Capacity Analysis of Multiple-Access OFDM Channels

Final Report of Project
"4G Wireless Access Technology"

Contract-Number:	G1-95003
Project Duration:	1 January 2006–31 December 2006
Funded by:	Industrial Technology Research Institute (ITRI), Zhudong, Taiwan

Author:	Stefan M. Moser
Organization:	Information Theory Laboratory Department of Communication Engineering National Chiao Tung University
Address:	Engineering Building IV, Room 727 1001 Ta Hsueh Rd. Hsinchu 300, Taiwan
E-mail:	stefan.moser@ieee.org

Abstract

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

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 (or *doubling* the SNR on a dB-scale)! Here, depending on the channel in use, "high SNR" typically starts somewhere between 30 to 80 dB. Since transmission in such a regime will be highly inefficient, it is crucial to avoid a system operating at such a rate. Hence we need 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 threshold between the power-efficient low-SNR and the power-inefficient high-SNR regime is the so-called *fading number* which is defined as the second term in the high-SNR asymptotic expansion of channel capacity.

The results of this report concern this fading number. We investigate a channel model based on a flat fading assumption without inter-symbol interference, which is a typical situation encountered when using a system based on orthogonal frequency division multiplexing (OFDM) or on orthogonal frequency division multiple-access (OFDMA). We assume several users at the transmitter side, but only one receiver (typical setup of a several mobile users communicating with one base-station), where all users and the receiver might have multiple antennas available. In the most general setup we do not restrict the fading to have a particular distribution (*i.e.*, it need not be Gaussian), but we only ask it to be a stationary, ergodic, finite-energy, and regular random process, possibly with memory both over time and space. For the sake of simplicity, however, we will introduce further restrictions or simplifying assumptions on the way.

The results can be grouped into two main chapters: firstly we investigate a single-user setup where we allow multiple-antennas both at transmitter and receiver. The fading is assumed to have no temporal memory, but the different antennas are allowed to have arbitrary correlation. The distribution of the fading is not restricted to be Gaussian, and not specified apart from the stationarity, ergodicity and regularity assumptions. In this setup we are able to derive the fading number precisely, *i.e.*, we are able to specify the exact asymptotic channel capacity in the limit when the available power tends to infinity, and we can give a good estimation of the threshold between the efficient low- to medium-SNR regime and the highly power-inefficient high-SNR regime.

This result is then specialized to the already 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. As a byproduct a new upper bound is derived on $g_m(\cdot)$, the expected value of a logarithm of a non-central chi-square random variable.

In a second chapter we investigate a simple two-user single-antenna setup assuming no memory and Gaussian fading. We prove that under the additional constraint that the users use circularly symmetric signaling, the sum-rate capacity of this multiple-access channel equals to the single-user capacity of the user with the better channel.

Keywords: Channel capacity, fading, fading number, flat fading channel, Gaussian fading, general fading, high signal-to-noise ratio, high SNR, joint estimation and detection, MAC, memory, MIMO, multiple-access channel, multiple-antenna, multiple-input multiple-output, multiple-user, non-central chi-square, non-coherent.

Contents

Acknowledgments

The author would like to thank Lai Kuei-Chiang and Lee Shuenn-Gi for their assistance during this project, and Su Yu-Ted for fruitful discussions.

This work was supported by the Industrial Technology Research Institute (ITRI), Zhudong, Taiwan, under contract G1-95003.

Part of the results presented here have been submitted to the *IEEE Transactions on Information Theory*.

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

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

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

- The signals of several transmitters arrive as a superposition at the receiver, *i.e.*, the different users act as interferers to each other.

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

The first two effects lead to a channel that not only adds noise to the transmitted signal (as this is the case for the traditional wired communication channel), but also changes the amplitude of the signal (so called *fading*) and introduces intersymbol interference. The latter effect can be combatted using appropriate transmissions schemes and coding like, *e.g.*, an orthogonal frequency division multiplexing (OFDM) system. Fading is more difficult to deal with. It has a strong impact on the performance of a system and is at high SNR the dominating source of transmission errors. Clever coding approaches are needed and one needs to take advantage of various types of *diversity* that is available.

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

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

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

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

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

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

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

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

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

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

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

- the free-space optical intensity channel [2], [6], [8];
- an optical intensity channel with input-dependent noise [2];
- the Poisson channel [2], [6], [8];

[1] As a matter of fact, this optimization is infeasible for most channels of interest.
[2] The technique works for general channels, not fading channels only.

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

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

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

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

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

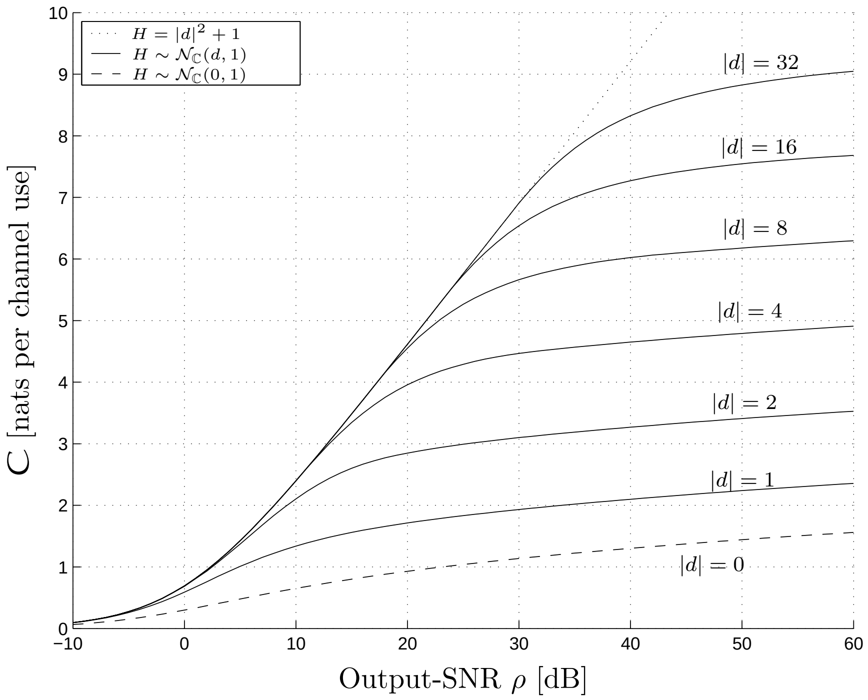

Figure 1.1: An upper bound on the capacity of a Rician fading channel as a function of the output-SNR $\rho = (1 + |d|^2)\text{SNR}$ for different values of the specular component d. The dotted line depicts the capacity of a Gaussian channel of equal output-SNR ρ, namely $\log(1 + \rho)$.

1.2 The Fading Number

In an attempt to quantify this threshold between efficient low- to medium-SNR regime and inefficient high-SNR regime more precisely, [7], [2] introduce the *fading number*. The fading number is defined as the second term in the high-SNR capacity, *i.e.*, at high SNR the channel capacity can be expressed as

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and in the SIMO case[4]

$$\chi(\mathbf{H}) = h_\lambda\bigl(\hat{\mathbf{H}}e^{\mathrm{i}\Theta}\bigr) - h(\mathbf{H}) + n_\mathrm{R}\mathsf{E}\bigl[\log\|\mathbf{H}\|^2\bigr] - \log 2 \tag{1.6}$$

(both are special cases from the corresponding situation with memory); and also for the MISO case [7], [2]

$$\chi\bigl(\mathbf{H}^\mathsf{T}\bigr) = \sup_{\hat{\mathbf{x}}}\left\{\log\pi + \mathsf{E}\Bigl[\log\bigl|\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}\bigr|^2\Bigr] - h\bigl(\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}\bigr)\right\}. \tag{1.7}$$

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

The most general situation of multiple antennas at both transmitter and receiver, however, has been solved so far only in the special situation 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

$$\chi(\mathbb{H}) = \log\frac{\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})} + n_\mathrm{R}\mathsf{E}\bigl[\log\|\mathbb{H}\hat{\mathbf{e}}\|^2\bigr] - h(\mathbb{H}\hat{\mathbf{e}}) \tag{1.8}$$

where $\hat{\mathbf{e}} \in \mathbb{C}^{n_\mathrm{T}}$ 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 3.2).

[4]For a precise definition of the used notation we refer to Section 2.1.

However, there are still many interesting cases open and unsolved. For example, the fading number in a multiple-user scenario is completely unknown, even for the SISO case. It is also interesting to further study the influence of multiple transmitter antennas on the fading number more in detail.

The remainder of this report is structured as follows: after some remarks about notation and a detailed mathematical definition of the channel model in the following chapter, we will firstly present the situation of a general single-user multiple-input multiple-output (MIMO) fading channel in Chapter 3. There in Section 3.1 we will derive some very interesting new results, *inter alia*,

- a lemma that proves that the capacity achieving input distribution in general must be circularly symmetric; and
- the fading number of general memoryless MIMO fading channels.

For the sake of clarity, some known results that are used in the proofs are discussed as well, *inter alia*, the concept of distributions that *escape to infinity*.

We then specialize these results to the already known fading numbers of SISO, SIMO, MISO, and rotation-commutative MIMO fading channels in Section 3.2. In Section 3.3 we investigate the situation of Gaussian fading processes and—as a byproduct—derive a new non-trivial upper bound on $g_m(\cdot)$, the function that describes the expected value of the logarithm of a non-central chi-square random variable.

In Chapter 4 we next study the simplest case of a multiple-user fading channel: a two-user Gaussian multiple-access fading channel where each user and the receiver have only one antenna and where the two fading paths are uncorrelated. We will show some bounds on the fading number in this scenario and show that if one of the users uses a circularly symmetric input then the fading number of this MAC coincides with the fading number of the single-user channel with higher capacity.

We will conclude in Chapter 5.

Some longer proofs can be found in the Appendices.

Chapter 2

Definitions and Notation

2.1 Notation

We try to use upper-case letters for random quantities and lower-case letters for their realizations. This rule, however, is broken when dealing with matrices and some constants. To better differentiate between scalars, vectors, and matrices we have resorted to using different fonts for the different quantities. Upper-case letters such as X are used to denote scalar random variables taking value in the reals $\mathbb{R}$ or in the complex plane $\mathbb{C}$. Their realizations are typically written in lower-case, *e.g.*, x. For random vectors we use bold face capitals, *e.g.*, $\mathbf{X}$ and bold lower-case for their realizations, *e.g.*, $\mathbf{x}$. Deterministic matrices are denoted by upper-case letters but of a special font, *e.g.*, H; and random matrices are denoted using another special upper-case font, *e.g.*, $\mathbb{H}$. If scalars or deterministic scalar functions are not denoted using Greek or lower-case letters, we use yet another font, *e.g.*, C for capacity or $\mathsf{F}(\cdot)$ for the spectral density function. The energy per symbol is denoted by $\mathcal{E}$ and the signal-to-noise ratio SNR is denoted by $\textsc{snr}$.

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

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

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

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

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

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

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

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

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

Note that for every matrix A

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

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

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

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

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

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

To be able to work with such *direction vectors* we shall need a differential entropy-like quantity for random vectors that take value on the unit sphere in $\mathbb{C}^m$: let λ denote the area measure on the unit sphere in $\mathbb{C}^m$. If a random vector $\hat{\mathbf{V}}$ takes value in the unit sphere and has the density $p^{\lambda}_{\hat{\mathbf{V}}}(\hat{\mathbf{v}})$ with respect to λ, then we shall let

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

if the expectation is defined.

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

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

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

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

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

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

The definition (2.7) can be easily extended to conditional entropies: if $\mathbf{W}$ is some random vector, and if conditional on $\mathbf{W} = \mathbf{w}$ the random vector $\hat{\mathbf{V}}$ has density $p^{\lambda}_{\hat{\mathbf{V}}|\mathbf{W}}(\hat{\mathbf{v}}|\mathbf{w})$ then we can define

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

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

Based on these definitions we have the following lemma:

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

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

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

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

Proof. Omitted. □

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

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

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

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

All rates specified in this report 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 m users each having n_i transmit antennas, $i = 1, \ldots, m$. The total number of transmit antennas is then

$$\sum_{i=1}^{m} n_i = n_{\mathrm{T}}. \tag{2.15}$$

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

$$\mathbf{Y}_k = \mathbb{H}_k \mathbf{x}_k + \mathbf{Z}_k. \tag{2.16}$$

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

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

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

$$\mathsf{E}\left[\|\mathbb{H}_k\|_{\mathrm{F}}^2\right] < \infty \tag{2.17}$$

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

$$h(\{\mathbb{H}_k\}) \triangleq \lim_{n \uparrow \infty} \frac{1}{n} h(\mathbb{H}_1, \ldots, \mathbb{H}_n) > -\infty. \tag{2.18}$$

Furthermore, we will restrict ourselves to the memoryless case, *i.e.*, we assume that $\{\mathbb{H}_k\}$ is IID with respect to 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 most of the time drop the time index k. In this case (2.16) simplifies to

$$\mathbf{Y} = \mathbb{H}\mathbf{x} + \mathbf{Z}. \tag{2.19}$$

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 $\mathbf{x}$. The different users are assumed to have access to a common clock (resulting in the common discrete time k), but are otherwise not allowed to cooperate, *i.e.*, the m subvectors of $\mathbf{x}$ are independent of each other.

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

$$\text{SNR} \triangleq \frac{\mathcal{E}}{\sigma^2}. \tag{2.20}$$

The total power then still must be split and distributed among all users, however, since we are only studying the sum rate we are not interested in that part at the moment. For simplicity we may assume that each user gets the same amount of power $\frac{\mathcal{E}}{m}$.

The (sum-rate) capacity $\mathsf{C}(\text{SNR})$ of the channel (2.16) and (2.19), respectively, is given by

$$\mathsf{C}(\text{SNR}) = \lim_{n\uparrow\infty} \frac{1}{n} \sup I\left(\mathbf{X}_1^n; \mathbf{Y}_1^n\right) = \sup I\left(\mathbf{X}; \mathbf{Y}\right) \tag{2.21}$$

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

$$\|\mathbf{X}\|^2 \leq \mathcal{E}, \quad \text{almost surely} \tag{2.22}$$

for a peak-power constraint, or

$$\mathsf{E}\left[\|\mathbf{X}\|^2\right] \leq \mathcal{E} \tag{2.23}$$

for an average-power constraint.

Specializing [7, Theorem 4.2], [2, Theorem 6.10] to memoryless MIMO fading, we have for $m = 1$

$$\varlimsup_{\text{SNR}\uparrow\infty} \Big\{ \mathsf{C}(\text{SNR}) - \log\log \text{SNR} \Big\} < \infty. \tag{2.24}$$

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

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

$$\chi(\mathbb{H}) \triangleq \varlimsup_{\text{SNR}\uparrow\infty} \Big\{ \mathsf{C}(\text{SNR}) - \log\log \text{SNR} \Big\}. \tag{2.25}$$

Prima facie the fading number depends on whether a peak-power constraint (2.22) or an average-power constraint (2.23) is imposed on the input. However, in the situation of $m = 1$ it will turn out that the memoryless MIMO fading number is identical for both cases.

In the following chapter we will consider a general single-user MIMO fading channel where the fading process is assumed to be memoryless, but otherwise is unrestricted.

Chapter 3

A General Single-User Memoryless MIMO Fading Channel

In this chapter we consider the special case of only one user $m = 1$. Moreover, we simplify the model to exclude temporal memory. Otherwise we assume the channel to be general without any particular assumption on the fading distribution, the number of antennas at transmitter and receiver, or on the correlation between the different antennas. Hence, we have a channel model that looks like (2.19) with either an average-power constraint (2.23) or a peak-power constraint (2.22) on the input.

3.1 Results

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

3.1.1 Escaping to Infinity

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

A sequence of input distributions parameterized by the allowed cost (in our case of fading channels the cost is the available power or the SNR, respectively) is said to *escape to infinity* if it assigns to every fixed compact set a probability that tends to zero as the allowed cost tends to infinity. 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 of importance since the asymptotic capacity of many channels of interest can only be achieved by input distributions that escape to infinity. As a matter of fact one can show that to achieve a mutual information of only identical asymptotic *growth rate* as the capacity, the input distribution *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], [2], [7].

Definition 3. *Let $\{Q_{\mathcal{E}}\}_{\mathcal{E}\geq 0}$ be a family of input distributions for the memoryless fading channel* (2.19)*, where this family is parameterized by the available average power $\mathcal{E}$ such that*

$$\mathsf{E}_{Q_{\mathcal{E}}}\left[\|\mathbf{X}\|^2\right] \leq \mathcal{E}, \qquad \mathcal{E} \geq 0. \tag{3.1}$$

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

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

We now have the following:

Lemma 4. *Let the memoryless MIMO fading channel be given as in* (2.19) *and let $W(\cdot|\cdot)$ denote the corresponding conditional channel law. Let $\{Q_{\mathcal{E}}\}_{\mathcal{E}\geq 0}$ be a family of input distributions satisfying the power constraint* (3.1) *and the condition*

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

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

Proof. A proof can be found in [3], [2]. □

Hence, when computing bounds on the fading number (which is part of the capacity in the limit when $\mathcal{E}$ tends to infinity, see (2.25)) we may assume that

$$\Pr\left[\|\mathbf{X}\|^2 \leq \mathcal{E}_0\right] = 0. \tag{3.4}$$

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 on it. The new proposed technique bases on a dual expression of mutual information that leads to an expression of capacity as a minimization instead of a maximization. This way it becomes much easier to find upper bounds.

Again, here we only state the upper bound in a form needed in the derivation of Theorem 8, for a more general form, for more mathematical details, and for all proofs we refer to [7], [2].

Lemma 5. *Consider a memoryless MISO fading channel with input $\mathbf{S} \in \mathbb{C}^{n_{\mathrm{R}}}$ and output $T \in \mathbb{C}$ such that*

$$T = \mathbf{H}^{\mathsf{T}}\mathbf{S} + Z. \tag{3.5}$$

Then the mutual information between input and output of the channel is upper-bounded as follows:

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

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

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

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

$$\mathbf{W} \stackrel{\mathscr{L}}{=} \mathbf{W} \cdot e^{\mathrm{i}\Theta} \tag{3.7}$$

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

In case of a random process we make the following definition: we say that a vector random process $\{\mathbf{W}_k\}$ is *circularly symmetric* if

$$\{\mathbf{W}_k\} \stackrel{\mathscr{L}}{=} \{\mathbf{W}_k e^{\mathrm{i}\Theta_k}\}, \tag{3.8}$$

i.e., the joint distribution defining $\{\mathbf{W}_k\}$ is identical to the joint distribution of a new process that is given as the product of the original process and a independent random process $\{e^{\mathrm{i}\Theta_k}\}$ where $\{\Theta_k\}$ is IID $\sim \mathcal{U}([0, 2\pi])$.

