparrow\infty} \Bigg\{ \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|\hat{\mathbf{X}}_0\big) + \log\pi + \epsilon_\nu + \delta(\kappa) + \epsilon \\
&\qquad\qquad + \lim_{n\uparrow\infty} \frac{1}{n-\kappa-3(\eta-1)} \sum_{k=\eta+\kappa}^{n-2\eta+2} I\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{X}}_k; \mathbf{H}_{-1}^{\mathsf{T}}\hat{\mathbf{X}}_{k-1},\ldots,\mathbf{H}_{-\kappa}^{\mathsf{T}}\hat{\mathbf{X}}_{k-\kappa}\,\big|\,\hat{\mathbf{X}}_{k-\kappa}^{k}\big) \\
&\qquad\qquad + \log\Gamma(\alpha,\nu/\beta) + \alpha\Big(\log\beta - \log\mathcal{E}_{\text{low}} - \xi\Big) + \frac{1}{\beta}\sup_{\|\hat{\mathbf{x}}\|=1}\mathsf{E}\big[|\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta} \\
&\qquad\qquad - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\Bigg\} && (5.63)\\
&\quad = \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|\hat{\mathbf{X}}_0\big) + \log\pi + \epsilon_\nu + \delta(\kappa) + \epsilon \\
&\qquad + \lim_{n\uparrow\infty} \frac{1}{n-\kappa-3(\eta-1)} \sum_{k=\eta+\kappa}^{n-2\eta+2} I\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{X}}_k; \mathbf{H}_{-1}^{\mathsf{T}}\hat{\mathbf{X}}_{k-1},\ldots,\mathbf{H}_{-\kappa}^{\mathsf{T}}\hat{\mathbf{X}}_{k-\kappa}\,\big|\,\hat{\mathbf{X}}_{k-\kappa}^{k}\big) \\
&\qquad + \varlimsup_{\mathcal{E}\uparrow\infty} \Bigg\{ \log\Gamma(\alpha,\nu/\beta) - \log\frac{1}{\alpha} + \alpha\Big(\log\beta - \log\mathcal{E}_{\text{low}} - \xi\Big) \\
&\qquad\qquad + \frac{1}{\beta}\sup_{\|\hat{\mathbf{x}}\|=1}\mathsf{E}\big[|\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta} \\
&\qquad\qquad + \log\frac{1}{\alpha} - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\Bigg\} && (5.64)\\
&\quad = \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|\hat{\mathbf{X}}_0\big) + \log\pi + \epsilon_\nu + \delta(\kappa) + \epsilon \\
&\qquad + \lim_{n\uparrow\infty} \frac{1}{n-\kappa-3(\eta-1)} \sum_{k=\eta+\kappa}^{n-2\eta+2} I\big(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{X}}_k; \mathbf{H}_{-1}^{\mathsf{T}}\hat{\mathbf{X}}_{k-1},\ldots,\mathbf{H}_{-\kappa}^{\mathsf{T}}\hat{\mathbf{X}}_{k-\kappa}\,\big|\,\hat{\mathbf{X}}_{k-\kappa}^{k}\big) \\
&\qquad + \log\big(1-e^{-\nu}\big) + \nu - \log\nu. && (5.65)
\end{aligned}$$

Here, (5.64) follows because, as mentioned above, we may assume that

$$\{\hat{\mathbf{X}}_k\} \perp\!\!\!\perp \{\|\mathbf{X}_k\|\}, \tag{5.66}$$

and from combining the mutual information and the differential entropy terms; in the last equation we have made the following choices on the free parameters α and β:

$$\alpha \triangleq \alpha(\mathcal{E}) = \frac{\nu}{\log\mathcal{E} + \log\sup_{\|\hat{\mathbf{x}}\|=1}\mathsf{E}[|\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}|^2]}; \tag{5.67}$$

$$\beta \triangleq \beta(\mathcal{E}) = \frac{1}{\alpha(\mathcal{E})}e^{\nu/\alpha}. \tag{5.68}$$

