Capacity Analysis of Multiple-Access OFDM Channels

Final Report of Project
"TW4G Wireless Access Technology"

Contract-Number:	G1-96001
Project Duration:	1 January 2007–31 December 2007
Funded by:	Industrial Technology Research Institute (ITRI), Zhudong, 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.

In this project wireless communication channels are investigated under the assumption that neither receiver nor transmitter has any *a priori* knowledge of the channel state, *i.e.*, the estimation schemes is incorporated into the system and its analysis. The goal is to find new insight into the fundamental limits of such communication systems, in particular, new results about its channel capacity. Since no assumptions are made about the estimation part of the system, the channel capacity often is known as *non-coherent channel capacity.*

Recent results show 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. Based on this observation the main goal of this project is the study of the threshold between the power-efficient low- to medium-SNR and the highly power-inefficient high-SNR regime. To that goal the *fading number* is derived defined as the second term in the high-SNR expansion of channel capacity. In particular new expressions are shown for the fading number of independent and identically distributed (IID) multiple-input multiple-output (MIMO) Gaussian fading channels with a scalar line-of-sight component and for the fading number of general MIMO fading channels with memory.

As a side-product closed-form expressions for so-far unknown expectations of a non-central chi-square distributed random variable are derived.

Keywords: Fading number, flat fading, high signal-to-noise ratio (high SNR), joint estimation and detection, memory, multiple-input multiple-output (MIMO), non-central chi-square, non-coherent channel capacity.

Contents

Acknowledgments

The author would like to thank Su Yu-Ted for fruitful discussions and Tu Shuchi for her invaluable help concerning the difficult Chinese language!

Chapter 1

Introduction

1.1 General Background

1.1.1 Wireless Communication

The importance of mobile communication systems 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 the 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, the reflectors, and 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, reflectors, and transmitter.
- The signals of several transmitters arrive as a superposition at the receiver, *i.e.*, the different users act as interferers to each other.
- 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 introduces intersymbol interference. The latter effect can be combatted using appropriate transmissions schemes and coding like, *e.g.*, an orthogonal frequency division multiplexing (OFDM) system. Fading is more difficult to deal with. It has a strong impact on the performance of a system and is the dominant source of transmission errors at high signal-to-noise ratios (SNR). Clever coding approaches are needed and one needs to take advantage of various types of available *diversity* [1].

To combat multiple-user interference we may use various multiple-access techniques like, *e.g.*, time-division multiple-access (TDMA), frequency-division multiple-access (FDMA), or code-division multiple-access (CDMA). The former two approaches separate times slots or frequency bands which are only to be used by one user. In the latter all users transmit at the same time and using the same frequency bandwidth, but using different spreading codes that are orthogonal to each other, so that the receiver can decode each user separately without having any interference (apart from a generally slightly higher noise power level), and then, once a particular user's signal has been decoded, can subtract the interference of this user from the received signal. Such an approach is also known as *successive cancellation*. Note that with respect to the inter-symbol interference an interesting approach is to use an orthogonal frequency division multiple-access (OFDMA) system that combines FDMA with OFDM.

In our analysis of these multiple-access fading channels we do not assume any particular multiple-access scheme, but instead are interested in the maximum possible sum of the rates of all users that is theoretically possible to transmit using a (possibly very elaborate) system, *i.e.*, we only consider the sum rate. In the following we use $\mathsf{C}(\textsc{snr})$ exchangeably for the maximum single-user rate (or *capacity*, see below) and the maximum multiple-user sum rate as a function of the total SNR in the channel, *i.e.*, the ratio of the total available power of all users and the noise power.

1.1.2 Joint Estimation and Detection

The time variant nature of the channel is probably the most difficult aspect of the channel. Currently, a wireless communication system usually uses training sequences that are regularly transmitted between real data bits 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 information carrying data at the same time for estimating the channel state. Such a *joint estimation and detection* approach will particularly be 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 a joint estimation and detection approach is that it allows fair and more realistic investigations of 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" [2] Claude E. Shannon proved that for every communication channel there exists a maximum 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 allow only for a very limited delay. Nevertheless, the capacity describes the ultimate limit of the 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 do not even need to make any 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 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.

1.1.3 Bounding Technique based on Duality

In [3] and extracts thereof published before [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 [3] 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

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

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

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 4.1.

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

- the free-space optical intensity channel [3], [7], [9];
- an optical intensity channel with input-dependent noise [3];
- the Poisson channel [3], [7], [9];
- 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 [3], [5], [8];
- 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], [3], [6];
- 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.

1.1.4 Double-Logarithmic Growth of the Capacity of Certain Fading Channels

Some of these results have been very unexpected. *E.g.*, it has been shown in [3] 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

[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.

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!

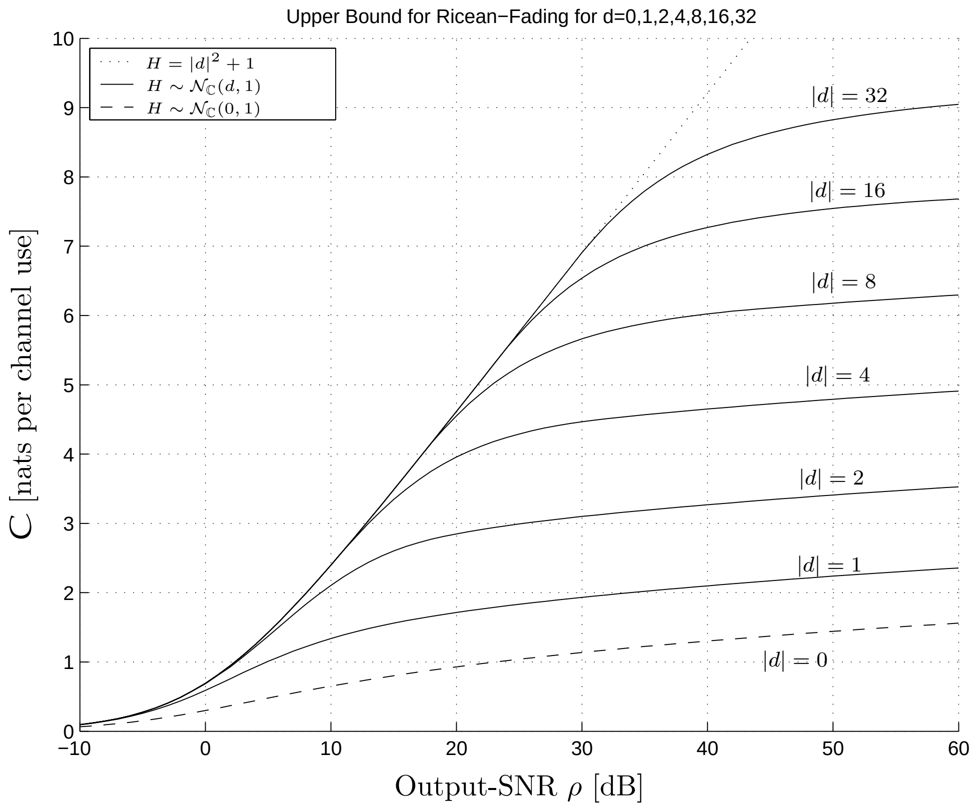

Figure 1.1: An upper bound on the capacity of a Rician fading channel as a function of the output-SNR $\rho = (1 + |d|^2)\text{SNR}$ for different values of the specular component d. The dotted line depicts the capacity of a Gaussian channel of equal output-SNR ρ, namely $\log(1 + \rho)$.

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 single-user Rayleigh fading channels with various numbers of receive antennas.

1.2 The Fading Number

In an attempt to more precisely quantify the mentioned threshold between the power-efficient and the power-inefficient regime, [8, Sec. IV.C] and [3, Sec. 6.5.2] define the *fading number* χ as the second term in the high-SNR asymptotic expansion of capacity, *i.e.*, at high SNR the channel capacity can be expressed as

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

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

Based on (1.1) we define the *high-SNR regime* to be the region where the $o(1)$-terms in (1.1) are negligible, *i.e.*, we say that a wireless communication system operates in the inefficient high-SNR regime if its capacity can be well approximated by

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

The important point to notice is that due to the extremely slow growth of $\log\log \text{SNR}$ the fading number χ is usually the dominant term in the lower range of the high-SNR regime. In other words, $\log\log \text{SNR}$ is only much larger than χ for extremely large values of SNR. An illustration of this behavior is shown in Figure 1.2.

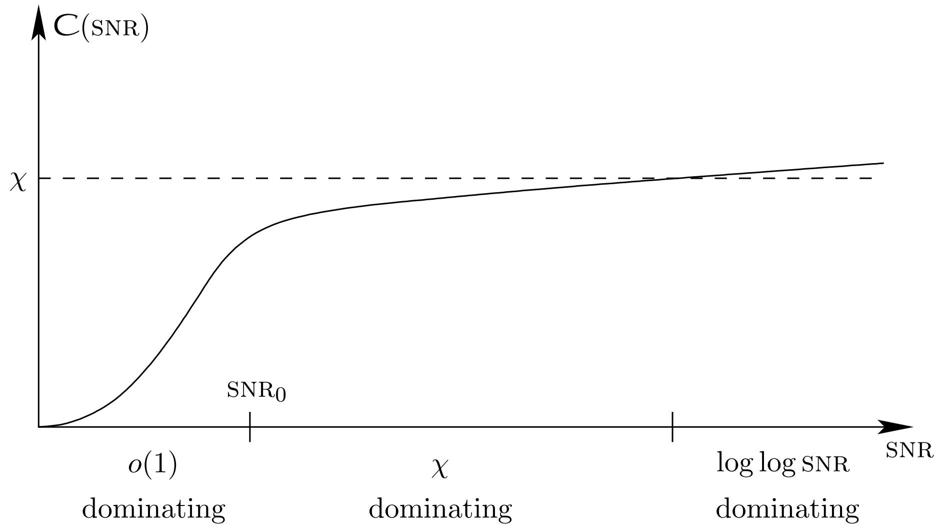

Figure 1.2: Illustration of the different regimes of a typical regular fading channel. At low SNR the $o(1)$ is dominant, in the lower range of the high-SNR regime the fading number χ is dominant, and only at very high SNR the $\log\log \text{SNR}$ terms takes the lead.

The fading number is therefore strongly connected to the point where the bend of the capacity curve occurs. As an example consider the following situation [16], [24]: 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:

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

$$\Longrightarrow \qquad \chi + 2.1 \text{ nats} \leq \mathsf{C}_0 \leq \chi + 3 \text{ nats}. \tag{1.4}$$

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

Conjecture 1 ([16], [24])**.** *A system that operates at rates appreciably above* $\chi + 2$ *nats is in the high-SNR regime and therefore extremely power-inefficient.*

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

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 are known in the special situation of general single-input single-output (SISO) fading channels with memory [8, Th. 4.41], [3, Th. 6.41]:

$$\chi\bigl(\{H_k\}\bigr) = \log \pi + \mathsf{E}\bigl[\log |H_0|^2\bigr] - h\bigl(\{H_k\}\bigr), \tag{1.5}$$

and of general SIMO fading channels with memory [4, Th. 1], [3, Th. 6.44]:

$$\chi\bigl(\{\mathbf{H}_k\}\bigr) = \chi_{\text{IID}}\left(\mathbf{H}_0 \,\middle|\, \mathbf{H}_{-\infty}^{-1}, \{\hat{\mathbf{H}}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\infty}\right). \tag{1.6}$$

Here $\chi_{\text{IID}}\bigl(\mathbf{H} \,\big|\, \mathbf{S}\bigr)$ denotes the memoryless SIMO fading number with partial side-information $\mathbf{S}$ at the receiver[4] [8, Note 4.31], [3, Eq. (6.194)]:

$$\chi_{\text{IID}}\bigl(\mathbf{H} \,\big|\, \mathbf{S}\bigr) = h_\lambda\bigl(\hat{\mathbf{H}} e^{\mathrm{i}\Theta} \,\big|\, \mathbf{S}\bigr) + n_{\text{R}} \mathsf{E}\bigl[\log \|\mathbf{H}\|^2\bigr] - \log 2 - h\bigl(\mathbf{H} \,\big|\, \mathbf{S}\bigr). \tag{1.7}$$

The fading number of the MISO fading channel has only been derived for the memoryless case [8, Th. 4.27], [3, Th. 6.27]:

$$\chi(\mathbf{H}^{\mathsf{T}}) = \sup_{\|\hat{\mathbf{x}}\|=1} \bigl\{\log \pi + \mathsf{E}\bigl[\log |\mathbf{H}^{\mathsf{T}}\hat{\mathbf{x}}|^2\bigr] - h(\mathbf{H}^{\mathsf{T}}\hat{\mathbf{x}})\bigr\}. \tag{1.8}$$

This fading number is achievable by inputs that can be expressed as the product of a constant unit vector in $\mathbb{C}^{n_{\text{T}}}$ and a circularly symmetric, scalar, complex random variable of the same law that achieves the memoryless SISO fading number [8]. Hence, the asymptotic capacity of a memoryless MISO fading channel is achieved by beam-forming where the beam-direction is chosen not to maximize the SNR, but the fading number.

For MISO fading with memory recently some bounds have been found [25], [26], [27], [28]:

$$\chi\bigl(\{\mathbf{H}_k^{\mathsf{T}}\}\bigr) \le \sup_{\hat{\mathbf{x}}_{-\infty}^0} \Bigl\{ \log \pi + \mathsf{E}\bigl[\log |\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0|^2\bigr] - h\bigl(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}_0 \,\big|\, \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\bigr)\Bigr\} \tag{1.9}$$

and

$$\chi\bigl(\{\mathbf{H}_k^{\mathsf{T}}\}\bigr) \ge \sup_{\hat{\mathbf{x}}} \Bigl\{ \log \pi + \mathsf{E}\bigl[\log |\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}|^2\bigr] - h\bigl(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}} \,\big|\, \{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}\}_{\ell=-\infty}^{-1}\bigr)\Bigr\}. \tag{1.10}$$

The most general situation of multiple antennas at both transmitter and receiver was first solved in the special situation of a memoryless fading process with a particular rotational symmetry: if every rotation of the input vector of the channel can be "undone" by a corresponding rotation of the output vector, and vice versa, then the fading number has been shown in [8], [3] to be

$$\chi(\mathbb{H}) = \log \frac{\pi^{n_{\text{R}}}}{\Gamma(n_{\text{R}})} + n_{\text{R}} \mathsf{E}\bigl[\log \|\mathbb{H}\hat{\mathbf{e}}\|^2\bigr] - h(\mathbb{H}\hat{\mathbf{e}}) \tag{1.11}$$

where $\hat{\mathbf{e}} \in \mathbb{C}^{n_{\text{T}}}$ is an arbitrary constant vector of unit length, and where n_{R} denotes the number of receive antennas. Such fading channels are called *rotation-commutative in the generalized sense* (for a detailed definition see Section 4.5).

Recently now we have been able to derive new expressions for general memoryless multiple-input multiple-output (MIMO) fading [29] and the most general single-user case of general MIMO fading with memory [30]. This report will discuss these cases more in detail.

[4]For an explanation of the notation used in this paper we refer to Section 2.1.

1.3 Non-Central Chi-Square Distribution

It is well known that adding several independent squared Gaussian random variables of identical variances yields a random variable that is non-central chi-square distributed. This distribution often shows up in information theory and communications. As an example we mention the situation of a fading channel as described above with a fading process consisting of independent unit-variance Gaussian components with or without mean. See Chapter 5 for more details.

While the special case of a squared Rayleigh distribution[5] is well understood in the sense that there exist closed-form expressions for more or less all interesting expected values, the more general situation of a non-central chi-square distribution is far more complex. Here, standard integration tools [31] or integration look-up tables [32] will very quickly cease to provide closed-form expressions for important expectations. So, for example, for a non-central chi-square distributed random variable V the values of $\mathsf{E}[\log V]$ are unknown.

Recently we have managed to improve on this problem: in this report we will state closed-form solutions to $\mathsf{E}[\log V]$ and $\mathsf{E}\big[\frac{1}{V^n}\big]$ for a non-central chi-square random variable V with an even number of degrees of freedom. Note that in practice we often have an even number of degrees of freedom because we usually consider *complex* Gaussian random variables consisting of *two real* Gaussian components.

We will see that these expectations are all related to a family of functions $g_m(\cdot)$ that is defined in Definition 4 in Chapter 3. There we will also derive some properties of these functions $g_m(\cdot)$ and state some tight upper and lower bounds.

1.4 Outline

The remainder 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 investigate the non-central chi-square distribution in Chapter 3: Section 3.1 summarizes the main results, *i.e.*, the closed-form expressions for some expectations, Sections 3.2 to 3.4 give some properties and bounds, and Section 3.5 contains all proofs.

In Chapter 4 we then summarize the newly derived expression of the fading number for general memoryless MIMO fading. We will only outline the proofs. In Chapter 5 we specialize these results to the situation of a MIMO Gaussian fading channel where all components of the fading matrix are independent unit-variance Gaussian. The results of this chapter rely on the results given in Chapters 3 and 4.

The most general case of general MIMO fading with memory is then treated in Chapter 6. There we put our main emphasis on two auxiliary results that are interesting by themselves: firstly we prove in Section 6.1 that the capacity-achieving input distribution for general MIMO fading with memory can be assumed to be circularly symmetric. Secondly, in Section 6.2, an even more general theorem is derived: we show that any stationary channel model will have a capacity-achieving input distribution that is basically stationary. The fading number for general MIMO fading with memory is then summarized in Section 6.3, however, we will only give the main ideas of the proof.

We conclude in Chapter 7. There in Section 7.1 we summarize our main contributions in this report and our publications stemming from them. In Section 7.2

[5] A squared Rayleigh random variable is actually exponentially distributed.

we then briefly discuss the results and in Section 7.3 we give some ideas of further research directions.

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.*, instead of C we use C to denote capacity. The energy per symbol is denoted by $\mathcal{E}$ and the signal-to-noise ratio SNR is denoted by $\textsc{snr}$. The $m \times m$ identity matrix is denoted by I_m, and $\mathsf{0}_{m\times n}$ stands for a $m \times n$ matrix with all components being zero.

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 maximum singular value of the matrix A.

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

$$\|\mathsf{A}\|_{\mathrm{F}} \triangleq \sqrt{\operatorname{tr}(\mathsf{A}^\dagger\mathsf{A})}. \tag{2.3}$$

Note that for every matrix A

$$\|\mathsf{A}\| \leq \|\mathsf{A}\|_{\mathrm{F}} \tag{2.4}$$

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

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

$$\hat{\mathbf{v}} \triangleq \frac{\mathbf{v}}{\|\mathbf{v}\|} \tag{2.5}$$

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

$$\|\hat{\mathbf{v}}\| = \|\hat{\mathbf{V}}\| = 1. \tag{2.6}$$

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

$$h_\lambda(\hat{\mathbf{V}}) \triangleq -\mathsf{E}\left[\log p^{\lambda}_{\hat{\mathbf{V}}}(\hat{\mathbf{V}})\right] \tag{2.7}$$

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

$$h_\lambda(\mathsf{U}\hat{\mathbf{V}}) = h_\lambda(\hat{\mathbf{V}}). \tag{2.8}$$

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

$$h_\lambda(\hat{\mathbf{V}}) = \log c_m, \tag{2.9}$$

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

$$c_m = \frac{2\pi^m}{\Gamma(m)}. \tag{2.10}$$

The definition (2.7) 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

$$h_\lambda\big(\hat{\mathbf{V}} \,\big|\, \mathbf{W} = \mathbf{w}\big) \triangleq -\mathsf{E}\left[\log p^{\lambda}_{\hat{\mathbf{V}}|\mathbf{W}}(\hat{\mathbf{V}}|\mathbf{W}) \,\middle|\, \mathbf{W} = \mathbf{w}\right] \tag{2.11}$$

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* $\mathbb{C}^\mu$ *and of differential entropy* $h(\mathbf{V})$*. Let* $\|\mathbf{V}\|$ *denote its norm and* $\hat{\mathbf{V}}$ *denotes its direction as in* (2.5)*. Then*

$$h(\mathbf{V}) = h(\|\mathbf{V}\|) + h_\lambda\big(\hat{\mathbf{V}} \,\big|\, \|\mathbf{V}\|\big) + (2\mu - 1)\mathsf{E}[\log \|\mathbf{V}\|] \tag{2.12}$$

$$= h_\lambda\big(\hat{\mathbf{V}}\big) + h\big(\|\mathbf{V}\| \,\big|\, \hat{\mathbf{V}}\big) + (2\mu - 1)\mathsf{E}[\log \|\mathbf{V}\|] \tag{2.13}$$

whenever all the quantities in (2.12) *and* (2.13)*, respectively, are defined. Here* $h(\|\mathbf{V}\|)$ *is the differential entropy of* $\|\mathbf{V}\|$ *when viewed as a real (scalar) random variable.*

Moreover, note that

$$h\big(\|\mathbf{V}\|^2\big) = h\big(\|\mathbf{V}\|\big) + \mathsf{E}[\log\|\mathbf{V}\|] + \log 2. \tag{2.14}$$

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, complex Gaussian random vector of covariance matrix $\mathsf{E}\big[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^{\dagger}\big] = \mathsf{K}$. Similarly, $\mathcal{N}_{\mathbb{R}}(\boldsymbol{\mu}, \mathsf{K})$ denotes a *real* Gaussian random vector. By $X \sim \mathcal{U}([a,b])$ we denote a random variable that is uniformly distributed on the interval $[a,b]$. The probability distribution of a random variable X or random vector $\mathbf{X}$ is denoted by Q_X or $Q_{\mathbf{X}}$, respectively.

Throughout this report $e^{\mathrm{i}\Theta}$ denotes a complex random variable that is uniformly distributed over the unit circle

$$e^{\mathrm{i}\Theta} \sim \text{Uniform on } \{z \in \mathbb{C} : |z| = 1\}. \tag{2.15}$$

When it appears in formulas with other random variables, $e^{\mathrm{i}\Theta}$ is always assumed to be independent of these other variables.

We use "$\stackrel{\mathscr{L}}{=}$" to denote "equal in law," and "$\triangleq$" stands for "is defined as." All rates specified in this report are in nats per channel use and $\log(\cdot)$ denotes the natural logarithmic function.

2.2 The Channel Model

We consider a channel with u users each having n_i transmit antennas, $i = 1, \ldots, u$. The total number of transmit antennas is then

$$\sum_{i=1}^{u} n_i = n_{\mathrm{T}}. \tag{2.16}$$

We then assume one receiver with n_{R} receive antennas whose time-k output $\mathbf{Y}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ is given by

$$\mathbf{Y}_k = \mathbb{H}_k \mathbf{x}_k + \mathbf{Z}_k. \tag{2.17}$$

Here $\mathbf{x}_k \in \mathbb{C}^{n_{\mathrm{T}}}$ denotes the time-k input vector consisting of u subvectors of length n_i for each user; the random matrix $\mathbb{H}_k \in \mathbb{C}^{n_{\mathrm{R}} \times n_{\mathrm{T}}}$ denotes the time-k fading matrix; and the random vector $\mathbf{Z}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ denotes the time-k additive noise vector.

We assume that the random vector process $\{\mathbf{Z}_k\}$ is a spatially and temporally white, zero-mean, circularly symmetric, complex Gaussian random process, *i.e.*, $\{\mathbf{Z}_k\}$ is temporally independent and identically distributed (IID) $\sim \mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \sigma^2 \mathsf{I}\big)$ for some $\sigma^2 > 0$. Here I denotes the identity matrix.

