ulm university universität
uulm

Non-Coherent Communication in Wireless Point-to-Point and Relay Channels: A Geometric Approach

DISSERTATION

zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS
(DR.-ING.)

der Fakultät für Ingenieurwissenschaften
und Informatik der Universität Ulm

von

ZORAN UTKOVSKI
AUS SKOPJE

Gutachter:	Prof. Dr.-Ing. Jürgen Lindner
	Prof. Dr. Petar Popovski
Amtierender Dekan:	Prof. Dr.-Ing. Michael Weber

Ulm, 07. September 2010

Za mojata najsakana Verce

Acknowledgments

I would like to express my gratitude to my supervisor, Prof. Juergen Lindner whose expertise, experience, and understanding, added considerably to the quality of my work. I particularly thank him for giving me the opportunity to do my dissertation at the Institute of Information Technology. My gratitude goes also to Dr. Werner Teich whose continuous support made the time spent at the university pleasant and significantly contributed to my graduate experience. I would like to thank Prof. Petar Popovski for acting as second examiner and for giving me the opportunity to visit the University of Aalborg, which resulted in a successful cooperation.

My thanks also go to my colleagues and the stuff at our Institute. I thank Christian Sgraja, Christian Pietsch, Marcus and Alex for the helpful discussions and the evenings at Dolce. I was happy to meet Ivan, with whom I spent nice time on conferences and outside the working hours, particularly when watching football. I thank Doris for the Ramstein CD and Jochem for the TV. I thank Dmitry for introducing me the Sunday football. I thank Matthias for playing Kubb together and generously helping me in all occasions, which I really appreciate. I thank Thanawat for sharing the passion for Manchester United. I thank Mohammed for the political discussions at Montreaux. I also thank Rui and Eva for the helpful discussions at the group meetings. Finally, I thank Sergey and Ksenya for the time spent together in different activities. I appreciate Sergey's 24/7 technical support and help with the thesis submission. I also thank the other former colleagues and the colleagues from Dialogue Systems.

I would also like to express my truly deep gratitude to my parents and my brother for their continuous support during my education. I appreciate their efforts, their patience, their sacrifice and their love. I also thank my broader family, especially my uncle who probably brought the interest for mathematics to me.

Finally, my deepest gratitude goes to my wife who was always there for me, in all the time during my graduation. I truly appreciate her endless and unconditional support, her love and sacrifice. She always had the necessary understanding for what I was doing and always believed in me. Without her, none of this would be possible.

Contents

Introduction

One of Shannon's most important ideas was the geometric interpretation of communications systems in general and messages, code words, and the encoding and decoding procedures in particular. Shannon exploited the geometric representation to derive the celebrated capacity formula for the AWGN (additive white Gaussian noise) channel. He used the observation that the noise turns the coded signal point into a cloud sphere of an essentially fixed radius, adhering basically to what is currently known as sphere hardening.

When moving from AWGN to fading channels and systems with, in general, more than one antenna, the analysis reveals that the problem of information transmission can be interpreted as a constrained packing problem in Riemannian manifolds. For example, in the case of unknown block Rayleigh fading channels, there is an insight generated by the concept of coordinate change which arises from the observation that the fading channel matrix does not change the subspace in which the transmitted signal resides, but it merely rotates and scales the bases of this subspace. However, the combined effect of noise and fading results in the perturbation of the signal subspace in a specific manner. Based on this geometric insight, it was shown that, at high SNR, the information carrying object is a linear subspace. For M transmit antennas and coherence time T, we can say that the relevant coding space is the set of M-dimensional linear subspaces of the, in general, complex T-dimensional space. This set can be given a structure of a manifold and is known as the Grassmann manifold $G^{\mathbb{C}}_{T,M}$. The differential structure of the Grassmann manifold yields non-coherent space-time codes based on sets in the tangent bundle of the manifold with connection given by the exponential map. Therefore, a treatment of the geometry and the structure of certain Riemannian manifolds is needed in order to formulate the problems and eventually construct codes for communication.

When moving from point-to-point systems to wireless networks, the situation becomes more difficult. One example are the ad hoc networks which are mobile peer-to-peer networks that operate without the assistance of preexisting infrastructure. Ad

hoc networks are the most general class of wireless networks and, at the same time, the most challenging to both quantify and design. Immediate applications of ad hoc networks include emergency networks, metropolitan mesh networks for broadband Internet access, and sensor networks. Besides in communications, ad hoc networks (and networks in general), play an important role in other fields, including biology, economics, and transportation. Information theory has been successful in describing the point-to point communication links and centralized networks. However, it has not been successfully applied to decentralized wireless networks. One of the reasons is the lack of a theory which describes the fundamental performance limits of mobile ad hoc networks. In contrary to the point-to-point communication links, where the capacity has been characterized in most of the practically relevant cases, the capacity characterization of wireless networks, and particularly mobile ad hoc networks, faces serious difficulties. This, on the other hand, has impeded the development and commercialization of many types of wireless networks.

The development of a general capacity theory for wireless networks is certainly a difficult task. However, a contribution to the problem solution might come from some underemployed techniques in capacity characterization. One possible approach is based on the geometric interpretation of the capacity, rather than the deployment of the typical algebraic tools. One of the advantages of the geometric approach is that, when put in a geometric context, the performance limits of communication systems have a more intuitive explanation. Second, sometimes is the formulation in geometric terms simpler.

These examples encourage the deployment of general geometric techniques for the description of wireless networks. Besides the development of the geometric framework, the construction of codes can also benefit from the geometric approach. This is motivated from some observations from point-to-point systems. Let us take for example the famous Alamouti code, initially developed for a quasi-static Rayleigh fading channel and communication system where the transmitter has two transmit antennas. The Alamouti code is a special case of a so-called orthogonal space-time block code. Orthogonal space-time block codes are known for having good diversity properties and allow for simple, symbol-per-symbol decoding. As far as the existence of orthogonal space-time block codes is concerned, for systems with more then two antennas it has been proved that these codes exist only in certain cases (number of antennas) and only for certain transmission rates (in terms of number of symbols per channel use). From algebraic point of view, it is not trivial to show why this is the case. However, another insight may be gained through the special representation of these codes via linear dispersion matrices. It can be shown that there is a connection between the linear dispersion matrices used to describe the space-time codes and special packings of subspaces in the Grassmann manifold. Moreover, there is an embedding of the Grassmann manifold with the chordal distance as metric, in a higher dimensional

Euclidean space where the chordal distance equals the Euclidean distance. After the embedding, the set of subspaces (or equivalently the set of dispersion matrices) represents the set of the vertices of an orthoplex in this higher-dimensional space. From geometry we know that the number of vertices of an orthoplex in certain dimension is fixed, which fixes the number of linear dispersion matrices which represent an orthogonal space-time code and thus explains the existence of these codes. We see that the geometric interpretation not only explains certain phenomena, but also can serve as a mean for code construction. Further, it provides bounds on what is achievable in a certain setup.

The described examples show that the geometric approach brings more intuition in the understanding of the fundamental limits of both point-to-point systems and wireless networks. Moreover, the geometric tools offer the possibility to construct high-dimensional codes which in certain cases mimic the optimal input distributions.

With the above mentioned, the focus in this thesis is mainly on the deployment of the geometric framework for the construction of codes for non-coherent communication over flat fading MIMO channels, both in the case of point-to-point systems as well as wireless networks with relays. The thesis is organized as follows. After the introduction, in the first chapter we present the theoretical preliminaries. We first describe the properties of Rayleigh fading channels and give examples of flat fading channels. The focus is on the block fading channel which is used as our system model throughout the thesis. We also give an introduction to space-time coding and analyze the effect of diversity offered by the deployment of multiple antennas. In Chapter 2 we provide the basics for non-coherent space-time coding. We summarize the capacity results and the performance analysis and design criteria for non-coherent space-time codes. Since the problem of non-coherent space-time coding can be interpreted as a constrained packing problem in Grassmann manifolds, we present the necessary preliminaries for the geometry of Grassmann manifolds. A more detailed introduction is given in the Appendix. The main contribution of the thesis is presented in Chapter 3, 4 and 5. In Chapter 3 we construct novel codes for the non-coherent point-to-point channels based on geometric methods in Grassmann manifolds and analyze their performance. The constructions exploit the differentiable structure of Grassmann manifolds. In Chapter 4 and Chapter 5 we focus on wireless networks with relays and address both the cases of one-way (uni-directional) and two-way (bi-directional) relaying. We present novel code constructions for both cases. For the case of two-way relaying we additionally derive rate bounds and prove the optimality of the amplify-and-forward scheme in the high SNR regime. We conclude the thesis in the last chapter.

The main novelty presented in this thesis is contained in Chapter 3, Chapter 4 and Chapter 5. Parts of this thesis were published in [53], [56], [55], [54], [52], [51], [49], [64], [57], [58], [59], [50].

Chapter 1

Basics

In this chapter we present some basic principles of wireless communications. We first introduce the Rayleigh fading channel, which is widely used to model mobile communication channels. Compared to the AWGN channel, the communication over Rayleigh fading channels is more challenging. We discuss fast and slow fading channels and their characteristics. We also present the block fading model which is the fading model of interest in this thesis and can be seen as a special case of a slow fading channel. The adverse effects of the Rayleigh fading can be mitigated by the introduction of multiple antenna systems, which we briefly describe. Further, we present the basics of space-time coding, which can be seen as technique that joints error-control coding, modulation, and transmit diversity for the multiple antenna systems.

1.1 Rayleigh Fading Channels

The wireless characteristic of the channel places fundamental limitations on the performance of wireless communication systems. Wireless channels are random and are not easily analyzed due to the diverse environments, the motion of the transmitter, the receiver, and the surrounding objects. In this section, characteristics of wireless

channels are discussed and the Rayleigh flat-fading channel model is explained in detail.

In a mobile wireless environment, the surrounding objects act as reflectors of electromagnetic waves. Due to these refections, electromagnetic waves travel along different paths of varying lengths and therefore have various amplitudes and phases. The linear superposition of these waves causes multiple fading at the receiver location, and the strength of the waves decreases as the distance between the transmitter and the receiver increases.

Traditionally, propagation modeling focuses on two aspects. Propagation models that predict the mean signal strength for an arbitrary transmitter-receiver separation distance are called *large-scale* propagation models since they characterize signal strength over large transmitter-receiver distances [40]. Propagation models that characterize the rapid fluctuations of the received signal strength over very short travel distances or short time durations are called *small scale* or *fading models* [40]. In this thesis, the focus is on fading models, which are more suitable for indoor and urban areas.

Small-scale fading is affected by many factors, such as multiple-path propagation, velocity of the transmitter and receiver, velocity of surrounding objects, and the transmission bandwidth of the signal. In this work, narrowband systems are considered, in which the bandwidth of the transmitted signal is smaller than the channel's coherence bandwidth, which is defined as the frequency range over which the channel fading process is correlated [40]. This type of fading is referred to as *flat fading* or *frequency non-selective fading*.

The Rayleigh distribution is commonly used to describe the statistical time-varying nature of the received envelope of a flat-fading signal. It is also used to model fading channels in this thesis. For a typical mobile wireless channel in indoor or urban areas, we may assume that the direct line-of-sight wave is obstructed and the receiver obtains only reflected waves from the surrounding objects. When the number of reflected waves is large, according to the central limit theorem, two quadrature components of the received signal are uncorrelated Gaussian random processes with mean zero and variance $\frac{1}{2}$. As a result, the envelope of the received signal at any time instant has a Rayleigh distribution and its phase is uniform between $-\pi$ and π. It is important to note that this model is applicable for narrowband signals, since then the signals traveling different paths can not be resolved in time at the receiver and they add together. This is the case when the signal bandwidth is sufficiently small, given the delays of the reflected signals.

We can say that there are two sources of noise at work: multiplicative noise that is associated with the Rayleigh fading, and the usual additive receiver noise. This makes the Rayleigh fading channel essentially different from the Additive White Gaussian Noise (AWGN) channel.

1.2 Fading Models of Interest

Rayleigh fading significantly deteriorates the performance of the communication system. It can be easily shown that under Rayleigh fading, the error probability of an uncoded modulation scheme decays only inversely proportional with the received signal-to-noise-ratio (SNR) [48]. Compared to the AWGN channel, where the error probability decays exponentially, this means that substantially more power should be added in order to obtain the same decrease in the error probability.

The underlying fading model is very important for the design of the communication system, since different fading models require different transmit and receive strategies. If the fading model is such that each transmitted symbol is affected independently, this effect can be averaged out by using long channel codes. This is the case of *fast fading*. The performance of this channel will then approach the performance of the AWGN channel. However, very often is the case that the channel is constant for a longer period of time which spans a lot of transit symbols. In this case, when a bad channel realization (small amplitude) takes place, a long seqzunce of transmit symbols will be affected and most certainly the whole sequence will be in error. The extreme example is the so called *slow fading* or *quasi-static fading* model where we basically assume that the channel is constant for the total duration of the transmission. In this case, channel coding can not average out the effects of the fading, since when the channel is bad, it remains bad for the duration of channel coding blocks of practical relevance. Strictly speaking, in this case the ergodic capacity of this channel is zero. Under this scenario, it makes sense to speak about the so-called *outage capacity*.

Let us look at the single-antenna example first. According to the slow fading model the channel gain is random but remains constant for all time, i. e. , $h[m] = h$ for all m. This models the slow fading situation where the delay requirement is short compared to the channel coherence time. The equivalent low-pass representation for a single-antenna communication system can thus be written as

$$y = \sqrt{p}sh + w \tag{1.1}$$

where h is the Rayleigh flat-fading channel coefficient, p is the transmit power and w is the noise at the receiver, which is Gaussian with zero-mean and unit-variance. s satisfies the power constraint $\mathrm{E}[|s|^2] = 1$. With this the average received signal-to-noise ratio (SNR) is p.

Conditional on a realization of the channel h, this is an AWGN channel with instantaneous received signal-to-noise ratio $p|h|^2$. Since $|h|$ is Rayleigh distributed, $|h|^2$ is exponentially distributed with probability density function

$$p(x) = \exp(-x),\ x > 0. \tag{1.2}$$

Thus, the probability that the receive SNR is less than a level σ is,

$$\text{Prob}\left[p|h|^2 \leq \sigma\right] = 1 - \exp(-\sigma/p). \tag{1.3}$$

We will mostly be interested in the high SNR regime. When the average SNR is high, ($p \gg 1$), we have

$$\text{Prob}\left[p|h|^2 \leq \sigma\right] \approx \frac{\sigma}{p}. \tag{1.4}$$

We can see that the probability that the received SNR is below a certain threshold decays only linearly with the SNR. This is also the typical error event, i. e . approximately the error probability of the transmission scheme.

With channel knowledge at the receiver, the maximum rate of reliable communication supported by this channel is [48]

$$R = \log_2(1 + p|h|^2) \text{ bits/s/Hz}. \tag{1.5}$$

This quantity is a function of the random channel gain h and is therefore random. Now suppose the transmitter encodes data at a rate R bits/s/Hz. If the channel realization h is such that $\log_2\left(1 + p|h|^2\right) \leq R$, then whatever the code used by the transmitter, the decoding error probability cannot be made arbitrarily small. The system is said to be in outage, and the outage probability is [48]

$$P_o(R) = \text{Prob}\left[\log_2\left(1 + p|h|^2\right) \leq R\right]. \tag{1.6}$$

We can say that there is a conceptual difference between the AWGN channel and the slow fading channel. In the former, one can send data at a positive rate (any rate less than the capacity C) while making the error probability as small as desired. This, however is not possible for the slow fading channel as long as the probability that the channel is in deep fade is non-zero. Thus, the capacity of the slow fading channel in the strict sense is zero. An alternative performance measure is the ϵ-outage capacity C_ϵ. This is the largest rate of transmission R such that the outage probability $P_o(R)$ is less than ϵ.

With the above said, the outage probability is given

$$P_o(R) = 1 - \exp\left(\frac{2^R - 1}{p}\right). \tag{1.7}$$

At high SNR,

$$P_o(R) \approx \frac{2^R - 1}{p}, \tag{1.8}$$

which is a result of the Taylor expansion. We observe that the outage probability decays as $1/p$, which is the same as the uncoded transmission. Thus, we see that

coding cannot significantly improve the error probability in a slow fading scenario. The reason is that while coding can average out the Gaussian white noise, it cannot average out the channel fade, which affects all the coded symbols. Thus, deep fade, which is the typical error event in the uncoded case, is also the typical error event in the coded case. Solving $P_o(R) = \epsilon$ yields

$$C_\epsilon = \log_2(1 + F^{-1}(1-\epsilon)p) \text{ bits/s/Hz}, \tag{1.9}$$

where F is the complementary cumulative distribution function of $|h|^2$, i.e., $F(x) = \text{Prob}\left[|h|^2 > x\right]$. For small ϵ we have,

$$C_\epsilon \approx \log_2(1 + \epsilon p). \tag{1.10}$$

If we compare with the capacity of the AWGN channel we get

$$C_\epsilon \approx C_{AWGN} - \log_2\left(\frac{1}{\epsilon}\right). \tag{1.11}$$

We can see that in the high SNR regime, we have a constant difference between the AWGN channel capacity and C_ϵ.

It is important to mention that the concept of outage and outage capacity is strictly speaking result of the definition of the slow fading channel, where we assume that the channel is constant for the whole duration of the transmission. In reality this may be realistic when the delay requirement is such that it is not possible to code over more then one block.

1.3 MIMO Transmission Systems

Multiple-input-multiple-output (MIMO) transmission systems have attracted attention in wireless communications, since they offer significant increases in data throughput and link range without additional bandwidth or transmit power. The technology figures prominently on the list of recent technical advances with a a chance of resolving the bottleneck of traffic capacity in future wireless networks.

MIMO systems can be defined simply. Given an arbitrary wireless communication system, we consider a link for which the transmitting end as well as the receiving end is equipped with multiple antenna elements, as illustrated in Fig. 1.1.

For a multiple antenna system with M transmit and N receive antennas, with the same transmit power, the system model equation is

$$\mathbf{y} = \sqrt{p}\mathbf{s}\mathbf{H} + \mathbf{w}, \tag{1.12}$$

where $\mathrm{E}[\mathbf{s}^*\mathbf{s}] = 1$. With this, the instantaneous received SNR is $\frac{p}{M}\sum_{i=1}^{M}\sum_{j=1}^{N}|h_{ij}|^2$.

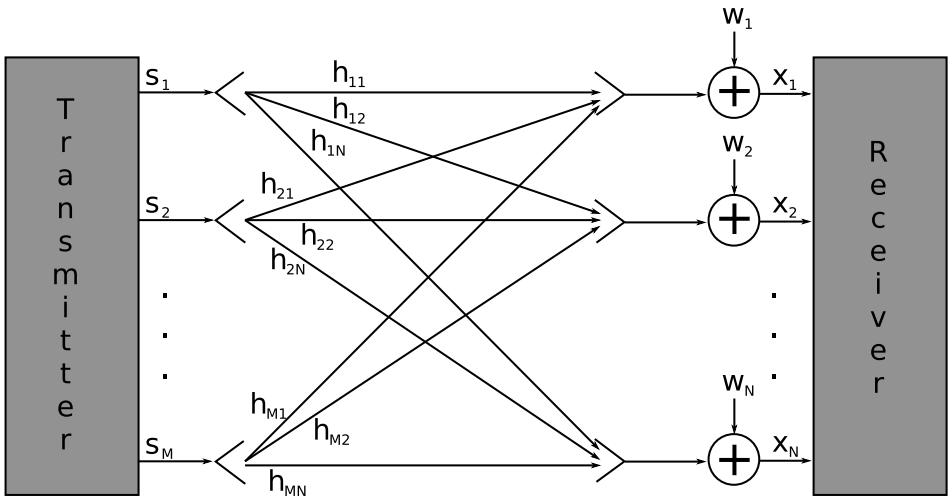

Figure 1.1: Diagram of a MIMO wireless transmission system. The transmitter and receiver are equipped with multiple antenna elements.

The above definition of MIMO transmission systems is quite general and independent of the channel model assumed. However, when it comes to the capacity and error performance of a certain MIMO system, the channel model plays a crucial role. We will thus focus on the effect of multiple antennas in our fading models of interest. We will start with the slow fading model, which we take as illustration for the extreme effects of the Rayleigh fading on the performance of the system. After this we will concentrate on the model of interest in this thesis which is the block fading model introduced by Marzetta and Hochwald [31].

1.3.1 Multiplexing

As already discussed, the capacity of different MIMO channels depends heavily on the channel model assumed and the degree of channel knowledge at the transmitter and/or at the receiver. In this sense MIMO channels with fast fading assumption differ substantially from MIMO channels with slow fading or block fading assumption. What is common for all models is that these systems offer certain spatial multiplexing gains. Roughly speaking, we can say that the multiplexing gain is the number of parallel streams that can be sent over the different antennas. More precisely, it is the factor in front of the $\log_2 SNR$ in the capacity expression. For example, for a fast fading Rayleigh MIMO channel, the capacity scales like $\min(M, N) \log_2 SNR$, where $\min(M, N)$, i. e. the factor in front of the logarithmic term, is the maximal multiplexing gain provided by the system. Another term to denote this factor is the degree of freedom of the system.

For slow fading and block fading channels, the multiplexing gain has a different value. In that context it also makes sense to look at the multiplexing gain and the diversity gain from a common perspective. We will come back to this question at the end of this chapter where we introduce the diversity-multiplexing-tradeoff.

1.3.2 Diversity

We have seen that under the slow fading assumption, both the capacity and the performance of the system in terms of error probability are significantly deteriorated. Hence, other techniques are needed in order to mitigate these effects. Diversity is one technique which can improve the performance of the system under these circumstances.

Diversity–Classical Definition

The basic idea of diversity is that, if two or more independent signals are sent and then fade in an uncorrelated manner, the probability that all the signals are simultaneously below a given level is much lower than the probability of any one signal being below that level. Thus, properly combining various signals reduces the severity of fading and improves reliability of transmission.

According to the domain where diversity is introduced, it can be classified into time diversity, frequency diversity and antenna diversity (space diversity). Time diversity can be achieved by transmitting identical messages in different time slots, which results in uncorrelated fading signals at the receiver. Frequency diversity can be achieved by using different frequencies to transmit the same message. The issue we are interested in is space or antenna diversity, which is typically implemented using multiple antennas at the transmitter or the receiver or both. The multiple antennas should be separated physically by a proper distance to obtain independent fading. Typically a separation of a few wavelengths is enough. Depending on whether multiple antennas are used for transmission or reception, space diversity can be classified into two categories: receive diversity and transmit diversity. To achieve receive diversity, multiple antennas are used at the receiver to obtain independent copies of the transmitted signals. The replicas are properly combined to increase the overall receive SNR and mitigate fading. There are many combining methods, for example, selection combining, switching combining, maximum ratio combining, and equal gain combining [37]. Transmit diversity is more difficult to implement than receive diversity due to the need for more signal processing at both the transmitter and the receiver. In addition, it is generally not easy for the transmitter to obtain information about the channel, which results in more complexity in the system design.