For this choice note that

$$\varlimsup_{\mathcal{E}\uparrow\infty} \left\{\log\Gamma(\alpha,\nu/\beta) - \log\frac{1}{\alpha}\right\} = \log\big(1-e^{-\nu}\big); \tag{5.69}$$

$$\varlimsup_{\mathcal{E}\uparrow\infty} \alpha\Big(\log\beta - \log\mathcal{E}_{\text{low}} - \xi\Big) = \nu; \tag{5.70}$$

$$\varlimsup_{\mathcal{E}\uparrow\infty}\left\{\frac{1}{\beta}\sup_{\|\hat{\mathbf{x}}\|=1}\mathsf{E}\big[|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}}|^2\big]\cdot\mathcal{E}+\frac{\nu}{\beta}\right\}=0; \tag{5.71}$$

$$\varlimsup_{\mathcal{E}\uparrow\infty}\left\{\log\frac{1}{\alpha}-\log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\right\}=-\log\nu. \tag{5.72}$$

(Compare with [7, Appendix VII], [2, Sec. B.5.9].)

To finish the derivation of the upper bound, we upper-bound the only term that still depends on n as follows:

$$\begin{aligned}
&\varliminf_{n\uparrow\infty}\frac{1}{n-\kappa-3(\eta-1)}\sum_{k=\eta+\kappa}^{n-2\eta+2} I\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_k;\mathbf{H}_{-1}^\mathsf{T}\hat{\mathbf{X}}_{k-1},\ldots,\mathbf{H}_{-\kappa}^\mathsf{T}\hat{\mathbf{X}}_{k-\kappa}\,\big|\,\hat{\mathbf{X}}_{k-\kappa}^{k}\big)\\
&\le \varliminf_{n\uparrow\infty}\frac{1}{n-\kappa-3(\eta-1)}\sum_{k=\eta+\kappa}^{n-2\eta+2}\sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^0}\\ \text{with marg. } Q_{\mathcal{E},\epsilon}}} I\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0;\{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,\hat{\mathbf{X}}_{-\kappa}^0\big) &(5.73)\\
&= \varliminf_{n\uparrow\infty}\left\{\sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^0}\\ \text{with marg. } Q_{\mathcal{E},\epsilon}}} I\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0;\{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,\hat{\mathbf{X}}_{-\kappa}^0\big)\cdot\frac{1}{n-\kappa-3(\eta-1)}\sum_{k=\eta+\kappa}^{n-2\eta+2}1\right\} &(5.74)\\
&= \sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^0}\\ \text{with marg. } Q_{\mathcal{E},\epsilon}}} I\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0;\{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,\hat{\mathbf{X}}_{-\kappa}^0\big). &(5.75)
\end{aligned}$$

Here again we use the fact given in Lemma 6 that the marginal distribution of $\{\hat{\mathbf{X}}_k\}$ is $Q_{\mathcal{E},\epsilon}$. We therefore are allowed to change notation accordingly and use the same $\hat{\mathbf{X}}_0$ as introduced above.

Next, we let ν go to zero. Note that $\epsilon_\nu \to 0$ as $\nu \downarrow 0$ as can be seen from (5.48). Note further that

$$\lim_{\nu\downarrow 0}\big\{\log\big(1-e^{-\nu}\big)-\log\nu\big\}=0. \tag{5.76}$$

Therefore, we get

$$\begin{aligned}
\chi(\{\mathbf{H}_k^\mathsf{T}\}) &\le \mathsf{E}\Big[\log|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0|^2\Big]-h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0\big|\hat{\mathbf{X}}_0\big)+\log\pi+\delta(\kappa)+\epsilon\\
&\quad+\sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^0}\\ \text{with marg. } Q_{\mathcal{E},\epsilon}}} I\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0;\{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,\hat{\mathbf{X}}_{-\kappa}^0\big) &(5.77)\\
&= \sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^0}\\ \text{with marg. } Q_{\mathcal{E},\epsilon}}}\Big\{\mathsf{E}\Big[\log|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0|^2\Big]-h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0\big|\hat{\mathbf{X}}_0\big)+\log\pi+\delta(\kappa)+\epsilon\\
&\qquad+I\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0;\{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}\,\big|\,\hat{\mathbf{X}}_{-\kappa}^0\big)\Big\} &(5.78)\\
&= \sup_{\substack{Q_{\hat{\mathbf{X}}_{-\kappa}^0}\\ \text{with marg. } Q_{\mathcal{E},\epsilon}}}\Big\{\log\pi+\mathsf{E}\Big[\log|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0|^2\Big]\\
&\qquad-h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0\,\big|\,\{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1},\hat{\mathbf{X}}_{-\kappa}^0\big)\Big\}+\delta(\kappa)+\epsilon &(5.79)
\end{aligned}$$

