Diplomarbeit

Capacity Analysis of MIMO Systems

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Diplom-Ingenieurs
unter der Leitung von

Dipl.-Ing. Dominik Seethaler
Ao. Prof. Dipl.-Ing. Dr. techn. Franz Hlawatsch
Institut für Nachrichtentechnik und Hochfrequenztechnik (E389)

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

von

Martin Wrulich, 9961105
Neubaugasse 47/19
1070 Wien

Wien, Jänner 2006 _____

Abstract

Digital communication using multiple-input multiple-output (MIMO) wireless links has recently emerged as one of the most significant technical breakthroughs in modern communications. This thesis presents an overview of some important theoretical concepts of MIMO systems. After describing the basic ideas of MIMO transmissions in the introduction, we mainly focused on information theoretical concepts and investigated the system capacity and the mutual information for finite symbol alphabets of some prominent MIMO ST designs. Furthermore, the error performance is studied, in order to derive a more complete understanding of MIMO system parameters. All analyses were performed under ideal identical independent fading conditions. At the end of this thesis, we related the system capacity and the error performance of MIMO systems to the framework of the diversity-multiplexing tradeoff. Each chapter is rounded by a number of simulations to deepen the understanding of the derived theoretical concepts.

Acknowledgements

To begin with, I would like to thank Dominik Seethaler, my supervisor, for his constant support and many suggestions, but also for his patience and gentleness in those times, where I had to slug through difficult problems.

I also have to thank Professor Franz Hlawatsch for his interest in my work (actually he was the one convincing me to deal with this subject) and his supply with interesting papers and ideas which nearly always found their way into my work.

I due Christian Mehlführer and Lukas Mayer a lot for sharing their knowledge with me and supporting me, whenever I had the need to discuss my emerged problems or successes during the work on the thesis. In this context, I have to say that my thank also goes to the whole institute of communications and radio-frequency engineering. It is a pleasure to work in such an innovative and friendly environment.

Of course, I am also grateful to my parents with their patience and love. Without them, this work would never have come into existence, and I will truly miss the phone calls, in which they asked for status reports.

Finally, I wish to thank the following: Gloria (for her love and her wonderful way of giving me self-confidence); Johannes (for his friendship and regular discussions about science itself); Sebastian (for his remarkably way of asking questions, whenever I was not prepared for it); Markus, Christian, Beate, Cornelia, Andreas, Daniel, Doreen, Eva, Laura, Lukas, ... (for all the good and bad times we had together); Seeed and Mick Jagger (they know why); *and* my sister Elisabeth (because she always was on my side and showed so much interest in my work, although it is not her profession).

Vienna, Austria
January 11, 2006

Martin Wrulich

Contents

List of Figures

Glossary

AILL	asymptotic-information-lossless
BER	bit error rate
cdf	cumulative density function
CSI	channel state information
GSM	global system for mobile communications
iid	independent identically distributed
ILL	information-lossless
ISI	intersymbol interference
LD	linear dispersion
LOS	line of sight
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
ML	maximum likelihood
MRC	maximum ratio combining
OSTBC	orthogonal space-time block codes
PC	personal computer
pdf	probability density function
PEP	pairwise error probability
PER	packet error rate
pmf	probability mass function
PSK	phase shift keying
QAM	quadrature amplitude modulation
SER	symbol error rate
SIMO	single-input multiple-output
SISO	single-input single-output
SM	spatial multiplexing

SNR	signal-to-noise ratio
ST	space-time
STBC	space-time block codes
STC	space-time coding
STTC	space-time trellis coding
SVD	singular value decomposition
ZMCSCG	zero-mean circularly symmetric complex Gaussian

Nomenclature

$\mathbf{H_{eff}}$	effective MIMO channel
$\mathcal{A}$	symbol alphabet
C	channel capacity
$\mathbf{C_x}$	covarince matrix of $\mathbf{x}$
$\lceil \cdot \rceil$	ceiling operation
E_s	mean symbol energy
$\|\cdot\|$	Frobenius norm (unless otherwise stated)
$(\cdot)^H$	complex conjugate (Hermitian transpose)
$H(\mathbf{x})$	entropy of a discrete random vector
$H(\mathbf{x},\mathbf{y})$	joint entropy of discrete random vectors
$H(\mathbf{y} \mid \mathbf{x})$	conditional entropy of discrete random vectors
$\mathbf{H}$	channel transfer matrix
$h(\mathbf{x})$	differential entropy of a continuous random vector
$h(\mathbf{x},\mathbf{y})$	joint differential entropy of continuous random vectors
$h(\mathbf{y} \mid \mathbf{x})$	conditional differential entropy of continuous random vectors
$h_{i,j}$	complex path gain from transmit antenna j to receive antenna i
$I(\mathbf{x};\mathbf{y})$	mutual information of discrete or continuous random variables
$I(\mathbf{x};\mathbf{y} \mid \mathbf{z})$	conditional mutual information
L	transmission time in symbol intervals for a block transmission
$\boldsymbol{\mu}_\mathbf{x}$	mean vector of $\mathbf{x}$
$\mathbf{N}$	noise block matrix
n_R	number of receive antennas
n_T	number of transmit antennas

$Q(\cdot)$	Q-function
ϱ	average SNR at receive antenna
$\mathbf{S}$	transmission block matrix
$\mathbf{s}$	transmit data vector
$(\cdot)^T$	transpose
$\mathbf{Y}$	receive block matrix
$\mathbf{y}$	receive data vector

1. Introduction

Wireless communications undergoes a dramatically change in recent years. More and more people are using modern communication services, thus increasing the need for more capacity in transmissions. Since bandwidth is a limited resource, the strongly increased demand in high transmission capacity has to be satisfied by a better use of existing frequency bands and channel conditions. One of the recent technical breakthroughs, which will be able to provide the necessary data rates, is the use of multiple antennas at both link ends. These systems are referred to as multiple-input multiple-output (MIMO) wireless systems. Initial theoretical studies from Foschini [1] and Telatar [2], as well as other pioneer works, have shown the potential of such systems.

Such MIMO systems are capable of realizing higher throughput without increasing bandwidth or transmit power. It is obvious that such a gain in transmissions rates and reliability comes at the cost of higher computational requirements. Fortunately, the feasibility of implementing the necessary signal processing algorithms is enabled by the corresponding increase of computational power of integrated circuits.

1.1. Why is MIMO Beneficial?

Motivated by these promising improvements, one question remains: why and how are these gains in rate and reliability possible? Basically, it turns out that there are two gains that can be realized by MIMO systems. They are termed as *diversity gain* and *spatial multiplexing gain*. First, to investigate the diversity gain in an introductory form, we take a look at the single input single output (SISO) system.

In the context of wireless transmissions, it is common knowledge that depending on the surrounding environment, a transmitted radio signal usually propagates through several different paths before it reaches the receiver, which is often referred to as *multipath propagation*. The radio signal received by the receiver antenna consists of the superposition of the various multipaths. If there is no line-of-sight (LOS) between the transmitter and the receiver, the attenuation coefficients corresponding to different paths are often assumed to be independent and identically distributed (iid). In this case the central limit theorem applies and the resulting path gain can be modeled as a complex Gaussian variable (which has an uniformly distributed phase and a Rayleigh distributed magnitude).

Due to this statistical behavior, the channel gain can sometimes become very small so that a reliable transmission is not always possible. To deal with this problem, communication engineers have thought of many possibilities to increase the so-called *diversity*. The higher the diversity is, the lower is the probability of a small channel gain.

Some common diversity techniques are time diversity and frequency diversity, where the same information is transmitted at different time instants or in different frequency bands, as well as spatial diversity, where one relies on the assumption that fading is at least partly independent between different points in space.

The concept of spatial diversity leads directly to an expansion of the SISO system. This enhancement is denoted as single-input multiple-output (SIMO) system. In such a system, we equip the receiver with multiple antennas. Doing so usually can be used to achieve a considerable performance gain, i.e. better link budget, but also co-channel interference can be better combatted. At the receiver, the signals are combined (i.e. if the phases of the transmission are known, in a coherent way) and the resulting advantage in performance is referred to as the diversity gain obtained from independent fading of the signal paths corresponding to the different antennas. This idea is well known and is used in many established communication systems, for example in the Global System for Mobile communications (GSM). It is clear that in the above described way, a base station can improve the uplink reliability and signal strength without adding any cost, size or power consumption to the mobile device.

As far as the ability to achieve performance in terms of diversity is concerned, system improvements are not only limited to the receiver side. If the transmitter side is also equipped with multiple antennas, we can either be in the multiple-input single-output (MISO) or multiple-input multiple-output (MIMO) case. A lot of research has been performed in recent years to exploit the possible performance gain of *transmit diversity*. The ways to achieve the predicted performance gain due to transmit diversity are various. Most of them are, loosely speaking, summarized under the concept of space-time coding (STC).

Besides the advantages of spatial diversity in MIMO systems, they can also offer a remarkably gain in terms of information rate or capacity [2]. This improvement is linked with the afore mentioned multiplexing gain. In fact, the advantages of MIMO are far more fundamental as it may have appeared to the reader so far. The underlying mathematical nature of MIMO systems, where data is transmitted over a matrix rather than a vector channel, creates new and enormous opportunities beyond the just described diversity effects. This was initially shown in [1], where the author points out how one may, under certain conditions, transmit a number of independent data streams *simultaneously* over the eigenmodes of a matrix channel, created by several transmit and receive antennas.

The gains achievable by a MIMO system in comparison to a SISO one can be described rigorously by information theory. A lot of research in the area of MIMO systems and STC is based on this mathematical framework introduced by Shannon [3]. The fundamental result of error free communication below a specific rate (depending on the actual signal-to-noise ratio (SNR)) in the limit of infinite length codes is also in the MIMO case an upper bound to all communication schemes. It can be used as a design criterion for transmission schemes as well as for comparison of different MIMO communication systems.

Overall, the potential increase in data rates and performance of wireless links offered by MIMO technology has proven to be so promising that we can except MIMO systems to be the cornerstone of many future wireless communication systems [4].

1.2. Topics Covered by This Diploma Thesis

As indicated by the title of this diploma thesis, MIMO communication systems will be investigated with special attention on information theoretic aspects. We tried to develop an objective look at the various different aspects of MIMO communication and it turned out that information theory is an appropriate tool with which an objective investigation of these systems is possible.

There has been a wide research in the area of MIMO with very different approaches. This work represents the topics we have investigated, and the basis literature therefore may be found in the bibliography.

To give a short overview, the thesis will start in Chapter 2 by discussing the MIMO system model, the channel capacity, and we will give a short introduction to maximum likelihood (ML) receivers. In Chapter 3, we investigate a very simple ST structure, the so-called spatial-multiplexing (SM) design, under the constraint of finite symbol alphabets. Chapter 4 introduces the theory of linear STBC and investigates how these systems behave in terms of system capacity and diversity gain. Finally, Chapter 5 treats the inherent tradeoff between the two performance measures: diversity and spatial-multiplexing gain. Additional material regarding proofs of different theorems or necessary definitions may be found in the Appendix.

2. MIMO Basics

The statistical nature of wireless communications and its various forms of appropriate description confronts us in the case of MIMO systems with an even more difficult problem. To be able to do a stringent analysis of MIMO systems and/or to make statements about performance gains, we need an adequate description of the underlying channel and its properties in terms of fading, time variance, linearity, correlation, etc. An adequate description of a MIMO channel is a research area of itself (see for example [5, 6]), and many publications have investigated the classification and description of MIMO transmission phenomena and their impact on MIMO performance parameters.

In this thesis, we are not interested in finding an optimal description for the MIMO channel in different scenarios, but we merely want to identify and analyze the key performance parameters of MIMO systems. To simplify matters, we will chose a very basic MIMO transmission model, which is not always satisfied in practice, but is strong enough to provide basic insights into MIMO communications while being sufficiently simple in its analytical representation.

This chapter explains the chosen MIMO transmission model, its analogies to a real communications environment, and the necessary assumptions to verify the choice of this representation. Furthermore, we investigate basic statistical properties of this model and derive necessary properties for a basic information theoretic analysis of MIMO systems. In addition, we can study fundamental issues of mutual information, and, of course, channel capacity.

With this first results in mind, we will take a closer look at the already mentioned *diversity* and *multiplexing gain*. The derived results will provide a basis for the MIMO system analysis in the subsequent chapters.

2.1. MIMO Transmission Model

We focus on a single-user communication model and consider a point-to-point link where the transmitter is equipped with n_T antennas and the receiver employs n_R antennas (see Figure 2.1). Next to the single user assumption in the depiction as point-to-point link, we suppose that no intersymbol interference (ISI) occurs. This implies that the bandwidth of the transmitted signal is very small and can be assumed frequency-flat (narrowband assumption), so that each signal path can be represented by a complex-valued gain factor. For practical purposes, it is common to model the channel as frequency-flat whenever the bandwidth of the system is smaller than the inverse of the delay spread of the channel; hence a wideband system operating where the delay spread is fairly small (for instance indoor scenes) may sometimes

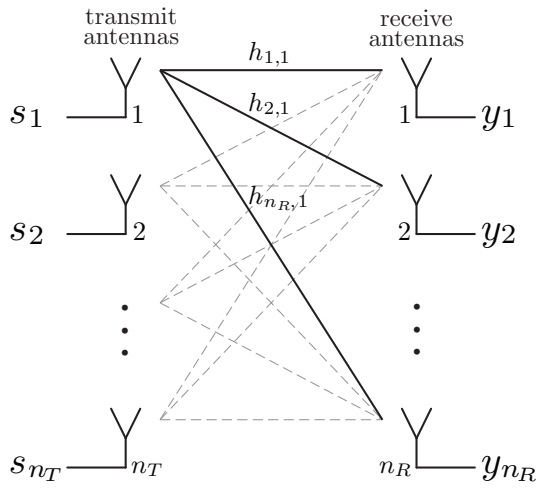

Figure 2.1.: A MIMO channel with n_T transmit and n_R receive antennas.

be considered as frequency-flat [7, 8]. If the channel is frequency selective, one could use an OFDM (orthogonal frequency-division multiplexing) system, to turn the MIMO channel into a set of parallel frequency-flat MIMO channels (see, e.g. [6, 9]), of which each obeys our stated assumptions.

In addition to these restrictions, we will further assume, that we are operating in a time-invariant setup. These assumptions allow us to use the standard complex-valued baseband representation of narrowband signals [10, 11] that can be written in a discrete form (omitting the dependency on time).

Now let $h_{i,j}$ be the complex-valued path gain from transmit antenna j to receive antenna i (the fading coefficient). If at a certain time instant the complex-valued signals $\{s_1, \ldots, s_{n_T}\}$ are transmitted via the n_T antennas, respectively, the received signal at antenna i can be expressed as

$$y_i = \sum_{j=1}^{n_T} h_{i,j} s_j + n_i,$$

where n_i represents additive noise, which will be treated later in this chapter. This linear relation can be easily written in a matrix framework. Thus, let $\mathbf{s}$ be a vector of size n_T containing the transmitted values, and $\mathbf{y}$ be a vector of size n_R containing the received values, respectively. Certainly, we have $\mathbf{s} \in \mathbb{C}^{n_T}$ and $\mathbf{y} \in \mathbb{C}^{n_R}$. Moreover, if we define the channel transfer matrix $\mathbf{H}$ as

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,n_T} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,n_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n_R,1} & h_{n_R,2} & \cdots & h_{n_R,n_T} \end{bmatrix},$$

we obtain

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}. \tag{2.1}$$

This is the same matrix notation as it is used in the majority of the publications in this field, e.g. [2]. This relation, denoting a transmission only over one symbol interval, is easily adapted to the case that several consecutive vectors $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_L\}$ are transmitted (here, L denotes the total number of symbol intervals used for transmission) over the channel. Therefore, we arrange the transmitted, the received and the noise vectors in the matrices

$$\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_L], \quad \mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_L], \quad \mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \cdots, \mathbf{n}_L],$$

respectively. The associated block transmission model is

$$\begin{bmatrix} y_{1,1} & \cdots & y_{1,L} \\ y_{2,1} & \cdots & y_{2,L} \\ \vdots & \ddots & \vdots \\ y_{n_R,1} & \cdots & y_{n_R,L} \end{bmatrix} = \begin{bmatrix} h_{1,1} & \cdots & h_{1,n_T} \\ h_{2,1} & \cdots & h_{2,n_T} \\ \vdots & \ddots & \vdots \\ h_{n_R,1} & \cdots & h_{n_R,n_T} \end{bmatrix} \begin{bmatrix} s_{1,1} & \cdots & s_{1,L} \\ s_{2,1} & \cdots & s_{2,L} \\ \vdots & \ddots & \vdots \\ s_{n_T,1} & \cdots & s_{n_T,L} \end{bmatrix} + \begin{bmatrix} n_{1,1} & \cdots & n_{1,L} \\ n_{2,1} & \cdots & n_{2,L} \\ \vdots & \ddots & \vdots \\ n_{n_R,1} & \cdots & n_{n_R,L} \end{bmatrix},$$

or equivalently,

$$\mathbf{Y} = \mathbf{HS} + \mathbf{N}.$$

2.1.1. Noise

After stating the general linear input-output relation of the MIMO channel under more or less general assumptions, we will now go a little bit into detail on the noise term of the transmission model (2.1).

In this thesis, the noise vectors $\{\mathbf{n}_l\}$ will be assumed to be spatially white circular Gaussian random variables with zero-mean and variance σ_N^2 per real and imaginary component. Thus,

$$\mathbf{n}_l \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, 2\sigma_N^2 \mathbf{I}),$$

where $\mathcal{N}_{\mathbb{C}}$ stands for a complex-valued multivariate Gaussian probability density function. Because we will need an exact definition of the complex-valued multivariate Gaussian probability density function, we will restate it here (compare [12, 11, 10]).

Definition 2.1.1 (Complex-valued Gaussian distribution)**.** *Let $\mathbf{x} \in \mathbb{C}^M$, then the probability density function (pdf) $f_{\mathbf{x}}(\boldsymbol{\xi})$ of $\mathbf{x}$ is given by*

$$f_{\mathbf{x}}(\boldsymbol{\xi}) = \frac{1}{\det(\pi \mathbf{C}_{\mathbf{n}})} \exp\left[-(\boldsymbol{\xi} - \boldsymbol{\mu}_{\mathbf{x}})^H \mathbf{C}_{\mathbf{x}}^{-1} (\boldsymbol{\xi} - \boldsymbol{\mu}_{\mathbf{x}}) \right],$$

where $\mathbf{C}_{\mathbf{x}} \triangleq \mathbb{E}\left\{ (\boldsymbol{\xi} - \boldsymbol{\mu}_{\mathbf{x}}) (\boldsymbol{\xi} - \boldsymbol{\mu}_{\mathbf{xk}})^H \right\}$ denotes the covariance matrix of $\mathbf{x}$, $\boldsymbol{\mu}_{\mathbf{x}} = \mathbb{E}\{\boldsymbol{\xi}\}$ denotes the mean vector of $\mathbf{x}$ and $(\cdot)^H$ stands for the complex conjugate (Hermitian transpose). Compactly, we write $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{C}_{\mathbf{x}})$.

There are at least two strong reasons for making the Gaussian assumption of the noise. First, Gaussian distributions tend to yield mathematical expressions that are relatively easy to deal with. Second, a Gaussian distribution of a disturbance term can often be motivated via the central limit theorem.

Throughout this thesis, we will also model the noise as temporally white. Although such an assumption is customary as well, it is clearly an approximation. In particular, $\mathbf{N}$ may contain interference consisting of modulated signals that are not perfectly white.

To conclude our examination of the noise term in our channel model, we summarize the statistical properties of the set of complex Gaussian vectors $\{\mathbf{n}_l\}, l = 1, \ldots, L$:

$$\mathbb{E}\{\mathbf{n}_l \mathbf{n}_l^H\} = 2\sigma_N^2 \mathbf{I},$$
$$\mathbb{E}\{\mathbf{n}_l \mathbf{n}_k^H\} = \mathbf{0}, \quad \text{for } l \neq k.$$

2.1.2. Fading

The elements of the matrix $\mathbf{H}$ correspond to the complex-valued channel gains between each transmit and receive antenna. For the purpose of assessing and predicting the performance of a communication system, it is necessary to postulate a statistical distribution of these elements [13]. This is also true to some degree for the design of well performing receivers, in the sense that knowledge of the statistical behavior of $\mathbf{H}$ could potentially be used to improve the performance of receivers.

Throughout this thesis, we will assume that the elements of the channel matrix $\mathbf{H}$ are zero-mean complex-valued Gaussian random variables with unit variance. This assumption is made to model the fading effects induced by local scattering in the absence of line-of-sight components. Consequently, the magnitudes of the channel gains $h_{i,j}$ have a Rayleigh distribution, or equivalently, $|h_{i,j}|^2$ are exponentially distributed [8, 14]. The presence of line-of-sight components can be modeled by letting $h_{i,j}$ have a Gaussian distribution with a non-zero mean (this is also called Ricean fading).

After having identified the possibilities to model the complex-valued channel path gains, it remains to check a possible correlation between these entries. In this work, we make a commonly made assumption on $\mathbf{H}$, i.e. that the elements of $\mathbf{H}$ are statistically independent. Although this assumption again tends to yield mathematical expressions that are easy to deal with, and allows the identification of fundamental performance limits, it is usually a rough approximation. In practice, the complex path gains $\{h_{i,j}\}$ are correlated by an amount that depends on the propagation environment as well as the polarization of the antenna elements and the spacing between them.

The channel correlation has a strong impact on the achievable system performance. Nevertheless, throughout this thesis, we will think of a rich scattering environment with enough antenna separation at the receiver and the transmitter, so that the entries of $\mathbf{H}$ can be assumed to be independent zero-mean complex Gaussian random variables with unit variance. This model is often popularly referred to as the iid (identically and independently distributed) Rayleigh fading MIMO channel model.

The fading itself will be modeled as block-fading, which means that the elements of $\mathbf{H}$ stay constant during the transmission of L data vectors $\mathbf{s}$ (or equivalently: during the whole transmission duration of $\mathbf{S}$) and change independently to another realization for the next block of L symbol periods. In practice, the duration L has to be shorter than the coherence time of

the channel, although in reality the channel path gains will change gradually. Nevertheless, we will use the block fading model for its simplicity.

2.1.3. Power Constraints, SNR Definition

The stated MIMO transmission model is now nearly ready to be investigated. What is still missing are declarations about the transmit power. Furthermore, we would like to derive expressions as a function of the signal-to-noise ratio (SNR) at the receiver, so we have to define it in terms of the already introduced quantities.

In the theoretical literature of MIMO systems, it is common to specify the power constraint on the input power in terms of an average power over the n_T transmit antennas. This may be written as

$$\frac{1}{n_T}\sum_{i=1}^{n_T}\mathbb{E}\Big\{|s_{i,l}|^2\Big\} = E_s, \quad \text{for } l = 1,\dots,L, \tag{2.2}$$

so that on average, we spend E_s in power at each transmit antenna. Here E_s denotes the mean symbol energy, as defined for example in [10], i.e. $E_s = \mathbb{E}\Big\{\big|s^{(i)}\big|^2\Big\}$ (here, i denotes the time index of the sent symbol), where the expectation is carried out over the symbol sequence (i.e. over i), which in case of a white symbol sequence reduces to an averaging over the symbol alphabet (see for example [11]).

Although this power constraint is a very common one, there is a variety of similar constraints that lead to the same basic information theoretic conclusions on MIMO transmission systems [15]. Since we will need other power constraints within this thesis, we will briefly restate them now. The power constraints can be written as

1. $\mathbb{E}\Big\{|s_{i,l}|^2\Big\} = E_s$, for $i = 1,\dots,n_T$ and $l = 1,\dots,L$, where no averaging over the transmit antennas is performed.
2. $\frac{1}{L}\sum_{l=1}^{L}\mathbb{E}\Big\{|s_{i,l}|^2\Big\} = E_s$, for $i = 1,\dots n_T$, what is quite similar to the power constraint (2.2), but here averaging is performed over time instead of space.
3. $\frac{1}{n_T\cdot L}\sum_{l=1}^{L}\sum_{i=1}^{n_T}\mathbb{E}\Big\{|s_{i,l}|^2\Big\} = E_s$, where we average over time and space. This can equivalently be expressed as $\frac{1}{n_T\cdot L}\mathbb{E}\big\{\operatorname{tr}\mathbf{S}\mathbf{S}^H\big\} = E_s$.

Since in most of our investigations, we want to derive expressions or curves depending on the SNR at a receive antenna, we will use a slightly adapted MIMO transmission model, in which we are using a redefinition of the power constraint. To motivate this, we would like to express the average signal-to-noise ratio at an arbitrary receive antenna. Because we transmit a total power of $n_T E_s$ over a channel with an average path gain of magnitude one[1] and a total noise power of $2\sigma_N^2$ at each receive antenna, we could state the SNR at a receive antenna as $\varrho = n_T E_s/(2\sigma_N^2)$. This would have the negative aspect, that our total transmitted power (and thus the receive SNR) is dependent on the number of transmit antennas. So, if we normalize the transmitted power by the number of transmit antennas n_T, we remove this small

[1] Because we defined our channel matrix $\mathbf{H}$ in the way that $\mathbb{E}\Big\{|h_{i,j}|^2\Big\} = 1$.

inconsistency. This also motivates a slightly different description of our MIMO transmission model:

$$\mathbf{Y} = \sqrt{\frac{\varrho}{n_T}}\mathbf{HS} + \mathbf{N}. \tag{2.3}$$

In this context, we have following constraints on our elements of the MIMO transmission model:

1. average magnitude of the channel path gains $\mathbb{E}\{\operatorname{tr}\mathbf{HH}^H\} = n_R n_T$,
2. average transmit power $\mathbb{E}\{\operatorname{tr}\mathbf{SS}^H\} = n_T L$ and
3. average noise variance $\mathbb{E}\{\operatorname{tr}\mathbf{NN}^H\} = n_R L$.

If these constraints are fulfilled, the factor $\sqrt{\varrho/n_T}$ ensures that ϱ is the average SNR at a receive antenna, independent of the number of transmit antennas (see for example also [16]).

2.2. Information Theoretic Background

Within this section we want to derive a basic understanding of the information theoretic theorems we need for an analysis of our MIMO transmission model. These findings are the basis for the identification of some of the performance gains already mentioned in the introduction. Furthermore, the explained concepts are crucial for the understanding of the investigations performed throughout the whole thesis.

We will not state any of the proofs of the following concepts and definitions. Some details may be found in the Appendix. For the proofs we refer to information theoretic works like [17, 18].

2.2.1. Introduction to Information Theory

Information theory is a very broad mathematical framework, which has its roots in communication theory, as founded by Shannon[2] in his well known paper [3]. An adequate description of all of the manifold applications of information theory would surely go beyond the scope of this diploma thesis. Nevertheless, it is of a great importance to define the basic concepts of information theory and explain its basic results in communication theory as they are needed throughout this work.

Within communication theory, information theory answers two fundamental questions: what is the ultimate data compression, and what is the ultimate transmission rate of any communications system [17]. Since a complete explanation of the basic definitions required for the subsequent development of the theory would again go beyond the scope of this thesis, we will only recapitulate the most important definitions. For the in communication theory educated reader, a repeat of those definitions would be rather uninformative, so we consequently moved them into the appendix, Section A.1.

[2] A popular scientific introduction to information theory and its applications is for example [19].

Figure 2.2.: A general communication system.

We will only work out the concept of capacity, which answers the second fundamental question concerning the ultimate transmission rate of any communications system. Therefore, we need to abstract the physical process of communication, as it can be seen in Figure 2.2. A sequence of source symbols (denoted as *message* in Figure 2.2) from some finite alphabet is mapped via an *encoder* on some sequence $\mathbf{x}$ of channel symbols, which then produces the output sequence $\mathbf{y}$ of the channel. The output sequence is random but has a distribution that depends on the specific input sequence. From the output sequence, we attempt to recover the transmitted message via a *decoder*.

Each of the possible input sequences induces a probability distribution on the output sequences. Since two different input sequences may give rise to the same output sequence, the inputs are confusable. By mapping the source messages into appropriate "widely spaced" input sequences to the channel, we can transmit a message with very low probability of confusion (or equivalently, error) at the decoder and reconstruct the source message at the output via the decoder. The maximum rate at which this can be done is called the capacity of the channel.

Definition 2.2.1 (Channel capacity)**.** *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be the input and output of a discrete vector channel with input alphabet* $\mathcal{X}$ *and output alphabet* $\mathcal{Y}$*, respectively. If the probability distribution of the output depends only on the input at that time, and is conditionally independent of previous channel inputs or outputs, the channel is said to be memoryless. This let us define the channel capacity of a discrete memoryless channel as*

$$C = \max_{p(\boldsymbol{\xi})} I(\mathbf{x}; \mathbf{y}),$$

where the maximum is taken over all possible input distributions $p(\boldsymbol{\xi})$*.*

Relying on this definition, we will recapitulate Shannon's second theorem, which gives an operational meaning to the definition of capacity as the number of bits we can transmit reliably over the channel. To do so, we need some basic definitions, which are (for the interested reader) given in the Appendix (Subsection A.1.5), so that a communication engineer does not have to read over them.

Nevertheless, for convenience, we repeat the definition of the code rate, because it is needed in a very direct sense for Shannon's second law.

Definition 2.2.2 (Rate of a (M, n) code)**.** *The rate* R *of an* (M, n) *code is*

$$R = \frac{\log M}{n} \ [bits]$$

per channel use.

Using this definition, we can relate codes of a given rate with their probability of error. In this context, we say that a rate R is *achievable* if there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes[3] such that the maximal probability of error ε tends to 0 as $n \to \infty$.

Following this concept, we can describe the capacity of a discrete memoryless channel as the supremum of all achievable rates. Thus, rates less than capacity yield arbitrarily small probability of error for sufficiently large block lengths. This leads directly to Shannon's second law, which is perhaps the most important theorem of information theory - the channel coding theorem.

Definition 2.2.3 (The channel coding theorem)**.** *All rates below capacity C are achievable. Specifically, for every rate $R < C$, there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes with maximum probability of error $\varepsilon \to 0$. Conversely, any sequence of $(\lceil 2^{nR} \rceil, n)$ codes with $\varepsilon \to 0$ must have $R \leq C$.*

To summarize the famous insights of the channel coding theorem, we can say that if one tries to transmit over a channel with capacity C with a rate $R \leq C$, there exists a code, such that $\epsilon \to 0$ for $n \to \infty$. In contrast, if one tries to transmit with a rate $R \geq C$, the probability of error is bound away from zero, i.e. $\varepsilon > 0$, for any code.

2.3. MIMO Information Theory

After having recapitulated the basic concepts of information theory, we now want to see how these can be applied to the analysis of a MIMO system. We will obtain expressions for the capacity of a MIMO channel and study its properties. Our intention is to offer a brief but consistent introduction into this topic.

2.3.1. Capacity of Deterministic MIMO Channels

We now study the capacity of a MIMO channel in the case that the channel matrix $\mathbf{H}$ is deterministic. Furthermore, we assume that the channel has a bandwidth of 1 Hz and fulfills all constraints of Section 2.1. Thus, we are investigating the vector transmission model

$$\mathbf{y} = \sqrt{\frac{\varrho}{n_T}}\mathbf{H}\mathbf{s} + \mathbf{n}. \tag{2.4}$$

In the following, we assume that the channel $\mathbf{H}$ is known to the receiver. This is a very common assumption, although in practice hard to realize. Channel knowledge at the receiver may be maintained via training and tracking, but time-varying environments can make it difficult to estimate the channel sufficiently exact.

The capacity of the MIMO channel is defined similar to Definition 2.2.1 as

$$C = \max_{p(\mathbf{s})} I(\mathbf{s}; \mathbf{y}). \tag{2.5}$$

[3]Here, $\lceil \cdot \rceil$ denotes the ceiling operation.

We start by using Equation (A.2) written as

$$I(\mathbf{s};\mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{s}), \tag{2.6}$$

where $H(\cdot)$ denotes the entropy, as defined in the Appendix[4]. Because $\mathbf{y}$ is specified through our linear MIMO transmission model, we can use the identity $H(\mathbf{y}|\mathbf{s}) = H(\mathbf{n}|\mathbf{s})$ (for the according theorem and proof, see Subsection A.1.4). Since according to our premises, the noise $\mathbf{n}$ and the transmit vector $\mathbf{s}$ are statistically independent, we can further write $H(\mathbf{y}|\mathbf{s}) = H(\mathbf{n})$. Therefore, Equation (2.6) simplifies to

$$I(\mathbf{s};\mathbf{y}) = H(\mathbf{y}) - H(\mathbf{n}).$$

By our assumptions about the noise term $\mathbf{n}$, the entropy $H(\mathbf{n})$ can be evaluated (see, e.g. [17, 18], or Subsection 3.1.1) as

$$H(\mathbf{n}) = \ln\det\left(\pi e\mathbf{C}_{\mathbf{n}}\right) = \ln\det\left(\pi e\mathbf{I}\right).$$

Thus, the maximization of the mutual information $I(\mathbf{s};\mathbf{y})$ reduces to a maximization of $H(\mathbf{y})$. To derive an expression for the entropy of $\mathbf{y}$, we first investigate its covariance matrix.

The covariance matrix of $\mathbf{y}$, $\mathbf{C}_{\mathbf{y}}$ satisfies

$$\mathbf{C}_{\mathbf{y}} = \mathbb{E}\{\mathbf{y}\mathbf{y}^H\} = \mathbb{E}\left\{\left(\sqrt{\frac{\varrho}{n_T}}\mathbf{H}\mathbf{s}+\mathbf{n}\right)\left(\sqrt{\frac{\varrho}{n_T}}\mathbf{H}\mathbf{s}+\mathbf{n}\right)^H\right\} = \frac{\varrho}{n_T}\mathbb{E}\{\mathbf{H}\mathbf{s}\mathbf{s}^H\mathbf{H}^H\} + \mathbb{E}\{\mathbf{n}\mathbf{n}^H\},$$

which can be further simplified to

$$\mathbf{C}_{\mathbf{y}} = \frac{\varrho}{n_T}\mathbf{H}\mathbb{E}\{\mathbf{s}\mathbf{s}^h\}\mathbf{H}^H + \mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \frac{\varrho}{n_T}\mathbf{H}\mathbf{C}_{\mathbf{s}}\mathbf{H}^H + \mathbf{C}_{\mathbf{n}},$$

where $\mathbf{C}_{\mathbf{s}}$ is the covariance matrix of $\mathbf{s}$. To evaluate the maximization of $H(\mathbf{y})$, we need the following theorem [2].

