

The Fading Number of Memoryless Multiple-Input Multiple-Output Fading Channels

Final Report of NSC Project
“Capacity Analysis of Joint Estimation and
Detection for MIMO Fading Channels”

Project-Number: NSC 95-2221-E-009-046
Project Duration: 1 August 2006 – 31 July 2007
Funded by: National Science Council, Taiwan
Author: Stefan M. Moser
Organization: Information Theory Laboratory
Department of Communication
Engineering
National Chiao Tung University
Address: Engineering Building IV, Room 727
1001 Ta Hsueh Rd.
Hsinchu 300, Taiwan
E-mail: stefan.moser@ieee.org

Abstract

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

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

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

New detection schemes, however, will not be sufficient for future mobile transmission systems in order to achieve the aimed high data rates. Another promising development step is to incorporate multiple antennas, both at transmitter and receiver. It therefore becomes important to evaluate the channel capacity of such a multiple-input multiple-output (MIMO) system.

In this report, both innovations—joint estimation and detection and multiple antennas—are studied together. The main result is an exact expression of the fading number of a MIMO fading channel without temporal memory. While we restrict ourselves to fading processes that are temporally independent and identically distributed (IID), we do allow dependencies between the fading of the different paths corresponding to the different antennas (spatial memory). Furthermore, the fading law is not restricted to be Gaussian, but is assumed to be a general regular law with spatial (but without temporal) memory. This result can be seen as a further step towards the final goal of the fading number of general MIMO fading channels with memory.

The result is also specialized to the known cases of single-input multiple-output (SIMO), multiple-input single-output (MISO), and single-input single-output (SISO) fading channels, as well as to the situation of Gaussian fading.

Keywords: Channel capacity, fading, fading number, flat fading, high SNR, joint estimation and detection, memory, MIMO, multiple-antenna, non-coherent detection.

Contents

Acknowledgments	1
1 Introduction	2
1.1 General Background	2
1.2 The Fading Number	6
2 Definitions and Notation	8
2.1 Notation	8
2.2 The Channel Model	10
3 Main Results	12
3.1 Preliminaries	12
3.1.1 Escaping to Infinity	12
3.1.2 An Upper Bound on Channel Capacity	13
3.1.3 Capacity-Achieving Input Distributions and Circular Symmetry	13
3.2 Fading Number of General Memoryless MIMO Fading Channels . . .	14
4 Special Cases	16
4.1 Some Known Special Cases	16
4.2 Gaussian Fading	18
4.2.1 Scalar Line-of-Sight Matrix	18
4.2.2 General Line-of-Sight Matrix	20
5 Proof of the Main Result	22
5.1 Derivation of an Upper Bound	23
5.2 Derivation of a Lower Bound	29
6 Discussion & Conclusion	34
A Some Derivations	36
A.1 Proof of Lemma 6	36
A.2 Derivation of Bounds (4.25)	36
A.3 Proof of Corollary 13	37
A.4 Proof of Proposition 14	38
A.5 Additional Derivation for the Proof of the Lower Bound	39
Bibliography	41

Acknowledgments

The author would like to thank Amos Lapidoth for long, fruitful discussions and Tobias Koch for giving the right hint to fix a serious bug in an earlier version of the proof.

This work was supported by the National Science Council under NSC 95-2221-E-009-046 and by the Industrial Technology Research Institute (ITRI), Zhudong, Taiwan, under contract G1-96001.

These results have been presented at the *2007 IEEE International Symposium on Information Theory (ISIT 2007)*, Nice, France, and at the *National Symposium on Telecommunications (NST 07)*, Taipei, Taiwan.

Chapter 1

Introduction

1.1 General Background

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 a old-fashioned wired telephone, the engineering technique behind it is very much different. The reason for this is that in a wireless communication system several physical effects occur that change the behavior of the channel completely compared with wired communication:

- The signal may find many different paths from the sender to the receiver via various different reflections (buildings, trees, etc.). Therefore the receiver receives multiple copies of the same signal, however, since each path has different length and different attenuation, the various copies of the signal will arrive at different times and with different strength.
- Since the transmitter and/or the receiver might be in motion while transmitting, a physical phenomenon called *Doppler effect* occurs: the frequency of the transmitted signal is shifted depending on the relative movement between receiver and transmitter.
- Since receiver and transmitter are moving and because the environment is permanently changing (*e.g.*, movements by wind, passing cars, people, etc.), the different signal paths are constantly changing.

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

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

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

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

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

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

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

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

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

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

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

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

In [2] and extracts thereof published separately [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14] large progress has been made in tackling this problem: a technique has been proposed for the derivation of upper bounds on channel capacity.² It is based on a dual expression for channel capacity where the maximization (of mutual information) over distributions on the channel input alphabet is replaced with a minimization (of average relative entropy) over distributions on the channel output alphabet. Every choice of an output distribution leads to an upper bound on mutual information. The chosen output distribution need not correspond to some distribution on the channel input. With a judicious choice of output distributions one can often derive tight upper bounds on channel capacity.

Furthermore, in [2] a technique has been proposed for the analysis of the *asymptotic* capacity of general cost-constrained channels, *i.e.*, the channel capacity in the limited of unrestricted cost.³ The technique is based on the observation that—under fairly mild conditions on the channel—every input distribution that achieves a mutual information with the same growth-rate in the cost constraint as the channel capacity must *escape to infinity*; *i.e.*, under such a distribution for some finite cost, the probability of the set of input symbols of lesser cost tends to zero as the cost constraint tends to infinity. For more details about this concept see Section 3.1.1.

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

- the free-space optical intensity channel [2], [6], [8];
- an optical intensity channel with input-dependent noise [2];
- the Poisson channel [2], [6], [8], [9];
- multiple-antenna flat fading channels with memory where the fading process is assumed to be *regular* (*i.e.*, of finite entropy rate⁴) and where the realization of the fading process is unknown at the transmitter and unknown (or only partially known) at the receiver [2], [4], [7];
- multiple-antenna flat fading channels with memory where the fading process may be *irregular* (*i.e.*, of possibly infinite entropy rate) and where the realization of the fading process is unknown (or only partially known) at the receiver [15], [16], [17], [18], [19];
- fading channels with feedback [20], [2], [5];
- non-coherent fading networks [21], [22];
- a phase noise channel [23], [24].

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

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

³In the situation of fading channels the cost is the available power.

⁴*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.

the logarithm of the SNR in the limit when the SNR tends to infinity) could be quantified.

Some of these results have been very unexpected. *E.g.*, it has been shown in [2] that regular fading processes have a capacity that grows only double-logarithmically in the SNR at high SNR. This means that at high power these channels become extremely power-inefficient in the sense that for every additional bit capacity the SNR needs to be squared or, respectively, on a dB-scale the SNR needs to be doubled! This behavior is independent of the particular law of the fading process, the law of the noise process, or the number of antennas at the transmitter or receiver. Moreover, the capacity-growth at high SNR is double-logarithmic irrespective whether there is memory in the fading process or not, and it even remains this slow when introducing *noiseless* feedback [20]! This is in stark contrast to the situation of additive noise channels and even to the so far known capacity results when assuming perfect 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!

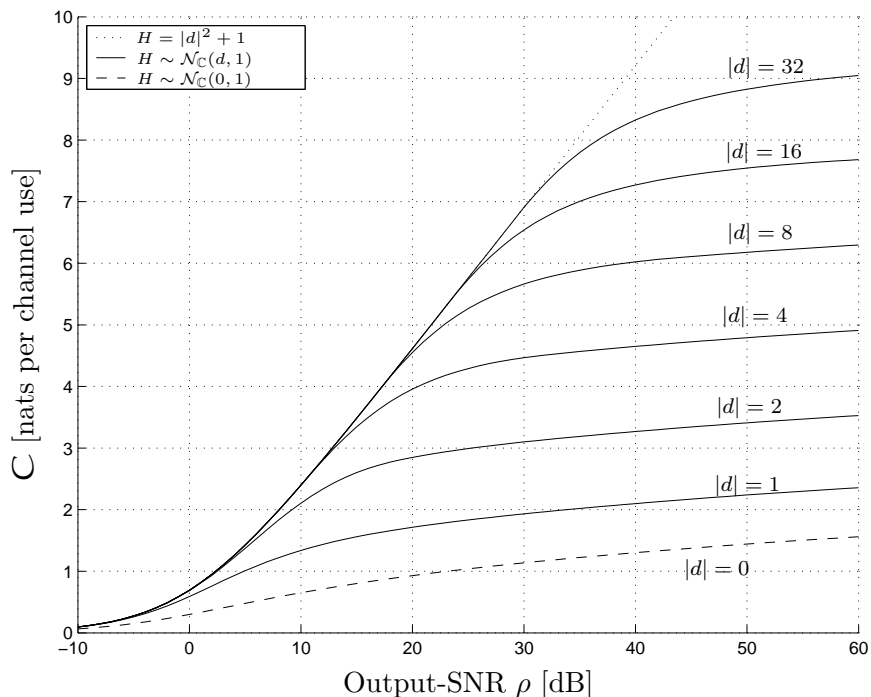


Figure 1: An upper bound on the capacity of a Rician fading channel for different values of the specular component d . The dotted line depicts the capacity of a Gaussian channel of equal output-SNR, namely $\log(1 + \rho)$.

Therefore the question arises whether in the case of non-coherent fading channels multiple antennas or feedback is useful at all. It turns out that although the asymptotic growth rate of capacity is unchanged by these parameters, they still do have a large (positive) influence on the system: the threshold above which the capacity growth changes from logarithmic to double-logarithmic is highly dependent on them! As an example Figure 1 shows the capacity of non-coherent Rician fading channels with various values of the line-of-sight component.

1.2 The Fading Number

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

$$C(\text{SNR}) = \log \log \text{SNR} + \chi + o(1), \quad (1.1)$$

where $o(1)$ denotes terms that tend to zero as the SNR tends to infinity and where χ is the fading number. For a mathematically more precise definition we refer to Section 2.2.

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

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

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

Now consider the following situation: assume for the moment that the threshold SNR_0 lies somewhere between 30 and 80 dB (it can be shown that this is a reasonable assumption for many channels that are encountered in practice). In this case, the threshold capacity $C_0 = C(\text{SNR}_0)$ must be somewhere in the following interval:

$$\log \log(30 \text{ dB}) + \chi \leq C_0 \leq \log \log(80 \text{ dB}) + \chi, \quad (1.3)$$

$$\implies \chi + 2.1 \text{ nats} \leq C_0 \leq \chi + 3 \text{ nats}. \quad (1.4)$$

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

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

Hence the fading number can be regarded as quality attribute of the channel: the larger the fading number is, the higher is the maximum rate at which the channel can be used without being extremely power-inefficient.

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

For memoryless fading channels, the fading number is known in the situation of only one antenna at transmitter and receiver (SISO)

$$\chi(H) = \log \pi + \mathbf{E}[\log |H|^2] - h(H); \quad (1.5)$$

in the situation of a SIMO fading channel⁵

$$\chi(\mathbf{H}) = h_\lambda(\hat{\mathbf{H}}e^{i\Theta}) + n_R \mathbf{E}[\log \|\mathbf{H}\|^2] - \log 2 - h(\mathbf{H}) \quad (1.6)$$

⁵For a precise definition of the notation used in this paper we refer to Section 2.1.

(both are special cases from the corresponding situation with memory); and also in the case of a multiple-input single-output (MISO) fading channel [7], [2]

$$\chi(\mathbf{H}^\top) = \sup_{\hat{\mathbf{x}}} \left\{ \log \pi + \mathbf{E} \left[\log |\mathbf{H}^\top \hat{\mathbf{x}}|^2 \right] - h(\mathbf{H}^\top \hat{\mathbf{x}}) \right\}. \quad (1.7)$$

The most general situation of multiple antennas at both transmitter and receiver, however, has been solved so far only in the special case of a particular rotational symmetry of the fading process: 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 [7], [2] to be

$$\chi(\mathbb{H}) = \log \frac{\pi^{n_R}}{\Gamma(n_R)} + n_R \mathbf{E} \left[\log \|\mathbb{H} \hat{\mathbf{e}}\|^2 \right] - h(\mathbb{H} \hat{\mathbf{e}}) \quad (1.8)$$

where $\hat{\mathbf{e}} \in \mathbb{C}^{n_T}$ 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 4.1).

In this report we will extend these results and derive the fading number of general (not necessarily Gaussian) memoryless MIMO fading channels.

The remaining of this report is structured as follows: after some remarks about notation and a detailed mathematical definition of the channel model in the following chapter, we will present the main result, *i.e.*, the fading number of the general memoryless MIMO fading channel, in Chapter 3. There also some concepts are introduced that are important in the analysis of channel capacity, *e.g.*, the concept of distributions that *escape to infinity*, and the relation between circular symmetry and capacity-achieving input distributions.

In Section 4.1 the known fading numbers of SISO, SIMO, MISO, and rotation-commutative MIMO fading channels are derived once more as special cases of the new general result from Chapter 3. In Section 4.2 we investigate the situation of Gaussian fading processes.

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

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.*, \mathbb{H} ; and random matrices are denoted using another special upper-case font, *e.g.*, \mathbb{C} . The capacity is denoted by \mathcal{C} , the energy per symbol by \mathcal{E} , and the signal-to-noise ratio SNR is denoted by SNR.

We use the shorthand H_a^b for $(H_a, H_{a+1}, \dots, H_b)$. For more complicated expressions, such as $(\mathbf{H}_a^\top \hat{\mathbf{x}}_a, \mathbf{H}_{a+1}^\top \hat{\mathbf{x}}_{a+1}, \dots, \mathbf{H}_b^\top \hat{\mathbf{x}}_b)$, we use the dummy variable ℓ to clarify notation: $\{\mathbf{H}_\ell^\top \hat{\mathbf{x}}_\ell\}_{\ell=a}^b$.