As for the matrix-valued fading process $\{\mathbb{H}_k\}$ we will not specify a particular distribution, but shall only assume that it is stationary, ergodic, of a finite-energy fading gain, *i.e.*,

$$\mathsf{E}\big[\|\mathbb{H}_k\|_{\mathrm{F}}^2\big] < \infty \tag{2.18}$$

and *regular*, *i.e.*, its differential entropy rate is finite:

$$h(\{\mathbb{H}_k\}) \triangleq \lim_{n\uparrow\infty} \frac{1}{n} h(\mathbb{H}_1, \ldots, \mathbb{H}_n) > -\infty. \tag{2.19}$$

I.e., unless explicitly stated we do not assume that $\{\mathbb{H}_k\}$ is Gaussian.

In the special situation when the fading process has no temporal memory, *i.e.*, when $\{\mathbb{H}_k\}$ is IID with respect to k, we will drop the time index k. In this case (2.17) simplifies to

$$\mathbf{Y} = \mathbb{H}\mathbf{x} + \mathbf{Z}. \tag{2.20}$$

Note that even if we assume that there is no temporal memory in the channel, we do not restrict the spatial memory, *i.e.*, the different fading components $H^{(i,j)}$ of the fading matrix $\mathbb{H}$ may be dependent.

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

We assume that the fading $\mathbb{H}$ and the additive noise $\mathbf{Z}$ are independent and of a joint law that does not depend on the channel input $\mathbf{x}$. The different users are assumed to have access to a common clock (resulting in the common discrete time k), but are otherwise not allowed to cooperate, *i.e.*, the u subvectors of $\mathbf{x}$ are independent of each other.

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 maximum allowed total instantaneous power in the former case, and to denote the allowed total average power in the latter case. For both cases we set

$$\text{SNR} \triangleq \frac{\mathcal{E}}{\sigma^2}. \tag{2.21}$$

The total power then still must be split and distributed among all users, however, since we are only studying the sum rate we are not interested in that part at the moment.

The (sum-rate) capacity $\mathsf{C}(\text{SNR})$ of the channel (2.17) and (2.20), respectively, is given by

$$\mathsf{C}(\text{SNR}) = \lim_{n\uparrow\infty} \frac{1}{n} \sup I\left(\mathbf{X}_1^n; \mathbf{Y}_1^n\right) = \sup I\left(\mathbf{X}; \mathbf{Y}\right) \tag{2.22}$$

where the second supremum is over the set of all probability distributions on $\mathbf{X}$ for which the u subvectors are independent and which satisfy the constraints, *i.e.*,

$$\|\mathbf{X}\|^2 \leq \mathcal{E}, \quad \text{almost surely} \tag{2.23}$$

for a peak-power constraint, or

$$\mathsf{E}\left[\|\mathbf{X}\|^2\right] \leq \mathcal{E} \tag{2.24}$$

for an average-power constraint.

Specializing [8, Th. 4.2], [3, Th. 6.10] to memoryless MIMO fading, we have for $u = 1$

$$\varlimsup_{\text{SNR}\uparrow\infty} \Big\{\mathsf{C}(\text{SNR}) - \log\log \text{SNR}\Big\} < \infty. \tag{2.25}$$

Note that [8, Th. 4.2], [3, Th. 6.10] is stated under the assumption of an average-power constraint only. However, since a peak-power constraint is more stringent than an average-power constraint, (2.25) also holds in the situation of a peak-power constraint.

In the case of multiple-users this still holds, since we may think of the u users as being a large transmitter with n_T transmit antennas. The additional constraint that the users cannot cooperate, *i.e.*, that the u subvectors have to be independent can then only decrease the capacity.

The fading number χ is now defined as in [8, Def. 4.6], [3, Def. 6.13] by

$$\chi(\mathbb{H}) \triangleq \varlimsup_{\text{SNR}\uparrow\infty} \Big\{ \mathsf{C}(\text{SNR}) - \log\log \text{SNR} \Big\}. \tag{2.26}$$

Prima facie the fading number depends on whether a peak-power constraint (2.23) or an average-power constraint (2.24) is imposed on the input. However, in the situation of $u = 1$ it will turn out that the memoryless MIMO fading number is identical for both cases.

Finally, we remark that for an arbitrary constant non-singular $n_\text{R} \times n_\text{R}$ matrix A and an arbitrary constant non-singular $n_\text{T} \times n_\text{T}$ matrix B

$$\chi\big(\{\mathsf{A}\mathbb{H}_k\mathsf{B}\}\big) = \chi\big(\{\mathbb{H}_k\}\big), \tag{2.27}$$

see [8, Lem. 4.7], [3, Lem. 6.14].

Chapter 3

Some Expectations of a Non-Central Chi-Square Distribution

In this chapter we derive some so-far unknown closed-form expressions for certain expectations of a non-central chi-square distributed random variable. These results are then applied in Chapter 5, however, since the non-central chi-square distribution is important in many other situations in communications, these results are interesting by themselves.

3.1 The Family of Functions $g_m(\cdot)$

We start with a short review of the non-central chi-square distribution. A non-negative real random variable is said to have a *non-central chi-square* distribution with n *degrees of freedom* and *non-centrality parameter* s^2 if it is distributed like

$$\sum_{j=1}^{n} (X_j + \mu_j)^2, \tag{3.1}$$

where $\{X_j\}_{j=1}^{n}$ are IID $\sim \mathcal{N}_{\mathbb{R}}(0,1)$ and the real constants $\{\mu_j\}_{j=1}^{n}$ satisfy

$$s^2 = \sum_{j=1}^{n} \mu_j^2. \tag{3.2}$$

(The distribution of (3.1) depends on the constants $\{\mu_j\}$ only via the sum of their squares.) The probability density function of such a distribution is given by [33, Ch. 29]

$$\frac{1}{2}\left(\frac{x}{s^2}\right)^{\frac{n-2}{4}} e^{-\frac{s^2+x}{2}} I_{n/2-1}\left(s\sqrt{x}\right), \qquad x \geq 0. \tag{3.3}$$

Here $I_\nu(\cdot)$ denotes the modified Bessel function of the first kind of order $\nu \in \mathbb{R}$, *i.e.*,

$$I_\nu(x) \triangleq \sum_{k=0}^{\infty} \frac{1}{k!\,\Gamma(\nu+k+1)} \left(\frac{x}{2}\right)^{\nu+2k}, \qquad x \geq 0 \tag{3.4}$$

(see [32, Eq. 8.445]).

If the number of degrees of freedom n is even, *i.e.*, if $n = 2m$ for some positive integer m, then the non-central chi-square distribution can also be expressed as a sum of the squared norms of *complex* Gaussian random variables.

Definition 3. *Let the random variable V have a non-central chi-square distribution with an even number $2m$ of degrees of freedom,* i.e.*,*

$$V \triangleq \sum_{j=1}^{m} \left|U_j + \mu_j\right|^2 \tag{3.5}$$

where $\{U_j\}_{j=1}^m$ are IID $\sim \mathcal{N}_{\mathbb{C}}(0,1)$, and $\{\mu_j\}_{j=1}^m$ are complex constants. Let further the non-centrality parameter s^2 be defined as

$$s^2 \triangleq \sum_{j=1}^{m} |\mu_j|^2. \tag{3.6}$$

Now we are ready for the main result of this chapter, *i.e.*, some closed-form expressions of expected values of V. We start with the definition of the following family of functions.

Definition 4. *The functions $g_m(\cdot)$ are defined as follows:*

$$g_m(\xi) \triangleq \begin{cases} \log(\xi) - \operatorname{Ei}(-\xi) + \sum_{j=1}^{m-1} (-1)^j \Big[e^{-\xi}(j-1)! \\ \qquad - \frac{(m-1)!}{j(m-1-j)!} \Big] \left(\frac{1}{\xi}\right)^j, & \xi > 0 \\ \psi(m), & \xi = 0 \end{cases} \tag{3.7}$$

for $m \in \mathbb{N}$, where $\operatorname{Ei}(\cdot)$ denotes the exponential integral function defined as

$$\operatorname{Ei}(-x) \triangleq -\int_x^\infty \frac{e^{-t}}{t}\,\mathrm{d}t, \qquad x > 0 \tag{3.8}$$

and $\psi(\cdot)$ is Euler's psi function given by

$$\psi(m) \triangleq -\gamma + \sum_{j=1}^{m-1} \frac{1}{j} \tag{3.9}$$

with $\gamma \approx 0.577$ denoting Euler's constant.

Note that $g_m(\xi)$ is continuous for all $\xi \geq 0$, *i.e.*, in particular

$$\lim_{\xi \downarrow 0} \left\{ \log(\xi) - \operatorname{Ei}(-\xi) + \sum_{j=1}^{m-1} (-1)^j \left[e^{-\xi}(j-1)! - \frac{(m-1)!}{j(m-1-j)!} \right] \left(\frac{1}{\xi}\right)^j \right\} = \psi(m) \tag{3.10}$$

for all $m \in \mathbb{N}$. Therefore its first derivative is defined for all $\xi \geq 0$ and can be evaluated to

$$g_m'(\xi) \triangleq \frac{\partial g_m(\xi)}{\partial \xi} = \frac{(-1)^m \Gamma(m)}{\xi^m} \left(e^{-\xi} - \sum_{j=0}^{m-1} \frac{(-1)^j}{j!} \xi^j \right) \tag{3.11}$$

(see [8, Eq. (417)], [3, Eq. (A.39)]). Note that $g_m'(\cdot)$ is also continuous, *i.e.*, in particular

$$\lim_{\xi \downarrow 0} \left\{ \frac{(-1)^m \Gamma(m)}{\xi^m} \left(e^{-\xi} - \sum_{j=0}^{m-1} \frac{(-1)^j}{j!} \xi^j \right) \right\} = \frac{1}{m} = g_m'(0). \tag{3.12}$$

Then we have the following result.

Theorem 5. *The expected value of the logarithm of a non-central chi-square random variable with an even number* $2m$ *of degrees of freedom is given as*

$$\boxed{\mathsf{E}[\log V] = g_m(s^2),} \tag{3.13}$$

where V *and* s^2 *are defined in* (3.5) *and* (3.6)*. Hence, we have the solution to the following integral:*

$$\int_0^\infty \log v \cdot \left(\frac{v}{s^2}\right)^{\frac{m-1}{2}} e^{-v-s^2} I_{m-1}(2s\sqrt{v})\,\mathrm{d}v = g_m(s^2) \tag{3.14}$$

for any $m \in \mathbb{N}$ *and* $s^2 \geq 0$.

Proof. A proof can be found in [8, Lem. 10.1], [3, Lem. A.6] □

Next we look at the reciprocal value.

Theorem 6. *Let* $n \in \mathbb{N}$ *with* $n < m$*. The expected value of the* n*-th power reciprocal value of a non-central chi-square random variable with an even number* $2m$ *of degrees of freedom is given as*

$$\boxed{\mathsf{E}\left[\frac{1}{V^n}\right] = \frac{(-1)^{n-1}}{(n-1)!} \cdot g_{m-n}^{(n)}(s^2), \qquad m > n} \tag{3.15}$$

where

$$g_m^{(\ell)}(\xi) = \frac{\partial^\ell g_m(\xi)}{\partial \xi^\ell} \tag{3.16}$$

denotes the ℓ*-th derivative of* $g_m(\cdot)$ *and where* V *and* s^2 *are defined in* (3.5) *and* (3.6)*. In particular, for* $m > 1$

$$\boxed{\mathsf{E}\left[\frac{1}{V}\right] = g'_{m-1}(s^2).} \tag{3.17}$$

Hence, we have the solution to the following integral:

$$\int_0^\infty \frac{1}{v^n} \cdot \left(\frac{v}{s^2}\right)^{\frac{m-1}{2}} e^{-v-s^2} I_{m-1}(2s\sqrt{v})\,\mathrm{d}v = \frac{(-1)^{n-1}}{(n-1)!} \cdot g_{m-n}^{(n)}(s^2) \tag{3.18}$$

for any $m, n \in \mathbb{N}$, $m > n$, *and any* $s^2 \geq 0$.

Note that in the cases where $m \leq n$, the expectation is unbounded.

Proof. See Section 3.5.1. □

3.2 Properties of $g_m(\cdot)$ and $g'_m(\cdot)$

In this section we will show that the family of functions $g_m(\cdot)$ and $g'_m(\cdot)$ are well-behaved.

Corollary 7. *The functions* $g_m(\cdot)$ *are monotonically strictly increasing and strictly concave in the interval* $[0, \infty)$ *for all* $m \in \mathbb{N}$.

Proof. From [8, Eqs. (415), (416)], [3, Eqs. (A.37), (A.38)] we know that

$$g'_m(\xi) = \frac{\partial g_m(\xi)}{\partial \xi} = e^{-\xi} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{1}{k+m} \cdot \xi^k, \tag{3.19}$$

$$g''_m(\xi) = \frac{\partial^2 g_m(\xi)}{\partial \xi^2} = -e^{-\xi} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{1}{(k+m)(k+m+1)} \cdot \xi^k, \tag{3.20}$$

i.e., the first derivative of $g_m(\cdot)$ is positive and the second derivative is negative. □

Corollary 8. *The function $g_m(\xi)$ is monotonically strictly increasing in m for all $\xi \geq 0$.*

Proof. Fix two arbitrary natural numbers $m_1, m_2 \in \mathbb{N}$ such that $m_1 < m_2$. Choose $\mu_1 = s$, $\mu_2 = \ldots = \mu_{m_2} = 0$, with an arbitrary $s \geq 0$. Let $\{U_j\}_{j=1}^{m_2}$ be IID $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. Then

$$\begin{aligned} g_{m_2}(s^2) &= \mathsf{E}\left[\log\left(\sum_{j=1}^{m_2} |U_j + \mu_j|^2\right)\right] &(3.21)\\ &= \mathsf{E}\left[\log\left(\sum_{j=1}^{m_1} |U_j + \mu_j|^2 + \sum_{j=m_1+1}^{m_2} |U_j + \mu_j|^2\right)\right] &(3.22)\\ &> \mathsf{E}\left[\log\left(\sum_{j=1}^{m_1} |U_j + \mu_j|^2\right)\right] &(3.23)\\ &= g_{m_1}(s^2), &(3.24)\end{aligned}$$

where the first equality follows from (3.13); the subsequent equality from splitting the sum into two parts; the subsequent inequality from dropping some positive terms; and the final equality again from (3.13). □

Corollary 9. *The functions $g'_m(\cdot)$ are positive, monotonically strictly decreasing, and strictly convex functions for all $m \in \mathbb{N}$.*

Proof. The positivity and the monotonicity follow from (3.19) and (3.20). To see the convexity, compute from (3.20)

$$g'''_m(\xi) = \frac{\partial^3 g_m(\xi)}{\partial \xi^3} = e^{-\xi} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{2}{(k+m)(k+m+1)(k+m+2)} \cdot \xi^k \tag{3.25}$$

which is positive. □

Corollary 10. *The function $g'_m(\xi)$ is monotonically strictly decreasing in m for all $\xi \geq 0$.*

Proof. Fix two arbitrary natural numbers $m_1, m_2 \in \mathbb{N}$ such that $m_2 > m_1 > 1$. Choose $\mu_1 = s$, $\mu_2 = \ldots = \mu_{m_2} = 0$, with an arbitrary $s \geq 0$. Let $\{U_j\}_{j=1}^{m_2}$ be IID

$\sim \mathcal{N}_{\mathbb{C}}(0,1)$. Then

$$g'_{m_2-1}(s^2) = \mathsf{E}\left[\frac{1}{\sum_{j=1}^{m_2} |U_j + \mu_j|^2}\right] \tag{3.26}$$

$$= \mathsf{E}\left[\frac{1}{\sum_{j=1}^{m_1} |U_j + \mu_j|^2 + \sum_{j=m_1+1}^{m_2} |U_j + \mu_j|^2}\right] \tag{3.27}$$

$$< \mathsf{E}\left[\frac{1}{\sum_{j=1}^{m_1} |U_j + \mu_j|^2}\right] \tag{3.28}$$

$$= g'_{m_1-1}(s^2), \tag{3.29}$$

where the first equality follows from (3.17); the subsequent equality from splitting the sum into two parts; the subsequent inequality from dropping some positive terms in the denominator; and the final equality again from (3.17). □

Theorem 11. *We have the following relation:*

$$g_{m+1}(\xi) = g_m(\xi) + g'_m(\xi) \tag{3.30}$$

for all $m \in \mathbb{N}$ and all $\xi \geq 0$.

Proof. See Section 3.5.2. □

Theorem 12. *We have the following relations:*

$$g'_{m+1}(\xi) = \frac{1}{\xi} - \frac{m}{\xi} g'_m(\xi), \qquad \xi > 0, \tag{3.31}$$

$$\text{and} \qquad g'_m(\xi) = \frac{1}{m} - \frac{\xi}{m} g'_{m+1}(\xi), \qquad \xi \geq 0, \tag{3.32}$$

for all $m \in \mathbb{N}$.

Proof. See Section 3.5.3. □

3.3 Bounds on $g_m(\cdot)$ and $g'_m(\cdot)$

In this section we derive some tight bounds on the functions $g_m(\cdot)$ and $g'_m(\cdot)$.

Theorem 13. *The function $g'_m(\cdot)$ can be bounded as follows:*

$$\frac{1}{\xi + m} \leq g'_m(\xi) \leq \min\left\{\frac{m+1}{m(\xi + m + 1)}, \frac{1}{\xi + m - 1}\right\}. \tag{3.33}$$

Note that for $\xi < m+1$ the first of the two upper bounds is tighter than second, while for $\xi > m + 1$ the second is tighter. Moreover, the first upper bound coincides with $g_m(\xi)$ for $\xi = 0$, and the second upper bound is asymptotically tight when ξ tends to infinity.

Proof. See Section 3.5.4. □

The bounds (3.33) are depicted in Figure 3.1 for the cases of $m = 1$, $m = 3$, and $m = 8$.

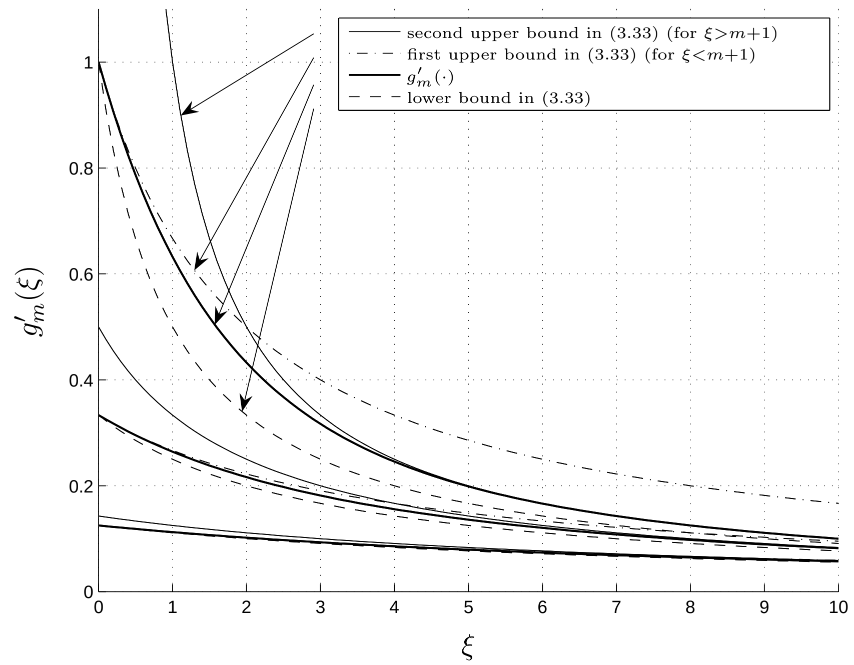

Figure 3.1: Upper and lower bounds on $g'_m(\cdot)$ according to (3.33) in Theorem 13. The top four curves correspond to $m = 1$, the middle four to $m = 3$, and the lowest group of four curves to $m = 8$.

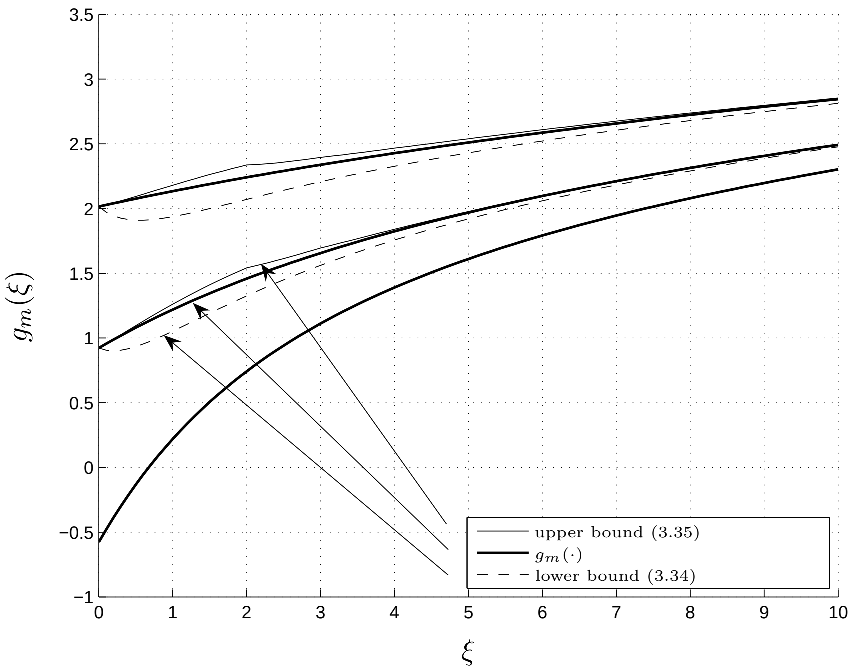

Figure 3.2: Upper and lower bounds on $g_m(\cdot)$ according to (3.34) and (3.35) in Theorem 14. The lowest curve corresponds to $m = 1$ (in this case all bounds coincides with $g_1(\cdot)$), the next three curves correspond to $m = 3$, and the top three to $m = 8$.

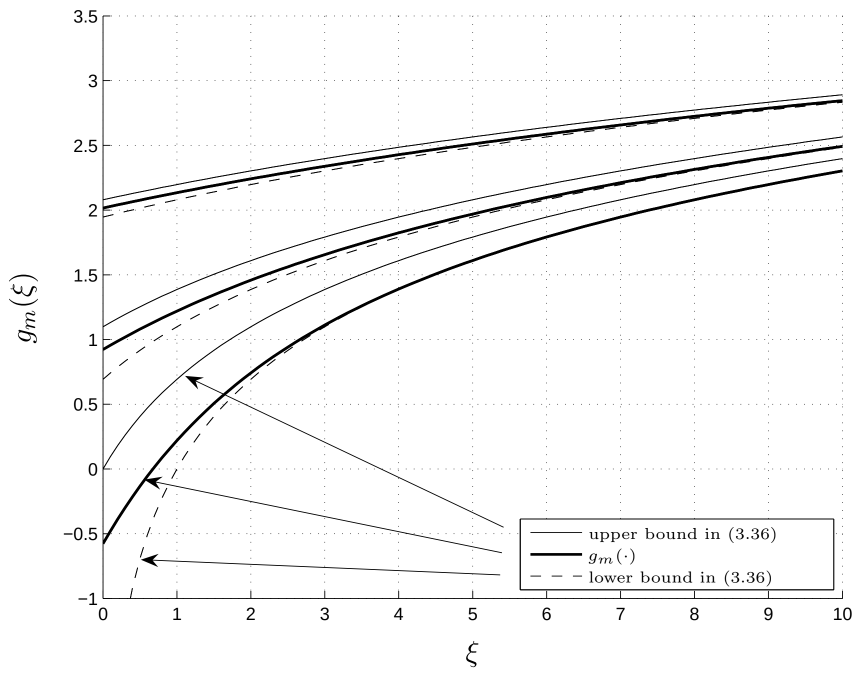

Figure 3.3: Upper and lower bounds on $g_m(\cdot)$ according to (3.36) in Theorem 14. The lowest three curves correspond to $m = 1$, the next three to $m = 3$, and the top three to $m = 8$.