$$= \sup_{\substack{Q_{\hat{\mathbf{X}}^0_{-\kappa}} \\ \text{with marg. } Q_{\mathcal{E},\epsilon}}} \mathsf{E}_{Q_{\hat{\mathbf{X}}^0_{-\kappa}}} \Bigg[\log \pi + \mathsf{E}\Big[\log |\mathbf{H}_0^\mathsf{T} \hat{\mathbf{X}}_0|^2 \Big| \hat{\mathbf{X}}_0 = \hat{\mathbf{x}}_0 \Big]$$

$$- h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=-\kappa}^{-1}, \hat{\mathbf{X}}^0_{-\kappa} = \hat{\mathbf{x}}^0_{-\kappa}\big) \Bigg] + \delta(\kappa) + \epsilon \tag{5.80}$$

$$\le \sup_{\hat{\mathbf{x}}^0_{-\kappa}} \Big\{ \log \pi + \mathsf{E}\big[\log |\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}}_0|^2\big] - h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}}_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}_\ell\}_{\ell=-\kappa}^{-1}\big) \Big\} + \delta(\kappa) + \epsilon. \tag{5.81}$$

Here, (5.78) follows since the first couple of terms only depend on the marginal distribution $Q_{\mathcal{E},\epsilon}$ which is kept constant for the maximization; the subsequent equality follows from the definition of mutual information; and in the last inequality the expectation is upper-bounded by the supremum.

Finally, we let κ go to infinity. The result now follows because ϵ is arbitrary and because $\delta(\kappa)$ tends to zero for $\kappa \to \infty$.

5.2 Derivation of the Lower Bound of Theorem 8

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

$$\mathbf{X}_k = X_k \cdot \hat{\mathbf{x}} \tag{5.82}$$

where $\hat{\mathbf{x}}$ is a deterministic unit vector (which is therefore known to both the receiver and transmitter) and where $\{X_k\}$ is an IID circularly symmetric random process with

$$\log |X_k|^2 \sim \mathcal{U}\left([\log x^2_{\min}, \log \mathcal{E}]\right), \tag{5.83}$$

where we choose $x^2_{\min}$ as

$$x^2_{\min} = \log \mathcal{E}. \tag{5.84}$$

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

$$\frac{1}{n} I\big(\mathbf{X}_1^n; Y_1^n\big) = \frac{1}{n}\sum_{k=1}^{n} I\big(\mathbf{X}_k; Y_1^n \,\big|\, \mathbf{X}_1^{k-1}\big) \tag{5.85}$$

$$\ge \frac{1}{n}\sum_{k=\kappa+1}^{n-\kappa} I\big(\mathbf{X}_k; Y_1^n \,\big|\, \mathbf{X}_1^{k-1}\big). \tag{5.86}$$

Then for every $\kappa+1 \le k \le n-\kappa$, we can use the fact that $\{X_k\}$ is IID and circularly symmetric to lower-bound $I\big(\mathbf{X}_k; Y_1^n \,\big|\, \mathbf{X}_1^{k-1}\big)$ as follows:

$$I\big(\mathbf{X}_k; Y_1^n \,\big|\, \mathbf{X}_1^{k-1}\big)$$

$$= I\big(X_k\hat{\mathbf{x}}; Y_1^n \,\big|\, \{X_\ell\hat{\mathbf{x}}\}_{\ell=1}^{k-1}\big) \tag{5.87}$$

$$= I\big(X_k; Y_1^n \,\big|\, X_1^{k-1}\big) \tag{5.88}$$

$$= I\big(X_k; X_1^{k-1}, Y_1^n\big) \tag{5.89}$$

$$\ge I\big(X_k; X_{k-\kappa}^{k-1}, Y_{k-\kappa}^{k-1}, Y_k\big) \tag{5.90}$$

$$= I\big(X_k; X_{k-\kappa}^{k-1}, Y_{k-\kappa}^{k-1}, Z_{k-\kappa}^{k-1}, Y_k\big) - \underbrace{I\big(X_k; Z_{k-\kappa}^{k-1} \,\big|\, X_{k-\kappa}^{k-1}, Y_{k-\kappa}^{k-1}, Y_k\big)}_{\le \epsilon(x_{\min},\kappa)} \tag{5.91}$$