Theorem 2.3.1 (Entropy-maximizing property of a Gaussian random variable)**.** *Suppose the complex random vector $\mathbf{x} \in \mathbb{C}^n$ is zero-mean and satisfies $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{C}_{\mathbf{x}}$. Then the entropy of $\mathbf{x}$ is maximized if and only if $\mathbf{x}$ is a circularly symmetric complex Gaussian random variable with $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{C}_{\mathbf{x}}$.*

Proof. Let $f_{\mathbf{x}}(\boldsymbol{\xi})$ be any density function satisfying $\int_{\mathbb{C}^n} f_{\mathbf{x}}(\boldsymbol{\xi})\xi_i\xi_j^* d\boldsymbol{\xi} = (\mathbf{C}_{\mathbf{x}})_{i,j}$, $1 \le i,j \le n$. Furthermore, let

$$f_{\mathbf{x},G}(\boldsymbol{\xi}) = \frac{1}{\pi\det\mathbf{C}_{\mathbf{x}}}\exp\left[-\boldsymbol{\xi}^H\mathbf{C}_{\mathbf{x}}^{-1}\boldsymbol{\xi}\right]$$

denote a joint complex Gaussian distribution with zero-mean. Now, we can observe that $\int_{\mathbb{C}^n} f_{\mathbf{x},G}(\boldsymbol{\xi})\xi_i\xi_j^* d\boldsymbol{\xi} = (\mathbf{C}_{\mathbf{x}})_{i,j}$, and that $\log f_{\mathbf{x},G}(\boldsymbol{\xi})$ is a linear combination of the terms $\xi_i\xi_j^*$.

[4]For notational simplicity, we will not distinguish between the differential entropy $h(\cdot)$ and the entropy $H(\cdot)$ as defined in the Appendix, because they share the same interpretation and the appliance of the correct entropy definition follows without confusion from the given random variable. This notation will be kept throughout all information theoretic analyses in this thesis.

This means that by the construction of $f_{\mathbf{x},G}(\boldsymbol{\xi})$, the integral $\int_{\mathbb{C}^n} f_{\mathbf{x},G}(\boldsymbol{\xi}) \log f_{\mathbf{x},G}(\boldsymbol{\xi}) d\boldsymbol{\xi}$ can be split up in integrals $\int_{\mathbb{C}^n} f_{\mathbf{x},G}(\boldsymbol{\xi}) \xi_i \xi_j^* d\boldsymbol{\xi}$, of which each yields the same as $\int_{\mathbb{C}^n} f_{\mathbf{x}}(\boldsymbol{\xi}) \xi_i \xi_j^* d\boldsymbol{\xi}$. Therefore, by construction, we have the identity $\int_{\mathbb{C}^n} f_{\mathbf{x},G}(\boldsymbol{\xi}) \log f_{\mathbf{x},G}(\boldsymbol{\xi}) d\boldsymbol{\xi} = \int_{\mathbb{C}^n} f_{\mathbf{x}}(\boldsymbol{\xi}) \log f_{\mathbf{x},G}(\boldsymbol{\xi}) d\boldsymbol{\xi}$. Then,

$$\begin{aligned} H(f_{\mathbf{x}}(\boldsymbol{\xi})) - H(f_{\mathbf{x},G}(\boldsymbol{\xi})) &= -\int_{\mathbb{C}^n} f_{\mathbf{x}}(\boldsymbol{\xi}) \log f_{\mathbf{x}}(\boldsymbol{\xi}) d\boldsymbol{\xi} + \int_{\mathbb{C}^n} f_{\mathbf{x},G}(\boldsymbol{\xi}) \log f_{\mathbf{x},G}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= -\int_{\mathbb{C}^n} f_{\mathbf{x}}(\boldsymbol{\xi}) \log f_{\mathbf{x}}(\boldsymbol{\xi}) d\boldsymbol{\xi} + \int_{\mathbb{C}^n} f_{\mathbf{x}}(\boldsymbol{\xi}) \log f_{\mathbf{x},G}(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= \int_{\mathbb{C}^n} f_{\mathbf{x}}(\boldsymbol{\xi}) \log \frac{f_{\mathbf{x},G}(\boldsymbol{\xi})}{f_{\mathbf{x}}(\boldsymbol{\xi})} d\boldsymbol{\xi} \leq 0, \end{aligned}$$

with equality if and only if $f_{\mathbf{x}}(\boldsymbol{\xi}) = f_{\mathbf{x},G}(\boldsymbol{\xi})$. Thus $H(f_{\mathbf{x}}(\boldsymbol{\xi})) \leq H(f_{\mathbf{x},G}(\boldsymbol{\xi}))$, which concludes the proof[5]. □

Accordingly, the differential entropy $H(\mathbf{y})$ is maximized when $\mathbf{y}$ is zero-mean circularly symmetric complex Gaussian (ZMCSCG) [6]. This, in turn implies that $\mathbf{s}$ must be a ZMCSCG vector, with distribution that is completely characterized by $\mathbf{C_s}$. The differential entropy $H(\mathbf{y})$ is thus given by

$$H(\mathbf{y}) = \log\det\left(\pi e \mathbf{C_y}\right).$$

Therefore, the mutual information $I(\mathbf{s};\mathbf{y})$, in case of a deterministic channel $\mathbf{H}$, reduces to

$$I(\mathbf{s};\mathbf{y}) = \log\det\left(\mathbf{I} + \frac{\varrho}{n_T}\mathbf{H}\mathbf{C_s}\mathbf{H}^H\right) \quad [\text{bps/Hz}].$$

This is the famous "log-det" formula, firstly derived by Telatar [2]. In principle, we could denote the derived mutual information as a capacity since we maximized over all possible input distributions. Nevertheless, the above derivation does not tell us how to choose the covariance matrix of $\mathbf{s}$ to get the maximum mutual information. Therefore we keep the above notation. Thus, following Equation (2.5) we write the capacity of the MIMO channel (within our power constraint) as

$$C(\mathbf{H}) = \max_{\operatorname{tr}\mathbf{C_s}=n_T} \log\det\left(\mathbf{I} + \frac{\varrho}{n_T}\mathbf{H}\mathbf{C_s}\mathbf{H}^H\right) \quad [\text{bps/Hz}]. \tag{2.7}$$

2.3.2. Capacity of Random MIMO Channels

For a fading channel, the channel matrix $\mathbf{H}$ is a random quantity and hence the associated channel capacity $C(\mathbf{H})$ is also a random variable. To deal with this circumstances, we define the ergodic channel capacity as the average of (2.7) over the distribution of $\mathbf{H}$.

Definition 2.3.2 (Ergodic MIMO channel capacity)**.** *The ergodic channel capacity of the MIMO transmission model (2.4) is given by*

$$C_E = \mathbb{E}\left\{\max_{\operatorname{tr}\mathbf{C_s}=n_T} \log\det\left(\mathbf{I} + \frac{\varrho}{n_T}\mathbf{H}\mathbf{C_s}\mathbf{H}^H\right)\right\}. \tag{2.8}$$

[5]For notational simplicity, we denote the differential entropy by $H(\cdot)$ instead of $h(\cdot)$.

According to our information theoretic basics, this capacity cannot be achieved unless coding is employed across an infinite number of independently fading blocks.

After having identified the channel capacity in a fading MIMO environment, it remains to evaluate the optimal input power distribution, or covariance matrix $\mathbf{C_s}$ that maximizes Equation (2.8). The maximization depends on an important condition, we have not taken into account yet. Before being able to compute the maximization, we have to clarify if the transmitter, the receiver, or both have perfect knowledge of the channel state information (CSI). This is equivalent to the constraint that the channel matrix $\mathbf{H}$ is perfectly known to any or both sides of the communication system.

If the channel $\mathbf{H}$ is known to the transmitter, the transmit correlation matrix $\mathbf{C_s}$ can be chosen to maximize the channel capacity for a given realization of the channel. The main tool for performing this maximization is a technique, which is commonly referred to as "water-filling" [8] or "water-pouring algorithm" [6, 20, 21, 22], which we will not restate here. Besides the performance gain achievable, this method implicates a complex system, because the CSI has to be fed back to the transmitter.

Therefore, we chose to focus on the case of perfect CSI on the receiver side and no CSI at the transmitter. Of course, this implies that the maximization of Equation (2.8) is now more restricted than in the previous case. Nevertheless, Telatar [2], among others showed that the optimal signal covariance matrix has to be chosen according to

$$\mathbf{C_s} = \mathbf{I}.$$

This means that the antennas should transmit uncorrelated streams with the same average power. With this result, the ergodic MIMO channel capacity reduces to

$$C_E = \mathbb{E}\left\{ \log\det\left(\mathbf{I} + \frac{\varrho}{n_T}\mathbf{H}\mathbf{H}^H \right) \right\}. \tag{2.9}$$

Clearly, this is not the Shannon capacity in a true sense, since as mentioned before, a genie with channel knowledge can choose a signal covariance matrix that outperforms $\mathbf{C_s} = \mathbf{I}$. Nevertheless, we shall refer to the expression in Equation (2.9) as the *ergodic channel capacity* with CSI at the receiver and no CSI at the transmitter.

Now that we have specified our MIMO transmission system in a consistent way, and having identified the corresponding ergodic MIMO channel capacity, we would like to derive another notation of the capacity formula. Therefore, we take a closer look at the term $\mathbf{H}\mathbf{H}^H$ in Equation (2.9).

The term $\mathbf{H}\mathbf{H}^H$ is a $n_R \times n_R$ positive semi-definite Hermitian matrix (compare [6, 23]). Let the eigendecomposition of $\mathbf{H}\mathbf{H}^H$ be $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$, where $\mathbf{Q}$ is a $n_R \times n_R$ matrix satisfying $\mathbf{Q}\mathbf{Q}^H = \mathbf{Q}^H\mathbf{Q} = \mathbf{I}$ and $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_{n_R}\}$ with $\lambda_i \geq 0$ denoting the ordered eigenvalues ($\lambda_i \geq \lambda_{i+1}$) of $\mathbf{H}\mathbf{H}^H$. Then the channel capacity can be expressed as

$$C_E = \mathbb{E}\left\{ \log\det\left(\mathbf{I} + \frac{\varrho}{n_T}\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H \right) \right\}.$$

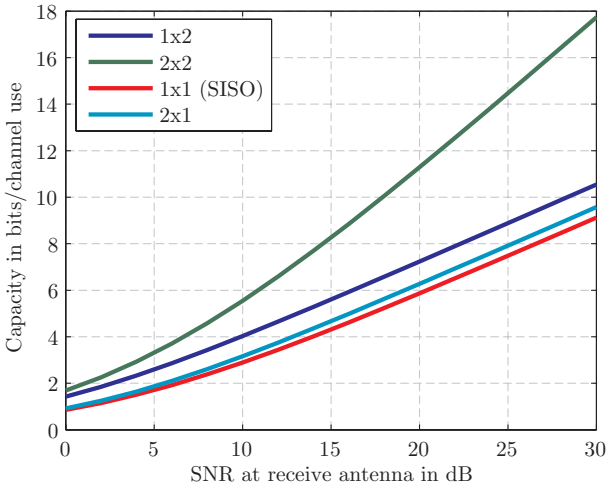

Figure 2.3.: Ergodic MIMO channel capacity versus the SNR with no CSI at the transmitter for various MIMO systems.

Using the identity $\det(\mathbf{I}+\mathbf{AB}) = \det(\mathbf{I}+\mathbf{BA})$ for matrices $\mathbf{A}$ of size $(m \times n)$ and $\mathbf{B}$ of size $(n \times m)$, together with the relation $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}$, the above equation simplifies to

$$C_E = \mathbb{E}\left\{\log\det\left(\mathbf{I} + \frac{\varrho}{n_T}\mathbf{\Lambda}\right)\right\} = \mathbb{E}\left\{\sum_{i=1}^{r}\log\left(1 + \frac{\varrho}{n_T}\lambda_i\right)\right\}, \tag{2.10}$$

where r is the rank of the channel $\mathbf{H}$. This expresses the capacity of the MIMO channel as the sum of the capacities of r SISO channels, each having a gain of $\lambda_i, i = 1, \ldots, r$.

Hence, the use of multiple antennas at the transmitter and receiver in a wireless link opens multiple scalar spatial pipes (also known as modes) between the transmitter and the receiver. This indicates the already mentioned *multiplexing gain*. To underline these insights, we did some numerical simulations, in which, according to our iid MIMO transmission model, we chose $\mathbf{H}$ to be formed by independent and Gaussian elements with unit variance. Figure 2.3 shows the ergodic MIMO channel capacity with no CSI at transmitter for various numbers of transmit and receive antennas. From this, we can see that the gain in capacity obtained by employing an extra receive antenna is around 3dB relative to the SISO system. This gain can be viewed as a consequence of the fact that the extra receive antenna effectively doubles the received power. The gain of a system with $n_T = 2, n_R = 1$ relative to the SISO system is small. As far as the ergodic channel capacity is concerned there is practically no benefit in adding an extra transmit antenna to the SISO system. Note also that the SIMO channel has a higher ergodic channel capacity than the MISO channel. Finally, the capacity of a system with $n_T = 2, n_R = 2$ is higher and faster growing with SNR than that of the SISO system. The growth of the ergodic channel capacity as a function of the number of antennas, which we observe in Figure 2.3, can be shown to obey a simple law. If we assume the channel $\mathbf{H}$ to be full rank, Equation (2.10) indicates that when the number of transmit and receive antennas are the same, the ergodic MIMO channel capacity increases linearly by the number of antennas.

In general, the capacity increases by the minimum of the number of transmit and receive antennas [20]. One can show that at high SNR, the ergodic channel capacity in terms of the received SNR can be described as

$$C_E \approx \min\{n_T, n_R\} \log\left(\frac{\varrho}{n_T}\right) + \sum_{k=|n_T-n_R|+1}^{\min\{n_T,n_R\}} \log(\chi_k), \tag{2.11}$$

where χ_k is a chi-squared random variable with $2k$ degrees of freedom [20]. Therefore, a 3dB increase in SNR results in $\min\{n_T, n_R\}$ extra bits of capacity at high SNR.

To further clarify our observation that the adding of transmit antennas to a system with a fixed number of receive antennas has a limited impact on the ergodic channel capacity, we investigate the ergodic capacity behavior for a large number of transmit antennas (see, e.g. [21]). In the mentioned case, using the law of large numbers, one can show that $\mathbf{H}^H\mathbf{H}/n_T \rightarrow \mathbf{I}$ almost surely. As a result, the ergodic channel capacity is $n_R \log(1+\varrho)$ for large n_T. This bound is rapidly reached, thus explaining the limited gain of adding extra transmit antennas. Similar investigations can be performed for a fixed number of transmit antennas, where the capacity gain for adding one additional receive antenna also gets smaller if the number of receive antennas gets large.

Now, it just remains to point out that a correlation of the entries of the channel matrix $\mathbf{H}$, as it might be induced by not well separated antennas at either the transmit or receiver side or by not sufficiently "good" scatterers, can of course influence the shape of the presented curves massively (see, e.g. [24], or [5]). In general, correlation of $\mathbf{H}$ reduces the gains obtained in MIMO channels, as long as we are investigating a MIMO system with perfect CSI on the receiver side. Recent research, as e.g. [25] show that if only partial CSI at the receiver is available, correlation may be used to improve capacity gains.

2.3.3. Outage Capacity

Since the MIMO channel capacity (2.7) is a random variable, it is meaningful to consider its statistical distribution. A particularly useful measure of its statistical behavior is the so-called outage capacity. Outage analysis quantifies the level of performance (in this case capacity) that is guaranteed with a certain level of reliability. In analogy to [6], we define:

Definition 2.3.3 (Outage MIMO channel capacity)**.** *The $q\%$ outage capacity $C_{out}(q)$ is defined as the information rate that is guaranteed for $(100-q)\%$ of the channel realizations, i.e.*

$$Pr(C(\mathbf{H}) \leq C_{out}(q)) = q\%.$$

The outage capacity is often a more relevant measure than the ergodic channel capacity, because it describes in some way the quality of the channel. This is due to the fact that the outage capacity measures how far the instantaneous rate supported by the channel is spread, in terms of probability. So if the rate supported by the channel is spread over a wide range, the outage capacity for a fixed probability level can get small, whereas the ergodic channel capacity may be high. To get further insights, we performed some numerical simulations, again based

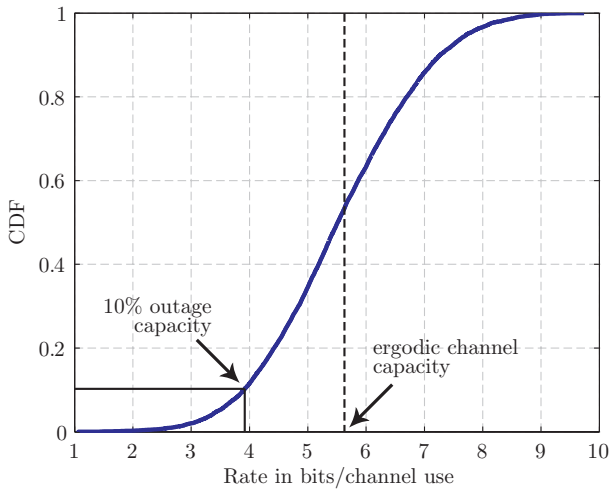

Figure 2.4.: CDF of channel capacity for the iid MIMO channel model with $n_T = n_R = 2$ at a SNR of 10dB.

on our iid channel model, as we did it in the simulations for Figure 2.3. Before showing an ensemble of outage capacity curves, we want to note that the outage capacity may be seen in the cumulative density function (cdf) of the instantaneous rate supported by the channel given by Equation (2.7). Figure 2.4 shows the cdf of the MIMO channel capacity $C(\mathbf{H})$ in the case of perfect CSI only on the receiver side. Note that the ergodic capacity is the mean channel capacity and is not necessarily equal to the median information rate. This figure shows that if the outage capacity and the ergodic channel capacity are largely separated, the slope of the cdf curve of the instantaneous rate will be small.

In the case of iid entries in $\mathbf{H}$, the outage channel capacity shows the same behavior versus the SNR, as the ergodic channel capacity does. To further investigate these relations, we simulated the 1% outage channel capacity for different antenna arrangements. The results are shown in Figure 2.5. Considering the outage capacity, a significant gain is obtained by employing an extra receive or transmit antenna (compared to the SISO channel). This gain is much larger than the corresponding gain in ergodic channel capacity.

2.3.4. Performance Limits

The previous capacity results can be illustrated in a variety of ways, but a particularly interesting comparison is obtained when the outage probability is plotted as a function of SNR for a given rate. If we consider a block-fading MIMO transmission model, we assume that the channel is randomly drawn from a given distribution and is held constant during the transmission of one codeword [6]. Now this means that for any non-zero signaling rate there is always a finite probability that the channel is unable to support it. If we use very large block size and optimal coding, the packet error rate (PER) performance will be binary - the packet is always

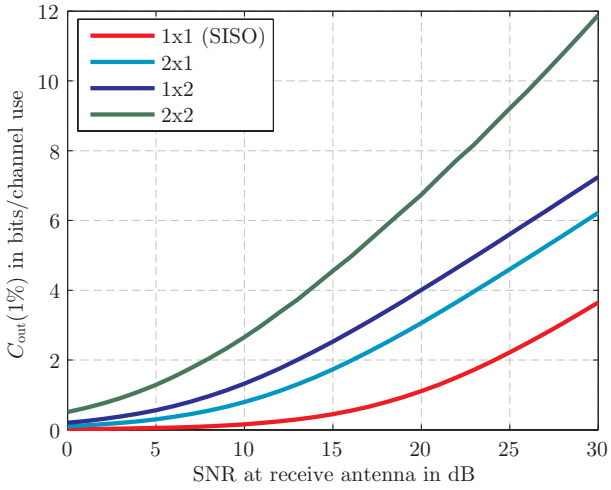

Figure 2.5.: 1% outage channel capacity versus the SNR for various antenna configurations.

decoded successfully if the channel supports the rate and is always in error otherwise. Therefore, if the transmitter does not know the channel, the PER will equal the outage probability for that signaling rate (outage capacity).

Hence, for a system with unity bandwidth transmitting packets with a bit rate R, the probability of a packet error can be lower bounded as

$$\Pr(\text{PE}) \geq \Pr(C(\mathbf{H}) < R) = \Pr\left[\log\det\left(\mathbf{I} + \frac{\varrho}{n_T}\mathbf{H}\mathbf{H}^H\right) < R\right].$$

To visualize these relations, we did some numerical simulations for different antenna constellations. The results are plotted in Figure 2.6. Notice that these curves imply that the PER cannot be zero and that it depends on the SNR much like bit error rate (BER) curves in uncoded (or suboptimally coded) AWGN channels. The magnitude of the slope of the PER curve has been shown to be $n_T n_R$ for fixed rate transmission and at high enough SNR (compare [26, 6], but also Chapter 5). To further clarify these coherences, we simulated the outage probability versus the SNR for different rates in a $n_T = 2, n_R = 2$ MIMO transmission system. The obtained surface (see Figure 2.7) is called the "signaling limit surface" [6] and represents the fundamental limit of fading channels, assuming optimal coding and large enough block size [14]. The region to the right of this surface is the achievable region, where practical signaling and receivers operate. We have seen that with optimal coding, for a given transmission rate, we can trade SNR for PER at $n_T n_R$ slope (this is the already mentioned *diversity gain*), and conversely for a fixed PER, we can trade SNR for transmission rate at $\min\{n_T, n_R\}$ slope (denoting the also already mentioned *multiplexing gain*). Thus, if we hold the rate constant and we increase SNR, the PER will decrease at $n_T n_R$ slope, what is the equivalency to Figure 2.6. On the other hand, if we fix the PER and increase SNR, in the limit of infinite SNR, the rate will increase at $\min\{n_T, n_R\}$ slope. This corresponds to Equation (2.11) and to Figure 2.3. These connections will be further investigated in Chapter 5.

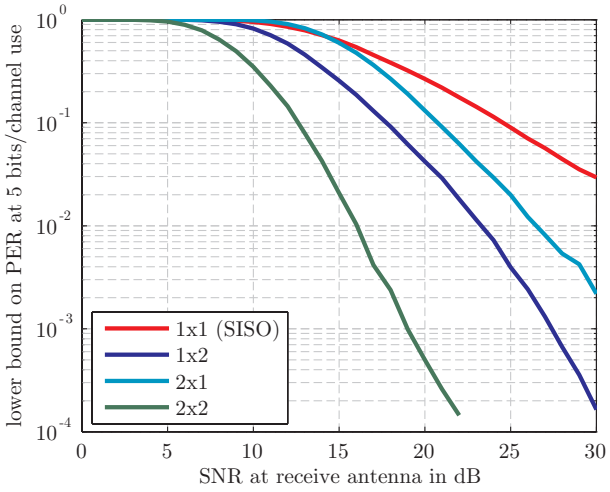

Figure 2.6.: Lower bound on PER at rate 5 bits/channel use for different antenna constellations.

2.4. MIMO Systems

Throughout this thesis, we want to investigate different MIMO *systems*. To clarify our notation, we now want to define what we understand under a so-called "MIMO system". Figure 2.1 shows the basic MIMO channel, which can be seen as a spatial multiplexing (SM) transmission system, since for a transmission, the source symbols are directly mapped in an independent manner onto the n_T transmit antennas. After passing the channel, the n_T independent data streams result in a vector of n_R complex receive symbols, which are used to receive the transmitted data symbols (for example, by using an optimal maximum likelihood (ML) receiver).

Having this in mind, one can ask the legitimate question, whether there exists a better way to map the data symbols onto the transmit antennas. If such an arbitrary mapping is allowed, our MIMO channel is enhanced by a so-called *space-time encoder* and a *space-time decoder*, whereas the latter includes the detection of the data symbols (see Figure 2.8). The term space-time results from the fact, that we allow the symbol mapping to act in the dimensions space and time, which means that each symbol can be spread over a number of symbol times and over a number of transmit antennas. Such a space-time (ST) coded system can be described by a number of performance parameters, which are summarized in the following:

1. Parameters based on mutual information:
   - Ergodic channel capacity,
   - outage capacity,
   - multiplexing gain.

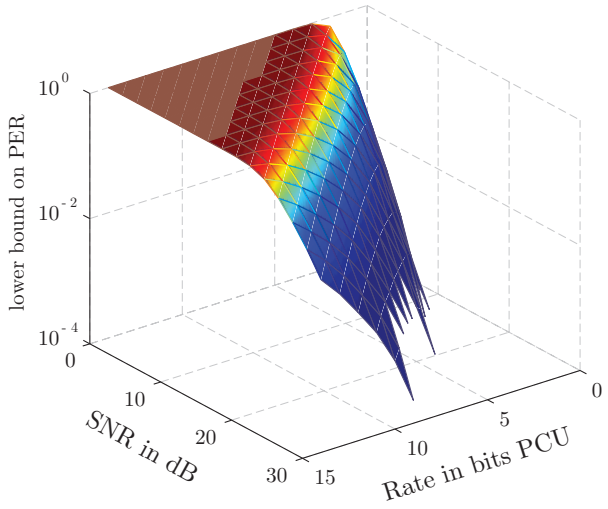

Figure 2.7.: Signaling limit surface for a $n_T = 2, n_R = 2$ MIMO channel.

2. Parameters concerning the error performance:
   - BER versus SNR performance,
   - diversity gain.
3. Parameters concerning the MIMO channel structuring:
   - Efficient encoding,
   - efficient decoding.
4. Tradeoffs between diversity and multiplexing gain.

Throughout this thesis, we focuse on information theoretic aspects, and we will mostly use the optimal ML receiver for decoding of the complex data symbols. Therefore, we will briefly recapitulate this important receiver type.

2.4.1. ML Receiver

In Section 2.1, we stated the block-wise MIMO transmission model. This means that we encode a number of data symbols into a block matrix $\mathbf{S}$, which is then transmitted using L symbol time instants. Beyond the scope of this thesis, there also exists another way of transmitting, which is called space-time trellis coding (STTC), [27, 28, 29, 30].

The ML receiver of the MIMO block transmission model from Section 2.1 can be stated as (see [22], but also [31]): Consider the MIMO block transmission model $\mathbf{Y} = \sqrt{\varrho/n_T}\mathbf{HS} + \mathbf{N}$ and let $\mathcal{S}$ denote the codebook of $\mathbf{S}$ (i.e., $\mathcal{S}$ denotes the set of all possible transmit matrices).

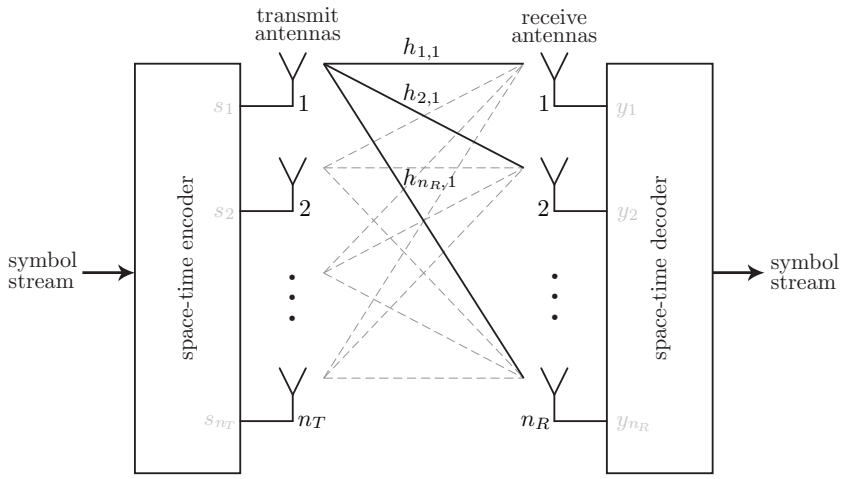

Figure 2.8.: A general MIMO system.

Then, the ML decision $\hat{\mathbf{S}}_{\mathrm{ML}}$ on the transmit matrix $\mathbf{S}$ with $\mathbf{H}$ perfectly known at the receiver is given by

$$\hat{\mathbf{S}}_{\mathrm{ML}} = \arg\min_{\mathbf{S}\in\mathcal{S}} \left\| \mathbf{Y} - \sqrt{\frac{\varrho}{n_T}}\mathbf{H}\mathbf{S} \right\|^2,$$

where $\|\cdot\|$ denotes the Frobenius norm (see Subsection A.3.1).

For convenience, we present a short proof, which follows the way of arguing in [11]. According to [11], the ML decision rule for $\mathbf{H}$ known at the receiver is given by

$$\mathbf{S}_{\mathrm{ML}} = \arg\max_{\mathbf{S}\in\mathcal{S}} f_{\mathbf{Y}|\mathbf{S},\mathbf{H}}(\boldsymbol{\eta}|\boldsymbol{\sigma}, \mathbf{H}_0),$$

where $f_{\mathbf{Y}|\mathbf{S},\mathbf{H}}(\boldsymbol{\eta}|\boldsymbol{\sigma}, \mathbf{H}_0)$ denotes the conditional pdf of $\mathbf{Y}$ given $\mathbf{S}$ and $\mathbf{H}$.

With the assumption that $\mathbf{N}$ is independent of $\mathbf{S}$ and $\mathbf{N}$, one can easily show that

$$f_{\mathbf{Y}|\mathbf{S},\mathbf{H}}(\boldsymbol{\eta}|\boldsymbol{\sigma}, \mathbf{H}_0) = f_{\mathbf{N}}\left(\boldsymbol{\eta} - \sqrt{\frac{\varrho}{n_T}}\mathbf{H}_0\boldsymbol{\sigma}\right).$$

Using the fact that the noise is spatially and temporally white, we obtain

$$f_{\mathbf{N}}(\boldsymbol{\nu}) = \prod_{i=1}^{L} f_{\mathbf{n}_i}(\boldsymbol{\nu}_i), \tag{2.12}$$

where $\mathbf{n}_i$ denotes the i-th column of $\mathbf{N}$. The pdf $f_{\mathbf{n}_i}(\boldsymbol{\nu}_i)$ has already been described in Definition 2.1.1. Since all noise vectors $\mathbf{n}_i$ are zero mean and have the same covariance matrix $\mathbf{C}_{\mathbf{n}_i} \equiv \mathbf{C}_{\mathbf{n}} = \mathbf{I}$, (2.12) reduces to

$$f_{\mathbf{N}}(\boldsymbol{\nu}) = \prod_{i=1}^{L} \frac{1}{\pi^{n_R}} \exp\left[-\mathbf{n}_i^H \mathbf{n}_i\right] = \frac{1}{\pi^{n_R L}} \exp\left[-\sum_{i=1}^{L} \mathbf{n}_i^H \mathbf{n}_i\right].$$

By using the identity $\|\mathbf{N}\|^2 = \sum_{i=1}^{L} \mathbf{n}_i^H \mathbf{n}_i$ we obtain $f_{\mathbf{N}}(\boldsymbol{\nu}) = 1/\pi^{n_R L} \exp\left[-\|\mathbf{N}\|^2\right]$. Thus, the ML receiver may be written as

$$\hat{\mathbf{S}}_{\mathrm{ML}} = \arg\max_{\mathbf{S}\in\mathcal{S}} \frac{1}{\pi^{n_R L}} \exp\left[-\left\|\mathbf{Y} - \sqrt{\frac{\varrho}{n_T}}\mathbf{H}\mathbf{S}\right\|^2\right] = \arg\min_{\mathbf{S}\in\mathcal{S}} \left\|\mathbf{Y} - \sqrt{\frac{\varrho}{n_T}}\mathbf{H}\mathbf{S}\right\|^2.$$

2.5. Diversity

So far, we studied how multiple antennas can enhance channel capacity. Now we discuss how antennas can also offer diversity. Diversity provides the receiver with multiple (ideally independent) observations of the same transmitted signal [6]. Each observation constitutes a diversity branch. With an increase in the number of independent diversity branches, the probability that all branches fade at the same time reduces. Thus, diversity techniques stabilize the wireless link leading to an improvement in link reliability or error rate.

To clarify matters, we will have a closer look at a very simple example, following the argumentation of [6]. Assume that we transmit a data symbol s drawn from a scalar constellation with unit average energy. This symbol is now transmitted in a way that we can provide M identically independently Rayleigh faded versions of this symbol to the receiver. If the fading is frequency-flat, the receiver sees

$$y_i = \sqrt{\frac{\varrho}{M}} h_i s + n_i, \quad i = 1, \ldots, M,$$

where ϱ is the average SNR for each of the M diversity branches and y_i is the received signal corresponding to the ith diversity branch. Furthermore, h_i denotes the channel path gain and n_i is additive ZMCSCG noise with variance 1 in the ith diversity branch, whereas the noise from different branches is assumed to be statistically independent.

If we provide a receiver with multiple versions of the transmitted symbol s, it can be shown that the post-processing SNR can be maximized by a technique called maximum ratio combining (MRC). With perfect CSI at receiver, the M signals are combined according to $z = \sum_{i=1}^{M} h_i^* y_i$, and thus the post-processing SNR η is given by $\eta = 1/M \sum_{i=1}^{M} |h_i|^2 \varrho$.

With ML detection, the corresponding probability of symbol error is given by [11, 10]

$$P_E \approx \bar{\mathcal{N}} Q\left(\sqrt{\frac{\eta d_{\min}^2}{2}}\right),$$

where $\bar{\mathcal{N}}$ denotes the number of nearest neighbors, $d_{\min}$ labels the minimum distance in the underlying scalar symbol constellation and $Q(\cdot)$ is the Q-function [10]. The error probability can be further bounded applying the Chernoff bound $Q(x) \leq \exp\left(-x^2/2\right)$:

$$P_e \leq \bar{\mathcal{N}} \exp\left[-\frac{\varrho d_{\min}^2}{4M} \sum_{i=1}^{M} |h_i|^2\right].$$

By averaging this instant error probability with respect to the fading gains $h_i, i = 1, \ldots, M$, the upper bound [7, 6]

$$\mathbb{E}\{P_e\} \leq \bar{\mathcal{N}} \prod_{i=1}^{M} \frac{1}{1 + \varrho d_{\min}^2/(4M)}$$

is obtained. In the high SNR regime, the preceding equation may be further simplified to

$$\mathbb{E}\{P_e\} \leq \bar{\mathcal{N}} \left(\frac{\varrho d_{\min}^2}{4M} \right)^{-M},$$

which makes it absolutely clear that diversity effects the slope of the symbol error rate (SER) curve. The slope of the SER curve on a log-log scale, compared to the slope of a SISO system terms the mentioned *diversity gain*. Clearly, multiple antennas on the transmitter and/or receiver side can lead to this kind of performance gain.