Theorem 14. *For the functions $g_m(\cdot)$ we state two sets of bounds. The first set is tighter for smaller values of m:*

$$g_m(\xi) \geq \log\xi - \mathrm{Ei}\left(-\xi\right) + \sum_{j=1}^{m-1} \frac{1}{\xi + j}, \tag{3.34}$$

$$g_m(\xi) \leq \log\xi - \mathrm{Ei}\left(-\xi\right) + \sum_{j=1}^{m-1} \min\left\{\frac{j+1}{j(\xi+j+1)}, \frac{1}{\xi+j-1}\right\}. \tag{3.35}$$

Secondly, we give a set of bounds that is tight for large values of m:

$$\log(\xi + m - 1) \leq g_m(\xi) \leq \log(\xi + m). \tag{3.36}$$

Note that this second set of bounds is very simple, i.e., *it is particularly interesting for further analysis.*

Proof. See Section 3.5.5. □

The bounds (3.34) and (3.35) are depicted in Figure 3.2 and the bounds (3.36) in Figure 3.3, both times for the cases of $m = 1$, $m = 3$, and $m = 8$.

3.4 Additional Properties

As an example of how the properties given in Section 3.2 and the bounds given in Section 3.3 can be used to derive further results we state the following corollary.

Corollary 15. *The functions $g_m\left(\frac{1}{\xi}\right)$ are monotonically strictly decreasing and convex in ξ for all $m \in \mathbb{N}$.*

Proof. We start as follows:

$$\frac{\partial}{\partial \xi} g_m\left(\frac{1}{\xi}\right) = -\frac{1}{\xi^2} \cdot g_m'\left(\frac{1}{\xi}\right) \tag{3.37}$$

$$= -\frac{1}{\xi^2}\left(g_{m+1}\left(\frac{1}{\xi}\right) - g_m\left(\frac{1}{\xi}\right)\right), \tag{3.38}$$

where the last equality follows from Theorem 11. From Corollary 9 we know that $g_m'(\cdot) > 0$, so that we can conclude from (3.37) that $g_m\left(\frac{1}{\xi}\right)$ is monotonically strictly decreasing.

To check convexity, we continue as follows:

$$\frac{\partial^2}{\partial \xi^2} g_m\left(\frac{1}{\xi}\right) = \frac{2}{\xi^3} g_{m+1}\left(\frac{1}{\xi}\right) + \frac{1}{\xi^4} g_{m+1}'\left(\frac{1}{\xi}\right) - \frac{2}{\xi^3} g_m\left(\frac{1}{\xi}\right) - \frac{1}{\xi^4} g_m'\left(\frac{1}{\xi}\right) \tag{3.39}$$

$$= \frac{2}{\xi^3} g_m'\left(\frac{1}{\xi}\right) + \frac{\xi - m\xi g_m'\left(\frac{1}{\xi}\right) - g_m'\left(\frac{1}{\xi}\right)}{\xi^4} \tag{3.40}$$

$$= g_m'\left(\frac{1}{\xi}\right) \cdot \frac{2\xi - m\xi - 1}{\xi^4} + \frac{1}{\xi^3} \tag{3.41}$$

$$\geq \frac{1}{\frac{1}{\xi} + m} \cdot \frac{2\xi - m\xi - 1}{\xi^4} + \frac{1}{\xi^3} \tag{3.42}$$

$$= \frac{2}{\xi^2(m\xi + 1)}. \tag{3.43}$$

Here, in the second equality we use Theorems 11 and 12; and the inequality follows from the lower bound of Theorem 13. Hence, the second derivative is positive and the statement proved. □

3.5 Proofs

3.5.1 Proof of Theorem 6

Let $n \in \mathbb{N}$ be arbitrary and assume that $m > n$. Then the required expectation can be written as

$$\mathsf{E}\left[\frac{1}{V^n}\right] = \int_0^\infty \frac{1}{v^n} \cdot \left(\frac{v}{s^2}\right)^{\frac{m-1}{2}} e^{-v-s^2} I_{m-1}(2s\sqrt{v})\,\mathrm{d}v. \tag{3.44}$$

Expressing $I_{m-1}(\cdot)$ as a power series (3.4)

$$I_{m-1}(z) = \sum_{k=0}^{\infty} \frac{1}{k!\,\Gamma(m+k)} \left(\frac{z}{2}\right)^{m-1+2k} \tag{3.45}$$

we obtain from [32, Eq. 3.351-3] (using that $m > n$)

$$\mathsf{E}\left[\frac{1}{V^n}\right] = \frac{1}{s^{m-1}} e^{-s^2} \sum_{k=0}^{\infty} \frac{1}{k!\,\Gamma(m+k)} s^{2k+m-1} \cdot \int_0^\infty v^{k+m-1-n} e^{-v}\,\mathrm{d}v \tag{3.46}$$

$$= e^{-s^2} \sum_{k=0}^{\infty} \frac{1}{k!\,(m+k-1)!} s^{2k} \cdot (k+m-1-n)! \tag{3.47}$$

$$= e^{-s^2} \sum_{k=0}^{\infty} \frac{1}{k!\,(k+m-1)\cdots(k+m-n)} s^{2k}. \tag{3.48}$$

Generalizing (3.19), (3.20), and (3.25) to the ℓ-th derivative, we have

$$g_m^{(\ell)}(\xi) = \frac{\partial^\ell g_m(\xi)}{\partial \xi^\ell} = (-1)^{\ell-1} e^{-\xi} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{(\ell-1)!}{(k+m)\cdots(k+m+\ell-1)} \cdot \xi^k. \tag{3.49}$$

Comparing this with (3.48) we see that

$$\mathsf{E}\left[\frac{1}{V^n}\right] = \frac{(-1)^{n-1}}{(n-1!)} \cdot g_{m-n}^{(n)}(s^2). \tag{3.50}$$

3.5.2 Proof of Theorem 11

Using (3.7) and (3.11) we get

$$\begin{aligned}
& g_m(\xi) + g_m'(\xi) \\
&\quad = \log(\xi) - \mathrm{Ei}\,(-\xi) + \sum_{j=1}^{m-1} (-1)^j e^{-\xi} (j-1)! \left(\frac{1}{\xi}\right)^j \\
&\qquad - \sum_{j=1}^{m-1} (-1)^j \frac{(m-1)!}{j(m-1-j)!} \left(\frac{1}{\xi}\right)^j + \frac{(-1)^m (m-1)!}{\xi^m} e^{-\xi} \\
&\qquad - \sum_{i=0}^{m-1} (-1)^{i+m} \frac{(m-1)!}{i!} \xi^{i-m} && (3.51) \\
&\quad = \log(\xi) - \mathrm{Ei}\,(-\xi) + \sum_{j=1}^{m} (-1)^j e^{-\xi} (j-1)! \left(\frac{1}{\xi}\right)^j \\
&\qquad - \sum_{j=1}^{m-1} (-1)^j \frac{(m-1)!}{j(m-1-j)!} \left(\frac{1}{\xi}\right)^j - \sum_{j=1}^{m} \underbrace{(-1)^{-j+2m}}_{=(-1)^j} \frac{(m-1)!}{(m-j)!} \xi^{-j} && (3.52) \\
&\quad = \log(\xi) - \mathrm{Ei}\,(-\xi) + \sum_{j=1}^{m} (-1)^j e^{-\xi} (j-1)! \left(\frac{1}{\xi}\right)^j \\
&\qquad - \sum_{j=1}^{m-1} (-1)^j (m-1)! \underbrace{\left(\frac{1}{j(m-1-j)!} + \frac{1}{(m-j)!}\right)}_{=\frac{m}{j(m-j)!}} \left(\frac{1}{\xi}\right)^j \\
&\qquad - (-1)^m \frac{(m-1)!}{0!} \left(\frac{1}{\xi}\right)^m && (3.53) \\
&\quad = \log(\xi) - \mathrm{Ei}\,(-\xi) + \sum_{j=1}^{m} (-1)^j e^{-\xi} (j-1)! \left(\frac{1}{\xi}\right)^j \\
&\qquad - \sum_{j=1}^{m-1} (-1)^j (m-1)! \frac{m}{j(m-j)!} \left(\frac{1}{\xi}\right)^j - (-1)^m \frac{m!}{m\, 0!} \left(\frac{1}{\xi}\right)^m && (3.54) \\
&\quad = \log(\xi) - \mathrm{Ei}\,(-\xi) + \sum_{j=1}^{m} (-1)^j e^{-\xi} (j-1)! \left(\frac{1}{\xi}\right)^j \\
&\qquad - \sum_{j=1}^{m} (-1)^j \frac{m!}{j(m-j)!} \left(\frac{1}{\xi}\right)^j && (3.55) \\
&\quad = g_{m+1}(\xi). && (3.56)
\end{aligned}$$

Here, the first equality follows from the definitions given in (3.7) and (3.11); in the subsequent equality we combine the second last term with the first sum and reorder the last summation by introducing a new counter-variable $j = m - i$; the subsequent three equalities follows from arithmetic rearrangements; and the final equality follows again from definition (3.7).

3.5.3 Proof of Theorem 12

Using (3.19) we get

$$\begin{aligned}
\frac{1}{\xi} - \frac{m}{\xi} g'_m(\xi) &= \frac{1}{\xi} e^{-\xi} \cdot e^{\xi} - \frac{m}{\xi} e^{-\xi} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{1}{k+m} \cdot \xi^k && (3.57)\\
&= \frac{1}{\xi} e^{-\xi} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \xi^k - \frac{1}{\xi} e^{-\xi} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{m}{k+m} \cdot \xi^k && (3.58)\\
&= \frac{1}{\xi} e^{-\xi} \sum_{k=0}^{\infty} \frac{1}{k!} \left(1 - \frac{m}{k+m}\right) \xi^k && (3.59)\\
&= e^{-\xi} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{k}{k+m} \cdot \xi^{k-1} && (3.60)\\
&= e^{-\xi} \sum_{k=1}^{\infty} \frac{1}{(k-1)!} \cdot \frac{1}{k+m} \cdot \xi^{k-1} && (3.61)\\
&= e^{-\xi} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{1}{k+m+1} \cdot \xi^k && (3.62)\\
&= g'_{m+1}(\xi). && (3.63)
\end{aligned}$$

Here, the first equality follows from (3.19); in the subsequent equality we use the series expansion of e^{ξ} which is valid for all $\xi \geq 0$; the subsequent two equalities follow from algebraic rearrangements; in the next equality we note that for $k = 0$ the terms in the sum are equal to zero; the second last equality then follows from renumbering the terms; and the last equality follows again from (3.19).

3.5.4 Proof of Theorem 13

We start with the lower bound and note that for $\xi = 0$ the bound is tight:

$$g'_m(0) = g_{m+1}(0) - g_m(0) = \psi(m+1) - \psi(m) = \frac{1}{m} \tag{3.64}$$

where the first equality follows from Theorem 11, the second from (3.7) and the final from (3.9); and

$$\left.\frac{1}{\xi + m}\right|_{\xi=0} = \frac{1}{m}. \tag{3.65}$$

Moreover we notice that the bound is asymptotically tight, too:

$$\lim_{\xi \uparrow \infty} g'_m(\xi) = 0; \tag{3.66}$$

$$\lim_{\xi \uparrow \infty} \frac{1}{\xi + m} = 0. \tag{3.67}$$

Hence, since additionally both functions $g'_m(\cdot)$ and $\frac{1}{\cdot+m}$ are monotonically strictly decreasing and strictly convex, they cannot cross. Moreover, it is shown[1] in Theorem 14 that $g_m(\xi) \le \log(\xi + m)$, so that we must have

$$g'_m(\xi) \ge \frac{\partial}{\partial \xi} \log(\xi + m) = \frac{1}{\xi + m}. \tag{3.68}$$

Next we turn to the first upper bound which will follow from the lower bound (3.33) derived above together with Theorem 12:

$$g'_{m-1}(\xi) = \frac{1 - \xi g'_m(\xi)}{m-1} \le \frac{1 - \xi \frac{1}{\xi+m}}{m-1} = \frac{m}{(m-1)(\xi+m)}. \tag{3.69}$$

To derive the second upper bound we use once again Theorem 12 and the lower bound (3.33) derived above:

$$\frac{1}{\xi+m-1} - g'_m(\xi) = \frac{1}{\xi+m-1} - \frac{1}{\xi} + \frac{m-1}{\xi} g'_{m-1}(\xi) \tag{3.70}$$

$$\ge \frac{1}{\xi+m-1} - \frac{1}{\xi} + \frac{m-1}{\xi} \frac{1}{\xi+m-1} \tag{3.71}$$

$$= 0, \tag{3.72}$$

where the first equality follows from Theorem 12 and the inequality from the lower bound in (3.33).

3.5.5 Proof of Theorem 14

The first set of bounds (3.34) and (3.35) follow directly from Theorems 11 and 13 and from the fact that

$$g_1(\xi) = \log \xi - \operatorname{Ei}(-\xi). \tag{3.73}$$

The upper bound in the second set of bounds (3.36) has been proven before in [29, App. B]. The proof is based on Jensen's inequality.

The lower bound in (3.36) follows from a slightly more complicated argument: Note that both $g_m(\cdot)$ and $\log(\cdot + m - 1)$ are monotonically strictly increasing and strictly concave (see Corollary 7). Hence, they can cross at most twice. Now asymptotically as $\xi \uparrow \infty$ the two functions coincide which corresponds to one of these "crossings." So they can only cross at most once more for finite ξ. For $\xi = 0$, we have

$$g_m(0) = \psi(m) > \log(m-1) \tag{3.74}$$

for all $m \in \mathbb{N}$ (where for $m = 1$ we take $\log 0 = -\infty$). Let's assume for the moment that there is another crossing for a finite value ξ_0. Then, for $\xi > \xi_0$, $\log(\xi + m - 1)$ is strictly larger than $g_m(\xi)$. However, since asymptotically they will coincide, the slope of $\log(\cdot + m - 1)$ must then be strictly smaller than the slope of $g_m(\cdot)$. But we know from Theorem 13 that

$$\frac{\partial}{\partial \xi} \log(\xi + m - 1) = \frac{1}{\xi + m - 1} \ge g'_m(\xi). \tag{3.75}$$

This is a contradiction which leads to the conclusion that there cannot be another crossing and $\log(\cdot + m - 1)$ must be a strict lower bound to $g_m(\cdot)$.

[1]Note that this part of the proof of Theorem 14 does not rely in any way on the theorem under consideration.

Chapter 4

A General Memoryless MIMO Fading Channel

In this chapter we consider the special case of only one user $u = 1$. Moreover, we simplify the model to exclude temporal memory. Otherwise we assume the channel to be general without any particular assumption on the fading distribution, the number of antennas at transmitter and receiver, or on the correlation between the different antennas. Hence, we have a channel model that looks like (2.20) with either an average-power constraint (2.24) or a peak-power constraint (2.23) on the input.

Before we can state our main result of this chapter, *i.e.*, the fading number of memoryless MIMO fading channels, we need to introduce three concepts: The first concerns probability distributions that escape to infinity, the second a technique of upper-bounding mutual information, and the third concept concerns circular symmetry.

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

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

In the following we will only state this result specialized to the situation at hand. For a more general description and for all proofs we refer to [4, Sec. VII.C.3], [3, Sec. 2.6].

Definition 16. *Let* $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ *be a family of input distributions for the memoryless fading channel* (2.20)*, where this family is parametrized by the available average power* $\mathcal{E}$ *such that*

$$\mathsf{E}_{Q_{\mathbf{X},\mathcal{E}}}\left[\|\mathbf{X}\|^2\right] \leq \mathcal{E}, \qquad \mathcal{E} \geq 0. \tag{4.1}$$

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

$$\lim_{\mathcal{E}\uparrow\infty} Q_{\mathbf{X},\mathcal{E}}\bigl(\|\mathbf{X}\|^2 \leq \mathcal{E}_0\bigr) = 0. \tag{4.2}$$

We now have the following proposition.

Proposition 17. *Let the memoryless MIMO fading channel be given as in* (2.20) *and let* $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ *be a family of distributions on the channel input that satisfy the power constraint* (4.1). *Let* $I(Q_{\mathbf{X},\mathcal{E}})$ *denote the mutual information between input and output of channel* (2.20) *when the input is distributed according to the law* $Q_{\mathbf{X},\mathcal{E}}$. *Assume that the family of input distributions* $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ *is such that the following condition is satisfied:*

$$\lim_{\mathcal{E}\uparrow\infty} \frac{I(Q_{\mathbf{X},\mathcal{E}})}{\log\log\mathcal{E}} = 1. \tag{4.3}$$

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

Proof. A proof can be found in [4, Th. 8, Rem. 9], [3, Cor. 2.8]. □

4.2 An Upper Bound on Channel Capacity

In [8], [3] a new approach of deriving 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 to it. The 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 20. For a more general form, for more mathematical details, and for all proofs we refer to [8, Sec. IV], [3, Sec. 2.4].

Lemma 18. *Consider a memoryless channel with input* $\mathbf{s} \in \mathbb{C}^{n_\mathrm{R}}$ *and output* $T \in \mathbb{C}$. *Then for an arbitrary distribution on the input* $\mathbf{S}$ *the mutual information between input and output of the channel is upper-bounded as follows:*

$$\begin{aligned} I(\mathbf{S};T) \leq{}& -h(T|\mathbf{S}) + \log\pi + \alpha\log\beta + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) \\ &+ (1-\alpha)\mathsf{E}\bigl[\log\bigl(|T|^2+\nu\bigr)\bigr] + \frac{1}{\beta}\mathsf{E}\bigl[|T|^2\bigr] + \frac{\nu}{\beta} \end{aligned} \tag{4.4}$$

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

Proof. A proof can be found in [8, Sec. IV], [3, Sec. 2.4]. □

4.3 Capacity-Achieving Input Distributions and Circular Symmetry

The final preliminary remark concerns circular symmetry. We say that a random vector $\mathbf{W}$ is *circularly symmetric* if

$$\mathbf{W} \overset{\mathscr{L}}{=} \mathbf{W}\cdot e^{\mathrm{i}\Theta} \tag{4.5}$$

where $\Theta \sim \mathcal{U}([0, 2\pi])$ is independent of $\mathbf{W}$ and where $\overset{\mathscr{L}}{=}$ stands for "equal in law". Note that this is not to be confused with *isotropically distributed*, which means that a vector has equal probability to point in every direction. Circular symmetry only concerns the phase of the components of a vector, not the vector's direction.

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

Lemma 19. *Assume a channel as given in* (2.20). *Then the capacity-achieving input distribution can be assumed to be circularly symmetric,* i.e., *the input vector* $\mathbf{X}$ *can be replaced by* $\mathbf{X}e^{\mathrm{i}\Theta}$, *where* $\Theta \sim \mathcal{U}([0, 2\pi])$ *is independent of every other random quantity.*

Proof. This is a special case of Proposition 30 that is proven in Chapter 6. □

4.4 Fading Number of a General Memoryless MIMO Fading Channel

We are now ready for the main result, *i.e.*, the fading number of a memoryless MIMO fading channel.

Theorem 20. *Consider a memoryless MIMO fading channel* (2.20) *where the random fading matrix* $\mathbb{H}$ *takes value in* $\mathbb{C}^{n_\mathrm{R} \times n_\mathrm{T}}$ *and satisfies*

$$h(\mathbb{H}) > -\infty \tag{4.6}$$

and

$$\mathsf{E}\left[\|\mathbb{H}\|_\mathrm{F}^2\right] < \infty. \tag{4.7}$$

Then, irrespective of whether a peak-power constraint (2.23) *or an average-power constraint* (2.24) *is imposed on the input, the fading number* $\chi(\mathbb{H})$ *is given by*

$$\boxed{\chi(\mathbb{H}) = \sup_{Q_{\hat{\mathbf{X}}}} \left\{ h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + n_\mathrm{R}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\right] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) \right\}} \tag{4.8}$$

where $\hat{\mathbf{X}}$ *denotes a random vector of unit length.*
Moreover, this fading number is achievable by a random vector $\mathbf{X} = \hat{\mathbf{X}} \cdot R$ *where* $\hat{\mathbf{X}}$ *is distributed according to the distribution that achieves the fading number in* (4.8) *and where* R *is a non-negative random variable such that*

$$\log R^2 \sim \mathcal{U}([\log\log\mathcal{E}, \log\mathcal{E}]). \tag{4.9}$$

Proof. The proof of Theorem 20 consists of two parts: firstly we derive an upper bound to the fading number assuming an average-power constraint (2.24) on the input. The key ingredients here are the ideas described above, *i.e.*, the fact that the capacity-achieving input distribution must escape to infinity and can be assumed to be circularly symmetric and the duality-based technique of bounding capacity.

Secondly we show that this upper bound can actually be achieved by an input that satisfies the peak-power constraint (2.23). Since a peak-power constraint is more restrictive than the corresponding average-power constraint, the theorem follows.

In the following we will merely give a short overview over the proof. For more details we refer to [29].

The upper bound relies strongly on Proposition 17 which says that the input can be assumed to take on large values only, *i.e.*, at high SNR the additive noise will become negligible so that we can bound

$$I(\mathbf{X};\mathbf{Y}) \lessapprox I(\mathbf{X};\mathbb{H}\mathbf{X}). \tag{4.10}$$

This term is then split into a term that only considers the magnitude of $\mathbb{H}\mathbf{X}$ and a term that takes into account the direction:

$$I(\mathbf{X};\mathbb{H}\mathbf{X}) = I(\mathbf{X};\|\mathbb{H}\mathbf{X}\|) + I\left(\mathbf{X};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\mathbf{X}\|\right). \tag{4.11}$$

For the first term—which is related to MISO fading—we then use the bounding technique of Lemma 18.

Because Proposition 17 only holds in the limit when $\mathcal{E}$ tends to infinity, we introduce an event $\|\mathbf{X}\|^2 > \mathcal{E}_0$ for some fixed $\mathcal{E}_0 \geq 0$ and condition everything on this event.

To derive a lower bound on capacity we choose a specific input distribution of the form

$$\mathbf{X} = R\cdot\hat{\mathbf{X}} \tag{4.12}$$

where the distribution of R is such that it achieves the fading number of a SIMO fading channel and where the distribution of $\hat{\mathbf{X}}$ is independent of R and will be only specified at the very end of the derivation (it will be chosen to maximize the fading number). We then split the mutual information into two terms:

$$I(\mathbf{X};\mathbf{Y}) = I\bigl(R;\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\bigr) + I\bigl(\hat{\mathbf{X}};\mathbf{Y}\bigr). \tag{4.13}$$

The first term (almost) corresponds to a SIMO fading channel with side-information for which the fading number is known. The second term is treated separately. □

Note that—even if it is not obvious on a first sight—the optimal choice of $Q_{\hat{\mathbf{X}}}$ is circularly symmetric. To see this note that for an arbitrary random unit vector $\hat{\mathbf{X}}$ and for some arbitrary random $e^{\mathrm{i}\Phi}$ that is independent of $\mathbb{H}$ (but possibly might be dependent on $\hat{\mathbf{X}}$) we have

$$\|\mathbb{H}\hat{\mathbf{X}}e^{\mathrm{i}\Phi}\| \overset{\mathscr{L}}{=} \|\mathbb{H}\hat{\mathbf{X}}\| \tag{4.14}$$

and

$$\begin{aligned} h\bigl(\mathbb{H}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}}\bigr) &= h\bigl(\mathbb{H}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}},e^{\mathrm{i}\Phi}\bigr) && (4.15)\\ &= h\bigl(\mathbb{H}\hat{\mathbf{X}}e^{\mathrm{i}\Phi}\,\big|\,\hat{\mathbf{X}}e^{\mathrm{i}\Phi},e^{\mathrm{i}\Phi}\bigr) && (4.16)\\ &= h\bigl(\mathbb{H}\hat{\mathbf{X}}'\,\big|\,\hat{\mathbf{X}}',e^{\mathrm{i}\Phi}\bigr) && (4.17)\\ &= h\bigl(\mathbb{H}\hat{\mathbf{X}}'\,\big|\,\hat{\mathbf{X}}'\bigr), && (4.18)\end{aligned}$$

where "$\overset{\mathscr{L}}{=}$" stands for "identical in law", where we have introduced $\hat{\mathbf{X}}' = \hat{\mathbf{X}}e^{\mathrm{i}\Phi}$, and where in the last equality we have used the fact that given $\hat{\mathbf{X}}'$, $e^{\mathrm{i}\Phi}$ and $\mathbb{H}\hat{\mathbf{X}}'$ are independent.