Hermitian conjugation is denoted by $(\cdot)^\dagger$, and $(\cdot)^\top$ stands for the transpose (without conjugation) of a matrix or vector. The trace of a matrix is denoted by $\text{tr}(\cdot)$.

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

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

$$\|\mathbf{A}\| \triangleq \max_{\|\hat{\mathbf{w}}\|=1} \|\mathbf{A}\hat{\mathbf{w}}\|. \quad (2.2)$$

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

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

$$\|\mathbf{A}\|_F \triangleq \sqrt{\text{tr}(\mathbf{A}^\dagger \mathbf{A})}. \quad (2.3)$$

Note that for every matrix \mathbf{A}

$$\|\mathbf{A}\| \leq \|\mathbf{A}\|_F \quad (2.4)$$

as can be verified by upper-bounding the squared magnitude of each of the components of $\mathbf{A}\hat{\mathbf{w}}$ 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*

$$\hat{\mathbf{v}} \triangleq \frac{\mathbf{v}}{\|\mathbf{v}\|} \quad (2.5)$$

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

$$\|\hat{\mathbf{v}}\| = \|\hat{\mathbf{V}}\| = 1. \quad (2.6)$$

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 $p_{\hat{\mathbf{V}}}^{\lambda}(\hat{\mathbf{v}})$ with respect to λ , then we shall let

$$h_{\lambda}(\hat{\mathbf{V}}) \triangleq -\mathbb{E}\left[\log p_{\hat{\mathbf{V}}}^{\lambda}(\hat{\mathbf{V}})\right] \quad (2.7)$$

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 \mathbf{U} is a deterministic unitary matrix, then

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

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

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

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

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

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 $p_{\hat{\mathbf{V}}|\mathbf{W}}^{\lambda}(\hat{\mathbf{v}}|\mathbf{w})$ then we can define

$$h_{\lambda}(\hat{\mathbf{V}} | \mathbf{W} = \mathbf{w}) \triangleq -\mathbb{E}\left[\log p_{\hat{\mathbf{V}}|\mathbf{W}}^{\lambda}(\hat{\mathbf{V}}|\mathbf{W}) \mid \mathbf{W} = \mathbf{w}\right] \quad (2.11)$$

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

Based on these definitions we have the following lemma:

Lemma 2. *Let \mathbf{V} be a complex random vector taking value in \mathbb{C}^m and having differential entropy $h(\mathbf{V})$. Let $\|\mathbf{V}\|$ denote its norm and $\hat{\mathbf{V}}$ denote its direction as in (2.5). Then*

$$h(\mathbf{V}) = h(\|\mathbf{V}\|) + h_{\lambda}(\hat{\mathbf{V}} | \|\mathbf{V}\|) + (2m - 1)\mathbb{E}[\log \|\mathbf{V}\|] \quad (2.12)$$

$$= h_{\lambda}(\hat{\mathbf{V}}) + h(\|\mathbf{V}\| | \hat{\mathbf{V}}) + (2m - 1)\mathbb{E}[\log \|\mathbf{V}\|] \quad (2.13)$$

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.

Proof. Omitted. □

We shall write $\mathbf{X} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \mathbf{K})$ if $\mathbf{X} - \boldsymbol{\mu}$ is a circularly symmetric, zero-mean, Gaussian random vector of covariance matrix $\mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\dagger] = \mathbf{K}$. 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 the paper $e^{i\Theta}$ denotes a complex random variable that is uniformly distributed over the unit circle

$$e^{i\Theta} \sim \text{Uniform on } \{z \in \mathbb{C} : |z| = 1\}. \quad (2.14)$$

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

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

2.2 The Channel Model

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

$$\mathbf{Y}_k = \mathbb{H}_k \mathbf{x}_k + \mathbf{Z}_k. \quad (2.15)$$

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

We assume that the random vectors $\{\mathbf{Z}_k\}$ are spatially and temporally white, zero-mean, circularly symmetric, complex Gaussian random vectors, *i.e.*, $\{\mathbf{Z}_k\} \sim \text{IID } \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma^2 \mathbf{I}_{n_R})$ for some $\sigma^2 > 0$. Here \mathbf{I}_{n_R} denotes the $n_R \times n_R$ 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.*,

$$\mathbb{E}[\|\mathbb{H}_k\|_{\text{F}}^2] < \infty \quad (2.16)$$

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

$$h(\{\mathbb{H}_k\}) \triangleq \lim_{n \uparrow \infty} \frac{1}{n} h(\mathbb{H}_1, \dots, \mathbb{H}_n) > -\infty. \quad (2.17)$$

Furthermore, we will restrict ourselves to the memoryless case, *i.e.*, we assume that $\{\mathbb{H}_k\}$ is IID with respect to time k . Since there is no memory in the channel, an IID input process $\{\mathbf{X}_k\}$ will be sufficient to achieve capacity and we will therefore drop the time index k hereafter, *i.e.*, (2.15) simplifies to

$$\mathbf{Y} = \mathbb{H}\mathbf{x} + \mathbf{Z}. \quad (2.18)$$

Note that while 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.

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} .

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

$$\text{SNR} \triangleq \frac{\mathcal{E}}{\sigma^2}. \quad (2.19)$$

The capacity $C(\text{SNR})$ of the channel (2.18) is given by

$$C(\text{SNR}) = \sup_{Q_{\mathbf{X}}} I(\mathbf{X}; \mathbf{Y}) \quad (2.20)$$

where the supremum is over the set of all probability distributions on \mathbf{X} satisfying the constraints, *i.e.*,

$$\|\mathbf{X}\|^2 \leq \mathcal{E}, \quad \text{almost surely} \quad (2.21)$$

for a peak-power constraint, or

$$\mathbb{E}[\|\mathbf{X}\|^2] \leq \mathcal{E} \quad (2.22)$$

for an average-power constraint.

Specializing [7, Th. 4.2], [2, Th. 6.10] to memoryless MIMO fading, we have

$$\overline{\lim}_{\text{SNR} \uparrow \infty} \left\{ C(\text{SNR}) - \log \log \text{SNR} \right\} < \infty. \quad (2.23)$$

Note that [7, Th. 4.2], [2, 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, (2.23) also holds in the situation of a peak-power constraint.

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

$$\chi(\mathbb{H}) \triangleq \overline{\lim}_{\text{SNR} \uparrow \infty} \left\{ C(\text{SNR}) - \log \log \text{SNR} \right\}. \quad (2.24)$$

Prima facie the fading number depends on whether a peak-power constraint (2.21) or an average-power constraint (2.22) is imposed on the input. However, it will turn out that the memoryless MIMO fading number is identical for both cases.

Chapter 3

Main Results

3.1 Preliminaries

Before we can state our main result 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.

3.1.1 Escaping to Infinity

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

A sequence of input distributions 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 [3, Sec. VII.C.3], [2, Sec. 2.6].

Definition 3. Let $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ be a family of input distributions for the memoryless fading channel (2.18), where this family is parametrized by the available average power \mathcal{E} such that

$$\mathbb{E}_{Q_{\mathbf{X},\mathcal{E}}}[\|\mathbf{X}\|^2] \leq \mathcal{E}, \quad \mathcal{E} \geq 0. \quad (3.1)$$

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

$$\lim_{\mathcal{E} \uparrow \infty} Q_{\mathbf{X},\mathcal{E}}(\|\mathbf{X}\|^2 \leq \mathcal{E}_0) = 0. \quad (3.2)$$

We now have the following:

Lemma 4. Let the memoryless MIMO fading channel be given as in (2.18) 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 (3.1). Let $I(Q_{\mathbf{X},\mathcal{E}})$ denote the mutual information between input and output of channel (2.18) 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:

$$\lim_{\mathcal{E}\uparrow\infty} \frac{I(Q_{\mathbf{X},\mathcal{E}})}{\log \log \mathcal{E}} = 1. \quad (3.3)$$

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

Proof. A proof can be found in [3, Th. 8, Rem. 9], [2, Cor. 2.8]. \square

3.1.2 An Upper Bound on Channel Capacity

In [7], [2] 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 8. For a more general form, for more mathematical details, and for all proofs we refer to [7, Sec. IV], [2, Sec. 2.4].

Lemma 5. *Consider a memoryless channel with input $\mathbf{s} \in \mathbb{C}^{n_{\text{R}}}$ 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:*

$$\begin{aligned} I(\mathbf{S}; T) &\leq -h(T|\mathbf{S}) + \log \pi + \alpha \log \beta + \log \Gamma\left(\alpha, \frac{\nu}{\beta}\right) \\ &\quad + (1 - \alpha)\mathbf{E}[\log(|T|^2 + \nu)] + \frac{1}{\beta}\mathbf{E}[|T|^2] + \frac{\nu}{\beta} \end{aligned} \quad (3.4)$$

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 [7, Sec. IV], [2, Sec. 2.4]. \square

3.1.3 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

$$\mathbf{W} \stackrel{\mathcal{L}}{=} \mathbf{W} \cdot e^{i\Theta} \quad (3.5)$$

where $\Theta \sim \mathcal{U}([0, 2\pi])$ is independent of \mathbf{W} and where $\stackrel{\mathcal{L}}{=}$ 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 6. *Assume a channel as given in (2.18). Then the capacity-achieving input distribution can be assumed to be circularly symmetric, i.e., the input vector \mathbf{X} can be replaced by $\mathbf{X}e^{i\Theta}$, where $\Theta \sim \mathcal{U}([0, 2\pi])$ is independent of every other random quantity.*

Proof. A proof is given in Appendix A.1. \square

Remark 7. Note that the proof of Lemma 6 relies only on the fact that the additive noise is assumed to be circularly symmetric.

3.2 Fading Number of General Memoryless MIMO Fading Channels

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

Theorem 8. Consider the memoryless MIMO fading channel (2.18) where the random fading matrix \mathbb{H} takes value in $\mathbb{C}^{n_R \times n_T}$ and satisfies

$$h(\mathbb{H}) > -\infty \quad (3.6)$$

and

$$\mathbb{E}[\|\mathbb{H}\|_{\mathbb{F}}^2] < \infty. \quad (3.7)$$

Then, irrespective of whether a peak-power constraint (2.21) or an average-power constraint (2.22) is imposed on the input, the fading number $\chi(\mathbb{H})$ is given by

$$\chi(\mathbb{H}) = \sup_{Q_{\hat{\mathbf{X}}}} \left\{ h_{\lambda} \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) + n_R \mathbb{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \right] - \log 2 - h(\mathbb{H}\hat{\mathbf{X}} | \hat{\mathbf{X}}) \right\}. \quad (3.8)$$

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, 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 (3.8) and where R is a non-negative random variable independent of $\hat{\mathbf{X}}$ such that

$$\log R^2 \sim \mathcal{U}([\log \log \mathcal{E}, \log \mathcal{E}]). \quad (3.9)$$

Proof. A proof is given in Chapter 5. \square

Note that—even if it might not be obvious at first sight—it is not hard to show that the distribution $Q_{\hat{\mathbf{X}}}$ that achieves the supremum in (3.8) is circularly symmetric. This is in agreement with Lemma 6.

The evaluation of (3.8) 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 next derive an upper bound to the fading number that is easier to evaluate.

To that goal firstly note that for an arbitrary constant non-singular $n_R \times n_R$ matrix \mathbf{A} and an arbitrary constant non-singular $n_T \times n_T$ matrix \mathbf{B}

$$\chi(\mathbf{A}\mathbb{H}\mathbf{B}) = \chi(\mathbb{H}), \quad (3.10)$$

see [7, Lem. 4.7], [2, Lem. 6.14]. Secondly, note that for an arbitrary random unit vector $\hat{\mathbf{Y}} \in \mathbb{C}^{n_R}$

$$h_\lambda(\hat{\mathbf{Y}}) \leq \log c_{n_R} = \log \frac{2\pi^{n_R}}{\Gamma(n_R)} \quad (3.11)$$

where c_{n_R} denotes the surface area of the unit sphere in \mathbb{C}^{n_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 9. *The fading number of a memoryless MIMO fading channel as given in Theorem 8 can be upper-bounded as follows:*

$$\chi(\mathbb{H}) \leq n_R \log \pi - \log \Gamma(n_R) + \inf_{\mathbf{A}, \mathbf{B}} \sup_{\hat{\mathbf{x}}} \left\{ n_R \mathbb{E} [\log \|\mathbf{A}\mathbb{H}\mathbf{B}\hat{\mathbf{x}}\|^2] - h(\mathbf{A}\mathbb{H}\mathbf{B}\hat{\mathbf{x}}) \right\} \quad (3.12)$$

where the infimum is over all non-singular $n_R \times n_R$ complex matrices \mathbf{A} and non-singular $n_T \times n_T$ complex matrices \mathbf{B} .

Proof. Using the two observations (3.10) and (3.11) we immediately get from Theorem 8:

$$\begin{aligned} \chi(\mathbb{H}) \leq \inf_{\mathbf{A}, \mathbf{B}} \sup_{Q_{\hat{\mathbf{x}}}} \mathbb{E}_{\hat{\mathbf{x}}} \left[n_R \log \pi - \log \Gamma(n_R) \right. \\ \left. + n_R \mathbb{E} \left[\log \|\mathbf{A}\mathbb{H}\mathbf{B}\hat{\mathbf{X}}\|^2 \mid \hat{\mathbf{X}} = \hat{\mathbf{x}} \right] \right. \\ \left. - h(\mathbf{A}\mathbb{H}\mathbf{B}\hat{\mathbf{X}} \mid \hat{\mathbf{X}} = \hat{\mathbf{x}}) \right]. \end{aligned} \quad (3.13)$$

The result now follows by noting that (3.12) can always be achieved by choosing $Q_{\hat{\mathbf{x}}}$ in (3.13) to be the distribution which with probability 1 takes on the value $\hat{\mathbf{x}}$ that achieves the maximum in (3.12). \square

This upper bound is possibly tighter than the upper bound given in [7, Lem. 4.14], [2, Lem. 6.16] because of the additional infimum over \mathbf{B} .