The answer to the question how we can achieve the maximum diversity gain $n_T n_R$ in a MIMO system is, among other related topics, part of the following chapters.

3. SM under Finite Symbol Alphabet Constraint

In Chapter 2 we gave an overview over some of the most important topics concerning MIMO transmission systems. Now we want to take a closer look at some of the introduced concepts and extend the information-theoretic insights. The results stated here represent a detailed derivation of a part of the work already performed in [32], or in [9].

A spatial multiplexing MIMO system describes a very simple, yet basic mapping method of the complex data symbols onto the transmit antennas. Reconsidering Figure 2.8, the ST encoder multiplexes the symbol stream onto the n_T transmit antennas. Thus, using such a transmission system, we are transmitting n_T independent data symbols. The according transmission relation can be written as

$$\mathbf{y} = \sqrt{\frac{\varrho}{n_T}}\mathbf{H}\mathbf{s} + \mathbf{n},$$

where $\mathbf{s}$ denotes the vector of multiplexed data symbols. Looking at the above equation, we can see that it is exactly the same relation, we presupposed in Subsection 2.3.1 to derive the MIMO channel capacity. This implies that the channel capacity (as well as the ergodic channel capacity) equals the system capacity. Unattached by this easy observation, it remains of interest to examine the behavior of the mutual information in the case of constraints on the symbol alphabet of the transmitted complex data symbols. In Subsection 2.3.1, we also showed that the capacity achieving input distribution of the complex data symbols has to be ZMCSCG. This implies that the symbols are drawn from an continuous constellation, i.e. $s_i \in \mathbb{C}$ for $i = 1, \ldots, n_T$. Equivalently, we could say that the transmit vector $\mathbf{s}$ is drawn from $\mathbb{C}^{n_T}$. Throughout this chapter, we restrict the complex data symbols to be drawn from a finite symbol alphabet. This can be stated as

$$s_i \in \mathcal{A}, \quad i = 1, \ldots, n_T,$$

where $\mathcal{A}$ denotes the symbol alphabet (like QAM or PSK, see e.g. [10]). The associated constraint on the transmitted data vector $\mathbf{s}$ can thus be written as

$$\mathbf{s} \in \mathcal{A}^{n_T} = \mathcal{S}.$$

For sake of notational simplicity, we drop the factor $\sqrt{\varrho/n_T}$ and operate only on the transmission relation $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$. This implies that we allow an arbitrary noise variance σ_N^2 and an arbitrary symbol energy E_s as also mentioned in the beginning of Subsection 2.1.3. Besides its notational benefit, we achieve a better comparability of our obtained result with the one derived in [32].

3.1. Evaluating the Mutual Information

The mathematical concepts used in this section belong to the basic ideas of information theory. Since we do not want to stress the patience of the information theory versed reader, we collected the necessary definitions in the Appendix, Section A.1.

The main goal of the current section is to derive an expression for the mutual information between the input $\mathbf{s}$ and the output $\mathbf{y}$ of the SM-MIMO system under the constraint $\mathbf{s} \in \mathcal{S}$ as defined above. We try to obtain a relation, which we can use to perform numerical simulations, as also done in [32, 9]. In our described MIMO transmission model, we assume that the receiver has perfect knowledge about the realization of $\mathbf{H}$. This knowledge can be analytical expressed in the way that the channel output consists of the pair $(\mathbf{y}, \mathbf{H})$. Accordingly, the relevant mutual information between input $\mathbf{s}$ and output $(\mathbf{y}, \mathbf{H})$, we are trying to find, is $I(\mathbf{s}; (\mathbf{y}, \mathbf{H}))$. To derive analytical expressions, we first try to rewrite the mutual information $I(\mathbf{s}; (\mathbf{y}, \mathbf{H}))$ in a way that we can identify terms which can be evaluated more easily. Therefore, we use the chain rule for mutual information (Definition A.1.11) to obtain

$$I(\mathbf{s}; (\mathbf{y}, \mathbf{H})) = I((\mathbf{y}, \mathbf{H}); \mathbf{s}) = I(\mathbf{y}; \mathbf{s}|\mathbf{H}) + I(\mathbf{H}; \mathbf{s}) = I(\mathbf{s}; \mathbf{H}) + I(\mathbf{y}; \mathbf{s}|\mathbf{H}), \tag{3.1}$$

where we used the symmetric property of the mutual information. Certainly, the result can also be obtained by using the entropy description of the mutual information and applying the chain rule for entropy (Definition A.1.10). By using relation (A.3), and taking into account that the transmit vector $\mathbf{s}$ and the channel matrix $\mathbf{H}$ are statistically independent, the mutual information $I(\mathbf{s}; \mathbf{H})$ can be simplified to

$$I(\mathbf{s}; \mathbf{H}) = H(\mathbf{s}) - H(\mathbf{s}|\mathbf{H}) = H(\mathbf{s}) - H(\mathbf{s}) = 0,$$

where the statistical independence of $\mathbf{s}$ and $\mathbf{H}$ is a direct consequence of our assumption that the transmitter has no CSI at all. Thus, following the arguments given in Subsection 2.3.2, we obtain

$$I(\mathbf{s}; (\mathbf{y}, \mathbf{H})) = I(\mathbf{y}; \mathbf{s}|\mathbf{H}).$$

Using Relation A.1, we get

$$I(\mathbf{s}; \mathbf{y}|\mathbf{H}) = H(\mathbf{s}|\mathbf{H}) - H(\mathbf{s}|\mathbf{y}, \mathbf{H}).$$

Using a straightforward combination of Definition A.1.6 and Definition A.1.5, we obtain

$$I(\mathbf{s}; \mathbf{y}|\mathbf{H}) = \int_{\Omega} f_{\mathbf{H}}(\mathbf{H}_0) H(\mathbf{s}|\mathbf{H} = \mathbf{H}_0) d\mathbf{H}_0 - \int_{\Omega} f_{\mathbf{H}}(\mathbf{H}_0) H(\mathbf{s}|\mathbf{y}, \mathbf{H} = \mathbf{H}_0) d\mathbf{H}_0,$$

where Ω denotes the support set of the channel matrix $\mathbf{H}$, and $\mathbf{H}_0$ denotes a specific channel realization. This can be further simplified to

$$\begin{aligned} I(\mathbf{s}; \mathbf{y}|\mathbf{H}) &= \int_{\Omega} f_{\mathbf{H}}(\mathbf{H}_0) \left[H(\mathbf{s}|\mathbf{H} = \mathbf{H}_0) - H(\mathbf{s}|\mathbf{y}, \mathbf{H} = \mathbf{H}_0) \right] d\mathbf{H}_0 \\ &= \int_{\Omega} f_{\mathbf{H}}(\mathbf{H}_0) I(\mathbf{s}; \mathbf{y}|\mathbf{H} = \mathbf{H}_0) d\mathbf{H}_0. \end{aligned}$$

Following the notation in [32] (compare also [17]), we express the mutual information $I(\mathbf{s};\mathbf{y}|\mathbf{H})$ in terms of an expectation, since the integral over $p(\mathbf{H}_0)$ may be interpreted as such, i.e.:

$$I(\mathbf{s};\mathbf{y}|\mathbf{H}) = \mathbb{E}\{I(\mathbf{s};\mathbf{y}|\mathbf{H}=\mathbf{H}_0)\}\,,$$

where the expectation is performed with respect to $\mathbf{H}$. Finally, we can write $I(\mathbf{s};(\mathbf{y},\mathbf{H}))$ as

$$I(\mathbf{s};(\mathbf{y},\mathbf{H})) = \mathbb{E}\{I(\mathbf{y};\mathbf{s}|\mathbf{H}=\mathbf{H}_0)\} = \mathbb{E}\{H(\mathbf{y}|\mathbf{H}=\mathbf{H}_0)\} - \mathbb{E}\{H(\mathbf{y}|\mathbf{s},\mathbf{H}=\mathbf{H}_0)\}\,, \tag{3.2}$$

and it remains to evaluate the terms $\mathbb{E}\{H(\mathbf{y}|\mathbf{s},\mathbf{H}=\mathbf{H}_0)\}$ and $\mathbb{E}\{H(\mathbf{y}|\mathbf{H}=\mathbf{H}_0)\}$.

3.1.1. Evaluation of $\mathbb{E}\{H(\mathbf{y}|\mathbf{s},\mathbf{H}=\mathbf{H}_0)\}$

To obtain an useable expression for $\mathbb{E}\{H(\mathbf{y}|\mathbf{s},\mathbf{H}=\mathbf{H}_0)\}$, let us first of all drop the expectation (and thus the integral over $f_{\mathbf{H}}(\mathbf{H}_0)$) and consider the term $H(\mathbf{y}|\mathbf{s},\mathbf{H}=\mathbf{H}_0)$. By a straightforward use of Definition A.1.2 we obtain the conditional differential entropy $H(\mathbf{y}|\mathbf{s},\mathbf{H}=\mathbf{H}_0)$ as

$$H(\mathbf{y}|\mathbf{s},\mathbf{H}=\mathbf{H}_0) = -\iint_{\Omega_{\mathbf{y},\mathbf{s}}} f_{\mathbf{y},\mathbf{s}|\mathbf{H}}(\mathbf{y},\mathbf{s}|\mathbf{H}) \log f_{\mathbf{y}|\mathbf{s},\mathbf{H}}(\mathbf{y}|\mathbf{s},\mathbf{H}) d\mathbf{y} d\mathbf{s},$$

where $\Omega_{\mathbf{y},\mathbf{s}}$ denotes the support region of the conditional pdf[1] $f_{\mathbf{y},\mathbf{s}|\mathbf{H}}(\mathbf{y},\mathbf{s}|\mathbf{H})$. Unfortunately, an analytical expression for the conditional pdf $f(\mathbf{y},\mathbf{s}|\mathbf{H})$ cannot be derived. Therefore, we use Definition A.1.5 to rewrite

$$H(\mathbf{y}|\mathbf{s},\mathbf{H}=\mathbf{H}_0) = \sum_{\mathbf{s}\in\mathcal{S}} p_{\mathbf{s}}(\mathbf{s}) H(\mathbf{y}|\mathbf{s}=\mathbf{s}_0,\mathbf{H}=\mathbf{H}_0),$$

with $\mathcal{S} = \mathcal{A}^{n_T}$ denoting the set of all possible transmitted data vectors $\mathbf{s}$. The occurring conditional entropy, with $\mathbf{s}$ and $\mathbf{H}$ already being observed as $\mathbf{s}_0$ and $\mathbf{H}_0$ respectively, $H(\mathbf{y}|\mathbf{s}=\mathbf{s}_0,\mathbf{H}=\mathbf{H}_0)$, is defined similar as in A.1.6, i.e.

$$H(\mathbf{y}|\mathbf{s}=\mathbf{s}_0,\mathbf{H}=\mathbf{H}_0) = -\int_{\Omega_{\mathbf{y}}} f_{\mathbf{y}|\mathbf{s},\mathbf{H}}(\mathbf{y}|\mathbf{s},\mathbf{H}) \log f_{\mathbf{y}|\mathbf{s},\mathbf{H}}(\mathbf{y}|\mathbf{s},\mathbf{H}) d\mathbf{y},$$

where $\Omega_{\mathbf{y}}$ denotes the support region of $f_{\mathbf{y}|\mathbf{s},\mathbf{H}}(\mathbf{y}|\mathbf{s},\mathbf{H})$. Similar to the arguments concerning the mutual information, the above integral can be interpreted as expectation over $\mathbf{y}$, although the expectation is not performed by integrating over $f_{\mathbf{y}}(\mathbf{y})$, but instead over $f_{\mathbf{y}|\mathbf{s},\mathbf{H}}(\mathbf{y}|\mathbf{s},\mathbf{H})$. Nevertheless, for notational comparability to [32], we keep this notation, and write

$$H(\mathbf{y}|\mathbf{s}=\mathbf{s}_0,\mathbf{H}=\mathbf{H}_0) = -\mathbb{E}\{\log f(\mathbf{y}|\mathbf{s},\mathbf{H})\}\,.$$

Having identified these relations, it remains to compute the conditional pdf $f_{\mathbf{y}|\mathbf{s},\mathbf{H}}(\mathbf{y}|\mathbf{s},\mathbf{H})$. Fortunately, this can be easily done within our MIMO transmission model assumptions (see Section 2.1). The conditional probability density function can be reformulated as

$$f_{\mathbf{y}|\mathbf{s},\mathbf{H}}(\mathbf{y}|\mathbf{s},\mathbf{H}) = f_{\mathbf{H}\mathbf{s}+\mathbf{n}|\mathbf{s},\mathbf{H}}(\mathbf{y}|\mathbf{s},\mathbf{H}) = f_{\mathbf{n}|\mathbf{s},\mathbf{H}}(\mathbf{y}-\mathbf{H}\mathbf{s}|\mathbf{s},\mathbf{H}) = f_{\mathbf{n}}(\mathbf{y}-\mathbf{H}\mathbf{s}), \tag{3.3}$$

[1] For notational simplicity, from now on we dropped the distinction between the random variable and the dependent variable of its pdf, e.g. $f_{\mathbf{x}}(\boldsymbol{\xi})$ is written as $f_{\mathbf{x}}(\mathbf{x})$. This only limits the amount of variables needed but does not lead to any confusion.

where we used the fact that the noise $\mathbf{n}$ and the transmitted vector $\mathbf{s}$ are statistically independent. According to our premises in Section 2.1, $\mathbf{n}$ is a joint complex Gaussian random variable (as in Definition 2.1.1), and we can write

$$\begin{aligned} H(\mathbf{y}|\mathbf{s}=\mathbf{s}_0,\mathbf{H}=\mathbf{H}_0) &= -\int_{\Omega_\mathbf{y}} f_{\mathbf{y}|\mathbf{s},\mathbf{H}}(\mathbf{y}|\mathbf{s},\mathbf{H})\log f_{\mathbf{y}|\mathbf{s},\mathbf{H}}(\mathbf{y}|\mathbf{s},\mathbf{H})d\mathbf{y} \\ &= -\int_{\Omega_\mathbf{y}} f_\mathbf{n}(\mathbf{y}-\mathbf{H}\mathbf{s})\log f_\mathbf{n}(\mathbf{y}-\mathbf{H}\mathbf{s})d\mathbf{y}. \end{aligned}$$

By a simple variable change $\mathbf{u}=\mathbf{y}-\mathbf{H}\mathbf{s}$, with $\partial\mathbf{y}/\partial\mathbf{u}=1$, we obtain

$$H(\mathbf{y}|\mathbf{s}=\mathbf{s}_0,\mathbf{H}=\mathbf{H}_0) = -\int_{\Omega_\mathbf{u}} f_\mathbf{n}(\mathbf{u})\log f_\mathbf{n}(\mathbf{u})d\mathbf{u} = -\mathbb{E}\{\log f_\mathbf{n}(\mathbf{u})\},$$

where we again used the notation in terms of an expectation over $\mathbf{u}$. Using Definition 2.1.1, this can be simplified to

$$H(\mathbf{y}|\mathbf{s}=\mathbf{s}_0,\mathbf{H}=\mathbf{H}_0) = -\mathbb{E}\{\log f_\mathbf{n}(\mathbf{u})\} = \ln\det(\pi\mathbf{C}_\mathbf{n}) + \mathbb{E}\left\{\mathbf{u}^H\mathbf{C}_\mathbf{n}^{-1}\mathbf{u}\right\}.$$

By using the identity $\mathbf{x}^H\mathbf{C}_\mathbf{n}^{-1}\mathbf{x} = \mathrm{tr}(\mathbf{x}\mathbf{x}^H\mathbf{C}_\mathbf{n}^{-1})$ and using the fact that $\mathbb{E}\{\cdot\}$ and $\mathrm{tr}(\cdot)$ commute, we obtain

$$H(\mathbf{y}|\mathbf{s}=\mathbf{s}_0,\mathbf{H}=\mathbf{H}_0) = \ln\det(\pi\mathbf{C}_\mathbf{n}) + \mathbb{E}\left\{\mathrm{tr}(\mathbf{u}\mathbf{u}^H)\mathbf{C}_\mathbf{n}^{-1}\right\} = \ln\det(\pi\mathbf{C}_\mathbf{n}) + \mathrm{tr}(\mathbb{E}\left\{\mathbf{u}\mathbf{u}^H\right\}\mathbf{C}_\mathbf{n}^{-1}).$$

Now, remembering that $\mathbf{u}=\mathbf{y}-\mathbf{H}\mathbf{s}$ and taking into account that we are momentarily considering $\mathbf{H}$ and $\mathbf{s}$ as deterministic variables, by following similar arguments as in [18], it follows that $\mathbb{E}\left\{\mathbf{u}\mathbf{u}^H\right\}$ carried out with respect to $\mathbf{u}$, and thus representing the covariance matrix $\mathbf{C}_\mathbf{u}$, is equal to $\mathbf{C}_\mathbf{n}$. This finally results in the well-known entropy (see e.g. [18, 17])

$$H(\mathbf{y}|\mathbf{s}=\mathbf{s}_0,\mathbf{H}=\mathbf{H}_0) = \ln\det(\pi\mathbf{C}_\mathbf{n}) + \mathrm{tr}\,\mathbf{I} = \ln\det(\pi e\mathbf{C}_\mathbf{n}) = H(\mathbf{n}).$$

This result is rather intuitive, since it corresponds directly to Theorem A.1.12 and reflects the fact that the uncertainty about $\mathbf{y}$ given $\mathbf{s}$ and $\mathbf{H}$ just depends on the noise vector $\mathbf{n}$.

Since $\mathbf{n}$ is iid Gaussian with variance $2\sigma_N^2$ per component, we have $\mathbf{C}_\mathbf{n} = 2\sigma_N^2\mathbf{I}$ and

$$\det\mathbf{C}_\mathbf{n} = \prod_{i=1}^{n_R} 2\sigma_N^2 = (2\sigma_N^2)^{n_R}.$$

Taking into account that $\det(\alpha\mathbf{X}) = \alpha^N\det\mathbf{X}$ for any square matrix $\mathbf{X}\in\mathbb{C}^{N\times N}$, we obtain

$$H(\mathbf{y}|\mathbf{s}=\mathbf{s}_0,\mathbf{H}=\mathbf{H}_0) = H(\mathbf{n}) = \ln\left[\pi^{n_R}e^{n_R}\left(2\sigma_N^2\right)^{n_R}\right] = n_R\ln\left(2\pi e\sigma_N^2\right).$$

Hence, we can state our result as

$$\mathbb{E}\{H(\mathbf{y}|\mathbf{s},\mathbf{H}=\mathbf{H}_0)\} = \mathbb{E}\left\{\sum_{\mathbf{s}\in\mathcal{S}} p_\mathbf{s}(\mathbf{s})H(\mathbf{y}|\mathbf{s}=\mathbf{s}_0,\mathbf{H}=\mathbf{H}_0)\right\} = \mathbb{E}\left\{\sum_{\mathbf{s}\in\mathcal{S}} p_\mathbf{s}(\mathbf{s})n_R\ln\left(2\pi e\sigma_N^2\right)\right\},$$

where the expectation is carried out with respect to $\mathbf{H}$, but can certainly be simplified to

$$\mathbb{E}\{H(\mathbf{y}|\mathbf{s},\mathbf{H}=\mathbf{H}_0)\} = n_R\ln\left(2\pi e\sigma_N^2\right).$$

To express the result in bits and not in nats, one has to change the base of the logarithm, i.e. $H_b(\mathbf{x}) = \log_b(a) H_a(\mathbf{x})$, where the subindex denotes the basis of the used logarithm in the calculation of the entropy, and thus

$$\mathbb{E}\{H(\mathbf{y}|\mathbf{s}, \mathbf{H} = \mathbf{H}_0)\} = \log_2(e)\, n_R \ln\left(2\pi e \sigma_N^2\right) \text{ [bits]}. \tag{3.4}$$

3.1.2. Evaluation of $\mathbb{E}\{H(\mathbf{y}|\mathbf{H} = \mathbf{H}_0)\}$

After having computed the first term $\mathbb{E}\{H(\mathbf{y}|\mathbf{s}, \mathbf{H} = \mathbf{H}_0)\}$ of the mutual information in Equation (3.2), we now evaluate the second term $\mathbb{E}\{H(\mathbf{y}|\mathbf{H} = \mathbf{H}_0)\}$. Similar as in Subsection 3.1.1, we firstly concentrate on $H(\mathbf{y}|\mathbf{H} = \mathbf{H}_0)$.

According to our Definition A.1.6 and Definition A.1.5, $H(\mathbf{y}|\mathbf{H} = \mathbf{H}_0)$ is given by

$$H(\mathbf{y}|\mathbf{H} = \mathbf{H}_0) = -\int_{\Omega_\mathbf{y}} f_{\mathbf{y}|\mathbf{H}}(\mathbf{y}|\mathbf{H}) \log f_{\mathbf{y}|\mathbf{H}}(\mathbf{y}|\mathbf{H}) d\mathbf{y} = -\mathbb{E}\left\{\log f_{\mathbf{y}|\mathbf{H}}(\mathbf{y}|\mathbf{H})\right\}, \tag{3.5}$$

where $\Omega_\mathbf{y}$ denotes the support set of the conditional pdf $f_{\mathbf{y}|\mathbf{H}}(\mathbf{y}|\mathbf{H})$ and the expectation is carried out with respect to $\mathbf{y}$, given $\mathbf{H}$. With the total probability theorem, the unknown probability density function $f_{\mathbf{y}|\mathbf{H}}(\mathbf{y}|\mathbf{H})$ can be expressed as

$$f(\mathbf{y}|\mathbf{H}) = \sum_{\mathbf{s} \in \mathcal{C}^{n_T}} p_\mathbf{s}(\mathbf{s}) f_{\mathbf{y}|\mathbf{H},\mathbf{s}}(\mathbf{y}|\mathbf{H}, \mathbf{s}),$$

so that we can use $f_{\mathbf{y}|\mathbf{H},\mathbf{s}}(\mathbf{y}|\mathbf{H}, \mathbf{s}) = f_\mathbf{n}(\mathbf{y} - \mathbf{H}\mathbf{s})$ (see Equation 3.3).

Now, it seems appropriate to specify the properties of the generation of $\mathbf{s}$, the pmf $p_\mathbf{s}(\mathbf{s})$. In our simulations, the marginal probability mass function $p_\mathbf{s}(\mathbf{s})$ is chosen uniformly, so that all possible transmit vectors $\mathbf{s} \in \mathcal{S}$ are drawn equally likely. Using n_T transmit antennas, the symbol vector $\mathbf{s}$ has to be of size $n_T \times 1$. Assuming symbol alphabet sizes of power two (e.g. QAM constellations), there are 2^{M_a} constellation points (i.e. the cardinality of the symbol alphabet is $|\mathcal{A}| = 2^{M_a}$), so that for each symbol (constellation point), M_a bits are assigned. This means that there are $(2^{M_a})^{n_T} = 2^{M_a n_T} = |\mathcal{S}|$ possible signal vectors. Thus

$$p_\mathbf{s}(\mathbf{s}) = \frac{1}{2^{M_a n_T}}, \quad \mathbf{s} \in \mathcal{A}^{n_T} = \mathcal{S}.$$

Using

$$f_\mathbf{n}(\mathbf{n}) = \frac{1}{(2\pi\sigma_N^2)^{n_R}} \exp\left[-\frac{\|\mathbf{n}\|^2}{2\sigma_N^2}\right]$$

(compare Definition 2.1.1), we obtain

$$f_{\mathbf{y}|\mathbf{H}}(\mathbf{y}|\mathbf{H}) = \sum_{\mathbf{s} \in \mathcal{C}^{n_T}} p_\mathbf{s}(\mathbf{s}) f_\mathbf{n}(\mathbf{y} - \mathbf{H}\mathbf{s}) = \frac{1}{2^{M_c n_T}} \frac{1}{(2\pi\sigma_N^2)^{n_R}} \sum_{\mathbf{s} \in \mathcal{S}} \exp\left[-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2}{2\sigma_N^2}\right].$$

Taking this result, we are able to evaluate $H(\mathbf{y}|\mathbf{H} = \mathbf{H}_0)$

$$\begin{aligned} H(\mathbf{y}|\mathbf{H} = \mathbf{H}_0) &= -\mathbb{E}\left\{\log f_{\mathbf{y}|\mathbf{H}}(\mathbf{y}|\mathbf{H})\right\} \\ &= -\mathbb{E}\left\{\log\left\{\frac{1}{2^{M_c n_T}} \frac{1}{(2\pi\sigma_N^2)^{n_R}} \sum_{\mathbf{s} \in \mathcal{S}} \exp\left[-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2}{2\sigma_N^2}\right]\right\}\right\}, \end{aligned}$$

where the expectation is taken over $\mathbf{y}$ given $\mathbf{H}$ as in Equation (3.5). Finally, we can use this expression to calculate $\mathbb{E}\{H(\mathbf{y}|\mathbf{H}=\mathbf{H}_0)\}$

$$\mathbb{E}\{H(\mathbf{y}|\mathbf{H}=\mathbf{H}_0)\} = -\mathbb{E}\left\{\mathbb{E}\left\{\log\left\{\frac{1}{2^{M_c n_T}}\frac{1}{(2\pi\sigma_N^2)^{n_R}}\sum_{\mathbf{s}\in\mathcal{S}}\exp\left[-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{s}\|^2}{2\sigma_N^2}\right]\right\}\right\}\right\}, \tag{3.6}$$

where the outer expectation with respect to $\mathbf{H}$ and the inner expectation is with respect to $\mathbf{y}$ given $\mathbf{H}$.

3.1.3. Result: Mutual Information for Finite Symbol Alphabets

The resulting mutual information $I(\mathbf{s};(\mathbf{y},\mathbf{H}))$ can accordingly be calculated by using Equation (3.2) together with results (3.4) and (3.6):

$$\begin{aligned} I(\mathbf{s};(\mathbf{y},\mathbf{H})) &= \mathbb{E}\{H(\mathbf{y}|\mathbf{H}=\mathbf{H}_0)\} - \mathbb{E}\{H(\mathbf{y}|\mathbf{s},\mathbf{H}=\mathbf{H}_0)\} \\ &= -\mathbb{E}\left\{\mathbb{E}\left\{\log\left\{\frac{1}{2^{M_c n_T}}\frac{1}{(2\pi\sigma_N^2)^{n_R}}\sum_{\mathbf{s}\in\mathcal{S}}\exp\left[-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{s}\|^2}{2\sigma_N^2}\right]\right\}\right\}\right\} \\ &\quad - \log_2(e)\, n_R \ln\left(2\pi e\sigma_N^2\right). \end{aligned} \tag{3.7}$$

In general, Equation (3.7) has no closed form expression. It may nevertheless be evaluated using numerical methods (like Monte Carlo simulations).

3.2. Numerical Simulations

To visualize the result (3.7), we did some Monte Carlo simulations by drawing a sufficient number of independent channel and noise realizations. Within Equation (3.7), we sum up different results of the Frobenius norm $\|\mathbf{y}-\mathbf{H}\mathbf{s}\|^2$ for all possible values of $\mathbf{s}$. If the number $|\mathcal{S}|$ is not too large, we can simulate the mutual information by averaging over all possible transmit vectors $\mathbf{s}$. The numerical complexity is very high, so that we wrote a distributed simulation in MATLAB. The computation over the SNR range from 0 to 30dB was performed on a cluster of five PCs. Since the simulations took on average between two and six days, we ask the reader to excuse the not perfectly smooth curves.

3.2.1. Simulation Results

We present two figures for a $n_T = n_R = 2$ (Figure 3.1) and a $n_T = n_R = 4$ (Figure 3.2) MIMO system, respectively. The ergodic channel capacity is also shown for comparison. It can be seen that the mutual information curves huddle against the ergodic channel capacity curve at low SNR, whereas they saturate in the high SNR region. The reason for this saturation is rather obvious, since in the high SNR region the influence of noise on the receiver gets arbitrarily low, so that the mutual information equals the entropy of the source times the number of

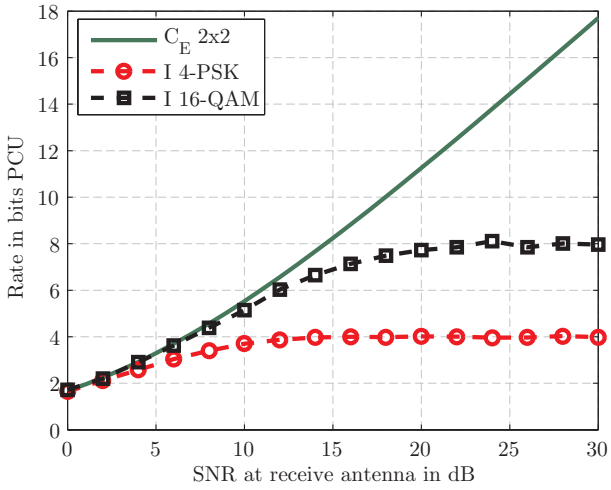

Figure 3.1.: Mutual information curves for finite symbol alphabets compared to the ergodic channel capacity for a $n_T = 2, n_R = 2$ system.

independently transmitted symbols n_T. Given a fixed symbol alphabet $\mathcal{A}$ with cardinality $|\mathcal{A}| = 2^{M_a}$ and with equally likely drawn symbols, this amounts to be $n_T \cdot \log|\mathcal{A}| = n_T M_a$. The equivalence of the mutual information curves to the ergodic channel capacity in the low SNR region also provides an indication to the already mentioned equality of the MIMO system capacity of the investigated SM system and the ergodic MIMO channel capacity. This provides an alternative proof of the conclusion that a SM design[2] is able to *achieve capacity* (the ergodic channel capacity and the MIMO system capacity for Gaussian inputs are coinciding). We summarize our finding of the saturating mutual information for finite symbol alphabets in the following subsection.

3.2.2. Bound for Mutual Information in the $\varrho \to \infty$ Case

Theorem 3.2.1 (Saturation value of mutual information in case of finite symbol alphabets)**.** *Let our MIMO transmission model be given by* $\mathbf{y} = \mathbf{H}\mathbf{s}+\mathbf{n}$ *(compare beginning of this chapter), and let* $\mathbf{s}$ *be drawn uniformly from a vector symbol alphabet* $\mathcal{S} = \mathcal{A}^{n_T}$*. Then the mutual information saturates for* $\varrho \to \infty$ *to*

$$\lim_{\varrho\to\infty} I(\mathbf{s};(\mathbf{y},\mathbf{H})) = M_a n_T.$$

Proof. Let us assume that we adjust the SNR solely by scaling the signal energy of the transmit vector $\mathbf{s}$. Therefore, we can exchange the limit $\varrho \to \infty$ by $E_s = \mathbb{E}\left\{\|\mathbf{s}\|^2\right\} \to \infty$. Because of this restriction, we only have to look at the term $\|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2$ and how it behaves as $E_s \to \infty$

[2] We sometimes use the term "design" to denote a specific ST system, as often done in literature.

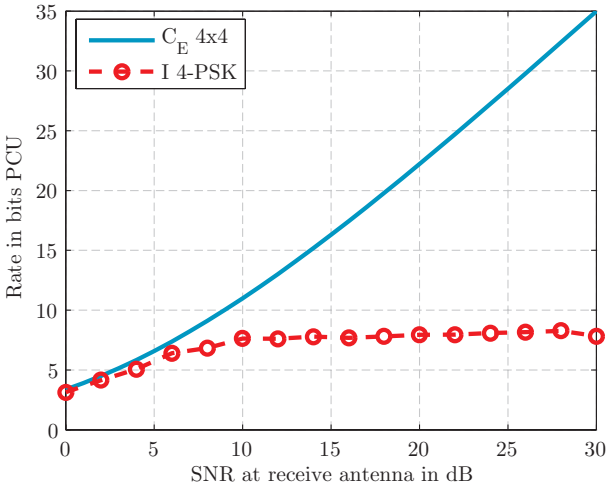

Figure 3.2.: Mutual information curves compared to the ergodic channel capacity of a $n_T = 4, n_R = 4$ system for a 4-PSK symbol alphabet.

(compare Equation (3.7)). If we assume an arbitrary vector $\mathbf{s}_0$ to be transmitted, the received vector gets $\mathbf{y} = \mathbf{H}\mathbf{s}_0 + \mathbf{n}$. Thus, we can write

$$\lim_{E_s\to\infty} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2 = \lim_{E_s\to\infty} \|\mathbf{H}(\mathbf{s}_0 - \mathbf{s}) + \mathbf{n}\|^2 = \begin{cases} \infty, & \text{for } \mathbf{s}_0 \neq \mathbf{s}, \\ \|\mathbf{n}\|^2, & \text{for } \mathbf{s}_0 = \mathbf{s}. \end{cases}$$

Thus, the sum over $\mathbf{s} \in \mathcal{S}$ in Equation (3.7) is equal to

$$\sum_{\mathbf{s}\in\mathcal{S}} \exp\left[-\frac{1}{2\sigma_N^2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2\right] = \exp\left[-\frac{\|\mathbf{n}\|^2}{2\sigma_N^2}\right].$$

To compute the two expectations, we note that with a fixed transmit vector $\mathbf{s}_0$ (thus $\mathbf{s}$ assumed to be given), the inner expectation with respect to $\mathbf{y}$ reduces to an expectation over $\mathbf{n}$. According to our presumptions, the expectation of $\|\mathbf{n}\|^2$ is $\mathbb{E}\left\{\|\mathbf{n}\|^2\right\} = n_R \cdot 2\sigma_N^2$, and thus we obtain

$$\begin{aligned}\lim_{E_s\to\infty} I(\mathbf{s};(\mathbf{y},\mathbf{H})) &= -\mathbb{E}\left\{\log \frac{1}{2^{M_a n_T}} \frac{1}{(2\pi\sigma_N^2)^{n_R}} \exp\left[-\frac{\mathbb{E}\left\{\|\mathbf{n}\|^2\right\}}{2\sigma_N^2}\right]\right\} - n_R \log(2\pi e\sigma_N^2) \\ &= M_a n_T + n_R \log(2\pi e\sigma_N^2) - n_R \log(2\pi e\sigma_N^2) = M_a n_T,\end{aligned}$$

where we used the logarithm to base 2. This concludes the proof. □

3.3. Error Performance

We now investigate the SM-based MIMO design with respect to error performance. Is SM capable of realizing the full diversity gain $n_T n_R$? The answer to this question will give us insight into the motivation of space-time coding (STC). We will introduce a general framework of STC error analysis, as introduced in [29] (compare also [28]). We present this theory in the general context of arbitrary ST block transmissions, which can be easily simplified to fit the SM design. As we want to derive expressions for the general case of MIMO ST block transmissions, we now rely on the MIMO block transmission model of Equation (2.3), i.e. $\mathbf{Y} = \sqrt{\varrho/n_T}\mathbf{H}\mathbf{S} + \mathbf{N}$.