Hence, in (4.8) the only term that depends on the choice of the phase distribution of $\hat{\mathbf{X}}$ is $h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right)$. This term is maximized for a circularly symmetric $\hat{\mathbf{X}}$ as can be seen as follows:

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) = I\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|};\hat{\mathbf{X}}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) \tag{4.19}$$

$$= I\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|};\hat{\mathbf{X}}\,\middle|\,e^{\mathrm{i}\Theta}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) \tag{4.20}$$

$$= I\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta};\hat{\mathbf{X}}\,\middle|\,e^{\mathrm{i}\Theta}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) \tag{4.21}$$

$$= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,e^{\mathrm{i}\Theta}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}},e^{\mathrm{i}\Theta}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) \tag{4.22}$$

$$= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,e^{\mathrm{i}\Theta}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}},e^{\mathrm{i}\Theta}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) \tag{4.23}$$

$$= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,e^{\mathrm{i}\Theta}\right) \tag{4.24}$$

$$\leq h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\right). \tag{4.25}$$

Here the first equality follows from the chain rule; the subsequent equality because $e^{\mathrm{i}\Theta}$ is independent of all other random quantities; (4.23) follows from scaling property of differential entropy by a constant of magnitude 1; and the final inequality from conditioning that reduces entropy.

Hence, a circularly symmetric input will always achieve a mutual information that is at least as large as any other input.

The evaluation of (4.8) can be pretty awkward mainly due to the first term, *i.e.*, the differential entropy with respect to the surface area measure λ. We therefore will derive next an upper bound to the fading number that is easier to evaluate.

To that goal firstly note that for an arbitrary constant non-singular $n_\mathrm{R} \times n_\mathrm{R}$ matrix A and an arbitrary constant non-singular $n_\mathrm{T} \times n_\mathrm{T}$ matrix B

$$\chi(\mathsf{A}\mathbb{H}\mathsf{B}) = \chi(\mathbb{H}) \tag{4.26}$$

see [8, Lem. 4.7], [3, Lem. 6.14]. Secondly, note that for an arbitrary random unit vector $\hat{\mathbf{Y}} \in \mathbb{C}^{n_\mathrm{R}}$

$$h_\lambda\big(\hat{\mathbf{Y}}\big) \leq \log c_{n_\mathrm{R}} = \log \frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})} \tag{4.27}$$

where c_{n_R} denotes the surface area of the unit sphere in $\mathbb{C}^{n_\mathrm{R}}$ as defined in (2.10) and where the upper bound is achieved with equality only if $\hat{\mathbf{Y}}$ is uniformly distributed on the sphere, *i.e.*, $\hat{\mathbf{Y}}$ is isotropically distributed.

Using these two observations we get the following upper bound on the fading number.

Corollary 21. *The fading number of a memoryless MIMO fading channel as defined in Theorem 20 can be upper-bounded as follows:*

$$\chi(\mathbb{H}) \le n_\mathrm{R} \log \pi - \log \Gamma(n_\mathrm{R}) + \inf_{\mathsf{A},\mathsf{B}} \sup_{\hat{\mathbf{x}}} \left\{ n_\mathrm{R} \mathsf{E}\left[\log \|\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{x}}\|^2\right] - h\bigl(\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{x}}\bigr) \right\} \tag{4.28}$$

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

Proof. From the two observations (4.26) and (4.27) mentioned above we immediately get from Theorem 20:

$$\chi(\mathbb{H}) \le \inf_{\mathsf{A},\mathsf{B}} \sup_{Q_{\hat{\mathbf{X}}}} \mathsf{E}_{\hat{\mathbf{X}}} \Biggl[n_\mathrm{R} \log \pi - \log \Gamma(n_\mathrm{R}) + n_\mathrm{R} \mathsf{E}\left[\log \|\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{X}}\|^2 \,\Big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}} \right] - h\bigl(\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}}\bigr) \Biggr]. \tag{4.29}$$

The result now follows by noting that (4.28) can always be achieved by choosing $Q_{\hat{\mathbf{X}}}$ in (4.29) to be the distribution which with probability 1 takes on the value $\hat{\mathbf{x}}$ that achieves the maximum in (4.28). □

This upper bound is possibly tighter than the upper bound given in [8, Lem. 4.14], [3, Lem. 6.16] because of the additional infimum over B.

4.5 Some Known Special Cases

In this section we will briefly show how some already known results of various fading numbers can be derived as special cases from this new more general result.

We start with the situation of a fading matrix that is *rotation-commutative in the generalized sense*, *i.e.*, the fading matrix $\mathbb{H}$ is such that for every constant unitary $n_\mathrm{T} \times n_\mathrm{T}$ matrix V_t there exists an $n_\mathrm{R} \times n_\mathrm{R}$ constant unitary matrix V_r such that

$$\mathsf{V}_r \mathbb{H} \stackrel{\mathscr{L}}{=} \mathbb{H}\mathsf{V}_t \tag{4.30}$$

where $\stackrel{\mathscr{L}}{=}$ stands for "has the same law"; and for every constant unitary $n_\mathrm{R} \times n_\mathrm{R}$ matrix V_r there exists a constant unitary $n_\mathrm{T} \times n_\mathrm{T}$ matrix V_t such that (4.30) holds [8, Def. 4.37], [3, Def. 6.37].

The property of rotation-commutativity for random matrices is a generalization of the isotropic distribution of random vectors, *i.e.*, we have the following lemma.

Lemma 22. *Let $\mathbb{H}$ be rotation-commutative in the generalized sense. Then the following two statements hold:*

- *If $\hat{\mathbf{X}} \in \mathbb{C}^{n_\mathrm{T}}$ is an isotropically distributed random vector that is independent of $\mathbb{H}$, then $\mathbb{H}\hat{\mathbf{X}} \in \mathbb{C}^{n_\mathrm{R}}$ is isotropically distributed.*
- *If $\hat{\mathbf{e}}, \hat{\mathbf{e}}' \in \mathbb{C}^{n_\mathrm{T}}$ are two constant unit vectors, then*

$$\|\mathbb{H}\hat{\mathbf{e}}\| \stackrel{\mathscr{L}}{=} \|\mathbb{H}\hat{\mathbf{e}}'\|, \quad \|\hat{\mathbf{e}}\| = \|\hat{\mathbf{e}}'\| = 1 \tag{4.31}$$

$$h\bigl(\mathbb{H}\hat{\mathbf{e}}\bigr) = h\bigl(\mathbb{H}\hat{\mathbf{e}}'\bigr), \quad \|\hat{\mathbf{e}}\| = \|\hat{\mathbf{e}}'\| = 1. \tag{4.32}$$

Proof. For a proof see, *e.g.*, [8, Lem. 4.38], [3, Lem. 6.38]. □

From Lemma 22 it immediately follows that in the situation of rotation-commutative fading the only term in the expression of the fading number (4.8) that depends on $Q_{\hat{\mathbf{X}}}$ is

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right).$$

This entropy is maximized if $\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}$ is uniformly distributed on the surface of the n_R-dimensional complex unit sphere, which can be achieved according to Lemma 22 by the choice of an isotropic distribution for $Q_{\hat{\mathbf{X}}}$. Then

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) = \log c_{n_\mathrm{R}} \tag{4.33}$$

where c_{n_R} is the surface area of a unit sphere in $\mathbb{C}^{n_\mathrm{R}}$, *i.e.*,

$$c_{n_\mathrm{R}} = \frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})}. \tag{4.34}$$

The fading number then becomes

$$\chi(\mathbb{H}) = \log\frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})} - \log 2 + n_\mathrm{R}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{e}}\|^2\right] - h(\mathbb{H}\hat{\mathbf{e}}) \tag{4.35}$$

where $\hat{\mathbf{e}}$ is an arbitrary constant unit vector in $\mathbb{C}^{n_\mathrm{T}}$.

In case of a SIMO fading channel, $\hat{\mathbf{X}} = e^{\mathrm{i}\Phi}$. Since the optimal distribution in (4.8) is a circularly symmetric unit random variable, we get $\hat{\mathbf{X}} = e^{\mathrm{i}\Theta}$ and (4.8) becomes:

$$\chi(\mathbf{H}) = h_\lambda\big(\hat{\mathbf{H}}e^{\mathrm{i}\Theta}\big) + \mathsf{E}\left[\log\|\mathbf{H}\|^2\right] - \log 2 - h(\mathbf{H}). \tag{4.36}$$

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

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

is identical to the distribution of $e^{\mathrm{i}\Theta}$. Since the optimal choice of $Q_{\hat{\mathbf{X}}}$ is circularly symmetric, we get

$$h_\lambda\left(\frac{\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}}{|\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}|}\right) = h_\lambda\big(e^{\mathrm{i}\Theta}\big) = \log 2\pi \tag{4.37}$$

and the fading number becomes

$$\begin{aligned}
\chi(\mathbf{H}^\mathsf{T}) &= \sup_{Q_{\hat{\mathbf{X}}}}\left\{\log 2\pi + \mathsf{E}\left[\log|\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}|^2\right] - \log 2 - h\big(\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}}\big)\right\} && (4.38)\\
&= \sup_{Q_{\hat{\mathbf{X}}}}\mathsf{E}_{\hat{\mathbf{X}}}\left[\log\pi + \mathsf{E}\left[\log|\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}|^2\,\Big|\,\hat{\mathbf{X}}=\hat{\mathbf{x}}\right] - h\big(\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}\,\big|\,\hat{\mathbf{X}}=\hat{\mathbf{x}}\big)\right] && (4.39)\\
&\le \sup_{\hat{\mathbf{x}}}\left\{\log\pi + \mathsf{E}\left[\log|\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}|^2\right] - h\big(\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}\big)\right\} && (4.40)
\end{aligned}$$

which can be achieved by an input

$$\hat{\mathbf{X}} = \hat{\mathbf{x}}e^{\mathrm{i}\Theta} \tag{4.41}$$

where $\hat{\mathbf{x}}$ is the deterministic unit vector that achieves the fading number in (4.40).

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

$$h_\lambda\bigl(e^{\mathrm{i}\Theta}\bigr) = \log 2\pi. \tag{4.42}$$

This yields

$$\chi(H) = \log 2\pi + \mathsf{E}\bigl[\log |H|^2\bigr] - \log 2 - h(H) \tag{4.43}$$

$$= \log \pi + \mathsf{E}\bigl[\log |H|^2\bigr] - h(H). \tag{4.44}$$

Chapter 5

An IID MIMO Gaussian Fading Channel with a Scalar Line-of-Sight Component

The expression of the fading number for a general MIMO fading channel given in Chapter 4 is very general. Unfortunately however, it is rather difficult to evaluate and is therefore partially hiding some insight. In this chapter we would like to evaluate the general expression of the memoryless MIMO fading number in the special situation of an IID Gaussian fading channel with a scalar line-of-sight component.[1] We will show that for a large line-of-sight component d the fading number basically grows like

$$\chi = n_{\mathrm{m}} \log |d|^2 \tag{5.1}$$

where

$$n_{\mathrm{m}} = \min\{n_{\mathrm{R}}, n_{\mathrm{T}}\} \tag{5.2}$$

is the degree of freedom as defined, *e.g.*, in [34]. Here n_{R} and n_{T} denote the number of antennas at the receiver and transmitter, respectively.

5.1 Channel Model

In this chapter we further specialize the memoryless channel model (2.20): we assume that the fading matrix $\mathbb{H}$ can be written as

$$\mathbb{H} = \mathsf{D} + \tilde{\mathbb{H}} \tag{5.3}$$

where all components of the $n_{\mathrm{R}} \times n_{\mathrm{T}}$ random matrix $\tilde{\mathbb{H}}$ are IID $\sim \mathcal{N}_{\mathbb{C}}(0,1)$ and where the constant $n_{\mathrm{R}} \times n_{\mathrm{T}}$ line-of-sight matrix D is scalar in the sense that, for $n_{\mathrm{R}} \leq n_{\mathrm{T}}$,

$$\mathsf{D} = d \begin{pmatrix} \mathsf{I}_{n_{\mathrm{R}}} & \mathsf{0}_{n_{\mathrm{R}} \times (n_{\mathrm{T}} - n_{\mathrm{R}})} \end{pmatrix} \tag{5.4}$$

or, for $n_{\mathrm{R}} > n_{\mathrm{T}}$,

$$\mathsf{D} = d \begin{pmatrix} \mathsf{I}_{n_{\mathrm{T}}} \\ \mathsf{0}_{(n_{\mathrm{R}} - n_{\mathrm{T}}) \times n_{\mathrm{T}}} \end{pmatrix}, \tag{5.5}$$

where $d \in \mathbb{C}$ is a constant.

[1] For precise definitions we refer to Section 5.1.

We know from Chapter 4 that the fading number is given as

$$\chi(\mathbb{H}) = \sup_{Q_{\hat{\mathbf{X}}}} \left\{ h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + n_\mathrm{R}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\right] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big)\right\}, \tag{5.6}$$

independently of the type of power constraint (2.23) or (2.24) that is imposed on the input. Here $\hat{\mathbf{X}}$ denotes a random vector of unit length and $Q_{\hat{\mathbf{X}}}$ denotes its probability law, *i.e.*, the supremum is taken over all distributions of the random unit-vector $\hat{\mathbf{X}}$. Note that the expectation in the second term is understood jointly over $\mathbb{H}$ and $\hat{\mathbf{X}}$.

Moreover, we know that this fading number is achievable by a random vector $\mathbf{X} = \hat{\mathbf{X}} \cdot R$ where $\hat{\mathbf{X}}$ is distributed according to the distribution that achieves the supremum in (5.6) and where R is a non-negative random variable independent of $\hat{\mathbf{X}}$ such that (4.9) is satisfied.

In the following we will try to evaluate the expression (5.6) in the situation of IID MIMO Gaussian fading with a scalar line-of-sight component as defined in (5.3)–(5.5).

5.2 Fading Number of an IID Gaussian Fading Channel with a Scalar Line-of-Sight Component

We make a case distinction: we start with the situation $n_\mathrm{R} \le n_\mathrm{T}$ which turns out to be easier to solve.

Theorem 23. *Assume $n_\mathrm{R} \le n_\mathrm{T}$ and a Gaussian fading matrix as given in* (5.3) *and* (5.4)*. Then*

$$\boxed{\chi(\mathbb{H}) = n_\mathrm{R} g_{n_\mathrm{R}}\big(|d|^2\big) - n_\mathrm{R} - \log\Gamma(n_\mathrm{R})} \tag{5.7}$$

where $g_m(\cdot)$ is defined in (3.7)*. The fading number is achievable by an input $\mathbf{X} = R\cdot\hat{\mathbf{X}}^*$ with $R \perp\!\!\!\perp \hat{\mathbf{X}}^*$, where the distribution of $R \in \mathbb{R}^+$ is specified in* (4.9) *and where*

$$\hat{\mathbf{X}}^* = \begin{pmatrix} \mathbf{\Xi}^* \\ \mathbf{0} \end{pmatrix} \tag{5.8}$$

with $\mathbf{\Xi}^* \in \mathbb{C}^{n_\mathrm{R}}$ being an isotropically distributed *unit vector.*

Proof. We write for the unit vector $\hat{\mathbf{X}}$ of the maximization in (5.6)

$$\hat{\mathbf{X}} = \begin{pmatrix} \mathbf{\Xi} \\ \mathbf{\Xi}' \end{pmatrix} \tag{5.9}$$

where $\mathbf{\Xi} \in \mathbb{C}^{n_\mathrm{R}}$ and $\mathbf{\Xi}' \in \mathbb{C}^{n_\mathrm{T}-n_\mathrm{R}}$. Note that from $\|\hat{\mathbf{X}}\|^2 = 1$ it follows that $\|\mathbf{\Xi}\|^2 \le 1$. Then

$$\mathbb{H}\hat{\mathbf{X}} = \mathsf{D}\hat{\mathbf{X}} + \tilde{\mathbb{H}}\hat{\mathbf{X}} \stackrel{\mathscr{L}}{=} d\mathbf{\Xi} + \tilde{\mathbf{H}} \tag{5.10}$$

where $\tilde{\mathbf{H}} \sim \mathcal{N}_\mathbb{C}(\mathbf{0}, \mathsf{I}_{n_\mathrm{R}})$. Hence,

$$h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) = h\big(\tilde{\mathbf{H}}\big) = n_\mathrm{R}\log\pi e; \tag{5.11}$$

$$n_\mathrm{R}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\right] = n_\mathrm{R} g_{n_\mathrm{R}}\big(|d|^2\|\mathbf{\Xi}\|^2\big) \tag{5.12}$$

$$\le n_\mathrm{R} g_{n_\mathrm{R}}\big(|d|^2\big); \tag{5.13}$$

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) \le \log\frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})}. \tag{5.14}$$

Here, (5.12) follows from the fact that $\|d\boldsymbol{\Xi}+\tilde{\mathbf{H}}\|^2$ is noncentral chi-square distributed and from (3.13). The inequality (5.13) follows from the monotonicity of $g_m(\cdot)$ and the fact that $\|\boldsymbol{\Xi}\|^2 \leq 1$. It is tight if $\|\boldsymbol{\Xi}\|^2 = 1$, *i.e.*, $\boldsymbol{\Xi}' = \mathbf{0}$. The inequality (5.14) follows from (2.9) and (2.10) and is tight if $\boldsymbol{\Xi}$ is uniformly distributed on the unit sphere in $\mathbb{C}^{n_\mathrm{R}}$ so that $\mathbb{H}\hat{\mathbf{X}}$ is isotropically distributed. The result now follows from (5.6). ☐

The case $n_\mathrm{R} > n_\mathrm{T}$ is more difficult since then (5.14) is in general not tight. We firstly need to introduce some notation. Note that

$$\mathbb{H}\hat{\mathbf{X}} = \mathsf{D}\hat{\mathbf{X}} + \tilde{\mathbb{H}}\hat{\mathbf{X}} \overset{\mathscr{L}}{=} \begin{pmatrix} d\hat{\mathbf{X}} \\ \mathbf{0} \end{pmatrix} + \tilde{\mathbf{H}} \tag{5.15}$$

where $\tilde{\mathbf{H}} \sim \mathcal{N}_\mathbb{C}(\mathbf{0}, \mathsf{I}_{n_\mathrm{R}})$. Let us split the vector $\tilde{\mathbf{H}}$ into two parts:

$$\tilde{\mathbf{H}} = \begin{pmatrix} \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \end{pmatrix} \tag{5.16}$$

where $\tilde{\mathbf{H}}_1 \sim \mathcal{N}_\mathbb{C}(\mathbf{0}, \mathsf{I}_{n_\mathrm{T}})$ and $\tilde{\mathbf{H}}_2 \sim \mathcal{N}_\mathbb{C}(\mathbf{0}, \mathsf{I}_{n_\mathrm{R}-n_\mathrm{T}})$ are two independent white Gaussian random vectors in $\mathbb{C}^{n_\mathrm{T}}$ and $\mathbb{C}^{n_\mathrm{R}-n_\mathrm{T}}$, respectively. Then we can write

$$\mathbb{H}\hat{\mathbf{X}} \overset{\mathscr{L}}{=} \begin{pmatrix} d\hat{\mathbf{X}} + \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \end{pmatrix}. \tag{5.17}$$

Next we define

$$S_1 \triangleq \left\|d\hat{\mathbf{X}} + \tilde{\mathbf{H}}_1\right\|^2, \tag{5.18}$$

$$S_2 \triangleq \left\|\tilde{\mathbf{H}}_2\right\|^2. \tag{5.19}$$

Note that S_1 is noncentral chi-square distributed with $2n_\mathrm{T}$ degrees of freedom and noncentrality parameter $\|d\hat{\mathbf{X}}\|^2 = |d|^2$ independently of the distribution of $\hat{\mathbf{X}}$, and that S_2 is central chi-square distributed with $2(n_\mathrm{R} - n_\mathrm{T})$ degrees of freedom. Moreover, S_1 and S_2 are independent of each other.

Theorem 24. *Assume $n_\mathrm{R} > n_\mathrm{T}$ and a Gaussian fading matrix as given in* (5.3) *and* (5.5)*. Then*

$$\boxed{\chi(\mathbb{H}) = n_\mathrm{T} g_{n_\mathrm{T}}\big(|d|^2\big) - n_\mathrm{T} - \log\Gamma(n_\mathrm{T}) + I\left(S_1; \frac{S_2}{S_1}\right)} \tag{5.20}$$

where $g_m(\cdot)$ is defined in (3.7)*. The fading number is achievable by an input $\mathbf{X} = R\cdot\hat{\mathbf{X}}^*$ with $R \perp\!\!\!\perp \hat{\mathbf{X}}^*$, where the distribution of $R \in \mathbb{R}^+$ is specified in* (4.9) *and where $\hat{\mathbf{X}}^*$ is an* isotropically distributed *unit vector.*

Proof. See Section 5.3.1. ☐

Unfortunately, we have not succeeded in deriving the term $I(S_1; S_2/S_1)$ precisely. Instead, we will state an upper and a lower bound that both do not depend on d. This shows that $I(S_1; S_2/S_1)$ is bounded in d.

Claim 25. *For $n_R > n_T$ let S_1 and S_2 be independent random variables defined in* (5.18) *and* (5.19)*, respectively. Then for $n_T > 1$*

$$0 \leq I\left(S_1; \frac{S_2}{S_1}\right) \leq (n_R - n_T - 1)\psi(n_R - n_T) \\ -(n_R - n_T - 1) - \log \frac{\Gamma(n_R - n_T)}{n_R - n_T} + \log \frac{n_T}{n_T - 1}, \tag{5.21}$$

and for $n_T = 1$

$$0 \leq I\left(S_1; \frac{S_2}{S_1}\right) \leq (n_R - 2)\psi(n_R - 1) - (n_R - 2) \\ - \log \frac{\Gamma(n_R - 1)}{n_R - 1} + \log \frac{\pi}{2} + 1. \tag{5.22}$$

Proof. See Section 5.3.2. □

Hence, combining Theorem 23, Theorem 24, and Claim 25 we get the following corollary.