Chapter 4

Special Cases

4.1 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 $n_T \times n_T$ matrix \mathbf{V}_t there exists an $n_R \times n_R$ constant unitary matrix \mathbf{V}_r such that

$$\mathbf{V}_r \mathbb{H} \stackrel{\mathcal{L}}{=} \mathbb{H} \mathbf{V}_t \quad (4.1)$$

where $\stackrel{\mathcal{L}}{=}$ stands for “has the same law”; and for every constant unitary $n_R \times n_R$ matrix \mathbf{V}_r there exists a constant unitary $n_T \times n_T$ matrix \mathbf{V}_t such that (4.1) holds [7, Def. 4.37], [2, 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 10. *Let \mathbb{H} be rotation-commutative in the generalized sense. Then the following two statements hold:*

- *If $\hat{\mathbf{X}} \in \mathbb{C}^{n_T}$ is an isotropically distributed random vector that is independent of \mathbb{H} , then $\mathbb{H}\hat{\mathbf{X}} \in \mathbb{C}^{n_R}$ is isotropically distributed.*
- *If $\hat{\mathbf{e}}, \hat{\mathbf{e}}' \in \mathbb{C}^{n_T}$ are two constant unit vectors, then*

$$\|\mathbb{H}\hat{\mathbf{e}}\| \stackrel{\mathcal{L}}{=} \|\mathbb{H}\hat{\mathbf{e}}'\|, \quad \|\hat{\mathbf{e}}\| = \|\hat{\mathbf{e}}'\| = 1 \quad (4.2)$$

$$h(\mathbb{H}\hat{\mathbf{e}}) = h(\mathbb{H}\hat{\mathbf{e}}'), \quad \|\hat{\mathbf{e}}\| = \|\hat{\mathbf{e}}'\| = 1. \quad (4.3)$$

Proof. For a proof see, *e.g.*, [7, Lem. 4.38], [2, Lem. 6.38]. □

From Lemma 10 it immediately follows that in the situation of rotation-commutative fading the only term in the expression of the fading number (3.8) 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 10

by the choice of an isotropic distribution for $Q_{\hat{\mathbf{X}}}$. Then according to (2.9) and (2.10)

$$h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) = \log \frac{2\pi^{n_R}}{\Gamma(n_R)}. \quad (4.4)$$

The expression of the fading number (3.8) then reduces to (1.8):

$$\chi(\mathbb{H}) = \log \frac{2\pi^{n_R}}{\Gamma(n_R)} - \log 2 + n_R \mathbb{E} [\log \|\mathbb{H}\hat{\mathbf{e}}\|^2] - h(\mathbb{H}\hat{\mathbf{e}}) \quad (4.5)$$

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

In case of a SIMO fading channel, the direction vector $\hat{\mathbf{X}}$ reduces to a phase term $e^{i\Phi}$. From Lemma 6 we know that an optimal choice of $e^{i\Phi}$ is circularly symmetric, such that (3.8) becomes

$$\chi(\mathbf{H}) = h_\lambda(\hat{\mathbf{H}}e^{i\Theta}) + n_R \mathbb{E} [\log \|\mathbf{H}\|^2] - \log 2 - h(\mathbf{H}). \quad (4.6)$$

Before we continue with the MISO case, we would like to remark that the only term in (3.8) that depends on the distribution of the phase of each component of $\hat{\mathbf{X}}$ is

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

Since we know from Lemma 6 that $\hat{\mathbf{X}}$ is circularly symmetric, we can therefore equivalently write

$$h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) = h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta} \right). \quad (4.7)$$

Turning to the MISO case now note that the distribution of

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

is identical to the distribution of $e^{i\Theta}$, independently of the distribution of \mathbf{H} and $\hat{\mathbf{X}}$. Hence,

$$h_\lambda \left(\frac{\mathbf{H}^T \hat{\mathbf{X}}}{|\mathbf{H}^T \hat{\mathbf{X}}|} e^{i\Theta} \right) = h_\lambda(e^{i\Theta}) = \log 2\pi. \quad (4.8)$$

The fading number (3.8) then becomes

$$\chi(\mathbf{H}^T) = \sup_{Q_{\hat{\mathbf{X}}}} \left\{ \log 2\pi + \mathbb{E} [\log |\mathbf{H}^T \hat{\mathbf{X}}|^2] - \log 2 - h(\mathbf{H}^T \hat{\mathbf{X}} | \hat{\mathbf{X}}) \right\} \quad (4.9)$$

$$= \sup_{Q_{\hat{\mathbf{X}}}} \mathbb{E}_{\hat{\mathbf{X}}} \left[\log \pi + \mathbb{E} [\log |\mathbf{H}^T \hat{\mathbf{x}}|^2 | \hat{\mathbf{X}} = \hat{\mathbf{x}}] - h(\mathbf{H}^T \hat{\mathbf{x}} | \hat{\mathbf{X}} = \hat{\mathbf{x}}) \right] \quad (4.10)$$

$$\leq \sup_{\hat{\mathbf{x}}} \left\{ \log \pi + \mathbb{E} [\log |\mathbf{H}^T \hat{\mathbf{x}}|^2] - h(\mathbf{H}^T \hat{\mathbf{x}}) \right\} \quad (4.11)$$

which can be achieved for a distribution of $Q_{\hat{\mathbf{X}}}$ that with probability 1 takes on the value $\hat{\mathbf{x}}$ that achieves the fading number in (4.11).

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

$$h_\lambda(e^{i\Theta}) = \log 2\pi \quad (4.12)$$

we get

$$\chi(H) = \log 2\pi + \mathbb{E}[\log |H|^2] - \log 2 - h(H) \quad (4.13)$$

$$= \log \pi + \mathbb{E}[\log |H|^2] - h(H). \quad (4.14)$$

4.2 Gaussian Fading

The evaluation of the fading number is rather difficult even for the usually simpler situation of Gaussian fading processes. However, we are able to give the exact value for some important special cases, and we will give bounds on some others.

Throughout this section we assume that the fading matrix \mathbb{H} can be written as

$$\mathbb{H} = \mathbf{D} + \tilde{\mathbb{H}} \quad (4.15)$$

where all components of $\tilde{\mathbb{H}}$ are independent of each other and zero-mean, unit-variance Gaussian distributed, and where \mathbf{D} denotes a constant line-of-sight matrix.

Note that for some constant unitary $n_R \times n_R$ matrix \mathbf{U} and some constant unitary $n_T \times n_T$ matrix \mathbf{V} the law of $\mathbf{U}\tilde{\mathbb{H}}\mathbf{V}$ is identical to the law of $\tilde{\mathbb{H}}$. Therefore, without loss of generality, we may restrict ourselves to matrices \mathbf{D} that are “diagonal”, *i.e.*, for $n_R \leq n_T$,

$$\mathbf{D} = \begin{pmatrix} \tilde{\mathbf{D}} & \mathbf{0}_{n_R \times (n_T - n_R)} \end{pmatrix} \quad (4.16)$$

or, for $n_R > n_T$,

$$\mathbf{D} = \begin{pmatrix} \tilde{\mathbf{D}} \\ \mathbf{0}_{(n_R - n_T) \times n_T} \end{pmatrix} \quad (4.17)$$

where $\tilde{\mathbf{D}}$ is a $\min\{n_R, n_T\} \times \min\{n_R, n_T\}$ diagonal matrix with the singular values of \mathbf{D} on the diagonal.

4.2.1 Scalar Line-of-Sight Matrix

We start with a scalar line-of-sight matrix, *i.e.*, we assume $\tilde{\mathbf{D}} = d\mathbf{I}$ where \mathbf{I} denotes the identity matrix.

Under these assumptions the fading number has been known already for $n_R = n_T = m$, in which case the fading matrix \mathbb{H} is rotation commutative [7], [2]:

$$\chi(\mathbb{H}) = mg_m(|d|^2) - m - \log \Gamma(m). \quad (4.18)$$

Here $g_m(\cdot)$ is a continuous, monotonically increasing, concave function defined as

$$g_m(\xi) \triangleq \begin{cases} \log(\xi) - \text{Ei}(-\xi) + \sum_{j=1}^{m-1} (-1)^j \left[e^{-\xi} (j-1)! \right. \\ \left. - \frac{(m-1)!}{j(m-1-j)!} \right] \left(\frac{1}{\xi} \right)^j, & \xi > 0 \\ \psi(m), & \xi = 0 \end{cases} \quad (4.19)$$

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

$$\text{Ei}(-x) \triangleq - \int_x^\infty \frac{e^{-t}}{t} dt, \quad x > 0 \quad (4.20)$$

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

$$\psi(m) \triangleq -\gamma + \sum_{j=1}^{m-1} \frac{1}{j} \quad (4.21)$$

with $\gamma \approx 0.577$ denoting Euler's constant. This function $g_m(\cdot)$ is the expected value of the logarithm of a non-central chi-square random variable, *i.e.*, for some Gaussian random variables $\{U_j\}_{j=1}^m \text{ IID } \sim \mathcal{N}_{\mathbb{C}}(0, 1)$ and for some complex constants $\{\mu_j\}_{j=1}^m$ we have

$$\mathbb{E} \left[\log \left(\sum_{j=1}^m |U_j + \mu_j|^2 \right) \right] = g_m(s^2) \quad (4.22)$$

where

$$s^2 \triangleq \sum_{j=1}^m |\mu_j|^2 \quad (4.23)$$

(see [25], [7, Lem. 10.1], [2, Lem. A.6] for more details and a proof).

We would like to emphasize that $g_m(\xi)$ is continuous for all $\xi \geq 0$, *i.e.*, in particular

$$\lim_{\xi \downarrow 0} \left\{ \log(\xi) - \text{Ei}(-\xi) + \sum_{j=1}^{m-1} (-1)^j \left[e^{-\xi} (j-1)! - \frac{(m-1)!}{j(m-1-j)!} \right] \left(\frac{1}{\xi} \right)^j \right\} = \psi(m) \quad (4.24)$$

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

$$\log \xi - \text{Ei}(-\xi) \leq g_m(\xi) \leq \log(m + \xi). \quad (4.25)$$

A derivation of (4.25) can be found in Appendix A.2.

We now consider the case where $n_{\text{R}} \leq n_{\text{T}}$:

Corollary 11. *Assume $n_{\text{R}} \leq n_{\text{T}}$ and a Gaussian fading matrix as given in (4.15). Let the line-of-sight matrix \mathbf{D} be given as*

$$\mathbf{D} = d \begin{pmatrix} \mathbf{I}_{n_{\text{R}}} & \mathbf{0}_{n_{\text{R}} \times (n_{\text{T}} - n_{\text{R}})} \end{pmatrix}. \quad (4.26)$$

Then

$$\chi(\mathbb{H}) = n_{\text{R}} g_{n_{\text{R}}}(|d|^2) - n_{\text{R}} - \log \Gamma(n_{\text{R}}) \quad (4.27)$$

where $g_m(\cdot)$ is defined in (4.19).

Proof. We write for the unit vector $\hat{\mathbf{X}}$

$$\hat{\mathbf{X}} = \begin{pmatrix} \mathbf{\Xi} \\ \mathbf{\Xi}' \end{pmatrix} \quad (4.28)$$

where $\mathbf{\Xi} \in \mathbb{C}^{n_{\text{R}}}$ and $\mathbf{\Xi}' \in \mathbb{C}^{n_{\text{T}} - n_{\text{R}}}$. Then

$$\mathbb{H} \hat{\mathbf{X}} = \mathbf{D} \hat{\mathbf{X}} + \tilde{\mathbb{H}} \hat{\mathbf{X}} = d \mathbf{\Xi} + \tilde{\mathbf{H}} \quad (4.29)$$

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

$$h(\mathbb{H}\hat{\mathbf{X}} \mid \hat{\mathbf{X}}) = h(\tilde{\mathbf{H}}) = n_{\text{R}} \log \pi e; \quad (4.30)$$

$$n_{\text{R}} \mathbb{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \right] = n_{\text{R}} g_{n_{\text{R}}}(|d|^2 \|\boldsymbol{\Xi}\|^2) \leq n_{\text{R}} g_{n_{\text{R}}}(|d|^2); \quad (4.31)$$

$$h_{\lambda} \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) \leq \log \frac{2\pi^{n_{\text{R}}}}{\Gamma(n_{\text{R}})}. \quad (4.32)$$

Here, the equality in (4.31) follows from the fact that $\|d\boldsymbol{\Xi} + \tilde{\mathbf{H}}\|^2$ is non-central chi-square distributed and from (4.22); the inequality in (4.31) follows from the monotonicity of $g_m(\cdot)$ and is tight if $\|\boldsymbol{\Xi}\| = 1$, *i.e.*, $\boldsymbol{\Xi}' = \mathbf{0}$; and the inequality in (4.32) follows from (2.9) and (2.10) and is tight if $\boldsymbol{\Xi}$ is uniformly distributed on the unit sphere in $\mathbb{C}^{n_{\text{R}}}$ so that $\mathbb{H}\hat{\mathbf{X}}$ is isotropically distributed. The result now follows from Theorem 8. \square

The case $n_{\text{R}} > n_{\text{T}}$ is more difficult since then (4.32) is in general not tight. We will only state an upper bound:

Proposition 12. *Assume $n_{\text{R}} > n_{\text{T}}$ and a Gaussian fading matrix as given in (4.15). Let the line-of-sight matrix \mathbf{D} be given as*

$$\mathbf{D} = d \begin{pmatrix} \mathbf{I}_{n_{\text{T}}} \\ \mathbf{0}_{(n_{\text{R}}-n_{\text{T}}) \times n_{\text{T}}} \end{pmatrix}. \quad (4.33)$$

Then

$$\chi(\mathbb{H}) \leq n_{\text{T}} \log \left(1 + \frac{|d|^2}{n_{\text{T}}} \right) + n_{\text{R}} \log n_{\text{R}} - n_{\text{R}} - \log \Gamma(n_{\text{R}}). \quad (4.34)$$

Proof. This result is a special case of Proposition 14 below (see also [7, (128)], [2, (6.224)]). \square

4.2.2 General Line-of-Sight Matrix

Next we assume Gaussian fading as defined in (4.15) with a general line-of-sight matrix \mathbf{D} having singular values $d_1, \dots, d_{\min\{n_{\text{R}}, n_{\text{T}}\}}$. Hence, $\tilde{\mathbf{D}}$, defined in (4.16) and (4.17), is given as

$$\tilde{\mathbf{D}} = \text{diag}(d_1, \dots, d_{\min\{n_{\text{R}}, n_{\text{T}}\}}) \quad (4.35)$$

where $|d_1| \geq |d_2| \geq \dots \geq |d_{\min\{n_{\text{R}}, n_{\text{T}}\}}| > 0$.