As already discussed, in the single antenna case the error probability is only inversely proportional to the SNR. In the following we evaluate shortly the effect of multiple transmit and receive antennas on the error probability. The outage probability is the probability that the SNR at the receiver is less than the level σ and is given by

$$Pr\left[\frac{p}{M}\sum_{i=1}^{M}\sum_{j=1}^{N}|h_{ij}|^2 \leq \sigma\right] = \left(1-\exp\left(-\frac{\sigma M}{p}\right)\right)^{MN} \approx \left(\frac{\sigma M}{p}\right)^{MN}, \qquad (1.13)$$

when the transmit power is high. The outage probability is thus inversely proportional to MN. Therefore, multiple-antenna systems have much lower error probability than single-antenna systems at high transmit power.

In this light, a key measure of the performance capability of a slow fading channel is the maximum diversity gain that can be extracted from it. For example, a slow i. i. d. Rayleigh faded MIMO channel with M transmit and N receive antennas has a maximum diversity gain of MN i. e. , for a fixed target rate R, the outage probability $P_o(R)$ decays like $1/SNR^{MN}$ at high SNR. This is the classical definition for diversity which appears in the first papers addressing the topic, e. g. [19], [46]. We will revisit the concept of diversity when we speak about the diversity-multiplexing-tradeoff at the end of this chapter.

1.4 Space-Time Coding

It is important to develop algorithms that take advantage of the spatial diversity provided by multiple antennas. Many algorithms with reasonable complexity and performance have been proposed, for example, the diversity techniques and diversity combining methods (see,[63],[45],[40],[37]). Among them, the most successful one is space-time coding, in which time (the natural dimension of digital communication data) is complemented with the spatial dimension inherent in the use of multiple spatially distributed antennas. By doing this, both the data rate and the performance are improved by many orders of magnitude with no extra cost of spectrum. We can say that space-time coding is a joint design of error-control coding, modulation, and transmit diversity. In this thesis we focus on space-time coding schemes for the block fading channel model. Therefore, in the following we first present this model in detail.

1.4.1 The Block Fading Channel

In the previous sections we have discussed two different channel models. The first one assumed that the channel changes between each transmitted symbol in an inde-

pendent realization. The other model was the quasi static or the slow fading model where the channel coherence time is larger than the delay requirement for the system and thus the channel is constant for the whole duration of the transmission. In reality, it is very often the case that the channel is constant for a finite period of time T and then changes to a different, independent realization. The propagation coefficients are assumed to be constant for T symbol periods, after which they change to new independent random values which they maintain for another T symbol periods, and so on. This is the block fading model introduced by Marzetta and Hochwald in [31]. This piecewise-constant fading process approximates, in a tractable manner, the behavior of a continuously fading process such as Jakes [22]. Furthermore, it is a very accurate representation of many time-division multiple access (TDMA), frequency-hopping, or block-interleaved systems [31]. The random propagation coefficients are modeled as independent, identically distributed, zero-mean, circularly symmetric complex Gaussian random variables of variance 1.

On the transmitter side, the information bits are encoded into $T \times M$ transmit matrix $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_M]$, where $\mathbf{s}_m = [s_{m,1}, s_{m,2}, \ldots, s_{m,T}]^{\mathrm{T}}$ is signal at the m-th antenna respectively. The transmit matrix is normalized to $\mathrm{tr}(\mathbf{S}^{\mathrm{H}}\mathbf{S}) = M$. The antennas then send the signals simultaneously to the receiver. Every receive antenna at the receiver obtains a signal that is a superposition of the signals from every transmit antenna through the fading coefficient. The received signal is also corrupted by noise. If we denote the noise at the n-th receive antenna by $\mathbf{w}_n$, which is i.i.d complex Gaussian $CN(0,1)$, then the received signal at the n-th receive antenna is

$$\mathbf{y}_n = \sqrt{\frac{\rho T}{M}} \sum_{m=1}^{M} \mathbf{s}_m h_{mn} + \mathbf{w}_n,$$

where ρ is the SNR at each receive antenna.

Let us denote the matrix of the received signal as $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N]$; the matrix of additive noise as $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_N]$; and the channel matrix as

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} & \ldots & h_{1N} \\ h_{21} & h_{22} & \ldots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M1} & h_{M2} & \ldots & h_{MN} \end{pmatrix}.$$

Then the system model can be written as

$$\mathbf{Y} = \sqrt{\frac{\rho T}{M}}\mathbf{S}\mathbf{H} + \mathbf{W}. \tag{1.14}$$

There has been much research on the topic of space-time coding and the design of space-time block codes. Most of the research concentrated on the case of the known

channel, and most of the well-known code constructions were designed for this channel. Examples include the orthogonal space-time codes [46] including the famous Alamouti code [2], unitary space-time codes [19] and later the codes obtained by algebraic methods (algebraic number theory) [5], [30], just to name few of them.

1.4.2 Space-Time Coding and Diversity-Multiplexing-Tradeoff

We have already argued that the key performance benefit of a fast fading MIMO channel is the spatial multiplexing capability it provides through the additional degrees of freedom. For example, the capacity of an i. i. d. Rayleigh fast fading channel scales like $\min(M,N)\log_2 SNR$, where $\min(M,N)$ is the number of spatial degrees of freedom in the channel. This fast fading (ergodic) capacity is achieved by averaging over the variation of the channel over time. In the slow fading scenario, no such averaging is possible and one cannot communicate at this rate reliably. Instead, the information rate allowed through the channel is a random variable fluctuating around the fast fading capacity. Nevertheless, one would still expect to be able to benefit from the increased degrees of freedom even in the slow fading scenario. Yet the maximum diversity gain provides no such indication; for example, both an $M \times N$ channel and an $MN \times 1$ channel have the same maximum diversity gain and yet one would expect the former to allow better spatial multiplexing than the latter. One needs something more than the maximum diversity gain to capture the spatial multiplexing benefit.

In [66] Zheng and Tse make the important observation that in order to achieve the maximum diversity gain, one needs to communicate at a fixed rate R, which becomes vanishingly small compared to the fast fading capacity at high SNR (which grows like $\min(M,N)\log_2 SNR$). Thus, one is actually sacrificing all the spatial multiplexing benefit of the MIMO channel to maximize the reliability. To reclaim some of that benefit, one would instead want to communicate at a rate $R = r\log_2 SNR$, which is a fraction of the fast fading capacity. In this spirit Zheng and Tse in [66] formulated the following diversity-multiplexing tradeoff (DMT) for a slow fading channel. A diversity gain $d^*(r)$ is achieved at multiplexing gain r if the rate is $R = r\log_2 SNR$ and the outage probability decays as $P_o(R) \approx SNR^{-d^*(r)}$. More precisely, the curve

$$d^*(r) = -\lim_{SNR\to\infty} \frac{\log_2 P_o(r\log_2 SNR)}{\log_2 SNR}, \tag{1.15}$$

gives the diversity-multiplexing tradeoff in slow fading channels.

We observe that this definition of diversity gain differs from the standard definition in the space-time coding literature presented in 1.3.2. In the standard formulation, diversity gain is an asymptotic performance metric of one fixed code. To be specific, the input of the fading channel is fixed to be a particular code, while SNR increases.

The speed that the error probability (of a maximum-likelihood (ML) detector) decays as SNR increases is called the diversity gain. In the formulation by Zheng and Tse, we notice that the channel capacity increases linearly with the SNR. Hence, in order to achieve a nontrivial fraction of the capacity at high SNR, the input data rate must also increase with the SNR, which requires a sequence of codebooks with increasing size. The diversity gain here is used as a performance metric of such a sequence of codes, which is formulated as a scheme. Under this formulation, any fixed code has 0 spatial multiplexing gain. Allowing both the data rate and the error probability scale with the SNR is the crucial element of this formulation and allows to talk about their tradeoff in a meaningful way.

Similarly, a diversity-multiplexing tradeoff for any space-time coding scheme can be formulated, with outage probabilities replaced by error probabilities. A space-time coding scheme is a family of codes, indexed by the SNR. It attains a multiplexing gain r and a diversity gain d if the data rate scales as $R = r \log_2 SNR$ and the error probability scales as $P_e \approx SNR^{-d}$ i. e .,

$$\lim_{SNR \to \infty} \frac{P_e}{\log_2 SNR} = -d, \tag{1.16}$$

The diversity-multiplexing tradeoff is fundamental property of space-time coding schemes. It is also probably the most important criterion when comparing two different space-time coding schemes (families of codes). For the code constructions presented in this thesis, the diversity-multiplexing tradeoff is out of scope. This is partially due to the fact that the constructions presented here rely mostly on a geometric approach, involving non-linear mapping, which makes the full characterization of the code properties difficult. Nevertheless, the diversity-multiplexing tradeoff analysis sheds light on the understanding of the problem of space-time coding and therefore we decided not to omit it from this introduction.

1.5 Chapter Summary

Fading Models

- Propagation models that characterize the rapid fluctuations of the received signal strength over very short travel distances or short time durations are called small scale or fading models.

- In narrowband systems, the bandwidth of the transmitted signal is smaller than the channel's coherence bandwidth. This type of fading is referred to as flat fading or frequency nonselective fading. The Rayleigh distribution is used to

describe the statistical timevarying nature of the received envelope of a flat-fading signal.

- If the fading model is such that each transmitted symbol is affected independently, this effect can be averaged out by using long channel codes. This is the case of *fast fading*.
- In the case of *slow fading* or *quasi-static fading* there is a maximum delay requirement, which practically means that the channel is constant for the total duration of the transmission.
- In slow fading, for a certain rate R, whatever the code used by the transmitter, the decoding error probability cannot be made arbitrarily small. The system is said to be in outage, and the outage probability is

$$P_o(R) = 1 - \exp\left(\frac{2^R - 1}{p}\right).$$

MIMO Transmission Systems

- For a multiple antenna system with M transmit and N receive antennas, with the same transmit power, the system equation is

$$\mathbf{y} = \sqrt{p}\mathbf{s}\mathbf{H} + \mathbf{w},$$

where $\mathrm{E}[\mathbf{s}^*\mathbf{s}] = 1$. The received SNR is $\frac{p}{M}\sum_{i=1}^{M}\sum_{j=1}^{N}|h_{ij}|^2$

Multiplexing

- MIMO systems offer a certain spatial multiplexing gains since parallel streams can be sent over the different antennas.
- For a fast fading Rayleigh MIMO channel, the capacity scales like $\min(M, N)\log_2 SNR$, where $\min(M, N)$ is the number of spatial degrees of freedom in the channel.

Diversity-Classical Definition

- If two or more independent samples of a signal are sent and then fade in an uncorrelated manner, the probability that all the samples are simultaneously below a given level is much lower than the probability of any one sample being below that level. Thus, properly combining various samples reduces the severity of fading and improves reliability of transmission. We say that in this case the system offers a certain degree of diversity.

- For a multiple antenna system with M transmit and N receive antennas, the maximal diversity gain is MN. This corresponds to the degree at which the error probability of a ML detector decays at high SNR.

Space-Time Coding

- Transmit diversity in multiple-antenna systems can be exploited by a coding scheme called space-time coding, which is a joint design of error-control coding, modulation, and transmit diversity.

Block Fading Model

- The block fading model assumes that the channel coefficients remain constant for T symbol periods before changing to a new independent realization.

- For a system with M transmit antennas and M receive antennas, the system model is given by

$$\mathbf{Y} = \sqrt{\frac{\rho T}{M}}\mathbf{SH} + \mathbf{W},$$

  where $\mathbf{S}$ is the matrix of transmitted signals, $\mathbf{H}$ is the matrix of fading coefficients, $\mathbf{W}$ is the matrix of the additive noise and $\mathbf{Y}$ is the matrix of received signals. With this identification, ρ is the SNR at each receive antenna.

Space-Time Coding and DMT

- In order to achieve the maximum diversity gain under the slow fading assumption, one needs to communicate at a fixed rate R, which becomes vanishingly small compared to the fast fading capacity at high SNR. Thus, one is actually sacrificing all the spatial multiplexing benefit of the MIMO channel to maximize the reliability.

- In order to reclaim some of that benefit, one should instead communicate at a rate $R = r \log_2 SNR$, which is a fraction of the fast fading capacity.

- A diversity gain $d^*(r)$ is achieved at multiplexing gain r if $R = r \log_2 SNR$ and $P_o(R) \approx SNR^{-d^*(r)}$. More precisely, the curve

$$d^*(r) = -\lim_{SNR\to\infty} \frac{\log_2 P_o(r \log_2 SNR)}{\log_2 SNR},$$

  gives the diversity-multiplexing tradeoff in slow fading channels.

- A space-time coding scheme attains a multiplexing gain r and diversity gain d if for the error probability it holds

$$\lim_{SNR\to\infty} \frac{P_e}{\log_2 SNR} = -d.$$

Chapter 2

Non-Coherent Space-Time Coding

In this chapter we summarize some of the main results in the field of non-coherent space-time coding and present the basics concepts of non-coherent transmission over block Rayleigh fading channels. Since the accent in this thesis is on the geometric interpretation of the communication problem, we also present the geometric preliminaries necessary to understand the properties of Grassmann manifolds and the code constructions that follow in the next chapters.

2.1 Introduction

If the time between signal fades is sufficiently long, then the transmitter can send training signals that allow the receiver to estimate the propagation coefficients accurately. With a mobile receiver, however, the time between fades may be too short to permit reliable estimation of the coefficients. A 100 km/h mobile operating at 1.9 GHz has a fading interval of about 3 ms, which for a symbol rate of 30 ksymb/sec corresponds to only about 100 symbol periods [31].

Besides the fact that the fading interval might be too short, the number of coefficients itself might be large and thus difficult to estimate. For example, for a 4×4 channel this means 16 impulse responses. Beside the overhead of preambles or test

sequences for measuring and tracking, there also might be unavoidable inaccuracies, which decrease the potential gain of the MIMO channel. Therefore it is straightforward to look for the alternative where no channel knowledge is available at all.

The main principles of non-coherent communication over block fading channels were presented by Marzetta and Hochwald in [31]. The main result in [31] states that for a MIMO communication system with M transmit antennas and N receive antennas operating non-coherently over the block-fading channel with block length T, the generic form of the input signals that enable communication at rates approaching the non-coherent ergodic capacity can be expressed as the product of an isotropically distributed $T \times M$ random unitary matrix and a diagonal $M \times M$ matrix $\mathbf{D}$ with real nonnegative entries. In [65] Zheng and Tse developed a geometric framework and derived the capacity of the non-coherent block fading MIMO channel for different number of transmit and receive antennas. They interpreted the problem of non-coherent space-time coding as sphere packing problem in Grassmann manifolds. These reults triggered further research relating the geometric properties of Grassmann manifolds with the problem of space-time code construction for the non-coherent channel. In this spirit sphere packing bounds were presented in [4]. Henkel further found tighter bounds in [16]. It is interesting to note that, besides for the problem of non-coherent space-time coding, Grassmann were addressed in the literature in other context, for example for solving optimization problems [10], where detailed introduction to the geometry of Grassmann manifolds is presented. Additionally, Grassmann manifolds find application in the problem of communication over channels with feedback, where the channel information fed back to the transmitter is quantized and represented by a set of elements of the Grassmann manifold [29]. The elements are chosen according to a criterion which maximizes a certain metric in the Grassmann manifold.

In the following we summarize the main results in this topic and set the principles for the construction of codes based on the geometric methods.

2.2 System Model

The channel model is the block fading model initiated by Marzetta and Hochwald [31], presented in 1.2. The channel coefficients are i.i.d. and remain constant for T symbol periods before changing to a new independent realization. The number of transmit antennas is M and the number of receive antennas is N. The model is given by

$$\mathbf{Y} = \sqrt{\frac{\rho T}{M}}\mathbf{\Phi H} + \mathbf{W}, \tag{2.1}$$

where $\mathbf{\Phi}$ is a $T \times M$ matrix of transmitted signals, $\mathbf{H}$ is a $M \times N$ matrix of fading coefficients, i.i.d complex Gaussian, $CN(0,1)$ and $\mathbf{W}$ is a $T \times N$ matrix of the additive noise, also i.i.d complex Gaussian, $CN(0,1)$. $\mathbf{Y}$ is a $T \times M$ matrix of received signals. With this identification, ρ is the SNR at each receive antenna.

An insight generated by the concept of coordinate change arises from the observation that the fading channel matrix does not change the subspace in which the transmitted signal resides but it merely rotates and scales the bases of this subspace [65]. However, the combined effect of noise and fading results in the perturbation of the signal subspace in a specific manner. Based on this geometric insight, it was shown in [65] that, at high SNR, the information carrying object is a linear subspace. That is, information about the transmitted data is contained in the subspace of the received signal and the particular orientation of the received signal vector within the subspace is "informationless". These observations suggest that, for the non-coherent channel, spectrally efficient signalling at high SNR requires the design of the bases of a set of linear signal subspaces rather than the design of the actual signal values. For M transmit antennas and coherence time T, we can say that the relevant coding space is the set of M-dimensional linear subspaces of the, in general, complex T-dimensional space. This set can be given a structure of a manifold and is known as the Grassmann manifold $G^{\mathbb{C}}_{T,M}$. The differential structure of the Grassmann manifold yields non-coherent space-time codes based on sets in the tangent bundle of the manifold , with connection given by the exponential map.

An example of how the channel acts on the transmit signal is given in Fig. 2.1. Let $\mathbf{s}_1, \mathbf{s}_2$ be two basis vectors of the two-dimensional subspace $\mathbf{\Omega_S}$ of $\mathbb{R}^3$. We can see that the channel rotates and scales the basis veciors of the two-dimensional subspace. However, the resulting vectors still span the same subspace.

2.3 Capacity Analysis and Geometric Interpretation

A detailed analysis of the capacity of non-coherent block fading channels is given in [65]. Here, without going into the full details of the capacity derivation, we will point out that the derivation is based on the change of the coordinate system, where the differential entropies involved in the capacity computation are calculated. Namely, the transmit matrix $\mathbf{\Phi}$ can be represented by the subspace spanned by its column vectors, $\mathbf{\Omega_\Phi}$ and an $M \times M$ matrix $\mathbf{C_\Phi}$ which specifies the column vectors of $\mathbf{\Phi}$ with respect to a canonical basis in $\mathbf{\Omega_\Phi}$. The motivation is the following. The channel $\mathbf{H}$ acts on the transmit matrix $\mathbf{\Phi}$ by scaling and linearly combining its columns. With this, the M-dimensional subspace of the complex space $\mathbb{C}^T$ spanned by the columns of $\mathbf{\Phi}$ does not change after the multiplication by the channel matrix. In other words, we can send information over the unknown channel by simply sending information

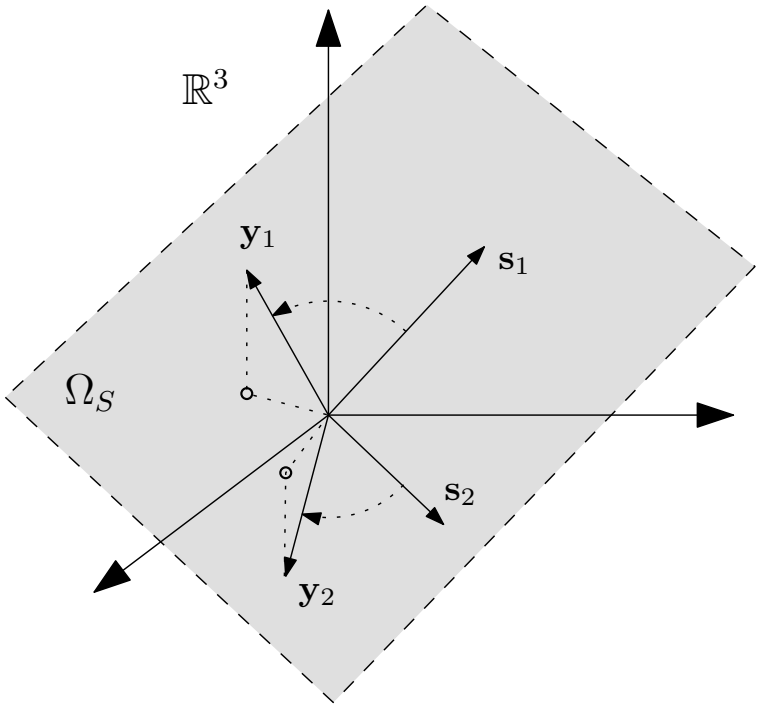

Figure 2.1: Rotation and scaling within the same linear subspace

about a subspace of the complex space $\mathbb{C}^T$ spanned by the columns of $\mathbf{\Phi}$. This naturally defines our space of transmit signals, or our coding space, to be the set of all M-dimensional linear subspaces of $\mathbb{C}^T$. This set has a structure of a manifold and is known as the Grassmann manifold $G_{T,M}^{\mathbb{C}}$.

We have already argued that the transmit matrix $\mathbf{\Phi}$ can be represented by the subspace spanned by its column vectors, $\mathbf{\Omega_\Phi}$ and an $M \times M$ matrix $\mathbf{C_\Phi}$. This motivates the transformation

$$\mathbf{\Phi} \rightarrow (\mathbf{C_\Phi}, \mathbf{\Omega_\Phi}) \tag{2.2}$$

which is a change of coordinate system $\mathbb{C}^{T\times M} \rightarrow \mathbb{C}^{M\times M} \times G_{T,M}^{\mathbb{C}}$. With this, the subspaces spanned by $\mathbf{\Phi}$ and $\mathbf{\Phi H}$, $\mathbf{\Omega_\Phi}$ and $\mathbf{\Omega_{\Phi H}}$ are identical.