$$\geq I\big(X_k; X_{k-\kappa}^{k-1}, Y_{k-\kappa}^{k-1}, Z_{k-\kappa}^{k-1}, Y_k\big) - \epsilon(x_{\min}, \kappa) \tag{5.92}$$
$$= I\big(X_k; \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=k-\kappa}^{k-1}, Y_k\big) - \epsilon(x_{\min}, \kappa) \tag{5.93}$$
$$= I\big(X_0; \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=-\kappa}^{-1}, Y_0\big) - \epsilon(x_{\min}, \kappa) \tag{5.94}$$
$$= I\big(X_0; Y_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=-\kappa}^{-1}\big) - \epsilon(x_{\min}, \kappa), \qquad \kappa+1 \leq k \leq n-\kappa. \tag{5.95}$$

Here the second equality follows because $\{X_k\}$ is chosen to be IID; in (5.90) we drop some arguments which reduces the mutual information; next we use the chain rule; in the subsequent inequality we lower-bound the second term by $-\epsilon(x_{\min}, \kappa)$ which is defined in Appendix B and is shown there to only depend on $x_{\min}$ and κ and to tend to zero as $x_{\min} \uparrow \infty$; in the subsequent equality we use $X_{k-\kappa}^{k-1}$ and $Z_{k-\kappa}^{k-1}$ in order to extract $\{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=k-\kappa}^{k-1}$ from $Y_{k-\kappa}^{k-1}$ and then drop $(\{X_\ell, Y_\ell, Z_\ell\}_{\ell=k-\kappa}^{k-1})$ since given $\{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=k-\kappa}^{k-1}$ it is independent of the other random variables; and the equality before last follows from stationarity.

Plugging (5.95) into (5.86) we get

$$\frac{1}{n} I\big(\mathbf{X}_1^n; Y_1^n\big) \geq \frac{1}{n} \sum_{k=\kappa+1}^{n-\kappa} \Big(I\big(X_0; Y_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=-\kappa}^{-1}\big) - \epsilon(x_{\min}, \kappa)\Big) \tag{5.96}$$
$$= \left(1 - \frac{2\kappa}{n}\right) \Big(I\big(X_0; Y_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=-\kappa}^{-1}\big) - \epsilon(x_{\min}, \kappa)\Big). \tag{5.97}$$

Letting n tend to infinity we obtain

$$\mathsf{C}(\mathcal{E}) \geq I\big(X_0; Y_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=-\kappa}^{-1}\big) - \epsilon(x_{\min}, \kappa) \tag{5.98}$$

where the first term on the RHS can be viewed as mutual information across a memoryless SISO fading channel with fading $H = \mathbf{H}_{\kappa+1}^\mathsf{T}\hat{\mathbf{x}}$ in the presence of the side-information $\{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=1}^{\kappa}$.

We next let the power grow to infinity $\mathcal{E} \to \infty$. Since the circularly symmetric law (5.83) achieves the fading number of IID SISO fading with side-information [7, Proposition 4.23], [2, Proposition 6.23] and since our choice (5.84) guarantees that $\epsilon(x_{\min}, \kappa)$ tend to zero as $\mathcal{E} \to \infty$ (see Appendix B) we obtain the bound

$$\chi(\{\mathbf{H}_k^\mathsf{T}\}) \geq \chi_{\text{IID}}\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}} \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=-\kappa}^{-1}\big) \tag{5.99}$$
$$= \log \pi + \mathsf{E}\big[\log |\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}}|^2\big] - h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{x}} \big| \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{x}}\}_{\ell=-\kappa}^{-1}\big). \tag{5.100}$$

Finally, we let κ go to infinity. The result now follows from choosing $\hat{\mathbf{x}}$ such as to maximize this lower bound to the fading number.

Chapter 6

Discussion & Conclusion

We have derived a new upper bound and a new lower bound on the fading number of a MISO fading channel of general law including memory. The fading number is the second term in the asymptotic expansion of channel capacity, *i.e.*, the fading number basically determines the capacity in the limit of infinite power.