To derive an upper bound on the error probability, we investigate the pairwise error probability (PEP) of the ML receiver. This is the probability that the receiver mistakes the transmitted codeword $\mathbf{S}^{(i)}$ for another codeword $\mathbf{S}^{(j)}$. According to [6, 10], the PEP of the ML receiver with perfect CSI, is given by

$$\Pr\left(\mathbf{S}^{(i)} \to \mathbf{S}^{(j)}|\mathbf{H}\right) = Q\left(\sqrt{\frac{\varrho\left\|\mathbf{H}(\mathbf{S}^{(i)} - \mathbf{S}^{(j)})\right\|^2}{2n_T}}\right).$$

According to [29], we can apply the Chernoff bound to obtain

$$\Pr\left(\mathbf{S}^{(i)} \to \mathbf{S}^{(j)}|\mathbf{H}\right) \leq \exp\left[-\frac{\varrho}{4n_T}\left\|\mathbf{H}\boldsymbol{\Delta}_{i,j}\right\|^2\right],$$

where we define $\boldsymbol{\Delta}_{i,j} = \mathbf{S}^{(i)} - \mathbf{S}^{(j)}$ to be the $n_T \times L$ codeword difference matrix. In [29], it is further shown that the PEP averaged over all iid channel realizations (Rayleigh distributed), may be upper-bounded by

$$\Pr\left(\mathbf{S}^{(i)} \to \mathbf{S}^{(j)}\right) \leq \prod_{k=1}^{r(\boldsymbol{\Gamma}_{i,j})}\left(1 + \frac{\varrho\lambda_k(\boldsymbol{\Gamma}_{i,j})}{4n_T}\right)^{-n_R},$$

where in analogy to [6], we introduced the matrix $\boldsymbol{\Gamma}_{i,j} = \boldsymbol{\Delta}_{i,j}\boldsymbol{\Delta}_{i,j}^H$. The terms $r(\boldsymbol{\Gamma}_{i,j})$ and $\lambda_k(\boldsymbol{\Gamma}_{i,j})$, denote the rank and the k-th non-zero eigenvalue of $\boldsymbol{\Gamma}_{i,j}$ respectively. In the high SNR regime, this may further be simplified to

$$\Pr\left(\mathbf{S}^{(i)} \to \mathbf{S}^{(j)}\right) \leq \left(\prod_{k=1}^{r(\boldsymbol{\Gamma}_{i,j})}\lambda_k(\boldsymbol{\Gamma}_{i,j})\right)^{-n_R}\left(\frac{\varrho}{4n_T}\right)^{-r(\boldsymbol{\Gamma}_{i,j})n_R}. \tag{3.8}$$

From the above analysis, we obtain two famous criteria for ST codeword construction, namely the *rank criterion* and the *determinant criterion*. To identify their meaning, we restate them in a compact form and point out their connection to the diversity gain:

1. Rank criterion: The rank criterion refers to the spatial diversity extracted by a ST code (and thus the diversity gain). In Equation (3.8), it can be seen that the ST code extracts a diversity of $r(\boldsymbol{\Gamma}_{i,j})n_R$ (the slope of the BER curve versus the SNR). Clearly, it follows that in order to extract the full spatial diversity of $n_R n_T$, the code should be designed such that the difference matrix $\boldsymbol{\Delta}_{i,j}$ between any pair $i \neq j$ of codeword matrices has full rank, and thus $r(\boldsymbol{\Gamma}_{i,j}) = n_T$.

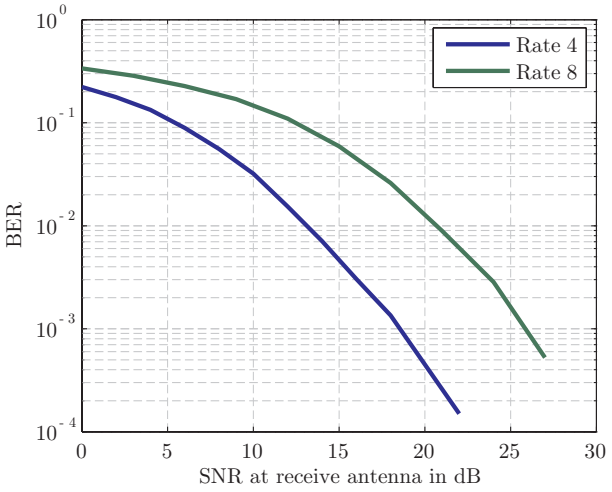

Figure 3.3.: BER curves of SM design over a $n_T = 2, n_R = 2$ channel with ML decoding for two different rates.

2. Determinant criterion: In contrast to the rank criterion, the determinant criterion optimizes the so-called *coding gain*. This can be visualized as a shift of the BER curve to the left and is a common terminus in the coding literature. If an error correcting code (e.g. turbo code) is used, the BER curve of the overall system will be shifted to the left and this is denoted as coding gain. From Equation (3.8) one sees that the term $\prod_{k=1}^{r(\mathbf{\Gamma}_{i,j})} \lambda_k(\mathbf{\Gamma}_{i,j})$ shifts the error probability (if plotted versus the SNR). This behavior can be seen as a coding gain, although no error correcting code is used. For a high coding gain in this context, one should maximize the minimum of the determinant of $\mathbf{\Gamma}_{i,j}$ over all possible pairs of codewords.

After having identified the rank criterion as the important criterion for achieving full diversity, we are now able to evaluate the diversity gain of our SM MIMO system. Equivalently to our used vector notation we can describe the system in terms of transmit block matrices $\mathbf{S}^{(i)}$, where all entries of $\mathbf{S}^{(i)}$ are independent. Then, the minimum distance pair of codeword matrices is given by a difference in only one entry. Clearly, the difference matrix $\mathbf{\Delta}_{i,j}$ for this codeword pair is a matrix containing only one non-zero element. This implies that the matrix $\mathbf{\Gamma}_{i,j}$ is rank one. Therefore, a SM design is only able to achieve a diversity gain of n_R instead of the full $n_T n_R$ diversity. Equivalently, the SM design only achieves a transmit diversity of 1. This observation motivates to search for ST coding schemes which are able to achieve full transmit diversity. Chapter 4 will discuss two important and well-known ST codes and analyzes them in terms of system capacity and diversity.

3.3.1. Numerical Simulations

To conclude this chapter, we will present a BER simulation of the SM design in a 2×2 MIMO channel for different rates, using a ML receiver. Figure 3.3 shows the obtained BER versus

SNR results for rate 4 (4-QAM) and 8 (16-QAM) respectively. We can observe that the slope of both curves achieve an order of two what equals the derived diversity gain of $n_R = 2$. Certainly the BER also depends on the size of the symbol alphabet, since in a larger symbol alphabet, the probability of confusing two codewords is larger - thus explaining the shift to the right of the rate 8 curve compared to the rate 4 curve.

4. Analysis of Space-Time Coded Systems

From our observations at the end of Chapter 3 it seems intuitive to spread our data symbols over time and space to achieve a higher transmit diversity (and accordingly a higher total diversity gain). The question of how to spread the data symbols in an efficient manner, such that transmit diversity is achieved, is the basis of ST coding. Of course, the rate should be kept as large as possible. Furthermore, simple receive processing can be seen as an additional design criterion.

In this chapter, we present a selection of well-known STC techniques. The range of results in this research area is certainly much broader than revealed in this thesis. Therefore, we refer the interested reader to [4, 33] and references therein. In particular, we will introduce the very universal class of linear space-time block codes (STBC). Within this class of STCs, we will investigate the special cases of orthogonal space-time block codes (OSTBC) and linear dispersion (LD) codes. An information theoretic analysis will be performed to identify the corresponding system capacities and analytical as well as numerical evaluations will be carried out to obtain measures of the associated error performance.

4.1. STBCs

Before we step into the detailed analysis of the mentioned STC techniques, we want to introduce a formal treatment of the already illustrated ST block transmission from Chapter 2.

Space-time block coding can be seen as a way of mapping a set of n_S complex data symbols $\{s_1, \ldots, s_{n_S}\} : s_n \in \mathcal{A}$, onto a matrix $\mathbf{S}$ of dimension $n_T \times L$ that is transmitted as described in Section 2.1 (see also [8]). In general, the mapping

$$\{s_1, \ldots, s_{n_S}\} \rightarrow \mathbf{S} \tag{4.1}$$

has no specific form (and can thus be nonlinear in general). For such a coding scheme, in which we transmit n_S complex symbols over L time intervals, we can define a transmission rate (in means of a symbol code rate) as:

Definition 4.1.1 (STBC transmission rate)**.** *Consider a STBC which transmits n_S complex data symbols during L symbol time instants. Then we define the STBC transmission rate as*

$$R_S \triangleq \frac{n_S}{L} \text{ [symbols/channel use]}.$$

A code is named "full-rate" if and only if $R_S = 1$.

Later throughout this chapter, we will need this definition for our information theoretic investigations. To narrow our focus throughout this thesis, we restrict ourselves to the case of *linear* STBC. These are relatively easy to treat and let us gain further insights in MIMO system design.

4.1.1. Linear Space-Time Block Codes

Linear STBCs are an important subclass of ST codes. In the case of linear STBCs, we choose the mapping (4.1) to be linear in the symbols $\{s_n\} \in \mathcal{A}$. Specifically, we define a linear STBC as:

Definition 4.1.2 (Mapping of a linear STBC)**.** *Let $\{\mathbf{A}_n, \mathbf{B}_n\}$, $n = 1, \ldots, n_S$ form a set of matrices of size $n_T \times L$ with domain $\mathbb{C}^{n_T \times L}$, respectively. Then the transmission matrix $\mathbf{S}$ of a linear STBC is formed according to*

$$\mathbf{S} = \sum_{n=1}^{n_S} \left(\mathrm{Re}\{s_n\}\, \mathbf{A}_n + j\ \mathrm{Im}\{s_n\}\, \mathbf{B}_n \right).$$

This definition was introduced in [16]. The set of matrices $\{\mathbf{A}_n\}$ and $\{\mathbf{B}_n\}$ can be interpreted as *modulation matrices*, because they "modulate" the real and imaginary part of the complex data symbols onto the transmitted matrix $\mathbf{S}$. Definition 4.1.2 could of course be written in other ways (compare [8]).

4.2. Orthogonal STBC

Orthogonal STBCs (OSTBCs) are an important subclass of linear STBCs. The underlying theory started with the famous work of Alamouti [34], which was extended to the general class of OSTBCs by Tarokh, et. al. (see [35]). An OSTBC is a linear STBC (as introduced in Definition 4.1.2) that has the following unitary property [8].

Definition 4.2.1 (OSTBC unitary property)**.** *Let the matrices $\mathbf{S}$ be MIMO block transmission matrices that are formed according to Definition 4.1.2. Then we define an OSTBC to be a ST code that fulfills*

$$\mathbf{S}\mathbf{S}^H = \sum_{n=1}^{n_S} |s_n|^2\, \mathbf{I}.$$

To explain the introduced relations, we stress a very popular example, the Alamouti code for the case of two transmit antennas as introduced in [34]. This code is given by

$$\mathbf{S} = \begin{bmatrix} s_1 & s_2^* \\ s_2 & -s_1^* \end{bmatrix}. \tag{4.2}$$

One easily verifies that the stated code matrix (or block transmission matrix) $\mathbf{S}$ fulfills the OSTBC unitary property from Definition 4.2.1. Furthermore, the Alamouti code can be

explained from the notation of linear STBC given in Definition 4.1.2. The modulation matrices $\{\mathbf{A}_n\}$ and $\{\mathbf{B}_n\}$ can be identified to be

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad \mathbf{A}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{B}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \mathbf{B}_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

The STBC transmission rate of the Alamouti code is $R_S = n_S/L = 2/2 = 1$, which means that with this ST design we are able to transmit one symbol at each time instant on average. The question how to construct orthogonal STBCs for a larger number of transmit antennas than two ($n_T > 2$) is rather difficult. One of the basic findings of Tarokh in [35] is that the construction of linear OSTBC is related to the theory of amicable orthogonal designs. This means that we relate our unitary property (Definition 4.2.1) to the set of modulation matrices.

Theorem 4.2.2 (OSTBC and amicable orthogonal designs)**.** *Let* $\mathbf{S}$ *be a matrix with structure given in Definition 4.2.1. Then*

$$\mathbf{S}\mathbf{S}^H = \sum_{n=1}^{n_S} |s_n|^2 \, \mathbf{I}_{n_T}$$

holds for all complex $\{s_n\}$ *if and only if* $\{\mathbf{A}_n, \mathbf{B}_n\}$ *is an amicable orthogonal design, satisfying*

$$\begin{aligned} &\mathbf{A}_n\mathbf{A}_n^H = \mathbf{I}_{n_T}, \quad \mathbf{B}_n\mathbf{B}_n^H = \mathbf{I}_{n_T}, \\ &\mathbf{A}_n\mathbf{A}_p^H = -\mathbf{A}_p\mathbf{A}_n^H, \quad \mathbf{B}_n\mathbf{B}_p^H = -\mathbf{B}_p\mathbf{B}_n^H, \quad \textit{for} \quad n \neq p, \\ &\mathbf{A}_n\mathbf{B}_p^H = \mathbf{B}_p\mathbf{A}_n^H, \end{aligned} \tag{4.3}$$

for $n, p = 1, \ldots, n_S$.

Proof. Is given in the Appendix (Subsection A.2.1). □

Although a very interesting field, the treatment of the theory of amicable orthogonal designs goes far beyond the scope of this thesis. For our analyses it is sufficient to restate one of the basic insights gained by the research performed in this field.

Theorem 4.2.3 (Nonexistence of full-rate OSTBC for $n_T > 2$)**.** *A full-rate* ($R_S = 1$) *OSTBC design for complex symbols exists only for* $n_T = 2$.

Proof. The proof is far from trivial. It is related to the Hurwitz-Radeon family of matrices and we do not state it here because it would give us no further insights. The interested reader is referred to [8] and references therein. □

4.2.1. Capacity Analysis of OSTBCs

After having introduced the important class of OSTBCs, we now analyze them in an information theoretic sense. The question we are asking is: Are OSTBC capable of achieving the ergodic channel capacity? Or equivalently: Equals the system capacity of OSTBCs the ergodic channel capacity?

Our derivations follow the arguments in [36], although we try to present a more detailed and contiguous analysis of the topic than performed in the mentioned paper. To find the system capacity of OSTBC MIMO systems, we first investigate how an OSTBC influences the *effective channel* (see also [8]). The effective channel $\mathbf{H}_{\text{eff}}$ is constructed in a way that we are able to rearrange the MIMO block transmission model from Equation (2.3) in a form that the data symbols appear in an unmodulated form in the relation, i.e. $\mathbf{y}' = \sqrt{\varrho/n_T}\mathbf{H}_{\text{eff}}\mathbf{s}' + \mathbf{n}'$. Naturally, the effective channel will reflect the modulation structure of the linear STBC (i.e. the choice of modulation matrices $\{\mathbf{A}_n\}$ and $\{\mathbf{B}_n\}$). In case of OSTBC MIMO systems, we already showed that the modulation matrices have to satisfy the constraints in Equation (4.3). Let us stay as general as possible and consider $\mathbf{S}$ only to be constructed as an arbitrary linear STBC. We write

$$\begin{aligned}\mathbf{y}' &= \text{vec}(\mathbf{Y}) \\ &= \text{vec}\left(\sqrt{\frac{\varrho}{n_T}}\mathbf{H}\mathbf{S} + \mathbf{N}\right) = \sqrt{\frac{\varrho}{n_T}}\text{vec}\left(\mathbf{H}\sum_{n=1}^{n_S}\left(\,\text{Re}\{s_n\}\,\mathbf{A}_n + j\;\text{Im}\{s_n\}\,\mathbf{B}_n\right)\right) + \text{vec}(\mathbf{N}) \\ &= \sqrt{\frac{\varrho}{n_T}}\left(\mathbf{H}_{\text{eff},a}\,\text{Re}\{\mathbf{s}\} + \mathbf{H}_{\text{eff},b}\,\text{Im}\{\mathbf{s}\}\right) + \text{vec}(\mathbf{N})\,,\end{aligned}$$

where we used the fact that $\text{vec}(\mathbf{A}+\mathbf{B}) = \text{vec}(\mathbf{A})+\text{vec}(\mathbf{B})$ and we defined

$$\mathbf{H}_{\text{eff},a} \triangleq \left[\text{vec}(\mathbf{H}\mathbf{A}_1)\,,\cdots,\text{vec}(\mathbf{H}\mathbf{A}_{n_S})\right],\quad \mathbf{H}_{\text{eff},b} \triangleq j\cdot\left[\text{vec}(\mathbf{H}\mathbf{B}_1)\,,\cdots,\text{vec}(\mathbf{H}\mathbf{B}_{n_S})\right]$$
$$\text{and } \mathbf{s} \triangleq [s_1,\ldots,s_{n_S}]^T.$$

The vectorized MIMO transmission model in the above relation can further be simplified to

$$\mathbf{y}' = \sqrt{\frac{\varrho}{n_T}}\mathbf{H}_{\text{eff}}\mathbf{s}' + \mathbf{n}',$$

where we defined

$$\mathbf{H}_{\text{eff}} \triangleq \left[\mathbf{H}_{\text{eff},a}, \mathbf{H}_{\text{eff},b}\right],\quad \mathbf{s}' \triangleq \begin{bmatrix}\text{Re}\{\mathbf{s}\}\\ \text{Im}\{\mathbf{s}\}\end{bmatrix}\quad\text{and}\quad \mathbf{n}' \triangleq \text{vec}(\mathbf{N})\,.$$

Having identified a possible representation of the MIMO block transmission model in case of a linear STBC by means of an effective channel, we are able to look at the consequences regarding the effective channel in case of an OSTBC. The result is stated in the following theorem.

Theorem 4.2.4 (Effective channel decoupling property of OSTBC)**.** *A linear code as in Definition 4.1.2 is an OSTBC if and only if the effective channel* $\mathbf{H}_{\mathit{eff}}$ *satisfies*

$$\mathbf{H}_{\mathit{eff}}^H\mathbf{H}_{\mathit{eff}} = \|\mathbf{H}\|^2\cdot\mathbf{I},$$

for all channels $\mathbf{H}$*.*

Proof. The proof is equivalent to the one in [8]. We just adapted it to our notation. We start by evaluating the Frobenius norm (see Subsection A.3.1)

$$\|\mathbf{H}_{\text{eff}}\mathbf{s}'\|^2 = \text{tr}\left\{\left(\mathbf{H}_{\text{eff}}\mathbf{s}'\right)^H\mathbf{H}_{\text{eff}}\mathbf{s}'\right\} = \text{tr}\left\{\mathbf{s}'^T\mathbf{H}_{\text{eff}}^H\mathbf{H}_{\text{eff}}\mathbf{s}'\right\} = \mathbf{s}'^T\mathbf{H}_{\text{eff}}^H\mathbf{H}_{\text{eff}}\mathbf{s}',$$

where we used the fact that since $\mathbf{H}_{\text{eff}}\mathbf{s}' = \text{vec}(\mathbf{HS})$ by construction, the trace operation can be dropped. Furthermore it follows that $\|\mathbf{H}_{\text{eff}}\mathbf{s}'\|^2 = \|\text{vec}(\mathbf{HS})\|^2 = \|\mathbf{HS}\|^2$. This can be easily shown by using Definition A.3.1. The next step is to include our knowledge about the structure of OSTBC (see Definition 4.2.1). Again using the definition of the Frobenius norm, we obtain

$$\|\mathbf{HS}\|^2 = \text{tr}\left\{\mathbf{HS}\left(\mathbf{HS}\right)^H\right\} = \text{tr}\left\{\mathbf{HSS}^H\mathbf{H}^H\right\},$$

and by usage of Definition 4.2.1 the above equation simplifies to

$$\|\mathbf{HS}\|^2 = \text{tr}\left\{\mathbf{H}\left(\sum_{i=1}^{n_S}|s_n|^2\,\mathbf{I}\right)\mathbf{H}^H\right\} = \sum_{i=1}^{n_S}|s_n|^2\,\text{tr}\left\{\mathbf{HH}^H\right\} = \|\mathbf{s}\|^2\,\|\mathbf{H}\|^2.$$

Since due to our premises, one easily verifies that $\|\mathbf{s}\|^2 = \|\mathbf{s}'\|^2$ and that $\|\mathbf{s}'\|^2 = \mathbf{s}'^T\mathbf{s}'$. With these insights we are able derive the following relation

$$\|\mathbf{H}_{\text{eff}}\mathbf{s}'\|^2 = \mathbf{s}'^T\mathbf{H}_{\text{eff}}^H\mathbf{H}_{\text{eff}}\mathbf{s}' = \|\mathbf{HS}\|^2 = \|\mathbf{s}'\|^2\,\|\mathbf{H}\|^2 = \mathbf{s}'^T\,\|\mathbf{H}\|^2\,\mathbf{s}'.$$

This implies that $\mathbf{H}_{\text{eff}}^H\mathbf{H}_{\text{eff}}$ must be equal to $\|\mathbf{H}\|^2\,\mathbf{I}$, which concludes the proof. □

Theorem 4.2.4 shows that by using an orthogonal STBC the effective channel will be orthogonal, irrespective of the channel realization. Thus, if the receiver knows the channel realization $\mathbf{H}$, it can form the effective channel and consequently use the effective transmission model $\mathbf{y}' = \sqrt{\varrho/n_T}\mathbf{H}_{\text{eff}}\mathbf{s}' + \mathbf{n}'$. Because of the orthogonality of the effective channel, we will see in Subsection 4.2.2 that the ML receiver decouples into n_S independent scalar decisions for which we have to form a data vector estimate $\hat{\mathbf{s}'} = \|\mathbf{H}\|^{-2}\,\text{Re}\{\mathbf{H}_{\text{eff}}^H\mathbf{y}'\}$. The multiplication of the received vector $\mathbf{y}'$ with $\mathbf{H}_{\text{eff}}^H$ can be seen as maximum ratio combining. Taking this first step of the receiver, we obtain

$$\mathbf{H}_{\text{eff}}^H\mathbf{y}' = \sqrt{\frac{\varrho}{n_T}}\mathbf{H}_{\text{eff}}^H\mathbf{H}_{\text{eff}}\mathbf{s}' + \mathbf{H}_{\text{eff}}^H\mathbf{n}' = \sqrt{\frac{\varrho}{n_T}}\,\|\mathbf{H}\|^2\,\mathbf{s}' + \mathbf{H}_{\text{eff}}^H\mathbf{n}', \tag{4.4}$$

where we used Theorem 4.2.4. Due to the fact that we are using an OSTBC, one can show that the elements of the noise after MRC (i.e. $\mathbf{H}_{\text{eff}}^H\mathbf{n}'$) are iid with variance $\|\mathbf{H}\|^2$ (compare [36] and [37]). By using the vectorized channel model, the receiver effectively uses $2n_S$ virtual antennas (since the vector $\mathbf{s}'$ has length $2n_S$). The effective transmit power after MRC can be obtained by calculating the covariance matrix of the MRC signal output $\sqrt{\frac{\varrho}{n_T}}\,\|\mathbf{H}\|^2\,\mathbf{s}'$

$$\mathbb{E}\left\{\frac{\varrho}{n_T}\,\|\mathbf{H}\|^4\,\mathbf{s}'\mathbf{s}'^T\right\} = \frac{\varrho}{n_T}\,\|\mathbf{H}\|^4\,\mathbb{E}\{\mathbf{s}'\mathbf{s}'^T\} = \frac{\varrho}{n_T}\,\|\mathbf{H}\|^4\,\frac{1}{2}\mathbf{I},$$

where we assumed that real and imaginary parts of the data symbols are uncorrelated and have an equal variance of 1/2. Thus, if the receiver combines two virtual antennas (the real and imaginary part of each data symbol contained in $\mathbf{s}$) respectively, we obtain the effective SNR at the receiver by

$$\frac{2\cdot\varrho/n_T\,\|\mathbf{H}\|^4\,(1/2)}{\|\mathbf{H}\|^2} = \frac{\varrho}{n_T}\,\|\mathbf{H}\|^2.$$

The main observation here is that (4.4) is effectively a set of n_S parallel independent Rayleigh fading channels with a SNR of $\frac{\varrho}{n_T}\|\mathbf{H}\|^2$ in each channel respectively. Since in each effective channel we are transmitting one data symbol for the duration of L time instances, the MIMO transmission rate is equal to $R_S = n_S/L$. Knowing these coherences, we can state our result on the derivation of the OSTBC MIMO system capacity (compare [36, 38]).

Theorem 4.2.5 (OSTBC MIMO system capacity)**.** *The MIMO system capacity of an arbitrary OSTBC design is given by*

$$C_{OSTBC} = \frac{n_S}{L}\log_2\left(1 + \frac{\varrho}{n_T}\|\mathbf{H}\|^2\right),$$

where the term n_S/L denotes the MIMO transmission rate R_S.

Proof. The proof follows the derivation of the n_S effective channels and their associated SNRs of $\frac{\varrho}{n_T}\|\mathbf{H}\|^2$. Since the capacity (in bits) of an AWGN channel is given by $C = \log_2(1 + \mathrm{SNR})$ (see [17]) and considering that the capacity of n_S parallel independent channels is n_S times the capacity of one channel, it follows that the capacity in bits per channel use (thus motivating the division by L) is $C = n_S/L\log_2(1 + \mathrm{SNR})$. This concludes the proof. □

After we have obtained the desired result, we can ask, whether an OSTBC is able to achieve the ergodic channel capacity. The answer to this question is given in the following theorem.

Theorem 4.2.6 (Capacity order of orthogonal STBC)**.** *Let $\mathbf{H}$ be an iid channel matrix according to our MIMO transmission model introduced in Section 2.1. Then*

$$C(\mathbf{H}) \geq C_{OSTBC}(\mathbf{H}).$$

The given inequality also holds in the ergodic case, i.e. $C_E = \mathbb{E}\{C(\mathbf{H})\} \geq C_{E,OSTBC} = \mathbb{E}\{C_{OSTBC}(\mathbf{H})\}$.

Proof. First we reformulate the expression for the channel capacity of a given channel realization. Let the singular value decomposition (SVD) of $\mathbf{H}$ be $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ (see Section A.3.2). Then the capacity can be written as

$$C(\mathbf{H}) = \log_2\det\left(\mathbf{I}_{n_R} + \frac{\varrho}{n_T}\mathbf{H}\mathbf{H}^H\right) = \log_2\det\left(\mathbf{I}_{n_R} + \frac{\varrho}{n_T}\mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^H\mathbf{U}^H\right).$$

Since $\mathbf{U}$ is an unitary matrix, thus obeying $\mathbf{U}^H\mathbf{U} = \mathbf{I}$, we do not change the above relation if we multiply it with $\det(\mathbf{U}^H\mathbf{U}) = \det\mathbf{U}^H\det\mathbf{U} = 1$. Thus it follows that

$$C(\mathbf{H}) = \log_2\det\mathbf{U}^H\det\left(\mathbf{I}_{n_R} + \frac{\varrho}{n_T}\mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^H\mathbf{U}^H\right)\det\mathbf{U} = \log_2\det\left(\mathbf{I}_{n_R} + \frac{\varrho}{n_T}\mathbf{\Sigma}\mathbf{\Sigma}^H\right).$$

By taking into account that $\mathbf{\Sigma}$ is diagonal, containing the singular values of $\mathbf{H}$, we obtain

$$C(\mathbf{H}) = \log_2\prod_{i=1}^{r}\left(1 + \frac{\varrho}{n_T}\lambda_i^2\right),$$

where r denotes the rank (and thus the number of non-zero singular values). Continuing our evaluation, the product can be expanded to

$$C(\mathbf{H}) = \log_2 \left(1 + \frac{\varrho}{n_T} \sum_{i=1}^{r} \lambda_i^2 + \frac{\varrho^2}{n_T^2} \sum_{i_1 \neq i_2}^{i_1 < i_2} \lambda_{i_1}^2 \lambda_{i_2}^2 + \frac{\varrho^3}{n_T^3} \sum_{i_1 \neq i_2 \neq i_3}^{i_1 < i_2 < i_3} \lambda_{i_1}^2 \lambda_{i_2}^2 \lambda_{i_3}^2 + \cdots + \frac{\varrho^r}{n_T^r} \prod_{i=1}^{r} \lambda_i^2 \right).$$

By considering the definition of the Frobenius norm (see Definition A.3.1), which states that $\|\mathbf{H}\|^2 = \sum_{i=1}^{r} \lambda_i^2$ and defining $\det\left(\mathbf{H}\mathbf{H}^H\right)_r$ to be the product of non-zero squared singular values, we finally obtain

$$C(\mathbf{H}) = \log_2 \left(1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2 + \cdots + \frac{\varrho^r}{n_T^r} \det\left(\mathbf{H}\mathbf{H}^H\right)_r \right). \tag{4.5}$$

Next, by using Equation (4.5) it follows that

$$\begin{aligned} C(\mathbf{H}) &= \log_2 \left(1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2 + \cdots + \frac{\varrho^r}{n_T^r} \det\left(\mathbf{H}\mathbf{H}^H\right)_r \right) \\ &\geq \log_2 \left(1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2 \right) \geq \frac{n_S}{L} \log_2 \left(1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2 \right) = C_{\mathrm{OSTBC}}(\mathbf{H}). \end{aligned}$$

The expansion to the ergodic case follows directly by considering C to be ≥ 0. This concludes the proof. □

We have seen so far that the instantaneous system capacity of OSTBC is in general suboptimal in terms of the channel capacity, and so is the ergodic OSTBC system capacity. Nevertheless our initial question is not fully answered, since we have not investigated under which conditions we can reach the equality in Theorem 4.2.6. Let us compute the capacity difference $\Delta C(\mathbf{H}) = C(\mathbf{H}) - C_{\mathrm{OSTBC}}(\mathbf{H})$ (compare [36]). By using the previous result, we can write

$$\Delta C(\mathbf{H}) = \log_2 \left(1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2 + S \right) - \frac{n_S}{L} \log_2 \left(1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2 \right),$$

where we set S to be $S \triangleq \left(\frac{\varrho}{n_T}\right)^2 \sum_{i_1 \neq i_2}^{i_1 < i_2} \lambda_{i_1}^2 \lambda_{i_2}^2 + \cdots \left(\frac{\varrho}{n_T}\right)^r \prod_{i=1}^{r} \lambda_i^2$. Straightforward manipulation leads to

$$\begin{aligned} \Delta C(\mathbf{H}) &= \log_2 \left(\frac{1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2 + S}{\left(1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2\right)^{n_S/L}} \right) \\ &= \log_2 \left[\left(1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2 \right)^{1 - n_S/L} \left(1 + \frac{S}{1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2} \right) \right], \end{aligned}$$

which can be further simplified to

$$\Delta C(\mathbf{H}) = \frac{L - n_S}{L} \log_2 \left(1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2 \right) + \log_2 \left(1 + \frac{S}{1 + \frac{\varrho}{n_T} \|\mathbf{H}\|^2} \right). \tag{4.6}$$

This result shows us that the difference is a function of the channel realization and thus is a random variable. Since this capacity difference is a function of the channel singular values, it can be used to answer the question when the OSTBC system capacity coincides with the channel capacity. The conclusion on this is summarized in the following theorem (see also [36]).

Theorem 4.2.7 (Capacity optimality of OSTBC)**.** *An OSTBC is optimal with respect to channel capacity when it is rate one and it is used over a channel of rank one.*

Proof. Assume that the channel is nontrivial and bounded, i.e., $0 < \|\mathbf{H}\|^2 < \infty$. Consider the capacity difference in (4.6). By inspection, the first logarithm term is zero if the code is rate one (i.e. $n_S = L$). The second logarithm term is zero if and only if $S = 0$. Since for $\|\mathbf{H}\|^2 > 0$ all quantities are positive, $S = 0$ implies that each constituent sum-of-product term $\sum_{i_1 \neq i_2 \cdots \neq i_k}^{i_1 < i_2 \cdots < i_k} \lambda_{i_1}^2 \lambda_{i_2}^2 \cdots \lambda_{i_k}^2$ $(2 \neq k \neq r)$ in the expression for S is zero. This follows directly by the following. Assume that all singular values except λ_1 are zero, thus describing a channel of rank one. By definition of the Frobenius norm, $\lambda_1^2 = \|\mathbf{H}\|^2$. Then each sum-of-products term $\sum_{i_1 \neq i_2 \cdots \neq i_k}^{i_1 < i_2 \cdots < i_k} \lambda_{i_1}^2 \lambda_{i_2}^2 \cdots \lambda_{i_k}^2 = 0$ because each product $\lambda_{i_1}^2 \lambda_{i_2}^2 \cdots \lambda_{i_k}^2$ consists of two or more singular values, at least one of which is zero. Therefore S is zero. If $\mathbf{H}$ is of rank 2, then $S = \lambda_1^2 \lambda_2^2 + 0$, which is non-zero. It follows, that $S = 0$ if and only if $\mathbf{H}$ is of rank 1. This concludes the proof. □

Despite the pleasing property that OSTBCs decouple the space-time channel into parallel independent AWGN channels, we showed that the structure imposed by orthogonal STBCs generally limits the maximal error free output (this is the obtained OSTBC system capacity) that can be achieved, regardless of the amount of outer channel coding that is employed. Theorem 4.2.7 describes the cases in which we could achieve channel capacity by an OSTBC system. Unfortunately the restriction of Theorem 4.2.3 (nonexistence of full-rate OSTBC for $n_T > 2$), reduces the cases in which OSTBC are optimal. The consequence of the combination of both theorems is that OSTBC can only achieve channel capacity in case of two transmit antennas and a rank 1 channel matrix. To visualize the obtained insights, some numerical simulations of the OSTBC system capacity are presented below.