The high SNR capacity (b/s/Hz) of this channel is [65]

$$C_{M,M} = M\left(1 - \frac{M}{T}\right)\log_2 \rho + c(M,M) + o(1) \tag{2.3}$$

where

$$c(M,M) = \frac{1}{T}\log_2 |G_{T,M}^{\mathbb{C}}| + M\left(1 - \frac{M}{T}\right)\log_2 \frac{T}{M\pi e} \quad + \left(1 - \frac{M}{T}\right) E\left[\log_2 \det(\mathbf{H}\mathbf{H}^H)\right], \tag{2.4}$$

is the term which does not depend on the SNR and $o(1)$ is a term which tends to 0, when $\rho \to \infty$. $|G^{\mathbb{C}}_{T,M}|$ is the volume of the Grassmann manifold $G^{\mathbb{C}}_{T,M}$ and appears in the capacity expression due to the coordinate transformation. Additionally, we have

$$E\left[\log_2 \det(\mathbf{H}\mathbf{H}^{\mathrm{H}})\right] = \sum_{i=1}^{M} \mathrm{E}\log_2 \chi^2_{2i}, \tag{2.5}$$

where χ^2_{2i} is a Chi-square randmom variable of degree $2i$. The pre-log factor $M\left(1-\frac{M}{T}\right)$ is the number of the degrees of freedom and plays a key role in the capacity at high SNR.

For the non-coherent channel where the fading coefficients are unknown, we can interpret the capacity by sphere packing in the Grassmann manifold. Since the subspace $\mathbf{\Omega}_{\Phi}$ is the object that we use to convey information, we view the transmitted signal in each coherence interval as a point in the Grassmann manifold $G^{\mathbb{C}}_{T,M}$. The channel matrix $\mathbf{H}$ scales the volume to be $\det(T\mathbf{H}\mathbf{H}^{\mathrm{T-M}})|G^{\mathbb{C}}_{T,M}|$. With codewords of length l, the received signal lies in the product space of l scaled Grassmann manifolds, with complex dimension $M(T-M)l$. The noise perturbs the signal in the sphere $S^{M(T-M)l}(\sqrt{M(T-M)l\sigma^2})$. If we denote as $\mathbf{H}_i$ the fading coefficient matrix in coherence interval i, in the high SNR region we can write the ratio of the two volumes

$$q = \frac{\prod_{i=1}^{l} \det(T\mathbf{H}_i\mathbf{H}_i^{\mathrm{H}})^{T-M}|G^{\mathbb{C}}_{T,M}|}{Vol(S^{M(T-M)l}(\sqrt{M(T-M)l\sigma^2}))}. \tag{2.6}$$

Using the formula for sphere volume and Stirling approximation [65], for the normalized high SNR capacity we have

$$\begin{aligned} C_{M,M} &= \frac{1}{T}\frac{l}{\log_2 q} \to \mathrm{E}\left[\log_2 \det(T\mathbf{H}\mathbf{H}^{\mathrm{H}})\right] + \log_2 |G^{\mathbb{C}}_{T,M}| - M\left(1-\frac{M}{T}\right)\log_2 \pi e\sigma^2 \\ &= M(1-\frac{M}{T})\log_2 \rho + c_{M,M}, \end{aligned} \tag{2.7}$$

which is exactly the high SNR capacity (without the $o(1)$ which vanishes at high SNR in (2.3).

We see that, similarly to the case of the AWGN channel, the geometric interpretation delivers the capacity of the non-coherent channel in the high SNR regime.

2.4 Differential Geometry Preliminaries

Here we basically review the properties of the Grassmann manifolds and present a basic introduction to the geometry of Grasmann manifolds. The approach and the definitions mainly follow [10], [16] and [6], where more detailed introduction to

Grassmann manifolds is presented. Here we mainly summarize some properties necessary to understand the problem of code construction in Grassmann manifolds and present the preliminaries necessary to understand the following chapters. A more detailed introduction to the topic is givven in the appendix.

2.4.1 Grassmann Manifolds

We can think of a manifold as a topological space which is second countable (has a countable topological basis) and is locally Euclidean, i.e. every point on the manifold has a neighborhood which is topologically equivalent (homeomorphic) to an open ball in R^n [6]. In this case the manifold is said to have a dimension n. In this context, the collection (set) of all M-dimensional linear subspaces of $\mathbb{C}^T$ can be given a structure of a manifold, which is known as the (complex) Grassmann manifold $G^{\mathbb{C}}_{T,M}$.

The Grassmann manifold stands in close connection to the Stiefel manifold. The (complex) Stiefel manifold $V^{\mathbb{C}}_{M,T}$ is the set of M orthonormal vectors in $\mathbb{C}^T$

$$V^{\mathbb{C}}_{M,T} := \{\mathbf{\Phi} \in \mathbb{C}^{T\times M} | \mathbf{\Phi}^H \mathbf{\Phi} = I_M\}. \tag{2.8}$$

The (complex) Grassmann manifold is formally defined as

$$G^{\mathbb{C}}_{T,M} := \{\langle \mathbf{\Phi} \rangle |\ \mathbf{\Phi}^{\mathrm{H}} \mathbf{\Phi} = \mathbf{I}_M\}, \tag{2.9}$$

where $\langle \mathbf{\Phi} \rangle$ denotes the subspace spanned by the columns of $\mathbf{\Phi}$ and $\mathbf{I}_M$ is the $M \times M$ identity matrix. We can think of the elements of the Grassmann manifold as an equivalent class of elements of the Stiefel manifold. In other words, each element of the Grassmann manifold corresponds to a set of elements of the Stiefel manifold whose columns spane the same subspace. There is a transitive action by the unitary group $\mathrm{U}(T)$ (group of unitary matrices) on $G^{\mathbb{C}}_{T,M}$ [6], [10], with isotropy group

$$\mathrm{H} = \begin{pmatrix} \mathrm{U}(M) & \mathbf{0} \\ \mathbf{0} & \mathrm{U}(T-M) \end{pmatrix}. \tag{2.10}$$

This justifies the quotient space representation [10]

$$G^{\mathbb{C}}_{T,M} \cong \mathrm{U}(T) / \begin{pmatrix} \mathrm{U}(M) & \mathbf{0} \\ \mathbf{0} & \mathrm{U}(T-M) \end{pmatrix}. \tag{2.11}$$

Intuitively, a quotient space is obtained from a topological space G when we define an equivalence relation on the space G. In the case of the Grassmann manifold, the equivalence relation is given by the requirement that points from G are equivalent if they span the same subspace. In other words, each element of the Grassmann manifold can be considered as an equivalent class of the unitary group $\mathrm{U}(T)$ consisting

of unitary matrices whose columns span the same subspace. This leads to straightforward calculation of the (real) dimension of the Grassmann manifold as

$$\dim_{\mathbb{R}} G^{\mathbb{C}}_{M,T} = T^2 - M^2 - (T-M)^2 = 2M(T-M), \tag{2.12}$$

since the dimension of the unitary group of $n \times n$ unitary matrices is $\dim_{\mathbb{R}} \mathrm{U}(n) = n^2$.

Let $\langle \mathbf{\Phi} \rangle, \langle \mathbf{\Psi} \rangle \in G^{\mathbb{C}}_{T,M}$ be two subspaces of $\mathbb{C}^T$ and let $\mathbf{\Psi}^{\mathrm{H}}\mathbf{\Phi} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}, \mathbf{U}, \mathbf{\Sigma}, \mathbf{V} \in \mathbb{C}^{M \times M}$ be the singular value decomposition of $\mathbf{\Psi}^{\mathrm{H}}\mathbf{\Phi}$, with $\mathbf{\Sigma}$ being a diagonal matrix of the singular values $\sigma_1, \ldots, \sigma_M$. Then, M principle angles $\theta_1, \ldots, \theta_M$ between $\langle \mathbf{\Phi} \rangle$ and $\langle \mathbf{\Psi} \rangle$ can be defined as

$$\theta_i = \operatorname{acos} \sigma_i. \tag{2.13}$$

A metric called geodesic distance can be defined by the set of principle angles between two subspaces [10]. The metric represents the length of the geodesic connecting two points of the Grassmann manifold. For $\langle \mathbf{\Phi} \rangle, \langle \mathbf{\Psi} \rangle \in G^{\mathbb{C}}_{T,M}$, the geodesic distance is given by

$$\mathrm{d_g}(\langle \mathbf{\Phi} \rangle, \langle \mathbf{\Psi} \rangle) = \sqrt{\textstyle\sum_{i=1}^{M} \theta_i^2}. \tag{2.14}$$

There is an embedding of the Grassmann manifold $G^{\mathbb{C}}_{T,M}$ in Euclidean space which is a result of associating elements of the Grassmann manifold with their projection matrices [7], [44] . For $\langle \mathbf{\Phi} \rangle \in G^{\mathbb{C}}_{M,T}$, there is an associate orthogonal projection map

$$\mathbf{P}_{\mathbf{\Phi}} := \mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}}, \ \mathbb{C}^T \to \langle \mathbf{\Phi} \rangle. \tag{2.15}$$

$\mathbf{P}_{\mathbf{\Phi}}$ is idempotent and hermitian and $\mathrm{tr}(\mathbf{P}_{\mathbf{\Phi}}) = M$. Thus $\mathbf{P}_{\mathbf{\Phi}}$ lies in a space of real dimension $T^2 - 1$. Additionally,

$$\left\| \mathbf{P}_{\mathbf{\Phi}} - \frac{T}{M}\mathbf{I}_T \right\|_{\mathrm{F}} = \sqrt{M(T-M)/T}, \tag{2.16}$$

which justifies the embedding

$$\begin{aligned} G^{\mathbb{C}}_{T,M} &\to S^{T^2-2}\left(\sqrt{M(T-M)/T}\right) \subset \mathbb{R}^{T^2-1}, \\ \langle \mathbf{\Phi} \rangle &\to \mathbf{P}_{\mathbf{\Phi}} - \frac{T}{M}\mathbf{I}_T. \end{aligned} \tag{2.17}$$

This justifies the definition of a topological "chordal" metric on the Grassmann manifold

$$\mathrm{d}_c(\langle \mathbf{\Phi} \rangle, \langle \mathbf{\Psi} \rangle) = \frac{1}{\sqrt{2}} \|\mathbf{P}_{\mathbf{\Phi}} - \mathbf{P}_{\mathbf{\Psi}}\|_F = \sqrt{\textstyle\sum_{i=1}^{M} \sin^2 \theta_i}. \tag{2.18}$$

In order to distinguish between an element of the Grassmann manifold and the particular matrix representation, the notation $\langle \mathbf{\Phi} \rangle$ was used to denote the element

(subspace) of the Grassmann manifold spanned by the columns of $\mathbf{\Phi}$. In order to simplify the notation, from now on we will omit the bracket and will identify the subspace with its matrix representative. However, we will have on mind that this representation is not unique.

2.4.2 Parametrization of the Grassmann Manifold

The unitary group $\mathrm{U}(T)$ is a Lie group, i.e a group which is also a differentiable manifold. Due to the differential structure, a tangent space can be constructed at every element of a Lie group [10], [6]. The tangent space at the identity element of a Lie group has the structure of a Lie algebra. The corresponding Lie algebra $\mathfrak{u}(T)$ of the unitary group $\mathrm{U}(T)$ is the tangent space at the identity element $\mathbf{I}_T$ ($T \times T$ identity matrix), and is given by the set of skew-hermittian matrices [10]

$$\mathfrak{u}(T) := \{\tilde{\mathbf{X}} \in \mathbb{C}^{T\times T} | \tilde{\mathbf{X}} = -\tilde{\mathbf{X}}^{\mathrm{H}}\}. \tag{2.19}$$

Thus, the elements of the Lie algebra $\mathfrak{u}(T)$ have the form

$$\tilde{\mathbf{X}} = \begin{pmatrix} \mathbf{A} & -\mathbf{B}^{\mathrm{H}} \\ \mathbf{B} & \mathbf{C} \end{pmatrix}, \tag{2.20}$$

where $\mathbf{A} = -\mathbf{A}^{\mathrm{H}}, \mathbf{C} = -\mathbf{C}^{\mathrm{H}}$ and $\mathbf{B} \in \mathbb{C}^{(T-M)\times M}$.

We have already represented the Grassmann manifold as a quotient space of the unitary group (2.11). Since the isotropy group H (2.10) is a subgroup of the unitary group $\mathrm{H} \subset \mathrm{U}(T)$ (and thus a Lie group itself), the Lie algebra $\mathfrak{h} \subset \mathfrak{u}(T)$ consists of matrices of the form

$$\mathbf{X}^{\|} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}, \tag{2.21}$$

where $\mathbf{A} \in \mathfrak{u}(M), \mathbf{C} \in \mathfrak{u}(T-M)$. Thus, $\mathfrak{u}(T)$ can be decomposed as $\mathfrak{u}(T) = \mathfrak{h} \cup \mathfrak{h}^{\perp}$, where the elements of $\mathfrak{h}^{\perp}$ have the form

$$\mathbf{X} = \begin{pmatrix} \mathbf{0} & -\mathbf{B}^{\mathrm{H}} \\ \mathbf{B} & \mathbf{0} \end{pmatrix}. \tag{2.22}$$

Due to the quotient space representation (2.11), tangents of $G^{\mathbb{C}}_{T,M}$ at the identity element $\mathbf{I}_{T,M} = \begin{pmatrix} \mathbf{I}_M \\ \mathbf{0} \end{pmatrix}$ are provided by the space $\mathfrak{h}^{\perp}$. Obviously, the tangent space is a vector space of dimension $D = 2M(T-M)$, which was to be expected, since its dimension should match the dimension of the Grassmann manifold $G^{\mathbb{C}}_{T,M}$, as in (2.12).

The Lie algebra determines the local structure of a Lie group via the *exponential map*. In the case of the unitary group $\mathrm{U}(T)$, the exponential map maps the Lie algebra

$\mathfrak{u}(T)$ to $\mathrm{U}(T)$. For a matrix Lie group, the exponential map coincides with the matrix exponential and is given by the ordinary series expansion: $\exp(\mathbf{X}) = \sum_{k=0}^{\infty} \frac{\mathbf{X}^k}{k!}$.

Thus, a point in the Grassmann manifold $\mathbf{\Phi}$ is obtained from a point $\mathbf{X}$ in the tangent space as

$$\mathbf{\Phi} = \exp(\mathbf{X})\mathbf{I}_{T,M}. \tag{2.23}$$

This relation provides a mapping from the tangent space to the manifold, and serves for construction of Grassmann codes from sets in the tangent space.

The exponential map itself is computationally inefficient. Fortunately, the representation of the tangents in the form (2.22), provides efficient computation of the exponential map. Given the tangent (2.22), the singular value decomposition of $\mathbf{B} \in \mathbb{C}^{(T-M)\times M}$, reads [16], [10]

$$\mathbf{B} = \mathbf{V}\mathbf{\Sigma}\mathbf{W}^{\mathrm{H}}, \tag{2.24}$$

where $\mathbf{V} \in \mathbb{C}^{(T-M)\times M}$ and has orthonormal columns, $\mathbf{\Sigma}$ is the matrix of singular values of $\mathbf{B}$ in decreasing order, and $\mathbf{W} \in \mathrm{U}(M)$. If we denote $\mathbf{\Phi} = \exp(X)\mathbf{I}_{T,M}$, it can be shown that [16], [10]

$$\mathbf{\Phi} = \begin{pmatrix} \mathbf{W}\mathbf{C}\mathbf{W}^{\mathrm{H}} \\ \mathbf{V}\mathbf{S}\mathbf{W}^{\mathrm{H}} \end{pmatrix}, \tag{2.25}$$

where $\mathbf{C} = \cos(\mathbf{\Sigma})$ and $\mathbf{S} = \sin(\mathbf{\Sigma})$ are diagonal matrices (sin and cos operate only on the diagonal of $\mathbf{\Sigma}$).

2.4.3 Performance Analysis and Design Criteria

For our purpose, we need to analyze the cases of both the unknown as well the known channel. This is due the fact that in some of the constructions methods, we will use codes designed for the coherent channel as basis for the construction of codes for the non-coherent channel. Therefore, it is important to summarize the design criteria for both cases.

We will use the Chernoff bound on the pairwise error probability as a starting point in the performance analysis. We will use a compact notation and represent equations which hold for both cases. Let $\mathbf{\Phi}, \mathbf{\Psi}$ are two elements of the Stiefel manifold $V_{M,T}^{\mathcal{C}}$, respectively the Grassmann manifold $G_{M,T}^{\mathcal{C}}$. The Chernoff bound on the PEP (Pairwise Error Probability) of mistaking $\mathbf{\Phi}$ for $\mathbf{\Psi}$ can be derived from [19] and gives

$$P(\mathbf{\Phi}, \mathbf{\Psi}) \leq \frac{1}{2}\left[\prod_{i=1}^{M}\left(1 + \varrho\sigma_i^2\right)\right]^{-N}, \tag{2.26}$$

where for the known channel we use the following notation

$$\boldsymbol{\Delta} = \overline{\boldsymbol{\Delta}} \triangleq \boldsymbol{\Phi} - \boldsymbol{\Psi} \tag{2.27}$$

$$\varrho = \bar{\varrho} \triangleq \frac{(\rho \frac{T}{M})^2}{4} \tag{2.28}$$

$$\sigma_i = \overline{\sigma}_i \triangleq \sigma_i(\overline{\boldsymbol{\Delta}}), \tag{2.29}$$

and for the unknown channel

$$\boldsymbol{\Delta} = \underline{\boldsymbol{\Delta}} \triangleq \boldsymbol{\Psi}^{\mathrm{H}} \boldsymbol{\Phi} \tag{2.30}$$

$$\varrho = \underline{\varrho} \triangleq \frac{(\rho \frac{T}{M})^2}{4(1 + \rho \frac{T}{M})} \tag{2.31}$$

$$\sigma_i = \sqrt{1 - \underline{\sigma}_i^2} \triangleq \sqrt{1 - \sigma_i^2(\underline{\Delta})}. \tag{2.32}$$

As a performance criterion we usually take the (pairwise) diversity defined as

$$Div(\boldsymbol{\Phi}, \boldsymbol{\Psi}) := \prod_{i=1}^{M} \left(1 + \varrho \sigma_i^2\right), \tag{2.33}$$

which can be written in the form

$$Div = \sum_{i=1}^{M} s_i \varrho^i. \tag{2.34}$$

The first and the last term of the expression are respectively called *diversity sum* and *diversity product*. The diversity sum is $d = \sum_{i=1}^{M} \sigma_i^2$, which for the known channel gives

$$\overline{d} = \|\overline{\Delta\Delta}^{\mathrm{H}}\|, \tag{2.35}$$

and for the unknown channel

$$\underline{d} = \sqrt{\mathrm{tr}(1 - \underline{\boldsymbol{\Delta}\boldsymbol{\Delta}}^{H})} = \sqrt{\textstyle\sum_{i=1}^{M}(1 - \underline{\sigma}_i^2)} = \textstyle\sum_{i=1}^{M} \sin^2 \theta_i, \tag{2.36}$$

which follows from (2.13). In the case of the known channel, the diversity sum coincides with the metric induced by the ML receiver, i.e. the Frobenius norm (Euclidean distance) between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. In the case of the unknown channel, the diversity sum also coincides with the ML receiver metrics, i.e equals the chordal distance between the subspaces $\langle \boldsymbol{\Phi} \rangle$ and $\langle \boldsymbol{\Psi} \rangle$. The diversity product is $p = \sigma_1 \sigma_2 \cdots \sigma_M$, which for the known channel gives

$$\overline{p} = \det(\overline{\boldsymbol{\Delta}\boldsymbol{\Delta}}^{\mathrm{H}}), \tag{2.37}$$

and for the unknown

$$\underline{p} = \det(1 - \underline{\mathbf{\Delta}}\underline{\mathbf{\Delta}}^{\mathrm{H}}) = \sqrt{\prod_{i=1}^{M}(1-\sigma_i^2)} = \sqrt{\prod_{i=1}^{M}\sin^2\theta_i}. \tag{2.38}$$

The diversity product has also an interpretation. In the case of the known channel it is the well-known determinant criterion [19]. In the literature addressing the topic of space-time block coding, there are two criteria which are used when designing space-time codes for the known channel. These are the rank and the determinant criterion, shortly summarized as:

The Rank Criterion: The minimum rank r of $\overline{\mathbf{\Delta}} = \mathbf{\Phi} - \mathbf{\Psi}$ taken over all distinct pairs $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is the diversity gain and should be maximized.

The Determinant Criterion: The minimum of the product $(\prod_{i=1}^{M}\overline{\sigma}_i^2)^{\frac{1}{M}}$ over all $\mathbf{\Phi}, \mathbf{\Psi}$ should be maximized. This is the coding gain.

In the case of the unknown channel the diversity product measures the non-negativity of the principle angles between $\langle\mathbf{\Phi}\rangle$ and $\langle\mathbf{\Psi}\rangle$ [16]. If we concentrate on the diversity sum and the diversity product, the design of codes for the non-coherent MIMO channel would involve searching for good packings in the Grassmann manifold with respect to the chordal distance, while trying to keep each principle angle as large as possible [16].

In the next chapter we will present code constructions for the non-coherent channel based on packings in Grassmann manifolds. We will mainly concentrate on the diversity sum (chordal distance) and the diversity product as the most relevant design criteria.

2.5 Decoding of Non-Coherent Space-Time Codes

According to the system model (2.1), the received matrix is zero-mean isotropically distributed Gaussian random matrix. Hence, for the conditional pdf $P(\mathbf{Y}|\mathbf{\Phi})$ we have [31]

$$\begin{aligned} P(\mathbf{Y}|\mathbf{\Phi}) &= \frac{\exp\left(-\mathrm{Tr}\left(\mathbf{Y}^{\mathrm{H}}\left(\frac{M}{\rho T}\mathbf{I}_T + \mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}}\right)^{-1}\mathbf{Y}\right)\right)}{\pi^{TN}\det^{N}\left(\frac{M}{\rho T}\mathbf{I}_T + \mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}}\right)}, \\ &= \frac{\exp\left(-\frac{\rho T}{M}Tr\left(\mathbf{Y}^{\mathrm{H}}\left(\mathbf{I}_T - \frac{1}{1+M/\rho T}XX^{\mathrm{H}}\right)\mathbf{Y}\right)\right)}{(\pi M/\rho T)\,(1+\rho T/M)^{MN}} \end{aligned} \tag{2.39}$$

A maximum likelihood detector tests the entire codebook, C, in search for the codeword $\mathbf{\Phi}$ that maximizes $P(\mathbf{Y}|\mathbf{\Phi})$

$$\hat{\mathbf{\Phi}} = \arg\max_{\mathbf{\Phi}\in C}\mathrm{Tr}\left(\mathbf{Y}^{\mathrm{H}}\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}}\mathbf{Y}\right) = \arg\min_{\mathbf{\Phi}\in C}\|\mathbf{Y}^{\mathrm{H}}\mathbf{\Phi}\|_{\mathrm{F}}^2. \tag{2.40}$$

We can see that the ML-decoding rule involves projection of the received matrix $\mathbf{Y}$ on all possible transmit subspaces.