Corollary 26. *The fading number of the IID MIMO Gaussian fading channel as defined in* (5.3)–(5.5) *is given by*

$$\boxed{\chi(\mathbb{H}) = n_m g_{n_m}\left(|d|^2\right) - n_m - \log \Gamma\left(n_m\right) + f(n_R, n_T; d)} \tag{5.23}$$

where

$$n_m = \min\{n_R, n_T\} \tag{5.24}$$

denotes the degree of freedom and where $f(n_R, n_T; d)$ depends primarily on n_T and n_R and is bounded in d:

$$0 \leq f(n_R, n_T; d) \leq (n_R - n_T - 1)\psi(n_R - n_T) \\ - (n_R - n_T - 1) - \log \frac{\Gamma(n_R - n_T)}{n_R - n_T} \\ + \log\left(\min\left\{\frac{n_T}{n_T - 1}, \frac{\pi e}{2}\right\}\right). \tag{5.25}$$

Using (3.36) *we therefore have for $|d| \gg 1$*

$$\chi \sim n_m \log\left(|d|^2 + n_m\right). \tag{5.26}$$

Hence, we have evaluated the expression of the fading number of a memoryless IID Gaussian fading channel with scalar line-of-sight component d. We have shown that for large d, the fading number grows like $n_m g_{n_m}\left(|d|^2\right)$ where $g_{n_m}\left(|d|^2\right)$ grows like $\log |d|^2$ and where n_m, denoted *degree of freedom*, is the smaller of the number of antennas at transmitter and receiver, respectively.

Moreover, we have shown that the optimal input is basically isotropically distributed, *i.e.*, for $n_R \leq n_T$

$$\hat{\mathbf{X}}^* = \begin{pmatrix} \mathbf{\Xi}^* \\ \mathbf{0} \end{pmatrix} \tag{5.27}$$

with $\mathbf{\Xi}^* \in \mathbb{C}^{n_R}$ being an *isotropically distributed* unit vector, and for $n_R > n_T$

$$\hat{\mathbf{X}}^* = \mathbf{\Xi}^* \tag{5.28}$$

with $\mathbf{\Xi}^* \in \mathbb{C}^{n_T}$ being an *isotropically distributed* unit vector.

5.3 Proofs

5.3.1 Proof of Theorem 24

We start with the derivation of an optimal input distribution. From (5.6) we know that the optimal input is given as $\mathbf{X} = R \cdot \hat{\mathbf{X}}$ and (4.9) specifies an optimal choice of R. It remains to prove that an isotropic unit vector $\hat{\mathbf{X}}$ is optimal. To that goal let V be an arbitrary (deterministic) unitary $n_\mathrm{T} \times n_\mathrm{T}$ matrix and define

$$\mathsf{U} \triangleq \begin{pmatrix} \mathsf{V} & \mathsf{0}_{n_\mathrm{T}\times(n_\mathrm{R}-n_\mathrm{T})} \\ \mathsf{0}_{(n_\mathrm{R}-n_\mathrm{T})\times n_\mathrm{T}} & \mathsf{I}_{n_\mathrm{R}-n_\mathrm{T}} \end{pmatrix}. \tag{5.29}$$

Note that U is a unitary $n_\mathrm{R} \times n_\mathrm{R}$ matrix. Now we compute

$$\begin{aligned} \mathsf{U}\mathbb{H}\hat{\mathbf{X}} &\overset{\mathscr{L}}{=} \begin{pmatrix} \mathsf{V} & \mathsf{0}_{n_\mathrm{T}\times(n_\mathrm{R}-n_\mathrm{T})} \\ \mathsf{0}_{(n_\mathrm{R}-n_\mathrm{T})\times n_\mathrm{T}} & \mathsf{I}_{n_\mathrm{R}-n_\mathrm{T}} \end{pmatrix} \begin{pmatrix} d\hat{\mathbf{X}} + \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \end{pmatrix} && (5.30) \\ &= \begin{pmatrix} d\mathsf{V}\hat{\mathbf{X}} + \mathsf{V}\tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \end{pmatrix} \overset{\mathscr{L}}{=} \begin{pmatrix} d\mathsf{V}\hat{\mathbf{X}} + \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \end{pmatrix} && (5.31) \\ &\overset{\mathscr{L}}{=} \begin{pmatrix} d\mathsf{V}\hat{\mathbf{X}} \\ \mathbf{0} \end{pmatrix} + \tilde{\mathbb{H}}\hat{\mathbf{X}}' = \begin{pmatrix} d\mathsf{V}\hat{\mathbf{X}} \\ \mathbf{0} \end{pmatrix} + \tilde{\mathbb{H}}\mathsf{V}\hat{\mathbf{X}} && (5.32) \\ &= \mathbb{H}\mathsf{V}\hat{\mathbf{X}}, && (5.33) \end{aligned}$$

where we replace $\mathbb{H}\hat{\mathbf{X}}$ by (5.15); where we use that $\mathsf{V}\tilde{\mathbf{H}}_1 \overset{\mathscr{L}}{=} \tilde{\mathbf{H}}_1$ for any unitary V; and where $\hat{\mathbf{X}}'$ is an arbitrary unit vector that we then choose to be $\mathsf{V}\hat{\mathbf{X}}$. Hence, for any choice of V we can find a U such that $\mathsf{U}\mathbb{H}\hat{\mathbf{X}} \overset{\mathscr{L}}{=} \mathbb{H}\mathsf{V}\hat{\mathbf{X}}$.

Let's now make V random, *i.e.*, we define a $n_\mathrm{T} \times n_\mathrm{T}$ unitary matrix $\mathbb{V}$ that is Haar distributed[2] and independent of $(\mathbb{H}, \mathbf{X}, \mathbf{Z})$. Further we define a random unitary $n_\mathrm{R} \times n_\mathrm{R}$ matrix $\mathbb{U}$ analogously to (5.29):

$$\mathbb{U} \triangleq \begin{pmatrix} \mathbb{V} & \mathsf{0}_{n_\mathrm{T}\times(n_\mathrm{R}-n_\mathrm{T})} \\ \mathsf{0}_{(n_\mathrm{R}-n_\mathrm{T})\times n_\mathrm{T}} & \mathsf{I}_{n_\mathrm{R}-n_\mathrm{T}} \end{pmatrix}. \tag{5.34}$$

Then we get:

$$\begin{aligned} &I\bigl(\mathbf{X}; \mathbb{H}\mathbf{X} + \mathbf{Z}\bigr) \\ &\quad= I\bigl(\mathbf{X}; \mathbb{H}\mathbf{X} + \mathbf{Z} \bigm| \mathbb{U}\bigr) && (5.35) \\ &\quad= I\bigl(\mathbb{V}\mathbf{X}; \mathbb{U}\mathbb{H}\mathbf{X} + \mathbb{U}\mathbf{Z} \bigm| \mathbb{U}\bigr) && (5.36) \\ &\quad= I\bigl(\mathbb{V}\mathbf{X}; \mathbb{H}\mathbb{V}\mathbf{X} + \mathbf{Z} \bigm| \mathbb{U}\bigr) && (5.37) \\ &\quad= I\bigl(\breve{\mathbf{X}}; \mathbb{H}\breve{\mathbf{X}} + \mathbf{Z} \bigm| \mathbb{U}\bigr) && (5.38) \\ &\quad= h\bigl(\mathbb{H}\breve{\mathbf{X}} + \mathbf{Z} \bigm| \mathbb{U}\bigr) - h\bigl(\mathbb{H}\breve{\mathbf{X}} + \mathbf{Z} \bigm| \breve{\mathbf{X}}, \mathbb{U}\bigr) && (5.39) \\ &\quad= h\bigl(\mathbb{H}\breve{\mathbf{X}} + \mathbf{Z} \bigm| \mathbb{U}\bigr) - h\bigl(\mathbb{H}\breve{\mathbf{X}} + \mathbf{Z} \bigm| \breve{\mathbf{X}}\bigr) && (5.40) \\ &\quad\le h\bigl(\mathbb{H}\breve{\mathbf{X}} + \mathbf{Z}\bigr) - h\bigl(\mathbb{H}\breve{\mathbf{X}} + \mathbf{Z} \bigm| \breve{\mathbf{X}}\bigr) && (5.41) \\ &\quad= I\bigl(\breve{\mathbf{X}}; \mathbb{H}\breve{\mathbf{X}} + \mathbf{Z}\bigr). && (5.42) \end{aligned}$$

Here, the first equality follows since $\mathbb{U}$ is independent of the other random quantities; the subsequent equality follows because given $\mathbb{U}$ also $\mathbb{V}$ is known and because mutual information is not changed when the arguments are multiplied by known invertible

[2] A random matrix $\mathbb{T}$ is Haar distributed if for any deterministic unitary matrix M we have that $\mathsf{M}\mathbb{T} \overset{\mathscr{L}}{=} \mathbb{T}$.

quantities; in the subsequent equality we use that $\mathbb{U}\mathbb{H}\hat{\mathbf{X}} \overset{\mathscr{L}}{=} \mathbb{H}\mathbb{V}\hat{\mathbf{X}}$ as shown above and that $\mathbb{U}\mathbf{Z} \overset{\mathscr{L}}{=} \mathbf{Z}$; in (5.38) we introduce $\check{\mathbf{X}} \triangleq \mathbb{V}\mathbf{X}$; the subsequent equality follows from the definition of mutual information; then we use the fact that $\mathbb{H}\check{\mathbf{X}} + \mathbf{Z}\big|\check{\mathbf{X}}$ is independent of $\mathbb{U}$ or $\mathbb{V}$; and the inequality follows from conditioning that cannot increase entropy.

Note that $\check{\mathbf{X}}$ is isotropically distributed independently of the distribution of $\mathbf{X}$. Hence, an isotropic input will achieve at least the same mutual information as any other input.

Next, we will derive the expression (5.20). To that goal we start with Lemma 2 and plug (2.14) into (2.12):

$$h_\lambda\big(\hat{\mathbf{V}} \,\big|\, \|\mathbf{V}\|\big) = h(\mathbf{V}) - h(\|\mathbf{V}\|^2) + \log 2 - (m-1)\mathsf{E}\big[\log \|\mathbf{V}\|^2\big]. \tag{5.43}$$

We now choose $\mathbf{V} \triangleq \mathbb{H}\hat{\mathbf{X}}$ and plug this expression into (5.6). Note that we will drop the supremum since we have proven above that the supremum is achieved by an isotropically distributed $\hat{\mathbf{X}}$.

$$\begin{aligned}
\chi &= I\left(\|\mathbb{H}\hat{\mathbf{X}}\|; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \|\mathbb{H}\hat{\mathbf{X}}\|\right) + n_{\mathrm{R}}\mathsf{E}\Big[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] \\
&\quad - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) &(5.44)\\
&= I\left(\|\mathbb{H}\hat{\mathbf{X}}\|; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + h\big(\mathbb{H}\hat{\mathbf{X}}\big) - h\big(\|\mathbb{H}\hat{\mathbf{X}}\|^2\big) + \mathsf{E}\Big[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] \\
&\quad - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) &(5.45)\\
&= I\left(\|\mathbb{H}\hat{\mathbf{X}}\|^2; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + I\big(\hat{\mathbf{X}}; \mathbb{H}\hat{\mathbf{X}}\big) - h\big(\|\mathbb{H}\hat{\mathbf{X}}\|^2\big) + \mathsf{E}\Big[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] &(5.46)\\
&= I\left(S_1 + S_2; \frac{\mathbb{H}\hat{\mathbf{X}}}{\sqrt{S_1+S_2}}\right) + I\big(\hat{\mathbf{X}}; \mathbb{H}\hat{\mathbf{X}}\big) - h\big(S_1+S_2\big) + \mathsf{E}[\log(S_1+S_2)]. &(5.47)
\end{aligned}$$

Here, the first equality follows from (5.6); in the subsequent equality we use (5.43); the next step follows from the definition of mutual information and the fact that squaring a non-negative argument of mutual information does not change its value; and in the final equality we apply our definitions (5.18) and (5.19).

The first term in (5.47) can be simplified as follows:

$$\begin{aligned}
&I\left(S_1 + S_2; \frac{\mathbb{H}\hat{\mathbf{X}}}{\sqrt{S_1+S_2}}\right) \\
&= I\left(S_1 + S_2; \frac{d\hat{\mathbf{X}} + \tilde{\mathbf{H}}_1}{\sqrt{S_1+S_2}}, \frac{\tilde{\mathbf{H}}_2}{\sqrt{S_1+S_2}}\right) &(5.48)\\
&= I\left(S_1 + S_2; \frac{d\hat{\mathbf{X}} + \tilde{\mathbf{H}}_1}{\sqrt{S_1+S_2}}, \frac{\tilde{\mathbf{H}}_2}{\sqrt{S_1+S_2}}, \frac{S_1}{S_1+S_2}, \frac{S_2}{S_1+S_2}\right) &(5.49)\\
&= I\left(S_1 + S_2; \frac{d\hat{\mathbf{X}} + \tilde{\mathbf{H}}_1}{\sqrt{S_1}}, \frac{\tilde{\mathbf{H}}_2}{\sqrt{S_2}}, \frac{S_2}{S_1}, \frac{S_1}{S_1+S_2}, \frac{S_2}{S_1+S_2}\right) &(5.50)\\
&= I\left(S_1 + S_2; \frac{d\hat{\mathbf{X}} + \tilde{\mathbf{H}}_1}{\sqrt{S_1}}, \frac{\tilde{\mathbf{H}}_2}{\sqrt{S_2}}, \frac{S_2}{S_1}\right) &(5.51)
\end{aligned}$$

$$= h(S_1+S_2) - h\left(S_1+S_2 \,\middle|\, \frac{d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1}{\sqrt{S_1}}, \frac{\tilde{\mathbf{H}}_2}{\sqrt{S_2}}, \frac{S_2}{S_1}\right) \tag{5.52}$$

$$= h(S_1+S_2) - h\left(S_1+S_2 \,\middle|\, \frac{S_2}{S_1}\right) \tag{5.53}$$

$$= h(S_1+S_2) - h\left(S_1\left(1+\frac{S_2}{S_1}\right) \,\middle|\, \frac{S_2}{S_1}\right) \tag{5.54}$$

$$= h(S_1+S_2) - h\left(S_1 \,\middle|\, \frac{S_2}{S_1}\right) - \mathsf{E}\left[\log\left(1+\frac{S_2}{S_1}\right)\right]. \tag{5.55}$$

Here the first equality follows from splitting the vector into two subvectors, one with n_T and one with $n_\mathrm{R}-n_\mathrm{T}$ components (see (5.15)); in the subsequent equality we compute the magnitudes of $\frac{d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1}{\sqrt{S_1+S_2}}$ and $\frac{\tilde{\mathbf{H}}_2}{\sqrt{S_1+S_2}}$ and add them to the arguments of the mutual information (this does not change mutual information since they are a deterministic function of the given arguments); in the subsequent equality we multiply $\frac{d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1}{\sqrt{S_1+S_2}}$ with $\sqrt{\frac{S_1+S_2}{S_1}}$ and $\frac{\tilde{\mathbf{H}}_2}{\sqrt{S_1+S_2}}$ with $\sqrt{\frac{S_1+S_2}{S_1}}$, and we add a new argument $\frac{S_1+S_2}{S_1}-1=\frac{S_2}{S_1}$; in the subsequent equality we drop the last two arguments of mutual information as they are deterministic functions of the other arguments. In (5.53) we note that $\hat{\mathbf{X}}$ is isotropically distributed and that therefore both $d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1$ and $\tilde{\mathbf{H}}_2$ are isotropic. Hence, the magnitudes S_1 and S_2 are independent of the directions $\frac{d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1}{\sqrt{S_1}}$ and $\frac{\tilde{\mathbf{H}}_2}{\sqrt{S_2}}$. And finally, the last equality follows from the scaling property of differential entropy of a *real* argument.

For the second term in (5.47) we note that it is independent of the last $n_\mathrm{R}-n_\mathrm{T}$ rows of $\mathbb{H}$:

$$I\big(\hat{\mathbf{X}};\mathbb{H}\hat{\mathbf{X}}\big) = I\left(\hat{\mathbf{X}};\begin{pmatrix} d\hat{\mathbf{X}} \\ \mathbf{0}\end{pmatrix}+\tilde{\mathbf{H}}\right) \tag{5.56}$$

$$= I\big(\hat{\mathbf{X}}; d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1\big); \tag{5.57}$$

and for the last term we get

$$\mathsf{E}[\log(S_1+S_2)] = \mathsf{E}[\log S_1] + \mathsf{E}\left[\log\left(1+\frac{S_2}{S_1}\right)\right]. \tag{5.58}$$

Putting this together we yield

$$\chi = I\big(\hat{\mathbf{X}}; d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1\big) + \mathsf{E}[\log S_1] - h\left(S_1 \,\middle|\, \frac{S_2}{S_1}\right) \tag{5.59}$$

$$= I\big(\hat{\mathbf{X}}; d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1\big) + \mathsf{E}[\log S_1] - h(S_1) + I\left(S_1;\frac{S_2}{S_1}\right). \tag{5.60}$$

Now note that the same derivation can be done for the situation $n_\mathrm{R}=n_\mathrm{T}$ in which case $S_2=0$. We get exactly the same result (5.60) apart from the last term $I\left(S_1;\frac{S_2}{S_1}\right)$ which is in this situation equal to zero. However, we know from Theorem 23 the exact value of the fading number for $n_\mathrm{R}=n_\mathrm{T}$, and hence we see that

$$I\big(\hat{\mathbf{X}}; d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1\big) + \mathsf{E}[\log S_1] - h(S_1) = n_\mathrm{T} g_{n_\mathrm{T}}\big(|d|^2\big) - n_\mathrm{T} - \log\Gamma(n_\mathrm{T}). \tag{5.61}$$

Plugging this into (5.60) then yields

$$\chi = n_\mathrm{T} g_{n_\mathrm{T}}\big(|d|^2\big) - n_\mathrm{T} - \log\Gamma(n_\mathrm{T}) + I\left(S_1;\frac{S_2}{S_1}\right) \tag{5.62}$$

which proves the claim.

5.3.2 Proof of Claim 25

The lower bounds follow trivially from the nonnegativity of mutual information. It only remains to derive the upper bounds. We start with the case $n_{\mathrm{T}} > 1$:

$$I\left(S_1; \frac{S_2}{S_1}\right) = h\left(\frac{S_2}{S_1}\right) - h\left(\frac{S_2}{S_1} \,\middle|\, S_1\right) \tag{5.63}$$

$$= h\left(\frac{S_2}{S_1}\right) - h(S_2|S_1) + \mathsf{E}[\log S_1] \tag{5.64}$$

$$= h\left(\frac{S_2}{S_1}\right) - h(S_2) + \mathsf{E}[\log S_1], \tag{5.65}$$

where the second equality follows from the behavior of differential entropy under scaling and where in the last equality we use that $S_1 \perp\!\!\!\perp S_2$.

It is straightforward to show that the differential entropy of a central chi-square random variable of $2(n_{\mathrm{R}} - n_{\mathrm{T}})$ degrees of freedom is

$$h(S_2) = (n_{\mathrm{R}} - n_{\mathrm{T}}) - (n_{\mathrm{R}} - n_{\mathrm{T}} - 1)\psi(n_{\mathrm{R}} - n_{\mathrm{T}}) + \log\Gamma(n_{\mathrm{R}} - n_{\mathrm{T}}) \tag{5.66}$$

with $\psi(\cdot)$ defined in (3.9). From Theorem 5 and Lemma 14 follows that

$$\mathsf{E}[\log S_1] = g_{n_{\mathrm{T}}}\left(|d|^2\right) \leq \log\left(n_{\mathrm{T}} + |d|^2\right). \tag{5.67}$$

The derivation of $h(S_2/S_1)$ is much harder. We content ourselves with an upper bound. Note that the entropy of a non-negative random variable T for a given mean $\mathsf{E}[T] = \mu$ is maximized by an exponential distribution [35, Ex. 12.2.5] yielding

$$h(T) = 1 + \log\mu. \tag{5.68}$$

Using the independence of S_1 and S_2, Theorem 6, and Lemma 13 we get

$$\mathsf{E}\left[\frac{S_2}{S_1}\right] = \mathsf{E}[S_2] \cdot \mathsf{E}\left[\frac{1}{S_1}\right] \tag{5.69}$$

$$= (n_{\mathrm{R}} - n_{\mathrm{T}}) \cdot g'_{n_{\mathrm{T}}-1}\left(|d|^2\right) \tag{5.70}$$

$$\leq (n_{\mathrm{R}} - n_{\mathrm{T}}) \cdot \frac{n_{\mathrm{T}}}{(n_{\mathrm{T}} - 1)(|d|^2 + n_{\mathrm{T}})} \tag{5.71}$$

and therefore

$$h\left(\frac{S_2}{S_1}\right) \leq 1 + \log\frac{n_{\mathrm{T}}(n_{\mathrm{R}} - n_{\mathrm{T}})}{(n_{\mathrm{T}} - 1)(|d|^2 + n_{\mathrm{T}})}. \tag{5.72}$$

The claimed result now follows from (5.65), (5.66), (5.67), and (5.72).

We see that for $n_{\mathrm{T}} = 1$ this bound is infinite. This is due to the implicit maximization in (5.68). We will now derive a tighter bound by maximizing entropy of a non-negative random variable T for a given expectation $\mathsf{E}\left[\sqrt{T}\right] = \nu$ instead of a given mean $\mathsf{E}[T]$. Note that due to Jensen's inequality $\mathsf{E}\left[\sqrt{T}\right] \leq \sqrt{\mathsf{E}[T]}$, *i.e.*, this bound will in general be tighter.[3] Using the approach given in [35] we readily see that $h(T)$ is maximized by a distribution

$$f_T(t) = \frac{2}{\nu^2} e^{-\frac{2\sqrt{t}}{\nu}} \tag{5.73}$$

[3]Unfortunately, we have not succeeded in evaluating it analytically for the cases $n_{\mathrm{T}} > 1$.

which leads to

$$h(T) = \log \nu^2 - \log 2 + 2. \tag{5.74}$$

In the situation $n_\mathrm{T} = 1$ we then continue as follows [32]:

$$\begin{aligned}
\mathsf{E}\left[\sqrt{\frac{S_2}{S_1}}\right] &= \mathsf{E}\left[\sqrt{S_2}\right] \cdot \mathsf{E}\left[\sqrt{\frac{1}{S_1}}\right] &(5.75)\\
&\le \sqrt{\mathsf{E}[S_2]} \cdot \mathsf{E}\left[\sqrt{\frac{1}{S_1}}\right] &(5.76)\\
&= \sqrt{n_\mathrm{R} - 1} \cdot \mathsf{E}\left[\sqrt{\frac{1}{S_1}}\right] &(5.77)\\
&= \sqrt{n_\mathrm{R} - 1}\sqrt{\pi} e^{-\frac{|d|^2}{2}} I_0\left(\frac{|d|^2}{2}\right) &(5.78)
\end{aligned}$$

where $I_0(\cdot)$ denotes the modified Bessel function of the first kind of order zero. Here the last equality can be derived using standard tools of integration [31] or integration tables [32].

Finally, we derive the following bound on the modified Bessel function.