Numerical Simulations

As mentioned before, we did some numerical simulations to visualize the difference between the OSTBC system capacity and the ergodic channel capacity. As an OSTBC we used the Alamouti design and $\mathbf{H}$ was assumed to be iid Gaussian. The curves represent results for the ergodic channel capacity and the OSTBC system capacity in the case of $n_T = 2$ transmit antennas and $n_R = 1, 2$ receive antennas. For these cases we used the Alamouti design given in Equation (4.2). The results of our simulations are plotted in Figure 4.1. It can be seen that the OSTBC design is optimal in terms of ergodic channel capacity in the case of $n_R = 1$. This corresponds to the unique case of $n_T = 2$, $R_S = 1$ and rank($\mathbf{H}$) equals to one, in which OSTBC is capacity optimal. This is because in the case of one receive antenna, the channel $\mathbf{H}$ is of rank one. In the case of $n_R = 2$ receive antennas, the picture changes dramatically. The curves of OSTBC system capacity and ergodic channel capacity do not coincide anymore, and even more, the OSTBC system capacity has a much lower slope. This reflects the fact that

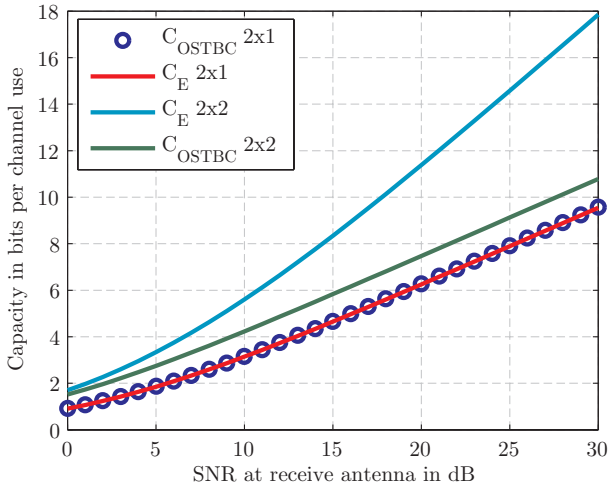

Figure 4.1.: Comparison of OSTBC system capacity with ergodic channel capacity for different antenna constellations.

OSTBCs have a smaller multiplexing gain. This also implies that OSTBC are not well suited for transmissions at high rates.

Finally, we computed the mutual information in case of finite symbol alphabets and OSTBC Alamouti coding. Therefore we used the equivalent transmission model from Subsection A.2.3. The results are plotted in Figure 4.2. In principle, we can identify the same properties as in the SM case (compare Figure 3.1). We note that the mutual information curves again huddle against the OSTBC system capacity curve at low SNR.

4.2.2. Error Performance of OSTBCs

After we had a close look on the system capacity of OSTBCs, we want to investigate how good or bad an OSTBC design behaves in terms of error performance (and thus diversity). Again, we rely on the usage of an optimal ML receiver. Besides the analysis of diversity, we will derive an alternative expression of the ML receiver, which shows clearly that in the case of OSTBC designs the ML decoding is very simple and efficient. We will conclude our analysis of OSTBC designs by providing some numerical BER simulations. First, we want to investigate the diversity gain of OSTBC designs.

Theorem 4.2.8 (Diversity gain of OSTBC systems)**.** *OSTBC systems achieve full diversity (i.e. a diversity gain of $n_T n_R$).*

Proof. Let us use the notation of Section 3.3. In [6] it is shown that the difference between two codewords $\mathbf{S}^{(i)}$ and $\mathbf{S}^{(j)}$ for $s_n \in \mathcal{A}$ is an orthogonal matrix $\mathbf{\Delta}_{i,j}$ in the case of an OSTBC design. Thus $r(\mathbf{\Gamma}_{i,j}) = n_T$, and accordingly a diversity gain of $n_T n_R$ is achieved. Further details may be found in [6]. □

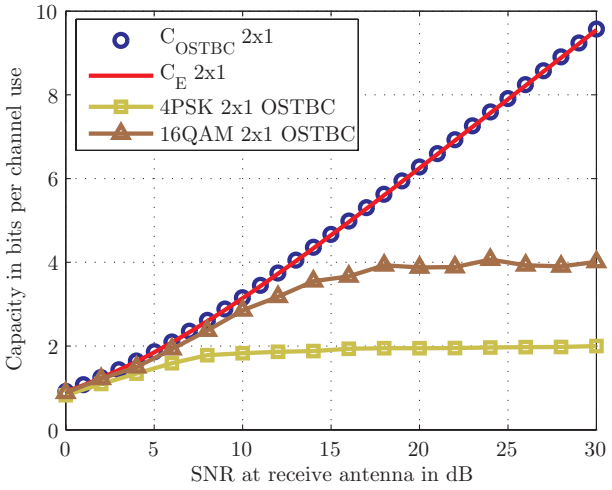

Figure 4.2.: Mutual information curves for OSTBC Alamouti coding and different sizes of symbol alphabets in the case of a $n_T = 2, n_R = 1$ channel.

ML Decision Decoupling

After having evaluated the diversity gain of OSTBC designs, we will now concentrate on the ML receiver. For OSTBCs, we show that the ML detector decouples into n_S scalar decisions, thus significantly reducing the computational complexity.

Reconsidering that the MIMO block transmission model $\mathbf{Y} = \sqrt{\varrho/n_T}\mathbf{HS} + \mathbf{N}$ (see Equation (2.3)) can equivalently be written as $\mathbf{y}' = \sqrt{\varrho/n_T}\mathbf{H}_{\text{eff}}\mathbf{s}' + \mathbf{n}'$, one can easily show that $\left\|\mathbf{Y} - \sqrt{\varrho/n_T}\mathbf{HS}\right\|^2$ equals to $\left\|\mathbf{y}' - \sqrt{\varrho/n_T}\mathbf{H}_{\text{eff}}\mathbf{s}'\right\|^2$. Accordingly, the ML receiver can choose which of these two metrics to minimize. Furthermore, in the case of OSTBCs the metric decouples (compare [8]). Following the notation in Subsection 2.4.1, the ML decision rule can thus be written as

$$\hat{\mathbf{s}}_{\text{ML}} = \arg\min_{\mathbf{s}'\in\mathcal{S}} \left\|\mathbf{y}' - \sqrt{\frac{\varrho}{n_T}}\mathbf{H}_{\text{eff}}\mathbf{s}'\right\|^2 = \arg\min_{\mathbf{s}'\in\mathcal{S}} \|\mathbf{H}\|^2 \left\|\tilde{\mathbf{s}}' - \sqrt{\frac{\varrho}{n_T}}\mathbf{s}'\right\|^2, \tag{4.7}$$

where $\tilde{\mathbf{s}}' = 1/\|\mathbf{H}\|^2 \operatorname{Re}\{\mathbf{H}_{\text{eff}}^H\mathbf{y}\}$ is a scaled version of the MRC receive vector. This implies that the ML detection is equivalent to solve n_S scalar detection problems, one for each symbol s_n.

To validate the above simplified ML decision rule, let us write the ML metric mentioned above as

$$\left\|\mathbf{y} - \sqrt{\frac{\varrho}{n_T}}\mathbf{H}_{\text{eff}}\mathbf{s}'\right\|^2 = \|\mathbf{y}\|^2 + \frac{\varrho}{n_T}\|\mathbf{H}\|^2\|\mathbf{s}'\|^2 - 2\sqrt{\frac{\varrho}{n_T}}\operatorname{Re}\{\mathbf{y}^H\mathbf{H}_{\text{eff}}\mathbf{s}'\}, \tag{4.8}$$

where we used the $\|\mathbf{H}_{\text{eff}}\mathbf{s}'\|^2 = \|\mathbf{H}\|^2\|\mathbf{s}'\|^2$ from the Proof of Theorem 4.2.4. Now, let us single out the term $\|\mathbf{H}\|^2$ and let us disregard the term $\|\mathbf{y}\|^2$, since it does not depend on $\mathbf{s}'$. Then,

we obtain

$$\arg\min_{\mathbf{s}'\in\mathcal{S}} \left\| \mathbf{y} - \sqrt{\frac{\varrho}{n_T}}\mathbf{H}_{\text{eff}}\mathbf{s}' \right\|^2 = \arg\min_{\mathbf{s}'\in\mathcal{S}} \|\mathbf{H}\|^2 \left(\frac{\varrho}{n_T} \|\mathbf{s}'\|^2 - 2\sqrt{\frac{\varrho}{n_T}} \frac{1}{\|\mathbf{H}\|^2} \operatorname{Re}\{\mathbf{y}^H \mathbf{H}_{\text{eff}}\} \mathbf{s}' \right).$$

If we identify $\tilde{\mathbf{s}}'$ as $1/\|\mathbf{H}\|^2 \operatorname{Re}\{\mathbf{H}_{\text{eff}}^H \mathbf{y}\}$, one easily verifies that we can write[1]

$$\arg\min_{\mathbf{s}'\in\mathcal{S}} \left\| \mathbf{y} - \sqrt{\frac{\varrho}{n_T}}\mathbf{H}_{\text{eff}}\mathbf{s}' \right\|^2 = \arg\min_{\mathbf{s}'\in\mathcal{S}} \arg\min_{\mathbf{s}'\in\mathcal{S}} \|\mathbf{H}\|^2 \left\| \tilde{\mathbf{s}}' - \sqrt{\frac{\varrho}{n_T}}\mathbf{s}' \right\|^2.$$

Equation 4.7 shows a very important property of OSTBC designs. Here, the ML receiver can be implemented in a very efficient way. For the sake of completeness, we note that a detailed evaluation of the ML metric $\left\| \mathbf{Y} - \sqrt{\varrho/n_T}\mathbf{H}\mathbf{S} \right\|^2$ and its decoupling in the case of OSTBC is given in the Appendix, Subsection A.2.2.

Numerical Simulations

After we have derived a basic understanding of the error performance behavior of OSTBC, we want to show some simulations we performed. The simulations given in Subsection 4.2.1 represent the performance of OSTBCs in terms of system capacity, whereas we now present some BER versus SNR results. We performed simulations for the Alamouti scheme from Equation (4.2) and for the SM design. The results are plotted in Figure 4.3. Since the Alamouti code is full-rate in this case ($n_T = 2$), we chose $\mathcal{A}$ to be a 16-QAM for the Alamouti code and a 4-QAM for the SM design. Thus, both schemes have a rate of 4 bits/channel use.

It can be seen that in the low SNR regime both systems have the same performance. However, in the high SNR regime, the Alamouti scheme performs much better than the SM design, which is due to the difference in the diversity gain.

4.3. Linear Dispersion Codes

The OSTBCs we investigated in the previous section gave us a first impression about how a MIMO system can improve the error performance. But, we also observed that its system capacity may be inferior to the corresponding ergodic channel capacity.

Hassibi and Hochwald showed in [16], among the introduction of the general class of linear STBC, that it is possible to construct a linear STBCs that achieves the ergodic channel capacity. In this Section, we want to have a closer look at these *LD codes*[2] and investigate how they perform in terms of system capacity and diversity.

[1] This can be seen by adding a constant term $\|\operatorname{Re}\{\mathbf{H}_{\text{eff}}\mathbf{y}\}\|^2 / \|\mathbf{H}\|^2$ (not depending on $\mathbf{s}'$).

[2] We refer to this MIMO code design as "LD codes", although the term originally refers to the whole class of linear STBCs. Nevertheless, the proposed techniques in [16] leads to a specific code structure, which is commonly referred to as LD-codes. This justifies our notation.

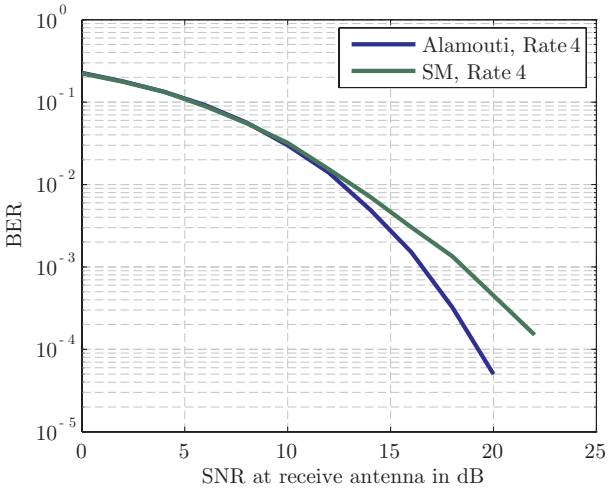

Figure 4.3.: BER comparison of the orthogonal Alamouti design and a SM design in case of a $n_R = 2, n_T = 2$ channel and a rate of 4.

4.3.1. Definition and Capacity Analysis

A LD code is given by the linear mapping of Definition 4.1.2, i.e.

$$\mathbf{S} = \sum_{n=1}^{n_S} \left(\mathrm{Re}\{s_n\}\, \mathbf{A}_n + j\ \mathrm{Im}\{s_n\}\, \mathbf{B}_n \right).$$

Thus, the associated rate of the LD code is given by $R = R_S \log_2 M_a = n_S/L \log_2 M_a$, where M_a denotes the size of the symbol alphabet. The design of the LD code depends crucially on the choices of the parameters L, n_S and the modulation matrices $\{\mathbf{A}_n, \mathbf{B}_n\}$. If we constrain the STBC block matrix to fulfill Definition 4.2.1, the mapping results in an orthogonal structure, as we investigated in Section 4.2. Nevertheless, this is only one possible way of choosing the modulation matrices. The question is, whether we can choose them in a way that the modulation matrices (or also sometimes called dispersion matrices) transmit some combination of each symbol from each antenna at every channel use, and therefore leading to desirable gains in terms of system capacity.

For the construction of the LD codes, Hassibi and Hochwald propose to choose the modulation matrices $\{\mathbf{A}_n, \mathbf{B}_n\}$ in a way to optimize a nonlinear information-theoretic criterion: the mutual information between the transmitted signal and the received signal. This criterion is very important for achieving high spectral efficiency with multiple antennas. The maximization of the mutual information is a problem which has to be done once for a given antenna constellation and desired rate once. To be able to optimize the modulation matrices, we have to derive an equivalent representation of the MIMO block transmission model

$$\mathbf{Y} = \sqrt{\frac{\varrho}{n_T}}\mathbf{H}\mathbf{S} + \mathbf{N}$$

in a similar way to Subsection 4.2.1. Because we want to obtain an expression that can be optimized in an efficient manner by means of numerical tools, we search for a real representation of the effective MIMO transmission model in that subsection. Therefore, let us transpose the MIMO block transmission model (this will result in a favorable expression of the resulting effective relation) and decompose the matrices into their real and imaginary parts. Doing so, we first obtain the transposed system equation

$$\mathbf{Y}^T = \sqrt{\frac{\varrho}{n_T}}\mathbf{S}^T\mathbf{H}^T + \mathbf{N}^T,$$

and dropping the $(\cdot)^T$ by simply redefining the affected matrices (they now have transposed dimensions) and performing the mentioned decomposition, we get

$$\mathbf{Y}_R + j\mathbf{Y}_I = \sqrt{\frac{\varrho}{n_T}} \sum_{n=1}^{n_S} \left[s_{R,n} \left(\mathbf{A}_{R,n} + j\mathbf{A}_{I,n} \right) + j s_{I,n} \left(\mathbf{B}_{R,n} + j\mathbf{B}_{I,n} \right) \right] \left(\mathbf{H}_R + j\mathbf{H}_I \right) + \mathbf{N}_R + j\mathbf{N}_I,$$

where we denoted the real ($\mathrm{Re}\{\cdot\}$) and imaginary ($\mathrm{Im}\{\cdot\}$) part of the matrices by $(\cdot)_R$ and $(\cdot)_I$ respectively.

Now let us denote the columns of $\mathbf{Y}_R$, $\mathbf{Y}_I$, $\mathbf{H}_R$, $\mathbf{H}_I$, $\mathbf{N}_R$ and $\mathbf{N}_I$ by $\mathbf{y}_{R,m}$, $\mathbf{y}_{I,m}$, $\mathbf{h}_{R,m}$, $\mathbf{h}_{I,m}$, $\mathbf{n}_{R,m}$ and $\mathbf{n}_{I,m}$ respectively, where $m = 1, \ldots, n_R$. With

$$\boldsymbol{\mathcal{A}}_n \triangleq \begin{bmatrix} \mathbf{A}_{R,n} & -\mathbf{A}_{I,n} \\ \mathbf{A}_{I,n} & \mathbf{A}_{R,n} \end{bmatrix}, \quad \boldsymbol{\mathcal{B}}_n \triangleq \begin{bmatrix} -\mathbf{B}_{I,n} & -\mathbf{B}_{R,n} \\ \mathbf{B}_{R,n} & -\mathbf{B}_{I,n} \end{bmatrix}, \quad \mathbf{h}_m \triangleq \begin{bmatrix} \mathbf{h}_{R,m} \\ \mathbf{h}_{I,m} \end{bmatrix},$$

we can form the system of real equations:

$$\begin{bmatrix} \mathbf{y}_{R,1} \\ \mathbf{y}_{I,1} \\ \vdots \\ \mathbf{y}_{R,n_R} \\ \mathbf{y}_{I,n_R} \end{bmatrix} = \sqrt{\frac{\varrho}{n_T}} \begin{bmatrix} \boldsymbol{\mathcal{A}}_1\mathbf{h}_1 & \boldsymbol{\mathcal{B}}_1\mathbf{h}_1 & \cdots & \boldsymbol{\mathcal{A}}_{n_S}\mathbf{h}_1 & \boldsymbol{\mathcal{B}}_{n_S}\mathbf{h}_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\mathcal{A}}_1\mathbf{h}_{n_R} & \boldsymbol{\mathcal{B}}_1\mathbf{h}_{n_R} & \cdots & \boldsymbol{\mathcal{A}}_{n_S}\mathbf{h}_{n_R} & \boldsymbol{\mathcal{B}}_{n_S}\mathbf{h}_{n_R}, \end{bmatrix} \begin{bmatrix} s_{R,1} \\ s_{I,1} \\ \vdots \\ s_{R,n_S} \\ s_{I,n_S} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{R,1} \\ \mathbf{n}_{I,1} \\ \vdots \\ \mathbf{n}_{R,n_R} \\ \mathbf{n}_{I,n_R} \end{bmatrix}.$$

Accordingly, the input-output relation of the MIMO channel using a linear STBC can be represented by

$$\mathbf{y}_{\mathrm{LD}} = \sqrt{\frac{\varrho}{n_T}}\mathbf{H}_{\mathrm{LD}}\mathbf{s}_{\mathrm{LD}} + \mathbf{n}_{\mathrm{LD}}, \tag{4.9}$$

where

$$\mathbf{y}_{\mathrm{LD}} \triangleq \begin{bmatrix} \mathbf{y}_{R,1} \\ \mathbf{y}_{I,1} \\ \vdots \\ \mathbf{y}_{R,n_R} \\ \mathbf{y}_{I,n_R} \end{bmatrix}, \quad \mathbf{s}_{\mathrm{LD}} \triangleq \begin{bmatrix} s_{R,1} \\ s_{I,1} \\ \vdots \\ s_{R,n_S} \\ s_{I,n_S} \end{bmatrix}, \quad \mathbf{n}_{\mathrm{LD}} \triangleq \begin{bmatrix} \mathbf{n}_{R,1} \\ \mathbf{n}_{I,1} \\ \vdots \\ \mathbf{n}_{R,n_R} \\ \mathbf{n}_{I,n_R} \end{bmatrix},$$

and

$$\mathbf{H}_{\mathrm{LD}} \triangleq \begin{bmatrix} \boldsymbol{\mathcal{A}}_1\mathbf{h}_1 & \boldsymbol{\mathcal{B}}_1\mathbf{h}_1 & \cdots & \boldsymbol{\mathcal{A}}_{n_S}\mathbf{h}_1 & \boldsymbol{\mathcal{B}}_{n_S}\mathbf{h}_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\mathcal{A}}_1\mathbf{h}_{n_R} & \boldsymbol{\mathcal{B}}_1\mathbf{h}_{n_R} & \cdots & \boldsymbol{\mathcal{A}}_{n_S}\mathbf{h}_{n_R} & \boldsymbol{\mathcal{B}}_{n_S}\mathbf{h}_{n_R} \end{bmatrix}.$$

This linear relation between the modulated data symbols (contained in $\mathbf{s}_{\mathrm{LD}}$ and the complex receive symbols (rearranged in $\mathbf{y}_{\mathrm{LD}}$) implies that we can draw two essential conclusions. First, we note that the relation from Equation (4.9) shows that the receiver has to solve a number of real equations with $n_R L$ observations of the transmission (these are the entries of $\mathbf{Y}$ or equivalently of $\mathbf{y}_{\mathrm{LD}}$) to obtain $2n_S$ values (the transmitted data symbols). Since we assume perfect channel knowledge at the receiver (from which we can build the effective channel $\mathbf{H}_{\mathrm{LD}}$ since the receiver also knows the set of modulation matrices $\{\mathbf{A}_n, \mathbf{B}_n\}$), the system of equations between transmitter and receiver is not undetermined as long as $n_S \leq n_R L$. The second conclusion is that it is possible to derive the mutual information of an arbitrary linear STBC in terms of the effective channel $\mathbf{H}_{\mathrm{LD}}$. This can be used to derive the system capacity of the proposed LD codes, which is the maximization of the mutual information with respect to the modulation matrices.

The mutual information of the effective input-output relation from Equation (4.9) can be easily derived following the arguments in Subsection 2.3.1. The obtained mutual information can thus be stated as

$$I_{\mathrm{LD}}(\mathbf{y}_{\mathrm{LD}}; \mathbf{s}_{\mathrm{LD}}) = \frac{1}{2L} \log\det\left(\mathbf{I}_{2n_R L} + \frac{\varrho}{n_T}\mathbf{H}_{\mathrm{LD}}\mathbf{C}_{\mathbf{s}_{\mathrm{LD}}}\mathbf{H}_{\mathrm{LD}}^T\right),$$

where $\mathbf{C}_{\mathbf{s}_{\mathrm{LD}}}$ denotes the covariance matrix of $\mathbf{s}_{\mathrm{LD}}$ and we used the subscript $2n_R L$ to denote the size of the identity matrix. The term $1/(2L)$ ensures that the mutual information is in bits per channel use (since the effective channel is real valued and spans L channel uses). To proceed in the arguments of Subsection 2.3.1, assuming no CSI at the transmitter, the mutual information can be maximized by choosing $\mathbf{C}_{\mathbf{s}_{\mathrm{LD}}} = \mathbf{I}_{2n_S}$. At this point we can use the same arguments as in Subsection 2.3.2 to obtain the ergodic mutual information. In terms of the maximization over all possible input distributions, it remains to maximize over the modulation matrices $\mathbf{A}_n$ and $\mathbf{B}_n$. Therefore, the system capacity of LD codes can be stated as following:

Theorem 4.3.1 (System capacity of LD codes)**.** *Consider the effective channel representation from Equation (4.9). Then the system capacity of the proposed LD code is given by (compare [16])*

$$C_{LD} = \max_{\mathbf{A}_n, \mathbf{B}_n, n=1,\ldots,n_S} \frac{1}{2L}\mathbb{E}\left\{\log\det\left(\mathbf{I}_{2n_R L} + \frac{\varrho}{n_T}\mathbf{H}_{LD}\mathbf{H}_{LD}^T\right)\right\}.$$

Proof. The proof is implicitly done following the arguments to derive the mutual information of arbitrary linear STBC in the previous paragraph. □

The question now is whether the system capacity of LD codes can be equal to the ergodic channel capacity. Since we constrained the number of complex data symbols n_S to fulfill $n_S \leq n_T L$ it is clear that $C_{\mathrm{LD}} \leq C_E$, because in terms of capacity, we would have to transmit $n_T L$ independent Gaussian data symbols to achieve the ergodic capacity (as shown in Subsection 2.3.2), but if we choose n_S to be less than $n_T L$, we will not be able to reach the ergodic capacity. How near we can reach the ergodic channel capacity depends on the specific choice of the modulation matrices. As an example, by choosing a system with an equal number of transmit and receive antennas ($n_R = n_T$), therefore setting $n_S = n_R L$ and fixing the modulation matrices to form a transmission block matrix $\mathbf{S}$ according to a subsequent use of the

channel by means of a SM design, we would achieve the ergodic channel capacity. Nevertheless, there may exist other solutions to the maximization problem that will have a desirable gain in terms of error performance, mentioned in the beginning of this section. According to [16], the number of complex data symbols n_S should be chosen according to $n_S = \min\{n_T, n_R\}L$ because (compare Subsection 2.3.2) in the high SNR regime the ergodic channel capacity scales effectively with the number $\min\{n_T, n_R\}L$ of degrees of freedom.

To completely specify the maximization problem it remains to rewrite the power constraint on $\mathbf{S}$ in terms of the modulation matrices $\{\mathbf{A}_n, \mathbf{B}_n\}$. To do so, we use the definition of our linear STBC structure (Definition 4.1.2) and insert it in the power constraint $\mathbb{E}\{\operatorname{tr}\mathbf{S}\mathbf{S}^H\} = n_T L$. If we assume the real and imaginary parts of the complex data symbols s_n to be independent with variance $1/2$ respectively, one can easily show that the power constraint on the modulation matrices is given by

$$\sum_{n=1}^{n_S}\left[\operatorname{tr}\left(\mathbf{A}_n\mathbf{A}_n^H\right) + \operatorname{tr}\left(\mathbf{B}_n\mathbf{B}_n^H\right)\right] = 2Ln_T.$$

According to [16], the above power constraint can be replaced with the stronger constraint

$$\mathbf{A}_n^H\mathbf{A}_n = \mathbf{B}_n^H\mathbf{B}_n = \frac{L}{n_S}\mathbf{I},$$

for $n = 1, \ldots, n_S$. This constraint forces the real and imaginary parts of the complex data symbols to be dispersed with equal energy in all spatial and temporal dimensions. Furthermore, the corresponding maximum mutual information (and thus the LD system capacity) will be less or equal to the system capacity for the original power constraint. Concerning this point, Hassibi and Hochwald found out that the more stringent constraint generally imposes only a small information-theoretic penalty while having the advantage of better gains in term of error performance (or diversity).

Optimization of the modulation matrices

After identifying the important issues concerning the maximization in Theorem 4.3.1 we now work out the details. In general, no closed expression can be pronounced for the modulation matrices. Thus, we are forced to use numerically methods. In literature a variety of methods exists (see, e.g. [39]) but we choose a gradient based method (like in [16]). The basics concerning this numerical method may be found in [40]. Basically, gradient methods try to find a local optimum by taking steps proportional to the gradient of the goal function at the current point of the iteration. This sort of algorithm is very simple and furthermore has the advantage that MATLAB provides a toolbox for the application of such a maximization. The success of a gradient based maximum search is generally limited if the underlying goal function is not convex. Unfortunately this applies to our goal function from Theorem 4.3.1 so that it cannot be guaranteed that we found the global maximum. Nevertheless we (in compliance with [16]) observed that the non-convexity of the system capacity does normally not cause much problems.

To use the gradient based methods, it remains to derive an analytical expression for the gradient. Our result is stated in the following theorem:

Theorem 4.3.2 (Gradient of LD system capacity)**.** *The gradients of the LD system capacity from Theorem 4.3.1 with respect to the real and imaginary parts of the modulations matrices* $\mathbf{A}_n$ *and* $\mathbf{B}_n$ *are given by*

$$\begin{aligned}
\left[\frac{\partial C(\mathbf{A}_{R,n})}{\partial \mathbf{A}_{R,n}}\right]_{i,j} &= \frac{\varrho}{n_T L}\mathbb{E}\{\mathrm{tr}\,\left(\mathbf{M}_{A,R}\mathbf{Z}^{-1}\right)\},\\
\left[\frac{\partial C(\mathbf{A}_{I,n})}{\partial \mathbf{A}_{I,n}}\right]_{i,j} &= \frac{\varrho}{n_T L}\mathbb{E}\{\mathrm{tr}\,\left(\mathbf{M}_{A,I}\mathbf{Z}^{-1}\right)\},\\
\left[\frac{\partial C(\mathbf{B}_{R,n})}{\partial \mathbf{B}_{R,n}}\right]_{i,j} &= \frac{\varrho}{n_T L}\mathbb{E}\{\mathrm{tr}\,\left(\mathbf{M}_{B,R}\mathbf{Z}^{-1}\right)\},\\
\left[\frac{\partial C(\mathbf{B}_{I,n})}{\partial \mathbf{B}_{I,n}}\right]_{i,j} &= \frac{\varrho}{n_T L}\mathbb{E}\{\mathrm{tr}\,\left(\mathbf{M}_{B,I}\mathbf{Z}^{-1}\right)\},
\end{aligned}$$

where the matrices $\mathbf{M}_{A,R}$*,* $\mathbf{M}_{A,I}$*,* $\mathbf{M}_{B,R}$ *and* $\mathbf{M}_{B,I}$ *are defined as*

$$\begin{aligned}
\mathbf{M}_{A,R} &= \mathbf{I}_{n_R}\otimes\left(\mathbf{I}_2\otimes\left(\xi_i\eta_j^T\right)\right)\mathrm{vec}(\mathbf{H}')\,\mathrm{vec}(\mathbf{H}')^T\left[\mathbf{I}_{n_R}\otimes\boldsymbol{\mathcal{A}}_n\right]^T\\
\mathbf{M}_{A,I} &= \mathbf{I}_{n_R}\otimes\left(\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\otimes\left(\xi_i\eta_j^T\right)\right)\mathrm{vec}(\mathbf{H}')\,\mathrm{vec}(\mathbf{H}')^T\left[\mathbf{I}_{n_R}\otimes\boldsymbol{\mathcal{A}}_n\right]^T\\
\mathbf{M}_{B,R} &= \mathbf{I}_{n_R}\otimes\left(\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\otimes\left(\xi_i\eta_j^T\right)\right)\mathrm{vec}(\mathbf{H}')\,\mathrm{vec}(\mathbf{H}')^T\left[\mathbf{I}_{n_R}\otimes\boldsymbol{\mathcal{B}}_n\right]^T\\
\mathbf{M}_{B,I} &= \mathbf{I}_{n_R}\otimes\left(\begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix}\otimes\left(\xi_i\eta_j^T\right)\right)\mathrm{vec}(\mathbf{H}')\,\mathrm{vec}(\mathbf{H}')^T\left[\mathbf{I}_{n_R}\otimes\boldsymbol{\mathcal{B}}_n\right]^T,
\end{aligned}$$

and $\mathbf{H}'$ *is given by* $\mathbf{H}' = [\mathbf{H}_R, \mathbf{H}_I]^T$*.*

Proof. The proof is given in the Appendix, Subsection A.2.4. □

Using gradient based optimization, we were able to compute a maximization for the case of $n_R = n_T = 2$, $L = 2$, $R = 4$ and $\varrho = 20\,\mathrm{dB}$. Our obtained result is given in Table 4.1. Because of the non-convexity of the goal function, we cannot guarantee that the found solution is optimal. Nevertheless, we observe in the following subsection that our solution achieves the ergodic channel capacity. Furthermore, the obtained solution is highly nonunique: Simply reordering the modulation matrices with respect to n gives another solution, as does pre- or post-multiplying all the matrices by the same unitary matrix. However, there is also another source of nonuniqueness. If we multiply our transmit vector $\mathbf{s}_{\mathrm{LD}}$ in the effective input-output relation from Equation (4.9) by a $2n_S \times 2n_S$ orthogonal matrix $\boldsymbol{\Phi}^T$ to obtain a new vector $\mathbf{s}'_{\mathrm{LD}} = \boldsymbol{\Phi}^T\mathbf{s}_{\mathrm{LD}}$ with entries that are still independent and have the same variance as $\mathbf{s}_{\mathrm{LD}}$. Thus we can write the effective input-output relation as

$$\mathbf{y}_{\mathrm{LD}} = \sqrt{\frac{\varrho}{n_T}}\mathbf{H}_{\mathrm{LD}}\boldsymbol{\Phi}\boldsymbol{\Phi}^T\mathbf{s}_{\mathrm{LD}} + \mathbf{n}_{\mathrm{LD}} = \sqrt{\frac{\varrho}{n_T}}\mathbf{H}'_{\mathrm{LD}}\mathbf{s}'_{\mathrm{LD}},$$

where we set $\mathbf{H}'_{\mathrm{LD}} = \mathbf{H}_{\mathrm{LD}}\boldsymbol{\Phi}$. Since the entries of $\mathbf{s}_{\mathrm{LD}}$ and $\mathbf{s}'_{\mathrm{LD}}$ have the same joint distribution, the maximum mutual information obtained from the channels $\mathbf{H}_{\mathrm{LD}}$ and $\mathbf{H}'_{\mathrm{LD}}$ are the same.

n	$\mathbf{A}_n$	$\mathbf{B}_n$
1	$\begin{bmatrix} -0.6237 + j0.2313 & -0.2237 + j0.0859 \\ 0.0123 - j0.2394 & -0.0266 + j0.6647 \end{bmatrix}$	$\begin{bmatrix} 0.2274 + j0.0879 & 0.0766 - j0.6593 \\ 0.4673 - j0.4713 & 0.2106 + j0.1228 \end{bmatrix}$
2	$\begin{bmatrix} -0.4074 + j0.0036 & 0.1405 + j0.5606 \\ -0.5245 + j0.2426 & -0.2531 - j0.3194 \end{bmatrix}$	$\begin{bmatrix} -0.0533 + j0.1833 & -0.6672 - j0.1355 \\ 0.4217 + j0.5346 & 0.1396 - j0.1302 \end{bmatrix}$
3	$\begin{bmatrix} -0.4524 - j0.6371 & 0.1788 - j0.1972 \\ 0.1303 - j0.2321 & -0.6526 + j0.0577 \end{bmatrix}$	$\begin{bmatrix} -0.5742 + j0.3630 & 0.1425 - j0.1364 \\ 0.0145 - j0.1956 & -0.0858 - j0.6736 \end{bmatrix}$
4	$\begin{bmatrix} -0.2723 - j0.0259 & -0.6185 - j0.1254 \\ 0.3985 + j0.5161 & -0.1721 - j0.2684 \end{bmatrix}$	$\begin{bmatrix} 0.3757 + j0.4191 & -0.3660 - j0.2267 \\ -0.4053 - j0.1377 & -0.5605 - j0.0221 \end{bmatrix}$

Table 4.1.: Optimized LD code for $n_R = n_T = 2$, $L = 2$ at $\varrho = 20\,\mathrm{dB}$.