The main drawback associated with the ML detector is the computational cost of having to examine all possible codewords in the codebook. In order to increase the computational efficiency, much work has concentrated on exploiting the suboptimal decoding at the receiver end. In this work we mostly concentrate on proving the potential of the geometric insight for construction of codes for the non-coherent channel. Less attention is devoted to the search for efficient decoding techniques. However, there are general techniques in the literature which enable decoding of Grassmann codes with reduced complexity, such as the method of Gohary and Davidson described in [13], [12]. Other techniques which shift the decision in the tangents space instead in the manifold can be used as well. One example is given in [1]. For Grassmann codes obtained from spherical codes, the decision may be further shifted to the sphere where the spherical code resides and perform the decoding on the sphere. For codes obtained from lattices, decoding might be performed in the lattice performing efficient decoding techniques developed for lattices. However, it is important to note that these techniques are usually suboptimal, since they require a mapping inverse to the exponential map. We remember that the exponential map and its inverse are non-linear, which affects the code structure. This effect can be partly mitigated by choosing a list of nearest neighbors in the tangent space, respectively on the sphere, which will be additionally tested in the manifold.

These techniques are quite general and can be used for decoding of different Grassmann codes. However, some of the codes proposed in this work have additional structure which can be exploited to further simplify the decoding. We will make a notice on this when we present constructions in the next chapters.

2.6 Chapter Summary

The Concept of Coordinate Change

- The system model for the non-coherent block MIMO channel is given as

$$\mathbf{Y} = \sqrt{\frac{\rho T}{M}}\mathbf{\Phi H} + \mathbf{W}.$$

- The fading channel matrix $\mathbf{H}$ does not change the subspace in which the transmitted signal resides. Thus, at high SNR, the information carrying object is a linear subspace.

- For M transmit antennas and coherence time T, the relevant coding space is the Grassmann manifold $G^{\mathbb{C}}_{T,M}$.

- The differential structure of the Grassmann manifold yields non-coherent space-time codes based on sets in the tangent bundle of the manifold , with connection given by the exponential map.

Capacity Analysis and Geometric Interpretation

- The way the channel acts on the transmit matrix $\mathbf{\Phi}$ motivates the transformation

$$\mathbf{\Phi} \rightarrow (\mathbf{C}_{\Phi}, \mathbf{\Omega}_{\Phi}),$$

  which is a change of coordinate system $\mathbb{C}^{T\times M} \rightarrow \mathbb{C}^{M\times M} \times G^{\mathbb{C}}_{T,M}$.

- The high SNR capacity (b/s/Hz) of this channel is

$$C_{M,M} = M\left(1 - \frac{M}{T}\right)\log_2 \rho + c(M,M) + o(1),$$

  where $M\left(1-\frac{M}{T}\right)$ are the degrees of freedom and $c_{M,M}$ is a term which depends on M and T but not on the SNR ρ.

The Grassmann Manifold $G^{\mathbb{C}}_{T,M}$

- The Grassmann Manifold $G^{\mathbb{C}}_{T,M}$ is a differentiable manifold and tangent space can be constructed at each element of the manifold.

- Tangents at the identity element have the form

$$\mathbf{X} = \begin{pmatrix} \mathbf{0} & -\mathbf{B}^{\mathrm{H}} \\ \mathbf{B} & \mathbf{0} \end{pmatrix},\ \mathbf{B} \in \mathbb{C}^{(T-M)\times M}.$$

- Elements of the Grassmann manifold can be represented by the matrices

$$\mathbf{\Phi} = \begin{pmatrix} \mathbf{WCW}^{\mathrm{H}} \\ \mathbf{VSW}^{\mathrm{H}} \end{pmatrix},$$

  where $\mathbf{C} = \cos(\mathbf{\Sigma})$ and $\mathbf{S} = \sin(\mathbf{\Sigma})$

Performance Analysis and Design Criteria

- As a performance criterion we take the (pairwise) diversity defined as

$$Div(\mathbf{\Phi}, \mathbf{\Psi}) := \prod_{i=1}^{M} \left(1 + \varrho\sigma_i^2\right),$$

which can be written in the form

$$Div = \sum_{i=1}^{M} s_i \varrho^i.$$

- The diversity sum is the first term in the expression and is given as

$$\underline{d} = \sqrt{\sum_{i=1}^{M} (\sin^2 \theta_i)}.$$

where θ_i are the principle angles between the subspaces $\mathbf{\Phi}$ and $\mathbf{\Psi}$.

- The diversity product is the last term in the expression and is the most important criterion at high SNR

$$\underline{p} = \det(1 - \underline{\mathbf{\Delta}\mathbf{\Delta}}^{\mathrm{H}}) = \sqrt{\prod_{i=1}^{M} \sin^2 \theta_i}.$$

Decoding

- The ML decoding rule is given as

$$\hat{\mathbf{\Phi}} = \arg \max_{\mathbf{\Phi} \in C} \mathrm{Tr} \left(\mathbf{Y}^{\mathrm{H}} \mathbf{\Phi} \mathbf{\Phi}^{\mathrm{H}} \mathbf{Y} \right) = \arg \min_{\mathbf{\Phi} \in C} \| \mathbf{Y}^{\mathrm{H}} \mathbf{\Phi} \|_{\mathrm{F}}^2.$$

Chapter 3

Code Constructions for the Non-Coherent Point-to-Point MIMO Channel

In this chapter we present novel code constructions for the point-to-point block fading MIMO channel without channel knowledge assumption. The code constructions are based on the geometric interpretation of the non-coherent space-time coding problem. The constructions exploit the geometry of the Grassmann manifold and its differentiable structure. Parts of this chapter were published in [53], [56], [55], [54] and [49].

3.1 Introduction

As discussed in Chapter 2, for a MIMO communication system with M transmit antennas and N receive antennas operating non-coherently over the block-fading channel with block length T, the generic form of the capacity achieving input signals are in a form of an isotropically distributed $T \times M$ random unitary matrix and a diagonal $M \times M$ matrix $\mathbf{D}$ with real nonnegative entries [31], [65]. While this structure of the input signals is capacity achieving irrespective of the values of the received SNR and the channel coherence time T, the distribution of the entries of the diagonal matrix $\mathbf{D}$

depends on these two factors. For example, it can be shown that at low SNR only one entry of $\mathbf{D}$ is non-zero when the transmitter is active.

In contrast to this, achieving capacity for high SNR scenarios requires the input signals to be in the form of isotropically distributed unitary matrices, provided that T satisfies $T \geq \min(M, N) + N$. In this case, setting $\mathbf{D}$ equal to the identity matrix achieves the high SNR ergodic capacity of the non-coherent channel. By comparing the degrees of freedom supported by the unitary component to those supported by the diagonal matrix $\mathbf{D}$, it can be concluded that even at moderate SNRs most of the information will be carried by the unitary component.

By assuming that the communication system operates in the moderate-to-high SNR region, one can gain insight into the manner in which the coherence time T, affects the achievable data rate. It was shown in [65] and [31] that for given M and N, the capacity of the non-coherent channel approaches that of the coherent one as T grows. From this one can conclude that if T is sufficiently long, the amount of time needed for the receiver to acquire a sufficiently accurate channel model becomes insignificant in comparison with the overall signalling interval. However, compared to the coherent communication model where the training preamble spans over hundreds or thousands of symbols, for non-coherent communication a coherence time of typically several time symbols is assumed.

The moderate-to-high SNR assumption also provides some insight into how the number of transmit antennas should be chosen for a given block length. In particular, given T and N for a system that satisfies $T \geq \min(M, N) + N$, the number of transmit antennas M, required to attain the maximum number of communication degrees of freedom is $M = \min(T/2, N)$ [65]. Under these assumptions, at moderate-to-high SNRs, unitary signalling is not only optimum from a capacity perspective, but it also possesses desirable performance characteristics.

In addition to unitary signalling, a few alternative approaches to non-coherent space-time coding for MIMO channels are available. Differential schemes which assume that the channel variation within two consecutive blocks is negligible were proposed in [20]. However, this assumption is not sufficiently accurate under the current block fading model, in which there is an independent channel realization for each block. In order to suit operation over independent block fading channels, signalling techniques that allow a portion of the coherence time to be used for training have been developed in [31]. These schemes comprise two phases: a training phase and a coherent communication phase. During the training phase the transmitter sends pilot symbols which are used by the receiver to estimate the channel. In order for the receiver to acquire a reasonably accurate model of the channel, the training phase must occupy a number of channel uses that is at least as large as the number of transmit antennas. Assuming that the channel estimate obtained during the training phase is sufficiently accurate; the receiver then switches to a coherent mode of operation

in which the remaining channel coherence time is used to detect the transmitted information coherently. Although training-based schemes were shown to achieve the maximum number of degrees of freedom available for communication at high SNR [65], they are still short of attaining the full channel capacity which involves an SNR independent term. This term can be particularly significant when a large number of transmit antennas is employed.

Guided by the results in [31] and [65], we will consider communication over the non-coherent channel through the design of signal constellations that directly mimic the high SNR capacity achieving isotropic distribution. For the construction, we will use the geometric insight developed in [65].

3.2 Geometric Construction of Grassmann Codes

For M transmit antennas and coherence time T, we already argued that the relevant coding space is the set of M-dimensional linear subspaces of the, in general, complex T-dimensional space, i.e. the Grassmann manifold $G^{\mathbb{C}}_{T,M}$. The Grassmann manifold is a quotient space of the unitary group and thus inherits its geometry [6]. The Grassmann manifold is thus a differentiable manifold and a tangent space can be constructed at each element (point) of the manifold. The differentiable structure of the Grassmann manifold yields non-coherent space-time codes based on sets in the tangent bundle of the manifold, with connection given by the exponential map. This was first noticed by Henkel in [15] and Kammoun et. al. in [26]. In the following we further exploit the geometric insight and present novel code constructions for the non-coherent MIMO channel which rely on the geometric interpretation.

3.2.1 Mapping from the Tangent Space

The differentiable structure of the Grassmann manifolds provides parameterization of the manifold with the tangent space at the identity $\mathbf{I}_{T,M}$. As already discussed, tangents of $G^{\mathbb{C}}_{T,M}$ at $\mathbf{I}_{T,M}$ have the form

$$\mathbf{X} = \begin{pmatrix} \mathbf{0} & -\mathbf{B}^{\mathrm{H}} \\ \mathbf{B} & \mathbf{0} \end{pmatrix}, \ \mathbf{B} \in \mathbb{C}^{(T-M)\times M}.$$

We observe that the element $\mathbf{X}$ in the tangent space is uniquely described by the matrix $\mathbf{B}$. Hence, sometimes we will equivalently refer to $\mathbf{B}$ as element from the tangent space.

If we choose a code in the tangent space, $\mathcal{X} = \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ of elements of the form (2.22), a Grassmann code, $\mathcal{C} \subset G^{\mathbb{C}}_{M,T}$ is constructed as

$$\mathcal{C} = \exp(\mathcal{X})\mathbf{I}_{T,M}. \tag{3.1}$$

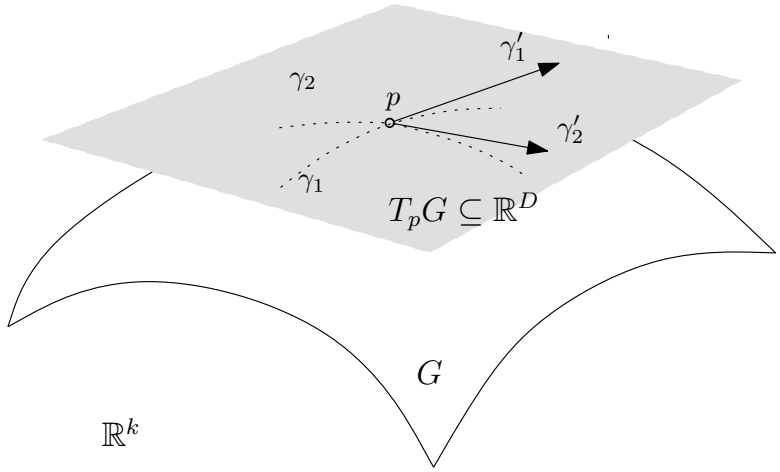

Figure 3.1: Representation of a tangent space of the manifold

As discussed, the exponential map can be performed in an efficient way. According to (2.25), given the SVD of $\mathbf{B}$

$$\mathbf{B} = \mathbf{V}\mathbf{\Sigma}\mathbf{W}^{\mathrm{H}},$$

where $V \in \mathbb{C}^{(T-M)\times M}$ and has orthonormal columns, $\mathbf{\Sigma}$ is the matrix of singular values of $\mathbf{B}$ in decreasing order, and $\mathbf{W} \in U(M)$, a codeword $\mathbf{\Phi} \in \mathcal{C}$, is given as

$$\mathbf{\Phi} = \begin{pmatrix} \mathbf{W}\mathbf{C}\mathbf{W}^{\mathrm{H}} \\ \mathbf{V}\mathbf{S}\mathbf{W}^{\mathrm{H}} \end{pmatrix}, \tag{3.2}$$

where $\mathbf{C} = \cos(\mathbf{\Sigma})$ and $\mathbf{S} = \sin(\mathbf{\Sigma})$.

This parametrization is the starting point for the construction of Grassmann codes. Grassmann codes can be obtained from any subset of the tangent space at the identity. However, one has to take into account the effect of the exponential map when analyzing the codes obtained by this method. In this sense, the Grassmann codes constructed in this way have to be investigated in terms of the known design criteria for codes for the non-coherent channel.

3.2.2 Code Properties after the Mapping

Here we will analyze the properties of Grassmann codes obtained by mapping constellations from the tangent space.

Let $\mathbf{B}_1, \mathbf{B}_2 \in \mathcal{B}$ be two points (elements) in the tangent space and let $\overline{\mathbf{\Delta}} = \mathbf{B}_1 - \mathbf{B}_2$. Note that we keep the same notation for $\overline{\mathbf{\Delta}}$ as in the case of the performance analysis of the coherent channel. Actually, as we will see, for the construction of Grassmann

codes for the non-coherent channel, we will usually use codes for the coherent channel which reside in the tangent space of the Grassmann manifold, which justifies the notation.

Now, let $\mathbf{\Phi}_1, \mathbf{\Phi}_2 \in \mathcal{C}$ be the corresponding non-coherent codewords obtained by the exponential map of $\mathbf{B}_1$ and $\mathbf{B}_2$. Then $\mathbf{\Phi}_1 = \begin{pmatrix} \mathbf{W}_1\mathbf{C}_1\mathbf{W}_1^{\mathrm{H}} \\ \mathbf{V}_1\mathbf{S}_1\mathbf{W}_1^{\mathrm{H}} \end{pmatrix}$ and $\mathbf{\Phi}_2 = \begin{pmatrix} \mathbf{W}_2\mathbf{C}_2\mathbf{W}_2^{\mathrm{H}} \\ \mathbf{V}_2\mathbf{S}_2\mathbf{W}_2^{\mathrm{H}} \end{pmatrix}$. If we denote $\underline{\mathbf{\Delta}} = \mathbf{\Phi}_2^{\mathrm{H}}\mathbf{\Phi}_1$, then according to (3.9) follows

$$\underline{\mathbf{\Delta}} = \mathbf{W}_2\mathbf{C}_2\mathbf{W}_2^{\mathrm{H}}\mathbf{W}_1\mathbf{C}_1\mathbf{W}_1^{H} + \mathbf{W}_2\mathbf{S}_2\mathbf{V}_2^{\mathrm{H}}\mathbf{V}_1\mathbf{S}_1\mathbf{W}_1^{\mathrm{H}}. \tag{3.3}$$

According to (2.33), the (pairwise) diversity is given as

$$\underline{Div}(\mathbf{\Phi_1}, \mathbf{\Phi_2}) = \prod_{i=1}^{M} \left(1 + \underline{\varrho}\left(1 - \sigma_i^2\left(\underline{\mathbf{\Delta}}\right)\right)\right) \tag{3.4}$$

On the other hand, for the codewords of the code $\mathcal{B}$ the (pairwise) diversity is given as

$$\overline{Div}(\mathbf{B}_1, \mathbf{B}_2) = \prod_{i=1}^{M} \left(1 + \bar{\varrho}\sigma_i^2(\bar{\mathbf{\Delta}})\right) \tag{3.5}$$

where $\underline{\varrho}$ and $\overline{\varrho}$ are given by (2.32) and 2.29 respectively. In general, it is difficult to give the exact link between the singular values of $\overline{\mathbf{\Delta}}$ and $\underline{\mathbf{\Delta}}$. An intuitive conclusion can be drawn when, for example, $\mathbf{\Phi}_2 = \mathbf{I}_{T,M}$ (or, equivalently, $\mathbf{B}_2 = \mathbf{0}$). Then $(\mathbf{B}_1 - \mathbf{B}_2)^{\mathrm{H}}(\mathbf{B}_1 - \mathbf{B}_2) = \mathbf{B}_1^{\mathrm{H}}\mathbf{B}_1$ and from (2.13) follows

$$\arccos\left[\sigma_i(\mathbf{I}_{T,M}^{\mathrm{H}}\mathbf{\Phi}_1)\right] = \sigma_i(\mathbf{B}_1),\ i = 1, \ldots, M \tag{3.6}$$

or, equivalently, the singular values of $\mathbf{B}_1$ are equal to the principle angles between the subspace $\langle\mathbf{\Phi}_1\rangle$ and the referent subspace $\langle\mathbf{I}_{T,M}\rangle$. Therefore, in this case, when (3.5) is maximized, so is (3.4). Although this does not exactly hold $\forall\mathbf{B}_i, \mathbf{B}_j \subset \mathcal{B}$, for codes with higher rates (and thus smaller principle angles), we may assume that a coherent code for which (3.5) is maximized, leads to a non-coherent code for which (3.4) is "fairly" maximized. This is basicaly the main motivation in [26] for construction of non-coherent codes from coherent codes residing on the tangent space of the Grassmann manifold.

3.3 Grassmann Codes from a High-Dimensional Spherical Code

The Grassmann manifold is a normal-homogeneous space [6]. Homogenity is a natural generalization of spherical symmetry [16]. In [15] Henkel proposed a construction

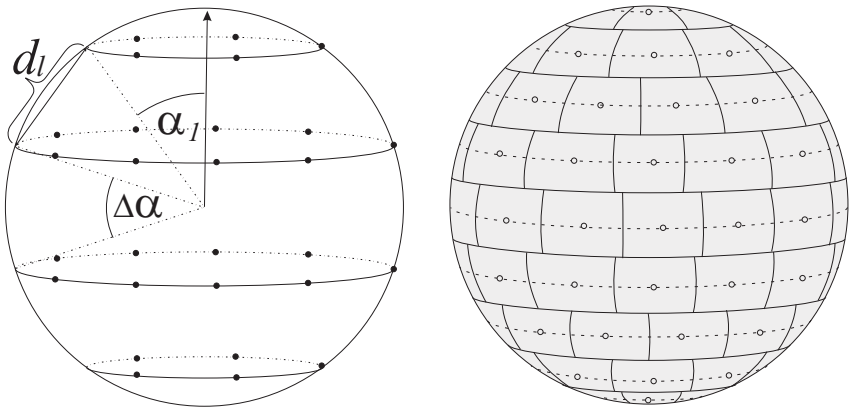

Figure 3.2: a) Construction of the spherical code. b) Decoding regions of the spherical code

based on wrapped spherical codes originaly introduced for vector quantization by Hamkins and Zeger [14]. Here we follow the same procedure described by Henkel, with the difference in the construction of the sperical code.

3.3.1 Construction of the Spherical Code

The underlying spherical code is a modified version of the spherical code introduced in [61], [60]. The codewords are points from the unit sphere $S^D \in \mathbb{R}^{D+1}$. As shown in Fig. 3.3.1, the points of the codebook are "uniformly" distributed on equidistant *layers* obtained as a result of the intersection of the sphere with parallel planes. The number of points on each layer is proportional to the layer's $D-1$ dimensional content (surface area). The choice of the layers is done in a way such that the obtained spherical code is antipodal. The procedure is recursive since the layers obtained in this way are also spheres, now of smaller dimension (one less) from the sphere they originate from. The same is done at every instance (dimension). With an appropriate choice of the distance d_L between the layers, a fairly uniform distribution of points on the surface of the sphere may be obtained. For a code with given dimension and rate, an appropriate value of the parameter d_L is found in an iterative procedure, starting with an initial value d_{L_o} which is obtained by approximating the surface area of the k-dimensional unit radius sphere with the area covered by N D-dimensional curved hypercubes with sidelength equal to d_{L_o}.

The set of layers at some instance k of the recursive procedure may equivalently be described by the set of angles $\alpha_i^{(k)} \in [0, \pi]$, $\quad i = 1, 2, \ldots, n$ and $\alpha_{i+1}^{(k)} - \alpha_i^k = \Delta\alpha^k$.

Another way to describe the process of point distribution is by a tree with nodes

denoting the layers of the code structure, where each codeword is represented as a path through the tree. According to this convention, the path (ordered sequence) $j_D, j_{D-1}, \cdots, j_1$ corresponds to a point from the codebook which belongs to the layer $L_j^{(D)}$ at stage D, $L_j^{(D-1)}$ at stage $D-1$, etc.

Decoding of the Spherical Code

This structure may be efficiently exploited by the receiver where the decision is based on successive "hard" decisions about affiliation to particular layers at each stage (dimension) of this procedure.

Let $\mathbf{y}_S$ be an input point at the spherical decoder and let $\Theta_D, \Theta_{D-1}, \ldots : \Theta_1$ be the equivalent representation with generalized spherical coordinates. Then, the angle representation of the layers can be used for efficient decoding. Namely, at each stage (dimension $D, D-1, \ldots$), we can decide about the affiliation of the point $\mathbf{y}_S$ to a particular layer by a simple scalar quantization (comparison with the angles of the layers) of the corresponding angle (in spherical coordinate representation). The decoding procedure can thus be represented by D successive scalar quantizations.

Obviously, the decoding procedure is suboptimal. However, it allows for efficient decoding of spherical codes of high dimension and various rate. Furthermore, the codewords do not need to be stored at the receiver, which is anyway not feasible for large codebooks. The decoder only has to perform the same indexing procedure as the encoder, following the same rules (using the same parameter of the sphere point distribution d_L). Obviously, very large constellations may be decoded.

3.3.2 Construction of the Space-Time Code

The $U(T)$-homogeneous structure of the Grassmann manifold provides relation to spherical codes. Any code on a half sphere $S^D \in \mathbb{R}^{D+1}$ can be transformed into a code on the Grassmann manifold [15] where

$$D \;=\; \dim_{\mathbb{R}} G_{M,T}^{\mathbb{C}} = 2M(T-M) \tag{3.7}$$

In this particular case, the spherical code is restricted to the northern hemisphere. The encoding procedure is the same one described in [15], shortly summarized in the following.