Lemma 27. *The modified Bessel function of the first kind of order zero is upper-bounded by*

$$I_0(\xi) \le \frac{e^\xi}{\sqrt{2\xi + 1}}, \qquad \forall\, \xi \ge 0. \tag{5.79}$$

Proof. Let $f(\xi) \triangleq \frac{e^\xi}{\sqrt{2\xi+1}}$. Computing the first and second derivatives

$$\begin{aligned}
f'(\xi) &= \frac{2\xi e^\xi}{(2\xi + 1)^{3/2}}, &(5.80)\\
f''(\xi) &= \frac{2(2\xi^2 + 1)e^\xi}{(2\xi + 1)^{5/2}}, &(5.81)
\end{aligned}$$

one easily sees that $f(\cdot)$ is strictly positive, strictly increasing, and strictly convex. Moreover, note that $f(0) = 1$, $f'(0) = 0$, $f''(0) = 2$, and

$$\lim_{\xi \uparrow \infty} \frac{f(\xi)}{e^\xi / \sqrt{2\pi\xi}} = \sqrt{\pi}. \tag{5.82}$$

On the other hand $I_0(\cdot)$ is strictly positive, strictly increasing, and strictly convex with $I_0(0) = 1$,

$$\begin{aligned}
\left.\frac{\partial}{\partial \xi} I_0(\xi)\right|_{\xi=0} &= I_1(0) = 0, &(5.83)\\
\left.\frac{\partial^2}{\partial \xi^2} I_0(\xi)\right|_{\xi \downarrow 0} &= I_0(\xi) - \left.\frac{1}{\xi} I_1(\xi)\right|_{\xi \downarrow 0} = \frac{1}{2}, &(5.84)
\end{aligned}$$

and

$$\lim_{\xi \uparrow \infty} \frac{I_0(\xi)}{e^\xi / \sqrt{2\pi\xi}} = 1. \tag{5.85}$$

Now note that two strictly convex functions can cross at most twice. In our case we have one intersection for $\xi = 0$. So there could be at most one more. However, we see from the first and second derivatives that for small $\xi > 0$, $f(\xi) > I_0(\xi)$. Moreover, from the limiting behavior we also see that $f(\xi) > I_0(\xi)$ for $\xi \uparrow \infty$. So no more crossing is possible and $f(\xi) \ge I_0(\xi)$ for $\xi \ge 0$. □

Using this bound we get from (5.78)

$$\mathsf{E}\left[\sqrt{\frac{S_2}{S_1}}\right] \le \sqrt{\frac{\pi(n_\mathrm{R}-1)}{|d|^2+1}} \tag{5.86}$$

and hence from (5.74)

$$h\left(\frac{S_2}{S_1}\right) \le \log \frac{\pi(n_\mathrm{R}-1)}{|d|^2+1} - \log 2 + 2. \tag{5.87}$$

The claimed result now follows from (5.65), (5.66), (5.67), and (5.87).

Chapter 6

A General MIMO Fading Channels with Memory

In this chapter we summarize some results concerning the most general single-user situation: a general MIMO fading channel with memory. We will firstly prove that for such a channel the optimum input distribution can be assumed to be circularly symmetric.

Then we will state an even more general result: we will show that the optimum input for any *stationary* channel model basically is also stationary. This result holds in a very general setting, *i.e.*, it is not restricted to fading channels, but can be applied to any stationary channel with an input alphabet $\mathbb{C}^{n_\mathrm{T}}$ and an output alphabet $\mathbb{C}^{n_\mathrm{R}}$.

Finally, we state an expression for the fading number of general MIMO fading with memory. There we only give an overview of the basic ideas behind a proof without any details. The proof relies on the two results described above as well as Proposition 17 and Lemma 18 given in Chapter 4.

6.1 Capacity-Achieving Input Distributions and Circular Symmetry

The following observation concerns circular symmetry. We say that a random vector $\mathbf{W}$ is *circularly symmetric* if

$$\mathbf{W} \stackrel{\mathscr{L}}{=} \mathbf{W}e^{\mathrm{i}\Theta} \tag{6.1}$$

where $\Theta \sim \mathcal{U}([0, 2\pi])$ is independent of $\mathbf{W}$ and where $\stackrel{\mathscr{L}}{=}$ stands for "equal in law". Note that this is not to be confused with *isotropically distributed*, which means that a vector has equal probability to point in every direction. Circular symmetry only concerns the phase of the components of a vector, not the vector's direction.

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

Definition 28. *A vector random process* $\{\mathbf{W}_k\}$ *is said to be* circularly symmetric *if*

$$\{\mathbf{W}_k\} \stackrel{\mathscr{L}}{=} \{\mathbf{W}_k e^{\mathrm{i}\Theta_k}\}, \tag{6.2}$$

where the process $\{\Theta_k\}$ *is IID* $\sim \mathcal{U}([0, 2\pi])$ *and independent of* $\{\mathbf{W}_k\}$.

Remark 29. *Note some subtleties of this definition: a random process being circularly symmetric does not only mean that for every time k the corresponding random vector* $\mathbf{W}_k$ *is circularly symmetric, but also that from past vectors* $\mathbf{W}_{-\infty}^{k-1}$ *one cannot* *gain any knowledge about the present phase,* i.e., *the phase is IID. On the other hand, however, knowing the phase of one component of* $\mathbf{W}_k$ *in general does yield certain knowledge about the phase of some other components at the* same time k.

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

Proposition 30. *Assume a channel as given in* (2.17)*. Then the capacity-achieving input process can be assumed to be circularly symmetric,* i.e., *the input vectors* $\{\mathbf{X}_k\}$ *can be replaced by* $\{\mathbf{X}_k e^{\mathrm{i}\Theta_k}\}$*, where the random process* $\{\Theta_k\}$ *is IID* $\sim \mathcal{U}([0,2\pi])$ *and independent of every other random quantity.*

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

$$
\begin{aligned}
&\frac{1}{n} I\big(\mathbf{X}_1^n;\mathbf{Y}_1^n\big) \\
&\quad = \frac{1}{n} I\big(\mathbf{X}_1^n;\mathbf{Y}_1^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) && (6.3)\\
&\quad = \frac{1}{n} I\big(\{\mathbf{X}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n;\{\mathbf{Y}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) && (6.4)\\
&\quad = \frac{1}{n} I\big(\{\mathbf{X}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n;\{\mathbb{H}_\ell \mathbf{X}_\ell e^{\mathrm{i}\Theta_\ell} + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) && (6.5)\\
&\quad = \frac{1}{n} I\big(\tilde{\mathbf{X}}_1^n;\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) && (6.6)\\
&\quad = \frac{1}{n} h\big(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) - \frac{1}{n} h\big(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \tilde{\mathbf{X}}_1^n, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) && (6.7)\\
&\quad = \frac{1}{n} h\big(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\big) - \frac{1}{n} h\big(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \tilde{\mathbf{X}}_1^n\big) && (6.8)\\
&\quad \le \frac{1}{n} h\big(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n\big) - \frac{1}{n} h\big(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \,\big|\, \tilde{\mathbf{X}}_1^n\big) && (6.9)\\
&\quad = \frac{1}{n} I\big(\tilde{\mathbf{X}}_1^n;\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n\big). && (6.10)
\end{aligned}
$$

Here the first equality follows because $\{\Theta_k\}$ is independent of every other random quantity; the third equality follows because $\{\mathbf{Z}_k\}$ is circularly symmetric; in the subsequent equality we define the new input $\tilde{\mathbf{X}}_\ell \triangleq \mathbf{X}_\ell e^{\mathrm{i}\Theta_\ell}$; and the inequality follows since conditioning reduces entropy.

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

Remark 31. *Note that the proof of Proposition 30 only relies on the fact that the additive noise is assumed to be circularly symmetric. Hence, for this proposition to hold the additive noise process does not need to be Gaussian distributed and may even have memory as long as it is circularly symmetric.*

6.2 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.* From an intuitive point of view it is clear that a stationary channel model should have a capacity-achieving input distribution that is also stationary. Unfortunately, we are not aware of a rigorous proof of this claim.

In [4, Lem. 5], [3, Lem. B.1] it is proven that—apart from edge effects—the optimum input distribution can be assumed to have equal marginals. Here we will extend this statement and prove that capacity can be approached up to an $\epsilon > 0$ by a distribution that looks stationary apart from edge effects.

Theorem 32. *Assume some general channel model with input* $\mathbf{x}_k \in \mathbb{C}^{n_\mathrm{T}}$ *and output* $\mathbf{Y}_k \in \mathbb{C}^{n_\mathrm{R}}$. *Let the channel model be stationary,* i.e., *for every choice of* $n \in \mathbb{N}$ *and distribution* $Q \in \mathcal{P}(\mathbb{C}^{n_\mathrm{T}\times n})$ *on* $\mathbf{X}_1^n$ *the mutual information* $I\bigl(\mathbf{X}_1^n;\mathbf{Y}_1^n\bigr)$ *does not change when the input block is shifted over time. Assume an average-power constraint* (2.24)*. Then the channel capacity is given by*

$$\mathsf{C}(\mathcal{E}) = \lim_{n\uparrow\infty} \frac{1}{n} \sup I\bigl(\mathbf{X}_1^n;\mathbf{Y}_1^n\bigr), \tag{6.11}$$

where the supremum is over all distributions on $\mathbf{X}_1^n$ *that satisfy the average-power constraint* (2.24)*. Finally assume that the channel model is such that a zero input has no influence on the mutual information in the following sense: for every* $n \in \mathbb{N}$, *for every* $k_1, k_2 \in \mathbb{N}$ *such that* $1 \le k_1 \le k_2 \le n$, *and for any distribution* $\tilde{Q} \in \mathcal{P}(\mathbb{C}^{n_\mathrm{T}\times(k_2-k_1+1)})$, *we have*

$$I\bigl(\mathbf{0}_1^{k_1-1}, \mathbf{X}_{k_1}^{k_2}, \mathbf{0}_{k_2+1}^{n}; \mathbf{Y}_1^n\bigr) = I\bigl(\mathbf{0}_1^{k_1-1+\kappa}, \mathbf{X}_{k_1+\kappa}^{k_2+\kappa}, \mathbf{0}_{k_2+1+\kappa}^{n}; \mathbf{Y}_1^n\bigr) \tag{6.12}$$

for every $\kappa \in \{-k_1+1, \ldots, n-k_2\}$, *where both* $\mathbf{X}_{k_1}^{k_2}$ *on the left hand side and* $\mathbf{X}_{k_1+\kappa}^{k_2+\kappa}$ *on the right hand side are distributed according to* $\tilde{Q}$.

Now fix some non-negative integer κ *and some power* $\mathcal{E}$ *with corresponding* $\mathrm{SNR} \triangleq \mathcal{E}/\sigma^2$. *Then for every fixed* $\epsilon > 0$ *there corresponds some positive integer* $\eta = \eta(\mathcal{E},\epsilon)$ *and some distribution* $Q_{\mathcal{E},\epsilon}^{\kappa+1} \in \mathcal{P}\bigl(\mathbb{C}^{n_\mathrm{T}\times(\kappa+1)}\bigr)$ *such that for a blocklength* n *sufficiently large there exists some input* $\mathbf{X}_1^n$ *satisfying the following:*

1. *The input* $\mathbf{X}_1^n$ *nearly achieves capacity in the sense that*

$$\frac{1}{n} I\bigl(\mathbf{X}_1^n;\mathbf{Y}_1^n\bigr) \ge \mathsf{C}(\mathcal{E}) - \epsilon. \tag{6.13}$$

2. *For every integer* μ *with* $0 \le \mu \le \kappa$, *every length-*$(\mu+1)$ *block of adjacent vectors*

$$(\mathbf{X}_\ell, \ldots, \mathbf{X}_{\ell+\mu}) \tag{6.14}$$

taken from

$$\mathbf{X}_\eta, \mathbf{X}_{\eta+1}, \ldots, \mathbf{X}_{n-2\eta+2} \tag{6.15}$$

has the same joint distribution $Q_{\mathcal{E},\epsilon}^{\mu+1}$, *where this distribution* $Q_{\mathcal{E},\epsilon}^{\mu+1}$ *is given as the corresponding marginal distribution of* $Q_{\mathcal{E},\epsilon}^{\kappa+1}$.

3. *In particular, all vectors in* (6.15) *have the same marginal* $Q_{\mathcal{E},\epsilon}^{1}$.

4. *The marginal distribution* $Q_{\mathcal{E},\epsilon}^{1}$ *gives rise to a second moment* $\mathcal{E}$:

$$\mathsf{E}\bigl[\|\mathbf{X}_\ell\|^2\bigr] = \mathcal{E}, \qquad \ell = \eta, \ldots, n-2\eta+2. \tag{6.16}$$

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

$$\mathsf{E}\bigl[\|\mathbf{X}_\ell\|^2\bigr] \le \mathcal{E}, \quad \ell \in \{1, \ldots, \eta-1\} \cup \{n-2\eta+3, \ldots, n\}. \tag{6.17}$$

Proof. The proof follows the same lines as the proofs of [4, Lem. 5] and [3, Lem. B.1]. It is based on a shift-and-mix argument using the fact that when using deterministic zeros at the input, the corresponding output yields zero information.

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

$$\mathsf{C}(\mathcal{E}) = \lim_{n\uparrow\infty} \frac{1}{n} \sup I(\mathbf{X}_1, \ldots, \mathbf{X}_n; \mathbf{Y}_1, \ldots, \mathbf{Y}_n) \tag{6.18}$$

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

$$\frac{1}{\eta} \sum_{\ell=1}^{\eta} \mathsf{E}\left[\|\mathbf{X}_\ell\|^2\right] = \mathcal{E} \tag{6.19}$$

and

$$\frac{1}{\eta} I(\mathbf{X}_1, \ldots, \mathbf{X}_\eta; \mathbf{Y}_1, \ldots, \mathbf{Y}_\eta) > \mathsf{C}(\mathcal{E}) - \frac{\epsilon}{2}. \tag{6.20}$$

For some non-negative integer $\mu \leq \kappa$, let $\mathbb{W}$ be a $n_\mathrm{T} \times \left(\eta \cdot \left\lceil \frac{\mu}{\eta} + 1 \right\rceil\right)$ random matrix whose distribution consists of $\left\lceil \frac{\mu}{\eta} + 1 \right\rceil$ independent $n_\mathrm{T} \times \eta$ blocks that are distributed according to Q^*:

$$Q_\mathbb{W}\left(\mathbf{w}_1, \ldots, \mathbf{w}_{\eta\left\lceil \frac{\mu}{\eta}+1 \right\rceil}\right) = \prod_{i=0}^{\left\lceil \frac{\mu}{\eta}+1 \right\rceil - 1} Q^*\left(\mathbf{w}_{i\cdot\eta+1}, \ldots, \mathbf{w}_{i\cdot\eta+\eta}\right), \qquad \mathbf{w}_\ell \in \mathbb{C}^{n_\mathrm{T}}. \tag{6.21}$$

Let $Q_{\mathcal{E},\epsilon}^{\mu+1}$ be the probability law on $\mathbb{C}^{n_\mathrm{T}\times(\mu+1)}$ that is a mixture of η different block-marginals of $Q_\mathbb{W}$, *i.e.*, for every Borel set $\mathcal{B} \subseteq \mathbb{C}^{n_\mathrm{T}\times(\mu+1)}$

$$Q_{\mathcal{E},\epsilon}^{\mu+1}(\mathcal{B}) \triangleq \frac{1}{\eta} \sum_{\ell=1}^{\eta} Q_\mathbb{W}\big[(\mathbf{W}_\ell, \ldots, \mathbf{W}_{\ell+\mu}) \in \mathcal{B}\big]. \tag{6.22}$$

In particular $Q_{\mathcal{E},\epsilon}^{\mu+1}$ can be computed as marginal from $Q_{\mathcal{E},\epsilon}^{\kappa+1}$ where for every Borel set $\mathcal{B} \subseteq \mathbb{C}^{n_\mathrm{T}\times(\kappa+1)}$

$$Q_{\mathcal{E},\epsilon}^{\kappa+1}(\mathcal{B}) \triangleq \frac{1}{\eta} \sum_{\ell=1}^{\eta} Q_\mathbb{W}\big[(\mathbf{W}_\ell, \ldots, \mathbf{W}_{\ell+\kappa}) \in \mathcal{B}\big]. \tag{6.23}$$

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

$$\begin{aligned} Q_{\mathcal{E},\epsilon}^{\mu+1}(\mathcal{B}) \triangleq{}& \frac{1}{\eta} \sum_{\ell=1}^{\eta-\mu} Q^*\big[(\mathbf{X}_\ell, \mathbf{X}_{\ell+1}, \ldots, \mathbf{X}_{\ell+\mu}) \in \mathcal{B}\big] \\ &+ \frac{1}{\eta} \sum_{\ell=\eta-\mu+1}^{\eta} Q^*\big[(\mathbf{X}_\ell, \ldots, \mathbf{X}_\eta) \in \mathcal{B}_{1,\ldots,\eta-\ell+1}\big] \\ &\qquad \cdot Q^*\big[(\mathbf{X}_1, \ldots, \mathbf{X}_{\ell-\eta+\mu}) \in \mathcal{B}_{\eta-\ell+2,\ldots,\eta}\big], \end{aligned} \tag{6.24}$$

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

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

$$\nu \triangleq \left\lfloor \frac{n-\eta+1}{\eta} \right\rfloor \tag{6.25}$$

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

$$\tilde{\mathbb{X}} \triangleq (\underbrace{\mathbf{0},\ldots,\mathbf{0}}_{\eta-1}, \underbrace{\mathbf{\Xi}_1^{(1)},\ldots,\mathbf{\Xi}_\eta^{(1)}}_{\eta}, \ldots, \ldots, \underbrace{\mathbf{\Xi}_1^{(\nu)},\ldots,\mathbf{\Xi}_\eta^{(\nu)}}_{\eta}, \mathbf{0}, \mathbf{0}, \ldots \tag{6.26}$$

so that

$$\tilde{\mathbf{X}}_\ell = \begin{cases} \mathbf{0} & \text{if } 1 \le \ell \le \eta - 1, \\ \mathbf{\Xi}_{(\ell \bmod \eta)+1}^{\lfloor \ell/\eta \rfloor} & \text{if } \eta \le \ell \le (\nu+1)\eta - 1, \\ \mathbf{0} & \text{if } \ell \ge (\nu+1)\eta, \end{cases} \tag{6.27}$$

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

$$\left\{ \left(\mathbf{\Xi}_1^{(j)}, \ldots, \mathbf{\Xi}_\eta^{(j)} \right) \right\}_{j=1}^{\nu} \text{ are IID} \sim Q^*. \tag{6.28}$$

Notice that since the lead-in and trailing zeros have no effect on our channel, the not-normalized 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 $0 \le t \le \eta - 1$). Consequently, if we define $\mathbf{X}_1, \ldots, \mathbf{X}_n$ by the mixture of the time shift of $\tilde{\mathbb{X}}$, *i.e.*,

$$\mathbf{X}_\ell = \tilde{\mathbf{X}}_{\ell+T}, \qquad 1 \le \ell \le n, \tag{6.29}$$

where

$$T \sim \mathcal{U}(\{0, \ldots, \eta - 1\}) \tag{6.30}$$

is independent of $\tilde{\mathbb{X}}$, then by the concavity of mutual information in the input distribution we obtain that the not-normalized 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

$$\frac{1}{n} I\left(\mathbf{X}_1^n; \mathbf{Y}_1^n\right) \ge \frac{\eta\nu}{n} \left(\mathsf{C}(\mathcal{E}) - \frac{\epsilon}{2} \right) \tag{6.31}$$

$$= \frac{\eta \left\lfloor \frac{n-\eta+1}{\eta} \right\rfloor}{n} \left(\mathsf{C}(\mathcal{E}) - \frac{\epsilon}{2} \right), \tag{6.32}$$

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

Except at the edges, the above mixture guarantees equal block-marginals as defined in (6.22) for every μ, $0 \le \mu \le \kappa$.

Note that η does not depend on the choice of κ (or on μ), however, the size of the edges where the theorem does not hold depends on κ.

Note further that by (6.19) we have for $\mu = 0$

$$\int_{\mathbb{C}^{n_{\mathrm{T}}}} \|\mathbf{x}\|^2 \,\mathrm{d}Q_{\mathcal{E},\epsilon}^1(\mathbf{x}) = \mathcal{E}. \tag{6.33}$$

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

Remark 33. *Neglecting the edge-effects for the moment, Theorem 32 basically says that, for every $\mu \leq \kappa$, every block of $\mu+1$* adjacent *vectors has the same distribution independent of the time shift. From this immediately follows that the distribution of every subset of (not necessarily adjacent) vectors of a $\mu+1$ block does not change when the vectors are shifted in time (simply marginalize those vectors out that are not member of the subset). Hence, Theorem 32* almost *proves that the capacity-achieving input distribution is stationary: the only problems are the edge effects and the fixed (but freely selectable) value of κ.*[1]

6.3 Fading Number of MIMO Fading Channels with Memory

We are now ready to state the expression of the fading number of MIMO fading with memory.

Theorem 34. *Consider a MIMO fading channel with memory* (2.17) *where the stationary and ergodic fading process $\{\mathbb{H}_k\}$ takes value in $\mathbb{C}^{n_R \times n_T}$ and satisfies $h(\{\mathbb{H}_k\}) > -\infty$ and $\mathsf{E}\left[\|\mathbb{H}_k\|_F^2\right] < \infty$. Then, irrespective of whether a peak-power constraint* (2.23) *or an average-power constraint* (2.24) *is imposed on the input, the fading number $\chi\big(\{\mathbb{H}_k\}\big)$ is given by*

$$\boxed{\begin{aligned}\chi\big(\{\mathbb{H}_k\}\big) = \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}} \\ \text{stationary} \\ \text{circ. sym.}}} \Bigg\{ & h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\,\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\infty}^{-1}\right) \\ & + n_R\mathsf{E}\left[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\right] \\ & - \log 2 - h\Big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\Big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\infty}^{0}\Big)\Bigg\}.\end{aligned}} \tag{6.34}$$

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

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

$$\mathbf{X}_k = R_k \cdot \hat{\mathbf{X}}_k, \tag{6.35}$$

where $\{\hat{\mathbf{X}}_k\} \in \mathbb{C}^{n_T}$ is a stationary and circularly symmetric unit-vector process with the distribution that achieves the maximum in (6.34)*, and $\{R_k\} \in \mathbb{R}_0^+$ is a scalar non-negative IID random process such that*

$$\log R_k^2 \sim \mathcal{U}([\log\log\mathcal{E}, \log\mathcal{E}]). \tag{6.36}$$

Note that this input satisfies the peak-power constraint (2.23) *(and therefore also the average-power constraint* (2.24)*).*

[1] As a matter of fact, one can choose κ arbitrarily large, however, note that the size of the edges where the lemma does not hold depends on κ.

Proof. The proof is long and obscured by many technical details. We will only give an outline here emphasizing the important key steps.

The proof consists of two parts: firstly we derive an upper bound on the fading number assuming an average-power constraint (2.24) on the channel input. The key ingredients for this part are the following four concepts:

- the optimal input distribution escapes to infinity (see Section 4.1);
- the optimal input distribution is circularly symmetric (see Section 6.1);
- the optimal input distribution is stationary (see Section 6.2);
- the capacity can be upper-bounded using a dual expression (see Section 4.2).

Secondly we derive a lower bound on the fading number by assuming one particular input distribution on the channel that satisfies the peak-power constraint (2.23). We then show that the fading that is achieved by this choice is identical to the upper bound derived before. Since a peak-power constraint is more restrictive than the corresponding average-power constraint, the theorem follows.