So, if we write

$$\begin{aligned} C_{\mathrm{LD}} &= \max_{\mathbf{A}_n, \mathbf{B}_n, n=1,\ldots,n_S} \frac{1}{2L} \mathbb{E}\left\{ \log\det\left(\mathbf{I}_{2n_R L} + \frac{\varrho}{n_T} \mathbf{H}'_{\mathrm{LD}} \mathbf{H}'^{T}_{\mathrm{LD}} \right) \right\} \\ &= \max_{\mathbf{A}_n, \mathbf{B}_n, n=1,\ldots,n_S} \frac{1}{2L} \mathbb{E}\left\{ \log\det\left(\mathbf{I}_{2n_R L} + \frac{\varrho}{n_T} \mathbf{H}_{\mathrm{LD}} \mathbf{\Phi}\mathbf{\Phi}^T \mathbf{H}^T_{\mathrm{LD}} \right) \right\}, \end{aligned}$$

one easily sees that since $\mathbf{\Phi}\mathbf{\Phi}^T = \mathbf{I}$, the multiplication by $\mathbf{\Phi}$ does not influence the mutual information. Furthermore, since $\mathbf{H}_{\mathrm{LD}}$ includes the modulation matrices $\{\mathbf{A}_n, \mathbf{B}_n\}$, we could redefine them in a way that the entries of $\mathbf{\Phi}$ are only contained in these redefined modulation matrices $\{\mathbf{A}'_n, \mathbf{B}'_n\}$. Thus, the transformation is another source of nonuniqueness.

Nevertheless, the mentioned transformation can be used to change the obtained dispersion code in a way that it satisfies other criteria (such as diversity) without sacrificing mutual system capacity. An example of a promising choice for $\mathbf{\Phi}$ will be given in Subsection 4.3.3.

4.3.2. Capacity Comparison

After we have defined the structure of LD codes and found a way to find good solutions for the modulation matrices to maximize the mutual information, we want to investigate how good the obtained code behaves compared to the ergodic channel capacity. Therefore we did some numerical simulations with the code given in Table 4.1, by numerically evaluating its system capacity and compare it to the ergodic channel capacity of a $n_T = n_R = 2$ MIMO channel. Our results are plotted in Figure 4.4. To emphasize the benefit of the optimized LD code compared to the Alamouti OSTBC code from Equation 4.2, we also plotted the Alamouti OSTBC system capacity curve.

The curves clearly show that the found LD code is able to achieve the ergodic capacity, although we optimized it just for a fixed SNR of 20dB. Furthermore, the benefit of the

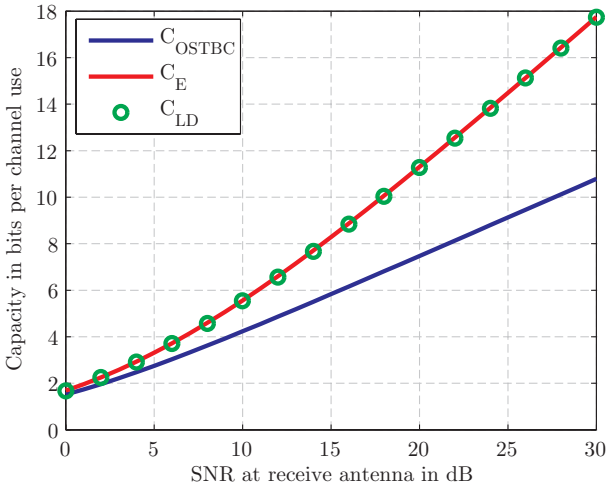

Figure 4.4.: Comparison of ergodic channel capacity in the $n_T = n_R = 2$ case with the system capacity of our optimized LD Code and the system capacity of the Alamouti OSTBC code.

proposed LD code in comparison to the Alamouti OSTBC in terms of capacity is underlined.

4.3.3. Error Performance of LD Codes

The result of the previous subsection showed the impressive advantage of LD codes in terms of capacity. Nevertheless, it remains to investigate the error performance of LD codes to give an objective answer to the question, whether the proposed LD code is superior to the Alamouti OSTBC.

Because of the nonuniqueness of the solutions found by the maximization so far, we are not able to present a general solution concerning the error performance and the diversity. But we want to point out an interesting aspect of codes for high rate transmissions (for which LD codes are more interesting, since the SNR gap to achieve the desired rate between OSTBCs and LD codes is large). Usually STC design is based on the rank criterion as stated in Section 3.3. This criterion only depends on matrix pairs and therefore does not exclude matrix designs with low spectral efficiencies. At high rates, the number of code matrices $\mathbf{S}$ in the constellation $\mathcal{S}$ is roughly exponential in the channel capacity at a given SNR. This number can be vary large, for example a $R = 16$ code for a $n_R = n_T = L = 4$ system effectively has $|\mathcal{S}| = 2^{RL} \approx 18.447 \cdot 10^{18}$ different code matrices. So even if the rank $r(\mathbf{\Gamma}_{i,j})$ is equal to one for many codeword pairs, the probability of encountering one of these matrix pairs may still be exceedingly small and thus the constellation performance may still be excellent. This reverses in the high SNR regime, since according to our mutual information simulations in Chapter 3, the mutual information saturates for a fixed symbol alphabet (thus reducing the relative spectral efficiency of the code

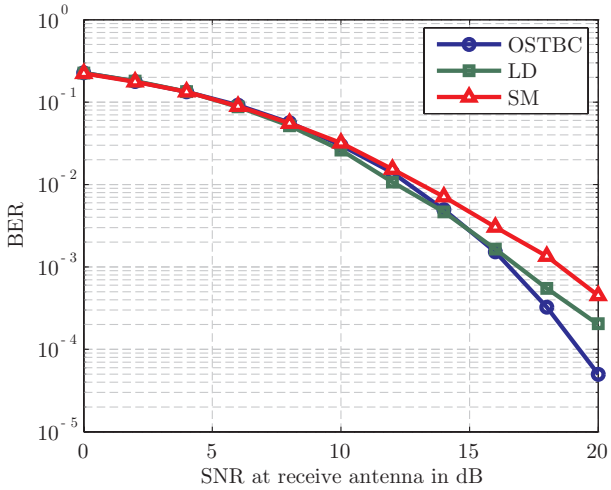

Figure 4.5.: BER performance comparison for a $n_T = n_R = 2$ MIMO channel at a rate of $R = 4$ bits/channel use.

compared with the channel capacity) and making a decoding error to a near neighbor more important.

Nevertheless we did some numerical simulations of the BER performance of our optimized LD code to visualize its error performance. Our first simulation is performed over a $n_T = n_R = 2$ MIMO channel with $L = 2$ channel uses. We are comparing the performance of three MIMO systems at $R = 4$bits/channel use. The results are plotted in Figure 4.5, where we chose a 16-QAM symbol alphabet for the OSTBC system and a 4-QAM symbol alphabet for the SM and the LD system respectively. We can draw the following conclusions from this result. First, the BER performance of the LD code is intersecting with the BER curve of the OSTBC. Second, the LD code always performs better than the SM system. The reason for the first observation is that in the medium SNR regime (approximately between 8 and 13dB), the BER performance of the codes is determined by the ability of the system to support the given rate. In this medium SNR region, it seems that we are discovering the gap between the OSTBC system capacity and the ergodic channel capacity. Furthermore, because we are transmitting two data symbols per channel use, we can use a smaller symbol alphabet to achieve the same rate. In the high SNR regime however, the pairwise error probability is the limiting factor of the BER performance and concerning this criterion (as we mentioned in the previous paragraph), the OSTBC is better designed, thus explaining the intersection. Nevertheless, the LD code is superior to the SM design. This can be explained because in the optimized LD design the data symbols are spread over space and time thus allowing a small coding gain (as defined in the discussion of the determinant criterion in Section 3.3). Finally, the proposed LD code seems to achieve a diversity gain of 2, which is equal to the diversity gain of the SM design. Thus, OSTBCs can exploit more diversity.

Our second example treats the same MIMO channel as before ($n_T = n_R = 2$), but at a different rate of $R = 8$bits/channel use. In this case we chose a 256-QAM as symbol alphabet for the

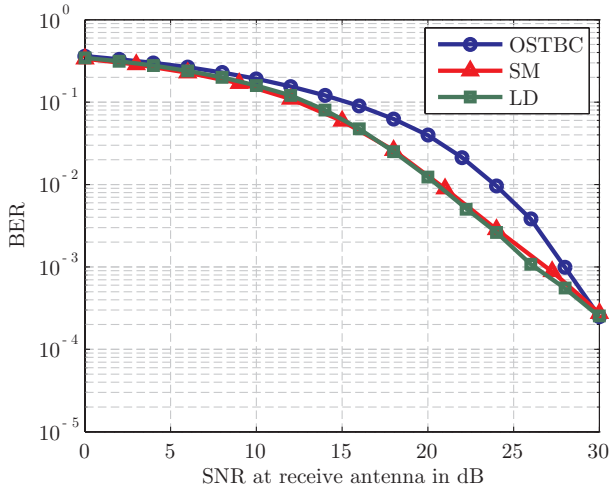

Figure 4.6.: BER performance comparison for a $n_T = n_R = 2$ MIMO channel at a rate of $R = 8$ bits/channel use.

OSTBC, and a 16-QAM symbol alphabet for the SM and the LD system, respectively. The results of our simulation are plotted in Figure 4.6. These curves support our observations from the last example, although they are much more pronounced. In the high rate regime, the limiting factor of the BER performance (in case of the ML receiver) is the spectral efficiency of the system. The difference between the SM and the optimized LD code is much smaller, thus indicating that the LD code is not able to realize a big coding gain in this case. Also, like observed in the previous example, the optimized LD code performs worse than the OSTBC in the high SNR regime, which is due to the lower diversity gain.

4.3.4. Number Theory Extension

After having identified the main drawback of the optimized LD code in terms of error performance, we investigate a special extension which can be used to improve the diversity gain of LD codes.

The main indication for doing so has been already observed by investigating the equality of the system capacity for a data symbol modulation by an orthogonal matrix $\mathbf{\Phi}^T$. In [41], the problem of ST diversity gain is related to algebraic number theory and the coding gain optimization to the theory of simultaneous Diophantine approximation in the geometry of numbers for the case of a $n_T = L = 2$ STBC. To relate their findings to the proposed LD codes, we first note that another solution to the maximization of the system capacity in Theorem 4.3.1 in the case of $n_T = n_R = L = 2$ with Q chosen to be four, is (compare [16])

$$\mathbf{A}_{2(k-1)+l} = \mathbf{B}_{2(k-1)+l} = \frac{1}{\sqrt{2}}\mathbf{D}^{k-1}\mathbf{\Pi}^{l-1}, \tag{4.10}$$

where we defined

$$\mathbf{D} \triangleq \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \mathbf{\Pi} \triangleq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

According to Definition 4.1.2, this results in a transmission block matrix structure of

$$\mathbf{S} = \frac{1}{\sqrt{2}} \begin{bmatrix} s_1 + s_3 & s_2 + s_4 \\ s_2 - s_4 & s_1 - s_3 \end{bmatrix}. \tag{4.11}$$

As mentioned above, in [41] another STBC structure is proposed, which has been optimized based on number theory. The explanation of the corresponding theoretical fundament is beyond the scope of this thesis, but we would like to give an idea of the basics. Therefore we repeat the following proposition from [41].

Theorem 4.3.3 (Number theory diversity gain optimum)**.** *If a STBC code structure*

$$\mathbf{S} = \frac{1}{\sqrt{2}} \begin{bmatrix} s_1 + \phi s_3 & \theta(s_2 + \phi s_4) \\ \theta(s_2 - \phi s_4) & s_1 - \phi s_3 \end{bmatrix},$$

with $\theta^2 = \phi$ is used, the corresponding diversity gain is maximized if ϕ is an algebraic number of degree ≥ 4 over $\mathbb{Q}[j]$ for all symbol constellations carved from $\mathbb{Z}[j]$. Here $\mathbb{Z}[j]$ denotes the ring of complex integers and $\mathbb{Q}[j]$ the field of complex rational.

Proof. Can be found in [41]. □

If we consider a code structure as in Theorem 4.3.3, we can write the coding gain of the STBC, $\delta(\phi)$, given by the determinant criterion of Subsection 3.3 as (compare [41]):

$$\delta(\phi) = \inf_{\mathbf{s} \neq (0,0,0,0)^T \in \mathbb{Z}[i]^4} \left(\det\left(\mathbf{S}\mathbf{S}^H\right)\right)^{1/2},$$

where we defined $\mathbf{s} = (s_1, \ldots, s_4)^T$ to be the vector of the transmitted complex data symbols and we defined $\mathbb{Z}[i]$ to be the ring of complex integers. In the case of four transmitted symbols, $\mathbf{s}$ is accordingly defined over the ring $\mathbb{Z}[i]^4$. In the case of finite symbol alphabets, the above equation simplifies to

$$\delta(\phi) = \min_{\mathbf{s} = \mathbf{s}_1 - \mathbf{s}_2, \mathbf{s}_1 \neq \mathbf{s}_2 \in \mathcal{S}} \left| s_1^2 - s_2^2\phi - s_3^2\phi^2 + s_4^2\phi^3 \right|,$$

where $\mathbf{s}_1$ and $\mathbf{s}_2$ denote a pair of possible transmit symbol vectors $\mathbf{s}$, drawn from the constellation $\mathcal{S}$. The main finding of [41] is that if ϕ is an algebraic number of degree ≥ 4 over $\mathbb{Q}[i]$, then one guarantees the maximum transmit diversity over all constellations carved from $\mathbb{Z}[i]^4$. Here, we denoted the ring of complex rational by $\mathbb{Q}[i]$. Without going to far into detail, we note that for ϕ to be algebraic, there exists an unique irreducible polynomial of degree n, which has ϕ as a root. Now if ϕ is an algebraic number of degree ≥ 4 (and thus the polynomial is of degree ≥ 4) over $\mathbb{Q}[i]$, then $\{1, \phi, \phi^2, \phi^3\}$ is a so-called "free set", if $\sum_{j=0}^{3} a_j \phi^j = 0$ (with $a_j \in \mathbb{Q}[i]$) results in $a_0 = a_1 = a_2 = a_3 = 0$. This guarantees that $\delta(\phi) \neq 0$ for all constellations carved from $\mathbb{Z}[i]^4$ and thus leads to the maximum transmit diversity. Furthermore, in [41], it is shown that if ϕ is an algebraic number of degree 2 over $\mathbb{Q}[i]$ and if $\phi^2 \in \mathbb{Q}[i]$, then

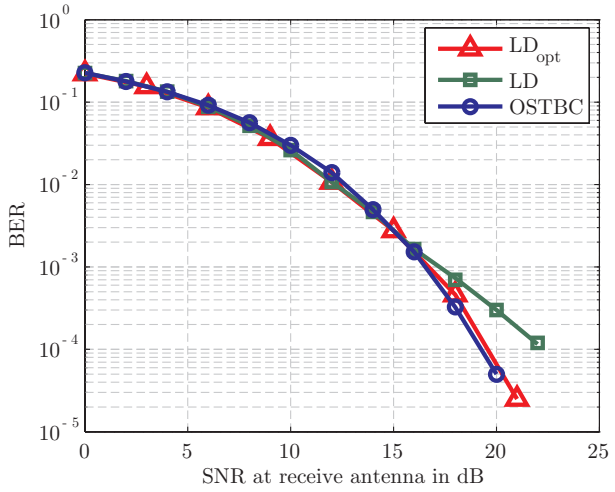

Figure 4.7.: BER performance comparison of the number theory optimized LD code at a rate $R = 4$ in a $n_R = n_T = 2$ MIMO channel.

one can also guarantee the maximum transmit diversity over all constellations carved from $\mathbb{Z}[i]^4$, which leads to the proposed code design.

Using the above STBC structure and comparing it with our structure of the LD code, one may see that these two merge in case of the following redefinition of the modulation matrices.

Theorem 4.3.4 (Number theory optimized LD code for $n_R = n_L = L = 2$)**.** *The modulation matrices of the LD code which corresponds to the number theory optimized STBC structure from Definition 4.3.3 are given by*

$$\mathbf{A}_{opt,1} = \mathbf{B}_{opt,1} = \mathbf{A}_1, \quad \mathbf{A}_{opt,2} = \mathbf{B}_{opt,2} = \theta\mathbf{A}_2,$$
$$\mathbf{A}_{opt,3} = \mathbf{B}_{opt,3} = \phi\mathbf{A}_3, \quad \mathbf{A}_{opt,4} = \mathbf{B}_{opt,4} = \theta\phi\mathbf{A}_4,$$

where the $\mathbf{A}_n, n = 1, \ldots, n_S$ *are defined as in Equation 4.10.*

Proof. The proof follows directly by inspection of the structure in Equation (4.11) and the one in Definition 4.3.3. □

Furthermore one may show that the new modulation matrices can be rewritten in a way that the effective channel (Equation (4.9)) can be rewritten in the form $\mathbf{y}_{\text{LD}} = \sqrt{\varrho/n_T}\mathbf{H}_{\text{LD}}\mathbf{\Phi}\mathbf{\Phi}^T\mathbf{s}_{\text{LD}} + \mathbf{n}_{\text{LD}}$ with $\mathbf{\Phi}$ being unitary. Thus according to our already stated arguments, the system capacity of the LD codes stays the same, though improving its diversity performance. The proof is given in the Appendix, Subsection A.2.5. According to [41], the algebraic number ϕ is optimized according to the used symbol alphabet. For the case of a 4-QAM, ϕ is given by $\phi = \exp(j/2)$.

With these insights we performed another numerical simulation of the BER performance of the LD code for the case $n_R = n_T = L = 2$ with $n_S = 4$ and using a 4-QAM symbol alphabet, thus resulting in a rate of $R = 4$bits/channel use. The obtained results are plotted in Figure 4.7. For comparison we also plotted the curves from the basic LD code and the curve of the Alamouti OSTBC for the same rate. We can clearly see that the LD code optimized by means of number theory performs equally well in terms of diversity as the OSTBC does (which means that it achieves full diversity), but offering the big advantage of a system capacity equally to the ergodic channel capacity.

To conclude this chapter we want to note that the optimization of the LD codes was always performed assuming an iid Gaussian model for the channel $\mathbf{H}$. If a correlated channel is assumed, the solution no longer is optimal in terms of system capacity. The interested reader may find additional information regarding the modified optimization for example in [42].

5. Diversity-Multiplexing Tradeoff

In the previous chapters we have discussed that different systems perform differently well in terms of system capacity (or equivalently multiplexing gain) and error performance (or equivalently diversity gain). Although we encountered that the LD system which is able to achieve the ergodic channel capacity has in general a bad error performance and that a OSTBC system performs vice versa, the question if there exists a tradeoff between these two performance measures remained unanswered.

In [26], Zheng and Tse established that there is a tradeoff between these two types of gains (multiplexing and diversity), i.e., how fast error probability can decay and how rapidly data rate can increase with SNR. To relate the notation of the cited paper, we note that the term *scheme* corresponds somehow to the term *system* used in the context of this thesis.

In this chapter we will show a short derivation of the optimal tradeoff (based on the proofs given in [26]) and its connection to the outage probability as well as the error probability. Furthermore we will try to establish a way to visualize the tradeoff by means of the outage probability (as also done in [43] and [44]). We then evaluate the tradeoffs achieved by OSTBC systems and LD systems treated in this thesis. We will conclude this chapter by providing an interesting connection of the diversity-multiplexing tradeoff to the theory of asymptotic-information lossless designs.

5.1. The Optimal Tradeoff

Within this section we provide the optimal tradeoff for a given MIMO channel (i.e. determined by the number of receive (n_R) and transmit (n_T) antennas), which is the upper bound achievable by any ST system. To do so, we formally follow the arguments given in [26], although only presenting the basic steps since a complete treatment of the underlying theory goes beyond the scope of this thesis. At the beginning, let us define the diversity gain d and the multiplexing gain r.

Definition 5.1.1 (Diversity gain and multiplexing gain)**.** *For a given SNR ϱ, let $R(\varrho)$ be the transmission rate and let $P_e(\varrho)$ be the packet error probability at that rate. Then a MIMO system achieves a spatial multiplexing gain r if*

$$r \triangleq \lim_{\varrho \to \infty} \frac{R(\varrho)}{\log \varrho},$$

and a diversity gain d if

$$d \triangleq -\lim_{\varrho \to \infty} \frac{\log P_e(\varrho)}{\log \varrho}.$$

These definitions are motivated by two observations we already made in this thesis. Since the performance gain at high SNR is dictated by the SNR exponent of the error probability, the above definition somehow "extracts" its exponent, which is the diversity gain we always referred to. Furthermore, in Subsection 2.3.2 we described the ergodic channel capacity behavior in the high SNR regime. The result suggests that the multiple-antenna channel can be viewed as $\min\{n_T, n_R\}$ parallel spatial channels - hence the number $\min\{n_T, n_R\}$ is the total number of degrees of freedom to communicate. The idea of transmitting independent information symbols in parallel through the spatial channels is called spatial multiplexing. Now let us think of a system for which we increase the data rate with SNR (for example by simply changing the size of the symbol alphabet depending on the SNR). Then we can write $R(\varrho)$ as the actual rate of the system at the given SNRs. Certainly, the maximum achievable rate $R(\varrho)$ of a system is its system capacity. So we can interpret the maximum spatial multiplexing gain $r_{\max}$ as the slope of the system capacity in the limit of $\varrho \to \infty$.

After having stated the two performance gains in terms of rate (or system capacity) and error probability (or diversity), we can investigate the optimum tradeoff between these two gains. Therefore let us denote $d^*(r)$ to be the optimal tradeoff achievable. It seems intuitive to define $d^*(r)$ to be the supremum of the diversity advantage at a given multiplexing gain r over all schemes, i.e.

$$d^*(r) \triangleq \sup d(r).$$

With this in mind, the maximum achievable diversity gain and the maximum achievable spatial multiplexing gain in a MIMO channel can be denoted as $d^*_{\max} = d^*(0)$ and $r^*_{\max} = \sup\{r : d^*(r) \geq 0\}$. For the derivation of the optimal tradeoff $d^*(r)$ we want to use a special notation also used by Zheng and Tse to denote an exponential equality, i.e., $f(x) \doteq x^b$ denotes

$$\lim_{x\to\infty} \frac{\log f(x)}{\log x} = b.$$

With this notation, diversity gain can also be written as $P_e(\varrho) \doteq \varrho^{-d}$. The notations $\dot{\leq}$ and $\dot{\geq}$ are defined similarly.

Before going into detail on the derivation of the optimal tradeoff, we want to note some important facts. The diversity gain defined for the optimal tradeoff differs from the one derived in Section 3.3, because the diversity gain defined there is an asymptotic performance metric of a system with a fixed rate. To be specific, until now, the speed that the error probability (of the ML detector) decays as SNR increases at a specific (but fixed) rate has been called the diversity gain. Now we relax the definition and allow a change of the rate of the transmission. This is done, because in the formulation of the diversity-multiplexing tradeoff, the ergodic channel capacity (or the system capacity) increases linearly with $\log \varrho$ and hence, in order to achieve a nontrivial fraction of the capacity at high SNR, the input data rate must also increase with SNR. Accordingly, this implies that for a given system, the symbol alphabet of the transmitted data symbols has to increase. Under this constraint, any fixed code (a system with fixed number of data symbols and symbol alphabet size) has a spatial multiplexing gain of 0. To see this, compare our results from Chapter 3, which show that the mutual information (or the actual data rate) for a fixed symbol alphabet saturates, thus in the limit of $\varrho \to \infty$ showing a slope of 0, which would be the associated multiplexing gain. On the other hand, if we fix the rate, the system could realize the maximum possible diversity gain (compare also Subsection 2.3.4). Over all, this means that in the context of the diversity-multiplexing

tradeoff, we operate with non-fixed rates of transmission and allow a trading of the decay of the error probability for a non-vanishing multiplexing gain.

Having gained a basic understanding of the used terminology, we can proceed in deriving the optimal tradeoff $d^*(r)$. A complete (and formally clean) derivation may be found in [26], but to gain a basic understanding of the tradeoff itself, we will state the important steps of the derivation. We start by noting that the probability of a packet error in a MIMO transmission may be upper bounded by the outage probability derived from the MIMO channel capacity (see Subsection 2.3.4 for details).

Theorem 5.1.2 (Outage probability bound on error rate)**.** *For any coding (or system) the average (over the channel realizations) probability of a packet error at rate $R = r \log \varrho$ is lower-bounded by*

$$P_e(\varrho) \geq p_{out}(R).$$

Proof. Let the transmitted block matrix $\mathbf{S}$ be $\in \mathbb{C}^{n_T \times L}$, which is uniformly drawn from a codebook $\mathcal{S}$. Since we assume the channel fading coefficients of $\mathbf{H}$ to be unknown at the transmitter, we assume that $\mathbf{S}$ is independent of $\mathbf{H}$. Conditioned on a specific channel realization $\mathbf{H} = \mathbf{H}_0$, we write the mutual information of the channel as $I(\mathbf{S}; \mathbf{Y}|\mathbf{H} = \mathbf{H}_0)$, and the probability of an error as $\Pr(\text{error}|\mathbf{H} = \mathbf{H}_0)$. Then by the usage of Fano's inequality (see Subsection A.1.6), we get

$$\begin{aligned} 1 + \Pr(\text{error}|\mathbf{H} = \mathbf{H}_0) \log |\mathcal{S}| &\geq H(\mathbf{S}|\mathbf{H} = \mathbf{H}_0) - I(\mathbf{S}; \mathbf{Y}|\mathbf{H} = \mathbf{H}_0) \\ &= H(\mathbf{S}) - I(\mathbf{S}; \mathbf{Y}|\mathbf{H} = \mathbf{H}_0). \end{aligned}$$

Since our rate is fixed by R bits per channel use, the size of $\mathcal{S}$ is $|\mathcal{S}| = 2^{RL}$, and we assumed $\mathbf{S}$ to be drawn uniformly from $\mathcal{S}$ (which implies that $H(\mathbf{S}) = \log |\mathcal{S}|$), the above equation simplifies to

$$1 + \Pr(\text{error}|\mathbf{H} = \mathbf{H}_0) RL \geq RL - I(\mathbf{S}; \mathbf{Y}|\mathbf{H} = \mathbf{H}_0),$$

which can be rewritten to

$$\Pr(\text{error}|\mathbf{H} = \mathbf{H}_0) \geq 1 - \frac{I(\mathbf{S}; \mathbf{Y}|\mathbf{H} = \mathbf{H}_0)}{RL} - \frac{1}{RL}.$$

Now average over $\mathbf{H}$ to get the average error probability $P_e(\varrho) = \mathbb{E}\{\Pr(\text{error}|\mathbf{H} = \mathbf{H}_0)\}$. Then, for any $\delta > 0$, and any $\mathbf{H}_0$ in the set $\mathcal{D}_\delta \triangleq \{\mathbf{H}_0 : I(\mathbf{S}; \mathbf{Y}|\mathbf{H} = \mathbf{H}_0) < R - \delta\}$ (which is exactly the definition of the outage event at rate $R-\delta$ and thus $\Pr(\mathcal{D}_\delta)$ denotes the outage probability), the probability of error is lower bounded by

$$P_e(\varrho) \geq \left(1 - \frac{R - \delta}{RL} - \frac{1}{RL}\right) \Pr(\mathcal{D}_\delta). \tag{5.1}$$

By taking $\delta \to 0$, $\Pr(\mathcal{D}_\delta)$ gets $p_{\text{out}}(R)$ and in the case of indefinitely long coding ($L \to \infty$), we obtain $P_e(\varrho) \geq p_{\text{out}}(R)$, which concludes the proof. □

With the outage probability as a lower bound on the error probability, Zheng and Tse showed in [26] that the outage probability at a given rate is exponentially equivalent to

$$p_{\text{out}}(R) \doteq \Pr\left[\log\det\left(\mathbf{I} + \frac{\varrho}{n_T}\mathbf{H}\mathbf{H}^H\right) < R\right] \doteq \Pr\left[\prod_{i=1}^{n}(1 + \varrho' \lambda_i) < \varrho'^r\right].$$

Let $\lambda_i = \varrho'^{-\alpha_i}$. Then, at high SNR, we have $(1 + \varrho'\lambda_i) \doteq \varrho'^{(1-\alpha_i)^+}$, and thus we can write

$$p_{\text{out}}(R) \doteq \Pr\left[\sum_{i=1}^{n}(1+\alpha_i)^+ < r\right],$$

where $\varrho' = \varrho/n_T$ (we will drop the $(\cdot)'$ in the following) and $\lambda_i, i = 1, \ldots, n$ denote the ordered singular values of $\mathbf{H}\mathbf{H}^H$. Furthermore, the rate R is expressed by $R = r \log \varrho$ and $(x)^+$ denotes $\max\{0, x\}$. By evaluating the probability density of α and taking the limit $\varrho \to \infty$, they showed that the outage probability is exponentially equal to

$$p_{\text{out}}(R) \doteq \varrho^{-d_{\text{out}}(r)},$$

as long as $0 \leq r \leq \min\{n_T, n_R\}$. The obtained diversity curve $d_{\text{out}}(r)$ is denoted by the subscript *out* since it refers to an upper bound on the optimal diversity-multiplexing tradeoff.

To further deepen our understanding of the tradeoff, we want to investigate its connection to the pairwise error probability (PEP). Without going too far into detail, we just want to state that Zheng and Tse showed in [26] that the PEP is exponentially equal to

$$\Pr\left(\mathbf{S}^{(i)} \to \mathbf{S}^{(j)}\right) \doteq \varrho^{-n_R \sum_{i=1}^{n_T}(1-\alpha_i)^+}.$$

The quantity $\sum_{i=1}^{n_T}(1-\alpha_i)^+$ is implicitly a function of the multiplexing gain r. As the rate R increases with SNR, the codebook and therefore the matrix $\mathbf{\Delta}_{i,j}$ changes, which in turn affects the α_i. The diversity curve obtained by the analysis of the PEP can be denoted by $d_{\text{G}}(r)$ and provides a lower bound on the optimal diversity-multiplexing tradeoff. Zheng and Tse showed that for block lengths $L \geq n_T + n_R - 1$ the lower and upper bound always coincide. Nevertheless, throughout this thesis, we will refer to the upper bound $d_{\text{out}}(r)$ if we talk about the optimal diversity-multiplexing tradeoff (as it is done, e.g. in [43]) even in the case of $L < n_T + n_R - 1$.

With this insight, the average error probability can be exponentially bounded by $\varrho^{-d_{\text{out}}(r)}$, but in addition, there is one more interesting fact to discover. Take Equation (5.1) and substitute R by $r \log \varrho$. Then in the limit of $\varrho \to \infty$, the bound holds even if we choose to code over a finite block length $L < \infty$. Thus, the optimal tradeoff curve gets tight for $\varrho \to \infty$ even for $L < \infty$. Thus, we can use the outage probability of the system capacity to derive the tradeoff curve achieved by a given MIMO system. In the case that we derive the outage probability from the ergodic channel capacity, $d_{\text{out}}(r)$ is the optimum tradeoff curve $d^*(r)$ (strictly speaking for $L \geq n_T + n_R - 1$), and by a consequent analysis of the α_i in the case of an iid channel $\mathbf{H}$, Zheng and Tse were able to compute $d^*(r)$.

Theorem 5.1.3 (Optimal diversity-multiplexing tradeoff)**.** *The optimal tradeoff curve $d^*(r)$ is given by the piece-wise linear function connecting the points $(k, d^*(k))$, $k = 0, \ldots, \min\{n_T, n_R\}$ and*

$$d^*(k) = (n_T - k)(n_R - k).$$

Proof. Is given in [26]. □

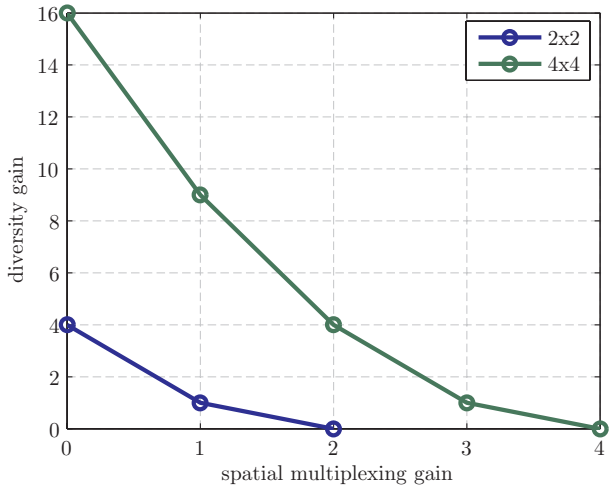

Figure 5.1.: Optimal diversity-multiplexing tradeoff curve for a $n_T = n_R = 2$ and a $n_T = n_R = 4$ MIMO channel.

In Figure 5.1, the optimal tradeoff curves of two different MIMO channels are depicted. In the case of the $n_T = n_R = 2$ MIMO channel, we can clearly see that the maximum achievable diversity gain $d^*_{\max}$ is four, which corresponds to the full diversity of $n_R n_T$ we mentioned in the previous chapters. Furthermore, we see that the maximum spatial multiplexing gain $r^*_{\max}$ is two, which corresponds to a simultaneous transmission of two symbols per channel use. In general, the tradeoff curve intersects the r axis at $\min\{n_T, n_R\}$. This means that the maximum achievable spatial multiplexing gain $r^*_{\max}$ is the total number of degrees of freedom provided by the channel. On the other hand, the curve intersects the d axis at the maximal diversity gain $d^*_{\max} = n_R n_T$ corresponding to the total number of independent fading coefficients.

To conclude, we want to note that the optimal tradeoff bridges the gap between the two design criteria diversity and spatial multiplexing we were talking about in the preceding chapters of this thesis. The tradeoff curve provides a more complete picture of the achievable performance over MIMO channels.