Note an important subtlety of this definition: a random process being circularly symmetric does not only mean that for every time k the corresponding random vector is circularly symmetric, but also that from past vectors one cannot gain any knowledge about the present phase, *i.e.*, the phase is IID.

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* (2.16)*. Then the capacity achieving input distribution can be assumed to be circularly symmetric,* i.e.*, the input vectors* $\{\mathbf{X}_k\}$ *can be replaced by* $\{\mathbf{X}_k e^{\mathrm{i}\Theta_k}\}$*, where* $\{\Theta_k\}$ *is IID* $\sim \mathcal{U}([0, 2\pi])$ *and independent of every other random quantity.*

Proof. The proof is given in Appendix A. □

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

3.1.4 Fading Number of a General Memoryless MIMO Fading Channel

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

Theorem 8. *Consider a memoryless MIMO fading channel* (2.19) *where the random fading matrix* $\mathbb{H}$ *takes value in* $\mathbb{C}^{n_R \times n_T}$ *and satisfies*

$$h(\mathbb{H}) > -\infty \tag{3.9}$$

and

$$\mathsf{E}\left[\|\mathbb{H}\|_F^2\right] < \infty. \tag{3.10}$$

Then, irrespective of whether a peak-power constraint (2.22) *or an average-power constraint* (2.23) *is imposed on the input, the fading number* $\chi(\mathbb{H})$ *is given by*

$$\boxed{\chi(\mathbb{H}) = \sup_{Q_{\hat{\mathbf{X}}}} \left\{ h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + n_R \mathsf{E}\left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2\right] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) \right\}} \tag{3.11}$$

where $\hat{\mathbf{X}}$ *denotes a random vector of unit length.*
Moreover, this fading number is achievable by a random vector $\mathbf{X} = \hat{\mathbf{X}} \cdot R$ *where* $\hat{\mathbf{X}}$ *is distributed according to the distribution that achieves the fading number in* (3.11) *and where* R *is a non-negative random variable such that*

$$\log R^2 \sim \mathcal{U}\left([\log\log \mathcal{E}, \log \mathcal{E}]\right). \tag{3.12}$$

Proof. A proof is given in Appendix B. □

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

$$\|\mathbb{H}\hat{\mathbf{X}}e^{i\Phi}\| \stackrel{\mathscr{L}}{=} \|\mathbb{H}\hat{\mathbf{X}}\| \tag{3.13}$$

and

$$\begin{aligned} h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) &= h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}, e^{i\Phi}\big) && (3.14)\\ &= h\big(\mathbb{H}\hat{\mathbf{X}}e^{i\Phi} \,\big|\, \hat{\mathbf{X}}e^{i\Phi}, e^{i\Phi}\big) && (3.15)\\ &= h\big(\mathbb{H}\hat{\mathbf{X}}' \,\big|\, \hat{\mathbf{X}}', e^{i\Phi}\big) && (3.16)\\ &= h\big(\mathbb{H}\hat{\mathbf{X}}' \,\big|\, \hat{\mathbf{X}}'\big), && (3.17) \end{aligned}$$

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

Hence, in (3.11) the only term that depends on the choice of the phase distribution of $\hat{\mathbf{X}}$ is $h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right)$. This term is maximized for a circularly symmetric $\hat{\mathbf{X}}$ as can be seen as follows:

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) = I\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|};\hat{\mathbf{X}}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) \tag{3.18}$$

$$= I\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|};\hat{\mathbf{X}}\,\middle|\,e^{\mathrm{i}\Theta}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) \tag{3.19}$$

$$= I\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta};\hat{\mathbf{X}}\,\middle|\,e^{\mathrm{i}\Theta}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) \tag{3.20}$$

$$= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,e^{\mathrm{i}\Theta}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}},e^{\mathrm{i}\Theta}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) \tag{3.21}$$

$$= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,e^{\mathrm{i}\Theta}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}},e^{\mathrm{i}\Theta}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) \tag{3.22}$$

$$= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,e^{\mathrm{i}\Theta}\right) \tag{3.23}$$

$$\leq h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\right). \tag{3.24}$$

Here the first equality follows from the chain rule; the subsequent equality because $e^{\mathrm{i}\Theta}$ is independent of all other random quantities; (3.22) follows from scaling property of differential entropy by a constant of magnitude 1; and the final inequality from conditioning that reduces entropy.

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

The evaluation of (3.11) can be pretty awkward mainly due to the first term, *i.e.*, the differential entropy with respect to the surface area measure λ. We therefore will derive next an upper bound to the fading number that is easier to evaluate.

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

$$\chi(\mathsf{A}\mathbb{H}\mathsf{B}) = \chi(\mathbb{H}) \tag{3.25}$$

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

$$h_\lambda\big(\hat{\mathbf{Y}}\big) \leq \log c_{n_\mathrm{R}} = \log\frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})} \tag{3.26}$$

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

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

Corollary 9. *The fading number of a memoryless MIMO fading channel as defined in Theorem 8 can be upper-bounded as follows:*

$$\chi(\mathbb{H}) \leq n_\mathrm{R} \log \pi - \log \Gamma(n_\mathrm{R}) + \inf_{\mathsf{A},\mathsf{B}} \sup_{\hat{\mathbf{x}}} \left\{ n_\mathrm{R} \mathsf{E}\left[\log \|\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{x}}\|^2\right] - h\left(\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{x}}\right) \right\} \tag{3.27}$$

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

Proof. From the two observations (3.25) and (3.26) mentioned above we immediately get from Theorem 8:

$$\chi(\mathbb{H}) \leq \inf_{\mathsf{A},\mathsf{B}} \sup_{Q_{\hat{\mathbf{X}}}} \mathsf{E}_{\hat{\mathbf{X}}} \Bigg[n_\mathrm{R} \log \pi - \log \Gamma(n_\mathrm{R}) + n_\mathrm{R} \mathsf{E}\left[\log \|\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{X}}\|^2 \,\Big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}} \right] \\ - h\left(\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{X}} \,\middle|\, \hat{\mathbf{X}} = \hat{\mathbf{x}}\right) \Bigg]. \tag{3.28}$$

The result now follows by noting that (3.27) can always be achieved by choosing $Q_{\hat{\mathbf{X}}}$ in (3.28) to be the distribution which with probability 1 takes on the value $\hat{\mathbf{x}}$ that achieves the maximum in (3.27). □

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

3.2 Some Known Special Cases

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

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

$$\mathsf{V}_r \mathbb{H} \stackrel{\mathscr{L}}{=} \mathbb{H} \mathsf{V}_t \tag{3.29}$$

where $\stackrel{\mathscr{L}}{=}$ stands for "has the same law"; and for every constant unitary $n_\mathrm{R} \times n_\mathrm{R}$ matrix V_r there exists a constant unitary $n_\mathrm{T} \times n_\mathrm{T}$ matrix V_t such that (3.29) holds [7, Definition 4.37], [2, Definition 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 $\hat{\mathbf{X}} \in \mathbb{C}^{n_\mathrm{T}}$ is an isotropically distributed random vector that is independent of $\mathbb{H}$, then $\mathbb{H}\hat{\mathbf{X}} \in \mathbb{C}^{n_\mathrm{R}}$ is isotropically distributed.*
- *If $\hat{\mathbf{e}}, \hat{\mathbf{e}}' \in \mathbb{C}^{n_\mathrm{T}}$ are two constant unit vectors, then*

$$\|\mathbb{H}\hat{\mathbf{e}}\| \stackrel{\mathscr{L}}{=} \|\mathbb{H}\hat{\mathbf{e}}'\|, \quad \|\hat{\mathbf{e}}\| = \|\hat{\mathbf{e}}'\| = 1 \tag{3.30}$$

$$h\left(\mathbb{H}\hat{\mathbf{e}}\right) = h\left(\mathbb{H}\hat{\mathbf{e}}'\right), \quad \|\hat{\mathbf{e}}\| = \|\hat{\mathbf{e}}'\| = 1. \tag{3.31}$$

Proof. For a proof see, *e.g.*, [7, Lemma 4.38], [2, Lemma 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 (3.11) 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

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) = \log c_{n_\mathrm{R}} \tag{3.32}$$

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

$$c_{n_\mathrm{R}} = \frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})}. \tag{3.33}$$

The fading number then becomes

$$\chi(\mathbb{H}) = \log\frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})} - \log 2 + n_\mathrm{R}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{e}}\|^2\right] - h(\mathbb{H}\hat{\mathbf{e}}) \tag{3.34}$$

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

In case of a SIMO fading channel, $\hat{\mathbf{X}} = e^{\mathrm{i}\Phi}$. Since the optimal distribution in (3.11) is a circularly symmetric unit random variable, we get $\hat{\mathbf{X}} = e^{\mathrm{i}\Theta}$ and (3.11) becomes:

$$\chi(\mathbf{H}) = h_\lambda\big(\hat{\mathbf{H}}e^{\mathrm{i}\Theta}\big) + \mathsf{E}\left[\log\|\mathbf{H}\|^2\right] - \log 2 - h(\mathbf{H}). \tag{3.35}$$

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

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

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

$$h_\lambda\left(\frac{\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}}{|\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}|}\right) = h_\lambda\big(e^{\mathrm{i}\Theta}\big) = \log 2\pi \tag{3.36}$$

and the fading number becomes

$$\chi(\mathbf{H}^\mathsf{T}) = \sup_{Q_{\hat{\mathbf{X}}}}\left\{\log 2\pi + \mathsf{E}\left[\log|\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}|^2\right] - \log 2 - h\big(\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}}\big)\right\} \tag{3.37}$$

$$= \sup_{Q_{\hat{\mathbf{X}}}}\mathsf{E}_{\hat{\mathbf{X}}}\left[\log\pi + \mathsf{E}\left[\log|\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}|^2\,\Big|\,\hat{\mathbf{X}} = \hat{\mathbf{x}}\right] - h\big(\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}\,\big|\,\hat{\mathbf{X}} = \hat{\mathbf{x}}\big)\right] \tag{3.38}$$

$$\leq \sup_{\hat{\mathbf{x}}}\left\{\log\pi + \mathsf{E}\left[\log|\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}|^2\right] - h\big(\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}\big)\right\} \tag{3.39}$$

which can be achieved by an input

$$\hat{\mathbf{X}} = \hat{\mathbf{x}}e^{\mathrm{i}\Theta} \tag{3.40}$$

where $\hat{\mathbf{x}}$ is the deterministic unit vector that achieves the fading number in (3.39).

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

$$h_\lambda\big(e^{\mathrm{i}\Theta}\big) = \log 2\pi. \tag{3.41}$$

This yields

$$\begin{aligned}\chi(H) &= \log 2\pi + \mathsf{E}\big[\log|H|^2\big] - \log 2 - h(H) && (3.42)\\ &= \log \pi + \mathsf{E}\big[\log|H|^2\big] - h(H). && (3.43)\end{aligned}$$

3.3 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 other.

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

$$\mathbb{H} = \mathsf{D} + \tilde{\mathbb{H}} \tag{3.44}$$

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

Note that for some constant unitary $n_{\mathrm{R}} \times n_{\mathrm{R}}$ matrix U and some constant unitary $n_{\mathrm{T}} \times n_{\mathrm{T}}$ matrix V the law of $\mathsf{U}\tilde{\mathbb{H}}\mathsf{V}$ is identical to the law of $\tilde{\mathbb{H}}$. Therefore, without loss of generality, we may restrict ourselves to matrices D that are "diagonal", *i.e.*, for $n_{\mathrm{R}} \le n_{\mathrm{T}}$,

$$\mathsf{D} = \begin{pmatrix} \tilde{\mathsf{D}} & \mathsf{0}_{n_{\mathrm{R}}\times(n_{\mathrm{T}}-n_{\mathrm{R}})}\end{pmatrix} \tag{3.45}$$

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

$$\mathsf{D} = \begin{pmatrix} \tilde{\mathsf{D}} \\ \mathsf{0}_{(n_{\mathrm{R}}-n_{\mathrm{T}})\times n_{\mathrm{T}}}\end{pmatrix} \tag{3.46}$$

where $\tilde{\mathsf{D}}$ is a $\min\{n_{\mathrm{R}}, n_{\mathrm{T}}\} \times \min\{n_{\mathrm{R}}, n_{\mathrm{T}}\}$ diagonal matrix with the singular values of D on the diagonal.

3.3.1 Scalar Line-of-Sight Matrix

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

Under these assumptions the fading number has been known already for $n_{\mathrm{R}} = n_{\mathrm{T}} = m$, in which case the fading matrix $\mathbb{H}$ is rotation-commutative [7], [2], and

$$\chi(\mathbb{H}) = m g_m\big(|d|^2\big) - m - \log\Gamma(m) \tag{3.47}$$

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

$$g_m(\xi) \triangleq \begin{cases} \log(\xi) - \mathrm{Ei}(-\xi) + \displaystyle\sum_{j=1}^{m-1}(-1)^j\left[e^{-\xi}(j-1)! - \frac{(m-1)!}{j(m-1-j)!}\right]\left(\frac{1}{\xi}\right)^j, & \xi > 0 \\ \psi(m), & \xi = 0 \end{cases} \tag{3.48}$$

for $m \in \mathbb{N}$. Here $\mathrm{Ei}(-\cdot)$ denotes the exponential integral function defined as

$$\mathrm{Ei}(-x) \triangleq -\int_x^\infty \frac{e^{-t}}{t}\,\mathrm{d}t, \qquad x > 0 \tag{3.49}$$

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

$$\psi(m) \triangleq -\gamma + \sum_{j=1}^{m-1} \frac{1}{j} \tag{3.50}$$

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

$$\mathsf{E}\left[\log\left(\sum_{j=1}^m \left|U_j + \mu_j\right|^2\right)\right] = g_m(s^2), \tag{3.51}$$

where

$$s^2 = \sum_{j=1}^m |\mu_j|^2 \tag{3.52}$$

(see [24], [7, Lemma 10.1], [2, Lemma A.6] for more details and a proof). For a plot of $g_m(\cdot)$ for various values of m see Figure 3.1.

We now consider the case where $n_\mathrm{R} \le n_\mathrm{T}$:

Corollary 11. *Assume* $n_\mathrm{R} \le n_\mathrm{T}$ *and a Gaussian fading matrix as given in* (3.44). *Let the line-of-sight matrix* D *be given as*

$$\mathsf{D} = d\left(\mathsf{I}_{n_\mathrm{R}} \quad \mathsf{0}_{n_\mathrm{R}\times(n_\mathrm{T}-n_\mathrm{R})}\right). \tag{3.53}$$

Then

$$\chi(\mathbb{H}) = n_\mathrm{R} g_{n_\mathrm{R}}\left(|d|^2\right) - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}) \tag{3.54}$$

where $g_m(\cdot)$ *is defined in* (3.48).

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

$$\hat{\mathbf{X}} = \begin{pmatrix} \boldsymbol{\Xi} \\ \boldsymbol{\Xi}' \end{pmatrix} \tag{3.55}$$

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

$$\mathbb{H}\hat{\mathbf{X}} = \mathsf{D}\hat{\mathbf{X}} + \tilde{\mathbb{H}}\hat{\mathbf{X}} = d\boldsymbol{\Xi} + \tilde{\mathbf{H}} \tag{3.56}$$

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

$$h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) = h\big(\tilde{\mathbf{H}}\big) = n_\mathrm{R}\log \pi e; \tag{3.57}$$

$$n_\mathrm{R}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\right] = n_\mathrm{R} g_{n_\mathrm{R}}\left(|d|^2\|\boldsymbol{\Xi}\|^2\right) \le n_\mathrm{R} g_{n_\mathrm{R}}\left(|d|^2\right); \tag{3.58}$$

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) \le \log\frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})}. \tag{3.59}$$

Here, the equality in (3.58) follows from the fact that $\|d\boldsymbol{\Xi} + \tilde{\mathbf{H}}\|^2$ is non-central chi-square distributed and from (3.51); the inequality in (3.58) follows from the monotonicity of $g_m(\cdot)$ [24] and is tight if $\|\boldsymbol{\Xi}\| = 1$, *i.e.*, $\boldsymbol{\Xi}' = \mathbf{0}$; and the inequality in (3.59) follows from (2.9) and is tight if $\boldsymbol{\Xi}$ is uniformly distributed on the unit sphere in $\mathbb{C}^{n_\mathrm{R}}$ such that $\mathbb{H}\hat{\mathbf{X}}$ is isotropically distributed. The result now follows from Theorem 8. □

The case $n_R > n_T$ is more difficult since then (3.59) is in general not tight. We will only state an upper bound:

Proposition 12. *Assume* $n_R > n_T$ *and a Gaussian fading matrix as given in* (3.44). *Let the line-of-sight matrix* D *be given as*

$$\mathsf{D} = d \begin{pmatrix} \mathsf{I}_{n_T} \\ \mathsf{0}_{(n_R - n_T)\times n_T} \end{pmatrix}. \tag{3.60}$$

Then

$$\chi(\mathbb{H}) \leq n_T \log\left(1 + \frac{|d|^2}{n_T}\right) + n_R \log n_R - n_R - \log\Gamma(n_R). \tag{3.61}$$

Proof. This result is a special case of Proposition 14 and has been published before in [7, (128)], [2, (6.224)]. □

3.3.2 General Line-of-Sight Matrix

Next we assume Gaussian fading as defined in (3.44) with a general line-of-sight matrix D having singular values $d_1, \ldots, d_{\min\{n_R, n_T\}}$. Hence, $\tilde{\mathsf{D}}$, defined in (3.45) and (3.46), is given as

$$\tilde{\mathsf{D}} = \operatorname{diag}\left(d_1, \ldots, d_{\min\{n_R,n_T\}}\right) \tag{3.62}$$

where $|d_1| \geq |d_2| \geq \ldots \geq |d_{\min\{n_R,n_T\}}|$.

We again start with the case $n_R \leq n_T$.