Outline of Upper Bound. To derive the upper bound we consider the average-power constraint (2.24). Similarly to the proof of the SIMO fading number with memory [4, Sec. VII], [3, Sec. B.5.9] we use the chain rule to write

$$\frac{1}{n} I\bigl(\mathbf{X}_1^n; \mathbf{Y}_1^n\bigr) = \frac{1}{n} \sum_{k=1}^{n} I\bigl(\mathbf{X}_1^n; \mathbf{Y}_k \bigm| \mathbf{Y}_1^{k-1}\bigr), \tag{6.37}$$

and would like to separate each term on the right-hand side (RHS) into terms that are memoryless and terms that take care of the memory. It has been shown in [8, App. IV], [4, Eq. (77)], [3, Eqs. (6.267)–(6.275)] that

$$\begin{aligned} I\bigl(\mathbf{X}_1^n; \mathbf{Y}_k \bigm| \mathbf{Y}_1^{k-1}\bigr) &\le I\bigl(\mathbf{X}_k; \mathbf{Y}_k\bigr) - I\bigl(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}\bigr) + I\bigl(\mathbb{H}_k; \mathbb{H}_1^{k-1}\bigr) && (6.38)\\ &\le I\bigl(\mathbf{X}_k; \mathbf{Y}_k\bigr) + I\bigl(\mathbb{H}_k; \mathbb{H}_1^{k-1}\bigr) && (6.39) \end{aligned}$$

which would nicely do the trick. Unfortunately, (6.39) is not tight for two reasons. Firstly, note that in the situation of only one transmit antenna it is possible to get a good estimate of the fading realizations by simply dividing the received vector $\mathbf{Y}_k$ by the decoded value of X_k:

$$\frac{\mathbf{Y}_k}{X_k} = \frac{\mathbf{H}_k X_k + \mathbf{Z}_k}{X_k} = \mathbf{H}_k + \frac{\mathbf{Z}_k}{X_k} \to \mathbf{H}_k, \qquad \text{for } |X_k| \uparrow \infty \tag{6.40}$$

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

$$\frac{\mathbf{Y}_k}{\|\mathbf{X}_k\|} = \frac{\mathbb{H}_k \mathbf{X}_k + \mathbf{Z}_k}{\|\mathbf{X}_k\|} = \mathbb{H}_k \hat{\mathbf{X}}_k + \frac{\mathbf{Z}_k}{\|\mathbf{X}_k\|} \to \mathbb{H}_k \hat{\mathbf{X}}_k, \qquad \text{for } \|\mathbf{X}_k\| \uparrow \infty. \tag{6.41}$$

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

The second reason why (6.39) is not tight is the term $I\bigl(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}\bigr)$ that (similar to the SIMO situation) we must not discard because it contains information about the past fading values even if we do not know the corresponding inputs. To see this note that from $\mathbf{Y}_k$ we can easily get

$$\frac{\mathbf{Y}_k}{\|\mathbf{Y}_k\|} = \frac{\mathbb{H}_k\mathbf{X}_k + \mathbf{Z}_k}{\|\mathbb{H}_k\mathbf{X}_k + \mathbf{Z}_k\|} = \frac{\mathbb{H}_k\hat{\mathbf{X}}_k + \frac{\mathbf{Z}_k}{\|\mathbf{X}_k\|}}{\left\|\mathbb{H}_k\hat{\mathbf{X}}_k + \frac{\mathbf{Z}_k}{\|\mathbf{X}_k\|}\right\|} \to \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}, \qquad \text{for } \|\mathbf{X}_k\| \uparrow \infty, \tag{6.42}$$

which is an estimate for the "direction" of the fading. However, note that similarly to (6.41) and unlike to the SIMO case we cannot gain full knowledge about $\frac{\mathbb{H}_k}{\|\mathbb{H}_k\|}$ because the fading is a matrix-valued process.

So we get the following bound instead:

$$I\big(\mathbf{X}_1^n; \mathbf{Y}_k \,\big|\, \mathbf{Y}_1^{k-1}\big) \le I\big(\mathbf{X}_k; \mathbf{Y}_k\big) - I\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{k-1}\right) + I\big(\mathbb{H}_k\hat{\mathbf{X}}_k; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1} \,\big|\, \hat{\mathbf{X}}_1^n\big). \tag{6.43}$$

Note that we have jumped over many details here, in particular, we need to rely on the observation of Proposition 17 that the capacity-achieving input distribution escapes to infinity ($\|\mathbf{X}_k\| \uparrow \infty$) in order to be able to discard the noise.

The first term in (6.43) corresponds to memoryless MIMO fading. Hence, we might use the knowledge of the memoryless MIMO fading number (4.8) or we could use the bounding techniques known from [8, Sec. IV.A], [3, Sec. 2.4] to get an upper bound on this term. Unfortunately, both approaches fail, the former because the memoryless MIMO fading number contains a maximization that will loosen the bound when introduced at this stage. The latter approach turns out to lead to an even less tight bound.

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

$$I\big(\mathbf{X}_k; \mathbf{Y}_k\big) = I\big(\mathbf{X}_k; \|\mathbf{Y}_k\|\big) + I\left(\mathbf{X}_k; \frac{\mathbf{Y}_k}{\|\mathbf{Y}_k\|} \,\middle|\, \|\mathbf{Y}_k\|\right) \tag{6.44}$$

and show that

$$I\left(\mathbf{X}_k; \frac{\mathbf{Y}_k}{\|\mathbf{Y}_k\|} \,\middle|\, \|\mathbf{Y}_k\|\right) \le I\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|, \hat{\mathbf{X}}_k; \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\right) \tag{6.45}$$

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

The first term in (6.44) almost looks like the mutual information between input and output of a memoryless MISO fading channel. We fix the problem that the output is non-negative by multiplying $\|\mathbf{Y}_k\|$ by an independent circularly symmetric phase Θ_k. Because we assume that Θ_k is independent of all other random quantities, particularly of $\mathbf{X}_k$, this does not change the mutual information.

The bound (6.43) then looks as follows:

$$\begin{aligned} &I\big(\mathbf{X}_1^n; \mathbf{Y}_k \,\big|\, \mathbf{Y}_1^{k-1}\big) \\ &\quad\le I\big(\mathbf{X}_k; \|\mathbf{Y}_k\| e^{\mathrm{i}\Theta_k}\big) + I\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|, \hat{\mathbf{X}}_k; \frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\right) \\ &\qquad - I\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}; \left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{k-1}\right) + I\big(\mathbb{H}_k\hat{\mathbf{X}}_k; \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1} \,\big|\, \hat{\mathbf{X}}_1^n\big). \end{aligned} \tag{6.46}$$

This bound still depends on the unknown capacity-achieving input distribution. In order to eliminate this dependence we need to maximize it over all joint distributions on $\mathbf{X}_1, \ldots, \mathbf{X}_n$ that satisfy the average-power constraint. Unfortunately, when we only consider one fixed k, this maximization will loosen our bound. The reason lies in the third term in (6.46) which can be loosely upper-bounded by zero. This loose upper bound can be achieved by the (obviously very bad) choice $\mathbf{X}_1 = \ldots = \mathbf{X}_{k-1} = \mathbf{0}$.

So it seems that we cannot consider each term of the sum in (6.37) separately. Fortunately, this is possible once we take Theorem 32 and Proposition 19 into account. It allows us to restrict ourselves to stationary and circularly symmetric input distributions which excludes the mentioned bad choice and will yield the following bound:

$$\begin{aligned}\frac{1}{n} I\big(\mathbf{X}_1^n; \mathbf{Y}_1^n\big) \leq \sup \Bigg\{ & I\big(\mathbf{X}_0; \|\mathbf{Y}_0\| e^{\mathrm{i}\Theta_0}\big) + I\left(\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|, \hat{\mathbf{X}}_0; \frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \right) \\ & - I\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|}; \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} \right\}_{\ell=-\infty}^{-1} \right) \\ & + I\big(\mathbb{H}_0 \hat{\mathbf{X}}_0; \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1} \,\big|\, \hat{\mathbf{X}}_{-\infty}^0\big) \Bigg\}, \end{aligned} \tag{6.47}$$

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

Note that Theorem 32 has also allowed us to get rid of the dependence on k, *i.e.*, we can let n tend to infinity. Then the way is free to loosen the power constraint $\mathcal{E} \uparrow \infty$ and to use the definition of the fading number (2.26).

We might now be tempted to use our knowledge about memoryless MISO fading. But again this approach fails due to the maximization in the expression of the memoryless MISO fading number (1.8). Instead we rely on Lemma 18 to get an upper bound on the first term in (6.47). This bound will look very similar to the memoryless MISO fading number, however, does not involve a local beam-forming maximization. To make the expressions easier to read, we will use here the notation $\tilde{\chi}$ to refer to this part of the bound.

Hence, we get the following:

$$\begin{aligned}\chi(\{\mathbb{H}_k\}) \lessapprox \sup \Bigg\{ & \tilde{\chi}_{\mathrm{MISO,IID}}\big(\|\mathbb{H}_0 \mathbf{X}_0 + \mathbf{Z}_0\| e^{\mathrm{i}\Theta_0}\big) + I\left(\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|, \hat{\mathbf{X}}_0; \frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|} \right) \\ & - I\left(\frac{\mathbb{H}_0 \hat{\mathbf{X}}_0}{\|\mathbb{H}_0 \hat{\mathbf{X}}_0\|}; \left\{ \frac{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\|} \right\}_{\ell=-\infty}^{-1} \right) \\ & + I\big(\mathbb{H}_0 \hat{\mathbf{X}}_0; \{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1} \,\big|\, \hat{\mathbf{X}}_{-\infty}^0\big) \Bigg\}. \end{aligned} \tag{6.48}$$

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

Note that in the whole derivation we rely on the fact that the input distribution does not take on any value smaller than an arbitrary $\mathcal{E}_0$. However, Proposition 17 only guarantees this in the limit when the power tends to infinity. In order to solve that problem we need to introduce the event $\{\|\mathbf{X}\|^2 \geq \mathcal{E}_0\}$ and condition everything on this event.

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

$$\mathbf{X}_k = R_k \cdot \hat{\mathbf{X}}_k. \tag{6.49}$$

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

$$\log R_k^2 \sim \mathcal{U}\bigl([\log x_{\min}^2, \log \mathcal{E}]\bigr), \tag{6.50}$$

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

$$x_{\min}^2 = \log \mathcal{E}. \tag{6.51}$$

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

Note that this choice of $\{R_k\}$ satisfies the peak-power constraint (2.23) and therefore also the average-power constraint (2.24).

We then again start with the chain rule and write

$$\frac{1}{n} I\bigl(\mathbf{X}_1^n; \mathbf{Y}_1^n\bigr) = \frac{1}{n} \sum_{k=1}^{n} I\bigl(\mathbf{X}_k; \mathbf{Y}_1^n \bigm| \mathbf{X}_1^{k-1}\bigr) \tag{6.52}$$

where we would like to treat each term separately. Note that for the same reason why we were not allowed to discard the term $I\bigl(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}\bigr)$ in the derivation of the upper bound, we are not allowed to discard the future outputs $\mathbf{Y}_{k+1}^n$ in the RHS of (6.52).

After some algebraic changes we get the following lower bound:

$$I\bigl(\mathbf{X}_k; \mathbf{Y}_1^n \bigm| \mathbf{X}_1^{k-1}\bigr) \gtrapprox I\bigl(\hat{\mathbf{X}}_k; \hat{\mathbf{Y}}_{k+1}^n \bigm| \mathbf{Y}_1^{k-1}, \hat{\mathbf{X}}_1^{k-1}\bigr) + I\bigl(\mathbf{X}_k; \mathbf{Y}_k \bigm| \hat{\mathbf{Y}}_{k+1}^n, \mathbf{Y}_1^{k-1}, \hat{\mathbf{X}}_1^{k-1}\bigr). \tag{6.53}$$

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

$$\mathbf{S}_k \triangleq \bigl(\hat{\mathbf{Y}}_{k+1}^n, \mathbf{Y}_1^{k-1}, \hat{\mathbf{X}}_1^{k-1}\bigr). \tag{6.54}$$

Contrary to the derivation of the upper bound that has been based on the memoryless MISO case, we will base the derivation of the lower bound on memoryless SIMO, *i.e.*, we split the second term in (6.53) into the following two parts:

$$I\bigl(\mathbf{X}_k; \mathbf{Y}_k \bigm| \mathbf{S}_k\bigr) = I\bigl(\hat{\mathbf{X}}_k; \mathbf{Y}_k \bigm| \mathbf{S}_k\bigr) + I\bigl(R_k; \mathbf{Y}_k \bigm| \hat{\mathbf{X}}_k, \mathbf{S}_k\bigr). \tag{6.55}$$

Now we have the problem that the second mutual information term in (6.55) does not correspond exactly to the SIMO situation since the input of the channel is real instead of complex. This is fixed by various arithmetic changes which at the end yield the following bound:

$$\begin{aligned} I\bigl(\mathbf{X}_k; \mathbf{Y}_k \bigm| \mathbf{S}_k\bigr) &\approx I\bigl(R_k e^{\mathrm{i}\Theta_k}; \mathbf{Y}_k e^{\mathrm{i}\Theta_k} \bigm| \hat{\mathbf{X}}_k, \mathbf{S}_k\bigr) \\ &\quad + h_\lambda\bigl(\hat{\mathbf{Y}}_k \bigm| \mathbf{S}_k\bigr) - h_\lambda\bigl(\hat{\mathbf{Y}}_k e^{\mathrm{i}\Theta_k} \bigm| \hat{\mathbf{X}}_k, \mathbf{S}_k\bigr) &(6.56) \\ &= I\bigl(\tilde{X}_k; \mathbb{H}_k \hat{\mathbf{X}}_k \tilde{X}_k + \mathbf{Z}_k \bigm| \hat{\mathbf{X}}_k, \mathbf{S}_k\bigr) \\ &\quad + h_\lambda\bigl(\hat{\mathbf{Y}}_k \bigm| \mathbf{S}_k\bigr) - h_\lambda\bigl(\hat{\mathbf{Y}}_k e^{\mathrm{i}\Theta_k} \bigm| \hat{\mathbf{X}}_k, \mathbf{S}_k\bigr). &(6.57) \end{aligned}$$

Note that our choice of R_k guarantees that $\tilde{X}_k \triangleq R_k e^{\mathrm{i}\Theta_k}$ achieves the fading number of memoryless SIMO fading with side-information [8, Prop. 4.30], [3, Prop. 6.30]. Hence, we get from (6.53) and (6.57)

$$\begin{aligned}\chi(\{\mathbb{H}_k\}) \gtrapprox\; & \chi_{\text{SIMO,IID}}\big(\mathbb{H}_k\hat{\mathbf{X}}_k \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) + h_\lambda\big(\hat{\mathbf{Y}}_k \,\big|\, \mathbf{S}_k\big) - h_\lambda\big(\hat{\mathbf{Y}}_k e^{\mathrm{i}\Theta_k} \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) \\ & + I\big(\hat{\mathbf{X}}_k; \hat{\mathbf{Y}}_{k+1}^n \,\big|\, \mathbf{Y}_1^{k-1}, \hat{\mathbf{X}}_1^{k-1}\big) & (6.58) \\ =\; & h_\lambda\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|} e^{\mathrm{i}\Theta_k} \,\middle|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\right) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}_k\hat{\mathbf{X}}_k\|^2\Big] - \log 2 \\ & - h\big(\mathbb{H}_k\hat{\mathbf{X}}_k \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) + h_\lambda\big(\hat{\mathbf{Y}}_k \,\big|\, \mathbf{S}_k\big) - h_\lambda\big(\hat{\mathbf{Y}}_k e^{\mathrm{i}\Theta_k} \,\big|\, \hat{\mathbf{X}}_k, \mathbf{S}_k\big) \\ & + I\big(\hat{\mathbf{X}}_k; \hat{\mathbf{Y}}_{k+1}^n \,\big|\, \mathbf{Y}_1^{k-1}, \hat{\mathbf{X}}_1^{k-1}\big), & (6.59)\end{aligned}$$

where we have used the expression of the fading number of memoryless SIMO fading with side-information (1.7).

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

The result now follows by showing that (6.59) is equivalent to the upper bound. This will follow from some arithmetic changes, from stationarity, and from the fact that we choose the distribution of $\{\hat{\mathbf{X}}_k\}$ to achieve the supremum given in the upper bound. □

6.4 Alternative Expressions and an Upper Bound

In the following we will state several equivalent expressions for the fading number given in (6.34). Depending on the context one particular form might be more convenient.

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

$$\chi_{\hat{\mathbf{X}}}(\mathbb{H}) \triangleq h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big). \qquad (6.60)$$

This corresponds to the situation where we additionally constrain the transmitter to use a fixed (possibly suboptimal) distribution on $\hat{\mathbf{X}}$, *i.e.*, the memoryless MIMO fading number is then given as (see (4.8))

$$\chi(\mathbb{H}) = \sup_{\substack{Q_{\hat{\mathbf{X}}} \\ \text{circ. sym.}}} \chi_{\hat{\mathbf{X}}}(\mathbb{H}). \qquad (6.61)$$

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

$$\chi\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) = \mathsf{E}_{\hat{\mathbf{X}}}\Big[\chi\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}}\big)\Big]. \qquad (6.62)$$

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

From (6.60) we next define the following natural extension: the *constrained memoryless MIMO fading number with partial side-information* $\mathbf{S}$ given a fixed distribution $Q_{\hat{\mathbf{X}}}$ on $\hat{\mathbf{X}}$ is defined as follows:

$$\chi_{\hat{\mathbf{X}}}\big(\mathbb{H}\,\big|\,\mathbf{S}\big) \triangleq h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\mathbf{S}\right) + n_{\mathrm{R}}\,\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}},\mathbf{S}\big). \tag{6.63}$$

Using these definitions we get the following alternative expressions.

Corollary 35. *The MIMO fading number with memory can be written in the following five equivalent forms:*

$$\begin{aligned}
\chi\big(\{\mathbb{H}_k\}\big) &= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\,\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\infty}^{-1}\right) + n_{\mathrm{R}}\,\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\Big] \\
&\qquad - \log 2 - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0\,\big|\,\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1},\hat{\mathbf{X}}_{-\infty}^{0}\big)\Bigg\} &(6.64)\\
&= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ \chi_{\hat{\mathbf{X}}_0}(\mathbb{H}_0) - I\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\infty}\right) \\
&\qquad + I\big(\mathbb{H}_0\hat{\mathbf{X}}_0;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}\,\big|\,\hat{\mathbf{X}}_{-\infty}^{0}\big)\Bigg\} &(6.65)\\
&= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ \chi_{\hat{\mathbf{X}}_0}\left(\mathbb{H}_0\,\middle|\,\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\infty},\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1},\hat{\mathbf{X}}_{-\infty}^{-1}\right) \\
&\qquad + I\left(\hat{\mathbf{X}}_0;\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{\infty}\,\middle|\,\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1},\hat{\mathbf{X}}_{-\infty}^{-1}\right)\Bigg\} &(6.66)\\
&= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ I\big(\{\hat{\mathbf{X}}_k\};\{\mathbb{H}_k\hat{\mathbf{X}}_k\}\big) + \mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|\Big] - \log 2 \\
&\qquad - h\left(\big\{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|\big\}\,\middle|\,\left\{\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\right\}\right)\Bigg\} &(6.67)\\
&= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ I\left(\{\hat{\mathbf{X}}_k\};\left\{\frac{\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k}\right\}_{r=2}^{n_{\mathrm{R}}}\right) - \log 2 + \mathsf{E}\Big[\log\big|\mathbf{H}_0^{(1)\mathsf{T}}\hat{\mathbf{X}}_0\big|^2\Big] \\
&\qquad + h\left(\big\{\Psi_k^{(1)}\big\}\,\middle|\,\big\{\Psi_k^{(r)}-\Psi_k^{(1)}\big\}_{r=2}^{n_{\mathrm{R}}},\big\{|\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k|\big\}_{r=1}^{n_{\mathrm{R}}}\right) \\
&\qquad - h\left(\{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k\}\,\middle|\,\left\{\frac{\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k}\right\}_{r=2}^{n_{\mathrm{R}}},\{\hat{\mathbf{X}}_k\}\right)\Bigg\}. &(6.68)
\end{aligned}$$

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

Moreover, in (6.67) *we have defined*

$$I\big(\{\hat{\mathbf{X}}_k\};\{\mathbb{H}_k\hat{\mathbf{X}}_k\}\big) \triangleq \lim_{n\uparrow\infty}\frac{1}{n}I\big(\hat{\mathbf{X}}_1^n;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^n\big); \tag{6.69}$$

$$h\left(\{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|\}\,\middle|\,\left\{\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\right\}\right) \triangleq \lim_{n\uparrow\infty}\frac{1}{n}h\left(\{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|\}_{\ell=1}^n\,\middle|\,\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^n\right); \tag{6.70}$$

and in (6.68)

$$I\left(\{\hat{\mathbf{X}}_k\};\left\{\frac{\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k}\right\}_{r=2}^{n_\mathrm{R}}\right) \triangleq \lim_{n\uparrow\infty}\frac{1}{n}I\left(\hat{\mathbf{X}}_1^n;\left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathrm{R}\\ \ell=1,\ldots,n}}\right); \tag{6.71}$$

$$\begin{aligned} &h\left(\big\{\Psi_k^{(1)}\big\}\,\middle|\,\big\{\Psi_k^{(r)}-\Psi_k^{(1)}\big\}_{r=2}^{n_\mathrm{R}},\big\{|\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k|\big\}_{r=1}^{n_\mathrm{R}}\right)\\ &\quad\triangleq \lim_{n\uparrow\infty}\frac{1}{n}h\left(\big\{\Psi_\ell^{(1)}\big\}_{\ell=1}^n\,\middle|\,\big\{\Psi_\ell^{(r)}-\Psi_\ell^{(1)}\big\}_{\substack{r=2,\ldots,n_\mathrm{R}\\ \ell=1,\ldots,n}},\big\{|\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell|\big\}_{\substack{r=1,\ldots,n_\mathrm{R}\\ \ell=1,\ldots,n}}\right); \end{aligned} \tag{6.72}$$

$$\begin{aligned} &h\left(\{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k\}\,\middle|\,\left\{\frac{\mathbf{H}_k^{(r)\mathsf{T}}\hat{\mathbf{X}}_k}{\mathbf{H}_k^{(1)\mathsf{T}}\hat{\mathbf{X}}_k}\right\}_{r=2}^{n_\mathrm{R}},\{\hat{\mathbf{X}}_k\}\right)\\ &\quad\triangleq \lim_{n\uparrow\infty}\frac{1}{n}h\left(\{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell\}_{\ell=1}^n\,\middle|\,\left\{\frac{\mathbf{H}_\ell^{(r)\mathsf{T}}\hat{\mathbf{X}}_\ell}{\mathbf{H}_\ell^{(1)\mathsf{T}}\hat{\mathbf{X}}_\ell}\right\}_{\substack{r=2,\ldots,n_\mathrm{R}\\ \ell=1,\ldots,n}},\hat{\mathbf{X}}_1^n\right). \end{aligned} \tag{6.73}$$

Proof. Omitted. □

Note that the expression (6.67) is interesting because it expresses the fading number without using the differential entropy-like quantity $h_\lambda(\cdot)$.

Since the evaluation of (6.34) is in general rather difficult, we will next state two upper bounds to the MIMO fading number that are usually easier to compute.