5.1.1. Visualizing the Tradeoff

To visualize the optimal tradeoff for the $n_R = n_T = 2$ MIMO channel depicted in Figure 5.1, we show the relationship between SNR, rate and outage probability by plotting p_{out} as functions of SNR for various rates. The result is plotted in Figure 5.2. Each curve represents how outage probability decays with SNR for a fixed rate R. As R increases, the curves shift to higher SNR. To see the diversity-multiplexing tradeoff for each value of r, we evaluate p_{out} as a function of SNR and $R = r \log_2 \varrho$ for a sequence of increasing SNR values and plot $p_{\text{out}}(r \log_2 \varrho)$ as a function of SNR for various r values. In Figure 5.3, several such curves are plotted for various values of r; each is labeled with the corresponding r and $d_{\text{out}}(r)$ values. Figure 5.2 is

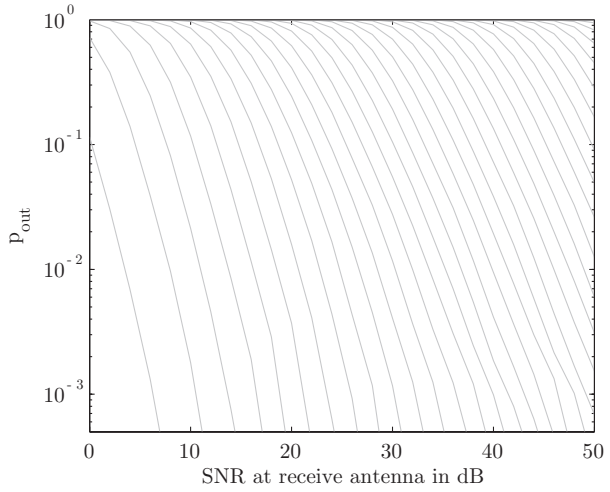

Figure 5.2.: Family of outage probability curves as functions of SNR for target rates of $R = 1, 2, \ldots, 40$ bits per channel use for the $n_T = n_R = 2$ MIMO channel.

overlaid as grey lines. For comparison purpose, we draw dashed lines with slopes $d^*(r)$ for the according multiplexing gain r values. According to Theorem 5.1.3, the solid and dashed curves have matching slopes for high SNR. We see that when R increases faster with SNR (i.e. r is larger), the corresponding outage probability decays slower over the SNR (i.e. d decreases). This is the fundamental diversity-multiplexing tradeoff.

To obtain further intuition, we perform the following approximation. Instead of $p_{\text{out}}(R) \doteq \varrho^{-d_{\text{out}}(r)}$, we replace the asymptotic equality $\doteq$ with an exact $=$. This approximation turns the smooth $p_{\text{out}}(R)$ curves into piecewise linear lines, since for growing SNR and a fixed rate, the multiplexing gain r decreases and thus, the exponent $d_{\text{out}}(r)$ is $3r-4$ for $r < 1$ and $r-2$ for $r \geq 1$. This results in the two different slopes of the outage probability curve. Figure 5.4 shows the linearized outage probability curves (solid black). For comparison (and as a visual proof to that the approximation is valid) we overlaid Figure 5.2 (dotted magenta). We observe that the SNR-$p_{\text{out}}(R)$ plane now has two distinct regions, each having a set of parallel lines. The upper-right half has denser lines, while the lower-left half has more sparse and steeper lines. These two regions correspond to the two linear pieces of the diversity-multiplexing tradeoff curve for the $n_R = n_T = 2$ MIMO channel. The boundary is the line $p_{\text{out}} = \varrho^{-1}$, which is the point labeled $r = 1, d = 1$ in the optimal tradeoff curve (compare Figure 5.1).

The slopes and gaps between the curves in Figure 5.4 lead to a concept called *local* diversity-multiplexing tradeoff, which is different from the global scale tradeoff we have defined. If we are operating at a certain $(R, \varrho, p_{\text{out}})$ point, and we increase the SNR, the local tradeoff characterizes the relationship between the incremental increase in rate and reduction of p_{out}. Thus, if we are in the upper-right region of Figure 5.4, and we spend all extra SNR on increasing rate and keep p_{out} constant, we can get 2 extra bits per channel use for every additional 3dB in SNR. If, on the other hand, we spend all the SNR on the reduction of p_{out} and keep the rate constant, we can get 2 orders of magnitude reduction for every additional 10dB in SNR. We

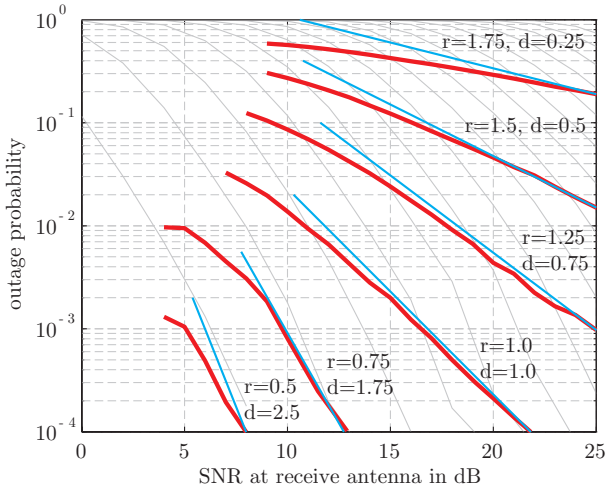

Figure 5.3.: Family of outage probability tradeoff curves $p_{\text{out}}(r \log_2 \varrho)$ as functions of SNR ϱ for various multiplexing gains r for a $n_R = n_T = 2$ MIMO channel (for a SM system).

can also get any linear combination of those two extremes because the lines are parallel. This corresponds to a straight line connecting the two points $(r, d) = (0, 2)$ and $(2, 0)$, which is the lower piece of the global tradeoff $d^*(r)$ from Figure 5.1 extended to $r = 0$. Similar arguments can be given for the lower left region of Figure 5.4, which results in a local tradeoff of a straight line connecting $(r, d) = (0, 4)$ and $(4/3, 0)$. Note that the maximum multiplexing gain of 2 is not achieved. Thus, for the system designer, different segments of the diversity-multiplexing tradeoff curve are important, depending at which SNR level and which target error rate the system operates (see also [43]).

5.2. Tradeoffs of STBCs

After we had a close look at the optimal tradeoff and its implications on system design, we want to investigate how well the already treated systems behave in terms of the diversity-multiplexing tradeoff. In the case of OSTBCs, the problem of the tradeoff curve can be solved analytically [26], whereas in the case of the LD system, only a numerical solution in analogy to Figure 5.3 can be pronounced.

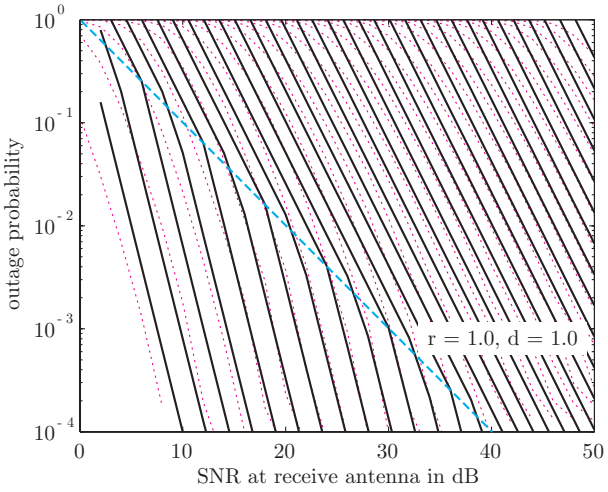

Figure 5.4.: Family of linearized outage probability curves as functions of SNR ϱ at various rates for the $n_T = n_R = 2$ MIMO channel.

5.2.1. Orthogonal STBCs

We are investigating the Alamouti OSTBC for $n_T = 2$ transmit antennas. As described in Section 4.2, the effective input output relation can be written as

$$\mathbf{y}' = \sqrt{\frac{\varrho}{2}}\mathbf{H}_{\text{eff}}\mathbf{s}' + \mathbf{n}',$$

which after MRC reduces to two independent channels with path gain $\sqrt{\varrho/2}\,\|\mathbf{H}\|^2$ respectively. As shown in [26], $\|\mathbf{H}\|^2$ is chi-square distributed with $2n_Tn_R$ degrees of freedom. Furthermore, it is shown that for small ε, $\Pr(\|\mathbf{H}\|^2 \leq \varepsilon) \approx \varepsilon^{n_Tn_R}$. As also mentioned in Section 4.2, conditioned on any realization of the channel matrix $\mathbf{H} = \mathbf{H}_0$, the Alamouti design has a system capacity of $\log(1 + \varrho\,\|\mathbf{H}_0\|^2/2)$. The outage event for this channel at a given rate R may thus be defined as

$$\left\{\mathbf{H}_0 : \log\left[1 + \frac{\varrho}{2}\,\|\mathbf{H}_0\|^2\right] < R\right\}.$$

Using the outage probability as lower bound on the error probability, one may see that (compare [26])

$$\begin{aligned} p_{\text{out}}(R) &= \Pr\left(\log\left[1 + \frac{\varrho}{2}\,\|\mathbf{H}\|^2\right] < R\right) = \Pr\left(1 + \frac{\varrho}{2}\,\|\mathbf{H}\|^2 < \varrho^r\right) \\ &\doteq \Pr\left(\|\mathbf{H}\|^2 \leq \varrho^{-(1-r)^+}\right) \doteq \varrho^{-n_Rn_T(1-r)^+}. \end{aligned}$$

Following the arguments of [26], this defines the tradeoff curve $d_{\text{Alamouti}}(r) = d_{\text{out}}(r) = n_Rn_T(1-r)^+ = n_R2(1-r)^+$.

Figure 5.6 shows the obtained tradeoff curve in comparison to the optimal tradeoff for the $n_T = n_R = 2$ MIMO channel. We can observe that the Alamouti OSTBC system is in general not

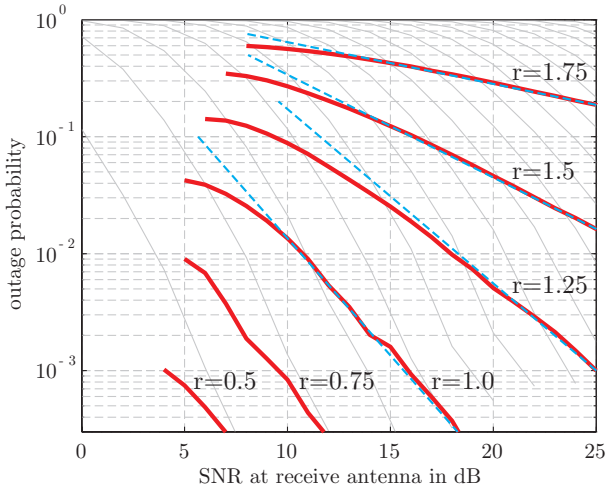

Figure 5.5.: Outage probability curves and limiting slopes of the LD system for a $n_T = n_R = 2$ MIMO channel.

r	1.75	1.5	1.25	1.0
$\tilde{d}$	0.23	0.45	0.74	1.0

Table 5.1.: Measured values of the diversity gain from Figure 5.5 for the optimized LD code of Table 4.1.

optimal. It achieves the maximum diversity gain, but falls below the optimum tradeoff curve for positive values of r. The maximum spatial multiplexing gain of $r_{\text{Alamouti,max}} = 1$ corresponds to the slope of the system capacity curve obtained in Section 4.2.

5.2.2. LD Code

We now investigate the diversity-multiplexing tradeoff of LD codes. Unfortunately, an analytical analysis of the tradeoff is far from trivial. Therefore, we base the evaluation of the tradeoff curve on numerical simulations in analogy to Figure 5.3, where we compute the outage probabilities of the LD system capacity. If drawn as a function of ϱ with $r \log \varrho$ increasing rates, we get an approximate tradeoff curve. Figure 5.5 shows our simulation results. From measuring the slopes of the drawn tangents, we obtain the values, which are pronounced in Table 5.1. The diversity gain values for $r < 1$ seem hard to measure correctly, thus we are not relying on them, but we propose another argument. We know from our previous investigations that the maximum diversity gain achievable by the LD codes in a $n_T = n_R = 2$ MIMO channel is approximately 2. If we round the diversity values obtained by measurement to the values denoted by $\tilde{d}$ in Table 5.1, we obtain a realistic tradeoff curve for the optimized LD system of Table 4.1. The obtained curve is shown in Figure 5.6 in comparison to the tradeoff of the Alamouti OSTBC and the optimal tradeoff. We can see that the tradeoff curve of the pro-

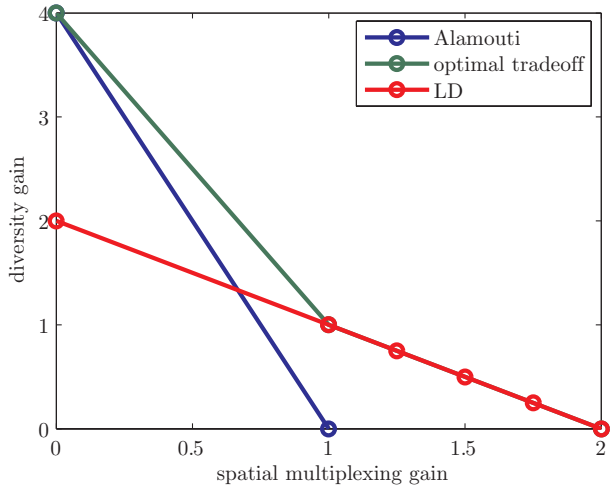

Figure 5.6.: Diversity-multiplexing tradeoff curve for the standard Alamouti design (compare Equation 4.2) and the optimized LD system (compare Table 4.1) for a $n_T = n_R = 2$ MIMO channel.

posed LD system coincides with the lower piece of the optimal tradeoff curve. Nevertheless, the maximum possible diversity gain of 4 is not achieved.

The fact that LD codes are able to achieve the maximum spatial multiplexing gain can be summarized in the so-called theory of *asymptotic-information-lossless* designs (see [45]). Here, a STBC design $\mathbf{S}_\mathrm{X}$ is defined to be an asymptotic-information-lossless (AILL) design for n_R receive antennas, if

$$\lim_{\varrho\to\infty} \frac{C(\varrho)}{C_\mathrm{X}(\varrho)} = 1,$$

where $C_\mathrm{X}(\varrho)$ denotes the system capacity of the design "X" (for example LD).

Obviously, OSTBCs are not AILL, since for $n_T > 2$ or $n_R \geq 2$ the system capacity has a lower slope and thus the above limit tends to infinity. In case of the LD codes however, one can see that they are AILL designs (compare [45]), since although we could not guarantee a system capacity that coincides with the ergodic channel capacity, the slope of the LD system capacity can be shown to be equal to the slope of the ergodic channel capacity. Thus, in the limit of $\varrho \to \infty$, the difference between the ergodic channel capacity and the LD system capacity vanishes, thus showing that LD systems are in fact AILL. In case of the number theory extended LD code from Definition 4.3.3, one may even be able to check that this system is information-lossless (ILL), what implies that the fraction $C(\varrho)/C_{\mathrm{LD,number\ theory}}(\varrho)$ is equal to 1 for all ϱ. This has been shown in [16], by proofing that the basic LD code structure of Equation (4.11) that is extended by $\mathbf{\Phi}$ to fulfill the error performance criteria, is an analytically optimal solution of the maximization problem.

Finally, we want to note that the STBC structure from Definition 4.3.3 (which in fact is a transformed LD system) can achieve the optimal tradeoff curve for all values of r. The

proof would go beyond the scope of this thesis, so we refer the interested reader to [41]. In addition recent, papers concerning the construction of ST systems are merely relying on the reachability of the optimal tradeoff curve instead of optimizing either the diversity or the spatial multiplexing gain. Examples therefore may be [46] or [33].

A. Appendix

A.1. Basic Definitions of Information Theory

The basic definitions given within this section are merely based on [17] and [18].

A.1.1. Entropy

We will first introduce the concept of entropy, which is a measure of uncertainty of a random variable.

Definition A.1.1 (Entropy of a discrete random vector)**.** *Let* $\mathbf{x}$ *be a discrete random vector with alphabet* $\mathcal{X}$ *and probability mass function* $p_{\mathbf{x}}(\boldsymbol{\xi}) = Pr(\mathbf{x} = \boldsymbol{\xi})$*,* $\boldsymbol{\xi} \in \mathcal{X}$*. Then, the entropy* $H(\mathbf{x})$ *of a discrete random vector* $\mathbf{x}$ *is defined as*

$$H(\mathbf{x}) \triangleq -\sum_{\boldsymbol{\xi} \in \mathcal{X}} p_{\mathbf{x}}(\boldsymbol{\xi}) \log p_{\mathbf{x}}(\boldsymbol{\xi}).$$

If the logarithm is chosen to have base 2, the entropy is expressed in bits. If the base of the logarithm is e, then the entropy is measured in *nats*. This disposition keeps valid for all following definitions. If $\mathbf{x}$ is not a discrete, but a continuous vector, we can define the differential entropy:

Definition A.1.2 (Entropy of a continuous random vector (differential entropy))**.** *Let* $\mathbf{x}$ *be a random continuous vector with cumulative distribution function* $F_{\mathbf{x}}(\boldsymbol{\xi}) = Pr(\mathbf{x} \leq \boldsymbol{\xi})$*. If* $\mathbf{x}$ *is continuous, then* $F_{\mathbf{x}}(\boldsymbol{\xi})$ *has to be continuous too. Furthermore let* $f_{\mathbf{x}}(\boldsymbol{\xi}) = \frac{\partial}{\partial \boldsymbol{\xi}} F_{\mathbf{x}}(\boldsymbol{\xi})$ *when the derivative is defined. If* $\int_{-\infty}^{\infty} f_{\mathbf{x}}(\boldsymbol{\xi}) d\boldsymbol{\xi} = 1$*, then* $f_{\mathbf{x}}(\boldsymbol{\xi})$ *is called the probability density function for* $\mathbf{x}$*. The set, where* $f_{\mathbf{x}}(\boldsymbol{\xi}) > 0$ *is called the support set* Ω *of* $\mathbf{x}$*:*

$$f_{\mathbf{x}}(\boldsymbol{\xi}) = \begin{cases} > 0, & \boldsymbol{\xi} \in \Omega, \\ = 0, & \boldsymbol{\xi} \notin \Omega. \end{cases}$$

The differential entropy $h(\mathbf{x})$ *of a continuous random vector* $\mathbf{x}$ *with pdf* $f(\boldsymbol{\xi})$ *is defined as*

$$h(\mathbf{x}) \triangleq -\int_{\Omega} f_{\mathbf{x}}(\boldsymbol{\xi}) \log f_{\mathbf{x}}(\boldsymbol{\xi}) d\boldsymbol{\xi}.$$

After we have defined the entropy of a single random vector, we will now extend the definition to a pair of random vectors. There is nothing really new in this definition because $(\mathbf{x}, \mathbf{y})$ can be considered to be a single random vector with larger size, but for the sake of completeness, we will state this definition too.

Definition A.1.3 (Joint entropy of a pair of discrete random vectors)**.** *Let, in analogy to Definition A.1.1, $\mathcal{X}$ and $\mathcal{Y}$ be the alphabets of $\mathbf{x}$ and $\mathbf{y}$ respectively. Furthermore, let $p_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ be the joint probability mass function (compare e.g. [12]). Then, the joint entropy $H(\mathbf{x},\mathbf{y})$ of a pair of discrete random vectors $(\mathbf{x},\mathbf{y})$ is defined as*

$$H(\mathbf{x},\mathbf{y}) \triangleq -\sum_{\boldsymbol{\xi}\in\mathcal{X}}\sum_{\boldsymbol{\eta}\in\mathcal{Y}} p_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi},\boldsymbol{\eta}) \log p_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi},\boldsymbol{\eta}).$$

And as in the discrete case, we can extend the definition of the differential entropy of a single random vector to several random vectors.

Definition A.1.4 (Joint differential entropy)**.** *Let $f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ be the joint probability density function (compare, e.g. [12]) of the pair of continuous random vectors $(\mathbf{x},\mathbf{y})$. Furthermore, let (in analogy to Definition A.1.2) $\Omega_{\boldsymbol{\xi},\boldsymbol{\eta}}$ be the support set of the joint probability density function:*

$$f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi},\boldsymbol{\eta}) = \begin{cases} > 0, & (\boldsymbol{\xi},\boldsymbol{\eta}) \in \Omega_{\boldsymbol{\xi},\boldsymbol{\eta}}, \\ = 0, & (\boldsymbol{\xi},\boldsymbol{\eta}) \notin \Omega_{\boldsymbol{\xi},\boldsymbol{\eta}}. \end{cases}$$

Then the differential entropy of this vector pair is defined as

$$h(\mathbf{x},\mathbf{y}) \triangleq -\int_{\Omega_{\boldsymbol{\xi},\boldsymbol{\eta}}} f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi},\boldsymbol{\eta}) \log f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi},\boldsymbol{\eta}) d\boldsymbol{\xi} d\boldsymbol{\eta}.$$

We also define the conditional entropy of a random vector given another random vector as the expected value of the entropies of the conditional distributions, averaged over the conditioning random variable.

Definition A.1.5 (Conditional entropy of discrete random vectors)**.** *If, in analogy to Definition A.1.3, the joint probability mass function of $(\mathbf{x},\mathbf{y})$ is given by $p_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi},\boldsymbol{\eta})$, with $\mathcal{X}$ and $\mathcal{Y}$ defining the alphabets of $\mathbf{x}$ and $\mathbf{y}$, respectively, then the conditional entropy $H(\mathbf{y}|\mathbf{x})$ is defined as*

$$H(\mathbf{y}|\mathbf{x}) \triangleq \sum_{\boldsymbol{\xi}\in\mathcal{X}} p_{\mathbf{x}}(\boldsymbol{\xi}) H(\mathbf{y}|\mathbf{x}=\boldsymbol{\xi}) = -\sum_{\boldsymbol{\xi}\in\mathcal{X}}\sum_{\boldsymbol{\eta}\in\mathcal{Y}} p_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi},\boldsymbol{\eta}) \log p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{\eta}|\boldsymbol{\xi}).$$

This definition includes the description of a slightly modified version of the conditional entropy of discrete random vectors. The term $H(\mathbf{y}|\mathbf{x}=\boldsymbol{\xi})$ denotes the entropy of $\mathbf{y}$ given that $\mathbf{x}=\boldsymbol{\xi}$ (i.e., $\mathbf{x}$ was already observed to be $\boldsymbol{\xi}$ [18]). This conditional entropy with already observed condition vector may be written as

$$H(\mathbf{y}|\mathbf{x}=\boldsymbol{\xi}) \triangleq -\sum_{\boldsymbol{\eta}\in\mathcal{Y}} p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{\eta}|\boldsymbol{\xi}) \log p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{\eta}|\boldsymbol{\xi}).$$

The Definition A.1.5 can of course be extended (in an equivalent manner as already twofold done) to the continuous case.

Definition A.1.6 (Conditional differential entropy of continuous random vectors)**.** *If $(\mathbf{x},\mathbf{y})$ has a joint density function $f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi},\boldsymbol{\eta})$ with a support set, equally defined as in Definition A.1.4, we can define the conditional differential entropy $h(\mathbf{y}|\mathbf{x})$ as*

$$h(\mathbf{y}|\mathbf{x}) \triangleq -\int_{\Omega_{\boldsymbol{\xi},\boldsymbol{\eta}}} f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi},\boldsymbol{\eta}) \log f_{\mathbf{y}|\mathbf{x}}(\boldsymbol{\eta}|\boldsymbol{\xi}) d\boldsymbol{\xi} d\boldsymbol{\eta}.$$

A.1.2. Mutual Information

The entropy of a random variable is a measure of the uncertainty of a random variable. It is a measure of the amount of information required on average to describe the random variable. Now we want to introduce a related concept: Mutual information.

Mutual information is a measure of the amount of information that one random variable contains about another random variable. It is the reduction of uncertainty of one random variable due to the knowledge of the other. Without going further into details on the interpretation of the mutual information, we specify its definition (compare also [47]):

Definition A.1.7 (Mutual information of discrete random variables)**.** *Consider two random variables* $\mathbf{x} \in \mathcal{X}$ *and* $\mathbf{y} \in \mathcal{Y}$*, with* $\mathcal{X}$ *and* $\mathcal{Y}$ *denoting the alphabet of* $\mathbf{x}$ *and* $\mathbf{y}$*, respectively. If* $p_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \boldsymbol{\eta})$ *denotes the joint probability mass function and* $p_{\mathbf{x}}(\boldsymbol{\xi})$ *and* $p_{\mathbf{y}}(\boldsymbol{\eta})$ *denoting the marginal probability mass functions of* $\mathbf{x}$ *and* $\mathbf{y}$*, respectively, then the mutual information* $I(\mathbf{x};\mathbf{y})$ *is defined as*

$$I(\mathbf{x};\mathbf{y}) \triangleq \sum_{\boldsymbol{\xi}\in\mathcal{X}} \sum_{\boldsymbol{\eta}\in\mathcal{Y}} p_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \boldsymbol{\eta}) \log \frac{p_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \boldsymbol{\eta})}{p_{\mathbf{x}}(\boldsymbol{\xi}) p_{\mathbf{y}}(\boldsymbol{\eta})}.$$

In the case of continuous random vectors, we can define the mutual information as:

Definition A.1.8 (Mutual information of continuous random variables)**.** *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *denote two random vectors with joint pdf* $f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \boldsymbol{\eta})$*. Furthermore, let* $\Omega_{\boldsymbol{\xi},\boldsymbol{\eta}}$ *be the support set of the joint pdf, as already introduced in Definition A.1.4 and let* $f_{\mathbf{x}}(\boldsymbol{\xi})$ *and* $f_{\mathbf{y}}(\boldsymbol{\eta})$ *denote the marginal pdfs of* $\mathbf{x}$ *and* $\mathbf{y}$*, respectively. Then the mutual information is defined as*

$$I(\mathbf{x};\mathbf{y}) \triangleq \int_{\Omega_{\boldsymbol{\xi},\boldsymbol{\eta}}} f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \boldsymbol{\eta}) \log \frac{f_{\mathbf{x},\mathbf{y}}(\boldsymbol{\xi}, \boldsymbol{\eta})}{f_{\mathbf{x}}(\boldsymbol{\xi}) f_{\mathbf{y}}(\boldsymbol{\eta})} d\boldsymbol{\xi} d\boldsymbol{\eta}.$$

With these definitions in mind, we can rewrite the mutual information in terms of entropies (for an intuitive interpretation see for example [12]). These relations are very important and often used in information theoretic analysis. For the case of discrete random vectors, we can write

$$I(\mathbf{x};\mathbf{y}) = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}), \tag{A.1}$$

where the second equation follows directly be using the symmetry property $I(\mathbf{x};\mathbf{y}) = I(\mathbf{y};\mathbf{x})$ of the mutual information. Thus $\mathbf{x}$ says as much about $\mathbf{y}$ as $\mathbf{y}$ says about $\mathbf{x}$.

With these relations, it is easy to state the last definition in this subsection, the conditional mutual information:

Definition A.1.9 (Conditional mutual information)**.** *Consider the three random vectors* $\mathbf{x}$*,* $\mathbf{y}$ *and* $\mathbf{z}$ *drawn either from a discrete or a continuous alphabet. The conditional mutual information in terms of entropy is defined as*

$$I(\mathbf{x};\mathbf{y}|\mathbf{z}) \triangleq H(\mathbf{x}|\mathbf{z}) - H(\mathbf{x}|\mathbf{y},\mathbf{z}). \tag{A.2}$$

In the case of continuous random vectors we have to exchange the entropy $H(\cdot)$ *with the differential entropy* $h(\cdot)$*.*

A.1.3. Chain Rules for Entropy and Mutual Information

We now want to restate the chain rules for entropy and mutual information. These relations show a possibility for expressing the entropy or the mutual information respectively. We first state the chain rule for the entropy of discrete random variables.

Definition A.1.10 (Chain rule for entropy of discrete random variables)**.** *Let* $\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_n$ *be drawn according to the joint probability mass function* $p_{\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_n)$*. Then the chain rule for entropy is given by*

$$H(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \sum_{i=1}^{n} H(\mathbf{x}_i | \mathbf{x}_{i-1}, \dots, \mathbf{x}_1).$$

The chain rule for entropy of continuous random variables can be stated fully equivalent. We only have to replace $H(\cdot)$ with $h(\cdot)$.

A similar chain rule can also be stated for the mutual information. Because the notation of the mutual information does not distinguish between discrete and continuous random variables, we again state the chain rule only once, since it applies in both cases.

Definition A.1.11 (Chain rule for mutual information)**.** *Consider a set of random variables* $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ *and* $\mathbf{y}$*. Then the mutual information can be written as*

$$I(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n; \mathbf{y}) = \sum_{i=1}^{n} I(\mathbf{x}_i; \mathbf{y} | \mathbf{x}_{i-1}, \dots, \mathbf{x}_1).$$

A.1.4. Relations of Entropy and Mutual Information

By using the chain rule for entropy from Definition A.1.10, we can derive another expression for the mutual information. It turns out that we can write

$$I(\mathbf{x}; \mathbf{y}) = H(\mathbf{x}) + H(\mathbf{y}) - H(\mathbf{x}, \mathbf{y}). \tag{A.3}$$

Another very useful relation in the context of the analysis of MIMO systems is the following:

Theorem A.1.12 (Entropy of the sum of two random variables)**.** *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be random variables that are drawn from alphabet* $\mathcal{X}$ *and* $\mathcal{Y}$ *respectively. Furthermore, let* $\mathbf{z} = \mathbf{x} + \mathbf{y}$*. Then the following holds*

$$H(\mathbf{z}|\mathbf{x}) = H(\mathbf{y}|\mathbf{x}).$$

Proof. First, let us indicate the probability mass functions of the random variables by $p_{\mathbf{x}}(\boldsymbol{\xi})$, $p_{\mathbf{y}}(\boldsymbol{\eta})$ and $p_{\mathbf{z}}(\boldsymbol{\zeta})$. By using the definition of the conditional entropy A.1.5, we write

$$H(\mathbf{z}|\mathbf{x}) = -\sum_{\boldsymbol{\xi} \in \mathcal{X}} p_{\mathbf{x}}(\boldsymbol{\xi}) \sum_{\boldsymbol{\zeta} \in \mathcal{Z}} p_{\mathbf{z}|\mathbf{x}}(\boldsymbol{\zeta}|\boldsymbol{\xi}) \log p_{\mathbf{z}|\mathbf{x}}(\boldsymbol{\zeta}|\boldsymbol{\xi}).$$

Now, we take a closer look at the conditional pmf

$$p_{\mathbf{z}|\mathbf{x}}(\boldsymbol{\zeta}|\boldsymbol{\xi}) = p_{\mathbf{x}+\mathbf{y}|\mathbf{x}}(\boldsymbol{\zeta}|\boldsymbol{\xi}),$$

where $\mathbf{x}$ can be treated as a deterministic value. Thus

$$p_{\mathbf{z}|\mathbf{x}}(\boldsymbol{\zeta}|\boldsymbol{\xi}) = p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{\zeta} - \mathbf{x}|\boldsymbol{\xi}),$$

and

$$H(\mathbf{z}|\mathbf{x}) = -\sum_{\boldsymbol{\xi}\in\mathcal{X}} p_{\mathbf{x}}(\boldsymbol{\xi}) \sum_{\boldsymbol{\zeta}\in\mathcal{Z}} p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{\zeta} - \mathbf{x}|\boldsymbol{\xi}) \log p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{\zeta} - \mathbf{x}|\boldsymbol{\xi}).$$

Within this equation, we can identify the following equality:

$$\sum_{\boldsymbol{\zeta}\in\mathcal{Z}} p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{\zeta} - \mathbf{x}|\boldsymbol{\xi}) \log p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{\zeta} - \mathbf{x}|\boldsymbol{\xi}) = H(\mathbf{y}|\mathbf{x} = \boldsymbol{\xi}),$$

which immediately results in

$$H(\mathbf{z}|\mathbf{x}) = -\sum_{\boldsymbol{\xi}\in\mathcal{X}} p_{\mathbf{x}}(\boldsymbol{\xi}) H(\mathbf{y}|\mathbf{x} = \boldsymbol{\xi}) = H(\mathbf{y}|\mathbf{x}).$$

This concludes the proof. □

A.1.5. Definitions Needed for Shannon's Second Theorem

We have to introduce the definition of a (M, n) code and a basic definition of the probability of error. To start with, we state the code definition:

Definition A.1.13 ((M, n) code)**.** *A (M, n) code for the channel $(\mathcal{X}, p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{\eta}|\boldsymbol{\xi}), \mathcal{Y})$ consists of the following*

1. *An index set $\{1, 2, \ldots, M\}$.*
2. *An encoding function $\mathbf{x}^n : \{1, 2, \ldots, M\} \to \mathcal{X}^n$, yielding codewords $\mathbf{x}^n(1), \mathbf{x}^n(2), \ldots, \mathbf{x}^n(M)$. Here $(\cdot)^n$ denotes that the channel is used n successive time instances for transmission (and thus, coding is performed over n time instances). The set of codewords is called the codebook.*
3. *A decoding function $g : \mathcal{Y}^n \to \{1, 2, \ldots, M\}$, which is a deterministic rule which assigns a guess to each possible received vector.*

Next, we repeat a definition of the probability of error:

Definition A.1.14 (Probability of error)**.** *Let $\varepsilon_i \triangleq Pr(g(\mathbf{y}^n) \neq i|\mathbf{x}^n = \mathbf{x}^n(i))$ be the conditional probability of error given that index i was sent.*

And in addition to the preceding definition, we state the maximal probability of error as:

Definition A.1.15 (Maximal probability of error)**.** *The maximal probability of error ε for a (M, n) code is defined as $\varepsilon \triangleq \max_{i\in\{1,2,\ldots,M\}} \varepsilon_i$.*

A.1.6. Fano's Inequality

Suppose we wish to estimate a random variable $\mathbf{x}$ with a distribution $p_{\mathbf{x}}(\boldsymbol{\xi})$ by using an observation of a random variable $\mathbf{y}$ which is related to $\mathbf{y}$ by the conditional distribution $p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{\eta}|\boldsymbol{\xi})$. From $\mathbf{y}$, we calculate a function $g(\mathbf{y}) = \hat{\mathbf{x}}$, which is an estimate of $\mathbf{x}$. We now wish to bound the probability that $\hat{\mathbf{x}} \neq \mathbf{x}$. The answer to this question is Fano's inequality.