The bounds are not identical, however, both bounds show the same structure involving the maximization of a deterministic beam-direction $\hat{\mathbf{x}}$, which suggests that beam-forming is optimal at high SNR. Be aware, however, that the beam-direction is not chosen to maximize the SNR, but to maximize the fading number.

The differences between the upper and lower bound lies in the details of the maximization: while in the lower bound one single direction unit vector $\hat{\mathbf{x}}$ is chosen for all time, the upper bound allows for different realizations of $\hat{\mathbf{x}}_k$ for different times k.

We are convinced that the lower bound is actually tight: intuition tells that for our stationary channel model a stationary input should be sufficient for achieving the capacity. As a matter of fact in the SISO and SIMO case it has been shown that actually an IID input suffices to achieve capacity at high SNR [7], [2], [3]. Furthermore, in the derivation of the upper bound we use obviously loose bounds on several places.

In the case of isotropically distributed fading the particular choice of direction has no influence on the fading process and therefore the upper and lower bounds coincide. Hence, we are able to specify the fading number of isotropically distributed MISO fading processes precisely.

In the important special case of Gaussian fading we could show that the bounds presented in [15] and [16] are special cases of the new bounds presented here, where the new upper bound (3.11) is in general tighter than (1.7).

Note that in the derivation of the upper bound it is tempting to use the known results about memoryless MISO fading (*e.g.*, in (5.42)). Unfortunately, this approach fails because the memoryless MISO fading number involves a supremum over the memoryless terms which afterward can not be incorporated anymore into a supremum over all terms. Another, so far unsuccessful attempt, has been to reduce the problem to a memoryless SISO situation. This leads to additional terms that take into account the direction of the MISO input, however, we have not been able to manipulate these terms such as to keep the upper bound tight.

We also would like to emphasize the importance of the preliminary results of Section 3.1, particularly, the concept of distributions that escape to infinity and the lemma about the stationarity of the capacity achieving input distribution. They give some additional information about the capacity achieving input distribution that turned out to be crucial in the derivation of the bounds. We are convinced that the more complex the channel model gets, the more one will need to rely on such auxiliary results.

In future we are going to try to improve these results further aiming at the precise derivation of the MISO fading number with memory. This will be a big step forward to our ultimate goal: the exact expression of the fading number of the most general fading channel model, *i.e.*, general MIMO fading with memory. Note that so far the fading number for fading with memory is known only in situations where the transmitter has just one antenna. Hence the importance of MISO, being the simplest non-trivial situation with multiple antennas at the transmitter side.

Appendix A

Proof of Lemma 6

The proof follows the same lines as the proofs of [3, Lemma 5] and [2, Lemma B.1].

The proof is by a simple shift-and-mix argument. Recalling that

$$\mathsf{C}(\mathcal{E}) = \lim_{n\uparrow\infty} \frac{1}{n} \sup I(\mathbf{X}_1, \ldots, \mathbf{X}_n; Y_1, \ldots, Y_n) \tag{A.1}$$

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 $p^* \in \mathcal{P}(\mathbb{C}^{n_\mathrm{T}\times\eta})$ such that if $(\mathbf{X}_1, \ldots, \mathbf{X}_\eta) \sim p^*$ then

$$\frac{1}{\eta} \sum_{\ell=1}^{\eta} \mathsf{E}\left[\|\mathbf{X}_\ell\|^2\right] = \mathcal{E} \tag{A.2}$$

and

$$\frac{1}{\eta} I(\mathbf{X}_1, \ldots, \mathbf{X}_\eta; Y_1, \ldots, Y_\eta) > \mathsf{C}(\mathcal{E}) - \frac{\epsilon}{2}. \tag{A.3}$$

Let Q be the probability law on $\mathbb{C}^{n_\mathrm{T}}$ that is the mixture of the η different marginals of p^*. That is, for every Borel set $\mathcal{B} \subset \mathbb{C}^{n_\mathrm{T}}$

$$Q(\mathcal{B}) = \frac{1}{\eta} \sum_{\ell=1}^{\eta} p^*(\mathbf{X}_\ell \in \mathcal{B}). \tag{A.4}$$