We again start with the case $n_{\text{R}} \leq n_{\text{T}}$.

Corollary 13. *Assume $n_{\text{R}} \leq n_{\text{T}}$ and a Gaussian fading matrix as given in (4.15). Let the line-of-sight matrix \mathbf{D} have singular values $d_1, \dots, d_{n_{\text{R}}}$, where $|d_1| \geq |d_2| \geq \dots \geq |d_{n_{\text{R}}}| > 0$. Then*

$$\chi(\mathbb{H}) \leq n_{\text{R}} g_{n_{\text{R}}}(\|\mathbf{D}\|^2) - n_{\text{R}} - \log \Gamma(n_{\text{R}}) \quad (4.36)$$

where $g_m(\cdot)$ is given in (4.19) and where $\|\mathbf{D}\|^2 = |d_1|^2$.

Proof. A proof is given in Appendix A.3. \square

The situation $n_{\text{R}} > n_{\text{T}}$ is again more complicated. We include this case in a new upper bound based on (3.12) which holds independently of the particular relation between n_{R} and n_{T} :

Proposition 14. *Assume a Gaussian fading matrix as given in (4.15) and let the line-of-sight matrix \mathbf{D} be general with singular values $d_1, \dots, d_{\min\{n_R, n_T\}}$, where $|d_1| \geq |d_2| \geq \dots \geq |d_{\min\{n_R, n_T\}}| > 0$. Then the fading number is upper-bounded as follows:*

$$\chi(\mathbb{H}) \leq \min\{n_R, n_T\} \log \frac{\delta^2}{\min\{n_R, n_T\}} + n_R \log n_R - n_R - \log \Gamma(n_R) \quad (4.37)$$

where

$$\delta^2 \triangleq (|d_1|^2 \cdots |d_{\min\{n_R, n_T\}}|^2)^{1/\min\{n_R, n_T\}} \cdot \left(1 + \frac{1}{|d_1|^2} + \cdots + \frac{1}{|d_{\min\{n_R, n_T\}}|^2} \right). \quad (4.38)$$

Proof. A proof is given in Appendix A.4. □

Chapter 5

Proof of the Main Result

The proof of Theorem 8 consists of two parts: firstly we derive an upper bound to the fading number assuming an average-power constraint (2.22) on the input. The key ingredients here are the preliminary results from Section 3.1.

In a second part we then show that this upper bound can actually be achieved by an input that satisfies the peak-power constraint (2.21). Since a peak-power constraint is more restrictive than the corresponding average-power constraint, the theorem follows.

Because the proof is rather technical, we will give a short overview to clarify the main ideas.

The upper bound relies strongly on Lemma 4 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

$$I(\mathbf{X}; \mathbf{Y}) \lesssim I(\mathbf{X}; \mathbb{H}\mathbf{X}). \quad (5.1)$$

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:

$$I(\mathbf{X}; \mathbb{H}\mathbf{X}) = I(\mathbf{X}; \|\mathbb{H}\mathbf{X}\|) + I\left(\mathbf{X}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \middle| \|\mathbb{H}\mathbf{X}\|\right). \quad (5.2)$$

For the first term—which is related to MISO fading—we then use the bounding technique of Lemma 5.

Because Lemma 4 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

$$\mathbf{X} = R \cdot \hat{\mathbf{X}} \quad (5.3)$$

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:

$$I(\mathbf{X}; \mathbf{Y}) = I(R; \mathbf{Y} | \hat{\mathbf{X}}) + I(\hat{\mathbf{X}}; \mathbf{Y}). \quad (5.4)$$

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.

5.1 Derivation of an Upper Bound

In the following we will use the notation $R \triangleq \|\mathbf{X}\|$ to denote the magnitude of the input vector \mathbf{X} , *i.e.*, we have $\mathbf{X} = R \cdot \hat{\mathbf{X}}$. Note that in this section we are not allowed to assume that R is independent of $\hat{\mathbf{X}}$.

From Lemma 4 we know that the capacity-achieving input distribution must escape to infinity. Hence, we fix an arbitrary finite $\mathcal{E}_0 \geq 0$ and define an indicator random variable as follows:

$$E \triangleq \begin{cases} 1 & \text{if } \|\mathbf{X}\|^2 \geq \mathcal{E}_0, \\ 0 & \text{otherwise.} \end{cases} \quad (5.5)$$

Let

$$p \triangleq \Pr[E = 1] = \Pr[\|\mathbf{X}\|^2 \geq \mathcal{E}_0] \quad (5.6)$$

where we know from Lemma 4 that

$$\lim_{\mathcal{E}_0 \uparrow \infty} p = 1. \quad (5.7)$$

We now bound as follows:

$$I(\mathbf{X}; \mathbf{Y}) \leq I(\mathbf{X}, E; \mathbf{Y}) \quad (5.8)$$

$$= I(E; \mathbf{Y}) + I(\mathbf{X}; \mathbf{Y} | E) \quad (5.9)$$

$$= H(E) - H(E | \mathbf{Y}) + I(\mathbf{X}; \mathbf{Y} | E) \quad (5.10)$$

$$\leq H(E) + I(\mathbf{X}; \mathbf{Y} | E) \quad (5.11)$$

$$= H_b(p) + pI(\mathbf{X}; \mathbf{Y} | E = 1) + (1 - p)I(\mathbf{X}; \mathbf{Y} | E = 0) \quad (5.12)$$

$$\leq H_b(p) + I(\mathbf{X}; \mathbf{Y} | E = 1) + (1 - p)\mathcal{C}(\mathcal{E}_0) \quad (5.13)$$

where

$$H_b(\xi) \triangleq -\xi \log \xi - (1 - \xi) \log(1 - \xi) \quad (5.14)$$

is the binary entropy function. Here, (5.8) follows from adding an additional random variable to mutual information; the subsequent two equalities follow from the chain rule and from the definition of mutual information (notice that we use entropy and not differential entropy because E is a binary random variable); in the subsequent inequality we rely on the non-negativity of entropy; and the last inequality follows from bounding $p \leq 1$ and from upper-bounding the mutual information term by the capacity \mathcal{C} for the available power which—conditional on $E = 0$ —is \mathcal{E}_0 .

We remark that even though $\mathcal{C}(\mathcal{E}_0)$ is unknown, we know that it is finite and independent of \mathcal{E} so that from (5.7) we have

$$\lim_{\mathcal{E}_0 \uparrow \infty} \{H_b(p) + (1 - p)\mathcal{C}(\mathcal{E}_0)\} = 0. \quad (5.15)$$

We continue with the second term of (5.13) as follows:

$$I(\mathbf{X}; \mathbf{Y} | E = 1) = I(\mathbf{X}; \mathbb{H}\mathbf{X} + \mathbf{Z} | E = 1) \quad (5.16)$$

$$\leq I(\mathbf{X}; \mathbb{H}\mathbf{X} + \mathbf{Z}, \mathbf{Z} | E = 1) \quad (5.17)$$

$$= I(\mathbf{X}; \mathbb{H}\mathbf{X}, \mathbf{Z} | E = 1) \quad (5.18)$$

$$= I(\mathbf{X}; \mathbb{H}\mathbf{X} | E = 1) + I(\mathbf{X}; \mathbf{Z} | \mathbb{H}\mathbf{X}, E = 1) \quad (5.19)$$

$$= I(\mathbf{X}; \mathbb{H}\mathbf{X} \mid E = 1) \quad (5.20)$$

$$= I\left(\mathbf{X}; \|\mathbb{H}\mathbf{X}\|, \frac{\mathbb{H}\mathbf{X}}{\|\mathbb{H}\mathbf{X}\|} \mid E = 1\right) \quad (5.21)$$

$$= I\left(\mathbf{X}; \|\mathbb{H}\mathbf{X}\|, \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid E = 1\right) \quad (5.22)$$

$$= I(\mathbf{X}; \|\mathbb{H}\mathbf{X}\| \mid E = 1) + I\left(\mathbf{X}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid \|\mathbb{H}\mathbf{X}\|, E = 1\right) \quad (5.23)$$

$$\leq I(\mathbf{X}; \|\mathbb{H}\mathbf{X}\|, e^{i\Theta} \mid E = 1) + I\left(\mathbf{X}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid \|\mathbb{H}\mathbf{X}\|, E = 1\right) \quad (5.24)$$

$$= I(\mathbf{X}; \|\mathbb{H}\mathbf{X}\|e^{i\Theta} \mid E = 1) + I\left(\mathbf{X}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid \|\mathbb{H}\mathbf{X}\|, E = 1\right). \quad (5.25)$$

Here, (5.17) follows from adding an additional random vector \mathbf{Z} to the argument of the mutual information; the subsequent equality from subtracting the known vector \mathbf{Z} from \mathbf{Y} ; the subsequent two equalities follow from the chain rule and the independence between the noise and all other random quantities; then we split $\mathbb{H}\mathbf{X}$ into magnitude and direction vector and use the chain rule again; (5.24) follows from adding a random variable to mutual information: we introduce $e^{i\Theta}$ that is independent of all the other random quantities and that is uniformly distributed on the complex unit circle; and the last equality holds because from $\|\mathbb{H}\mathbf{X}\|e^{i\Theta}$ we can easily get back $\|\mathbb{H}\mathbf{X}\|$ and $e^{i\Theta}$.

We next apply Lemma 5 to the first term in (5.25), *i.e.*, we choose $\mathbf{S} = \mathbf{X}$ and $T = \|\mathbb{H}\mathbf{X}\|e^{i\Theta}$. Note that we need to condition everything on the event $E = 1$:

$$\begin{aligned} & I(\mathbf{X}; \|\mathbb{H}\mathbf{X}\|e^{i\Theta} \mid E = 1) \\ & \leq -h(\|\mathbb{H}\mathbf{X}\|e^{i\Theta} \mid \mathbf{X}, E = 1) + \log \pi + \alpha \log \beta \\ & \quad + \log \Gamma\left(\alpha, \frac{\nu}{\beta}\right) + (1 - \alpha)\mathbb{E}[\log(\|\mathbb{H}\mathbf{X}\|^2 + \nu) \mid E = 1] \\ & \quad + \frac{1}{\beta}\mathbb{E}[\|\mathbb{H}\mathbf{X}\|^2 \mid E = 1] + \frac{\nu}{\beta} \end{aligned} \quad (5.26)$$

where $\alpha, \beta > 0$, and $\nu \geq 0$ can be chosen freely, but must not depend on \mathbf{X} .

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

$$\begin{aligned} & h(\|\mathbb{H}\mathbf{X}\|e^{i\Theta} \mid \mathbf{X} = \mathbf{x}, E = 1) \\ & = h_\lambda(e^{i\Theta} \mid \mathbf{X} = \mathbf{x}, E = 1) + h(\|\mathbb{H}\mathbf{X}\| \mid e^{i\Theta}, \mathbf{X} = \mathbf{x}, E = 1) \\ & \quad + \mathbb{E}[\log \|\mathbb{H}\mathbf{X}\| \mid \mathbf{X} = \mathbf{x}, E = 1] \end{aligned} \quad (5.27)$$

$$\begin{aligned} & = \log 2\pi + h(\|\mathbb{H}\mathbf{X}\| \mid \mathbf{X} = \mathbf{x}, E = 1) \\ & \quad + \mathbb{E}[\log \|\mathbb{H}\mathbf{X}\| \mid \mathbf{X} = \mathbf{x}, E = 1] \end{aligned} \quad (5.28)$$

where we have used that $e^{i\Theta}$ is independent of all other random quantities and uniformly distributed on the unit circle. Taking the expectation over \mathbf{X} conditional

on $E = 1$ we then yield

$$\begin{aligned} & h(\|\mathbb{H}\mathbf{X}\|e^{i\Theta} \mid \mathbf{X}, E = 1) \\ &= \log 2\pi + h(\|\mathbb{H}\mathbf{X}\| \mid \mathbf{X}, E = 1) + \mathbb{E}[\log \|\mathbb{H}\mathbf{X}\| \mid E = 1] \end{aligned} \quad (5.29)$$

$$\begin{aligned} &= \log 2\pi + h(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \mid \hat{\mathbf{X}}, R, E = 1) \\ &+ \mathbb{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \mid E = 1 \right] \end{aligned} \quad (5.30)$$

$$\begin{aligned} &= \log 2\pi + h(\|\mathbb{H}\hat{\mathbf{X}}\| \mid \hat{\mathbf{X}}, R, E = 1) + \mathbb{E}[\log R \mid E = 1] \\ &+ \mathbb{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\| \mid E = 1 \right] + \mathbb{E}[\log R \mid E = 1] \end{aligned} \quad (5.31)$$

$$\begin{aligned} &= \log 2\pi + h(\|\mathbb{H}\hat{\mathbf{X}}\| \mid \hat{\mathbf{X}}, E = 1) + 2\mathbb{E}[\log R \mid E = 1] \\ &+ \mathbb{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\| \mid E = 1 \right] \end{aligned} \quad (5.32)$$

where the second equality follows from the definition of $R \triangleq \|\mathbf{X}\|$; where the third equality follows from the scaling property of entropy with a *real* argument; and where the last equality follows because given $\hat{\mathbf{X}}$, $\|\mathbb{H}\hat{\mathbf{X}}\|$ is independent of R .