First, we fix the north pole $\mathbf{n}$ of the D dimensional half-sphere; then we project the spherical code $\mathcal{C}_S$ on the tangent space $T_N(S^D)$ at the north pole (orthogonal projection) and scale to length equal to the geodesics emanating from $\mathbf{n}$ to the considered point of the spherical code $\mathcal{C}_S$; in the next step, tangential code points from $T_{\mathbf{n}}S^D$ are mapped to the tangent space of the manifold $T_{I_{T,M}}(G_{T,M}^{\mathbb{C}})$ by choosing orthonormal

bases in both spaces (this is allowed, since both tangent spaces are of same dimension); in the last step, we apply exponential map from $T_{I_{T,M}}(G^{\mathbb{C}}_{T,M})$ to the Grassmann manifold to obtain the space time code (having on mind the quotient space representation). The exponential map is conducted in a computationally efficient way, as given by (2.25).

Grassmannian constellations with structure

The exponential map (and its inverse) preserve geodesics emanating from the point at which the tangent space is constructed (the north pole $\mathbf{n}$ in the case of the (half) sphere and the identity point $\mathbf{I}_{T,M}$ in the case of the Grassmann manifold.

With this, the circular (layer) structure of the spherical code will be transfered to the tangent space of the (half) sphere and thus to the tangent space of the Grassmann manifold. Finally, this structure will apply to the Grassmann manifold itself, meaning that points from the spherical code that belong to the same layer (sphere of dimension $D-1$) with particular latitude, will be equally distant (geodesic distance) from $\mathbf{I}_{T,M}$. The mapping, in general, distorts the chordal and geodesic distances between the points from the spherical code. However, since it preserves the geodesic distances to the fixed point where the tangent space is constructed, it also preserves the relative position of the points in the codewords, meaning that the structure is preserved. This imposes a certain "layer" structure to the Grassmannian constellation obtained with the mapping procedure, which is similar to the structure of the spherical code it originates from.

3.3.3 Decoding

The decoding procedure is the inverse encoding procedure. However, there are some specifics worth to be mentioned. Let $\mathbf{\Psi}$ be the received $T \times N$ matrix, as given by (2.1). The M-dimensional subspace detection can be done by performing QR decomposition of $\mathbf{\Psi}$. Let $\hat{\mathbf{\Phi}} \in \mathbb{C}^{T\times M}$ be the result of the decomposition (after deletion of the last $T-M$ columns). In the absence of noise, $\hat{\mathbf{\Phi}}$ spans the same subspace as the transmitted matrix $\mathbf{\Phi}$. The result of the channel (after the decomposition) can be seen as a right multiplication by a unitary matrix $\mathbf{U}_{ch} \in U(M)$. With this and (2.25), $\hat{\mathbf{\Phi}}$ becomes

$$\hat{\mathbf{\Phi}} = \begin{pmatrix} \mathbf{W}(\mathbf{C})\mathbf{W}^{\mathrm{H}}\mathbf{U}_{ch} \\ \mathbf{V}_1(\mathbf{S})\mathbf{W}^{\mathrm{H}}\mathbf{U}_{ch} \end{pmatrix} \tag{3.8}$$

This helps us to perform the inverse mapping to the tangent space. If we remember, the corresponding tangent $\mathbf{X}$ is determined by the matrix $\mathbf{B}$, where $\mathbf{B} = \mathbf{V}_1\mathbf{\Sigma}\mathbf{W}^{\mathrm{H}}$. Thus, $\mathbf{B}$ can be read off from $\hat{\mathbf{\Phi}}$. Additionally, $\mathbf{U}_{ch}$ can be found and $\hat{\mathbf{\Phi}}\mathbf{U}_{ch}^{-1}$ gives $\mathbf{\Phi}$. The same applies with the noise with the remark that the noise will change the

M-dimensional subspace. It should be noted that loss of performance is expected because of the properties of the inverse mapping, since only specific geodesic distances are preserved. However, the aim was to do the final decoding on the surface of the D-dimensional sphere, and not on the Grassmann manifold, in order to exploit the efficient decoding mechanism.

It is interesting to point out that, for smaller constellations that can be stored, it is possible to perform the decoding with the decoding algorithm for Grassmannian lattices, presented in [13], [12]. In this context, our method can be seen a practical way of obtaining Grassmannian codes with certain structure (codewords distributed on layers). Once again, this is feasible for smaller size constellations and consequently there is a limit in the spectral efficiency in terms of sent bits per channel use.

3.3.4 Examples and Simulation Results

It may be expected that ”good” spherical codes would lead to ”good” space time codes. The original spherical code has also been tested in another scenario (source coding of high-dimensional Gaussian sources) and the results show that it compares well to some of the best spherical codes designed for that purpose ([61], table 1) such as the wrapped spherical codes introduced by Hamkins and Zeger, [14]. With this

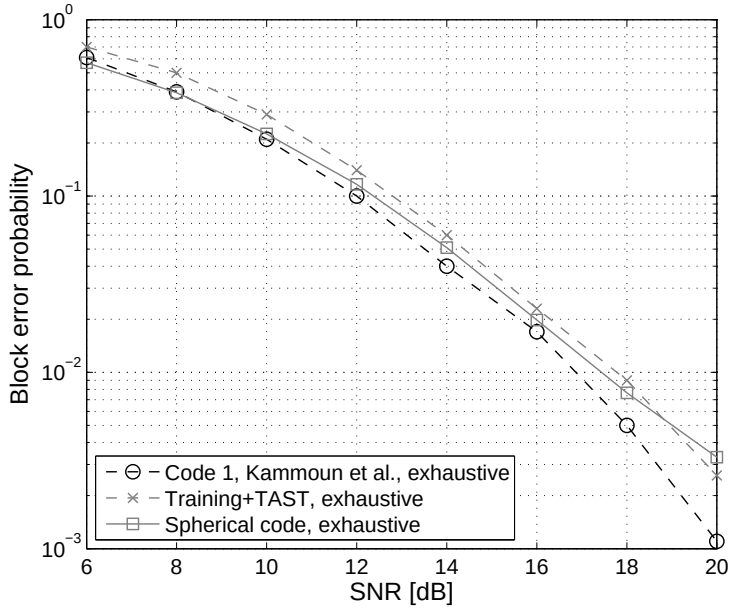

Figure 3.3: Performance comparison of non-coherent space time codes: $M = 2, N = 2, T = 4, \eta = 2$ bits/c.u.

on mind, the next step was to test the performance in the multiple transmit antenna channel. Because of the specific construction and no storage requirements, codes of larger blocks (coherence time T) can be designed. Additionally, higher code rates can be achieved, leading to spectral efficient space time codes. The simulation results are given in Fig. 3.3 and Fig. 3.4.

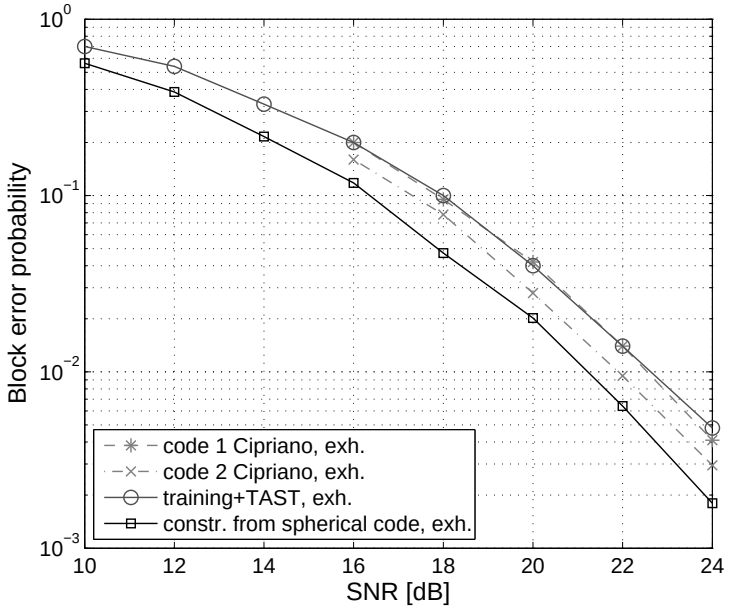

Figure 3.4: Performance comparison of non-coherent space time codes: $M = 2, N = 2, T = 6, \eta = 4$ bits/c.u.

3.4 Grassmann Codes from Stiefel Codes

The non-linearity of the exponential map is the main obstacle in the exact derivation of the diversity properties of the Grassmann codes obtained by the exponential map. In the following we analyze a construction from constellations in the tangent space which allows for closed form analysis of the performance of the Grassmann codes.

3.4.1 Code Construction

Let us observe the special class of Grassmann codes obtained by (2.23), where $\mathbf{B}^{\mathrm{H}}\mathbf{B} = \mathbf{I}_k$, i.e. $\mathcal{B}$ is a code in the Stiefel manifold, $\mathcal{B} \subset V_{k,n-k}^{\mathbb{C}}$, $(n \geq 2k)$. This corresponds to codes constructed from matrices with orthogonal columns. Then, $\mathbf{B} = \mathbf{V}_1\mathbf{I}_k\mathbf{V}_2^H$.

If we take $\mathbf{X}$ in the form (2.22) and a scaling factor α then, according to (2.25), the exponential map gives

$$\mathbf{\Phi} = \exp(\alpha\mathbf{X})\mathbf{I}_{n,k} = \begin{pmatrix} \mathbf{V}_2(a\mathbf{I}_k)\mathbf{V}_2^H \\ \mathbf{V}_1(b\mathbf{I}_k)\mathbf{V}_2^H \end{pmatrix} = \begin{pmatrix} a\mathbf{I}_k \\ b\mathbf{B} \end{pmatrix}, \tag{3.9}$$

where $a = \cos\alpha$ and $b = \sin\alpha$.

We note that similar constructions are presented in [19]. The difference here is the scaling factor α, which gives additional degree of freedom. In [34] a similar construction is presented, where orthogonal space-time codes are brought in relation to packings on Grassmann manifolds. It can be shown [34] that (3.9) corresponds to packings in the Grassmann manifold with pairwise equal principle angles between subspaces. It is interesting to see that the codes here are obtained based on the geometric approach and on first sight are independent from the constructions in [19] and [34].

3.4.2 Code Properties

Let $\mathbf{B}_1, \mathbf{B}_2$ be two codewords from the code $\mathcal{B}$. The corresponding Grassmann codewords obtained after the exponential map will be then

$$\mathbf{\Phi}_1 = \begin{pmatrix} a\mathbf{I}_k \\ b\mathbf{B}_1 \end{pmatrix}, \mathbf{\Phi}_2 = \begin{pmatrix} a\mathbf{I}_k \\ b\mathbf{B}_2 \end{pmatrix}, \tag{3.10}$$

where $a^2 + b^2 = 1$ and $\Phi_1, \Phi_2 \in G_{k,n}^{\mathbb{C}}$. In this case $\underline{\mathbf{\Delta}}$ becomes

$$\underline{\mathbf{\Delta}} = \mathbf{\Phi}_1^{\mathrm{H}}\mathbf{\Phi}_2 = a^2\mathbf{I}_k + b^2\mathbf{B}_1^{\mathrm{H}}\mathbf{B}_2 \tag{3.11}$$

We recall for the (pairwise) non-coherent diversity we have

$$\begin{aligned} \underline{Div}(\mathbf{\Phi}_1, \mathbf{\Phi}_2) &= \prod_{i=1}^{M} \left[1 + \underline{\varrho}\left(1 - \underline{\sigma}_i^2(\underline{\mathbf{\Delta}})\right)\right] \\ &= \prod_{i=1}^{M} \left[1 + \underline{\varrho}\left(1 - \lambda_i(\underline{\mathbf{\Delta}\mathbf{\Delta}}^{\mathrm{H}})\right)\right], \end{aligned}$$

where $\lambda_i(\underline{\mathbf{\Delta}\mathbf{\Delta}}^{\mathrm{H}})$ are the eigenvalues of $\underline{\mathbf{\Delta}\mathbf{\Delta}}^{\mathrm{H}}$. (3.10) yields

$$\begin{aligned} \underline{\mathbf{\Delta}\mathbf{\Delta}}^{\mathrm{H}} &= (a^4 + b^4)\mathbf{I}_k + a^2b^2\left(\mathbf{B}_1^{\mathrm{H}}\mathbf{B}_2 + \mathbf{B}_2^{\mathrm{H}}\mathbf{B}_1\right) \\ &= \mathbf{I}_k - 2a^2b^2\left(\mathbf{I}_k - \frac{\mathbf{B}_1^{\mathrm{H}}\mathbf{B}_2 + \mathbf{B}_2^{\mathrm{H}}\mathbf{B}_1}{2}\right). \end{aligned} \tag{3.12}$$

Since $\mathbf{B}_1^{\mathrm{H}}\mathbf{B}_2 + \mathbf{B}_2^{\mathrm{H}}\mathbf{B}_2$ is Hermitian, for $\lambda_i(\underline{\mathbf{\Delta}\mathbf{\Delta}}^{\mathrm{H}})$ it holds

$$\lambda_i(\underline{\mathbf{\Delta}\mathbf{\Delta}}^{\mathrm{H}}) = 1 - 2a^2b^2\left[1 - \lambda_i\left(\frac{\mathbf{B}_1^{\mathrm{H}}\mathbf{B}_2 + \mathbf{B}_2^{\mathrm{H}}\mathbf{B}_1}{2}\right)\right]. \tag{3.13}$$

The diversity then becomes

$$\underline{Div}(\mathbf{\Phi}_1, \mathbf{\Phi}_2) = \prod_{i=1}^{M}\left\{1 + \underline{\varrho}2a^2b^2\left[1 - \lambda_i\left(\frac{\mathbf{B}_1^{H}\mathbf{B}_2 + \mathbf{B}_2^{H}\mathbf{B}_1}{2}\right)\right]\right\}. \tag{3.14}$$

Since

$$1 - \lambda_i\left(\frac{\mathbf{B}_1^{H}\mathbf{B}_2 + \mathbf{B}_2^{H}\mathbf{B}_1}{2}\right) = \frac{1}{2}\sigma_i^2(\overline{\Delta}) \doteq \frac{1}{2}\bar{\sigma}_i^2 \tag{3.15}$$

we have

$$\underline{Div}(\mathbf{\Phi}_1, \mathbf{\Phi}_2) = \prod_{i=1}^{k}\left(1 + a^2b^2\ \underline{\varrho}\bar{\sigma}_i^2\right). \tag{3.16}$$

This provides a relation between $\underline{Div}(\mathbf{\Phi}_1, \mathbf{\Phi}_2)$ and $\overline{Div}(\mathbf{B}_1, \mathbf{B}_2)$, having on mind that

$$\overline{Div}(\mathbf{B}_1, \mathbf{B}_2) := \prod_{i=1}^{k}\left(1 + \bar{\varrho}\bar{\sigma}_i^2\right), \tag{3.17}$$

where

$$\underline{\varrho} \doteq \frac{(\rho\frac{n}{k})^2}{4(1+\rho\frac{n}{k})}$$

and

$$\bar{\varrho} \doteq \frac{\rho(n-k)}{4k}$$

If we concentrate only the last term of the expression in the non-coherent case, i.e. on the diversity product, we see that it differs from the "determinant" criterion [46], [30] for construction of coherent codes only in the scaling factor. The "determinant" criterion is related to the coding gain of a coherent space-time code. This means that the properties of the non-coherent code $\mathcal{C}$ in terms of the diversity product depend on the properties of the coherent code $\mathcal{B}$ in terms of the determinant criterion. In other words, Grassmann codes constructed from coherent Stiefel codes with maximized coding gain, will have the diversity product maximized.

These codes will be used in the further text as basis for the construction of other Grassmann codes, for both the point-to-point as well as for the wireless networks with relays. Therefore, we will addres the performance and decoding of these codes in the context of the codes where they appear as building blocks.

3.5 Grassmann Codes based on Recursive Construction

The capacity analysis of non-coherent space-time codes shows that block lengths $T \geq 2M$ allow for maximum use of the available degrees of freedom [65]. Additionally, the coherence length of the channel, although considered small for channels where non-coherent codes are of interest, is still relatively large with respect to the number of transmit antennas, $T_c >> M$. Although this is a motivation for employment of codes with large block lengths, the constructions in the literature are usually limited to $T = 2M$.

A strong argument for increasing the block length is given by Henkel in [16], where it is proven that a non-coherent space time code $\mathcal{C} \in G_{M,T}^{\mathbb{C}}$ with rate $R = (\log_2 |\mathcal{C}|)/T$ exists, such that for the minimum distance it holds

$$d_{min} \sim \sqrt{\frac{T}{M}} \cdot \left(\frac{1}{2}\right)^{\frac{TR+1}{2M(T-M)}}. \tag{3.18}$$

Thus, for a fixed rate R (spectral efficiency), $SNR\ \rho \geq 1$ and $T \geq 2M$ [16], the performance is expected to grow at least proportionally to $\sqrt{\frac{T}{M}}$. This motivates the construction of higher-dimensional non-coherent codes with $T >> M$, unlike the usual design with $T = 2M$.

In [26], [1], non-coherent codes are constructed from coherent codes residing in the tangent space of the identity point. There, under some assumptions, the problem of non-coherent code construction is identified with the problem of construction of a fully-diverse coherent code [26], [30].

On the other hand, in [17], a construction of higher-dimensional coherent codes from non-coherent codes is proposed. The coherent code is obtained as a product of a non-coherent code and a low-dimensional unitary code. This leads to a decrease of the dimensionality of the code construction problem, since the coherent code resides in the Stiefel manifold $V_{M,T}^{\mathbb{C}}$ which has a higher dimension compared to the Grassmann manifold $G_{M,T}^{\mathbb{C}}$.

A natural remark arising from these two works is that, under some conditions, high-dimensional non-coherent codes can be constructed from coherent or non-coherent codes of lower dimension, in a recursive procedure. This is also noted by Henkel in [16], as a remark following the analysis of the coherent codes constructed from non-coherent ones.

In this spirit we propose a recursive code construction of high-dimensional Grassmann codes. A low-dimensional coherent or non-coherent code is chosen as a starting point.

3.5.1 Code Construction

Let us construct a non-coherent code for M transmit antennas and coherent time $T > 2M$. We start with a code in the tangent space (code for the coherent channel) $\mathcal{B}_0 \subset \mathbb{C}^{M\times M}$. According to (2.22), a code $\mathcal{X}_0$ can be constructed in the tangent space of $G^{\mathbb{C}}_{M,T_0}$, $(T_0 = 2M)$ as

$$\mathcal{X}_0 = \begin{pmatrix} \mathbf{0} & -\mathcal{B}_0^H \\ \mathcal{B}_0 & \mathbf{0} \end{pmatrix}. \tag{3.19}$$

Then, a non-coherent code $\mathcal{C}_0 \subset G^{\mathbb{C}}_{M,T_0}$ is constructed as

$$\mathcal{C}_0 = \exp(\alpha_0 \mathcal{X}_0)\mathbf{I}_{T_0,M}, \tag{3.20}$$

where α_0 is a scaling factor which is parameter of the construction. The non-coherent code $\mathcal{C}_0$ yields a coherent code $\mathcal{B}_1$, by multiplication (set product) with an $M \times M$ unitary code $\mathcal{U}_0$. As examples we can take the Alamouti [2] code or an $Sp(M/2)$ code [23]. The $Sp(2)$ code was proposed by Hassibi in [23]. The code $\mathcal{B}_1$ is given as

$$\mathcal{B}_1 = \mathcal{C}_0 \times \mathcal{U}_0. \tag{3.21}$$

With this, $\mathcal{B}_1$ is a subset of the Stiefel manifold, $\mathcal{B}_1 \subset V^{\mathbb{C}}_{M,T_0}$. Further, $\mathcal{B}_1$ yields a non-coherent code $\mathcal{C}_1 \subset G^{\mathbb{C}}_{M,T_1}$

$$\mathcal{C}_1 = \exp(\alpha_1 \mathcal{X}_1)\mathbf{I}_{T_1,M},\ T_1 = 3M. \tag{3.22}$$

The procedure can continue in the same way, increasing the dimensionality of the non-coherent code with each cycle. Given $\mathcal{B}_0$ and unitary codes $\mathcal{U}_0, \mathcal{U}_1, \ldots, \mathcal{U}_{k-1}$, we have

$$\begin{aligned} \mathcal{C}_i &= \exp(\alpha_i \mathcal{X}_i)\mathbf{I}_{T_i,M}\,, \\ \mathcal{X}_i &= \begin{pmatrix} \mathbf{0} & -\mathcal{B}_i^H \\ \mathcal{B}_i & \mathbf{0} \end{pmatrix}, \\ \text{and} \quad \mathcal{B}_i &= \mathcal{C}_{i-1} \times \mathcal{U}_{i-1}\ ,\ i = 1, 2, \ldots, k. \end{aligned} \tag{3.23}$$

The recursive code construction is summarized as follows

$$\mathcal{B}_0 \overset{\exp(\alpha_0)}{\longrightarrow} \mathcal{C}_0 \overset{\times\mathcal{U}_0}{\longrightarrow} \mathcal{B}_1 \overset{\exp(\alpha_1)}{\longrightarrow} \cdots \overset{\exp(\alpha_k)}{\longrightarrow} \mathcal{C}_k, \tag{3.24}$$

where $\mathcal{B}_i \subset \mathbb{C}^{(i+1)M\times M}$ and $\mathcal{C}_i \subset \mathbb{C}^{(i+2)M\times M}$. The block length increases for M in each cycle, $T_i = (i+2)M$, $i \geq 0$. We observe that for the special case when $\mathcal{B}_0^{\mathrm{H}}$ is an unitary code, i. e. $\mathbf{B}_0^{\mathrm{H}}\mathbf{B}_0 = \mathbf{I}_M$, a code for M transmit antennas and coherence time

T_k, $\mathcal{C}_k \subset G^{\mathbb{C}}_{M,T_k}$, has the form

$$\mathcal{C}_k = \begin{pmatrix} a_k \cdot \mathbf{I}_M \\ b_k \cdot a_{k-1} \cdot \mathcal{U}_{k-1} \\ \vdots \\ b_k \cdots b_2 \cdot a_1 \cdot \mathcal{U}_1 \times \cdots \times \mathcal{U}_{k-1} \\ b_k \cdots b_1 \cdot a_0 \cdot \mathcal{U}_0 \times \mathcal{U}_1 \times \cdots \times \mathcal{U}_{k-1} \\ b_k \cdots b_1 \cdot b_0 \cdot \mathcal{B}_0 \times \mathcal{U}_0 \times \mathcal{U}_1 \times \cdots \times \mathcal{U}_{k-1} \end{pmatrix}. \tag{3.25}$$

This can be easily shown by using (3.9). The scalars $a_0, a_1, \ldots, a_k$ and $b_0, b_1, \ldots, b_k$ are parameters of the exponential map, satisfying $a_i = \cos\alpha_i$ and $b_i = \sin\alpha_i$.