Corollary 13. *Assume* $n_R \leq n_T$ *and a Gaussian fading matrix as given in* (3.44). *Let the line-of-sight matrix* D *have singular values* $d_1, \ldots, d_{n_R}$*, where* $|d_1| \geq |d_2| \geq \ldots \geq |d_{n_R}|$*. Then*

$$\chi(\mathbb{H}) \leq n_R g_{n_R}\left(\|\mathsf{D}\|^2\right) - n_R - \log\Gamma(n_R) \tag{3.63}$$

where $g_m(\cdot)$ *is given in* (3.48) *and where* $\|\mathsf{D}\|^2 = |d_1|^2$*. Furthermore,*

$$\chi(\mathbb{H}) \geq n_R \mathsf{E}\left[g_{n_R}\left(\frac{1}{\frac{|\hat{X}^{(1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_R)}|^2}{|d_{n_R}|^2}}\right)\right] - n_R - \log\Gamma(n_R) \tag{3.64}$$

$$\geq n_R g_{n_R}\left(\frac{1}{\frac{1}{|d_1|^2} + \cdots + \frac{1}{|d_{n_R}|^2}}\right) - n_R - \log\Gamma(n_R) \tag{3.65}$$

where the expectation is over an isotropically distributed random unit vector $\hat{\mathbf{X}} \in \mathbb{C}^{n_R}$.

Proof. A proof is given in Appendix C. □

The situation $n_R > n_T$ is again more complicated. We include this case in a new upper bound based on (3.27) which holds independently of the particular relation of n_R and n_T:

Proposition 14. *Assume a Gaussian fading matrix as given in* (3.44) *and let the line-of-sight matrix* D *be general with singular values* $d_1, \ldots, d_{\min\{n_R,n_T\}}$*. Then the fading number is upper-bounded as follows:*

$$\chi(\mathbb{H}) \leq \min\{n_R, n_T\} \log \frac{\delta^2}{\min\{n_R, n_T\}} + n_R \log n_R - n_R - \log\Gamma(n_R) \tag{3.66}$$

where

$$\delta^2 \triangleq \left(|d_1|^2 \cdot \dots \cdot |d_{\min\{n_\mathrm{R},n_\mathrm{T}\}}|^2\right)^{1/\min\{n_\mathrm{R},n_\mathrm{T}\}} \left(1 + \frac{1}{|d_1|^2} + \dots + \frac{1}{|d_{\min\{n_\mathrm{R},n_\mathrm{T}\}}|^2}\right). \tag{3.67}$$

Proof. A proof is given in Appendix D. □

Note that from Corollary 11 and from Proposition 14 we can derive an interesting new upper bound on $g_m(\cdot)$:

Lemma 15. *The function* $g_m(\cdot)$ *as defined in* (3.48) *can be upper-bounded as follows:*

$$g_m(\xi) \leq \log(m + \xi) \tag{3.68}$$

$\xi \geq 0$, $m \in \mathbb{N}$.

This bound is actually quite tight as can be seen from Figure 3.1.

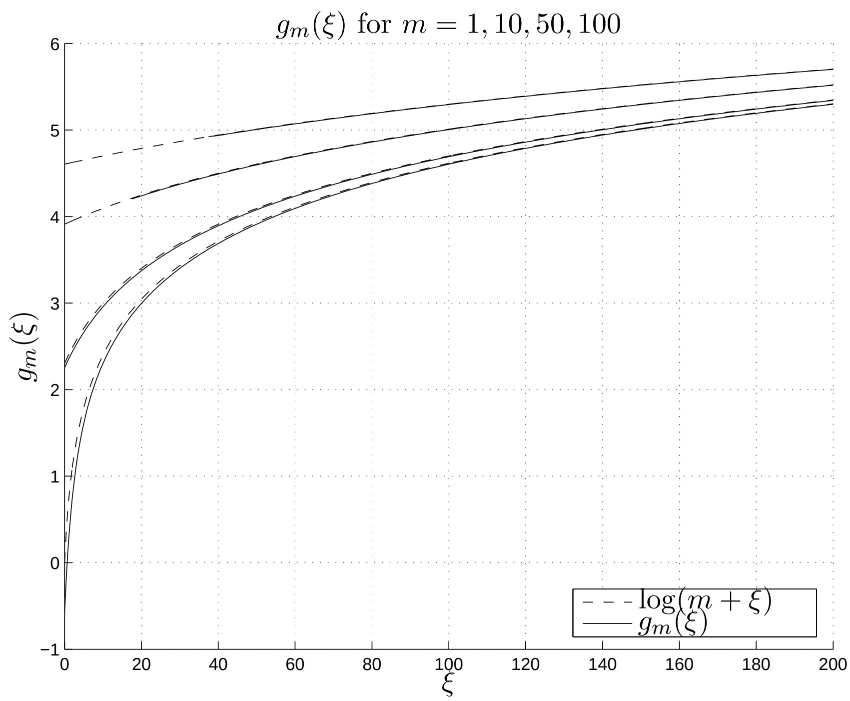

Figure 3.1: The function $g_m(\cdot)$ and its corresponding upper bound for $m = 1, 10, 50, 100$.

Proof. Assume $n_\mathrm{R} \leq n_\mathrm{T}$ and $d_1 = d_2 = \ldots = d_{n_\mathrm{R}} = d$. Then δ^2 in (3.67) is given as $\delta^2 = |d|^2 + n_\mathrm{R}$. Hence, we get from Corollary 11 and from Proposition 14

$$\chi = n_\mathrm{R} g_{n_\mathrm{R}}\left(|d|^2\right) - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}) \tag{3.69}$$

$$\leq n_\mathrm{R} \log \frac{|d|^2 + n_\mathrm{R}}{n_\mathrm{R}} + n_\mathrm{R} \log n_\mathrm{R} - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}) \tag{3.70}$$

$$= n_\mathrm{R} \log(|d|^2 + n_\mathrm{R}) - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}) \tag{3.71}$$

which proves the claim. □

Chapter 4

A Two-User SISO Gaussian Multiple-Access Fading Channel

In this chapter we consider the special case of a memoryless two-user Gaussian multiple-access fading channel where each transmitter and the receiver have only one antenna *i.e.*, $m = 2$, $n_1 = n_2 = 1$ so that $n_{\mathrm{T}} = 2$, and $n_{\mathrm{R}} = 1$, and where the two channel paths suffer from independent fading. The channel output $Y \in \mathbb{C}$ can then be written as

$$\begin{aligned} Y &= H^{(1)}x^{(1)} + H^{(2)}x^{(2)} + Z && (4.1)\\ &= d^{(1)}x^{(1)} + \tilde{H}^{(1)}x^{(1)} + d^{(2)}x^{(2)} + \tilde{H}^{(2)}x^{(2)} + Z, && (4.2)\end{aligned}$$

where $x^{(i)} \in \mathbb{C}$ denotes the input of user i, $i = 1, 2$; where the random variables $H^{(i)}$ describe Gaussian fading

$$\tilde{H}^{(i)} + d^{(i)} = H^{(i)} \sim \mathcal{N}_{\mathbb{C}}\Big(d^{(i)}, 1\Big), \qquad i = 1, 2, \tag{4.3}$$

(hence, $\tilde{H}^{(i)}$ are zero-mean, circularly symmetric, complex Gaussian random variables with variance 1) and are assumed to be independent

$$H^{(1)} \perp\!\!\!\perp H^{(2)}; \tag{4.4}$$

and where $Z \sim \mathcal{N}_{\mathbb{C}}\big(0, \sigma^2\big)$ denotes additive, zero-mean, circularly symmetric Gaussian noise.

Note that given $X^{(1)} = x^{(1)}$ and $X^{(2)} = x^{(2)}$ the channel output is Gaussian distributed:

$$Y \sim \mathcal{N}_{\mathbb{C}}\Big(d^{(1)}x^{(1)} + d^{(2)}x^{(2)}, \big|x^{(1)}\big|^2 + \big|x^{(2)}\big|^2 + \sigma^2\Big). \tag{4.5}$$

4.1 Preliminaries

It is important to note that our multiple-access channel (MAC) model can be considered as a special case of a single-user multiple-input single-output (MISO) fading channel with two antennas at the transmitter where, as additional constraint, cooperation between the transmitter antennas is prohibited. Hence, any known result on the MISO Gaussian fading channel are implicitly upper bounds to the situation considered here. In particular we know that the high-SNR sum-rate capacity can be upper-bounded as follows:

$$I\big(X^{(1)}, X^{(2)}; Y\big) \leq \mathsf{C}_{\mathrm{MISO}}(\mathcal{E}) = \log\log\frac{\mathcal{E}}{\sigma^2} + \chi_{\mathrm{MISO}} + o(1), \tag{4.6}$$

where $o(1)$ denotes terms that tend to zero as $\mathcal{E}$ tends to infinity and where χ_{MISO} denotes the fading number of memoryless MISO Gaussian fading:

$$\begin{aligned}\chi_{\text{MISO}} &= \log \|\mathbf{d}\|^2 - \text{Ei}\bigl(-\|\mathbf{d}\|^2\bigr) - 1 && (4.7)\\ &= \log\left(\bigl|d^{(1)}\bigr|^2 + \bigl|d^{(2)}\bigr|^2\right) - \text{Ei}\Bigl(-\bigl|d^{(1)}\bigr|^2 + \bigl|d^{(2)}\bigr|^2\Bigr) - 1. && (4.8)\end{aligned}$$

Here $\text{Ei}(-\cdot)$ is the exponential integral function defined as

$$\text{Ei}(-x) \triangleq -\int_x^\infty \frac{e^{-t}}{t}\,\text{d}t, \qquad x > 0. \tag{4.9}$$

On the other hand can we always lower-bound the sum rate of the MAC by switching off one of the users:

$$\begin{aligned}&I\bigl(X^{(1)}, X^{(2)}; Y\bigr)\\ &\quad\geq \max_i \mathsf{C}_{i,\text{SISO}}(\mathcal{E}) && (4.10)\\ &\quad= \max_{i=1,2}\left\{\log\log\frac{\mathcal{E}}{\sigma^2} + \chi_{i,\text{SISO}} + o(1)\right\} && (4.11)\\ &\quad= \log\log\frac{\mathcal{E}}{\sigma^2} + \max_{i=1,2}\chi_{i,\text{SISO}} + o(1) && (4.12)\\ &\quad= \log\log\frac{\mathcal{E}}{\sigma^2} + \max_{i=1,2}\left\{\log\left(\bigl|d^{(i)}\bigr|^2\right) - \text{Ei}\Bigl(-\bigl|d^{(i)}\bigr|^2\Bigr) - 1\right\} + o(1) && (4.13)\\ &\quad= \log\log\frac{\mathcal{E}}{\sigma^2} + \log\left(d_{\max}^2\right) - \text{Ei}\bigl(-d_{\max}^2\bigr) - 1 + o(1), && (4.14)\end{aligned}$$

where the last equality follows from the monotonicity of $\log(\cdot) - \text{Ei}(-\cdot)$ and where

$$d_{\max} \triangleq \max\left\{\bigl|d^{(1)}\bigr|, \bigl|d^{(2)}\bigr|\right\}. \tag{4.15}$$

Hence, we can write

$$\mathsf{C}_{\text{MAC}}(\mathcal{E}) = \log\log\frac{\mathcal{E}}{\sigma^2} + \log\left(d_{\text{MAC}}^2\right) - \text{Ei}\bigl(-d_{\text{MAC}}^2\bigr) - 1 + o(1) \tag{4.16}$$

where we have introduced d_{MAC} to be a non-negative real number satisfying

$$\max\left\{\bigl|d^{(1)}\bigr|, \bigl|d^{(2)}\bigr|\right\} \leq d_{\text{MAC}} \leq \sqrt{\bigl|d^{(1)}\bigr|^2 + \bigl|d^{(2)}\bigr|^2}. \tag{4.17}$$

Hence, the fading number of Gaussian MAC is defined and given as

$$\begin{aligned}\chi_{\text{MAC}} &\triangleq \lim_{\mathcal{E}\to\infty}\left\{\mathsf{C}_{\text{MAC}}(\mathcal{E}) - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} && (4.18)\\ &= \log\left(d_{\text{MAC}}^2\right) - \text{Ei}\bigl(-d_{\text{MAC}}^2\bigr) - 1, && (4.19)\end{aligned}$$

for some d_{MAC} that still needs to be determined. In the following we will study d_{MAC}.

4.2 An Upper Bound on the Sum-Rate Capacity and Fading Number

As in Chapter 3 we will rely on the new technique of deriving upper bounds on channel capacity that has been introduced in [7], [2]. Since a MAC is basically a

MISO channel with the additional constraint of non-cooperation, we can use the results from Lemma 5:

$$\begin{aligned}
& I\big(1X^{(1)}, X^{(2)}; Y\big) \\
&\quad \leq -h\big(Y \,\big|\, X^{(1)}, X^{(2)}\big) + \log \pi + \alpha \log \beta + \log \Gamma\left(\alpha, \frac{\nu}{\beta}\right) \\
&\qquad + (1-\alpha)\mathsf{E}\big[\log\big(|Y|^2 + \nu\big)\big] + \frac{1}{\beta}\mathsf{E}\big[|Y|^2\big] + \frac{\nu}{\beta} && (4.20)\\
&\quad \leq -h\big(Y \,\big|\, X^{(1)}, X^{(2)}\big) + \log \pi + \alpha \log \beta + \log \Gamma\left(\alpha, \frac{\nu}{\beta}\right) \\
&\qquad + (1-\alpha)\mathsf{E}\big[\log |Y|^2\big] + \epsilon_\nu + \frac{1}{\beta}\mathsf{E}\big[|Y|^2\big] + \frac{\nu}{\beta} && (4.21)\\
&\quad = -\mathsf{E}\Big[\log \pi e \left(\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2\right)\Big] + \log \pi + \alpha \log \beta + \log \Gamma\left(\alpha, \frac{\nu}{\beta}\right) \\
&\qquad + (1-\alpha)\mathsf{E}\Big[\mathsf{E}\Big[\log |Y|^2 \,\Big|\, X^{(1)} = x^{(1)}, X^{(2)} = x^{(2)}\Big]\Big] + \epsilon_\nu \\
&\qquad + \frac{1}{\beta}\mathsf{E}\Big[\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2 + \big|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\big|^2\Big] + \frac{\nu}{\beta} && (4.22)\\
&\quad = -\mathsf{E}\Big[\log\left(\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2\right)\Big] - 1 + \alpha \log \beta + \log \Gamma\left(\alpha, \frac{\nu}{\beta}\right) \\
&\qquad + (1-\alpha)\mathsf{E}\Big[\log\left(\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2\right)\Big] \\
&\qquad + (1-\alpha)\mathsf{E}\left[\log \frac{\big|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2} - \mathrm{Ei}\left(-\frac{\big|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2}\right)\right] \\
&\qquad + \epsilon_\nu + \frac{1}{\beta}\mathsf{E}\Big[\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2 + \big|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\big|^2\Big] + \frac{\nu}{\beta} && (4.23)\\
&\quad = -1 + \mathsf{E}\left[\log \frac{\big|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2} - \mathrm{Ei}\left(-\frac{\big|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2}\right)\right] \\
&\qquad + \alpha\Bigg(\log \beta - \mathsf{E}\Big[\log\left(\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2\right)\Big] \\
&\qquad\qquad - \mathsf{E}\left[\log \frac{\big|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2} - \mathrm{Ei}\left(-\frac{\big|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2}\right)\right]\Bigg) \\
&\qquad + \log \Gamma\left(\alpha, \frac{\nu}{\beta}\right) + \epsilon_\nu \\
&\qquad + \frac{1}{\beta}\mathsf{E}\Big[\big|X^{(1)}\big|^2 + \big|X^{(2)}\big|^2 + \sigma^2 + \big|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\big|^2\Big] + \frac{\nu}{\beta}. && (4.24)
\end{aligned}$$

Here the first inequality follows from Lemma 5; in the subsequent equality we assume $0 < \alpha < 1$ such that $1 - \alpha > 0$ and define

$$\begin{aligned}
\epsilon_\nu \triangleq \sup_{x^{(1)}, x^{(2)}} \Big\{ & \mathsf{E}\Big[\log\big(|Y|^2 + \nu\big) \,\Big|\, X^{(1)} = x^{(1)}, X^{(2)} = x^{(2)}\Big] \\
& - \mathsf{E}\Big[\log |Y|^2 \,\Big|\, X^{(1)} = x^{(1)}, X^{(2)} = x^{(2)}\Big] \Big\}, && (4.25)
\end{aligned}$$

such that

$$\begin{aligned}
& (1-\alpha)\mathsf{E}\big[\log\big(|Y|^2 + \nu\big)\big] \\
&\quad = (1-\alpha)\mathsf{E}\big[\log |Y|^2\big] + (1-\alpha)\left(\mathsf{E}\big[\log\big(|Y|^2 + \nu\big)\big] - \mathsf{E}\big[\log |Y|^2\big]\right) && (4.26)
\end{aligned}$$

$$
\begin{aligned}
&\leq (1-\alpha)\mathsf{E}\big[\log|Y|^2\big] \\
&\quad + (1-\alpha)\sup_{x^{(1)},x^{(2)}}\Big\{\mathsf{E}\Big[\log\big(|Y|^2+\nu\big)\ \Big|\ X^{(1)}=x^{(1)},X^{(2)}=x^{(2)}\Big] \\
&\qquad\qquad - \mathsf{E}\Big[\log|Y|^2\ \Big|\ X^{(1)}=x^{(1)},X^{(2)}=x^{(2)}\Big]\Big\} && (4.27)\\
&= (1-\alpha)\mathsf{E}\big[\log|Y|^2\big] + (1-\alpha)\epsilon_\nu && (4.28)\\
&\leq (1-\alpha)\mathsf{E}\big[\log|Y|^2\big] + \epsilon_\nu; && (4.29)
\end{aligned}
$$

in the subsequent equality we use the fact that given $X^{(1)}=x^{(1)}$ and $X^{(2)}=x^{(2)}$ the channel output is Gaussian distributed according to (4.5); in the subsequent equality we evaluate the expected logarithm of a non-central chi-square random variable as derived in [24], [7, Lemma 10.1], [2, Lemma A.6] and also discussed in (3.51); and the last equality follows from simple algebraic rearrangements.

Next we bound the following expressions:

$$
\mathsf{E}\Big[\log\Big(\big|X^{(1)}\big|^2+\big|X^{(2)}\big|^2+\sigma^2\Big)\Big] \geq \log\sigma^2; \qquad (4.30)
$$

$$
\mathsf{E}\left[\log\frac{\big|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2+\big|X^{(2)}\big|^2+\sigma^2} - \mathrm{Ei}\left(-\frac{\big|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2+\big|X^{(2)}\big|^2+\sigma^2}\right)\right] \geq -\gamma; \qquad (4.31)
$$

and

$$
\begin{aligned}
&\mathsf{E}\Big[\big|X^{(1)}\big|^2+\big|X^{(2)}\big|^2+\sigma^2+\big|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\big|^2\Big] \\
&\quad\leq \mathcal{E}+\sigma^2+\mathsf{E}\Big[\big|d^{(1)}X^{(1)}\big|^2+\big|d^{(2)}X^{(2)}\big|^2\Big] && (4.32)\\
&\quad\leq \mathcal{E}+\sigma^2+\max\Big\{\big|d^{(1)}\big|^2,\big|d^{(1)}\big|^2\Big\}\mathsf{E}\Big[\big|X^{(1)}\big|^2+\big|X^{(2)}\big|^2\Big] && (4.33)\\
&\quad\leq \mathcal{E}+\sigma^2+\max\Big\{\big|d^{(1)}\big|^2,\big|d^{(1)}\big|^2\Big\}\mathcal{E} && (4.34)\\
&\quad\triangleq \big(1+d_{\max}^2\big)\mathcal{E}+\sigma^2. && (4.35)
\end{aligned}
$$

Here, (4.30) follows from dropping some non-negative terms; (4.31) follows because $\log\xi-\mathrm{Ei}(-\xi)\geq-\gamma$ where $\gamma\approx 0.57$ denotes Euler's constant; and to derive (4.35) we used the Schwarz inequality and the fact that the input needs to satisfy the average-power constraint.