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

$$\epsilon_\nu \triangleq \sup_{\|\mathbf{x}\|^2 \geq \mathcal{E}_0} \left\{ \mathbb{E}[\log(\|\mathbb{H}\mathbf{x}\|^2 + \nu)] - \mathbb{E}[\log \|\mathbb{H}\mathbf{x}\|^2] \right\} \quad (5.33)$$

such that

$$\begin{aligned} & (1 - \alpha)\mathbb{E}[\log(\|\mathbb{H}\mathbf{X}\|^2 + \nu) \mid E = 1] \\ &= (1 - \alpha)\mathbb{E}[\log(\|\mathbb{H}\mathbf{X}\|^2 + \nu) - \log \|\mathbb{H}\mathbf{X}\|^2 \mid E = 1] \\ &+ (1 - \alpha)\mathbb{E}[\log \|\mathbb{H}\mathbf{X}\|^2 \mid E = 1] \end{aligned} \quad (5.34)$$

$$\begin{aligned} &\leq (1 - \alpha) \sup_{\|\mathbf{x}\|^2 \geq \mathcal{E}_0} \left\{ \mathbb{E}[\log(\|\mathbb{H}\mathbf{x}\|^2 + \nu)] - \mathbb{E}[\log \|\mathbb{H}\mathbf{x}\|^2] \right\} \\ &+ (1 - \alpha)\mathbb{E}[\log \|\mathbb{H}\mathbf{X}\|^2 \mid E = 1] \end{aligned} \quad (5.35)$$

$$= (1 - \alpha)\epsilon_\nu + (1 - \alpha)\mathbb{E}[\log \|\mathbb{H}\mathbf{X}\|^2 \mid E = 1] \quad (5.36)$$

$$\leq \epsilon_\nu + (1 - \alpha)\mathbb{E}[\log \|\mathbb{H}\mathbf{X}\|^2 \mid E = 1]. \quad (5.37)$$

Note that in the first inequality we have made use of the fact that $E = 1$, *i.e.*, that $\|\mathbf{X}\|^2 \geq \mathcal{E}_0$.

Finally, we bound

$$\frac{1}{\beta}\mathbb{E}[\|\mathbb{H}\mathbf{X}\|^2 \mid E = 1] = \frac{1}{\beta}\mathbb{E}[\|\mathbb{H}\hat{\mathbf{X}}\|^2 \cdot R^2 \mid E = 1] \quad (5.38)$$

$$\leq \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbb{E}[\|\mathbb{H}\hat{\mathbf{x}}\|^2 \cdot R^2 \mid E = 1] \quad (5.39)$$

$$= \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbb{E}[\|\mathbb{H}\hat{\mathbf{x}}\|^2] \cdot \mathbb{E}[R^2 \mid E = 1] \quad (5.40)$$

$$\leq \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbb{E}[\|\mathbb{H}\hat{\mathbf{x}}\|^2] \cdot \frac{\mathcal{E}}{p} \quad (5.41)$$

where we have used the fact that R needs to satisfy the average-power constraint (2.22) to get the following bound:

$$\mathcal{E} \geq \mathbb{E}[R^2] \quad (5.42)$$

$$= p\mathbb{E}[R^2 \mid E = 1] + (1 - p)\mathbb{E}[R^2 \mid E = 0] \quad (5.43)$$

$$\geq p\mathbb{E}[R^2 \mid E = 1]. \quad (5.44)$$

Plugging (5.32), (5.37), and (5.41) into (5.26) we yield

$$\begin{aligned}
& I(\mathbf{X}; \|\mathbb{H}\mathbf{X}\| e^{i\Theta} \mid E = 1) \\
& \leq -\log 2 - h(\|\mathbb{H}\hat{\mathbf{X}}\| \mid \hat{\mathbf{X}}, E = 1) - 2\mathbb{E}[\log R \mid E = 1] \\
& \quad - \mathbb{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\| \mid E = 1 \right] + \alpha \log \beta + \log \Gamma \left(\alpha, \frac{\nu}{\beta} \right) \\
& \quad + (1 - \alpha) \mathbb{E} \left[\log \|\mathbb{H}\mathbf{X}\|^2 \mid E = 1 \right] + \epsilon_\nu \\
& \quad + \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbb{E} \left[\|\mathbb{H}\hat{\mathbf{x}}\|^2 \right] \frac{\mathcal{E}}{p} + \frac{\nu}{\beta}. \tag{5.45}
\end{aligned}$$

Next we continue with the second term in (5.25):

$$\begin{aligned}
& I \left(\mathbf{X}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid \|\mathbb{H}\mathbf{X}\|, E = 1 \right) \\
& = h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid \|\mathbb{H}\mathbf{X}\|, E = 1 \right) \\
& \quad - h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R, \hat{\mathbf{X}}, R, E = 1 \right) \tag{5.46}
\end{aligned}$$

$$\begin{aligned}
& = h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid \|\mathbb{H}\mathbf{X}\|, E = 1 \right) \\
& \quad - h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid \|\mathbb{H}\hat{\mathbf{X}}\|, \hat{\mathbf{X}}, R, E = 1 \right) \tag{5.47}
\end{aligned}$$

$$\begin{aligned}
& \leq h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid E = 1 \right) \\
& \quad - h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid \|\mathbb{H}\hat{\mathbf{X}}\|, \hat{\mathbf{X}}, E = 1 \right). \tag{5.48}
\end{aligned}$$

Here, the last inequality follows because conditioning cannot increase entropy and because given $\hat{\mathbf{X}}$ and $\|\mathbb{H}\hat{\mathbf{X}}\|$, the term $\mathbb{H}\hat{\mathbf{X}}/\|\mathbb{H}\hat{\mathbf{X}}\|$ does not depend on R .

Hence, using (5.48), (5.45), and (5.25) in (5.13) we get

$$\begin{aligned}
& I(\mathbf{X}; \mathbf{Y}) \\
& \leq H_b(p) + (1 - p)C(\mathcal{E}_0) - \log 2 - h(\|\mathbb{H}\hat{\mathbf{X}}\| \mid \hat{\mathbf{X}}, E = 1) \\
& \quad - 2\mathbb{E}[\log R \mid E = 1] - \mathbb{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\| \mid E = 1 \right] + \alpha \log \beta \\
& \quad + \log \Gamma \left(\alpha, \frac{\nu}{\beta} \right) + (1 - \alpha) \mathbb{E} \left[\log \|\mathbb{H}\mathbf{X}\|^2 \mid E = 1 \right] \\
& \quad + \epsilon_\nu + \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbb{E} \left[\|\mathbb{H}\hat{\mathbf{x}}\|^2 \right] \frac{\mathcal{E}}{p} + \frac{\nu}{\beta} + h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid E = 1 \right) \\
& \quad - h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid \|\mathbb{H}\hat{\mathbf{X}}\|, \hat{\mathbf{X}}, E = 1 \right) \tag{5.49} \\
& = H_b(p) + (1 - p)C(\mathcal{E}_0) - \log 2 - h(\mathbb{H}\hat{\mathbf{X}} \mid \hat{\mathbf{X}}, E = 1) \\
& \quad + (2n_R - 1) \mathbb{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\| \mid E = 1 \right] - 2\mathbb{E}[\log R \mid E = 1] \\
& \quad - \mathbb{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\| \mid E = 1 \right] + \alpha \log \beta + \log \Gamma \left(\alpha, \frac{\nu}{\beta} \right)
\end{aligned}$$

$$\begin{aligned}
& + 2\mathbb{E}[\log \|\mathbb{H}\mathbf{X}\| \mid E = 1] - \alpha\mathbb{E}[\log \|\mathbb{H}\mathbf{X}\|^2 \mid E = 1] + \epsilon_\nu \\
& + \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbb{E}[\|\mathbb{H}\hat{\mathbf{x}}\|^2] \frac{\mathcal{E}}{p} + \frac{\nu}{\beta} + h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid E = 1 \right) \tag{5.50}
\end{aligned}$$

$$\begin{aligned}
& = h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid E = 1 \right) - h(\mathbb{H}\hat{\mathbf{X}} \mid \hat{\mathbf{X}}, E = 1) \\
& + n_R \mathbb{E}[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \mid E = 1] - \log 2 + \log \Gamma \left(\alpha, \frac{\nu}{\beta} \right) \\
& + \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbb{E}[\|\mathbb{H}\hat{\mathbf{x}}\|^2] \frac{\mathcal{E}}{p} + \frac{\nu}{\beta} + \epsilon_\nu + H_b(p) + (1-p)\mathcal{C}(\mathcal{E}_0) \\
& + \alpha \left(\log \beta - \mathbb{E}[\log \|\mathbb{H}\mathbf{X}\|^2 \mid E = 1] \right) \tag{5.51}
\end{aligned}$$

$$\begin{aligned}
& \leq h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \mid E = 1 \right) - h(\mathbb{H}\hat{\mathbf{X}} \mid \hat{\mathbf{X}}, E = 1) \\
& + n_R \mathbb{E}[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \mid E = 1] - \log 2 + \log \Gamma \left(\alpha, \frac{\nu}{\beta} \right) \\
& + \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbb{E}[\|\mathbb{H}\hat{\mathbf{x}}\|^2] \frac{\mathcal{E}}{p} + \frac{\nu}{\beta} + \epsilon_\nu + H_b(p) + (1-p)\mathcal{C}(\mathcal{E}_0) \\
& + \alpha \left(\log \beta - \log \mathcal{E}_0 - \xi \right). \tag{5.52}
\end{aligned}$$

Here, (5.50) follows again from a conditional version of Lemma 2 similar to (5.27)–(5.32) which allows us to combine the fourth and the last term in (5.49); in the subsequent equality we arithmetically rearrange the terms; and the final inequality follows from the following bound:

$$\mathbb{E}[\log \|\mathbb{H}\mathbf{X}\|^2 \mid E = 1] \geq \inf_{\|\mathbf{x}\|^2 \geq \mathcal{E}_0} \mathbb{E}[\log \|\mathbb{H}\mathbf{x}\|^2] \tag{5.53}$$

$$= \log \mathcal{E}_0 + \inf_{\hat{\mathbf{x}}} \mathbb{E}[\log \|\mathbb{H}\hat{\mathbf{x}}\|^2] \tag{5.54}$$

$$\triangleq \log \mathcal{E}_0 + \xi \tag{5.55}$$

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

$$-\infty < \xi < \infty \tag{5.56}$$

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

Note that (5.52) does not depend on the distribution of R anymore, but only on $\hat{\mathbf{X}}$! Hence, we can get an upper bound on capacity by taking the supremum over all possible distributions $Q_{\hat{\mathbf{x}}}$. This then gives us the following upper bound on the fading number:

$$\chi(\mathbb{H}) = \overline{\lim}_{\mathcal{E} \uparrow \infty} \left\{ \mathcal{C}(\mathcal{E}) - \log \left(1 + \log \left(1 + \frac{\mathcal{E}}{\sigma^2} \right) \right) \right\} \tag{5.57}$$

$$= \overline{\lim}_{\mathcal{E} \uparrow \infty} \left\{ \sup_{Q_{\mathbf{x}, \mathcal{E}}} I(\mathbf{X}; \mathbf{Y}) - \log \left(1 + \log \left(1 + \frac{\mathcal{E}}{\sigma^2} \right) \right) \right\} \tag{5.58}$$

$$\leq \overline{\lim}_{\mathcal{E} \uparrow \infty} \left\{ \sup_{Q_{\hat{\mathbf{x}}}} \left\{ h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h(\mathbb{H}\hat{\mathbf{X}} \mid \hat{\mathbf{X}}) - \log 2 \right. \right.$$

$$\begin{aligned}
& + n_{\text{R}} \mathbf{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \right] + \log \Gamma \left(\alpha, \frac{\nu}{\beta} \right) \\
& + \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbf{E} \left[\|\mathbb{H}\hat{\mathbf{x}}\|^2 \right] \frac{\mathcal{E}}{p} + \frac{\nu}{\beta} + \epsilon_{\nu} + H_{\text{b}}(p) \\
& + (1-p)\mathbf{C}(\mathcal{E}_0) + \alpha \left(\log \beta - \log \mathcal{E}_0 - \xi \right) \Bigg\} \\
& - \log \left(1 + \log \left(1 + \frac{\mathcal{E}}{\sigma^2} \right) \right) \Bigg\} \tag{5.59}
\end{aligned}$$

$$\begin{aligned}
& = \overline{\lim}_{\mathcal{E} \uparrow \infty} \left\{ \sup_{Q_{\hat{\mathbf{x}}}} \left\{ h_{\lambda} \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h(\mathbb{H}\hat{\mathbf{X}} | \hat{\mathbf{X}}) - \log 2 \right. \right. \\
& \quad \left. \left. + n_{\text{R}} \mathbf{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \right] \right\} + \log \Gamma \left(\alpha, \frac{\nu}{\beta} \right) \right. \\
& \quad \left. + \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbf{E} \left[\|\mathbb{H}\hat{\mathbf{x}}\|^2 \right] \frac{\mathcal{E}}{p} + \frac{\nu}{\beta} + \epsilon_{\nu} + H_{\text{b}}(p) \right. \\
& \quad \left. + (1-p)\mathbf{C}(\mathcal{E}_0) + \alpha \left(\log \beta - \log \mathcal{E}_0 - \xi \right) \right. \\
& \quad \left. - \log \left(1 + \log \left(1 + \frac{\mathcal{E}}{\sigma^2} \right) \right) \right\} \tag{5.60}
\end{aligned}$$

$$\begin{aligned}
& = \sup_{Q_{\hat{\mathbf{x}}}} \left\{ h_{\lambda} \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h(\mathbb{H}\hat{\mathbf{X}} | \hat{\mathbf{X}}) - \log 2 \right. \\
& \quad \left. + n_{\text{R}} \mathbf{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \right] \right\} \\
& + \overline{\lim}_{\mathcal{E} \uparrow \infty} \left\{ \log \Gamma \left(\alpha, \frac{\nu}{\beta} \right) - \log \frac{1}{\alpha} \right. \\
& \quad \left. + \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbf{E} \left[\|\mathbb{H}\hat{\mathbf{x}}\|^2 \right] \frac{\mathcal{E}}{p} + \frac{\nu}{\beta} + \epsilon_{\nu} + H_{\text{b}}(p) \right. \\
& \quad \left. + (1-p)\mathbf{C}(\mathcal{E}_0) + \alpha \left(\log \beta - \log \mathcal{E}_0 - \xi \right) \right. \\
& \quad \left. + \log \frac{1}{\alpha} - \log \left(1 + \log \left(1 + \frac{\mathcal{E}}{\sigma^2} \right) \right) \right\} \tag{5.61}
\end{aligned}$$

$$\begin{aligned}
& = \sup_{Q_{\hat{\mathbf{x}}}} \left\{ h_{\lambda} \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h(\mathbb{H}\hat{\mathbf{X}} | \hat{\mathbf{X}}) + n_{\text{R}} \mathbf{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \right] \right. \\
& \quad \left. - \log 2 \right\} + \log (1 - e^{-\nu}) + \nu + \epsilon_{\nu} - \log \nu. \tag{5.62}
\end{aligned}$$