3.5.2 Properties of the Recursive Codes

We have seen that the recursive construction involves two basic steps. The first one is the construction of a coherent code from a non-coherent one, performed by right multiplication with unitary matrices. The code obtained in this way is a subset from the Stiefel manifold, as described before. The second step is the construction of a non-coherent code from a coherent code by performing the exponential map. Therefore, in order to evaluate the properties of the recursive code, we will separate the analysis into two steps. First, we will analyze the properties of coherent codes obtained from non-coherent codes. Then, we will analyze the properties of non-coherent codes obtained from coherent codes.

Properties of the coherent codes obtained from non-coherent codes

Let $\mathbf{B}_1 = \mathbf{\Phi}_1\mathbf{U}_1$ and $\mathbf{B}_2 = \mathbf{\Phi}_2\mathbf{U}_2$ be two codewords of the coherent code $\mathcal{B}$ obtained from a non-coherent code $\mathcal{C}$, by multiplication with unitary matrices, i.e. codewords from a coherent unitary code. Since $\mathbf{B}_1^{\mathrm{H}}\mathbf{B}_1 = \mathbf{I}$ and $\mathbf{B}_2^{\mathrm{H}}\mathbf{B}_2 = \mathbf{I}$, $\mathbf{B}_1$ and $\mathbf{B}_2$ are elements from the Stiefel manifold, $\mathbf{B}_1, \mathbf{B}_2 \in V^{\mathbb{C}}_{M,T}$. According to (2.33), the (pairwise) diversity is given by

$$\begin{aligned} \overline{Div}(\mathbf{B}_1, \mathbf{B}_2) := \prod_{i=1}^{M} \left(1 + \bar{\varrho}\bar{\sigma}_i^2\right) &= \prod_{i=1}^{M} \left[1 + \bar{\varrho}\lambda_i\left(\bar{\Delta}^{\mathrm{H}}\bar{\Delta}\right)\right] \\ &= \prod_{i=1}^{M}\left[1 + 2\bar{\varrho}\left(1 - \lambda_i\left(\frac{\underline{\Delta} + \underline{\Delta}^H}{2}\right)\right)\right] \end{aligned} \tag{3.26}$$

where $\underline{\Delta} = \mathbf{U}_2^{\mathrm{H}}\mathbf{\Phi}_2^{\mathrm{H}}\mathbf{\Phi}_1\mathbf{U}_1$. For the eigenvalues of $\frac{\underline{\Delta}+\underline{\Delta}^{\mathrm{H}}}{2}$ it holds [36]

$$\lambda_i\left(\frac{\underline{\Delta} + \underline{\Delta}^{\mathrm{H}}}{2}\right) \leq \sigma_i(\underline{\Delta}). \tag{3.27}$$

So, the following inequality holds

$$\overline{Div}(\mathbf{B}_1, \mathbf{B}_2) := \prod_{i=1}^{M} \left(1 + \bar{\varrho}\bar{\sigma}_i^2\right) \geq \prod_{i=1}^{M} \{1 + 2\bar{\varrho}\,[1 - \sigma_i(\underline{\mathbf{\Delta}})]\} \tag{3.28}$$

$$\geq \prod_{i=1}^{M} \left\{1 + \bar{\varrho}\left[1 - \sigma_i^2(\underline{\mathbf{\Delta}})\right]\right\} \tag{3.29}$$

$$= \prod_{i=1}^{M} \left(1 + \bar{\varrho}\underline{\sigma}_i^2\right) \tag{3.30}$$

$$\geq \prod_{i=1}^{M} \left(1 + \underline{\varrho}\underline{\sigma}_i^2\right) = \underline{Div}(\mathbf{\Phi}_1, \mathbf{\Phi}_2) \tag{3.31}$$

In other words, when non-coherent codes are used as coherent, the diversity does not decrease.

Properties of the non-coherent codes obtained from coherent codes

Let $\mathbf{B}_1, \mathbf{B}_2 \in V_{T,T-M}^{\mathbb{C}}$ be two codewords of a Stiefel code constructed for the known channel. According to (3.4), for the non-coherent diversity we have

$$\underline{Div}(\mathbf{\Phi}_1, \mathbf{\Phi}_2) = \prod_{i=1}^{M} \left(1 + a^2 b^2\, \underline{\varrho}\bar{\sigma}_i^2\right).$$

This provides a relation between $\underline{Div}(\mathbf{\Phi}_1, \mathbf{\Phi}_2)$ and $\overline{Div}(\mathbf{B}_1, \mathbf{B}_2)$, having on mind that

$$\overline{Div}(\mathbf{B_1}, \mathbf{B_2}) := \prod_{i=1}^{M} \left(1 + \bar{\varrho}\bar{\sigma}_i^2\right)$$

and $\underline{\varrho}$ and $\bar{\varrho}$ are given by (2.32) and (2.29) respectively. Finally, at high SNR ρ we have

$$\underline{Div}(\mathbf{\Phi}_1, \mathbf{\Phi}_2) \approx \prod_{i=1}^{M} \left(1 + a^2 b^2 \frac{T}{T-M}\, \bar{\varrho}\bar{\sigma}_i^2\right). \tag{3.32}$$

We can see that the terms for the non-coherent and the coherent channel differ in the scaling factor $a^2 b^2 \frac{T}{T-M}$. If we ignore the scaling factor for a moment, we observe that the properties of the non-coherent code depend on the eigenvalues of $\overline{\Delta}$, i.e. on the properties of the coherent code $\mathcal{B}$.

3.5.3 Decoding

We perform ML-decoding by using exhaustive search over the codebook in order to demonstarte the potential of the recursive construction. The simplified decoding of

the codes of the form (3.25) has to be further investigated. It seems that the block structure offers potential for simplified decoding. The general algorithms suggested in [13], [12] can also be used. Additionally, decoding can be performed in the tangent space, as described in [1].

3.5.4 Examples and Simulation Results

Here we present simulation results for several worked-out examples. The simulations show that the code performance fits the theoretical prediction given by (3.18), which is result of the good distance properties of the codes obtained in this way. As expected, the increase of the block length could lead to a significant improvement of the performance of Grassmann codes for the non-coherent MIMO system.

Recursive construction in $G_{2,6}^{\mathbb{C}}$, $\eta = 2$ bits/c.u.

Here, two different codes are taken as starting point for the recursive construction. The first one is the 2×2 coherent code $\mathcal{C}_c^{(0)}$ of rate 4, constructed in [30], also used in [26], [1], with codewords

$$\mathbf{B} = \frac{1}{\sqrt{2}} \begin{pmatrix} s_1 + \theta s_2 & \phi(s_3 + \theta s_4) \\ \phi(s_3 - \theta s_4) & s_1 - \theta s_2 \end{pmatrix}, \tag{3.33}$$

where $\phi^2 = \theta = e^{i\pi/4}$ and s_i, $i = 1, 2, 3, 4$ are the 4 information QPSK symbols. This is a version of the Golden code introduced in [5]. The code used for multiplication is the rate 2 unitary (Alamouti) code $\mathcal{U}$ [2], with codewords

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{pmatrix}, \tag{3.34}$$

where s_1, s_2 are taken from the 4-QPSK symbols. The scaling factors used in the exponential map are $\alpha_0 = 0.566$ and $\alpha_1 = 1$.

The final code of block length $T = 6$ is constructed as follows

$$\mathcal{C}_c^{(0)} \xrightarrow{\exp} \mathcal{C}_n^{(0)} \xrightarrow{\times \mathcal{U}} \mathcal{C}_c^{(1)}, \xrightarrow{\exp} \mathcal{C}_n^{(1)}, \ \mathcal{C}_n^{(1)} \subset G_{2,6}^{\mathbb{C}}. \tag{3.35}$$

The rate (spectral efficiency) of the obtained code $\mathcal{C}_n^{(1)}$ is

$$\eta = \log_2 \frac{(|\mathcal{C}_c^{(0)}| \cdot |\mathcal{U}|)}{T} = 2 \,\text{bits/channel use}.$$

The simulation results are shown in Fig. 3.5. As expected, the performance increases with increase of the block length. For $T = 6$, compared to $T = 4$, the gain is approximately 2 dB. The results fit well the theoretical prediction given by (3.18).

The second code is a rate 2 non-coherent code of block length $T = 4$, constructed from the spherical code described in 3.3. The construction is flexible and allows for design of non-coherent codes of different rates. The scaling factor of the exponential map is here also, $\alpha = 1$. The same effect of performance increase can be observed with the increase of the block length T.

Recursive construction in $G_{2,8}^{\mathbb{C}}$, $\eta = 2$ bits/channel use

Here the construction procedure is the same as in the case of $G_{2,6}^{\mathbb{C}}$, only adding one additional cycle to obtain codes of block length $T = 8$. The scaling factor for the additional exponential map is chosen $\alpha_2 = 1$. The starting codes are the same as in the previous examples.

$$\mathcal{C}_n^{(1)} \xrightarrow{\times \mathcal{U}} \mathcal{C}_c^{(2)} \xrightarrow{\exp} \mathcal{C}_n^{(2)},\ \mathcal{C}_n^{(2)} \subset G_{2,8}^{\mathbb{C}}. \tag{3.36}$$

As shown in Fig. 3.5, for $T = 8$, compared to $T = 4$, the gain is approximately 3 dB, which fits the prediction given by (3.18).

Recursive construction in $G_{2,10}^{\mathbb{C}}$, $\eta = 1.6$ bits/channel use

Here the starting code is the 2×2 Alamouti code of rate 2. The same code is used for multiplication in each step. With the recursive procedure, a code with block length $T = 10$ and rate (spectral efficiency) 1.6 is constructed. The scaling factor α is the same, $\alpha = 1$, every time when exponential map takes place. This construction is particularly interesting, since it depends on only one parameter, i.e. the same code is used in all steps of the procedure. The performance is shown in Fig. 3.5.

3.6 Grassmann Space-Time Codes from Lattices

The parametrization of the Grassmann manifold $G_{T,M}^{\mathbb{C}}$ with the set of $T \times T$ skew-hermitian matrices of the form (2.22), gives rise to non-coherent codes constructed from a set of elements (points) of the tangent space. Since the tangent space can be identified with the set of matrices $\mathcal{B} = \left\{ \mathbf{B},\ \mathbf{B} \in \mathbb{C}^{M \times (T-M)} \right\}$, canonical embedding of the tangent space in the real space $\mathbb{R}^{2M(T-M)}$ is justified. Hence, non-coherent space-time codes can be constructed from sets in $\mathbb{R}^{2M(T-M)}$. One way to do this is to choose a set which arises from a (sphere) packing in $\mathbb{R}^{2M(T-M)}$. In this sense, a lattice packing is favored, due to the simple construction and the potential for decoding.

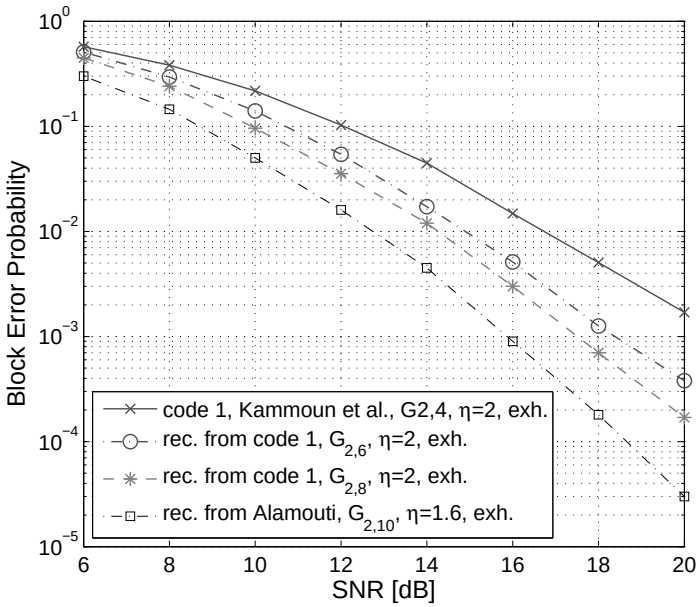

Figure 3.5: Performance comparison of non-coherent space time codes obtained from a recursive construction, with block lengths $T = 4, T = 6$ and $T = 8$, $M = 2, N = 2, \eta = 2$ bits per channel use.

3.6.1 Code Construction

As lattice Λ in $\mathbb{R}^n$ we will understand a discrete subgroup of $\mathbb{R}^n$ which spans the real vector space $\mathbb{R}^n$. Some of the densest sphere packings in certain dimensions (e.g. 2, 8 and 24) are shown to be lattice packings.

Let Λ be a lattice in $\mathbb{R}^D$, where $D = 2M(T - M)$. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_D$ be a basis for the D-dimensional tangent space with elements matrices of the form

$$\mathbf{X} = \begin{pmatrix} \mathbf{0} & -\mathbf{B}^{\mathrm{H}} \\ \mathbf{B} & \mathbf{0} \end{pmatrix}, \ \mathbf{B} \in \mathbb{C}^{(T-M)\times M}. \tag{3.37}$$

A lattice point $\mathbf{x} = (x_1, x_2, \ldots, x_D) \in \Lambda$ becomes a point from the tangent space under the transformation

$$\mathbf{X} = \sum_{i=1}^{D} x_i \mathbf{X}_i. \tag{3.38}$$

Let us denote by $\mathcal{C}_\Lambda$ a subset of the lattice $\mathcal{C}_\Lambda \subset \Lambda$. Let $\mathcal{X}$ be the image of $\mathcal{C}_\Lambda$ after the basis transformation. $\mathcal{X}$ is then of the form

$$\mathcal{X} = \begin{pmatrix} \mathbf{0} & -\mathcal{B}^{\mathrm{H}} \\ \mathcal{B} & \mathbf{0} \end{pmatrix}, \ \mathcal{B} \subset \mathbb{C}^{(T-M)\times M}, \tag{3.39}$$

and is uniquely identified by the set $\mathcal{B}$. The Grassmann code $\mathcal{C}$ is obtained after mapping $\mathcal{X}$ (or equivalently $\mathcal{B}$), on the Grassmann manifold via the exponential map. For $\mathbf{X} \in \mathcal{X}$, a codeword $\mathbf{\Phi} \in \mathcal{C}$ is obtained as

$$\mathbf{\Phi} = \exp(\alpha \mathbf{X})\, \mathbf{I}_{T,M}, \tag{3.40}$$

where α is a scaling factor which ensures that the Grassmannian constellation is not folded up with the exponential mapping [26]. This implies a constraint on the singular values of the elements of $\mathcal{B}$, i.e.

$$\max_k \sigma_k(\alpha \mathbf{B}_i) \leq \frac{\pi}{2}, \; \forall \mathbf{B}_i \in \mathcal{B}. \tag{3.41}$$

The important question is what happens with the lattice structure after the mapping on the manifold. More precisely, the performance of the resulting code has to be evaluated in terms of the already established criteria for construction of non-coherent codes for multiple antenna channels.

The representation of the geodesic distance in the space of tangents at the identity gives us a hint about the structure of the code obtained by the mapping. For $\mathbf{\Phi} = \exp(\mathbf{X})\mathbf{I}_{T,M}$, for the geodesic distance between $\mathbf{\Phi}$ and $\mathbf{I}_{T,M}$ we have

$$\mathrm{d_g}(\mathbf{\Phi}, \mathbf{I}_{T,M}) = \|\mathbf{B}\|_{\mathrm{F}} = \frac{1}{\sqrt{2}} \|\mathbf{X}\|_{\mathrm{F}}. \tag{3.42}$$

This is easily shown by using the definition of principle angles and the representation of $\mathbf{\Phi}$ in the form (2.25). According to this, for the principle angles between $\mathbf{\Phi}$ and $\mathbf{I}_{T,M}$ we have

$$\cos\theta_i = \sigma_i(\mathbf{I}_{T,M}^{\mathrm{H}} \mathbf{\Phi}) = \cos\sigma_i(\mathbf{B}), \tag{3.43}$$

which for the geodesic distance gives

$$\mathrm{d_g}(\mathbf{\Phi}, \mathbf{I}_{T,M}) = \sqrt{\textstyle\sum_{i=1}^{M} \sigma_i^2(\mathbf{B})} = \sqrt{\mathrm{tr}(\mathbf{B}^{\mathrm{H}}\mathbf{B})} = \|\mathbf{B}\|_{\mathrm{F}}. \tag{3.44}$$

Similarly, we can calculate the geodesic distance between arbitrary subspaces $\mathbf{\Phi}, \mathbf{\Psi} \in G_{T,M}^{\mathbb{C}}$. Let $\tilde{\mathbf{\Psi}} = \left[\mathbf{\Psi}\mathbf{\Psi}^{\perp}\right]$ be the unitary extension of $\mathbf{\Psi}$, where $\mathbf{\Psi}^{\perp}$ is the orthogonal complement of $\mathbf{\Psi}$. Then the geodesic distance is given by $\mathrm{d_g}(\mathbf{\Phi}, \mathbf{\Psi}) = \mathrm{d_g}(\tilde{\mathbf{\Psi}}^{-1}\mathbf{\Phi}, \mathbf{I}_{T,M})$, because a left multiplication by an unitary matrix is an isometric transformation.

Now, let us define a sphere of radius r in $\mathbb{R}^n$,

$$\mathcal{S}(r) := \{\mathbf{b} \in \mathbb{R}^n \mid \|\mathbf{b}\| = r\}. \tag{3.45}$$

Due to the embedding of the tangent space in $\mathbb{R}^D$ where $D = 2M(T-M)$, the sphere can equivalently be represented with the set of matrices

$$\mathcal{S}(r) := \left\{\mathbf{B} \in \mathbb{C}^{M\times(T-M)} \mid \|\mathbf{B}\|_{\mathrm{F}} = r\right\}. \tag{3.46}$$

According to (2.22), after the exponential map, the sphere $\mathcal{S}(r)$ will be mapped to a subset of the Grassmann Manifold

$$\mathcal{S}_G(r) := \left\{ \mathbf{\Phi} \mid \mathrm{d}_g(\mathbf{\Phi}, \mathbf{I}_{T,M}) = r \right\}. \tag{3.47}$$

We will call this set a *geodesic sphere*, using the analogy with spheres in Euclidean space. We will have on mind that the dimension of the set is $D-1$, since it is obtained by mapping a $D-1$-dimensional sphere from the tangent space. The former observation leads to the conclusion that it is possible to impose a certain structure on the Grassmann code.

In an equivalent representation, a lattice can be partitioned into spherical codes, by grouping all vectors with the same norm. Let us by $\Lambda(m)$ denote the subset of the lattice corresponding to the m-th sphere. Then the spherical codes $\Lambda(m)$ form an exhaustive partitioning of Λ when m ranges over the positive integers

$$\Lambda = \bigcup_{m=1}^{\infty} \Lambda(m). \tag{3.48}$$

Let us by Λ_{G} denote the image of the lattice Λ on the Grassmann manifold under the exponential map. According to (2.22), the exponential map maps $\Lambda(m)$ to a subset $\Lambda_{\mathrm{G}}(m)$ of the Grassmann manifold

$$\begin{aligned} \Lambda(m) &\rightarrow \Lambda_{\mathrm{G}}(m) \subset G_{M,T}^{\mathbb{C}}, \\ \mathbf{x} &\rightarrow \mathbf{\Phi}\,,\ \mathrm{d}_g(\mathbf{\Phi}, \mathbf{I}_{M,T}) = r(m), \end{aligned} \tag{3.49}$$

where $\Lambda_{\mathrm{G}}(m)$, $m = 1, 2, \ldots$ partition Λ_{G}

$$\Lambda_{\mathrm{G}} = \bigcup_{m=1}^{\infty} \Lambda_{\mathrm{G}}(m). \tag{3.50}$$

Obviously, the structure of the lattice Λ is transferred to the Grassmann manifold in the sense that Λ_{G} can be partitioned into so-called geodesic spheres. A code on the lattice $\mathcal{C}_\Lambda$ can be constructed by choosing subsets of the lattice spherical codes.

Let $\mathcal{C}_\Lambda(m) \subset \Lambda(m)$ be a code on the m-th sphere (spherical code) with radius $r(m)$. Then $\mathcal{C}_\Lambda$ is given as

$$\mathcal{C}_\Lambda = \bigcup_{m \in \mathcal{M}} \mathcal{C}_\Lambda(m), \tag{3.51}$$

where $\mathcal{M} \subset \mathbb{N} \cup \{0\}$. According to (3.49), a code on the Grassmann manifold (Grassmann code) $\mathcal{C}_{\mathrm{G}}$ is obtained after the exponential map

$$\mathcal{C}_{\mathrm{G}} = \bigcup_{m \in \mathcal{M}} \mathcal{C}_{\mathrm{G}}(m), \tag{3.52}$$

where $\mathcal{C}_{\mathrm{G}}(m)$ is the image of $\mathcal{C}_\Lambda(m)$. According to the previous discussion, the structure of the lattice code is transferred to the code on the Grassmann manifold.

Lattice rotation

While describing the construction of Grassmann codes from lattices, we have mainly commented on the geodesic distance between the subspaces (codewords). Nevertheless, the chordal and the geodesic distance are locally equivalent [16]. This is intuitively supported by the fact that the Grassmann manifold (as any manifold) locally looks like Euclidean space. The former discussion about the geodesic distance does not hold for the chordal distance in general, but gives us an insight about the chordal distance distribution among the subspaces.