By (A.2) we have

$$\int_{\mathbb{C}^{n_\mathrm{T}}} \|\mathbf{x}\|^2 \,\mathrm{d}Q(\mathbf{x}) = \mathcal{E}. \tag{A.5}$$

Let n now be given. We shall next describe the required input distribution as follows. Let

$$\nu = \left\lfloor \frac{n - \eta + 1}{\eta} \right\rfloor \tag{A.6}$$

and let the infinite sequence $\tilde{\mathbb{X}}$ of random n_T-vectors be defined by

$$\tilde{\mathbb{X}} = (\underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{\eta-1}, \underbrace{\mathbf{\Xi}_1^{(1)}, \ldots, \mathbf{\Xi}_\eta^{(1)}}_{\eta}, \ldots, \ldots, \underbrace{\mathbf{\Xi}_1^{(\nu)}, \ldots, \mathbf{\Xi}_\eta^{(\nu)}}_{\eta}, \mathbf{0}, \mathbf{0}, \ldots \tag{A.7}$$

so that

$$\tilde{\mathbf{X}}_\ell = \begin{cases} \mathbf{0} & \text{if } 1 \leq \ell \leq \eta - 1, \\ \mathbf{\Xi}_{(\ell \mod \eta)+1}^{\lfloor \ell/\eta \rfloor} & \text{if } \eta \leq \ell \leq (\nu+1)\eta - 1, \\ \mathbf{0} & \text{if } \ell \geq (\nu+1)\eta, \end{cases} \tag{A.8}$$

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

$$\left\{\left(\mathbf{\Xi}_1^{(j)},\ldots,\mathbf{\Xi}_\eta^{(j)}\right)\right\}_{j=1}^{\nu} \sim \text{IID } p^*. \tag{A.9}$$

Notice that since the lead-in and trailing zeros have no effect on our channel, the unnormalized mutual information induced by $\tilde{\mathbb{X}}$ is lower-bounded by $\nu\eta(\mathsf{C}(\mathcal{E})-\epsilon/2)$. Again, since the lead-in and trailing zeros are of no consequence, this same mutual information results if we shift $\tilde{\mathbb{X}}$ by t, (provided that $0 \leq t \leq \eta-1$). Consequently, if we define $\mathbf{X}_1,\ldots,\mathbf{X}_n$ by the mixture of the time shift of $\tilde{\mathbb{X}}$, *i.e.*,

$$\mathbf{X}_\ell = \tilde{\mathbf{X}}_{\ell+T}, \qquad 1 \leq \ell \leq n, \tag{A.10}$$

where

$$T \sim \mathcal{U}\left(\{0,\ldots,\eta-1\}\right) \tag{A.11}$$

is independent of $\tilde{\mathbb{X}}$, then by the concavity of mutual information in the input distribution we obtain that the unnormalized mutual information induced by $\mathbf{X}_1^n$ is lower-bounded by $\nu\eta(\mathsf{C}(\mathcal{E})-\epsilon/2)$, so that the normalized mutual information satisfies

$$\frac{1}{n} I\left(\mathbf{X}_1^n; Y_1^n\right) \geq \frac{\eta\nu}{n}\left(\mathsf{C}(\mathcal{E}) - \frac{\epsilon}{2}\right) \tag{A.12}$$

$$= \frac{\eta\left\lfloor\frac{n-\eta+1}{\eta}\right\rfloor}{n}\left(\mathsf{C}(\mathcal{E}) - \frac{\epsilon}{2}\right), \tag{A.13}$$

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

Except at the edges, the above mixture guarantees equal marginals of average power $\mathcal{E}$. The power in the edges can be smaller than $\mathcal{E}$ because of the mixture with deterministic zero vectors.

Appendix B

Additional Derivation for the Proof of the Lower Bound

In the derivation of the lower bound to the fading number we need to find the following upper bound

$$I\bigl(X_k; Z_{k-\kappa}^{k-1} \bigm| X_{k-\kappa}^{k-1} Y_{k-\kappa}^{k-1}, Y_k\bigr) \le \epsilon(x_{\min}, \kappa) \tag{B.1}$$

and to show that $\epsilon(x_{\min}, \kappa)$ only depends on $x_{\min}$ and κ and tends to zero as $x_{\min}$ tends to infinity.