Theorem A.1.16 (Fano's inequality)**.** *Let $P_e = \Pr(\hat{\mathbf{x}} \neq \mathbf{x})$, and let $\mathcal{X}$ denote the alphabet of $\mathbf{x}$. Then Fano's inequality is given as*

$$H(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(\mathbf{x}|\mathbf{y}).$$

This inequality can be weakened to

$$1 + P_e \log |\mathcal{X}| \geq H(\mathbf{x}|\mathbf{y}).$$

Proof. Is given in [17]. □

A.2. Further Details on some Evaluations

A.2.1. Proof of Theorem 4.2.2

Considering the unitary property of OSTBCs (Definition 4.2.1), we have

$$\begin{aligned}
\mathbf{S}\mathbf{S}^H &= \sum_{n=1}^{n_S}\sum_{p=1}^{n_S} \left(\operatorname{Re}\{s_n\}\,\mathbf{A}_n + j\operatorname{Im}\{s_n\}\,\mathbf{B}_n\right)\left(\operatorname{Re}\{s_p\}\,\mathbf{A}_p + j\operatorname{Im}\{s_p\}\,\mathbf{B}_p\right)^H \\
&= \sum_{n=1}^{n_S}\left(\operatorname{Re}\{s_n\}^2\,\mathbf{A}_n\mathbf{A}_n^H + \operatorname{Im}\{s_n\}^2\,\mathbf{B}_n\mathbf{B}_n^H\right) \\
&+ \sum_{n=1}^{n_S}\sum_{p=1,p>n}^{n_S}\left(\operatorname{Re}\{s_n\}\operatorname{Re}\{s_p\}\left(\mathbf{A}_n\mathbf{A}_p^H + \mathbf{A}_p\mathbf{A}_n^H\right) + \operatorname{Im}\{s_n\}\operatorname{Im}\{s_p\}\left(\mathbf{B}_n\mathbf{B}_p^H + \mathbf{B}_p\mathbf{B}_n^H\right)\right) \\
&+ j\sum_{n=1}^{n_S}\sum_{p=1}^{n_S}\operatorname{Im}\{s_n\}\operatorname{Re}\{s_p\}\left(\mathbf{B}_n\mathbf{A}_p^H - \mathbf{A}_p\mathbf{B}_n^H\right).
\end{aligned}$$

With this equation, one can easily see that Theorem 4.2.2 holds, whenever Equation (4.3) is satisfied, which concludes the proof.

A.2.2. OSTBC ML Detection Decoupling

We want to show that the ML metric $\|\mathbf{Y} - \mathbf{H}\mathbf{S}\|^2$ can be decoupled in a way that the ML decision rule can be performed in a linear way and each symbol s_n can be decided independently of the other symbols s_p for an arbitrary $p \neq n$. The derivation is based on the results stated

in [35], which was the original work performed in the field of OSTBC, but also uses [8], as well as [11] and [23].

First, we start by using an extension of a norm relation, well known in literature

$$\|\mathbf{Y}-\mathbf{HS}\|^2 = \|\mathbf{Y}\|^2 + \|\mathbf{HS}\|^2 - 2\,\mathrm{Re}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HS}\}\},$$

where, by use of Definition 4.2.1, one easily sees that $\|\mathbf{HS}\|^2 = \|\mathbf{s}\|^2\,\|\mathbf{H}\|^2$ (see also the proof of Theorem 4.2.4). The vector of transmitted symbols s_n, $n = 1,\ldots,n_S$ is denoted by $\mathbf{s}$. Through usage of Definition 4.1.2, we obtain

$$\begin{aligned}
2\,\mathrm{Re}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HS}\}\} &= 2\,\mathrm{Re}\left\{\mathrm{tr}\left\{\mathbf{Y}^H\mathbf{H}\sum_{n=1}^{n_S}(\,\mathrm{Re}\{s_n\}\,\mathbf{A}_n + j\,\mathrm{Im}\{s_n\}\,\mathbf{B}_n)\right\}\right\}\\
&\overset{1}{=} 2\,\mathrm{tr}\left\{\mathrm{Re}\left\{\mathbf{Y}^H\mathbf{H}\sum_{n=1}^{n_S}(\,\mathrm{Re}\{s_n\}\,\mathbf{A}_n + j\,\mathrm{Im}\{s_n\}\,\mathbf{B}_n)\right\}\right\}\\
&= 2\,\mathrm{tr}\left\{\mathrm{Re}\left\{\sum_{n=1}^{n_S}\mathrm{Re}\{s_n\}\,\mathbf{Y}^H\mathbf{HA}_n + j\sum_{i=1}^{n_S}\mathrm{Im}\{s_n\}\,\mathbf{Y}^H\mathbf{HB}_n\right\}\right\}\\
&\overset{2}{=} 2\,\mathrm{tr}\left\{\sum_{n=1}^{n_S}\mathrm{Re}\{s_n\}\,\mathrm{Re}\{\mathbf{Y}^H\mathbf{HA}_n\} - \sum_{n=1}^{n_S}\mathrm{Im}\{s_n\}\,\mathrm{Im}\{\mathbf{Y}^H\mathbf{HB}_n\}\right\}\\
&\overset{3}{=} 2\,\mathrm{Re}\left\{\mathrm{tr}\left\{\sum_{n=1}^{n_S}\mathrm{Re}\{s_n\}\,\mathbf{Y}^H\mathbf{HA}_n\right\}\right\} - 2\,\mathrm{Im}\left\{\mathrm{tr}\left\{\sum_{n=1}^{n_S}\mathrm{Im}\{s_n\}\,\mathbf{Y}^H\mathbf{HB}_n\right\}\right\}\\
&\overset{4}{=} 2\sum_{n=1}^{n_S}\mathrm{Re}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HA}_n\}\}\,\mathrm{Re}\{s_n\} - 2\sum_{n=1}^{n_S}\mathrm{Im}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HB}_n\}\}\,\mathrm{Im}\{s_n\},
\end{aligned}$$

where equality 1 holds, because $\mathrm{Re}\{\cdot\}$ and $\mathrm{tr}\,\{\cdot\}$ commute, equality 2 uses the fact that $\mathrm{Re}\{ja\} = -\,\mathrm{Im}\{a\}$ for an arbitrary complex number (or matrix) a and equality 3 and 4 use the fact that the trace operation is linear.

Now, we are able to reformulate the ML metric as

$$\begin{aligned}
&\|\mathbf{Y}-\mathbf{HS}\|^2\\
&= \|\mathbf{Y}\|^2 - 2\sum_{n=1}^{n_S}\mathrm{Re}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HA}_n\}\}\,\mathrm{Re}\{s_n\} + 2\sum_{n=1}^{n_S}\mathrm{Im}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HB}_n\}\}\,\mathrm{Im}\{s_n\} + \|\mathbf{H}\|^2\,\|\mathbf{s}\|^2\\
&= \sum_{n=1}^{n_S}\left(-2\,\mathrm{Re}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HA}_n\}\}\,\mathrm{Re}\{s_n\} + 2\,\mathrm{Im}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HB}_n\}\}\,\mathrm{Im}\{s_n\} + |s_n|^2\,\|\mathbf{H}\|^2\right) + \mathrm{const.}\\
&= \|\mathbf{H}\|^2\sum_{n=1}^{n_S}\left(-2\frac{\mathrm{Re}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HA}_n\}\}}{\|\mathbf{H}\|^2}\,\mathrm{Re}\{s_n\} + 2\frac{\mathrm{Im}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HB}_n\}\}}{\|\mathbf{H}\|^2}\,\mathrm{Im}\{s_n\} + |s_n|^2\right) + \mathrm{const.},
\end{aligned}$$

and by amending the complete square in the brace by

$$\frac{\mathrm{Re}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HA}_n\}\}}{\|\mathbf{H}\|^2} + \frac{\mathrm{Im}\{\mathrm{tr}\,\{\mathbf{Y}^H\mathbf{HB}_n\}\}}{\|\mathbf{H}\|^2},$$

which does not depend on s_n and can therefore be accumulated with the const. term, we can write

$$\|\mathbf{Y}-\mathbf{HS}\|^2 = \|\mathbf{H}\|^2 \sum_{n=1}^{n_S} \left| s_n - \frac{\mathrm{Re}\{\mathrm{tr}\{\mathbf{Y}^H\mathbf{H}\mathbf{A}_n\}\} - j\,\mathrm{Im}\{\mathrm{tr}\{\mathbf{Y}^H\mathbf{H}\mathbf{B}_n\}\}}{\|\mathbf{H}\|^2} \right|^2 + \text{const.}$$

A.2.3. Effective Channels for Alamouti STC ($n_T = 2$)

For the sake of completeness, we state that in case of using the Alamouti OSTBC design of Equation (4.2), an equivalent effective channel for $n_R = 1$ receive antennas may be given by

$$\mathbf{H} = [h_1, h_2] \rightarrow \mathbf{H}_{\text{eff}} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix},$$

where we used an equivalent MIMO transmission relation $\mathbf{y} = \sqrt{\varrho/n_T}\mathbf{H}_{\text{eff}}\mathbf{s} + \mathbf{n}$. For $n_R = 1$ we have, $\mathbf{s} = [s_1, s_2]^T$ and $\mathbf{n} = [n_1, n_2^*]^T$. In the case of $n_R = 2$ receive antennas, the effective channel may be written as

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \end{bmatrix} \rightarrow \mathbf{H}_{\text{eff}} = \begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \\ h_{1,2}^* & -h_{1,1}^* \\ h_{2,2}^* & -h_{2,1}^* \end{bmatrix},$$

where in contrast to the $n_R = 1$ case, we set $\mathbf{n} = [n_{1,1}, n_{2,1}, n_{1,2}^*, n_{2,2}^*]^T$.

A.2.4. Proof of Theorem 4.3.2

Let us define $\mathbf{H}'$ to be

$$\mathbf{H}' \triangleq \begin{bmatrix} \mathbf{H}_R \\ \mathbf{H}_I \end{bmatrix}.$$

Then, the system capacity of the LD codes (Theorem 4.3.1), which is our goal function, can be rearranged to

$$\begin{aligned} C = \frac{1}{2L}\mathbb{E}\Bigg\{ \log\det\Bigg(& \mathbf{I}_{2n_RL} \\ & + \frac{\varrho}{n_T}\sum_{n=1}^{n_S} (\mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_n)\,\mathrm{vec}(\mathbf{H}')\,\mathrm{vec}(\mathbf{H}')^T\,(\mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_n)^T + (\boldsymbol{\mathcal{B}}_n \leftarrow \boldsymbol{\mathcal{A}}_n)\Bigg)\Bigg\}, \end{aligned} \tag{A.4}$$

with $(\boldsymbol{\mathcal{B}}_n \leftarrow \boldsymbol{\mathcal{A}}_n)$ denoting the first term of the sum with $\boldsymbol{\mathcal{A}}_n$ replaced by $\boldsymbol{\mathcal{B}}_n$. To compute the gradient, we state the definition (i.e. for $\mathbf{A}_{R,n}$) of the differential quotient

$$\left[\frac{\partial C(\mathbf{A}_{R,n})}{\partial \mathbf{A}_{R,n}}\right]_{i,j} = \lim_{\delta\to 0} \frac{C(\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T) - C(\mathbf{A}_{R,n})}{\delta}, \tag{A.5}$$

with $\xi_i \in \mathbb{Z}^L$ and $\eta_j \in \mathbb{Z}^{n_T}$ denoting vectors filled with zeros except of position i, respectively j, where it is set to one. Furthermore, we define

$$\mathbf{Z} \triangleq \mathbf{I}_{2n_RL} + \frac{\varrho}{n_T} \sum_{n=1}^{n_S} \left(\mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_n\right) \mathrm{vec}(\mathbf{H}') \, \mathrm{vec}(\mathbf{H}')^T \left(\mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_n\right)^T + \left(\boldsymbol{\mathcal{B}}_n \leftarrow \boldsymbol{\mathcal{A}}_n\right),$$

which is a function of $\mathbf{A}_{R,n}$, $\mathbf{A}_{I,n}$, $\mathbf{B}_{R,n}$ and $\mathbf{B}_{I,n}$. If we exchange $\mathbf{A}_{R,n}$ by $\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T$, we simply denote this by

$$\begin{aligned}
&\mathbf{Z}(\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T) = \mathbf{I}_{2n_RL} \\
&+ \frac{\varrho}{n_T} \sum_{n=1}^{n_S} \mathbf{I}_{n_R} \otimes \left(\boldsymbol{\mathcal{A}}_n + \mathbf{I}_2 \otimes (\delta\xi_i\eta_j^T)\right) \mathrm{vec}(\mathbf{H}') \, \mathrm{vec}(\mathbf{H}')^T \left[\mathbf{I}_{n_R} \otimes \left(\boldsymbol{\mathcal{A}}_n + \mathbf{I}_2 \otimes (\delta\xi_i\eta_j^T)\right)\right]^T \\
&+ \left(\boldsymbol{\mathcal{B}}_n \leftarrow \boldsymbol{\mathcal{A}}_n\right)_{\mathrm{old}},
\end{aligned}$$

where the subindex old in $\left(\boldsymbol{\mathcal{B}}_n \leftarrow \boldsymbol{\mathcal{A}}_n\right)_{\mathrm{old}}$ denotes that the terms with replaced $\boldsymbol{\mathcal{A}}_n$ by $\boldsymbol{\mathcal{B}}_n$ are still the same as in (A.4). Straightforward manipulation with the usage of the associative property of the Kronecker product (i.e., $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C}$) leads to

$$\begin{aligned}
\mathbf{Z}(\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T) &= \mathbf{I}_{2n_RL} + \frac{\varrho}{n_T} \sum_{n'=1, n' \neq n}^{n_S} \left(\mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_{n'}\right) \mathrm{vec}(\mathbf{H}') \, \mathrm{vec}(\mathbf{H}')^T \left(\mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_{n'}\right)^T \\
&+ \frac{\varrho}{n_T} \mathbf{I}_{n_R} \otimes \left(\boldsymbol{\mathcal{A}}_n + \mathbf{I}_2 \otimes (\delta\xi_i\eta_j^T)\right) \mathrm{vec}(\mathbf{H}') \, \mathrm{vec}(\mathbf{H}')^T \left[\mathbf{I}_{n_R} \otimes \left(\boldsymbol{\mathcal{A}}_n + \mathbf{I}_2 \otimes (\delta\xi_i\eta_j^T)\right)\right]^T \\
&+ \left(\boldsymbol{\mathcal{B}}_n \leftarrow \boldsymbol{\mathcal{A}}_n\right)_{\mathrm{old}}.
\end{aligned} \tag{A.6}$$

Now we define some matrices to simplify the notation. The middle term in the preceding formula may be written as

$$\begin{aligned}
&\frac{\varrho}{n_T} \left[\mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_n + \mathbf{I}_{n_R} \otimes \left(\mathbf{I}_2 \otimes \left(\delta\xi_i\eta_j^T\right)\right)\right] \mathrm{vec}(\mathbf{H}') \, \mathrm{vec}(\mathbf{H}')^T \left[\mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_n + \mathbf{I}_{n_R} \otimes \left(\mathbf{I}_2 \otimes \left(\delta\xi_i\eta_j^T\right)\right)\right]^T \\
&= \frac{\varrho}{n_T} \left[\mathbf{X}_1 + \mathbf{X}_2\right] \mathbf{Y} \left[\mathbf{X}_1 + \mathbf{X}_2\right]^T,
\end{aligned}$$

with $\mathbf{X}_1 \triangleq \mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_n$, $\mathbf{X}_2 \triangleq \mathbf{I}_{n_R} \otimes \left(\mathbf{I}_2 \otimes \left(\delta\xi_i\eta_j^T\right)\right)$ and $\mathbf{Y} \triangleq \mathrm{vec}(\mathbf{H}') \, \mathrm{vec}(\mathbf{H}')^T$. Again, straightforward algebra leads to

$$\begin{aligned}
&\frac{\varrho}{n_T} \left[\mathbf{X}_1\mathbf{Y} + \mathbf{X}_2\mathbf{Y}\right] \left[\mathbf{X}_1 + \mathbf{X}_2\right]^T = \frac{\varrho}{n_T} \left[\mathbf{X}_1\mathbf{Y} + \mathbf{X}_2\mathbf{Y}\right] \left[\mathbf{X}_1^T + \mathbf{X}_2^T\right] \\
&= \frac{\varrho}{n_T} \left[\mathbf{X}_1\mathbf{Y}\mathbf{X}_1^T + \mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T + \mathbf{X}_1\mathbf{Y}\mathbf{X}_2^T + \mathbf{X}_2\mathbf{Y}\mathbf{X}_2^T\right].
\end{aligned}$$

Now because $\mathbf{X}_1\mathbf{Y}\mathbf{X}_1^T = \left(\mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_n\right) \mathrm{vec}(\mathbf{H}') \, \mathrm{vec}(\mathbf{H}')^T \left(\mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_n\right)^T$ is exactly the missing part in the sum of Equation (A.6), the problem reduces to

$$\mathbf{Z}(\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T) = \mathbf{Z} + \frac{\varrho}{n_T} \left[\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T + \mathbf{X}_1\mathbf{Y}\mathbf{X}_2^T + \mathbf{X}_2\mathbf{Y}\mathbf{X}_2^T\right].$$

Here, it is interesting to see that $\mathbf{X}_1\mathbf{Y}\mathbf{X}_2^T = \left(\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T\right)^T$, which is a direct consequence of the symmetry of $\mathbf{Y} = \mathbf{Y}^T$. Now we can express the nominator $C(\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T) - C(\mathbf{A}_{R,n})$ of

Equation (A.5) by

$$
\begin{aligned}
&C(\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T) - C(\mathbf{A}_{R,n}) \\
&= \frac{1}{2L}\mathbb{E}\left\{\log\det\left(\mathbf{Z} + \frac{\varrho}{n_T}\left[\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T + \left(\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T\right)^T + \mathbf{X}_2\mathbf{Y}\mathbf{X}_2^T\right]\right)\right\} - \frac{1}{2L}\mathbb{E}\{\mathbf{Z}\}\,.
\end{aligned}
$$

The linearity of the expectation operator together with the identity $\log\det(\cdot) = \operatorname{tr}\log(\cdot)$ allows us to state

$$
\begin{aligned}
&C(\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T) - C(\mathbf{A}_{R,n}) \\
&= \frac{1}{2L}\mathbb{E}\left\{\operatorname{tr}\log\left(\mathbf{Z} + \frac{\varrho}{n_T}\left[\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T + \left(\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T\right)^T + \mathbf{X}_2\mathbf{Y}\mathbf{X}_2^T\right]\right) - \operatorname{tr}\log\left(\mathbf{Z}\right)\right\},
\end{aligned}
$$

where $\log(\cdot)$ denotes the generalized matrix logarithm. The commutativity of the trace operator $\operatorname{tr}(\mathbf{A}+\mathbf{B}) = \operatorname{tr}\mathbf{A} + \operatorname{tr}\mathbf{B}$ results in

$$
\begin{aligned}
&C(\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T) - C(\mathbf{A}_{R,n}) \\
&= \frac{1}{2L}\mathbb{E}\left\{\operatorname{tr}\left[\log\left(\mathbf{Z} + \frac{\varrho}{n_T}\left[\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T + \left(\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T\right)^T + \mathbf{X}_2\mathbf{Y}\mathbf{X}_2^T\right]\right) - \log\left(\mathbf{Z}\right)\right]\right\} \\
&= \frac{1}{2L}\mathbb{E}\left\{\operatorname{tr}\left[\log\left(\mathbf{Z} + \frac{\varrho}{n_T}\left[\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T + \left(\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T\right)^T + \mathbf{X}_2\mathbf{Y}\mathbf{X}_2^T\right]\right)\mathbf{Z}^{-1}\right]\right\} \\
&= \frac{1}{2L}\mathbb{E}\left\{\operatorname{tr}\log\left[\mathbf{I}_{2n_RL} + \frac{\varrho}{n_T}\left[\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T + \left(\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T\right)^T + \mathbf{X}_2\mathbf{Y}\mathbf{X}_2^T\right]\mathbf{Z}^{-1}\right]\right\}.
\end{aligned}
$$

Now we focus again on the terms $\mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T$ and $\mathbf{X}_2\mathbf{Y}\mathbf{X}_2^T$. Because of the linearity of the matrix multiplications, we can write these terms as

$$
\begin{aligned}
\delta\cdot\mathbf{M}_1 \triangleq \mathbf{X}_2\mathbf{Y}\mathbf{X}_1^T &= \delta\cdot\mathbf{I}_{n_R}\otimes\left(\mathbf{I}_2\otimes\left(\xi_i\eta_j^T\right)\right)\operatorname{vec}(\mathbf{H}')\operatorname{vec}(\mathbf{H}')^T\left[\mathbf{I}_{n_R}\otimes\boldsymbol{\mathcal{A}}_n\right]^T \\
\delta^2\cdot\mathbf{M}_2 \triangleq \mathbf{X}_2\mathbf{Y}\mathbf{X}_2^T &= \delta^2\cdot\left[\mathbf{I}_{n_R}\otimes\left(\mathbf{I}_2\otimes\left(\xi_i\eta_j^T\right)\right)\right]\operatorname{vec}(\mathbf{H}')\operatorname{vec}(\mathbf{H}')^T\left[\mathbf{I}_{n_R}\otimes\left(\mathbf{I}_2\otimes\left(\xi_i\eta_j^T\right)\right)\right]^T,
\end{aligned}
$$

which leads us to

$$
C(\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T) - C(\mathbf{A}_{R,n}) = \frac{1}{2L}\mathbb{E}\left\{\operatorname{tr}\log\left[\mathbf{I}_{2n_RL} + \frac{\varrho\delta}{n_T}\left[\mathbf{M}_1 + \mathbf{M}_1^T + \delta\mathbf{M}_2\right]\mathbf{Z}^{-1}\right]\right\}.
$$

Finally, using the identity $\log(\mathbf{I}+\mathbf{A}) = \mathbf{A} - \frac{1}{2}\mathbf{A}^2 + \cdots$, together with the linearity of the trace operator, i.e. $\operatorname{tr}(c\mathbf{A}) = c\operatorname{tr}\mathbf{A}$, we can write

$$
\begin{aligned}
&C(\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T) - C(\mathbf{A}_{R,n}) \\
&= \frac{1}{2L}\frac{\varrho\delta}{n_T}\mathbb{E}\left\{\operatorname{tr}\left[\left(\mathbf{M}_1 + \mathbf{M}_1^T + \delta\mathbf{M}_2\right)\mathbf{Z}^{-1} - \frac{\varrho\delta}{2n_T}\left(\mathbf{M}_1 + \mathbf{M}_1^T + \delta\mathbf{M}_2\right)^2\mathbf{Z}^{-2} + \cdots\right]\right\}.
\end{aligned}
$$

In the differential quotient, all terms with δ of order higher than one will vanish because of the limit, thus leading to

$$
\left[\frac{\partial C(\mathbf{A}_{R,n})}{\partial\mathbf{A}_{R,n}}\right]_{i,j} = \lim_{\delta\to 0}\frac{C(\mathbf{A}_{R,n} + \delta\xi_i\eta_j^T) - C(\mathbf{A}_{R,n})}{\delta} = \frac{\varrho}{2n_TL}\mathbb{E}\left\{\operatorname{tr}\left[\left(\mathbf{M}_1 + \mathbf{M}_1^T\right)\mathbf{Z}^{-1}\right]\right\},
$$

with $\mathbf{M}_1 = \mathbf{I}_{n_R} \otimes \left(\mathbf{I}_2 \otimes \left(\xi_i \eta_j^T\right)\right) \operatorname{vec}(\mathbf{H}') \operatorname{vec}(\mathbf{H}')^T \left[\mathbf{I}_{n_R} \otimes \boldsymbol{\mathcal{A}}_n\right]^T$. By using the identity $(\mathbf{X}^{-1})^T = (\mathbf{A}^T)^{-1}$ and the symmetry of $\mathbf{Z}$, i.e. $\mathbf{Z}^T = \mathbf{Z}$, we can simplify our result to

$$\left[\frac{\partial C(\mathbf{A}_{R,n})}{\partial \mathbf{A}_{R,n}}\right]_{i,j} = \frac{\varrho}{n_T L} \mathbb{E}\left\{\operatorname{tr}\left(\mathbf{M}_1 \mathbf{Z}^{-1}\right)\right\}.$$

Now this is exactly the same as stated in Theorem 4.3.2. The derivation of the other gradients is performed in complete analogy.

A.2.5. Proof of Presentability and Orthogonality of Φ

The proof is straightforward. Let us use the definition of the new modulation matrices from Theorem 4.3.4. Furthermore, bring the linear STBC mapping from Definition 4.1.2 back in mind. Then, for example the second symbol s_2 is modulated via the linear relation

$$\mathbf{A}_2 \theta s_{2,R} + j\mathbf{B}_2 \theta s_{2,I} = \mathbf{A}_2 \theta s_{2,R} + j\mathbf{A}_2 \theta s_{2,I}.$$

Now let us decompose the complex number θ into its real and imaginary parts and rearrange the above relation to

$$\mathbf{A}_2(\theta_R + j\theta_I)s_{2,R} + j\mathbf{A}_2(\theta_R + j\theta_I)s_{2,I} = \mathbf{A}_2(\theta_R s_{2,R} - \theta_I s_{2,I}) + j\mathbf{A}_2(\theta_I s_{2,R} + \theta_R s_{2,I}).$$

Similar relations can be easily obtained for $n = 1, 3$ and 4. Using these, one easily sees that the matrix $\boldsymbol{\Phi}$ that transforms the original $\mathbf{s}_{\mathrm{LD}}$ into the corresponding $\mathbf{s}'_{\mathrm{LD}}$ so that the above relation (among the others obtainable for $n = 1, 3, 4$) is fulfilled, is given by

$$\boldsymbol{\Phi}^T = \operatorname{diag}\left(\mathbf{I}_2, \begin{bmatrix} \theta_R & -\theta_I \\ \theta_I & \theta_R \end{bmatrix}, \begin{bmatrix} \phi_R & -\phi_I \\ \phi_I & \phi_R \end{bmatrix}, \begin{bmatrix} (\theta\phi)_R & -(\theta\phi)_I \\ (\theta\phi)_I & (\theta\phi)_R \end{bmatrix}\right).$$

This proves the presentability by $\boldsymbol{\Phi}$. To prove its orthogonality, we note that the transpose $\operatorname{diag}(\mathbf{X}_1, \ldots, \mathbf{X}_M)^T$ equals $\operatorname{diag}(\mathbf{X}_1^T, \ldots, \mathbf{X}_M^T)$ for any set of block matrices $\mathbf{X}_i, i = 1, \ldots, M$. Thus, $\boldsymbol{\Phi}\boldsymbol{\Phi}^T$ is orthogonal if and only if all block matrices inside the $\operatorname{diag}(\cdot)$ operator are orthogonal. This can easily be verified by the fact that ϕ is always chosen to be of unit magnitude (see [41]). This concludes the proof.

A.3. Review of some Mathematical Concepts

A.3.1. Frobenius Norm of a Matrix

Definition A.3.1 (Frobenius norm)**.** *The Frobenius norm of a matrix* $\mathbf{X}$ *with size* $m \times n$ *is defined as (see [23])*

$$\|\mathbf{X}\|^2 \triangleq \sum_{i=1}^{m} \sum_{j=1}^{n} |x_{ij}|^2 = \operatorname{tr}\left\{\mathbf{X}\mathbf{X}^H\right\} = \operatorname{tr}\left\{\mathbf{X}^H\mathbf{X}\right\} = \sum_{i=1}^{\min\{m,n\}} \lambda_i^2, \tag{A.7}$$

where we used the cyclic property of the trace and λ_i *denotes the* i*-th singular value of* $\mathbf{X}$*.*

A.3.2. Singular Value Decomposition

Suppose $\mathbf{M}$ being a $m \times n$ matrix with elements from $\mathbb{R}$ or $\mathbb{C}$. Then there exists a factorization of the form [23]

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H,$$

where $\mathbf{V}$ is a $m \times m$ unitary matrix over $\mathbb{R}^{m\times m}$ or $\mathbb{C}^{m\times m}$, describing the rows of $\mathbf{M}$ with respect to the base vectors associated with the singular values, $\mathbf{\Sigma}$ is a $m \times n$ matrix with singular values on the main diagonal, all other entries zero and $\mathbf{V}^H$ denotes the complex transpose of $\mathbf{V} \in \mathbb{R}^{m\times m}$ or $\mathbb{C}^{m\times m}$, an $n \times n$ matrix, which describes the columns of $\mathbf{M}$ with respect to the base vectors associated with the singular values.

Bibliography

[1] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41 – 59, 1996.

[2] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585 – 595, November 1999.

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

[4] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of mimo space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 281 – 302, April 2003.

[5] D. W. Bliss, K. W. Forsythe, A. O. Hero, and A. F. Yegulalp, "Environmental issues for mimo capacity," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2128 – 2142, September 2002.

[6] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[7] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[8] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge University Press, 2003.

[9] S. H. Müller-Weinfurtner, "Coding approaches for multiple antenna transmission in fast fading and ofdm," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2442 – 2450, October 2002.

[10] J. G. Proakis, *Digital Communications*. McGraw-Hill Book Co., 3 ed., 1995.

[11] F. Hlawatsch, "Übertragungsverfahren 1+2," 2002.

[12] H. Weinrichter and F. Hlawatsch, "Grundlagen nachrichtentechnischer signale," 2002.

[13] H. Bölcskei and A. Paulraj, *The Communications Handbook*, ch. Multiple-Input Multiple-Output (MIMO) Wireless Systems, p. 22. CRC Press, 2nd ed., 1997.

[14] A. J. Paulraj, D. Gore, and R. U. Nabar, "Performance limits in fading mimo channels," in *The 5th International Symposium on Wireless Personal Multimedia Communications*, vol. 1, pp. 7–11, 2002.

[15] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, pp. 139 – 157, January 1999.

[16] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, vol. 48, pp. 1804 – 1824, July 2002.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. III, John Wiley & Sons, Inc., 1991.

[18] F. Hlawatsch, "Information theory for communication engineers," 2003.

[19] J. R. Pierce, *An Introduction to Information Theory*. Dover Publications, Inc., 2nd ed., 1980.

[20] H. Jafarkhani, *space-time coding, theory and practice*. Cambridge University Press, 2005.

[21] M. Jankiraman, *space-time codes and MIMO systems*. Artech House, 2004.

[22] B. Vucetic and J. Yuan, *Space-Time Coding*. John Wiley & Sons, Inc., 2003.

[23] I. N. Bronstein and K. A. Semendjajew, *Teubner-Taschenbuch der Mathematik*. B.G. Teubner Verlagsgesellschaft, 1996.

[24] C. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in mimo wireless systems under correlated fading," *IEEE Transactions on Information Theory*, vol. 48, pp. 637 – 650, March 2002.

[25] S. A. Jafar and A. Goldsmith, "Multiple-antenna capacity in correlated rayleigh fading with channel covariance information," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 990 – 997, May 2005.

[26] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1073 – 1096, May 2003.

[27] A. F. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1459 – 1478, October 1998.

[28] D. M. Ionescu, "On space-time code design," *IEEE Transactions on Wireless Communications*, vol. 2, pp. 20 – 28, January 2003.

[29] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744 – 765, March 1998.

[30] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Transactions on Communications*, vol. 47, pp. 199 – 207, February 1999.

[31] E. G. Larsson, P. Stoica, and J. Li, "On maximum-likelihood detection and decoding for space-time coding systems," *IEEE Transactions on Signal Processing*, vol. 50, pp. 937 – 944, April 2002.

[32] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, pp. 389 – 399, March 2003.

[33] H. E. Gamal and M. O. Damen, "Universal space-time coding," *IEEE Transactions on Information Theory*, vol. 49, pp. 1097 – 1119, May 2003.

[34] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Select Areas in Communications*, vol. 16, pp. 1451 – 1458, October 1998.

[35] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, pp. 1456 – 1467, July 1999.

[36] S. Sandhu and A. Paulraj, "Space-time block codes: A capacity perspective," *IEEE Communications Letters*, vol. 4, pp. 384 – 386, December 2000.

[37] G. Ganesan and P. Stoica, "Space-time diversity using orthogonal and amicable orthogonal designs," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 2561 – 2564, June 2000.

[38] K. Tanaka, R. Matsumoto, and T. Uyematsu, "Maximum mutual information of space-time block codes with symbolwise decodability," in *International Symposium on Information Theory and its Applications*, pp. 1025 – 1030, 2004.

[39] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[40] I. N. Bronstein and K. A. Semendjajew, *Teubner-Taschenbuch der Mathematik Teil II*. B.G. Teubner Verlagsgesellschaft, 1995.

[41] M. O. Damen, A. Tewfik, and J.-C. Belfiore, "A construction of a space-time code based on number theory," *IEEE Transactions on Information Theory*, vol. 48, pp. 753 –760, March 2002.

[42] A. M. Sayeed, J. H. Kotecha, and Z. Hong, "Capacity-optimal structured linear dispersion codes for correlated mimo channels," in *IEEE Vehicular Technology Conference*, vol. 3, pp. 1623 – 1627, September 2004.

[43] H. Yao, *Efficient Signal, Code, and Receiver Designs for MIMO Commnunication Systems*. PhD thesis, Massachusetts Institute of Technology, June 2003.

[44] H. Yao and G. W. Wornell, "Structured space-time block codes with optimal diversity-multiplexing tradeoff and minimum delay," in *IEEE Global Telecommunications Conference*, vol. 4, pp. 1941 – 1945, 2003.

[45] V. Shashidhar, B. S. Rajan, and P. V. Kumar, "Asymptotic-information-lossless designs and diversity-multiplexing tradeoff," in *IEEE Global Telecommunications Conference*, vol. 1, pp. 366 – 370, 2004.

[46] L. Dai, S. Sfar, and K. B. Letaif, "Towards a better diversity-multiplexing tradeoff in mimo systems," in *IEEE Taiwan/Hong Kong Joint Workshop on Information Theory and Communications*, 2005.

[47] N. M. Blachman, "The amount of information that y gives about x," *IEEE Transactions on Information Theory*, vol. 14, pp. 27 – 31, January 1968.