Here the first two equalities follows from the definition of the fading number (2.24); the subsequent inequality from (5.52); (5.60) follows because the parameters α , β , and ν must not depend on the input distribution $Q_{\hat{\mathbf{x}}}$ (however, note that we are allowed to let them depend on \mathcal{E}); the subsequent equality follows since the first four terms do not depend on \mathcal{E} ; and in the last equality we have used (5.15) and made the following choices on the free parameters α and β :

$$\alpha \triangleq \alpha(\mathcal{E}) = \frac{\nu}{\log \mathcal{E} + \log \sup_{\hat{\mathbf{x}}} \mathbf{E} [\|\mathbb{H}\hat{\mathbf{x}}\|^2]} \tag{5.63}$$

$$\beta \triangleq \beta(\mathcal{E}) = \frac{1}{\alpha(\mathcal{E})} e^{\nu/\alpha} \quad (5.64)$$

for some constant $\nu \geq 0$. For this choice note that

$$\overline{\lim}_{\mathcal{E} \uparrow \infty} \left\{ \log \Gamma \left(\alpha, \frac{\nu}{\beta} \right) - \log \frac{1}{\alpha} \right\} = \log(1 - e^{-\nu}); \quad (5.65)$$

$$\overline{\lim}_{\mathcal{E} \uparrow \infty} \alpha \left(\log \beta - \log \mathcal{E}_0 - \xi \right) = \nu; \quad (5.66)$$

$$\overline{\lim}_{\mathcal{E} \uparrow \infty} \left\{ \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathbb{E} [\|\mathbb{H}\hat{\mathbf{x}}\|^2] \frac{\mathcal{E}}{p} + \frac{\nu}{\beta} \right\} = 0; \quad (5.67)$$

$$\overline{\lim}_{\mathcal{E} \uparrow \infty} \left\{ \log \frac{1}{\alpha} - \log \left(1 + \log \left(1 + \frac{\mathcal{E}}{\sigma^2} \right) \right) \right\} = -\log \nu. \quad (5.68)$$

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

To finish the derivation of the upper bound, we let ν go to zero. Note that $\epsilon_\nu \rightarrow 0$ as $\nu \downarrow 0$ as can be seen from (5.33). Note further that

$$\lim_{\nu \downarrow 0} \{ \log(1 - e^{-\nu}) - \log \nu \} = 0. \quad (5.69)$$

Therefore, we get

$$\chi(\mathbb{H}) \leq \sup_{Q_{\hat{\mathbf{x}}}} \left\{ h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h(\mathbb{H}\hat{\mathbf{X}} | \hat{\mathbf{X}}) + n_{\mathbb{R}} \mathbb{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \right] - \log 2 \right\}. \quad (5.70)$$

5.2 Derivation of a Lower Bound

To derive a lower bound on capacity (or the fading number, respectively) we choose a specific input distribution. Let \mathbf{X} be of the form

$$\mathbf{X} = R \cdot \hat{\mathbf{X}}. \quad (5.71)$$

Here $\hat{\mathbf{X}} \in \mathbb{C}^{n_{\mathbb{T}}}$ is assumed to be a random unit vector that is circularly symmetric, but whose exact distribution will be specified later. The random variable $R \in \mathbb{R}_0^+$ is chosen to be independent of $\hat{\mathbf{X}}$ and such that

$$\log R^2 \sim \mathcal{U}([\log x_{\min}^2, \log \mathcal{E}]) \quad (5.72)$$

where we choose x_{\min}^2 as

$$x_{\min}^2 \triangleq \log \mathcal{E}. \quad (5.73)$$

Note that this choice of R satisfies the peak-power constraint (2.21) and therefore also the average-power constraint (2.22).

Using such an input to our MIMO fading channel we get the following lower bound to channel capacity:

$$C(\mathcal{E}) \geq I(\mathbf{X}; \mathbf{Y}) \quad (5.74)$$

$$= I(R, \hat{\mathbf{X}}; \mathbf{Y}) \quad (5.75)$$

$$= I(\hat{\mathbf{X}}; \mathbf{Y}) + I(R; \mathbf{Y} | \hat{\mathbf{X}}) \quad (5.76)$$

$$= I(\hat{\mathbf{X}}; \mathbf{Y}) + I(R; \mathbf{Y} e^{i\Theta} | \hat{\mathbf{X}}) - I(R; \mathbf{Y} e^{i\Theta} | \hat{\mathbf{X}}) + I(R; \mathbf{Y} | \hat{\mathbf{X}}) \quad (5.77)$$

$$= I(\hat{\mathbf{X}}; \mathbf{Y}) + I(R, e^{i\Theta}; \mathbf{Y} e^{i\Theta} | \hat{\mathbf{X}}) - I(e^{i\Theta}; \mathbf{Y} e^{i\Theta} | \hat{\mathbf{X}}, R) - I(R; \mathbf{Y} e^{i\Theta} | \hat{\mathbf{X}}) + I(R; \mathbf{Y} | \hat{\mathbf{X}}). \quad (5.78)$$

Here we have introduced a new random variable $\Theta \sim \mathcal{U}([0, 2\pi])$ which is assumed to be independent of every other random quantity.

The last two terms can be rearranged as follows:

$$\begin{aligned} & -I(R; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) + I(R; \mathbf{Y} | \hat{\mathbf{X}}) \\ & = -h(\mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) + h(\mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}, R) + h(\mathbf{Y} | \hat{\mathbf{X}}) \\ & \quad - h(\mathbf{Y} | \hat{\mathbf{X}}, R) \end{aligned} \quad (5.79)$$

$$\begin{aligned} & = -h(\mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) + h(\mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}, R) + h(\mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}, e^{i\Theta}) \\ & \quad - h(\mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}, R, e^{i\Theta}) \end{aligned} \quad (5.80)$$

$$= -I(e^{i\Theta}; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) + I(e^{i\Theta}; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}, R). \quad (5.81)$$

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

Combining this with (5.78) we yield

$$\mathcal{C}(\mathcal{E}) \geq I(\hat{\mathbf{X}}; \mathbf{Y}) + I(R, e^{i\Theta}; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) - I(e^{i\Theta}; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) \quad (5.82)$$

$$= I(\hat{\mathbf{X}}; \mathbf{Y}) + I(Re^{i\Theta}; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) - I(e^{i\Theta}; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) \quad (5.83)$$

where the last equality follows because from $Re^{i\Theta}$ the random variables R and $e^{i\Theta}$ can be gained back.

We continue with bounding the first term in (5.83):

$$I(\hat{\mathbf{X}}; \mathbf{Y}) = I(\hat{\mathbf{X}}; \mathbf{Y}, \mathbf{Z}) - \underbrace{I(\hat{\mathbf{X}}; \mathbf{Z} | \mathbf{Y})}_{\leq \epsilon(x_{\min})} \quad (5.84)$$

$$\geq I(\hat{\mathbf{X}}; \mathbf{Y}, \mathbf{Z}) - \epsilon(x_{\min}) \quad (5.85)$$

$$= I(\hat{\mathbf{X}}; \mathbb{H}\hat{\mathbf{X}}R) - \epsilon(x_{\min}) \quad (5.86)$$

$$= I\left(\hat{\mathbf{X}}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}, \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R\right) - \epsilon(x_{\min}) \quad (5.87)$$

$$\begin{aligned} & = I\left(\hat{\mathbf{X}}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + I\left(\hat{\mathbf{X}}; \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \left| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right.\right) \\ & \quad - \epsilon(x_{\min}). \end{aligned} \quad (5.88)$$

Here the first equality follows from the chain rule; in the subsequent inequality we lower-bound the second term by $-\epsilon(x_{\min})$ which is defined in Appendix A.5 and is shown there to be independent of the input distribution $Q_{\mathbf{X}}$ and to tend to zero as $x_{\min} \uparrow \infty$; in the subsequent equality we use \mathbf{Z} in order to extract $\mathbb{H}\hat{\mathbf{X}}R$ from \mathbf{Y} and then drop (\mathbf{Y}, \mathbf{Z}) since given $\mathbb{H}\hat{\mathbf{X}}R$ it is independent of the other random variables; and the last equality follows again from the chain rule.

Similarly, we bound the third term in (5.83):

$$\begin{aligned} & I(e^{i\Theta}; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) \\ & \leq I(e^{i\Theta}; \mathbf{Y}e^{i\Theta}, \mathbf{Z}e^{i\Theta} | \hat{\mathbf{X}}) \end{aligned} \quad (5.89)$$

$$= I(e^{i\Theta}; \mathbb{H}\mathbf{X}e^{i\Theta}, \mathbf{Z}e^{i\Theta} | \hat{\mathbf{X}}) \quad (5.90)$$

$$= I(e^{i\Theta}; \mathbb{H}\mathbf{X}e^{i\Theta} | \hat{\mathbf{X}}) + I(e^{i\Theta}; \mathbf{Z}e^{i\Theta} | \mathbb{H}\mathbf{X}e^{i\Theta}, \hat{\mathbf{X}}) \quad (5.91)$$

$$= I(e^{i\Theta}; \mathbb{H}\mathbf{X}e^{i\Theta} | \hat{\mathbf{X}}) \quad (5.92)$$

$$= I \left(e^{i\Theta}; \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R, \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta} \middle| \hat{\mathbf{X}} \right) \quad (5.93)$$

$$= I \left(e^{i\Theta}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta} \middle| \hat{\mathbf{X}} \right) \\ + I \left(e^{i\Theta}; \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta}, \hat{\mathbf{X}} \right). \quad (5.94)$$

Hence, plugging these results into (5.83) we get:

$$C(\mathcal{E}) \geq I(Re^{i\Theta}; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) + I \left(\hat{\mathbf{X}}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) \\ + I \left(\hat{\mathbf{X}}; \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) \\ - I \left(e^{i\Theta}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta} \middle| \hat{\mathbf{X}} \right) \\ - I \left(e^{i\Theta}; \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta}, \hat{\mathbf{X}} \right) - \epsilon(x_{\min}). \quad (5.95)$$

We next bound the third and fifth mutual information term in (5.95):

$$I \left(\hat{\mathbf{X}}; \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - I \left(e^{i\Theta}; \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta}, \hat{\mathbf{X}} \right) \\ = h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}, \hat{\mathbf{X}} \right) \\ - h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta}, \hat{\mathbf{X}} \right) \\ + h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta}, \hat{\mathbf{X}}, e^{i\Theta} \right) \quad (5.96)$$

$$= h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}, \hat{\mathbf{X}} \right) \\ - h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta}, \hat{\mathbf{X}} \right) \\ + h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}, \hat{\mathbf{X}} \right) \quad (5.97)$$

$$= h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta}, \hat{\mathbf{X}} \right) \quad (5.98)$$

$$\geq h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta} \right) \quad (5.99)$$

$$= h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h \left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \middle| \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) \quad (5.100)$$

$$= 0. \quad (5.101)$$

Here, the inequality follows from conditioning that reduces entropy; and the second last equality holds because we have assumed $\hat{\mathbf{X}}$ to be circularly symmetric, *i.e.*, $\hat{\mathbf{X}}$ “destroys” the random phase shift of $e^{i\Theta}$.

Therefore, we are left with the following bound:

$$\begin{aligned} C(\mathcal{E}) &\geq I(Re^{i\Theta}; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) + I\left(\hat{\mathbf{X}}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) \\ &\quad - I\left(e^{i\Theta}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{i\Theta} \middle| \hat{\mathbf{X}}\right) - \epsilon(x_{\min}). \end{aligned} \quad (5.102)$$

Now, we rewrite the second and third term as follows:

$$\begin{aligned} &I\left(\hat{\mathbf{X}}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - I\left(e^{i\Theta}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{i\Theta} \middle| \hat{\mathbf{X}}\right) \\ &= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \middle| \hat{\mathbf{X}}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{i\Theta} \middle| \hat{\mathbf{X}}\right) \\ &\quad + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{i\Theta} \middle| \hat{\mathbf{X}}, e^{i\Theta}\right) \end{aligned} \quad (5.103)$$

$$\begin{aligned} &= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \middle| \hat{\mathbf{X}}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{i\Theta} \middle| \hat{\mathbf{X}}\right) \\ &\quad + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \middle| \hat{\mathbf{X}}\right) \end{aligned} \quad (5.104)$$

$$= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{i\Theta} \middle| \hat{\mathbf{X}}\right) \quad (5.105)$$

where the second equality follows from (2.8) with a choice $\mathbf{U} = e^{-i\theta}\mathbf{1}_{n_R}$ and from the fact that $e^{i\Theta}$ is independent of all other random quantities.