To that goal we bound as follows:

$$\begin{aligned}
&I\bigl(X_k; Z_{k-\kappa}^{k-1} \bigm| X_{k-\kappa}^{k-1}, Y_{k-\kappa}^{k-1}, Y_k\bigr) \\
&\quad = h\bigl(Z_{k-\kappa}^{k-1} \bigm| X_{k-\kappa}^{k-1}, Y_{k-\kappa}^{k-1}, Y_k\bigr) - h\bigl(Z_{k-\kappa}^{k-1} \bigm| X_{k-\kappa}^{k-1}, Y_{k-\kappa}^{k-1}, X_k, Y_k\bigr) && \text{(B.2)} \\
&\quad \le h\bigl(Z_{k-\kappa}^{k-1}\bigr) - h\bigl(Z_{k-\kappa}^{k-1} \bigm| X_{k-\kappa}^{k-1}, Y_{k-\kappa}^{k-1}, X_k, Y_k, Z_k\bigr) && \text{(B.3)} \\
&\quad = h\bigl(Z_{k-\kappa}^{k-1}\bigr) - h\bigl(Z_{k-\kappa}^{k-1} \bigm| X_{k-\kappa}^{k-1}, Y_{k-\kappa}^{k-1}, \mathbf{H}_k^{\mathsf{T}}\hat{\mathbf{x}}\bigr) && \text{(B.4)} \\
&\quad \le h\bigl(Z_{k-\kappa}^{k-1}\bigr) - \inf_{\substack{|x_{k-\kappa}|\ge x_{\min},\ldots,\\ |x_{k-1}|\ge x_{\min}}} h\Bigl(Z_{k-\kappa}^{k-1} \Bigm| \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}} \cdot x_\ell + Z_\ell\}_{\ell=k-\kappa}^{k-1}, \mathbf{H}_k^{\mathsf{T}}\hat{\mathbf{x}}\Bigr) && \text{(B.5)} \\
&\quad = h\bigl(Z_{k-\kappa}^{k-1}\bigr) - h\bigl(Z_{k-\kappa}^{k-1} \bigm| \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}} \cdot x_{\min} + Z_\ell\}_{\ell=k-\kappa}^{k-1}, \mathbf{H}_k^{\mathsf{T}}\hat{\mathbf{x}}\bigr) && \text{(B.6)} \\
&\quad = I\left(Z_{k-\kappa}^{k-1}; \left\{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}} + \frac{Z_\ell}{x_{\min}}\right\}_{\ell=k-\kappa}^{k-1}, \mathbf{H}_k^{\mathsf{T}}\hat{\mathbf{x}}\right) && \text{(B.7)} \\
&\quad = I\left(Z_1^{\kappa}; \left\{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}} + \frac{Z_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa}, \mathbf{H}_{\kappa+1}^{\mathsf{T}}\hat{\mathbf{x}}\right) && \text{(B.8)} \\
&\quad = I\left(\left\{\frac{Z_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa}; \left\{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}} + \frac{Z_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa} \middle| \mathbf{H}_{\kappa+1}^{\mathsf{T}}\hat{\mathbf{x}}\right) && \text{(B.9)} \\
&\quad = h\left(\left\{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}} + \frac{Z_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa} \middle| \mathbf{H}_{\kappa+1}^{\mathsf{T}}\hat{\mathbf{x}}\right) - h\bigl(\{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}\}_{\ell=1}^{\kappa} \bigm| \mathbf{H}_{\kappa+1}^{\mathsf{T}}\hat{\mathbf{x}}\bigr) && \text{(B.10)} \\
&\quad \triangleq \epsilon(x_{\min}, \kappa). && \text{(B.11)}
\end{aligned}$$

Here (B.3) follows from conditioning that reduces entropy; in the subsequent equality we use X_k and Z_k in order to extract $\mathbf{H}_k^{\mathsf{T}}\hat{\mathbf{x}}$ from Y_k, and then we drop (X_k, Y_k, Z_k) since given $\mathbf{H}_k^{\mathsf{T}}\hat{\mathbf{x}}$ it is independent of the other random variables; and (B.8) follows from stationarity.