Moreover, we bound

$$
\begin{aligned}
&\mathsf{E}\left[\log\frac{\big|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2+\big|X^{(2)}\big|^2+\sigma^2} - \mathrm{Ei}\left(-\frac{\big|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2+\big|X^{(2)}\big|^2+\sigma^2}\right)\right] \\
&\quad\leq \mathsf{E}\left[\log\frac{\big|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2+\big|X^{(2)}\big|^2} - \mathrm{Ei}\left(-\frac{\big|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2+\big|X^{(2)}\big|^2}\right)\right], \qquad (4.36)
\end{aligned}
$$

which follows from the monotonicity of $\log\xi-\mathrm{Ei}(-\xi)$.

Together with (4.24) we then get

$$
\begin{aligned}
&I\big(X^{(1)},X^{(2)};Y\big) \\
&\quad\leq -1+\mathsf{E}\left[\log\frac{\big|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2+\big|X^{(2)}\big|^2} - \mathrm{Ei}\left(-\frac{\big|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\big|^2}{\big|X^{(1)}\big|^2+\big|X^{(2)}\big|^2}\right)\right]+\epsilon_\nu \\
&\qquad + \alpha\big(\log\beta-\log\sigma^2+\gamma\big)+\log\Gamma\left(\alpha,\frac{\nu}{\beta}\right)+\frac{1}{\beta}\Big(\big(1+d_{\max}^2\big)\mathcal{E}+\sigma^2\Big)+\frac{\nu}{\beta} \qquad (4.37)
\end{aligned}
$$

$$\leq \sup_{Q_{X^{(1)}}\cdot Q_{X^{(2)}}} \Bigg\{ \log \mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2+\left|X^{(2)}\right|^2}\right] - \mathrm{Ei}\left(-\mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2+\left|X^{(2)}\right|^2}\right]\right) - 1 \Bigg\} + \epsilon_\nu$$
$$+ \alpha\big(\log\beta - \log\sigma^2 + \gamma\big) + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) + \frac{1}{\beta}\Big(\big(1+d_{\max}^2\big)\mathcal{E}+\sigma^2\Big) + \frac{\nu}{\beta}. \quad (4.38)$$

Here the last inequality follows from Jensen's inequality, the fact that $\log\xi - \mathrm{Ei}(-\xi)$ is concave, and by taking the supremum over all distributions that satisfy the average-power constraint (2.23).

We will now make the following choices of the free parameters α and β:

$$\alpha \triangleq \alpha(\mathcal{E}) = \frac{\nu}{\log\Big(\big(1+d_{\max}^2\big)\mathcal{E}+\sigma^2\Big)} \quad (4.39)$$
$$\beta \triangleq \beta(\mathcal{E}) = \frac{1}{\alpha(\mathcal{E})} e^{\nu/\alpha(\mathcal{E})} \quad (4.40)$$

for some constant $\nu \geq 0$, which leads to the following asymptotic behavior:

$$\overline{\lim_{\mathcal{E}\uparrow\infty}} \left\{\log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) - \log\frac{1}{\alpha}\right\} = \log\left(1-e^{-\nu}\right); \quad (4.41)$$
$$\overline{\lim_{\mathcal{E}\uparrow\infty}}\, \alpha\Big(\log\beta - \log\sigma^2 + \gamma\Big) = \nu; \quad (4.42)$$
$$\overline{\lim_{\mathcal{E}\uparrow\infty}} \left\{\frac{1}{\beta}\Big(\big(1+d_{\max}^2\big)\mathcal{E}+\sigma^2\Big) + \frac{\nu}{\beta}\right\} = 0; \quad (4.43)$$
$$\overline{\lim_{\mathcal{E}\uparrow\infty}} \left\{\log\frac{1}{\alpha} - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} = -\log\nu. \quad (4.44)$$

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

Hence, we have derived the following upper bound on the fading number of a Gaussian MAC:

$$\chi_{\mathrm{MAC}} = \lim_{\mathcal{E}\to\infty} \left\{\mathsf{C}(\mathcal{E}) - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} \quad (4.45)$$
$$\leq \lim_{\mathcal{E}\to\infty} \Bigg\{ \sup_{Q_{X^{(1)}}\cdot Q_{X^{(2)}}} \Bigg\{ \log \mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2+\left|X^{(2)}\right|^2}\right] - \mathrm{Ei}\left(-\mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2+\left|X^{(2)}\right|^2}\right]\right) - 1 \Bigg\}$$
$$+ \epsilon_\nu + \alpha\big(\log\beta - \log\sigma^2 + \gamma\big) + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right)$$
$$+ \frac{1}{\beta}\Big(\big(1+d_{\max}^2\big)\mathcal{E}+\sigma^2\Big) + \frac{\nu}{\beta} - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right) \Bigg\} \quad (4.46)$$
$$\leq \lim_{\mathcal{E}\to\infty} \sup_{Q_{X^{(1)}}\cdot Q_{X^{(2)}}} \Bigg\{ \log \mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2+\left|X^{(2)}\right|^2}\right] - \mathrm{Ei}\left(-\mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2+\left|X^{(2)}\right|^2}\right]\right) - 1 \Bigg\}$$
$$+ \epsilon_\nu + \nu + \log\left(1-e^{-\nu}\right) - \log\nu, \quad (4.47)$$

where the supremum is over distributions $Q_{X^{(1)}} \cdot Q_{X^{(2)}}$ that satisfy the average-power constraint.

By letting ν tend to zero which makes sure that $\epsilon_\nu \to 0$ as can be seen from (4.25) we finally get the following bound:

Theorem 16. *The fading number of a two-user Gaussian fading MAC as defined in* (4.5) *is upper-bounded as follows:*

$$\chi_{\text{MAC}} \leq \lim_{\mathcal{E}\to\infty} \sup_{Q_{X^{(1)}} \cdot Q_{X^{(2)}}} \left\{ \log \mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2 + \left|X^{(2)}\right|^2} \right] - \operatorname{Ei}\left(-\mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2 + \left|X^{(2)}\right|^2} \right] \right) - 1 \right\}. \tag{4.48}$$

The problem is therefore reduced to find a bound on the expression

$$\mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2 + \left|X^{(2)}\right|^2} \right]. \tag{4.49}$$

First note the following:

$$\sup_{Q_{X^{(1)}} \cdot Q_{X^{(2)}}} \mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2 + \left|X^{(2)}\right|^2} \right] \leq \sup_{x^{(1)},x^{(2)}} \frac{\left|d^{(1)}x^{(1)} + d^{(2)}x^{(2)}\right|^2}{\left|x^{(1)}\right|^2 + \left|x^{(2)}\right|^2} \tag{4.50}$$

$$= \sup_{\mathbf{x}} \frac{\left|\mathbf{d}^{\mathsf{T}}\mathbf{x}\right|^2}{\|\mathbf{x}\|^2} \tag{4.51}$$

$$= \sup_{\mathbf{x}} \frac{\mathbf{x}\left(\mathbf{d}^{*}\mathbf{d}^{\mathsf{T}}\right)\mathbf{x}^{\dagger}}{\|\mathbf{x}\|^2} \tag{4.52}$$

$$= \lambda_{\max}\left(\mathbf{d}^{*}\mathbf{d}^{\mathsf{T}}\right) \tag{4.53}$$

where $\lambda_{\max}\left(\mathbf{d}^{*}\mathbf{d}^{\mathsf{T}}\right)$ denotes the maximum eigenvalue of the matrix $\mathbf{d}^{*}\mathbf{d}^{\mathsf{T}}$. Here the last equality follows from the Rayleigh-Ritz Theorem [25, Theorem 4.2.2]. The maximum eigenvalue of $\mathbf{d}^{*}\mathbf{d}^{\mathsf{T}}$ can be easily computed to be $\|\mathbf{d}\|^2$. Hence,

$$\sup_{Q_{X^{(1)}} \cdot Q_{X^{(2)}}} \mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2 + \left|X^{(2)}\right|^2} \right] \leq \left|d^{(1)}\right|^2 + \left|d^{(2)}\right|^2. \tag{4.54}$$

Note that this maximum can be achieved if cooperation is allowed: choose $\mathbf{X}$ as

$$\mathbf{X} = \frac{\mathbf{d}}{\|\mathbf{d}\|} \cdot \tilde{X} \tag{4.55}$$

where $\tilde{X}$ is a circularly symmetric input with a magnitude such that

$$\log|\tilde{X}| \sim \mathcal{U}\left([\log\log\mathcal{E}, \log\mathcal{E}]\right) \tag{4.56}$$

(see the corresponding results of MISO fading, [2], [7]).

This agrees with (4.16) and (4.17).

4.3 Loose Bound On MAC Fading Number

Unfortunately, the upper bound (4.54) is also achievable for independent $X^{(1)} \perp\!\!\!\perp X^{(2)}$:

Lemma 17. *Let $X^{(1)} \perp\!\!\!\perp X^{(2)}$ and assume $\mathsf{E}\bigl[X^{(1)}\bigr] = \mathsf{E}\bigl[X^{(2)}\bigr] = 0$. Moreover, assume*

$$\mathsf{E}\Bigl[\bigl|X^{(1)}\bigr|^2 + \bigl|X^{(2)}\bigr|^2\Bigr] \le \mathcal{E}. \tag{4.57}$$

Then

$$\lim_{\mathcal{E}\to\infty} \sup_{Q_{X^{(1)}}\cdot Q_{X^{(2)}}} \mathsf{E}\left[\frac{\bigl|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\bigr|^2}{\bigl|X^{(1)}\bigr|^2 + \bigl|X^{(2)}\bigr|^2}\right] = \bigl|d^{(1)}\bigr|^2 + \bigl|d^{(2)}\bigr|^2. \tag{4.58}$$

Proof. The right-hand side of (4.58) is for sure an upper bound on the left-hand side because of (4.54). We now show that it is also a lower bound. To that goal choose the following distributions on the inputs:

$$X^{(i)} = \begin{cases} -d^{(i)*} & \text{with probability } p^{(i)} \triangleq \frac{\mathcal{E}}{\mathcal{E}+2|d^{(i)}|^2}, \\ \frac{d^{(i)*}p^{(i)}}{1-p^{(i)}} & \text{with probability } 1-p^{(i)}, \end{cases} \tag{4.59}$$

for $i = 1, 2$. This choice is zero-mean and of variance $\frac{\mathcal{E}}{2}$ and leads to

$$\lim_{\mathcal{E}\to\infty} \mathsf{E}\left[\frac{\bigl|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\bigr|^2}{\bigl|X^{(1)}\bigr|^2 + \bigl|X^{(2)}\bigr|^2}\right] = \bigl|d^{(1)}\bigr|^2 + \bigl|d^{(2)}\bigr|^2. \tag{4.60}$$

For more details see Appendix E. □

However, note that the choice of $Q_{X^{(1)}} \cdot Q_{X^{(2)}}$ given in the proof of Lemma 17 is performing very poorly when considering the total mutual information instead of only expression (4.49). Indeed, one can show that the upper bound (4.54) is loose:

Theorem 18. *The fading number of a two-user Gaussian fading MAC is strictly smaller than the corresponding MISO Gaussian fading:*

$$\chi_{\text{MAC}} < \log\Bigl(\bigl|d^{(1)}\bigr|^2 + \bigl|d^{(2)}\bigr|^2\Bigr) - \text{Ei}\Bigl(-\bigl|d^{(1)}\bigr|^2 - \bigl|d^{(2)}\bigr|^2\Bigr) - 1. \tag{4.61}$$

Proof. For simplicity, firstly consider the case where $d^{(1)} = d^{(2)} = d$:

$$\begin{aligned} \frac{\bigl|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\bigr|^2}{\bigl|X^{(1)}\bigr|^2 + \bigl|X^{(2)}\bigr|^2} &= |d|^2 \frac{\bigl|X^{(1)} + X^{(2)}\bigr|^2}{\bigl|X^{(1)}\bigr|^2 + \bigl|X^{(2)}\bigr|^2 + \sigma^2} && (4.62) \\ &= |d|^2 \frac{\bigl|X^{(1)}\bigr|^2 + \bigl|X^{(2)}\bigr|^2 + 2\bigl|X^{(1)}\bigr| \cdot \bigl|X^{(2)}\bigr| \cos(\Phi_1 - \Phi_2)}{\bigl|X^{(1)}\bigr|^2 + \bigl|X^{(2)}\bigr|^2} && (4.63) \\ &= |d|^2 \left(1 + \frac{2|X|\cos\Phi}{1+|X|^2}\right) && (4.64) \end{aligned}$$

where we have introduced Φ_1 and Φ_2 to denote the phase of $X^{(1)}$ and $X^{(2)}$, respectively, and $|X| = \bigl|X^{(1)}\bigr| / \bigl|X^{(2)}\bigr|$ and $\Phi = \Phi_1 - \Phi_2$. Note that the maximum of $2|d|^2$ can only be achieved if $|X| = 1$ and $\Phi = 0$. But since the two users are independent, this is only possible if $X^{(1)}$ and $X^{(2)}$ become almost deterministic.

The same argument also holds if one considers the general case of general $d^{(1)}$, $d^{(2)}$.

Next, let's investigate the situation of almost deterministic inputs: Let $\boldsymbol{\xi} = (\xi, \xi)^{\mathsf{T}} \in \mathbb{C}^2$ be a deterministic input vector, fix $\delta > 0$ and define

$$\mathcal{V}_\delta \triangleq \left\{ \mathbf{x} : \frac{|x^{(1)}|}{|\xi|} \in (1-\delta, 1+\delta), \frac{|x^{(2)}|}{|\xi|} \in (1-\delta, 1+\delta), \right. \\ \left. (\angle x^{(1)} - \angle \xi) \in (-\delta, \delta), (\angle x^{(2)} - \angle \xi) \in (-\delta, \delta) \right\}. \quad (4.65)$$

Let $p \triangleq \Pr(\mathcal{V}_\delta)$. Hence, to achieve the MISO fading number in the MAC situation we need that $p \to 1$ as $\mathcal{E} \to \infty$.

We will now show that in this case capacity actually tends to zero. To that goal we will use the dual-based technique of finding upper bounds on capacity introduced in [7] and [2]. We choose as an output distribution

$$R(y) = pW(y|\boldsymbol{\xi}) + (1-p)R_\Gamma(y) \quad (4.66)$$

where $R_\Gamma(\cdot)$ denotes the output distribution that has been used to derive the bound given in Lemma 5, and where $W(\cdot|\cdot)$ denotes the conditional channel law. Then, for $\mathbf{x} \in \mathcal{V}_\delta$

$$\begin{aligned} &D\big(W(\cdot|\mathbf{x}) \big\| R(\cdot)\big) \\ &\quad = \int W(y|\mathbf{x}) \log \frac{W(y|\mathbf{x})}{pW(y|\boldsymbol{\xi}) + (1-p)R_\Gamma(y)}\,\mathrm{d}y &(4.67)\\ &\quad \le \int W(y|\mathbf{x}) \log \frac{W(y|\mathbf{x})}{pW(y|\boldsymbol{\xi})}\,\mathrm{d}y &(4.68)\\ &\quad = -\log p - H(Y|\mathbf{X}=\mathbf{x}) - \int W(y|\mathbf{x}) \log \frac{1}{\pi(\|\boldsymbol{\xi}\|^2+\sigma^2)} e^{-\frac{|y-\mathbf{d}^\dagger\boldsymbol{\xi}|^2}{\|\boldsymbol{\xi}\|^2+\sigma^2}}\,\mathrm{d}y &(4.69)\\ &\quad = -\log p - \log\frac{\|\mathbf{x}\|^2+\sigma^2}{\|\boldsymbol{\xi}\|^2+\sigma^2} + \frac{\|\mathbf{x}\|^2+\sigma^2}{\|\boldsymbol{\xi}\|^2+\sigma^2} - 1 + \frac{|\mathbf{d}^\dagger\mathbf{x} - \mathbf{d}^\dagger\boldsymbol{\xi}|^2}{\|\boldsymbol{\xi}\|^2+\sigma^2}. &(4.70)\end{aligned}$$

For $\mathbf{x} \notin \mathcal{V}_\delta$,

$$\begin{aligned} &D\big(W(\cdot|\mathbf{x}) \big\| R(\cdot)\big) \\ &\quad = \int W(y|\mathbf{x}) \log \frac{W(y|\mathbf{x})}{pW(y|\boldsymbol{\xi}) + (1-p)R_\Gamma(y)}\,\mathrm{d}y &(4.71)\\ &\quad \le \int W(y|\mathbf{x}) \log \frac{W(y|\mathbf{x})}{(1-p)R_\Gamma(y)}\,\mathrm{d}y &(4.72)\\ &\quad \le -\log(1-p) - h\big(Y \,\big|\, \mathbf{X}=\mathbf{x}\big) + \log\pi + \alpha\log\beta \\ &\qquad + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) + (1-\alpha)\mathsf{E}\big[\log\big(|Y|^2+\nu\big) \,\big|\, \mathbf{X}=\mathbf{x}\big] \\ &\qquad + \frac{1}{\beta}\mathsf{E}\big[|Y|^2 \,\big|\, \mathbf{X}=\mathbf{x}\big] + \frac{\nu}{\beta} &(4.73)\end{aligned}$$

analogously to Lemma 5.