The diversity product, on the other hand, can be controlled by choosing an appropriate rotation of the lattice, which will affect the distribution of the principle angles after the mapping on the manifold. One way to perform such rotation is described in [18]. As the axis of rotation the diagonal $\mathbf{e} = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^D$ is chosen. Rotation around $\mathbf{e}$ is performed by the matrix $\mathbf{R}$ given by

$$\mathbf{R} = \mathbf{W}_{\mathbf{e}}^{\mathrm{T}} \mathbf{R}_1 \mathbf{W}_{\mathbf{e}} \tag{3.53}$$

where $\mathbf{W}_e$ is a unitary matrix with its first row $\mathbf{e}/\sqrt{D}$ and its j-th row

$$\begin{pmatrix} 1^{(j-1)} & -(j-1) & 0^{(D-j)} \end{pmatrix} / \sqrt{j(j-1)}, \tag{3.54}$$

where $0^{(D-j)}$ is a vector of all zeros of length $D-j$. The matrix $\mathbf{R}_1$ performs a rotation around the axis $\mathbf{e}_1 = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^D$ of angle ϕ. The matrix $\mathbf{R}_1$ is

$$\mathbf{R}_1 = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & \exp(\phi \mathbf{U}) \end{pmatrix}, \tag{3.55}$$

where $\mathbf{U}$ being the $(D-1) \times (D-1)$ antisymmetric matrix with ones in the upper triangular part

$$\mathbf{U} = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ -1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ -1 & \cdots & -1 & 0 \end{pmatrix}. \tag{3.56}$$

3.6.2 Some Special Lattices

The E_8 (Gosset) lattice

For the system with $M = 2$, $N = 2$ and $T = 4$, the E_8 lattice can serve for the construction since the dimension of the Grassmann manifold $G_{T,M}^{\mathbb{C}}$ is $2T(T-M) = 8$. The E_8 lattice is the densest lattice packing in $\mathbb{R}^8$. It can be given explicitly by the set

of points $\mathrm{E}_8 \in \mathbb{R}^8$ such that all the coordinates are integers or all the coordinates are half-integers, and the sum of the eight coordinates is an even integer. In symbols,

$$\begin{aligned} \mathrm{E}_8 &= \left\{ (x_i) \in \mathbb{Z}^8 : \sum_i x_i \equiv 0 \pmod 2 \right\} \\ &\cup \left\{ (x_i) \in (\mathbb{Z} + \tfrac{1}{2})^8 : \sum_i x_i \equiv 0 \pmod 2 \right\}. \end{aligned} \tag{3.57}$$

The E_8 lattice is closely related to the Hamming $(8,4)$ code and can actually be constructed from it. The E_8 lattice can be partitioned into spheres, $\mathrm{E}_8 = \bigcup_{m=1}^{\infty} \mathrm{E}_8(m)$, where $\mathrm{E}_8(m)$ is a spherical code with vectors having norm $\sqrt{2m}$. If we denote by $\mathcal{C}_{\mathrm{E}_8}(m)$ a code on the m-th sphere, then a code on the lattice can be defined as $\mathcal{C}_{\mathrm{E}_8} = \bigcup_{m\in\mathcal{M}} \mathcal{C}_{\mathrm{E}_8}(m)$, where $\mathcal{M} \subset \mathbb{N}$. After the mapping (3.49), the Grassmann code will be $\mathcal{C}_{\mathrm{G}} = \bigcup_{m\in\mathcal{M}} \mathcal{C}_{\mathrm{G}}(m)$.

The Barnes-Wall lattice, Λ_{16}

The Barnes-Wall lattice Λ_{16} can serve for construction of codes for the non-coherent channel with $M = 2$ and coherence time $T = 6$. It can be constructed from the Reed-Mueller code $\mathcal{C}_{\mathrm{RM}}(16,5,8)$. Every vector of the lattice is congruent (modulo 2) to a codeword of the $\mathcal{C}_{\mathrm{RM}}(16,5,8)$ code and the sum of all the coordinates of is a multiple of four

$$\Lambda_{16} = \left\{ \mathbf{x} \in \mathbb{Z}^{16} \mid \exists \mathbf{c} \in \mathcal{C}_{RM},\ \mathbf{x} \equiv \mathbf{c} \pmod 2 \right\} \tag{3.58}$$

$$\cap \left\{ \mathbf{x} \mid \sum_i x_i \equiv 0 \pmod 4 \right\} \tag{3.59}$$

The Λ_{16} lattice can be partitioned into spheres, $\Lambda_{16} = \bigcup_{m=1}^{\infty} \Lambda_{16}(m)$, where $\Lambda_{16}(m)$ is a spherical code with vectors having norm $2\sqrt{2m}$. If we denote by $\mathcal{C}_{\Lambda_{16}}(m)$ a code on the m-th sphere, then a code on the lattice is defined as $\mathcal{C}_{\Lambda_{16}} = \bigcup_{m\in\mathcal{M}} \mathcal{C}_{\Lambda_{16}}(m)$, where $\mathcal{M} \subset \mathbb{N}$. After the mapping (3.49), the Grassmann code will be $\mathcal{C}_{\mathrm{G}} = \bigcup_{m\in\mathcal{M}} \mathcal{C}_{\mathrm{G}}(m)$.

The Leech Lattice, Λ_{24}

The Leech lattice can be explicitly constructed as the set of vectors of the form $2^{-3/2}(x_1, x_2, ..., x_{24})$ where the x_i are integers such that $x_1 + x_2 + \cdots + x_{24} \equiv 4x_1 \equiv 4x_2 \equiv \cdots \equiv 4x_{24} \pmod 8$ and for each fixed residue class modulo 4, the 24 bit word, whose 1's correspond to the coordinates i such that x_i belongs to this residue class, is a word in the extended $(24,12)$ binary Golay code. Similarly as in the case of the E_8 lattice, we can partition the Leech lattice into spherical codes. The radius of the

m-th sphere is $r(m) = \sqrt{2(m+1)}$. The spherical codes $\Lambda_{24}(m)$ form an exhaustive partitioning of Λ_{24}, $\Lambda_{24} = \bigcup_{m=1}^{\infty} \Lambda_{24}(m)$.

If we denote by $\mathcal{C}_{\Lambda_{24}}(m)$ a code on the m-th sphere, then a code on the lattice can be defined as $\mathcal{C}_{\Lambda_{24}} = \bigcup_{m\in\mathcal{M}} \mathcal{C}_{\Lambda_{24}}(m)$, where $\mathcal{M} \subset \mathbb{N}$. After the mapping (3.49), the Grassmann code will be $\mathcal{C}_{\mathrm{G}} = \bigcup_{m\in\mathcal{M}} \mathcal{C}_{\mathrm{G}}(m)$.

3.6.3 Decoding

The ML-decoding rule for non-coherent codes is given by (2.40) and is based on the chordal distance measure. The decoding here was performed by exhaustive search, in order to demonstrate the potential benefit of the introduced codes. Suboptimal decoding as described in [1] can be applied in this case as well. There the decision is shifted to the tangent space, instead to be performed in the manifold. Additional improvement is obtained by using a list of candidates in the tangent space which is mapped back in the manifold and compared based on the ML-decoding rule. With the decision taken to the tangent space, the decoding simplifies to decoding on lattices, for which efficient techniques are found in the literature, specifically for the lattices of interest here, E_8, Λ_{16} and Λ_{24}. This simplified decoding performs near the ML decoding in the middle-SNR region [1]. For large blocks T the performance in this region is satisfactory, due to the trend of performance increase with the block length T. Additionally, the lattice codewords do not need to be stored, which enables employment of codes of large cardinality (rate).

Another method for simplified decoding would be the technique introduced in [13], [12]which could benefit from the structure imposed on the Grassmann code after mapping the lattice from the tangent space to the Grassmann manifold.

3.6.4 Examples and Simulation Results

M=2, N=2, T=4

A non-coherent code is constructed from the first three spheres of the E_8 lattice after rotating and mapping the lattice code on the Grassmann manifold $G_{T,M}^{\mathbb{C}}$. The scaling factors are $\alpha_1 = 0.5$, $\alpha_2 = 0.575$ and $\alpha_3 = 0.525$ for the three spheres respectively. The rotation angle is $\phi = \pi/4$. The number of points on the spheres is 240, 2160 and 6720 respectively, which in total gives a code of cardinality 9120 and rate $\eta = 3.2887$ bits per channel use. The results are compared with the Grassmann code obtained from the Golden code [26],[1] with 8QAM (star QAM) with 4096 codewords in total and rate $\eta = 3$ bits per channel use. From Fig. 3.6 we can see that the performance is similar, which favors the code from the E_8 lattice, due to the higher rate. The performance

is further improved if we modify the mapping as in Section 3.7 and use a code from the first sphere of the E_8 lattice, as shown in Fig. 3.6.

M=2, N=2, T=6

A non-coherent code is constructed from the first sphere of the Barnes-Wall lattice Λ_{16}. The scaling factor is $\alpha = 1$. The rotation angle is $\phi = 0.725$. The number of points on the sphere is 4320, which gives a rate of $\eta = 2.013$ bits per channel use. In order to support the prediction that the performance grows with the increase of the block length, the performance is compared to codes for the channel with coherence time $T = 4$. The codes used for comparison are the Grassmann code based on the Golden code from 4PSK.

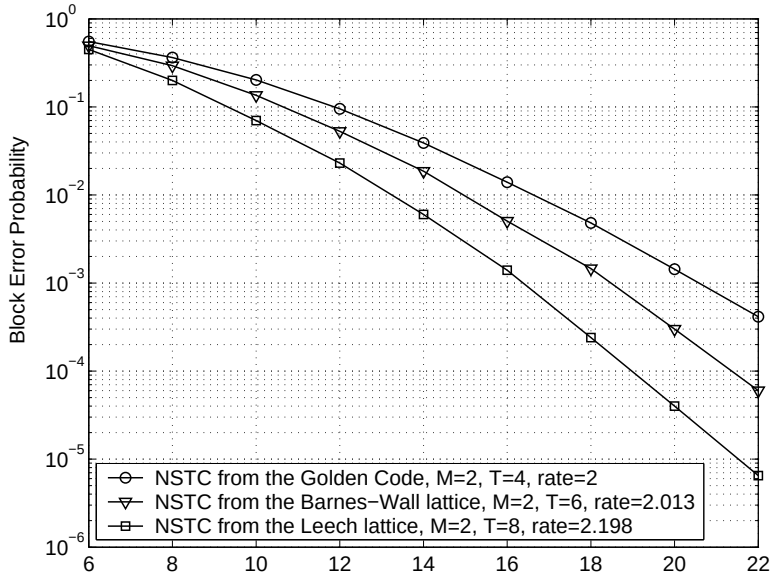

Figure 3.6: Performance comparison for non-coherent space-time codes for $M = 2$ transmit antennas and coherence time $T = 4,\ 6$ and 8.

M=2, N=2, T=8

A non-coherent code is constructed from the first sphere of the Leech lattice Λ_{16}. The scaling factor is $\alpha = 0.5$. The rotation angle is $\phi = 0.7$. The number of points on the sphere is 196560, which gives a rate of $\eta = 2.198$ bits per channel use. In order to support the prediction that the performance grows with the increase of the block length, the performance is compared to codes for the channel with coherence time

$T = 4$ and $T = 6$. The codes used for comparison are the Grassmann code based on the Golden code from 4PSK and rate $\eta = 2$ bits per channel use and the code constructed from the first sphere of the Barnes-Wall λ_{16} lattice. The results are shown in Fig. 3.6. The code is superior to the other codes, probably mainly due to the very special properties of the Leech lattice.

3.7 Modified Mapping from the Tangent Space

We have already argued that the exponential map is non-linear, making the properties of the Grassmann code difficult to control. However, for points in the vicinity of the point where the tangent space is constructed (e.g. the identity), the mapping is almost linear. This property can be used in order to modify the construction. The idea is to construct tangent spaces at several points of the Grassmann manifold (instead of one) and use a small scaling factor α in order to keep the exponential map almost linear. The resulting code would represent a union of "smaller" codes centered at different points of the Grassmann manifold.

3.7.1 Code Construction

In this context we propose the following construction. First, a code $\mathcal{C}_t \subset G^{\mathbb{C}}_{M,T}$ "centered" at $\mathbf{I}_{T,M}$ is constructed. The code $\mathcal{C}_t$ can be any Grassmann code, for example one obtained from a lattice by the exponential map, with a suitable choice of the scaling factor α. Then, another code $\mathcal{C}_u = \{\mathbf{Q}_1, \ldots, \mathbf{Q}_{N_u}\} \subset G^{\mathbb{C}}_{M,T}$ is chosen and for each $\mathbf{Q} \in \mathcal{C}_u$, the code $\mathcal{C}_t$ is translated along the geodesic connecting $\mathbf{I}_{T,M}$ and $\mathbf{Q}$.

The translation of $\mathcal{C}_t$ from $\mathbf{I}_{T,M}$ to $\mathbf{Q}$ is performed by left multiplication by unitary matrix $\tilde{\mathbf{Q}} = [\mathbf{Q}\, \mathbf{Q}^{\perp}]$. In terms of mapping, this can be interpreted as a modification of the "classical" exponential map performed at $\mathbf{I}_{T,M}$. Finally, the Grassmann code $\mathcal{C}$ is represented as the union of the codes "centered" at $\mathbf{Q}_1, \ldots, \mathbf{Q}_{N_u}$,

$$\mathcal{C} = (\tilde{\mathbf{Q}}_1 \mathcal{C}_t) \cup (\tilde{\mathbf{Q}}_2 \mathcal{C}_t) \cup \cdots \cup (\tilde{\mathbf{Q}}_{N_u} \mathcal{C}_t). \tag{3.60}$$

The size of $\mathcal{C}$ is $|\mathcal{C}| = |\mathcal{C}_t| \cdot |\mathcal{C}_u|$ and the rate is $\eta = (\log_2 |\mathcal{C}|)/T$.

Left multiplication of an element of $\mathcal{C}_t$ by unitary matrix $\tilde{\mathbf{Q}} \in U(T)$, transforms it into another element of the Grassmann manifold. With this, the identity $\mathbf{I}_{T,M}$ is transformed into $\mathbf{Q} = \tilde{\mathbf{Q}}\mathbf{I}_{T,M}$. Additionally, for $\mathbf{\Phi}_1, \mathbf{\Phi}_2 \in \mathcal{C}_t$, the transformation leaves the principle angles between $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ unchanged (and thus the chordal distance), since $(\tilde{\mathbf{Q}}\mathbf{\Phi}_2)^{\mathrm{H}}(\tilde{\mathbf{Q}}\mathbf{\Phi}_1) = \mathbf{\Phi}_2^{\mathrm{H}}\mathbf{\Phi}_1$. The choice of $\mathcal{C}_u$ should correspond to a "uniform" distribution (packing) on the manifold, with respect to the chordal distance. Since the number of elements of $\mathcal{C}_u$ is typically small, they can be chosen by linear programming methods, for example the one used by Conway and Sloane [7], or the alternate

projection method introduced in [47]. The maximum number of equidistant points in $G^{\mathbb{C}}_{M,T}$ is T^2 [7]. This is a result of the canonical embedding (2.17) of $G^{\mathbb{C}}_{M,T}$ with chordal distance into a sphere $S^{D-1} \in \mathbb{R}^D$, where $D = T^2 - 1$, and the simplex bound. With a proper choice od $\mathcal{C}_u$, the code performance is dominated by the properties of $\mathcal{C}_t$ (chordal distance and diversity).

3.7.2 Examples and Simulation Results

We present results for $M = 2, N = 2$ and $T = 4$. We take a code $\mathcal{C}_u$ of cardinality 16. This means that we take 16 "equidistant" points and construct tangents space at each point. For the second code we take two examples. The first one is the spherical code presented in Section 3.3. The second one is the code constructed from the E_8 lattice. For the choice of the parameter, $\alpha = 0.35$ "overlap" between the N_u regions is avoided and the performance is optimized.

Comparison was made with other codes from the literature such as the TAST-code [30] and the codes from the authors of [1], for $\eta = 3$ b.c.u. The performance is shown in Fig. 3.7 and Fig. 3.8.

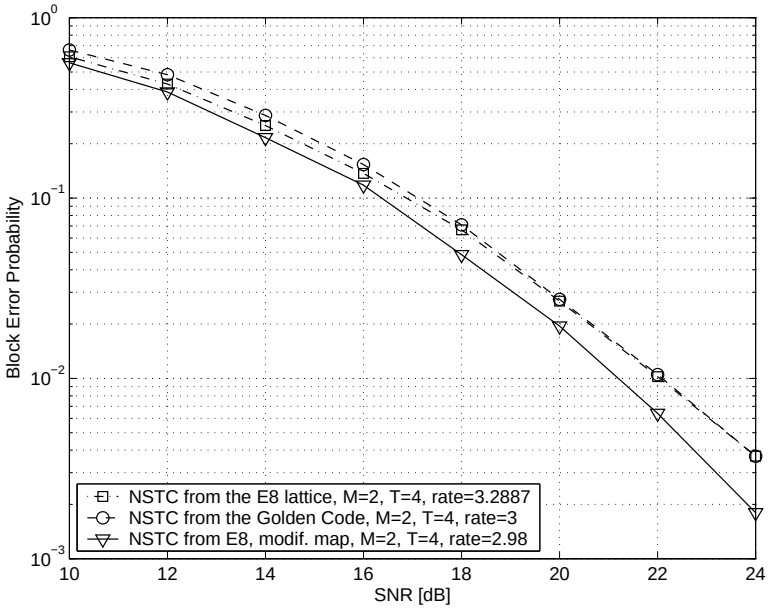

Figure 3.7: Performance comparison of non-coherent space-time codes for $M = 2$ transmit antennas and coherence time $T = 4$.

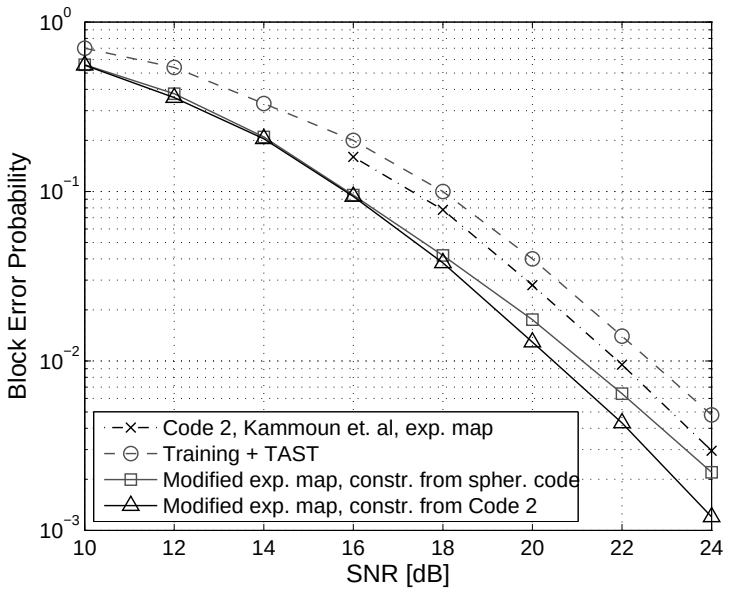

Figure 3.8: Performance comparison of non-coherent space-time codes for $M = 2$ transmit antennas and coherence time $T = 4$.

3.8 Grassmann Codes and Lie Groups: Discussion

The Grassmann codes constructed from constellations in the tangent space at the identity fully exploit the degrees of freedom (dimension) of the tangent space, since no constraint is imposed on the matrix $\mathbf{B}$. This enables construction of spectral efficient Grassmann constellations. However, in the general case, we can not guarantee nor full diversity neither optimal distribution with respect to the chordal distance since, due to the nonlinearity of the exponential map, only few things may be said for the Grassmann codes obtained in this way. Rotation may improve the diversity product distribution between the different Grassmann codewords, as in the case of the construction from lattices, but the rotation angle is obtained heuristically, and does not guarantee optimization of the diversity product between the codewords. However, for a special class of codes, some relations can be established.

The key to the construction of Grassmann codes was the coset representation of the Grasmmann manifold, i.e. the representation of $G^{\mathbb{C}}_{T,M}$ as a quotient space of the unitary group $U(T)$. Instead of using the unitary group $U(T)$, we can constrain ourselves to a subgroup of the unitary group, aiming at obtaining Grassmann codes with special properties. This, of course, limits the number of degrees of freedom (dimension) available for the construction of Grassmann codes.

Let us take for example the symplectic group $Sp(n)$, i.e. the quaternionic unitary group $U(n, \mathbb{H})$. This group is a Lie group and can be equivalently be represented as the group of $2n \times 2n$ matrices $\mathbf{S}$ which satisfy the unitary condition $\mathbf{S}\mathbf{S}^{\mathrm{H}} = \mathbf{S}^{\mathrm{H}}\mathbf{S} = \mathbf{I}_{2n}$ and the symplectic condition $\mathbf{S}^{\mathrm{T}}\mathbf{J}_{2n}\mathbf{S} = \mathbf{J}_{2n}$, where $\mathbf{J}_{2n} = \begin{pmatrix} \mathbf{0} & \mathbf{I}_n \\ -\mathbf{I}_n & \mathbf{0} \end{pmatrix}$. The Lie algebra $\mathfrak{s}_p(n)$ contains $n \times n$ quaternionic matrices of the type $\mathbf{X} = -\mathbf{X}^{\mathrm{T}}$. An $n \times n$ matrix with entries from the quaternions can be equivalently represented by an $2n \times 2n$ matrix with complex entries by using the matrix representation of quaternions

$$a + bi + cj + dk \equiv \begin{pmatrix} a + bi & c + di \\ -c + di & a - bi \end{pmatrix}. \tag{3.61}$$

Let us by G_S denote the subset of $G^{\mathbb{C}}_{k,n}$, $G_S \subset G^{\mathbb{C}}_{k,n}$ which is isomorphic to the quotient space of the symplectic group

$$G_S \cong Sp(n) / \begin{pmatrix} \mathrm{U}(k) & \mathbf{0} \\ \mathbf{0} & \mathrm{U}(2n-k) \end{pmatrix}. \tag{3.62}$$

Tangents of this subset are provided by $\mathfrak{s}_p(n) \cap \mathfrak{b}^{\perp}$, where $\mathfrak{b}$ has the form (2.22). Having the tangents, Grassmann codes can be constructed by using the exponential map, or the modified mapping as in 3.7. We leave this construction for future work with the comment that there is an obvious trade-off between the full exploitation of the available degrees of freedom (dimension) and the potential for simplified decoding.

3.9 Chapter Summary

Geometric Methods for Construction of Non-Coherent Codes

- Achieving capacity for high SNR scenarios requires the input signals to be in the form of isotropically distributed unitary matrices, provided that T satisfies $T \geq \min(M, N) + N$.
- If we choose a code in the tangent space, $\mathcal{X} = \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, a Grassmann code, $\mathcal{C} \subset G^{\mathbb{C}}_{M,T}$, is constructed as

$$\mathcal{C} = \exp(\alpha\mathcal{X})\mathbf{I}_{T,M}.$$

Grassmann Codes from Spherical Codes

- The Grassmann manifold is a normal-homogeneous space. Homogeneity is a natural generalization of spherical symmetry.

- Grassmann codes can be constructed from spherical codes.

- A construction from a high-dimensional spherical code designed for vector quantization is presented. The points on the sphere are distributed on "equidistant" layers by a recursive procedure.