From [7, Lemma 6.11], [2, Lemma A.19] we conclude that for every realization of $\mathbf{H}_{\kappa+1}^{\mathsf{T}}\hat{\mathbf{x}}$ the expression

$$h\left(\left\{\mathbf{H}_{\ell}^{\mathsf{T}}\hat{\mathbf{x}}+\frac{Z_{\ell}}{x_{\min}}\right\}_{\ell=1}^{\kappa}\,\middle|\,\mathbf{H}_{\kappa+1}^{\mathsf{T}}\hat{\mathbf{x}}=\mathbf{h}_{\kappa+1}^{\mathsf{T}}\hat{\mathbf{x}}\right) \tag{B.12}$$

converges monotonically in $x_{\min}$ to $h\big(\{\mathbf{H}_{\ell}^{\mathsf{T}}\hat{\mathbf{x}}\}_{\ell=1}^{\kappa}\,\big|\,\mathbf{H}_{\kappa+1}^{\mathsf{T}}\hat{\mathbf{x}}=\mathbf{h}_{\kappa+1}^{\mathsf{T}}\hat{\mathbf{x}}\big)$. By the Monotone Convergence Theorem (MCT) [25] this is also true when we average over $\mathbf{H}_{\kappa+1}^{\mathsf{T}}\hat{\mathbf{x}}$.

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October 1948.

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

[3] A. Lapidoth and S. M. Moser, "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.

[4] ——, "The fading number of SIMO fading channels with memory," in *Proceedings IEEE International Symposium on Information Theory and its Applications (ISITA)*, Parma, Italy, October 10–13, 2004, pp. 287–292.

[5] ——, "Feedback increases neither the fading number nor the pre-log," in *Proceedings Twenty-Third IEEE Convention of Electrical & Electronics Engineers in Israel (IEEEI)*, Herzlia, Israel, September 6–7, 2004, pp. 213–215.

[6] ——, "Bounds on the capacity of the discrete-time Poisson channel," in *Proceedings Forty-First Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, Illinois, October 1–3, 2003, pp. 201–210.

[7] ——, "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 asymptotic capacity of the discrete-time Poisson channel," in *Proceedings Winter School on Coding and Information Theory*, Monte Verità, Ascona, Switzerland, February 24–27, 2003.

[9] ——, "Capacity bounds via duality with applications to multi-antenna systems on flat fading channels," Signal and Information Processing Laboratory, ETH Zurich, Tech. Rep., June 25, 2002, preprint.

[10] ——, "On the fading number of multi-antenna systems over flat fading channels with memory and incomplete side information," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Lausanne, Switzerland, June 30 – July 5, 2002, p. 478.

[11] ——, "On the fading number of multi-antenna systems," in *Proceedings IEEE Information Theory Workshop (ITW)*, Cairns, Australia, September 2–7, 2001, pp. 110–111.

[12] ——, "Convex-programming bounds on the capacity of flat-fading channels," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Washington DC, USA, June 24–29, 2001, p. 52.

[13] ——, "Limits on reliable communication over flat-fading channels," in *Proceedings Winter School on Coding and Information Theory*, Schloss Reisensburg, Günzburg, University of Ulm, Germany, December 17–20, 2000.

[14] 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.

[15] 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.

[16] ——, "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.

[17] A. Lapidoth, "On the high SNR capacity of stationary Gaussian fading channels," in *Proceedings Forty-First Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, Illinois, October 1–3, 2003, pp. 410–419.

[18] 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.

[19] A. Lapidoth and S. M. Moser, "On non-coherent fading channels with feedback," in *Proceedings Winter School on Coding and Information Theory*, Bratislava, Slovakia, February 20–25, 2005, pp. 113–118.

[20] A. Lapidoth, "On the high-SNR capacity of noncoherent networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3025–3036, September 2005.

[21] ——, "On the capacity of non-coherent fading networks," in *Proceedings Third Joint Workshop on Communications and Coding (JWCC)*, Donnini-Firenze, Italy, October 14–17, 2004.

[22] ——, "Capacity bounds via duality: a phase noise example," in *Proceedings Second Asian-European Workshop on Information Theory*, Breisach, Germany, June 26–29, 2002.

[23] ——, "On phase noise channels at high SNR," in *Proceedings IEEE Information Theory Workshop (ITW)*, Bangalore, India, October 20–25, 2002, pp. 1–4.

[24] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge University Press, 1985.

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