This leaves us with

$$\begin{aligned} C(\mathcal{E}) &\geq I(Re^{i\Theta}; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) \\ &\quad - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{i\Theta} \middle| \hat{\mathbf{X}}\right) - \epsilon(x_{\min}). \end{aligned} \quad (5.106)$$

We next let the power grow to infinity $\mathcal{E} \rightarrow \infty$ and use the definition of the fading number (2.24). Since $Re^{i\Theta}$ is circularly symmetric with a magnitude distributed according to (5.72), we know from [7, (108) and Th. 4.8], [2, (6.194) and Th. 6.15], that $Re^{i\Theta}$ achieves the fading number of a memoryless SIMO fading channel with partial side-information. In our situation we have

$$I(Re^{i\Theta}; \mathbf{Y}e^{i\Theta} | \hat{\mathbf{X}}) = I(Re^{i\Theta}; \mathbb{H}\hat{\mathbf{X}}Re^{i\Theta} + \mathbf{Z} | \hat{\mathbf{X}}) \quad (5.107)$$

$$= I(Re^{i\Theta}; \mathbb{H}\hat{\mathbf{X}}Re^{i\Theta} + \mathbf{Z}, \hat{\mathbf{X}}) \quad (5.108)$$

where $\hat{\mathbf{X}}$ serves as partial receiver side-information (that is independent of the SIMO input $Re^{i\Theta}$). Note that a random vector \mathbf{A} is said to contain only *partial* side-information about \mathbf{B} if $h(\mathbf{B}|\mathbf{A}) > -\infty$, *i.e.*, in our case we need

$$h(\mathbb{H}\hat{\mathbf{X}} | \hat{\mathbf{X}}) > -\infty \quad (5.109)$$

which is satisfied since we assume that $h(\mathbb{H}) > -\infty$ and $\mathbf{E}[\|\mathbb{H}\|_{\mathbb{F}}^2] < \infty$ (see [7, Lem. 6.6], [2, Lem. A.14]).

Hence,

$$\begin{aligned} \chi(\mathbb{H}) \geq & \overline{\lim}_{\mathcal{E} \uparrow \infty} \left\{ I(Re^{i\Theta}; \mathbb{H}\hat{\mathbf{X}}Re^{i\Theta} + \mathbf{Z} \mid \hat{\mathbf{X}}) + h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) \right. \\ & - h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta} \mid \hat{\mathbf{X}} \right) - \epsilon(x_{\min}) \\ & \left. - \log \left(1 + \log \left(1 + \frac{\mathcal{E}}{\sigma^2} \right) \right) \right\} \end{aligned} \quad (5.110)$$

$$\begin{aligned} = & \overline{\lim}_{\mathcal{E} \uparrow \infty} \left\{ I(Re^{i\Theta}; \mathbb{H}\hat{\mathbf{X}}Re^{i\Theta} + \mathbf{Z} \mid \hat{\mathbf{X}}) \right. \\ & \left. - \log \left(1 + \log \left(1 + \frac{\mathcal{E}}{\sigma^2} \right) \right) - \epsilon(x_{\min}) \right\} \\ & + h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta} \mid \hat{\mathbf{X}} \right) \end{aligned} \quad (5.111)$$

$$= \chi(\mathbb{H}\hat{\mathbf{X}} \mid \hat{\mathbf{X}}) + h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta} \mid \hat{\mathbf{X}} \right) \quad (5.112)$$

$$\begin{aligned} = & h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta} \mid \hat{\mathbf{X}} \right) + n_{\mathbb{R}} \mathbf{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \right] - \log 2 \\ & - h(\mathbb{H}\hat{\mathbf{X}} \mid \hat{\mathbf{X}}) + h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{i\Theta} \mid \hat{\mathbf{X}} \right) \end{aligned} \quad (5.113)$$

$$= h_\lambda \left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) + n_{\mathbb{R}} \mathbf{E} \left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \right] - \log 2 - h(\mathbb{H}\hat{\mathbf{X}} \mid \hat{\mathbf{X}}). \quad (5.114)$$

Here in (5.112) we have used the fact that our choice (5.73) guarantees that $\epsilon(x_{\min})$ tends to zero as $\mathcal{E} \rightarrow \infty$ (see Appendix A.5) and that we achieve the SIMO fading number for a channel with input $Re^{i\Theta}$ and output $\mathbb{H}\hat{\mathbf{x}}Re^{i\Theta} + \mathbf{Z}$; the subsequent equality follows from the fading number of a memoryless SIMO fading channel where the receiver has access to some partial side-information [7, (108)], [2, (6.194)]:

$$\chi(\mathbf{H}|\mathbf{S}) = h_\lambda(\hat{\mathbf{H}}e^{i\Theta} \mid \mathbf{S}) + n_{\mathbb{R}} \mathbf{E} \left[\log \|\mathbf{H}\|^2 \right] - \log 2 - h(\mathbf{H}|\mathbf{S}). \quad (5.115)$$

The result now follows by choosing the distribution $Q_{\hat{\mathbf{X}}}$ such as to maximize the lower bound (5.114) to the fading number.

Chapter 6

Discussion & Conclusion

We have derived 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 have precisely specified the behavior of the channel capacity asymptotically when the power grows to infinity. The result shows 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 non-negative random variable is such that (3.9) is satisfied. This is the well-known choice that 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.

We have then derived some new results for the important special situation of Gaussian fading. For the case of a scalar line-of-sight matrix (4.26) assuming at least as many transmit as receive antennas $n_R \leq n_T$ we have been able to state the fading number precisely:

$$\chi = n_R g_{n_R}(|d|^2) - n_R - \log \Gamma(n_R) \quad (6.1)$$

where $g_m(\cdot)$ denotes the expected value of a non-central chi-square random variable (see (4.19)). We see that the asymptotic capacity only depends on the number of receive antennas and is growing proportionally to $n_R \log |d|^2$.

For a general line-of-sight matrix we have shown an upper bound that grows like $\min\{n_R, n_T\} \log \delta^2$ where δ^2 is a certain kind of average of all singular values of the line-of-sight matrix (see (4.37) and (4.38)).

Once more we would like to emphasize that even though all results on the fading number are asymptotic results for the theoretical situation of infinite power, they are still of relevance for finite SNR values: it has been shown that the approximation

$$C(\text{SNR}) \approx \log(1 + \log(1 + \text{SNR})) + \chi \quad (6.2)$$

holds already for moderate values of the SNR. Moreover, since usually the double-logarithmic term is so slowly growing, the fading number χ will be dominating over $\log(1 + \log(1 + \text{SNR}))$ for these moderate SNR values. Hence, the threshold $C_0 \triangleq C(\text{SNR}_0)$ is closely related to the fading number. We have seen in Conjecture 1 that once the capacity is appreciably above $\chi + 2$ nats, the approximation (6.2) is likely to be valid and the transmission system in the highly inefficient log log-regime.

Therefore, the fading number can be seen as an indicator of the maximal rate at which power efficient communication is possible on the channel. For a further discussion about the practical relevance of the fading number we refer to [17], [16], [26].

In future we are going to try to improve these results further aiming at the precise derivation of the MIMO fading number with memory.

Appendix A

Some Derivations

A.1 Proof of Lemma 6

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

$$I(\mathbf{X}; \mathbf{Y}) = I(\mathbf{X}; \mathbf{Y} | e^{i\Theta}) \quad (\text{A.1})$$

$$= I(\mathbf{X}e^{i\Theta}; \mathbf{Y}e^{i\Theta} | e^{i\Theta}) \quad (\text{A.2})$$

$$= I(\mathbf{X}e^{i\Theta}; \mathbb{H}\mathbf{X}e^{i\Theta} + \mathbf{Z} | e^{i\Theta}) \quad (\text{A.3})$$

$$= I(\tilde{\mathbf{X}}; \mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} | e^{i\Theta}) \quad (\text{A.4})$$

$$= h(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} | e^{i\Theta}) - h(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} | \tilde{\mathbf{X}}, e^{i\Theta}) \quad (\text{A.5})$$

$$= h(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} | e^{i\Theta}) - h(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} | \tilde{\mathbf{X}}) \quad (\text{A.6})$$

$$\leq h(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z}) - h(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} | \tilde{\mathbf{X}}) \quad (\text{A.7})$$

$$= I(\tilde{\mathbf{X}}; \mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z}). \quad (\text{A.8})$$

Here the first equality follows because Θ is independent of every other random quantity; the third equality follows because \mathbf{Z} is circularly symmetric; in the subsequent equality we substitute $\tilde{\mathbf{X}} = \mathbf{X}e^{i\Theta}$; 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.

A.2 Derivation of Bounds (4.25)

In this appendix we will derive the bounds (4.25) on $g_m(\cdot)$. We start with the upper bound which follows directly from (4.22) and (4.23) and from Jensen's inequality:

$$g_m(s^2) = \mathbb{E} \left[\log \left(\sum_{j=1}^m |U_j + \mu_j|^2 \right) \right] \quad (\text{A.9})$$

$$\leq \log \left(\sum_{j=1}^m \mathbb{E} [|U_j + \mu_j|^2] \right) \quad (\text{A.10})$$

$$= \log \left(\sum_{j=1}^m (1 + |\mu_j|^2) \right) \quad (\text{A.11})$$

$$= \log(m + s^2). \quad (\text{A.12})$$

For the lower bound we also start with (4.22) and choose $\mu_1 = s$ and $\mu_2 = \dots = \mu_m = 0$. Then we get

$$g_m(s^2) = \mathbb{E} \left[\log \left(\sum_{j=1}^m |U_j + \mu_j|^2 \right) \right] \quad (\text{A.13})$$

$$\geq \mathbb{E} \left[\log \left(|U_1 + \mu_1|^2 \right) \right] \quad (\text{A.14})$$

$$= g_1(s^2) \quad (\text{A.15})$$

$$= \log s^2 - \text{Ei}(-s^2). \quad (\text{A.16})$$

Here, (A.14) follows from dropping some non-negative terms in the sum; and in the subsequent two equalities we use the definition of $g_1(\cdot)$.

A.3 Proof of Corollary 13

We choose a constant $n_T \times n_T$ matrix \mathbf{B} as follows

$$\mathbf{B} \triangleq \text{diag} \left(\frac{1}{d_1}, \dots, \frac{1}{d_{n_R}}, \frac{1}{d_1}, \dots, \frac{1}{d_1} \right) \quad (\text{A.17})$$

and then we note that for a unit vector $\hat{\mathbf{x}} = (\hat{x}^{(1)}, \dots, \hat{x}^{(n_T)})^\top$

$$\mathbb{H}\mathbf{B}\hat{\mathbf{x}} = \mathbf{D}\mathbf{B}\hat{\mathbf{x}} + \tilde{\mathbb{H}}\mathbf{B}\hat{\mathbf{x}} = \begin{pmatrix} \hat{x}^{(1)} \\ \vdots \\ \hat{x}^{(n_R)} \end{pmatrix} + \tilde{\mathbb{H}}\mathbf{B}\hat{\mathbf{x}} \triangleq \boldsymbol{\xi} + \tilde{\mathbf{H}} \quad (\text{A.18})$$

where $\tilde{\mathbf{H}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \zeta^2(\hat{\mathbf{x}}) \mathbf{I}_{n_R})$ with

$$\zeta^2(\hat{\mathbf{x}}) \triangleq \frac{|\hat{x}^{(1)}|^2}{|d_1|^2} + \dots + \frac{|\hat{x}^{(n_R)}|^2}{|d_{n_R}|^2} + \frac{|\hat{x}^{(n_R+1)}|^2}{|d_1|^2} + \dots + \frac{|\hat{x}^{(n_T)}|^2}{|d_1|^2} \quad (\text{A.19})$$

and where $\boldsymbol{\xi} \in \mathbb{C}^{n_R}$ with $\|\boldsymbol{\xi}\| \leq 1$. Therefore,

$$h(\mathbb{H}\mathbf{B}\hat{\mathbf{X}} \mid \hat{\mathbf{X}} = \hat{\mathbf{x}}) = n_R \log \pi e \zeta^2(\hat{\mathbf{x}}); \quad (\text{A.20})$$

$$\mathbb{E} [\log \|\mathbb{H}\mathbf{B}\hat{\mathbf{x}}\|^2] = \log \zeta^2(\hat{\mathbf{x}}) + g_{n_R} \left(\frac{\|\boldsymbol{\xi}\|^2}{\zeta^2(\hat{\mathbf{x}})} \right) \quad (\text{A.21})$$

(where the last equality follows from (4.22)); and hence

$$\begin{aligned} & n_R \mathbb{E} [\log \|\mathbb{H}\mathbf{B}\hat{\mathbf{X}}\|^2] - h(\mathbb{H}\mathbf{B}\hat{\mathbf{X}} \mid \hat{\mathbf{X}}) \\ &= n_R \mathbb{E} \left[g_{n_R} \left(\frac{|\hat{X}^{(1)}|^2 + \dots + |\hat{X}^{(n_R)}|^2}{\zeta^2(\hat{\mathbf{X}})} \right) \right] - n_R \log \pi e. \end{aligned} \quad (\text{A.22})$$

The upper bound on the fading number now follows from (3.10); from Theorem 8 by upper-bounding the h_λ -term by $\log c_{n_R}$; and from the additional observations that $g_m(\cdot)$ is a monotonically increasing function, that

$$|\hat{X}^{(1)}|^2 + \dots + |\hat{X}^{(n_R)}|^2 \leq 1 \quad (\text{A.23})$$

and that

$$\begin{aligned}\zeta^2(\hat{\mathbf{X}}) &= \frac{|\hat{X}^{(1)}|^2}{|d_1|^2} + \dots + \frac{|\hat{X}^{(n_R)}|^2}{|d_{n_R}|^2} \\ &\quad + \frac{|\hat{X}^{(n_R+1)}|^2}{|d_1|^2} + \dots + \frac{|\hat{X}^{(n_T)}|^2}{|d_1|^2}\end{aligned}\tag{A.24}$$

$$\geq \frac{|\hat{X}^{(1)}|^2}{|d_1|^2} + \dots + \frac{|\hat{X}^{(n_T)}|^2}{|d_1|^2}\tag{A.25}$$

$$= \frac{1}{|d_1|^2} \left(|\hat{X}^{(1)}|^2 + \dots + |\hat{X}^{(n_T)}|^2 \right)\tag{A.26}$$

$$= \frac{1}{|d_1|^2} = \frac{1}{\|\mathbf{D}\|^2}\tag{A.27}$$

where the inequality follows since $|d_1| \geq |d_2| \geq \dots \geq |d_{n_R}|$.