Hence, we have

$$\begin{aligned} I(\mathbf{X};Y) &\le \mathsf{E}\big[D\big(W(\cdot|\mathbf{X}) \big\| R(\cdot)\big)\big] &(4.74)\\ &= p\mathsf{E}\big[D\big(W(\cdot|\mathbf{X}) \big\| R(\cdot)\big) \,\big|\, \mathbf{X} \in \mathcal{V}_\delta\big] \\ &\quad + (1-p)\mathsf{E}\big[D\big(W(\cdot|\mathbf{X}) \big\| R(\cdot)\big) \,\big|\, \mathbf{X} \notin \mathcal{V}_\delta\big] &(4.75)\end{aligned}$$

$$
\begin{aligned}
&\leq -p\log p - p\mathsf{E}\left[\log\frac{\|\mathbf{X}\|^2+\sigma^2}{\|\boldsymbol{\xi}\|^2+\sigma^2}\,\middle|\,\mathbf{X}\in\mathcal{V}_\delta\right] + p\mathsf{E}\left[\frac{\|\mathbf{X}\|^2+\sigma^2}{\|\boldsymbol{\xi}\|^2+\sigma^2}\,\middle|\,\mathbf{X}\in\mathcal{V}_\delta\right]\\
&\quad - p + p\mathsf{E}\left[\frac{|\mathbf{d}^\dagger\mathbf{x}-\mathbf{d}^\dagger\boldsymbol{\xi}|^2}{\|\boldsymbol{\xi}\|^2+\sigma^2}\,\middle|\,\mathbf{X}\in\mathcal{V}_\delta\right] - (1-p)\log(1-p)\\
&\quad - (1-p)h\big(Y\,\big|\,\mathbf{X},\mathbf{X}\notin\mathcal{V}_\delta\big) + (1-p)\log\pi + (1-p)\alpha\log\beta\\
&\quad + (1-p)\log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) + (1-p)(1-\alpha)\mathsf{E}\big[\log\big(|Y|^2+\nu\big)\,\big|\,\mathbf{X}\notin\mathcal{V}_\delta\big]\\
&\quad + (1-p)\frac{1}{\beta}\mathsf{E}\big[|Y|^2\,\big|\,\mathbf{X}\notin\mathcal{V}_\delta\big] + (1-p)\frac{\nu}{\beta}. \qquad (4.76)
\end{aligned}
$$

To simplify this bound note that every $\mathbf{x}\in\mathcal{V}_\delta$ can be easily upper- or lower-bounded by means of δ. Moreover, note

$$
h\big(Y\,\big|\,\mathbf{X},\mathbf{X}\notin\mathcal{V}_\delta\big) = \log\pi + 1 + \mathsf{E}\big[\log\big(\|\mathbf{X}\|^2+\sigma^2\big)\,\big|\,\mathbf{X}\notin\mathcal{V}_\delta\big], \qquad (4.77)
$$

$$
\mathsf{E}\big[|Y|^2\,\big|\,\mathbf{X}\notin\mathcal{V}_\delta\big] = \mathsf{E}\big[\|\mathbf{X}\|^2+\sigma^2\,\big|\,\mathbf{X}\notin\mathcal{V}_\delta\big] \leq \frac{\mathcal{E}}{(1-p)} + \sigma^2, \qquad (4.78)
$$

$$
\begin{aligned}
\mathsf{E}\big[\log|Y|^2\,\big|\,\mathbf{X}\notin\mathcal{V}_\delta\big] &= \mathsf{E}\big[\log\big(\|\mathbf{X}\|^2+\sigma^2\big)\,\big|\,\mathbf{X}\notin\mathcal{V}_\delta\big]\\
&\quad + \mathsf{E}\left[\log\frac{|\mathbf{d}^\dagger\mathbf{X}|^2}{\|\mathbf{X}\|^2+\sigma^2} - \mathrm{Ei}\left(-\frac{|\mathbf{d}^\dagger\mathbf{X}|^2}{\|\mathbf{X}\|^2+\sigma^2}\right)\,\middle|\,\mathbf{X}\notin\mathcal{V}_\delta\right] \qquad (4.79)\\
&\leq \mathsf{E}\big[\log\big(\|\mathbf{X}\|^2+\sigma^2\big)\,\big|\,\mathbf{X}\notin\mathcal{V}_\delta\big] + \log\|\mathbf{d}\|^2 - \mathrm{Ei}\big(-\|\mathbf{d}\|^2\big), \qquad (4.80)\\
\mathsf{E}\big[\log|Y|^2\,\big|\,\mathbf{X}\notin\mathcal{V}_\delta\big] &= \mathsf{E}\big[\log\big(\|\mathbf{X}\|^2+\sigma^2\big)\,\big|\,\mathbf{X}\notin\mathcal{V}_\delta\big]\\
&\quad + \mathsf{E}\left[\log\frac{|\mathbf{d}^\dagger\mathbf{X}|^2}{\|\mathbf{X}\|^2+\sigma^2} - \mathrm{Ei}\left(-\frac{|\mathbf{d}^\dagger\mathbf{X}|^2}{\|\mathbf{X}\|^2+\sigma^2}\right)\,\middle|\,\mathbf{X}\notin\mathcal{V}_\delta\right] \qquad (4.81)\\
&\geq \log\sigma^2 - \gamma. \qquad (4.82)
\end{aligned}
$$

Finally, we make the following choice of the free parameters:

$$
\nu = 0, \qquad (4.83)
$$

$$
\beta = \frac{1}{\alpha}\left(\frac{\mathcal{E}}{1-p} + \sigma^2\right), \qquad (4.84)
$$

$$
\alpha = \frac{1}{\log\mathcal{E}}, \qquad (4.85)
$$

and note that $\log\Gamma(\alpha)\geq 0$ for all $\alpha>0$. Then we get

$$
\begin{aligned}
I(\mathbf{X};Y) &\leq -p\log p - p\log\frac{2(1-\delta)^2|\xi|^2+\sigma^2}{2|\xi|^2+\sigma^2} + p\frac{2(1+\delta)^2|\xi|^2+\sigma^2}{2|\xi|^2+\sigma^2}\\
&\quad - p + p\frac{|d^{(1)}((1+\delta)\xi e^{\mathrm{i}\delta}-\xi) + d^{(2)}((1+\delta)\xi e^{\mathrm{i}\delta}-\xi)|^2}{2|\xi|^2+\sigma^2}\\
&\quad - (1-p)\log(1-p) - (1-p) - (1-p)\alpha\log\alpha\\
&\quad + (1-p)\alpha\log\left(\frac{\mathcal{E}}{1-p}+\sigma^2\right) + \log\Gamma(\alpha)\\
&\quad + (1-p)\big(\log\|\mathbf{d}\|^2 - \mathrm{Ei}\big(-\|\mathbf{d}\|^2\big)\big) - (1-p)\alpha(\log\sigma^2-\gamma)\\
&\quad + (1-p)\alpha \qquad (4.86)
\end{aligned}
$$

and

$$
\begin{aligned}
\chi_{\mathrm{MAC}} &= \lim_{\mathcal{E}\to\infty}\left\{\mathsf{C}(\mathcal{E}) - \log\left(1+\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} \qquad (4.87)\\
&\leq -\log\frac{2(1-\delta)^2|\xi|^2+\sigma^2}{2|\xi|^2+\sigma^2} + \frac{2(1+\delta)^2|\xi|^2+\sigma^2}{2|\xi|^2+\sigma^2}
\end{aligned}
$$

$$- 1 + \frac{|d^{(1)}((1+\delta)\xi e^{\mathrm{i}\delta} - \xi) + d^{(2)}((1+\delta)\xi e^{\mathrm{i}\delta} - \xi)|^2}{2|\xi|^2 + \sigma^2} \tag{4.88}$$

$$\to 0 \qquad \text{as } \delta \to 0. \tag{4.89}$$

Here we have used that $\log \Gamma(\alpha) + \log \alpha \to 0$ as $\alpha \to 0$. □

4.4 Circular Symmetry

Unfortunately, we have not yet been able to prove that the optimal input for both users is circularly symmetric. The problem lies in the fact that the users need to be independent of each other, *i.e.*, the proof of Lemma 6 breaks down since there each antenna has the same circularly symmetric random phase.

However, we are able to prove the following proposition:

Proposition 19. *Assume that at least one user uses an input with*

$$\mathsf{E}\left[e^{\mathrm{i}\Phi} \,\middle|\, |X|\right] = 0, \tag{4.90}$$

where Φ denotes the phase and $|X|$ the magnitude of the input. Then the Gaussian MAC fading number is given by

$$d_{\text{MAC}} = \max\left\{\left|d^{(1)}\right|, \left|d^{(2)}\right|\right\}, \tag{4.91}$$

i.e., *the high-SNR capacity of a MAC corresponds to the high-SNR capacity of the better of the two channels if the other user is switched off.*

In particular this holds if at least one user uses an input that is circularly symmetric.

Proof. Note that (4.90) implies the following:

$$\begin{aligned}
&\mathsf{E}\left[X^{(1)}\left(X^{(2)}\right)^* \,\middle|\, \left|X^{(1)}\right|, \left|X^{(2)}\right|\right] \\
&\quad = \mathsf{E}\left[\left|X^{(1)}\right| \cdot e^{\mathrm{i}\Phi^{(1)}} \cdot \left|X^{(2)}\right| \cdot e^{-\Phi^{(2)}} \,\middle|\, \left|X^{(1)}\right|, \left|X^{(2)}\right|\right] &(4.92)\\
&\quad = \left|X^{(1)}\right| \cdot \left|X^{(2)}\right| \cdot \mathsf{E}\left[e^{\mathrm{i}\Phi^{(1)}} e^{-\Phi^{(2)}} \,\middle|\, \left|X^{(1)}\right|, \left|X^{(2)}\right|\right] &(4.93)\\
&\quad = \left|X^{(1)}\right| \cdot \left|X^{(2)}\right| \cdot \mathsf{E}\left[e^{\mathrm{i}\Phi^{(1)}} \,\middle|\, \left|X^{(1)}\right|\right] \cdot \mathsf{E}\left[e^{-\Phi^{(2)}} \,\middle|\, \left|X^{(2)}\right|\right] &(4.94)\\
&\quad = 0. &(4.95)
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2 + \left|X^{(2)}\right|^2}\right] \\
&\quad = \mathsf{E}\left[\mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2 + \left|X^{(2)}\right|^2} \,\middle|\, \left|X^{(1)}\right|, \left|X^{(2)}\right|\right]\right] &(4.96)\\
&\quad = \mathsf{E}\left[\frac{\mathsf{E}\left[\left|d^{(1)}X^{(1)} + d^{(2)}X^{(2)}\right|^2 \,\middle|\, \left|X^{(1)}\right|, \left|X^{(2)}\right|\right]}{\left|X^{(1)}\right|^2 + \left|X^{(2)}\right|^2}\right] &(4.97)\\
&\quad = \mathsf{E}\left[\frac{1}{\left|X^{(1)}\right|^2 + \left|X^{(2)}\right|^2}\left(\left|d^{(1)}\right|^2\left|X^{(1)}\right|^2 + \left|d^{(2)}\right|^2\left|X^{(2)}\right|^2\right.\right. \\
&\qquad\qquad \left. + d^{(1)}\left(d^{(2)}\right)^* \mathsf{E}\left[X^{(1)}\left(X^{(2)}\right)^* \,\middle|\, \left|X^{(1)}\right|, \left|X^{(2)}\right|\right]\right.
\end{aligned}$$

$$+ \left(d^{(1)}\right)^{*} d^{(2)} \mathsf{E}\left[\left(X^{(1)}\right)^{*} X^{(2)} \,\middle|\, \left|X^{(1)}\right|, \left|X^{(2)}\right| \right] \Bigg) \Bigg] \tag{4.98}$$

$$= \mathsf{E}\left[\frac{\left|d^{(1)}\right|^2 \left|X^{(1)}\right|^2 + \left|d^{(2)}\right|^2 \left|X^{(2)}\right|^2}{\left|X^{(1)}\right|^2 + \left|X^{(2)}\right|^2} \right] \tag{4.99}$$

$$\leq \sup_{\left|x^{(1)}\right|, \left|x^{(2)}\right|} \frac{\left|d^{(1)}\right|^2 \left|x^{(1)}\right|^2 + \left|d^{(2)}\right|^2 \left|x^{(2)}\right|^2}{\left|x^{(1)}\right|^2 + \left|x^{(2)}\right|^2} \tag{4.100}$$

$$= \sup_{t \geq 0} \frac{\left|d^{(1)}\right|^2 + \left|d^{(2)}\right|^2 t}{1 + t} \tag{4.101}$$

$$= \max\left\{ \left|d^{(1)}\right|^2, \left|d^{(2)}\right|^2 \right\}, \tag{4.102}$$

where the last equality follows from the fact that $\frac{\left|d^{(1)}\right|^2 + \left|d^{(2)}\right|^2 t}{1+t}$ is monotonically increasing or decreasing depending on whether $\left|d^{(1)}\right| \geq \left|d^{(2)}\right|$ or not.

The theorem now follows from the fact that this upper bound can be achieved by a simple scheme where the user with the worse channel is switched off.

To show that if one user uses a circularly symmetric input, then the other user also should use a circularly symmetric input, we rely on the corresponding result from the single-user situation: in a single-user situation the optimal input is circularly symmetric as shown in Lemma 6. Hence, if the user with the worse channel uses a circularly symmetric input, then the optimum can only be achieved if the magnitude of this user is deterministically equal to zero and the other user then needs to use a circularly symmetric input in order to achieve the single-user capacity. □

Chapter 5

Discussion & Conclusions

The topic under study in this project is the theoretical upper limit on the rate of reliable transmission that can be achieved over a wireless, mobile communication system. This upper limit is characterized by the *channel capacity* in the case of a single-user setup or the *capacity region* in the situation of multiple users. We assume an OFDM system such that for our analysis we can neglect inter-symbol interference, but only need to deal with the problem of *fading*. We define a discrete-time version of a corresponding channel model where we try to make as few assumptions as possible in order to keep the results as general as possible. Hence, we do not make any particular assumption about the distribution of the fading process (*i.e.*, not necessarily Gaussian!), allow in general for an arbitrary number of antennas both at transmitter and receiver, and consider multiple users. We also allow memory, both over time and space, *i.e.*, the different fading coefficients of different antennas and at different times might be dependent. The only restriction applied is the so-called *regularity assumption*. To explain this assumption slightly imprecisely in an engineering way we might say that we ask the fading process to be fully random in the sense that even with the knowledge of all past fading realizations and of all fading realizations of neighboring antennas one cannot predict the actual value of the fading precisely. The prediction error might be very small once one knows the past and the neighboring fading values, but it is still non-zero. Considering the fact that in reality no measurement is absolutely perfect this assumption seems to be realistic.

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

$$\chi \triangleq \lim_{\text{SNR}\to\infty} \Big\{ \mathsf{C}(\text{SNR}) - \log\big(1+\log\big(1+\text{SNR}\big)\big) \Big\}, \tag{5.1}$$

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

$$\mathsf{C}(\text{SNR}) = \log(1+\log(1+\text{SNR})) + \chi + o(1), \tag{5.2}$$

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

We motivate our interest in the high-SNR regime and the fading number by arguing that the knowledge of the high-SNR capacity behavior has a very strong impact on practical considerations when designing such mobile communication systems, even if one does never operate a system at such high SNR levels. The argument is based on the observation that the high-SNR behavior (5.2) is extremely poor because the capacity grows only double-logarithmically in the SNR. in other words this means that in the regime where the approximation

$$\mathsf{C}(\textsc{snr}) \approx \log(1+\log(1+\textsc{snr})) + \chi \tag{5.3}$$

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

To answer this question the fading number can give an interesting answer: pulling ourselves by our bootstraps, let us consider for the moment that (5.3) starts to be valid for an SNR somewhere in the range[1] of 30 to 80 dB. In spite of this rather large range the function $\log(1+\log(1+\textsc{snr}))$ will vary only between 2 and 3 nats. Hence, the capacity will vary between $2+\chi$ and $3+\chi$ nats. Therefore, we can conclude that once the capacity is appreciably above $\chi+2$ nats, the approximation (5.3) is likely to be valid. Therefore, the fading number can be seen as an indicator of the maximum rate at which power efficient communication is possible on the channel.

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

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

In the process of the study we further restrict the channel model to some special cases, however, always keeping in mind the ultimate goal of an analysis of the unrestricted case. We here basically study two special cases:

- a single-user memoryless MIMO fading channel, and
- a two-user memoryless SISO Gaussian multiple-access fading channel.

5.1 A Single-User Memoryless MIMO Fading Channel

In the former we are able to derive the fading number of a MIMO fading channel of general fading law including spatial, but without temporal memory. Since the fading number is the second term after the double-logarithmic term of the high-SNR expansion of channel capacity, this means that we precisely specify the behavior of the channel capacity asymptotically when the power grows to infinity. We further show that the asymptotic capacity can be achieved by an input that consists of the product of two independent random quantities: a circularly symmetric random unit vector (the *direction*) and a non-negative (*i.e.*, real) random variable (the *magnitude*). The distribution of the random direction is chosen such as to maximize the fading number and therefore depends on the particular law of the fading process. The distribution of the magnitude is chosen such that

$$\log R^2 \sim \mathcal{U}\left([\log\log\mathcal{E}, \log\mathcal{E}]\right). \tag{5.4}$$

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.

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

The derivation of this result is based on three main techniques or observations: firstly, we know that the capacity achieving input distribution must *escape to infinity* which means that at high SNR no finite energy input symbols should be used anymore. Secondly, we use the dual-based approach of deriving upper bounds to channel capacity as firstly described in [26] without application and as firstly applied in [7] and [2]. Finally we prove that in general the capacity achieving input distribution must be circularly symmetric. The proof only relies on the fact that the additive noise is circularly symmetric so that this result holds in a most general setting. To understand this result note that a circularly symmetric input can pack more information into the phase than any other phase distribution. Since the additive noise does not favor any direction there is no need to protect some direction more than others and therefore it is optimal to pack as much information into the phase as possible.

Note that this result has been actually known for a long time, but it seems that part of the literature has not really been aware of it. It is mentioned for example in [7, Section IV.D.6].

We then derive some new results for the important special situation of Gaussian fading. For the case of a scalar line-of-sight matrix (3.53) assuming at least as many transmit as receive antennas $n_{\mathrm{R}} \leq n_{\mathrm{T}}$ we have been able to state the fading number precisely:

$$\chi = n_{\mathrm{R}} g_{n_{\mathrm{R}}}\left(|d|^2\right) - n_{\mathrm{R}} - \log \Gamma(n_{\mathrm{R}}), \tag{5.5}$$

where $g_m(\cdot)$ denotes the expected value of a non-central chi-square random variable [24]. We see that the asymptotic capacity only depends on the number of receive antennas and is growing proportionally to $n_{\mathrm{R}} \log |d|^2$.