Corollary 36. *The fading number of a MIMO fading channel with memory as defined in Theorem 34 can be upper-bounded as follows:*

$$\begin{aligned} \chi\big(\{\mathbb{H}_k\}\big) &\leq n_\mathrm{R}\log\pi-\log\Gamma(n_\mathrm{R})\\ &\quad+\inf_{\mathsf{A},\mathsf{B}}\sup_{\hat{\mathbf{x}}_{-\infty}^0}\Big\{n_\mathrm{R}\mathsf{E}\big[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}\|^2\big]-h\big(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}\,\big|\,\{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\big)\Big\} &(6.74)\\ &\leq n_\mathrm{R}\log\pi-\log\Gamma(n_\mathrm{R})\\ &\quad+\inf_{\mathsf{A},\mathsf{B}}\sup_{\hat{\mathbf{x}}}\Big\{n_\mathrm{R}\mathsf{E}\big[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}\|^2\big]-h\big(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}\,\big|\,\mathbb{H}_{-\infty}^{-1}\big)\Big\} &(6.75) \end{aligned}$$

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

Proof. Note that from conditioning that reduces entropy and from (2.9) and (2.10)

$$h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\,\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\infty}^{-1}\right)\leq h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\right)\leq\log c_{n_\mathrm{R}}=\log\frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})} \tag{6.76}$$

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

We now get from Theorem 34 and from (2.27):

$$\chi\bigl(\{\mathbb{H}_k\}\bigr) = \inf_{\mathsf{A},\mathsf{B}} \chi\bigl(\{\mathsf{A}\mathbb{H}_k\mathsf{B}\}\bigr) \tag{6.77}$$

$$\leq \inf_{\mathsf{A},\mathsf{B}} \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigl\{ n_\mathrm{R}\log\pi - \log\Gamma(n_\mathrm{R}) + n_\mathrm{R}\mathsf{E}\Bigl[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{X}}_0\|^2\Bigr]$$

$$- h\Bigl(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{X}}_0 \Bigm| \{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\infty}^0\Bigr)\Bigr\} \tag{6.78}$$

$$= \inf_{\mathsf{A},\mathsf{B}} \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \mathsf{E}\Bigl[n_\mathrm{R}\log\pi - \log\Gamma(n_\mathrm{R}) + n_\mathrm{R}\mathsf{E}\Bigl[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{X}}_0\|^2 \Bigm| \hat{\mathbf{X}}_0 = \hat{\mathbf{x}}_0\Bigr]$$

$$- h\Bigl(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{X}}_0 \Bigm| \{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \{\hat{\mathbf{X}}_\ell = \hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{0}\Bigr)\Bigr] \tag{6.79}$$

$$\leq \inf_{\mathsf{A},\mathsf{B}} \sup_{\hat{\mathbf{x}}_{-\infty}^0} \Bigl\{ n_\mathrm{R}\log\pi - \log\Gamma(n_\mathrm{R}) + n_\mathrm{R}\mathsf{E}\bigl[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}_0\|^2\bigr]$$

$$- h\Bigl(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}_0 \Bigm| \{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}\Bigr)\Bigr\} \tag{6.80}$$

$$\leq \inf_{\mathsf{A},\mathsf{B}} \sup_{\hat{\mathbf{x}}_{-\infty}^0} \Bigl\{ n_\mathrm{R}\log\pi - \log\Gamma(n_\mathrm{R}) + n_\mathrm{R}\mathsf{E}\bigl[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}_0\|^2\bigr]$$

$$- h\Bigl(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}_0 \Bigm| \{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}, \mathbb{H}_{-\infty}^{-1}\Bigr)\Bigr\} \tag{6.81}$$

$$= \inf_{\mathsf{A},\mathsf{B}} \sup_{\hat{\mathbf{x}}} \Bigl\{ n_\mathrm{R}\log\pi - \log\Gamma(n_\mathrm{R}) + n_\mathrm{R}\mathsf{E}\bigl[\log\|\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}\|^2\bigr]$$

$$- h\Bigl(\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}} \Bigm| \mathbb{H}_{-\infty}^{-1}\Bigr)\Bigr\}. \tag{6.82}$$

Here, the first equality follows from (2.27); the subsequent inequality from (6.34) and (6.76); the subsequent equality holds due to linearity of expectation; in the subsequent inequality we upper-bound the expectation by the supremum over all possible values (this proves (6.74)); then the subsequent inequality is due to conditioning that reduces entropy; and the final equality holds because conditional on $\mathbb{H}_{-\infty}^{-1}$, $\mathsf{A}\mathbb{H}_0\mathsf{B}\hat{\mathbf{x}}_0$ is independent of $\{\mathsf{A}\mathbb{H}_\ell\mathsf{B}\hat{\mathbf{x}}_\ell\}_{\ell=-\infty}^{-1}$. □

6.5 Some Special Cases

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

6.5.1 Memoryless Fading

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

The expression for the fading number of a memoryless MIMO fading channel (4.8) can be derived from (6.34) as follows: firstly note that only the first and the last term in (6.34) are influenced by memory. However, once we assume that there is no memory in the fading process $\{\mathbb{H}_k\}$, the past can only influence the present values via some memory in the input process $\{\hat{\mathbf{X}}_k\}$. Now note that the fourth term is conditioned on the input of the past *and of the present*. Hence, the past has no influence on this term either. Finally, note that the first term can be upper-bounded by dropping the conditioning, and note further that this upper-bound can actually be achieved if the input is chosen to be IID. Hence, an optimum choice of $Q_{\{\hat{\mathbf{X}}_k\}}$ will be memoryless, and (4.8) follows.

6.5.2 MISO Fading With Memory

Next we are going to study the special case of MISO fading with memory whose fading number has been unknown so far. If we specialize Theorem 34 to the situation of only one antenna at the receiver we get the following corollary.

Corollary 37. *Consider a MISO fading channel with memory where the stationary and ergodic fading process* $\{\mathbf{H}_k\}$ *takes value in* $\mathbb{C}^{n_\mathrm{T}}$ *and satisfies* $h(\{\mathbf{H}_k\}) > -\infty$ *and* $\mathsf{E}\big[\|\mathbf{H}_k\|^2\big] < \infty$*. Then, irrespective of whether a peak-power constraint* (2.23) *or an average-power constraint* (2.24) *is imposed on the input, the fading number* $\chi\big(\{\mathbf{H}_k^\mathsf{T}\}\big)$ *is given by*

$$\chi\big(\{\mathbf{H}_k^\mathsf{T}\}\big) = \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}}} \left\{ \log\pi + \mathsf{E}\Big[\log|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0|^2\Big] - h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\infty}^0\big) \right\} \tag{6.83}$$

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

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

$$\mathbf{X}_k = R_k \cdot \hat{\mathbf{X}}_k \cdot e^{\mathrm{i}\Theta_k}. \tag{6.84}$$

Here $\{\hat{\mathbf{X}}_k\} \in \mathbb{C}^{n_\mathrm{T}}$ *is a stationary unit-vector process with the distribution that maximizes* (6.83)*;* $\{R_k\} \in \mathbb{R}_0^+$ *is a scalar non-negative IID random process satisfying* (6.36)*; and* $\{\Theta_k\}$ *is IID* $\sim \mathcal{U}([0, 2\pi))$ *as defined in Definition 28.*

Proof. This follows directly from Theorem 34 by the observation that independently of the distribution of $\{\mathbf{H}_k\}$ and $\{\hat{\mathbf{X}}_k\}$, in the MISO case the distribution of

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

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

$$h_\lambda\left(\frac{\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0}{\|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0\|} e^{\mathrm{i}\Theta_0} \,\middle|\, \left\{ \frac{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell}{\|\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\|} e^{\mathrm{i}\Theta_\ell} \right\}_{\ell=-\infty}^{-1} \right) = \log 2\pi. \tag{6.85}$$

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

Remark 38. *We would like to remind you that in the case of a MISO fading process without memory the optimal input uses beam-forming with a deterministic direction that maximizes the fading number, see* (1.8)*. Once the fading process has memory this is not the case anymore. However, it is straightforward to derive the upper bound* (1.9) *and the lower bound* (1.10) *that are of beam-forming type* [26]*: the upper bound follows by upper-bounding the expectation over* $\hat{\mathbf{X}}_{-\infty}^0$ *by the supremum over* $\hat{\mathbf{x}}_{-\infty}^0$*. For the lower bound we choose the following stationary and circularly symmetric distribution on* $\{\hat{\mathbf{X}}_k\}$*:*

$$\hat{\mathbf{X}}_k = \hat{\mathbf{x}} e^{\mathrm{i}\Theta_k} \tag{6.86}$$

where $\hat{\mathbf{x}}$ *is the deterministic direction that achieves the maximum in* (1.10)*.*

6.5.3 SIMO and SISO Fading With Memory

In the situation with only one antenna at the transmitter the input vector $\mathbf{X}_k$ is reduced to a random variable X_k and therefore the input direction $\hat{\mathbf{X}}_k$ to a phase $e^{\mathsf{i}\Phi_k}$. Hence the expression (6.34) gets simplified considerably by the fact that there is only one choice of a circularly symmetric distribution of $e^{\mathsf{i}\Phi_k}$:

$$\hat{\mathbf{X}}_k = e^{\mathsf{i}\Phi_k} = e^{\mathsf{i}\Theta_k}, \quad \forall\, k, \tag{6.87}$$

i.e., the supremum disappears.

The fading number of a SIMO fading channel with memory follows then from (6.34) in a straightforward way from the fact that

$$h\big(\mathbf{H}_0 e^{\mathsf{i}\Theta_0} \,\big|\, \{\mathbf{H}_\ell e^{\mathsf{i}\Theta_\ell}\}_{\ell=-\infty}^{-1}, \{e^{\mathsf{i}\Theta_\ell}\}_{\ell=-\infty}^{0}\big) = h\big(\mathbf{H}_0 \,\big|\, \mathbf{H}_{-\infty}^{-1}\big). \tag{6.88}$$

The expression (1.6) can be derived from the alternative form (6.66).

For the fading number of a SISO fading channel with memory (1.5) we use

$$h_\lambda\big(e^{\mathsf{i}\Theta_0} \,\big|\, \{e^{\mathsf{i}\Theta_\ell}\}_{\ell=-\infty}^{-1}\big) = \log 2\pi. \tag{6.89}$$

Chapter 7

Summary & Discussion

7.1 Summary

Our main contributions are as follows:

- For a non-central chi-square distributed random variable V with an even number of degrees of freedom we have derived closed-form expressions for
  - $\mathsf{E}[\log V]$,
  - $\mathsf{E}\left[\frac{1}{V^n}\right]$, for $n \in \mathbb{N}$.
- We have shown that all these expressions depend on a family of functions $g_m(\cdot)$ and have derived properties and bounds for these functions.
- We have given an expression for the fading number of general memoryless MIMO fading channels where the fading process is assumed to be IID over time, but have arbitrary spatial dependencies.
- We have specialized this expression for the situation of spatially and temporally IID Gaussian MIMO fading channels with a scalar line-of-sight component.
- We have shown that an optimum input distribution for the most general single-user case, *i.e.*, for the general MIMO fading channel with memory, can be assumed to be circularly symmetric.
- We have shown that for a general *stationary* channel model with input alphabet $\mathbb{C}^{n_\mathrm{T}}$ and output alphabet $\mathbb{C}^{n_\mathrm{R}}$ an optimum input basically is also stationary.
- We have derived an expression for the fading number of a general MIMO fading channel with memory.

We have published one journal paper:

- Stefan M. Moser: "The Fading Number of Memoryless Multiple-Input Multiple-Output Fading Channels," *IEEE Transactions on Information Theory*, vol. 53, no. 7, pp. 2652–2666, July 2007 [29],

and submitted two more:

- Stefan M. Moser: "Some Expectations of a Non-Central Chi-Square Distribution With an Even Number of Degrees of Freedom," submitted, April 27, 2007 [36],

- Stefan M. Moser: "The Fading Number of Multiple-Input Multiple-Output Fading Channels with Memory," submitted, April 6, 2007 [30].

Furthermore, we presented our results at the following conferences:

- Stefan M. Moser: "The Fading Number of Multiple-Input Multiple-Output Fading Channels with Memory," Proceedings *2007 International Symposium on Information Theory (ISIT'07)*, pp. 521–525, Nice, France, Jun. 24–30, 2007 [37],
- Stefan M. Moser: "The Fading Number of Multiple-Input Multiple-Output Fading Channels with Memory," Poster session *2007 Fall Workshop on Information Theory and Communications*, Penghu, Taiwan, Aug. 2–3, 2007 [38],
- Stefan M. Moser: "The Fading Number of IID MIMO Gaussian Fading Channels with a Scalar Line-of-Sight Component," Proceedings *45th Annual Allerton Conference on Communication, Control, and Computing*, Allerton House, Monticello, IL, USA, Sep. 26–28, 2007 [39],
- Stefan M. Moser: "Some Expectations of a Non-Central Chi-Square Distribution With an Even Number of Degrees of Freedom," Proceedings *2007 IEEE International Region 10 Conference (TENCON 2007)*, Taipei, Taiwan, Oct. 30 – Nov. 2, 2007 [40],
- Stefan M. Moser: "The Fading Number of Memoryless Multiple-Input Multiple-Output Fading Channels," Poster session *2007 National Symposium on Telecommunications*, Taipei, Taiwan, Nov. 23–24, 2007 [41].

7.2 Discussion

The topic under study in this project is the theoretical upper limit on the rate of reliable transmission that can be achieved over a wireless, mobile communication system. This upper limit is characterized by the *channel capacity* in the case of a single-user setup or the *capacity region* in the situation of multiple users. We assume an OFDM system such that for our analysis we can neglect inter-symbol interference, but only need to deal with the problem of *fading*.

We define a discrete-time version of a corresponding channel model where we try to make as few assumptions as possible in order to keep the results as general as possible. Hence, we do not make any particular assumption about the distribution of the fading process (*i.e.*, not necessarily Gaussian!), allow in general for an arbitrary number of antennas both at transmitter and receiver, and consider multiple users. We also allow memory, both over time and space, *i.e.*, the different fading coefficients of different antennas and at different times might be dependent.

The only restriction applied is the so-called *regularity assumption*. To explain this assumption slightly imprecisely in an engineering way we might say that we ask the fading process to be fully random in the sense that even with the knowledge of all past fading realizations and of all fading realizations of neighboring antennas one cannot predict the actual value of the fading precisely. The prediction error might be very small once one knows the past and the neighboring fading values, but it is still non-zero. Considering the fact that in reality no measurement is absolutely perfect this assumption seems to be realistic.

7.2.1 Fading Number

We concentrate our study to the high- and highest-SNR regime where the available power becomes very large. In particular we are interested in the *fading number* χ:

$$\chi \triangleq \lim_{\text{SNR}\to\infty} \Big\{ \mathsf{C}(\text{SNR}) - \log\big(1 + \log\big(1 + \text{SNR}\big)\big) \Big\}, \tag{7.1}$$

i.e., in the second term of the high-SNR asymptotic expansion of capacity. Note that the capacity at high SNR can be written as

$$\mathsf{C}(\text{SNR}) = \log(1 + \log(1 + \text{SNR})) + \chi + o(1), \tag{7.2}$$

where $o(1)$ denotes terms that tend to zero as the SNR tends to infinity.

We motivate our interest in the high-SNR regime and the fading number by arguing that the knowledge of the high-SNR capacity behavior has a very strong impact on practical considerations when designing such mobile communication systems, even if one does never operate a system at such high SNR levels. The argument is based on the observation that the high-SNR behavior (7.2) is extremely poor because the capacity grows only double-logarithmically in the SNR. in other words this means that in the regime where the approximation

$$\mathsf{C}(\text{SNR}) \approx \log(1 + \log(1 + \text{SNR})) + \chi \tag{7.3}$$

is valid, any additional bit of capacity requires a *squaring* of the SNR, or, on a dB-scale, *doubling the dB-value* of the SNR. This behavior must be avoided in any practical communication system, *i.e.*, a system must be designed to operate at lower SNRs. Hence, the question arises to where the *threshold* lies between this highly inefficient high-SNR regime and the normal low- to medium-SNR regime.

To answer this question the fading number can give an interesting answer: pulling ourselves by our bootstraps, let us consider for the moment that (7.3) starts to be valid for an SNR somewhere in the range[1] of 30 to 80 dB. In spite of this rather large range the function $\log(1+\log(1+\text{SNR}))$ will vary only between 2 and 3 nats. Hence, the capacity will vary between $2 + \chi$ and $3 + \chi$ nats. Therefore, we can conclude that once the capacity is appreciably above $\chi + 2$ nats, the approximation (7.3) is likely to be valid. Therefore, the fading number can be seen as an indicator of the maximum rate at which power efficient communication is possible on the channel.

Note that while the term $\log(1+\log(1+\text{SNR}))$ remains always the same independently of the details of the channel model, the value of the fading number strongly depends on the specific assumptions of the channel model like fading distribution, number of antennas at transmitter and receiver, number of users, type of memory in the channel, etc.

For a further discussion about the practical relevance of the fading number we also refer to [16] and [15].

In the process of our study we further restrict the channel model to some special cases, however, always keeping in mind the ultimate goal of an analysis of the unrestricted case. In this report we concentrated on two special cases:

- a general single-user memoryless MIMO fading channel and its important special case of an IID Gaussian fading, and
- a general single-user MIMO fading channel with memory.

[1] This assumption is reasonable for many channels encountered in practice.

7.2.2 A Single-User Memoryless MIMO Fading Channel

In the former case we are able to derive the fading number of a MIMO fading channel of general fading law including spatial, but without temporal memory. Since the fading number is the second term after the double-logarithmic term of the high-SNR expansion of channel capacity, this means that we precisely specify the behavior of the channel capacity asymptotically when the power grows to infinity. We further show that the asymptotic capacity can be achieved by an input that consists of the product of two independent random quantities: a circularly symmetric random unit vector (the *direction*) and a non-negative (*i.e.*, real) random variable (the *magnitude*). The distribution of the random direction is chosen such as to maximize the fading number and therefore depends on the particular law of the fading process. The distribution of the magnitude has the standard logarithmically uniform distribution (6.36) that has been used in previous publications about the fading number since it also achieves the fading number in the SISO and SIMO case and is also used in the MISO case where it is multiplied by a constant beam-direction $\hat{\mathbf{x}}$. All these special cases follow nicely from this new result.

The derivation of this result is based on three main techniques or observations: firstly, we know that the capacity achieving input distribution must *escape to infinity* which means that at high SNR no finite energy input symbols should be used anymore. Secondly, we use the dual-based approach of deriving upper bounds to channel capacity as firstly described in [42] without application and as firstly applied in [8] and [3]. Finally we prove that in general the capacity achieving input distribution must be circularly symmetric.

We then investigate the important special situation of Gaussian fading (5.3) where we assume a scalar line-of-sight matrix (5.4)–(5.5) and are able to show that this fading number basically grows like

$$\chi = n_{\mathrm{m}} g_{n_{\mathrm{m}}}\big(|d|^2\big) \tag{7.4}$$

where

$$n_{\mathrm{m}} = \min\{n_{\mathrm{R}}, n_{\mathrm{T}}\} \tag{7.5}$$

is called *degree of freedom*.

7.2.3 A Single-User MIMO Fading Channel with Memory

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

We have shown that the MIMO fading number is achievable by an input process that can be written as a product of two independent processes: an IID non-negative "magnitude" process and a stationary and circularly symmetric "direction" process. The former has again the standard logarithmically uniform distribution (6.36). It escapes to infinity as guaranteed by Proposition 17.

The "direction" process depends on the particular law of the fading process, *i.e.*, it needs to be chosen such as to maximize the fading number. The expression of the fading number is therefore not given in a completely closed form but still contains a maximization. However, one has to be aware that we have not specified the fading process in closed form either, *i.e.*, we do not believe it possible to further simply (6.34) without making more detailed assumptions about $\{\mathbb{H}_k\}$.

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

Moreover, the proof relies on a lemma which says that in general the capacity-achieving input distribution must be circularly symmetric. The proof of this lemma only relies on the fact that the additive noise is circularly symmetric. Therefore this result holds in a most general setting. Note that a circularly symmetric input can pack more information into the phase than any other phase distribution. Since the additive noise does not favor any direction, there is no need to protect some direction more than others and it is optimal to pack as much information into the phase as possible.

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

7.2.4 Non-Central Chi-Square Distribution

As a side product we were also able to derive some important and so-far unknown closed-form expressions of expectations of a non-central chi-square distributed random variable V:

$$\mathsf{E}[\log V] = g_m(s^2), \tag{7.6}$$

$$\mathsf{E}\left[\frac{1}{V^n}\right] = \frac{(-1)^{n-1}}{(n-1)!} \cdot g_{m-n}^{(n)}(s^2), \qquad m > n, \tag{7.7}$$

where $2m$ is the degree of freedom of V and s^2 is its non-centrality. The family of functions $g_m(\cdot)$ is defined in (3.7). We have shown that these functions are continuous, monotonically increasing, and concave and have derived tight upper and lower bounds.

7.3 Outlook

There still remain many unanswered questions. In particular, we have restricted our study recently more on the single-user case for technical reasons: the fact that in a multiple-user case some subsets of the transmitter antennas are not allowed to cooperate makes the analysis considerably more difficult. All the more we would be very much interested in results in the more general multiple-user case and we will try to extend our analysis to this case.

Another direction of further study is the specialization of the results in Chapter 6. The expression of the MIMO fading number with memory is rather difficult due to the inherent maximization. We would like to make some further assumptions about the fading process like, *e.g.*, a Gaussian distribution in order to evaluate the expression and get a closed-form. Similarly, it would be interesting to get a fully closed-form expression for Theorem 24, without the mutual information term in (5.20).

Bibliography

[1] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, 2002.

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

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

[4] 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.

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

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

[7] A. Lapidoth and S. M. Moser, "Bounds on the capacity of the discrete-time Poisson channel," in *Proceedings Forty-First Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, IL, USA, October 1–3, 2003, pp. 201–210.

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

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

[10] A. Lapidoth and S. M. Moser, "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] A. Lapidoth and S. M. Moser, "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] A. Lapidoth and S. M. Moser, "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] A. Lapidoth and S. M. Moser, "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] T. Koch and A. Lapidoth, "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, IL, USA, 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] A. Lapidoth, "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] A. Lapidoth, "Capacity bounds via duality: a phase noise example," in *Proceedings Second Asian-European Workshop on Information Theory*, Breisach, Germany, June 26–29, 2002.

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

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

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

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

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

[28] S. M. Moser, "Capacity analysis of multiple-access ofdm channels," Department of Communication Engineering, National Chiao Tung University (NCTU), Zhudong, Taiwan, final report of project4G Wireless Access Technology G1-95003, from January 1, to December 31, 2006, funded by Industrial Technology Research Institute (ITRI).

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

[30] S. M. Moser, "The fading number of multiple-input multiple-output fading channels with memory," April 2007, *IEEE Transactions on Information Theory.* [Online]. Available: http://moser.cm.nctu.edu.tw/publications.html

[31] M. B. Monagan, K. O. Geddes, K. M. Heal, G. Labahn, S. M. Vorkoetter, J. McCarron, and P. DeMarco, *Maple 10 Programming Guide*. Waterloo ON, Canada: Maplesoft, 2005.

[32] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed., A. Jeffrey, Ed. Academic Press, San Diego, 1996.

[33] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed. John Wiley & Sons, 1995, vol. 2.

[34] R. Etkin and D. Tse, "Degrees of freedom in underspread MIMO fading channels," in *IEEE International Symposium on Information Theory (ISIT)*, Yokohama, Japan, June 29 – July 4, 2003, p. 323.

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

[36] S. M. Moser, "Some expectations of a non-central chi-square distribution with an even number of degrees of freedom," April 2007, submitted. [Online]. Available: http://moser.cm.nctu.edu.tw/publications.html

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

[38] S. M. Moser, "The fading number of multiple-input multiple-output fading channels with memory," in *Poster Session 2007 Fall Workshop on Information Theory and Communications*, Penghu, Taiwan, August 2–3, 2007.

[39] S. M. Moser, "The fading number of IID MIMO Gaussian fading channels with a scalar line-of-sight component," in *Proceedings Forty-Fifth Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, IL, USA, September 26–28, 2007.

[40] S. M. Moser, "Some expectations of a non-central chi-square distribution with an even number of degrees of freedom," in *IEEE International Region 10 Conference (TENCON)*, Taipei, Taiwan, October 31 – November 2, 2007.

[41] S. M. Moser, "The fading number of memoryless multiple-input multiple-output fading channels," in *Poster Session 2007 National Symposium on Telecommunication*, Taipei, Taiwan, November 23–24, 2007.

[42] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Studia Scientarum Mathematicarum Hungarica*, vol. 2, pp. 291–292, 1967.