A.4 Proof of Proposition 14

This upper bound is based on the upper bound given in Corollary 9 for a choice of $\mathbf{B} = \mathbf{I}_{n_T}$. If $n_R > n_T$ we choose for \mathbf{A}

$$\mathbf{A} \triangleq \text{diag} \left(\frac{a}{d_1}, \dots, \frac{a}{d_{n_T}}, b, \dots, b \right)\tag{A.28}$$

with

$$b \triangleq \left(\frac{\delta^2}{n_T} \right)^{\frac{n_T}{2n_R}}\tag{A.29}$$

for δ as given in (4.38), and with a such that $\det \mathbf{A} = 1$, *i.e.*,

$$a \triangleq (d_1 \cdots d_{n_T})^{\frac{1}{n_T}} \cdot b^{\frac{n_T - n_R}{n_T}}.\tag{A.30}$$

For such a choice we note that

$$\begin{aligned}\mathbf{A}\mathbb{H}\hat{\mathbf{x}} &= a \begin{pmatrix} \hat{\mathbf{x}} \\ \mathbf{0} \end{pmatrix} + \left(\mathcal{N}_{\mathbb{C}} \left(0, \frac{|a|^2}{|d_1|^2} \right), \dots, \mathcal{N}_{\mathbb{C}} \left(0, \frac{|a|^2}{|d_{n_T}|^2} \right), \right. \\ &\quad \left. \mathcal{N}_{\mathbb{C}}(0, b^2), \dots, \mathcal{N}_{\mathbb{C}}(0, b^2) \right)^{\top}\end{aligned}\tag{A.31}$$

so that

$$\mathbb{E} [\|\mathbf{A}\mathbb{H}\hat{\mathbf{x}}\|^2] = \delta^2 (b^2)^{\frac{n_T - n_R}{n_T}} + (n_R - n_T)b^2\tag{A.32}$$

$$= n_R \left(\frac{\delta^2}{n_T} \right)^{\frac{n_T}{n_R}}.\tag{A.33}$$

Hence, using Jensen's inequality and the fact that $\det \mathbf{A} = 1$ we get

$$\begin{aligned}n_R \mathbb{E} [\log \|\mathbf{A}\mathbb{H}\hat{\mathbf{x}}\|^2] &- h(\mathbf{A}\mathbb{H}\hat{\mathbf{x}}) \\ &\leq n_R \log \mathbb{E} [\|\mathbf{A}\mathbb{H}\hat{\mathbf{x}}\|^2] - \log \det \mathbf{A} - h(\mathbb{H}\hat{\mathbf{x}})\end{aligned}\tag{A.34}$$

$$= n_R \log \left(n_R \left(\frac{\delta^2}{n_T} \right)^{n_T/n_R} \right) - n_R \log \pi e.\tag{A.35}$$

Plugging this into the upper bound (3.12) of Corollary 9 we yield

$$\begin{aligned}\chi &\leq n_{\text{R}} \log \pi - \log \Gamma(n_{\text{R}}) + n_{\text{R}} \log n_{\text{R}} + n_{\text{T}} \log \left(\frac{\delta^2}{n_{\text{T}}} \right) \\ &\quad - n_{\text{R}} \log \pi e\end{aligned}\tag{A.36}$$

$$= n_{\text{T}} \log \left(\frac{\delta^2}{n_{\text{T}}} \right) + n_{\text{R}} \log n_{\text{R}} - \log \Gamma(n_{\text{R}}) - n_{\text{R}}.\tag{A.37}$$

If $n_{\text{R}} \leq n_{\text{T}}$ we choose for \mathbf{A}

$$\mathbf{A} = \text{diag} \left(\frac{a}{d_1}, \dots, \frac{a}{d_{n_{\text{R}}}} \right)\tag{A.38}$$

with a such that $\det \mathbf{A} = 1$, *i.e.*,

$$a \triangleq (d_1 \dots d_{n_{\text{R}}})^{\frac{1}{n_{\text{R}}}}.\tag{A.39}$$

For such a choice we note that

$$\begin{aligned}\mathbf{A}\mathbb{H}\hat{\mathbf{x}} &= a \left(\hat{x}^{(1)}, \dots, \hat{x}^{(n_{\text{R}})} \right)^{\text{T}} \\ &\quad + \left(\mathcal{N}_{\mathbb{C}} \left(0, \frac{|a|^2}{|d_1|^2} \right), \dots, \mathcal{N}_{\mathbb{C}} \left(0, \frac{|a|^2}{|d_{n_{\text{R}}}|^2} \right) \right)^{\text{T}}\end{aligned}\tag{A.40}$$

so that

$$\begin{aligned}\mathbb{E} [\|\mathbf{A}\mathbb{H}\hat{\mathbf{x}}\|^2] &= |a|^2 \left(|\hat{x}^{(1)}|^2 + \dots + |\hat{x}^{(n_{\text{R}})}|^2 \right) \\ &\quad + \frac{|a|^2}{|d_1|^2} + \dots + \frac{|a|^2}{|d_{n_{\text{R}}}|^2}\end{aligned}\tag{A.41}$$

$$\leq \delta^2\tag{A.42}$$

where we have bounded $|\hat{x}^{(1)}|^2 + \dots + |\hat{x}^{(n_{\text{R}})}|^2 \leq 1$. Hence, using Jensen's inequality and the fact that $\det \mathbf{A} = 1$ we get

$$\begin{aligned}n_{\text{R}} \mathbb{E} [\log \|\mathbf{A}\mathbb{H}\hat{\mathbf{x}}\|^2] - h(\mathbf{A}\mathbb{H}\hat{\mathbf{x}}) \\ \leq n_{\text{R}} \log \mathbb{E} [\|\mathbf{A}\mathbb{H}\hat{\mathbf{x}}\|^2] - \log \det \mathbf{A} - h(\mathbb{H}\hat{\mathbf{x}})\end{aligned}\tag{A.43}$$

$$\leq n_{\text{R}} \log \delta^2 - n_{\text{R}} \log \pi e.\tag{A.44}$$

Plugging this into the upper bound (3.12) of Corollary 9 we yield

$$\chi \leq n_{\text{R}} \log \pi - \log \Gamma(n_{\text{R}}) + n_{\text{R}} \log \delta^2 - n_{\text{R}} \log \pi e\tag{A.45}$$

$$= n_{\text{R}} \log \frac{\delta^2}{n_{\text{R}}} + n_{\text{R}} \log n_{\text{R}} - \log \Gamma(n_{\text{R}}) - n_{\text{R}}.\tag{A.46}$$

The result now follows by combining (A.37) and (A.46).

A.5 Additional Derivation for the Proof of the Lower Bound

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

$$I(\hat{\mathbf{X}}; \mathbf{Z} | \mathbf{Y}) \leq \epsilon(x_{\min})\tag{A.47}$$

and to show that $\epsilon(x_{\min})$ does not depend on the input distribution $Q_{\mathbf{X}}$ and tends to zero as x_{\min} tends to infinity.

Such a bound can be found as follows:

$$I(\hat{\mathbf{X}}; \mathbf{Z} | \mathbf{Y}) = h(\mathbf{Z} | \mathbf{Y}) - h(\mathbf{Z} | \mathbf{Y}, \hat{\mathbf{X}}) \quad (\text{A.48})$$

$$\leq h(\mathbf{Z}) - h(\mathbf{Z} | \mathbf{Y}, \hat{\mathbf{X}}, R) \quad (\text{A.49})$$

$$= h(\mathbf{Z}) - h(\mathbf{Z} | \mathbb{H}\hat{\mathbf{X}}R + \mathbf{Z}, \hat{\mathbf{X}}, R) \quad (\text{A.50})$$

$$\leq h(\mathbf{Z}) - \inf_{\hat{\mathbf{x}}} \inf_{r \geq x_{\min}} h(\mathbf{Z} | \mathbb{H}\hat{\mathbf{x}}r + \mathbf{Z}) \quad (\text{A.51})$$

$$= h(\mathbf{Z}) - \inf_{\hat{\mathbf{x}}} h(\mathbf{Z} | \mathbb{H}\hat{\mathbf{x}}x_{\min} + \mathbf{Z}) \quad (\text{A.52})$$

$$= \sup_{\hat{\mathbf{x}}} I(\mathbf{Z}; \mathbb{H}\hat{\mathbf{x}}x_{\min} + \mathbf{Z}) \quad (\text{A.53})$$

$$= \sup_{\hat{\mathbf{x}}} I\left(\frac{\mathbf{Z}}{x_{\min}}; \mathbb{H}\hat{\mathbf{x}} + \frac{\mathbf{Z}}{x_{\min}}\right) \quad (\text{A.54})$$

$$= \sup_{\hat{\mathbf{x}}} \left\{ h\left(\mathbb{H}\hat{\mathbf{x}} + \frac{\mathbf{Z}}{x_{\min}}\right) - h(\mathbb{H}\hat{\mathbf{x}}) \right\} \quad (\text{A.55})$$

$$\triangleq \epsilon(x_{\min}) \quad (\text{A.56})$$

where we have used the fact that we have chosen R such that $R \geq x_{\min}$. Note that (A.55) does not depend on the input \mathbf{X} anymore. The convergence

$$\lim_{x_{\min} \uparrow \infty} \epsilon(x_{\min}) = 0 \quad (\text{A.57})$$

follows from [7, Lem. 6.11], [2, Lem. A.19].

Bibliography

- [1] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October 1948.
- [2] S. M. Moser, “Duality-based bounds on channel capacity,” Ph.D. dissertation, Swiss Federal Institute of Technology (ETH), Zurich, October 2004, Diss. ETH No. 15769. [Online]. Available: <http://moser.cm.nctu.edu.tw/>
- [3] A. Lapidoth and S. M. Moser, “The fading number of single-input multiple-output fading channels with memory,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 437–453, February 2006.
- [4] —, “The fading number of SIMO fading channels with memory,” in *Proceedings IEEE International Symposium on Information Theory and its Applications (ISITA)*, Parma, Italy, October 10–13, 2004, pp. 287–292.
- [5] —, “Feedback increases neither the fading number nor the pre-log,” in *Proceedings Twenty-Third IEEE Convention of Electrical & Electronics Engineers in Israel (IEEEI)*, Herzlia, Israel, September 6–7, 2004, pp. 213–215.
- [6] —, “Bounds on the capacity of the discrete-time Poisson channel,” in *Proceedings Forty-First Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, IL, USA, October 1–3, 2003, pp. 201–210.
- [7] —, “Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2426–2467, October 2003.
- [8] —, “The asymptotic capacity of the discrete-time Poisson channel,” in *Proceedings Winter School on Coding and Information Theory*, Monte Verità, Ascona, Switzerland, February 24–27, 2003.
- [9] —, “Duality-based bounds on the capacity of the discrete-time Poisson channel,” October 2007, submitted. [Online]. Available: <http://moser.cm.nctu.edu.tw/publications.html>
- [10] —, “Capacity bounds via duality with applications to multi-antenna systems on flat fading channels,” Signal and Information Processing Laboratory, ETH Zurich, Tech. Rep., June 25, 2002, preprint.
- [11] —, “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.

- [12] —, “On the fading number of multi-antenna systems,” in *Proceedings IEEE Information Theory Workshop (ITW)*, Cairns, Australia, September 2–7, 2001, pp. 110–111.
- [13] —, “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.
- [14] —, “Limits on reliable communication over flat-fading channels,” in *Proceedings Winter School on Coding and Information Theory*, Schloss Reinsensburg, Günzburg, University of Ulm, Germany, December 17–20, 2000.
- [15] 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.
- [16] 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.
- [17] —, “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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] A. Lapidoth, “On the high-SNR capacity of noncoherent networks,” *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3025–3036, September 2005.
- [22] —, “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.
- [23] —, “Capacity bounds via duality: a phase noise example,” in *Proceedings Second Asian-European Workshop on Information Theory*, Breisach, Germany, June 26–29, 2002.
- [24] —, “On phase noise channels at high SNR,” in *Proceedings IEEE Information Theory Workshop (ITW)*, Bangalore, India, October 20–25, 2002, pp. 1–4.

- [25] A. Lapidoth and S. M. Moser, “The expected logarithm of a noncentral chi-square random variable,” website. [Online]. Available: <http://moser.cm.nctu.edu.tw/explog.html>
- [26] A. Lapidoth, “The fading number and degrees of freedom: A peace pipe,” presentation, Shushan Purim, Israel, March 27, 2005.