For a general line-of-sight matrix we derived an upper bound (3.66) that grows like $\min\{n_{\mathrm{R}}, n_{\mathrm{T}}\} \log \delta^2$ where δ^2 is a certain kind of average (3.67) of all singular values of the line-of-sight matrix. For $n_{\mathrm{R}} \leq n_{\mathrm{T}}$ we also specify some lower bounds.

As a byproduct based on these results, we have been able to give a new upper bound on the function $g_m(\cdot)$.

We very much hope that the result derived here will be helpful in our attempts of deriving the fading number for the MIMO case including temporal memory.

5.2 A Two-User Memoryless SISO Gaussian Multiple-Access Fading Channel

In the second case of a two-user memoryless SISO Gaussian multiple-access fading channel, we have not yet been able to derive the asymptotic capacity precisely. The difficulty lies in the fact that the inputs of the two users must be independent (in contrast to the MISO situation where the two antennas at the transmitter can cooperate). We derive some upper and lower bounds to the fading number which relate the MAC situation to the situation of MISO fading with two antennas at the transmitter. Basically the MAC is a case between the pessimistic situation where the sum-rate capacity is identical to the single-user capacity of the user with the best channel (which means that the other user cannot transmit at all when the highest total sum rate shall be achieved) and the optimistic case where the same sum rate can be achieved as if cooperation between the users were allowed.

We show that the latter optimistic case cannot be achieved, *i.e.*, the MAC fading number is strictly smaller than the corresponding MISO fading number. Moreover, we show that when at least one user uses a circular symmetric input, then the fading number is reduced to the pessimistic situation where the user with the worse channel must be switched off. It then follows from the results of single-user channels that the optimum input must be circularly symmetric.

To derive the fading number precisely, some more future work is required. One promising approach is the attempt of proving that a circular symmetric input is also optimal in the MAC situation, which would then give the fading number based on the above described result.

5.3 Outlook

As mentioned there are two questions that follow immediately from the results presented here and that we would like to solve:

- What is the exact expression of the fading number of a general single-user MIMO fading channel with memory?
- Is a circularly symmetric input optimal in the situation of a two-user SISO Gaussian multiple-access fading channel? If not, what is the exact expression of the fading number in this situation?

Both questions are rather difficult: in the first problem we probably will need to prove that a stationary channel model has a capacity achieving input distribution that is stationary. Note that so far the fading number of channels with memory is only known in cases with only one antenna at the transmitter. In that situation it turned out that IID inputs are sufficient in order to achieve the capacity at high SNR. We suspect that this might not be the case anymore once we have several antennas at the transmitter. The reason for this is that the channel state estimation in case of only one transmit antenna is relatively simple: the receiver decodes the input and then divides the received vector by this decoded value. There is no need for additional structure in the input. However, if there are several transmitter antennas one cannot get all fading values directly from the knowledge of the received vector and the decoded transmitted vector. Hence, additional structure in the input might be needed to get a better estimate.

The main problem in the second question lies in the fact that the two users must be independent. This constraints makes many techniques and approaches that have been successful in the single-user situation useless. At the moment the most promising path seems to be an attempt in proving that circular symmetry is the optimal choice for both users.

Appendix A

Proof of Lemma 6

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

$$\begin{aligned}
&\frac{1}{n} I\bigl(\mathbf{X}_1^n; \mathbf{Y}_1^n\bigr) \\
&\quad= \frac{1}{n} I\bigl(\mathbf{X}_1^n; \mathbf{Y}_1^n \bigm| \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\bigr) && \text{(A.1)}\\
&\quad= \frac{1}{n} I\bigl(\{\mathbf{X}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n; \{\mathbf{Y}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n \bigm| \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\bigr) && \text{(A.2)}\\
&\quad= \frac{1}{n} I\bigl(\{\mathbf{X}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n; \{\mathbb{H}_\ell \mathbf{X}_\ell e^{\mathrm{i}\Theta_\ell} + \mathbf{Z}_\ell\}_{\ell=1}^n \bigm| \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\bigr) && \text{(A.3)}\\
&\quad= \frac{1}{n} I\bigl(\tilde{\mathbf{X}}_1^n; \{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \bigm| \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\bigr) && \text{(A.4)}\\
&\quad= \frac{1}{n} h\bigl(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \bigm| \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\bigr) - \frac{1}{n} h\bigl(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \bigm| \tilde{\mathbf{X}}_1^n, \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\bigr) && \text{(A.5)}\\
&\quad= \frac{1}{n} h\bigl(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \bigm| \{e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^n\bigr) - \frac{1}{n} h\bigl(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \bigm| \tilde{\mathbf{X}}_1^n\bigr) && \text{(A.6)}\\
&\quad\le \frac{1}{n} h\bigl(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n\bigr) - \frac{1}{n} h\bigl(\{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n \bigm| \tilde{\mathbf{X}}_1^n\bigr) && \text{(A.7)}\\
&\quad= \frac{1}{n} I\bigl(\tilde{\mathbf{X}}_1^n; \{\mathbb{H}_\ell \tilde{\mathbf{X}}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^n\bigr). && \text{(A.8)}
\end{aligned}$$

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

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

Appendix B

Proof of Theorem 8

The proof of Theorem 8 consists of two parts: in a first part we show that the right-hand side of (3.11) is an upper bound to the fading number and in a second part we show that this upper bound can actually be achieved.

B.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 $\mathbf{X} = R \cdot \hat{\mathbf{X}}$.

Fix an arbitrary $\epsilon > 0$. Let $\{Q_{\mathbf{X}}\}_{\mathcal{E}}$ be a sequence of input distributions parameterized by the available power $\mathcal{E}$ that satisfy the average-power constraint (2.23) and that almost achieve capacity in the sense that

$$I(\mathbf{X}; \mathbf{Y}) \geq \mathsf{C}(\mathcal{E}) - \epsilon, \qquad \forall \mathcal{E} > 0. \tag{B.1}$$

Then we know from Lemma 4 that $\{Q_{\mathbf{X}}\}_{\mathcal{E}}$ must escape to infinity, *i.e.*, for an arbitrary $\mathcal{E}_0 \geq 0$

$$\lim_{\mathcal{E} \to \infty} Q_{\mathbf{X}}\bigl(R^2 < \mathcal{E}_0\bigr) = 0. \tag{B.2}$$

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

$$E \triangleq \begin{cases} 1 & \text{if } R^2 \geq \mathcal{E}_0, \\ 0 & \text{otherwise,} \end{cases} \tag{B.3}$$

and let

$$p \triangleq \Pr[E = 1] = \Pr\bigl[R^2 \geq \mathcal{E}_0\bigr]. \tag{B.4}$$

Then it follows from (B.2) that

$$\lim_{\mathcal{E} \to \infty} p = 1. \tag{B.5}$$

Using these definitions we start to upper-bound mutual information as follows:

$$\begin{aligned}
I(\mathbf{X}; \mathbf{Y}) &\leq I(\mathbf{X}, E; \mathbf{Y}) && \text{(B.6)} \\
&= I(E; \mathbf{Y}) + I\bigl(\mathbf{X}; \mathbf{Y} \bigm| E\bigr) && \text{(B.7)} \\
&= H(E) - H\bigl(E \bigm| \mathbf{Y}\bigr) + I\bigl(\mathbf{X}; \mathbf{Y} \bigm| E\bigr) && \text{(B.8)} \\
&\leq H(E) + I\bigl(\mathbf{X}; \mathbf{Y} \bigm| E\bigr) && \text{(B.9)} \\
&= H_{\mathrm{b}}(p) + pI\bigl(\mathbf{X}; \mathbf{Y} \bigm| E = 1\bigr) + (1-p)I\bigl(\mathbf{X}; \mathbf{Y} \bigm| E = 0\bigr) && \text{(B.10)} \\
&\leq H_{\mathrm{b}}(p) + I\bigl(\mathbf{X}; \mathbf{Y} \bigm| E = 1\bigr) + (1-p)\mathsf{C}_{\mathrm{IID}}(\mathcal{E}_0). && \text{(B.11)}
\end{aligned}$$

Here the first inequality follows from adding an additional term to mutual information; the subsequent equality from the chain rule; the subsequent equality from the definition of mutual information; then the subsequent inequality follows from the non-negativity of entropy (note that since E is binary, we do not have differential entropies!); and the final inequality follows by upper-bounding $p \leq 1$, the fact that mutual information is non-negative, and by upper-bounding the mutual information by the corresponding capacity where we note that an average-power constraint is less stringent than a peak-power constraint, *i.e.*, conditional on $R^2 < \mathcal{E}_0$ the mutual information can be at most the capacity for a peak-power constraint $\mathcal{E}_0$ which in turn is upper-bounded by $\mathsf{C}_{\text{IID}}(\mathcal{E}_0)$.

We then continue by bounding the second term in (B.11):

$$\begin{aligned}
I\big(\mathbf{X};\mathbf{Y}\,\big|\,E=1\big) &\leq I\big(\mathbf{X};\mathbf{Y},\mathbf{Z}\,\big|\,E=1\big) && \text{(B.12)}\\
&= I\big(\mathbf{X};\mathbb{H}\mathbf{X},\mathbf{Z}\,\big|\,E=1\big) && \text{(B.13)}\\
&= I\big(\mathbf{X};\mathbb{H}\mathbf{X}\,\big|\,E=1\big) + I\big(\mathbf{X};\mathbf{Z}\,\big|\,\mathbb{H}\mathbf{X},E=1\big) && \text{(B.14)}\\
&= I\big(\mathbf{X};\mathbb{H}\mathbf{X}\,\big|\,E=1\big) && \text{(B.15)}\\
&= I\left(\mathbf{X};\|\mathbb{H}\mathbf{X}\|,\frac{\mathbb{H}\mathbf{X}}{\|\mathbb{H}\mathbf{X}\|}\,\middle|\,E=1\right) && \text{(B.16)}\\
&= I\left(\mathbf{X};\|\mathbb{H}\mathbf{X}\|,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,E=1\right) && \text{(B.17)}\\
&= I\big(\mathbf{X};\|\mathbb{H}\mathbf{X}\|\,\big|\,E=1\big) + I\left(\mathbf{X};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\mathbf{X}\|,E=1\right) && \text{(B.18)}\\
&= I\big(\mathbf{X};\|\mathbb{H}\mathbf{X}\|,e^{\mathrm{i}\Theta}\,\big|\,E=1\big) - I\big(\mathbf{X};e^{\mathrm{i}\Theta}\,\big|\,\|\mathbb{H}\mathbf{X}\|,E=1\big)\\
&\quad + I\left(\mathbf{X};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\mathbf{X}\|,E=1\right) && \text{(B.19)}\\
&= I\big(\mathbf{X};\|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}\,\big|\,E=1\big) + I\left(\mathbf{X};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\mathbf{X}\|,E=1\right). && \text{(B.20)}
\end{aligned}$$

Here the first inequality 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; in (B.19) we use the chain rule and we introduce $e^{\mathrm{i}\Theta}$ that is independent of all the other random quantities and that is uniformly distributed on the complex unit circle; and the last equality follows from the independence of $e^{\mathrm{i}\Theta}$ from all other random quantities.

We next would like to apply Lemma 5 to the first of the two terms in (B.20), *i.e.*, we choose $\mathbf{S} = \mathbf{X}$ and $T = \|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}$:

$$\begin{aligned}
&I\big(\mathbf{X};\|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}\,\big|\,E=1\big)\\
&\quad\leq -h\big(\|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}\,\big|\,\mathbf{X},E=1\big) + \log\pi + \alpha\log\beta + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right)\\
&\qquad + (1-\alpha)\mathsf{E}\big[\log\big(\|\mathbb{H}\mathbf{X}\|^2+\nu\big)\,\big|\,E=1\big] + \frac{1}{\beta}\mathsf{E}\big[\|\mathbb{H}\mathbf{X}\|^2\,\big|\,E=1\big] + \frac{\nu}{\beta}, \qquad \text{(B.21)}
\end{aligned}$$

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

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

$$\begin{aligned}
&-h\big(\|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta} \,\big|\, \mathbf{X}, E=1\big) \\
&\quad = -\log 2\pi - h\big(\|\mathbb{H}\mathbf{X}\| \,\big|\, \mathbf{X}, e^{\mathrm{i}\Theta}, E=1\big) - \mathsf{E}[\log\|\mathbb{H}\mathbf{X}\| \mid E=1] && \text{(B.22)}\\
&\quad = -\log 2\pi - h\big(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R \,\big|\, \hat{\mathbf{X}}, R, E=1\big) - \mathsf{E}[\log\|\mathbb{H}\mathbf{X}\| \mid E=1] && \text{(B.23)}\\
&\quad = -\log 2\pi - h\big(\|\mathbb{H}\hat{\mathbf{X}}\| \,\big|\, \hat{\mathbf{X}}, R, E=1\big) - \mathsf{E}[\log R \mid E=1] \\
&\qquad - \mathsf{E}[\log\|\mathbb{H}\mathbf{X}\| \mid E=1] && \text{(B.24)}\\
&\quad = -\log 2\pi - h\big(\|\mathbb{H}\hat{\mathbf{X}}\| \,\big|\, \hat{\mathbf{X}}\big) - \mathsf{E}[\log R \mid E=1] - \mathsf{E}[\log\|\mathbb{H}\mathbf{X}\| \mid E=1], && \text{(B.25)}
\end{aligned}$$

where the second equality follows from the definition of $R = \|\mathbf{X}\|$; 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

$$\epsilon_\nu \triangleq \sup_{\|\mathbf{x}\|^2 \geq \mathcal{E}_0} \Big\{ \mathsf{E}\big[\log\big(\|\mathbb{H}\mathbf{x}\|^2 + \nu\big)\big] - \mathsf{E}\big[\log\|\mathbb{H}\mathbf{x}\|^2\big] \Big\}, \tag{B.26}$$

such that

$$\begin{aligned}
&(1-\alpha)\mathsf{E}\big[\log\big(\|\mathbb{H}\mathbf{X}\|^2+\nu\big) \,\big|\, E=1\big] \\
&\quad = (1-\alpha)\mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|^2 \,\big|\, E=1\big] \\
&\qquad + (1-\alpha)\Big(\mathsf{E}\big[\log\big(\|\mathbb{H}\mathbf{X}\|^2+\nu\big) \,\big|\, E=1\big] - \mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|^2 \,\big|\, E=1\big]\Big) && \text{(B.27)}\\
&\quad \leq (1-\alpha)\mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|^2 \,\big|\, E=1\big] \\
&\qquad + (1-\alpha)\sup_{\|\mathbf{x}\|^2\geq\mathcal{E}_0}\Big\{\mathsf{E}\big[\log\big(\|\mathbb{H}\mathbf{x}\|^2+\nu\big)\big] - \mathsf{E}\big[\log\|\mathbb{H}\mathbf{x}\|^2\big]\Big\} && \text{(B.28)}\\
&\quad = (1-\alpha)\mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|^2 \,\big|\, E=1\big] + (1-\alpha)\epsilon_\nu && \text{(B.29)}\\
&\quad \leq (1-\alpha)\mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|^2 \,\big|\, E=1\big] + \epsilon_\nu. && \text{(B.30)}
\end{aligned}$$

Moreover we bound

$$\begin{aligned}
\frac{1}{\beta}\mathsf{E}\big[\|\mathbb{H}\mathbf{X}\|^2 \,\big|\, E=1\big] &= \frac{1}{\beta}\mathsf{E}\Big[\|\mathbb{H}\hat{\mathbf{X}}\|^2\cdot R^2 \,\Big|\, E=1\Big] && \text{(B.31)}\\
&\leq \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\cdot R^2 \,\big|\, E=1\big] && \text{(B.32)}\\
&= \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\cdot\mathsf{E}\big[R^2 \,\big|\, E=1\big] && \text{(B.33)}\\
&\leq \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\cdot\mathcal{E}, && \text{(B.34)}
\end{aligned}$$

where we have used the fact that R needs to satisfy the average-power constraint (2.23).

Plugging (B.25), (B.30), and (B.34) into (B.21) yields

$$\begin{aligned}
&I\big(\mathbf{X}; \|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta} \,\big|\, E=1\big) \\
&\quad \leq -h\big(\|\mathbb{H}\hat{\mathbf{X}}\| \,\big|\, \hat{\mathbf{X}}\big) - \log 2 - \mathsf{E}[\log R \mid E=1] - \mathsf{E}[\log\|\mathbb{H}\mathbf{X}\| \mid E=1] \\
&\qquad + \alpha\log\beta + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) + (1-\alpha)\mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|^2 \,\big|\, E=1\big] + \epsilon_\nu \\
&\qquad + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta}. && \text{(B.35)}
\end{aligned}$$

Next we continue with the second term in (B.20):

$$
\begin{aligned}
& I\left(\mathbf{X}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \|\mathbb{H}\mathbf{X}\|, E=1\right) \\
& \quad = h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \|\mathbb{H}\hat{\mathbf{X}}\|\cdot R, E=1\right) \\
& \qquad - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \|\mathbb{H}\hat{\mathbf{X}}\|\cdot R, \hat{\mathbf{X}}, R, E=1\right) && \text{(B.36)} \\
& \quad = h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \|\mathbb{H}\hat{\mathbf{X}}\|\cdot R, E=1\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \|\mathbb{H}\hat{\mathbf{X}}\|, \hat{\mathbf{X}}, R, E=1\right) && \text{(B.37)} \\
& \quad \le h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \|\mathbb{H}\hat{\mathbf{X}}\|, \hat{\mathbf{X}}\right). && \text{(B.38)}
\end{aligned}
$$

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

Hence, using (B.38) and (B.35) in (B.20) we get

$$
\begin{aligned}
& I\big(\mathbf{X};\mathbf{Y} \,\big|\, E=1\big) \\
& \quad \le -h\big(\|\mathbb{H}\hat{\mathbf{X}}\| \,\big|\, \hat{\mathbf{X}}\big) - \log 2 - \mathsf{E}[\log R \mid E=1] - \mathsf{E}[\log\|\mathbb{H}\mathbf{X}\| \mid E=1] + \alpha\log\beta \\
& \qquad + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) + (1-\alpha)\mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|^2 \,\big|\, E=1\big] + \epsilon_\nu + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\cdot\mathcal{E} \\
& \qquad + \frac{\nu}{\beta} + 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|\, \|\mathbb{H}\hat{\mathbf{X}}\|, \hat{\mathbf{X}}\right) && \text{(B.39)} \\
& \quad = -h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) + (2n_{\mathrm{R}}-1)\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|\Big] - \log 2 - \mathsf{E}[\log R \mid E=1] \\
& \qquad - \mathsf{E}[\log\|\mathbb{H}\mathbf{X}\| \mid E=1] + \alpha\log\beta + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) + 2\mathsf{E}[\log\|\mathbb{H}\mathbf{X}\| \mid E=1] \\
& \qquad - \alpha\mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|^2 \,\big|\, E=1\big] + \epsilon_\nu + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta} \\
& \qquad + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) && \text{(B.40)} \\
& \quad = h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2 + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) \\
& \qquad + \alpha\Big(\log\beta - \mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|^2 \,\big|\, E=1\big]\Big) + \epsilon_\nu + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta} && \text{(B.41)} \\
& \quad \le h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2 + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) \\
& \qquad + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi\Big) + \epsilon_\nu + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta}. && \text{(B.42)}
\end{aligned}
$$

Here, (B.40) follows from a conditional version of Lemma 2; and the final inequality follows from the following bound:

$$
\begin{aligned}
\mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|^2 \,\big|\, E=1\big] &\ge \inf_{\|\mathbf{x}\|^2\ge\mathcal{E}_0}\mathsf{E}\big[\log\|\mathbb{H}\mathbf{x}\|^2\big] && \text{(B.43)} \\
&= \log\mathcal{E}_0 + \inf_{\hat{\mathbf{x}}}\mathsf{E}\big[\log\|\mathbb{H}\hat{\mathbf{x}}\|^2\big] && \text{(B.44)} \\
&\triangleq \log\mathcal{E}_0 + \xi, && \text{(B.45)}
\end{aligned}
$$

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

$$-\infty < \xi < \infty \tag{B.46}$$

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

Note that (B.42) does not depend on the distribution of R anymore, but only on $Q_{\hat{\mathbf{X}}}$! Hence, we can get an upper bound on capacity by taking the supremum over all possible distributions of $\hat{\mathbf{X}}$.

Taking everything together we now have the following bound on channel capacity:

$$\begin{aligned}
\mathsf{C}(\mathcal{E}) &\leq I(\mathbf{X};\mathbf{Y}) + \epsilon \\
&\leq H_{\mathrm{b}}(p) + (1-p)\mathsf{C}_{\mathrm{IID}}(\mathcal{E}_0) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\big(\mathbb{H}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}}\big) \\
&\quad + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2 + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi\Big) \\
&\quad + \epsilon_\nu + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta} + \epsilon.
\end{aligned}$$

Hence we get an upper bound on the fading number as follows:

$$\begin{aligned}
\chi(\mathbb{H}) &= \varlimsup_{\mathcal{E}\uparrow\infty}\left\{\mathsf{C}(\mathcal{E}) - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} \qquad \text{(B.47)}\\
&\leq \varlimsup_{\mathcal{E}\uparrow\infty}\left\{h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\big(\mathbb{H}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}}\big) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2 + \epsilon\right. \\
&\qquad + H_{\mathrm{b}}(p) + (1-p)\mathsf{C}_{\mathrm{IID}}(\mathcal{E}_0) + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi\Big) \\
&\qquad \left. + \epsilon_\nu + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta} - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} \qquad \text{(B.48)}\\
&\leq \varlimsup_{\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\big(\mathbb{H}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}}\big) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2\right\} + \epsilon\right. \\
&\qquad + H_{\mathrm{b}}(p) + (1-p)\mathsf{C}_{\mathrm{IID}}(\mathcal{E}_0) + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi\Big) \\
&\qquad \left. + \epsilon_\nu + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta} - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} \qquad \text{(B.49)}\\
&= \sup_{Q_{\hat{\mathbf{X}}}}\left\{h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\big(\mathbb{H}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}}\big) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2\right\} + \epsilon \\
&\quad + \varlimsup_{\mathcal{E}\uparrow\infty}\left\{H_{\mathrm{b}}(p) + (1-p)\mathsf{C}_{\mathrm{IID}}(\mathcal{E}_0) + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) - \log\frac{1}{\alpha}\right. \\
&\qquad + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi\Big) + \epsilon_\nu + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\cdot\mathcal{E} + \frac{\nu}{\beta} \\
&\qquad \left. + \log\frac{1}{\alpha} - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} \qquad \text{(B.50)}\\
&= \sup_{Q_{\hat{\mathbf{X}}}}\left\{h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\big(\mathbb{H}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}}\big) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2\right\} \\
&\quad + \epsilon + \log\left(1 - e^{-\nu}\right) + \nu + \epsilon_\nu - \log\nu. \qquad \text{(B.51)}
\end{aligned}$$

Here in (B.49) we upper-bound four terms by maximizing over the distribution of $\hat{\mathbf{X}}$ which does not depend on $\mathcal{E}$. Hence in the subsequent equality we can take those terms out of the limit. The last equality then follows from (B.5) and the following choices of the free parameters α and β:

$$\alpha \triangleq \alpha(\mathcal{E}) = \frac{\nu}{\log \mathcal{E} + \log \sup_{\hat{\mathbf{x}}} \mathsf{E}[\|\mathbb{H}\hat{\mathbf{x}}\|^2]}; \tag{B.52}$$

$$\beta \triangleq \beta(\mathcal{E}) = \frac{1}{\alpha(\mathcal{E})} e^{\nu/\alpha(\mathcal{E})}, \tag{B.53}$$

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

$$\overline{\lim_{\mathcal{E}\uparrow\infty}} \left\{ \log \Gamma\left(\alpha, \frac{\nu}{\beta}\right) - \log \frac{1}{\alpha} \right\} = \log\left(1 - e^{-\nu}\right); \tag{B.54}$$

$$\overline{\lim_{\mathcal{E}\uparrow\infty}}\, \alpha \Big(\log \beta - \log \mathcal{E}_0 - \xi \Big) = \nu; \tag{B.55}$$

$$\overline{\lim_{\mathcal{E}\uparrow\infty}} \left\{ \frac{1}{\beta} \sup_{\hat{\mathbf{x}}} \mathsf{E}\left[\|\mathbb{H}\hat{\mathbf{x}}\|^2\right] \cdot \mathcal{E} + \frac{\nu}{\beta} \right\} = 0; \tag{B.56}$$

$$\overline{\lim_{\mathcal{E}\uparrow\infty}} \left\{ \log \frac{1}{\alpha} - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right) \right\} = -\log \nu. \tag{B.57}$$

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

To finish the derivation of the upper bound, we let ν go to zero. Note that $\epsilon_\nu \to 0$ as $\nu \downarrow 0$ as can be seen from (B.26). Note further that

$$\lim_{\nu\downarrow 0} \Big\{ \log\left(1 - e^{-\nu}\right) - \log \nu \Big\} = 0. \tag{B.58}$$

Therefore, we get

$$\chi(\mathbb{H}) \leq \sup_{Q_{\hat{\mathbf{X}}}} \left\{ h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) + n_{\mathrm{R}} \mathsf{E}\Big[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2 \right\} + \epsilon. \tag{B.59}$$

The upper bound now follows since ϵ is arbitrary.

B.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

$$\mathbf{X} = R \cdot \hat{\mathbf{X}}. \tag{B.60}$$

Here $\hat{\mathbf{X}} \in \mathbb{C}^{n_{\mathrm{T}}}$ 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

$$\log R^2 \sim \mathcal{U}\left([\log x_{\min}^2, \log \mathcal{E}] \right), \tag{B.61}$$

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

$$x_{\min}^2 = \log \mathcal{E}. \tag{B.62}$$

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

$$\begin{aligned}
\mathsf{C}(\mathcal{E}) &\geq I(\mathbf{X};\mathbf{Y}) && \text{(B.63)}\\
&\geq I\big(R,\hat{\mathbf{X}};\mathbf{Y}\big) && \text{(B.64)}\\
&= I\big(\hat{\mathbf{X}};\mathbf{Y}\big) + I\big(R;\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\big) && \text{(B.65)}\\
&= I\big(\hat{\mathbf{X}};\mathbf{Y}\big) + I\big(R;\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) - I\big(R;\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + I\big(R;\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\big) && \text{(B.66)}\\
&= I\big(\hat{\mathbf{X}};\mathbf{Y}\big) + I\big(R,e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) - I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},R\big)\\
&\quad - I\big(R;\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + I\big(R;\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\big). && \text{(B.67)}
\end{aligned}$$

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:

$$\begin{aligned}
&-I\big(R;\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + I\big(R;\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\big)\\
&\quad = -h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},R\big) + h\big(\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\big) - h\big(\mathbf{Y}\,\big|\,\hat{\mathbf{X}},R\big) && \text{(B.68)}\\
&\quad = -h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},R\big) + h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},e^{\mathrm{i}\Theta}\big)\\
&\qquad - h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},R,e^{\mathrm{i}\Theta}\big) && \text{(B.69)}\\
&\quad = -I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},R\big). && \text{(B.70)}
\end{aligned}$$

Here the second equality follows because $e^{\mathrm{i}\Theta}$ 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.

Putting this into (B.67) yields

$$\begin{aligned}
\mathsf{C}(\mathcal{E}) &\geq I\big(\hat{\mathbf{X}};\mathbf{Y}\big) + I\big(R,e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) - I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) && \text{(B.71)}\\
&= I\big(\hat{\mathbf{X}};\mathbf{Y}\big) + I\big(Re^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) - I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big). && \text{(B.72)}
\end{aligned}$$

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

We continue with bounding the first term in (B.72):

$$\begin{aligned}
I\big(\hat{\mathbf{X}};\mathbf{Y}\big) &= I\big(\hat{\mathbf{X}};\mathbf{Y},\mathbf{Z}\big) - \underbrace{I\big(\hat{\mathbf{X}};\mathbf{Z}\,\big|\,\mathbf{Y}\big)}_{\leq\epsilon(x_{\min})} && \text{(B.73)}\\
&\geq I\big(\hat{\mathbf{X}};\mathbf{Y},\mathbf{Z}\big) - \epsilon(x_{\min}) && \text{(B.74)}\\
&= I\big(\hat{\mathbf{X}};\mathbb{H}\hat{\mathbf{X}}R\big) - \epsilon(x_{\min}) && \text{(B.75)}\\
&= I\left(\hat{\mathbf{X}};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|},\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\right) - \epsilon(x_{\min}) && \text{(B.76)}\\
&= I\left(\hat{\mathbf{X}};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + I\left(\hat{\mathbf{X}};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - \epsilon(x_{\min}). && \text{(B.77)}
\end{aligned}$$

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 B.3 and is shown there to only depend on $x_{\min}$ 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 $\mathbf{Y}$ and $\mathbf{Z}$ 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.

Next we bound the third term in (B.72):

$$
\begin{aligned}
&-I\bigl(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\bigm|\hat{\mathbf{X}}\bigr)\\
&\quad\geq -I\bigl(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta},\mathbf{Z}e^{\mathrm{i}\Theta}\bigm|\hat{\mathbf{X}}\bigr) && \text{(B.78)}\\
&\quad= -I\bigl(e^{\mathrm{i}\Theta};\mathbb{H}\mathbf{X}e^{\mathrm{i}\Theta},\mathbf{Z}e^{\mathrm{i}\Theta}\bigm|\hat{\mathbf{X}}\bigr) && \text{(B.79)}\\
&\quad= -I\bigl(e^{\mathrm{i}\Theta};\mathbb{H}\mathbf{X}e^{\mathrm{i}\Theta}\bigm|\hat{\mathbf{X}}\bigr)-I\bigl(e^{\mathrm{i}\Theta};\mathbf{Z}e^{\mathrm{i}\Theta}\bigm|\mathbb{H}\mathbf{X}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\bigr) && \text{(B.80)}\\
&\quad= -I\bigl(e^{\mathrm{i}\Theta};\mathbb{H}\mathbf{X}e^{\mathrm{i}\Theta}\bigm|\hat{\mathbf{X}}\bigr) && \text{(B.81)}\\
&\quad= -I\left(e^{\mathrm{i}\Theta};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right) && \text{(B.82)}\\
&\quad= -I\left(e^{\mathrm{i}\Theta};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right)-I\left(e^{\mathrm{i}\Theta};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\right). && \text{(B.83)}
\end{aligned}
$$

Note that in (B.81) we used the fact that $\mathbf{Z}$ is circularly symmetric.

Hence, plugging these results into (B.72) we get:

$$
\begin{aligned}
C(\mathcal{E}) &\geq I\bigl(Re^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\bigm|\hat{\mathbf{X}}\bigr)+I\left(\hat{\mathbf{X}};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right)+I\left(\hat{\mathbf{X}};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right)-\epsilon(x_{\min})\\
&\quad -I\left(e^{\mathrm{i}\Theta};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right)-I\left(e^{\mathrm{i}\Theta};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\right) && \text{(B.84)}\\
&= I\bigl(Re^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\bigm|\hat{\mathbf{X}}\bigr)+I\left(\hat{\mathbf{X}};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right)-I\left(e^{\mathrm{i}\Theta};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right)\\
&\quad +I\left(\hat{\mathbf{X}};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right)-I\left(e^{\mathrm{i}\Theta};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\right)\\
&\quad -\epsilon(x_{\min}), && \text{(B.85)}
\end{aligned}
$$

where in (B.85) we only rearranged the order of the terms.

We next bound the last two mutual information terms in (B.85):

$$
\begin{aligned}
&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^{\mathrm{i}\Theta};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\right)\\
&\quad= 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)\\
&\qquad -h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{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^{\mathrm{i}\Theta},\hat{\mathbf{X}},e^{\mathrm{i}\Theta}\right) && \text{(B.86)}\\
&\quad= 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)\\
&\qquad -h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{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) && \text{(B.87)}\\
&\quad= 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^{\mathrm{i}\Theta},\hat{\mathbf{X}}\right) && \text{(B.88)}\\
&\quad= h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},e^{\mathrm{i}\Theta}\right)-h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\right) && \text{(B.89)}
\end{aligned}
$$

$$\geq h\left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \,\middle|\, \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}, e^{\mathrm{i}\Theta}\right) - h\left(\|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \,\middle|\, \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\right) \tag{B.90}$$

$$= -I\left(e^{\mathrm{i}\Theta}; \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \,\middle|\, \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\right) \tag{B.91}$$

$$= -I\left(e^{\mathrm{i}\Theta}; \|\mathbb{H}\hat{\mathbf{X}}\| \cdot R \,\middle|\, \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) \tag{B.92}$$

$$= 0. \tag{B.93}$$

Here, the inequality follows from conditioning the reduces entropy; 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 $e^{\mathrm{i}\Theta}$; and the last equality follows since Θ is independent of any other random quantity.

Therefore, we are left over with the following bound:

$$\mathsf{C}(\mathcal{E}) \geq I\big(Re^{\mathrm{i}\Theta}; \mathbf{Y}e^{\mathrm{i}\Theta} \,\big|\, \hat{\mathbf{X}}\big) + I\left(\hat{\mathbf{X}}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - I\left(e^{\mathrm{i}\Theta}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right) - \epsilon(x_{\min}), \tag{B.94}$$

and continue by rewriting the second and third term as follows:

$$\begin{aligned}
&I\left(\hat{\mathbf{X}}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - I\left(e^{\mathrm{i}\Theta}; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right) \\
&\quad = h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \hat{\mathbf{X}}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right) \\
&\qquad + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}, e^{\mathrm{i}\Theta}\right) &\text{(B.95)} \\
&\quad = h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \hat{\mathbf{X}}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right) \\
&\qquad + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \hat{\mathbf{X}}\right) &\text{(B.96)} \\
&\quad = h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right), &\text{(B.97)}
\end{aligned}$$

which leaves us with

$$\mathsf{C}(\mathcal{E}) \geq I\big(Re^{\mathrm{i}\Theta}; \mathbf{Y}e^{\mathrm{i}\Theta} \,\big|\, \hat{\mathbf{X}}\big) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right) - \epsilon(x_{\min}). \tag{B.98}$$

We next let the power grow to infinity $\mathcal{E} \to \infty$ and use the definition of the fading number

$$\chi(\mathbb{H}) = \varlimsup_{\mathcal{E}\uparrow\infty} \left\{\mathsf{C}(\mathcal{E}) - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right)\right\}. \tag{B.99}$$

Since $Re^{\mathrm{i}\Theta}$ is circularly symmetric with a magnitude according to (B.61), we know from [7, (108)], [2, (6.194)], that $Re^{\mathrm{i}\Theta}$ achieves the fading number of a memoryless SIMO fading channel with partial side-information. In our situation we have

$$I\big(Re^{\mathrm{i}\Theta}; \mathbf{Y}e^{\mathrm{i}\Theta} \,\big|\, \hat{\mathbf{X}}\big) = I\big(Re^{\mathrm{i}\Theta}; \mathbb{H}\hat{\mathbf{X}}Re^{\mathrm{i}\Theta} + \mathbf{Z} \,\big|\, \hat{\mathbf{X}}\big) \tag{B.100}$$

where $\hat{\mathbf{X}}$ serves as partial receiver side-information (that is independent of the SIMO input $Re^{\mathrm{i}\Theta}$). 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.*,

$$h\big(\mathbb{H}\hat{\mathbf{X}}\,\big|\hat{\mathbf{X}}\big) > -\infty, \tag{B.101}$$

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

Hence, using the fading number of a memoryless SIMO fading channel where the receiver has access to some partial side-information [7, (108)], [2, (6.194)]:

$$\chi(\mathbf{B}|\mathbf{A}) = h_\lambda\big(\hat{\mathbf{B}}e^{\mathrm{i}\Theta}\,\big|\,\mathbf{A}\big) + n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{B}\|^2\big] - \log 2 - h(\mathbf{B}|\mathbf{A}), \tag{B.102}$$

we get

$$\begin{aligned}
\chi(\mathbb{H}) &\geq \varlimsup_{\mathcal{E}\uparrow\infty}\Bigg\{ I\big(Re^{\mathrm{i}\Theta};\mathbb{H}\hat{\mathbf{X}}Re^{\mathrm{i}\Theta}+\mathbf{Z}\,\big|\,\hat{\mathbf{X}}\big) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right) \\
&\qquad\qquad - \epsilon(x_{\min}) - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\Bigg\} && \text{(B.103)}\\
&= \varlimsup_{\mathcal{E}\uparrow\infty}\Bigg\{ I\big(Re^{\mathrm{i}\Theta};\mathbb{H}\hat{\mathbf{X}}Re^{\mathrm{i}\Theta}+\mathbf{Z}\,\big|\,\hat{\mathbf{X}}\big) - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right) - \epsilon(x_{\min})\Bigg\} \\
&\quad + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right) && \text{(B.104)}\\
&= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}}\big) \\
&\quad + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right) && \text{(B.105)}\\
&= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}}\,\big|\,\hat{\mathbf{X}}\big). && \text{(B.106)}
\end{aligned}$$

Here in (B.105) we have used the fact that our choice (B.62) guarantees that $\epsilon(x_{\min})$ tends to zero as $\mathcal{E} \to \infty$ (see Section B.3) and that we achieve the SIMO fading number for a channel with input $Re^{\mathrm{i}\Theta}$ and output $\mathbb{H}\hat{\mathbf{X}}Re^{\mathrm{i}\Theta} + \mathbf{Z}$.

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

B.3 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

$$I\big(\hat{\mathbf{X}};\mathbf{Z}\,\big|\,\mathbf{Y}\big) \leq \epsilon(x_{\min}) \tag{B.107}$$

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

Such a bound can be found as follows:

$$\begin{aligned}
I\big(\hat{\mathbf{X}};\mathbf{Z}\,\big|\,\mathbf{Y}\big) &= h(\mathbf{Z}|\mathbf{Y}) - h\big(\mathbf{Z}\,\big|\,\mathbf{Y},\hat{\mathbf{X}}\big) && \text{(B.108)}\\
&\le h(\mathbf{Z}) - h\big(\mathbf{Z}\,\big|\,\mathbf{Y},\hat{\mathbf{X}},R\big) && \text{(B.109)}\\
&= h(\mathbf{Z}) - h\big(\mathbf{Z}\,\big|\,\mathbb{H}\hat{\mathbf{X}}R+\mathbf{Z},\hat{\mathbf{X}},R\big) && \text{(B.110)}\\
&\le h(\mathbf{Z}) - \inf_{\hat{\mathbf{x}}}\inf_{r\ge x_{\min}} h\big(\mathbf{Z}\,\big|\,\mathbb{H}\hat{\mathbf{x}}r+\mathbf{Z}\big) && \text{(B.111)}\\
&= h(\mathbf{Z}) - \inf_{\hat{\mathbf{x}}} h\big(\mathbf{Z}\,\big|\,\mathbb{H}\hat{\mathbf{x}}x_{\min}+\mathbf{Z}\big) && \text{(B.112)}\\
&= \sup_{\hat{\mathbf{x}}} I\big(\mathbf{Z};\mathbb{H}\hat{\mathbf{x}}x_{\min}+\mathbf{Z}\big) && \text{(B.113)}\\
&= \sup_{\hat{\mathbf{x}}} I\left(\frac{\mathbf{Z}}{x_{\min}};\mathbb{H}\hat{\mathbf{x}}+\frac{\mathbf{Z}}{x_{\min}}\right) && \text{(B.114)}\\
&= \sup_{\hat{\mathbf{x}}}\left\{h\left(\mathbb{H}\hat{\mathbf{x}}+\frac{\mathbf{Z}}{x_{\min}}\right)-h\big(\mathbb{H}\hat{\mathbf{x}}\big)\right\} && \text{(B.115)}\\
&\triangleq \epsilon(x_{\min}). && \text{(B.116)}
\end{aligned}$$

The convergence now follows from [7, Lemma 6.11], [2, Lemma A.19].

Appendix C

Proof of Corollary 13

We choose a constant $n_T \times n_T$ matrix B as follows:

$$\mathsf{B} = \operatorname{diag}\left(\frac{1}{d_1}, \dots, \frac{1}{d_{n_R}}, \frac{1}{d_1}, \dots, \frac{1}{d_1}\right) \tag{C.1}$$

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

$$\mathbb{H}\mathsf{B}\hat{\mathbf{x}} = \mathsf{D}\mathsf{B}\hat{\mathbf{x}} + \tilde{\mathbb{H}}\mathsf{B}\hat{\mathbf{x}} = \begin{pmatrix} \hat{x}^{(1)} \\ \vdots \\ \hat{x}^{(n_R)} \end{pmatrix} + \tilde{\mathbb{H}}\mathsf{B}\hat{\mathbf{x}} \triangleq \boldsymbol{\xi} + \tilde{\mathbf{H}} \tag{C.2}$$

where $\tilde{\mathbf{H}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma \mathsf{I}_{n_R})$ with

$$\sigma^2 = \frac{|\hat{x}^{(1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{x}^{(n_R)}|^2}{|d_{n_R}|^2} + \frac{|\hat{x}^{(n_R+1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{x}^{(n_T)}|^2}{|d_1|^2} \tag{C.3}$$

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

$$h\big(\mathbb{H}\mathsf{B}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}}\big) = n_R \log \pi e \sigma^2; \tag{C.4}$$

$$\mathsf{E}\big[\log \|\mathbb{H}\mathsf{B}\hat{\mathbf{x}}\|^2\big] = \log \sigma^2 + g_{n_R}\left(\frac{\|\boldsymbol{\xi}\|^2}{\sigma^2}\right); \tag{C.5}$$

and hence

$$\begin{aligned} & n_R \mathsf{E}\Big[\log \|\mathbb{H}\mathsf{B}\hat{\mathbf{X}}\|^2\Big] - h\big(\mathbb{H}\mathsf{B}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) \\ &= n_R \mathsf{E}\left[g_{n_R}\left(\frac{|\hat{X}^{(1)}|^2 + \cdots + |\hat{X}^{(n_R)}|^2}{\frac{|\hat{X}^{(1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_R)}|^2}{|d_{n_R}|^2} + \frac{|\hat{X}^{(n_R+1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_T)}|^2}{|d_1|^2}}\right)\right] - n_R \log \pi e. \end{aligned} \tag{C.6}$$

The upper bound on the fading number now follows from Theorem 8 by upper-bounding the h_λ-term by $\log c_{n_R}$; by

$$|\hat{X}^{(1)}|^2 + \cdots + |\hat{X}^{(n_R)}|^2 \leq 1; \tag{C.7}$$

by

$$\begin{aligned}
&\frac{|\hat{X}^{(1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}(n_{\mathrm{R}})|^2}{|d_{n_{\mathrm{R}}}|^2} + \frac{|\hat{X}^{(n_{\mathrm{R}}+1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_{\mathrm{T}})}|^2}{|d_1|^2} \\
&\quad \geq \frac{|\hat{X}^{(1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_{\mathrm{R}})}|^2}{|d_1|^2} + \frac{|\hat{X}^{(n_{\mathrm{R}}+1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_{\mathrm{T}})}|^2}{|d_1|^2} && \text{(C.8)} \\
&\quad = \frac{1}{|d_1|^2}\left(|\hat{X}^{(1)}|^2 + \cdots + |\hat{X}^{(n_{\mathrm{T}})}|^2\right) && \text{(C.9)} \\
&\quad = \frac{1}{|d_1|^2} = \frac{1}{\|\mathsf{D}\|^2} && \text{(C.10)}
\end{aligned}$$

where the inequality follows since $|d_1| \geq |d_2| \geq \ldots \geq |d_{n_{\mathrm{R}}}|$; and from the fact that $g_m(\cdot)$ is a monotonically increasing function.

To derive the lower bounds we choose a particular distribution on $\hat{\mathbf{X}}$:

$$\hat{\mathbf{X}} \triangleq \begin{pmatrix} \hat{\boldsymbol{\Xi}} \\ \mathbf{0} \end{pmatrix} \tag{C.11}$$

where $\hat{\boldsymbol{\Xi}}$ is chosen to be uniformly distributed on the unit sphere in $\mathbb{C}^{n_{\mathrm{R}}}$. This choice ensures that $\mathbb{H}\mathsf{B}\hat{\mathbf{X}}$ is isotropically distributed, *i.e.*,

$$h_\lambda\left(\frac{\mathbb{H}\mathsf{B}\hat{\mathbf{X}}}{\|\mathbb{H}\mathsf{B}\hat{\mathbf{X}}\|}\right) = \log\frac{2\pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})} \tag{C.12}$$

and that in (C.6)

$$\begin{aligned}
&\mathsf{E}\left[g_{n_{\mathrm{R}}}\left(\frac{|\hat{X}^{(1)}|^2 + \cdots + |\hat{X}^{(n_{\mathrm{R}})}|^2}{\frac{|\hat{X}^{(1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_{\mathrm{R}})}|^2}{|d_{n_{\mathrm{R}}}|^2} + \frac{|\hat{X}^{(n_{\mathrm{R}}+1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_{\mathrm{T}})}|^2}{|d_1|^2}}\right)\right] \\
&\qquad\qquad = \mathsf{E}\left[g_{n_{\mathrm{R}}}\left(\frac{1}{\frac{|\hat{X}^{(1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_{\mathrm{R}})}|^2}{|d_{n_{\mathrm{R}}}|^2}}\right)\right]. && \text{(C.13)}
\end{aligned}$$

The (in general not tight) lower bound (3.65) follows from (3.64) by upper-bounding $|\hat{X}^{(\ell)}|^2 \leq 1$.

Appendix D

Proof of Proposition 14

This upper bound is based on the upper bound given in Corollary 9 for a choice of $\mathsf{B} = \mathsf{I}_{n_\mathrm{T}}$. If $n_\mathrm{R} > n_\mathrm{T}$ we choose for A

$$\mathsf{A} = \mathrm{diag}\left(\frac{a}{d_1}, \ldots, \frac{a}{d_{n_\mathrm{T}}}, b, \ldots, b\right) \tag{D.1}$$

with a such that $\det \mathsf{A} = 1$ and with

$$b^2 = \left(\frac{\delta^2}{n_\mathrm{T}}\right)^{n_\mathrm{T}/n_\mathrm{R}} \tag{D.2}$$

for δ as given in (3.67). For such a choice we get

$$n_\mathrm{R}\mathsf{E}\left[\log \|\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\|^2\right] - h\big(\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\big) \leq n_\mathrm{R} \log \mathsf{E}\left[\|\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\|^2\right] - \log\det\mathsf{A} - h\big(\mathbb{H}\hat{\mathbf{x}}\big) \tag{D.3}$$

$$= n_\mathrm{R} \log\left(n_\mathrm{R}\left(\frac{\delta^2}{n_\mathrm{T}}\right)^{n_\mathrm{T}/n_\mathrm{R}}\right) - n_\mathrm{R}\log \pi e \tag{D.4}$$

where the inequality follows from Jensen. Plugging this into (3.27) yields

$$\chi \leq n_\mathrm{R}\log\pi - \log\Gamma(n_\mathrm{R}) + n_\mathrm{R}\log n_\mathrm{R} + n_\mathrm{T}\log\left(\frac{\delta^2}{n_\mathrm{T}}\right) - n_\mathrm{R}\log\pi e \tag{D.5}$$

$$= n_\mathrm{T}\log\left(\frac{\delta^2}{n_\mathrm{T}}\right) + n_\mathrm{R}\log n_\mathrm{R} - \log\Gamma(n_\mathrm{R}) - n_\mathrm{R}. \tag{D.6}$$

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

$$\mathsf{A} = \mathrm{diag}\left(\frac{a}{d_1}, \ldots, \frac{a}{d_{n_\mathrm{R}}}\right) \tag{D.7}$$

with a such that $\det\mathsf{A} = 1$. For such a choice we get

$$n_\mathrm{R}\mathsf{E}\left[\log \|\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\|^2\right] - h\big(\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\big)$$

$$\leq n_\mathrm{R} \log \mathsf{E}\left[\|\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\|^2\right] - \log\det\mathsf{A} - h\big(\mathbb{H}\hat{\mathbf{x}}\big) \tag{D.8}$$

$$= n_\mathrm{R}\log\Bigg(\left(|d_1|^2\cdot\cdots\cdot|d_{n_\mathrm{R}}|^2\right)^{1/n_\mathrm{R}} \tag{D.9}$$

$$\cdot\left(\frac{1}{|d_1|^2} + \cdots + \frac{1}{|d_{n_\mathrm{R}}|^2} + \left|\hat{x}^{(1)}\right|^2 + \cdots + \left|\hat{x}^{(n_\mathrm{R})}\right|^2\right)\Bigg)$$

$$- h\big(\mathbb{H}\hat{\mathbf{x}}\big) \tag{D.10}$$

$$\leq n_\mathrm{R}\log\left(\left(|d_1|^2\cdot\cdots\cdot|d_{n_\mathrm{R}}|^2\right)^{1/n_\mathrm{R}}\left(\frac{1}{|d_1|^2} + \cdots + \frac{1}{|d_{n_\mathrm{R}}|^2} + 1\right)\right) - n_\mathrm{R}\log\pi e \tag{D.11}$$

$$= n_\mathrm{R}\log\delta^2 - n_\mathrm{R}\log\pi e \tag{D.12}$$

where the first inequality follows from Jensen, and the second by upper-bounding $|\hat{x}^{(1)}|^2 + \cdots + |\hat{x}^{(n_{\mathrm{R}})}|^2 \leq 1$. Plugging this into (3.27) yields

$$\begin{aligned} \chi &\leq n_{\mathrm{R}} \log \pi - \log \Gamma(n_{\mathrm{R}}) + n_{\mathrm{R}} \log \delta^2 - n_{\mathrm{R}} \log \pi e && \text{(D.13)} \\ &= n_{\mathrm{R}} \log \delta^2 - \log \Gamma(n_{\mathrm{R}}) - n_{\mathrm{R}} && \text{(D.14)} \\ &= n_{\mathrm{R}} \log \frac{\delta^2}{n_{\mathrm{R}}} + n_{\mathrm{R}} \log n_{\mathrm{R}} - \log \Gamma(n_{\mathrm{R}}) - n_{\mathrm{R}}. && \text{(D.15)} \end{aligned}$$

The result now follows by combining (D.6) and (D.15).

Appendix E

Proof of Lemma 17

Note that for the given choice of distribution we get

$$\mathsf{E}\left[X^{(i)}\right] = -d^{(i)*} \cdot p^{(i)} + \frac{d^{(i)*}p^{(i)}}{1-p^{(i)}} \cdot \left(1-p^{(i)}\right) = 0 \tag{E.1}$$

and

$$\mathsf{E}\left[\left|X^{(i)}\right|^2\right] = \left|-d^{(i)*}\right|^2 \cdot \frac{\mathcal{E}}{\mathcal{E}+2|d^{(i)}|^2} + \left|\frac{d^{(i)*}\mathcal{E}\cdot(\mathcal{E}+2|d^{(i)}|^2)}{(\mathcal{E}+2|d^{(i)}|^2)\cdot 2|d^{(i)}|^2}\right|^2 \cdot \frac{2|d^{(i)}|^2}{\mathcal{E}+2|d^{(i)}|^2} \tag{E.2}$$

$$= \frac{|d^{(i)}|^2\mathcal{E}}{\mathcal{E}+2|d^{(i)}|^2} + \frac{\mathcal{E}^2}{4|d^{(i)}|^2} \cdot \frac{2|d^{(i)}|^2}{\mathcal{E}+2|d^{(i)}|^2} \tag{E.3}$$

$$= \frac{|d^{(i)}|^2\mathcal{E}}{\mathcal{E}+2|d^{(i)}|^2} + \frac{\mathcal{E}^2/2}{\mathcal{E}+2|d^{(i)}|^2} \tag{E.4}$$

$$= \frac{\mathcal{E}}{2} \cdot \frac{2|d^{(i)}|^2+\mathcal{E}}{\mathcal{E}+2|d^{(i)}|^2} = \frac{\mathcal{E}}{2}. \tag{E.5}$$

Moreover we have

$$\mathsf{E}\left[\frac{\left|d^{(1)}X^{(1)}+d^{(2)}X^{(2)}\right|^2}{\left|X^{(1)}\right|^2+\left|X^{(2)}\right|^2}\right]$$

$$= p^{(1)}p^{(2)}\frac{\left|-d^{(1)}d^{(1)*}-d^{(2)}d^{(2)*}\right|^2}{\left|d^{(1)*}\right|^2+\left|d^{(2)*}\right|^2} + p^{(1)}\left(1-p^{(2)}\right)\frac{\left|-d^{(1)}d^{(1)*}+d^{(2)}\frac{d^{(2)*}\mathcal{E}}{2|d^{(2)}|^2}\right|^2}{\left|d^{(1)*}\right|^2+\left|\frac{d^{(2)*}\mathcal{E}}{2|d^{(2)}|^2}\right|^2}$$

$$+ \left(1-p^{(1)}\right)p^{(2)}\frac{\left|d^{(1)}\frac{d^{(1)*}\mathcal{E}}{2|d^{(1)}|^2}-d^{(2)}d^{(2)*}\right|^2}{\left|\frac{d^{(1)*}\mathcal{E}}{2|d^{(1)}|^2}\right|^2+\left|d^{(2)*}\right|^2}$$

$$+ \left(1-p^{(1)}\right)\left(1-p^{(2)}\right)\frac{\left|d^{(1)}\frac{d^{(1)*}\mathcal{E}}{2|d^{(1)}|^2}+d^{(2)}\frac{d^{(2)*}\mathcal{E}}{2|d^{(2)}|^2}\right|^2}{\left|\frac{d^{(1)*}\mathcal{E}}{2|d^{(1)}|^2}\right|^2+\left|\frac{d^{(2)*}\mathcal{E}}{2|d^{(2)}|^2}\right|^2} \tag{E.6}$$

$$= p^{(1)}p^{(2)}\left(\left|d^{(1)}\right|^2+\left|d^{(2)}\right|^2\right) + p^{(1)}\left(1-p^{(2)}\right)\frac{\left(|d^{(1)}|^2-\frac{\mathcal{E}}{2}\right)^2}{\left|d^{(1)}\right|^2+\frac{\mathcal{E}^2}{4|d^{(2)}|^2}}$$

$$+ \left(1-p^{(1)}\right)p^{(2)}\frac{\left(\frac{\mathcal{E}}{2}-|d^{(2)}|^2\right)^2}{\frac{\mathcal{E}^2}{4|d^{(1)}|^2}+\left|d^{(2)}\right|^2} + \left(1-p^{(1)}\right)\left(1-p^{(2)}\right)\frac{\mathcal{E}^2}{\frac{\mathcal{E}^2}{4|d^{(1)}|^2}+\frac{\mathcal{E}^2}{4|d^{(2)}|^2}} \tag{E.7}$$

$$\geq p^{(1)}p^{(2)}\left(\left|d^{(1)}\right|^2+\left|d^{(2)}\right|^2\right) \tag{E.8}$$

$$= \frac{\mathcal{E}}{\mathcal{E}+2|d^{(1)}|^2}\cdot\frac{\mathcal{E}}{\mathcal{E}+2|d^{(2)}|^2}\left(\left|d^{(1)}\right|^2+\left|d^{(2)}\right|^2\right) \tag{E.9}$$

$$\to \left|d^{(1)}\right|^2+\left|d^{(2)}\right|^2. \tag{E.10}$$

Bibliography

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

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

[3] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 437–453, February 2006.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[24] A. Lapidoth and S. M. Moser, "The expected logarithm of a non-central chi-square random variable," website. [Online]. Available: http://moser.cm.nctu.edu.tw/explog.html

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

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