Grassmann Codes from Stiefel Codes

- Class of Grassmann codes obtained by the exponential map, where $\mathbf{B}^{\mathrm{H}}\mathbf{B} = \mathbf{I}_k$

$$\mathbf{\Phi} = \exp(\alpha \mathbf{X})\mathbf{I}_{n,k} = \begin{pmatrix} \mathbf{V}_2(a\mathbf{I}_k)\mathbf{V}_2^H \\ \mathbf{V}_1(b\mathbf{I}_k)\mathbf{V}_2^H \end{pmatrix} = \begin{pmatrix} a\mathbf{I}_k \\ b\mathbf{B} \end{pmatrix},$$

  where $a = \cos\alpha$ and $b = \sin\alpha$.

Grassmann Codes based on Recursive Construction

- For a fixed rate R (spectral efficiency), SNR $\rho \geq 1$ and $T \geq 2M$ [16], the performance of a Grassmann code is expected to grow at least proportionally to $\sqrt{\frac{T}{M}}$. This motivates the construction of higher-dimensional non-coherent codes with $T >> M$, unlike the usual design with $T = 2M$.

- The recursive code construction is summarized as follows

$$\mathcal{B}_0 \overset{\exp(\alpha_0)}{\longrightarrow} \mathcal{C}_0 \overset{\times \mathcal{U}_0}{\longrightarrow} \mathcal{B}_1 \overset{\exp(\alpha_1)}{\longrightarrow} \cdots \overset{\exp(\alpha_k)}{\longrightarrow} \mathcal{C}_k,$$

- For the special case when $\mathcal{B}_0$ is an unitary code, a code for M transmit antennas and coherence time T_k, $\mathcal{C}_k \subset G^{\mathbb{C}}_{M,T_k}$, has the form

$$\mathcal{C}_k = \begin{pmatrix} a_k \cdot \mathbf{I}_M \\ b_k \cdot a_{k-1} \cdot \mathcal{U}_{k-1} \\ \vdots \\ b_k \cdots b_2 \cdot a_1 \cdot \mathcal{U}_1 \times \cdots \times \mathcal{U}_{k-1} \\ b_k \cdots b_1 \cdot a_0 \cdot \mathcal{U}_0 \times \mathcal{U}_1 \times \cdots \times \mathcal{U}_{k-1} \\ b_k \cdots b_1 \cdot b_0 \cdot \mathcal{B}_0 \times \mathcal{U}_0 \times \mathcal{U}_1 \times \cdots \times \mathcal{U}_{k-1} \end{pmatrix}.$$

  The scalars $a_0, a_1, \ldots, a_k$ and $b_0, b_1, \ldots, b_k$ are parameters of the exponential map, satisfying $a_i = \cos\alpha_i$ and $b_i = \sin\alpha_i$.

Grassmann Space-Time Codes from Lattices

- The tangent space at the identity $\mathbf{I}_{T,M}$ can be can be identified with the set of matrices $\mathcal{B} = \left\{ \mathbf{B},\ \mathbf{B} \in \mathbb{C}^{M\times(T-M)} \right\}$. Hence, canonical embedding of the tangent space in the real space $\mathbb{R}^{2M(T-M)}$ is justified.
- Non-coherent space-time codes can be constructed from sets in $\mathbb{R}^{2M(T-M)}$.
- Construction from a lattice $\Lambda \subset \mathbb{R}^{2M(T-M)}$
- The construction is followed by lattice rotation for improving the diversity product.
- Examples: constructions from the Gosset lattice E_8, Barnes-Wall lattice, λ_{16} and the Leech lattice, Λ_{24}.

Modified Mapping from the Tangent Space

- The exponential map is non-linear.
- However, in the vicinity of the identity, the exponential map is almost linear.
- Code construction by translation of a smaller code along geodesics in $G^{\mathbb{C}}_{T,M}$.
- The code $\mathcal{C}$ is constructed as the union of the codes "centered" at predefined points, $\mathbf{Q}_1, \ldots, \mathbf{Q}_{N_u}$,

$$\mathcal{C} = (\tilde{\mathbf{Q}}_1 \mathcal{C}_t) \cup (\tilde{\mathbf{Q}}_2 \mathcal{C}_t) \cup \cdots \cup (\tilde{\mathbf{Q}}_{N_u} \mathcal{C}_t).$$

Grassmann Codes and Lie Groups

- The key to the construction of Grassmann codes was the coset representation of the Grasmmann manifold, i.e. the representation of $G^{\mathbb{C}}_{T,M}$ as a quotient space of the unitary group $U(T)$.
- Instead of using the unitary group $U(T)$, we can constrain ourselves to a subgroup of the unitary group, aiming at obtaining Grassmann codes with special properties. There is trade-off between spectral efficiency and complexity.
- Example: The symplectic group $Sp(n)$, i.e. the quaternionic unitary group $U(n, \mathbb{H})$. This group is a Lie group andthe Lie algebra $\mathfrak{s}_p(n)$ contains $n \times n$ quaternionic matrices of the type $\mathbf{X} = -\mathbf{X}^{\mathrm{T}}$.

- Code constructions from the subset $G_S \subset G_{k,n}^{\mathbb{C}}$ which is isomorphic to the quotient space of the symplectic group

$$G_S \cong Sp(n) / \begin{pmatrix} \mathrm{U}(k) & \mathbf{0} \\ \mathbf{0} & \mathrm{U}(2n-k) \end{pmatrix}.$$

Chapter 4

Wireless Relay Networks: One-Way Relaying and Distributed Space-Time Coding

In this chapter we present novel constructions of distributed space-time codes for non-coherent transmission in wireless relay networks. Distributed space-time coding is a cooperative strategy which exploits the spatial diversity provided by the relays in order to help the transmission. We focus on the one-way (uni-directional) relaying protocol with terminals operating in half-duplex regime. We show that codes based on packings in Grassmann manifolds can be used as distributed space-time codes. Parts of this chapter have been published in [52], [51].

4.1 Introduction

The use of multiple antennas can increase the capacity and the reliability of point-to-point wireless communication links using space-time coding. With the increased interest in ad-hoc networks, several cooperation based methods which exploit spatial diversity provided by antennas of different users have been developed, e.g. [41],[42],

[28], [32], [11], [3]. The improvement is called cooperative diversity and is achieved by having different users in the network cooperate in some way.

Among the first cooperative strategies are amplify-and-forward [28] and decode-and-forward, [3]. An enhanced cooperative strategy, based on coherent distributed space-time coding was introduced in [24]. The scheme requires neither decoding nor channel information at the relays, however, requires full knowledge at the receiver of the channel from the transmitter to the relays and from the relays to the receiver.

Here we focus on non-coherent operation, assuming no channel knowledge at neither the terminals nor at the relays. We will address a one-way relaying protocol for information exchange in wireless relay networks, where the information exchange requires 4 time blocks. In this chapter we generalize the concept of distributed space-time coding to the non-coherent case. The approach here differs to the approach in [25], where distributed differential space-time coding is introduced, since we concentrate on genuine or "pure" non-coherent communication.

Without channel knowledge, the problem of distributed space-time coding becomes more specific. The intuition that the information should be carried by subspaces still holds here and hence we will use some of the results obtained for the non-coherent point-to-point MIMO channel. However, as we will see, we will have to modify the techniques in order to suit to the specifics of the relaying networks.

We recall that in the case of point-to-point MIMO, the problem of non-coherent space-time coding can be interpreted as a constrained packing problem in Grassmann manifolds [65]. This means that the codewords of the non-coherent code should carry information about linear subspaces of a higher-dimensional (complex) space. When generalizing this concept to wireless relay networks, the distributed space-time code obtained at the relays should carry the same structure, i.e. the relays should perform operations such that the distributed code is a subset of a Grassmann manifold.

Although the performance analysis of the distributed non-coherent codes for wireless relay networks is similar to the one of simple point-to-point non-coherent MIMO communication, the fact that the code should be build in a distributive way at the relays, makes this problem specific and different from the coherent distributed space-time coding and non-coherent space-time coding for point-to-point MIMO. In the following we address the specifics of non-coherent distributed space time-coding and propose code constructions based on the geometric approach introduced in Chapter 3.

4.2 System Model

The system model approach mainly follows [24], [33]. We consider a wireless network with $K + 2$ nodes, where K nodes act as relays (in parallel). The relays have low

power and limited computational resources, so that they do not perform decoding. The other two nodes can be distinguished as being capable to encode and decode the data as well as, in general, to send and receive with multiple antennas. The transmitter and the receiver have M respectively N antennas. We assume that the relays have single antennas. Actually, since the relays do not perform decoding, we can assume that this covers the general model where each relay has K_i antennas, used for both transmission and reception. Indeed, since the transmit and received signals at different antennas of the same relay can be processed and designed independently, the network can be transformed to a network with $K = \sum K_i$ single-antenna relays by designing the transmit signal at every antenna of every relay according to the received signal at that antenna only. The equivalent system model holds also for the case when the transmitter and the receiver do not have multiple antennas, but several nodes cooperate in order to transmit (encode) and receive (decode) the information and effectively act as one transmitter and one receiver with multiple antennas. This is a realistic scenario for networks with nodes having power (energy) limitations and particularly for sensor networks where the sensors send common data. Thus, we can consider the above scenario to be quite general for wireless networks.

We recall that the capacity analysis of point-to-point non-coherent MIMO channels reveals that there is no gain from capacity perspective if there are more transmit then receive antennas. In the case of distributed space-time coding for wireless relay networks the relays effectively help the transmitter and create a virtual point-to-point channel with M' transmit antennas, where M' depends on M and K, as we will see in the worked-out examples later in this chapter. Therefore, we will assume that the number of receive antennas is $N = M'$, unless otherwise stated.

The communication scheme is half-duplex, meaning that the network nodes do not transmit and receive simultaneously. The transmission is done in two steps. In the first step the information is encoded into $T_1 \times M$ transmit matrix $\mathbf{S}$, normalized as $E\{\mathrm{tr}(\mathbf{S}^H\mathbf{S})\} = M$. The channel between the transmitter and the relays is assumed constant for T_1 time instants. If we denote by P_1 the average transmit power for one transmission, then the transmit signal is $\sqrt{P_1 T_1/M}\mathbf{S}$. With this, the average power used at the transmitter for the T_1 transmissions is $P_1 T_1$.

Let us denote the signal transmitted by the m-th antenna as $\mathbf{s}_m$, and the channel coefficient between the m-th transmit antenna and the i-th relay as f_{mi}. Further, let us denote the received signal at the i-th relay at time instant τ as $r_{i\tau}$ and the additive noise as $v_{i\tau}$. The channel coefficients and the noise are complex Gaussian with zero mean and unit-variance. The received signal at the relays is then written as

$$\mathbf{r}_i = \sqrt{P_1 T_1/M} \sum_{m=1}^{M} f_{mi}\mathbf{s}_m + \mathbf{v}_n, \tag{4.1}$$

where $\mathbf{r}_i = (r_{i1} \;\; r_{i2} \;\; \cdots \;\; r_{iT_1})^T$ and $\mathbf{v}_i = (v_{i1} \;\; v_{i2} \;\cdots\; v_{iT_1})^T$. In the second step, from

time instant T_1 to time instant $T_1 + T_2$, the relay i sends $t_{i1}, \ldots, t_{iT_2}$. The received signal at the n-th antenna at time instant $T_1 + \tau$ is $y_{\tau n}$ and the additive noise is $w_{\tau n}$ respectively. The channel coefficient from the i-th relay to the n-th receive antenna is g_{in}. The received signal at antenna n is then

$$\mathbf{y}_n = \sum_{m=1}^{R} g_{in}\mathbf{t}_i + \mathbf{z}_n, \tag{4.2}$$

where $\mathbf{y}_n = (y_{1n}\ y_{2n}\ \cdots\ y_{T_2n})^T$, $\mathbf{z}_n = (w_{1n}\ w_{2n}\ \cdots\ w_{T_2n})^T$ and the transmit signal at relay i is

$$\mathbf{t}_i = \sqrt{\frac{P_2T_2}{(P_1+1)T_1}}\left(\mathbf{A_i}\mathbf{r}_i + \mathbf{A}_i'\mathbf{r}_i^*\right),\ i = 1, 2, \ldots, K. \tag{4.3}$$

Here for the relay matrices we assume $\mathbf{A}_i = \mathbf{0}$ or $\mathbf{A}_i' = \mathbf{0}$. The more general approach would assume that the matrix $\begin{pmatrix} \Re(\mathbf{A}_i + \mathbf{A}_i') & -\Im(\mathbf{A}_i - \mathbf{A}_i') \\ \Im(\mathbf{A}_i + \mathbf{A}_i') & \Re(\mathbf{A}_i - \mathbf{A}_i') \end{pmatrix}$ is a $2T_2 \times 2T_1$ orthogonal matrix. Here we assume the simplified case. For details referee to [24]. $\mathbf{A}_i$ and $\mathbf{A}_i'$ are $T_2 \times T_1$ matrices with orthonormal columns. This ensures an average transmit power at relay i of P_2T_2, i.e. an average power per transmission instant of P_2. It also makes the protocol equitable between different users and different time instants. The choice of different T_1 and T_2 adds more flexibility to the problem of non-coherent distributed space-time coding, compared to the standard approaches in the literature, e.g. [24][33], where $T_1 = T_2$ is assumed.

Let us now denote $\mathbf{f}_i = (f_{1i}\ f_{2i}\ \cdots\ f_{Mi})^T$ and $\mathbf{g}_i = (g_{i1}\ g_{i2}\ \cdots\ g_{iN})^T$. The equivalent channel model between the transmitter and the receiver can now be written as

$$\mathbf{Y} = \sqrt{\frac{P_1P_2T_2}{(P_1+1)M}}\mathbf{\Phi}\mathbf{H} + \mathbf{W}, \tag{4.4}$$

where

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_N \end{pmatrix}, \tag{4.5}$$

$$\mathbf{\Phi} = \begin{pmatrix} \hat{\mathbf{A}}_1\hat{\mathbf{S}} & \hat{\mathbf{A}}_2\hat{\mathbf{S}} & \cdots & \hat{\mathbf{A}}_K\hat{\mathbf{S}} \end{pmatrix}, \tag{4.6}$$

$$\mathbf{H} = \begin{pmatrix} \hat{\mathbf{f}}_1\mathbf{g}_1^T & \hat{\mathbf{f}}_2\mathbf{g}_2^T & \cdots & \hat{\mathbf{f}}_K\mathbf{g}_K^T \end{pmatrix}^T, \tag{4.7}$$

$$\mathbf{W} = \begin{pmatrix} \sqrt{\frac{P_2T_2}{(P_1+1)T_1}}\sum_{i=1}^{K} g_{i1}\hat{\mathbf{A}}_i\hat{\mathbf{v}}_i + \mathbf{z}_1 \\ \vdots \\ \sqrt{\frac{P_2T_2}{(P_1+1)T_1}}\sum_{i=1}^{K} g_{iN}\hat{\mathbf{A}}_i\hat{\mathbf{v}}_i + \mathbf{z}_N \end{pmatrix}^T, \tag{4.8}$$

and the relay matrices are given as

$$\hat{\mathbf{A}}_i = \mathbf{A}_i,\ \hat{\mathbf{f}}_i = \mathbf{f}_i,\ \hat{\mathbf{v}}_i = \mathbf{v}_i,\ \hat{\mathbf{S}} = \mathbf{S} \text{ if } \mathbf{A}_i' = \mathbf{0}, \tag{4.9}$$

$$\hat{\mathbf{A}}_i = \mathbf{A}_i',\ \hat{\mathbf{f}}_i = \mathbf{f}_i^*,\ \hat{\mathbf{v}}_i = \mathbf{v}_i^*,\ \hat{\mathbf{S}} = \mathbf{S}^* \text{ if } \mathbf{A}_i = \mathbf{0}. \tag{4.10}$$

We can say that the problem of non-coherent distributed space-time coding is a problem of choice of the relay matrices $\hat{\mathbf{A}}_i$ as well as a suitable code $\hat{\mathcal{S}} = \{\hat{\mathbf{S}} \mid \hat{\mathbf{S}} \in \mathrm{C}^{T_1 \times M}\}$ for the first part of the transmission, such that the codewords $\mathbf{\Phi}$ of the distributed non-coherent space-time code describe elements of a Grassmann manifold. Practically, non-coherent communication where the information is carried by subspaces is possible whenever the network parameters (M and K) and the relay matrices are such that it is possible to construct a distributed Grassmann code.

4.3 Pairwise Error Probability

The equivalent system model of the relay network can be considered as a non-coherent system model with $M' = MK$ transmit antennas, N receive antennas and coherence time T_2. In the special case of interest here, (4.4) can be rewritten as

$$\mathbf{Y} = \sqrt{\frac{P_1 P_2 K T_2}{(P_1+1)M'}}\mathbf{\Phi}\mathbf{H} + \mathbf{W}, \tag{4.11}$$

where the noise is of the form (4.10). Then the PEP Chernoff bound will have the form

$$P(\mathbf{\Phi}_1, \mathbf{\Phi}_2) \leq \mathrm{E}_{g_{ji}} \left\{ \frac{1}{2} \prod_{i=1}^{N} \det{}^{-1} \left[\mathbf{I}_{M'} + \underline{\varrho}_i (\mathbf{I}_{M'} - \underline{\mathbf{\Delta}}^{\mathrm{H}} \underline{\mathbf{\Delta}}) \right] \right\}, \tag{4.12}$$

where

$$\underline{\varrho}_i = \frac{(\rho_i \frac{T_2}{M'})^2}{4(1+\rho_i \frac{T_2}{M'})}, \tag{4.13}$$

and

$$\begin{aligned} \rho_i &= \frac{\frac{P_1 P_2 K T_2}{P_1+1}}{T_2 + T_1 \cdot \frac{P_2 T_2}{(1+P_1)T_1} \sum_{j=1}^{K} |g_{ji}|^2} \\ &= \frac{P_1 P_2 K}{1 + P_1 + P_2 \sum_{j=1}^{K} |g_{ji}|^2}. \end{aligned} \tag{4.14}$$

The problem is simplified if we assume that $\sum_{j=1}^{K} |g_{ji}|^2 \approx K$. This approximation is valid for large K. Then (4.12) becomes

$$P(\mathbf{\Phi}_1, \mathbf{\Phi}_2) \leq \frac{1}{2} \det \left[\mathbf{I}_{M'} + \underline{\varrho} (\mathbf{I}_{M'} - \underline{\mathbf{\Delta}}^{\mathrm{H}} \underline{\mathbf{\Delta}}) \right]^{-N}, \tag{4.15}$$

where

$$\underline{\varrho} = \frac{(\rho \frac{T_2}{M'})^2}{4(1+\rho \frac{T_2}{M'})} \text{ and } \rho = \frac{P_1 P_2 K}{1 + P_1 + P_2 K}. \tag{4.16}$$

For the code constructions presented here, we will assume a finite number of relays. Therefore, we will have in mind that the above analysis is only approximately true. Nevertheless, the same analogy would hold as in the case of coherent distributed space-time coding, see for example [24] for further details.

4.4 Optimum Power Allocation

We denote the power of the transmitter when it transmits as P_1, and the power of each relay when it transmits as P_2. Then the total average power used in the network for the $T_1 + T_2$ symbol periods is $P_1T_1 + KP_2T_2$ or, equivalently the average power used for one symbol period (also including the time when the device is not active) is

$$P = \frac{P_1T_1}{T_1 + T_2} + K\frac{P_2T_2}{T_1 + T_2}. \tag{4.17}$$

One natural question is how to allocate power between the transmitter and the relays if P is fixed. In this section, we find the optimum power allocation such that the PEP is minimized. The pairwise error probability (4.12) is minimized when ϱ is maximized. It can be shown that (see appendix A.3), for a given average system power P, ϱ is maximized when

$$P_1 = P_2K \cdot \sqrt{\frac{T_2}{T_1}}. \tag{4.18}$$

From (4.17) and (4.18), we get

$$P_1 = \frac{T_1 + T_2}{T_1 + \sqrt{T_1T_2}}P, \quad P_2 = \frac{T_1 + T_2}{(T_2 + \sqrt{T_1T_2})K} \tag{4.19}$$

4.5 Code Construction

As already described, the problem of non-coherent distributed space-time coding is the problem of choosing suitable codes $\hat{\mathbf{S}}$ for the first part of the transmission as well as relay matrices $\hat{\mathbf{A}}_i$ to make the distributed space-time code to be a Grassmann code. Basically, every code which can be constructed in a distributed way at the relays, can be used. It is necessary that the columns of the transmit codeword matrix can be obtained from the transmitted signal by linear transformation, or alternatively, by linear transformation of its' complex conjugate. Here, we present the construction of non-coherent distributed space-time codes (NDSTCs) based on Alamouti and Sp(2) codes, as discussed in Chapter 3.

The constructions proposed here are based on the geometric constructions presented in Sec. 3.4 and 3.5. We recall that when according to the geometric method,

Grassmann codes are designed from a code $\mathcal{B}$ in the tangent space at the identity element of the Grassmann manifold. We recall that according to Sec. 3.4, when the codewords of $\mathcal{B}$ fulfill $\mathbf{B}^{\mathrm{H}}\mathbf{B} = \mathbf{I}$, the Grassmann codes obtained after the exponential map have the form

$$\mathbf{\Phi} = \begin{pmatrix} a \cdot \mathbf{I} \\ b \cdot \mathbf{B} \end{pmatrix}. \tag{4.20}$$

where $a = \cos\alpha$ and $b = \sin\alpha$.

In the following we present constructions where $\mathbf{B}$ comes for from Alamouti and $Sp(2)$ codes. For these codes this prerequisite can be fulfilled and the Grassmann codes will have the above form. The codes in this form can be used as distributed codes with the appropriate choice of the relay matrices.

4.5.1 Construction from Alamouti Code

As an example, we consider the case with a transmitter equipped with a single transmit antenna and two relays which assist the transmission, each equipped with single antenna. In this case, a non-coherent code derived from the Alamouti code [2] can be used.

According to Sec. 3.4, the non-coherent code obtained by exponential map of an Alamouti code can be written in the form of

$$\mathbf{\Phi} = \begin{pmatrix} \cos\alpha & 0 \\ 0 & \cos\alpha \\ \sin\alpha \cdot s_1 & -\sin\alpha \cdot s_2^* \\ \sin\alpha \cdot s_2 & \sin\alpha \cdot s_1^* \end{pmatrix}. \tag{4.21}$$

We note that if the signal sent from the transmitter is in the form

$$\mathbf{S} = \begin{pmatrix} \cos\alpha \\ 0 \\ \sin\alpha \cdot s_1 \\ \sin\alpha \cdot s_2 \end{pmatrix}, \tag{4.22}$$

and the relay matrices are chosen as follows

$$\mathbf{A}_1 = \mathbf{I}_4, \mathbf{A}_1' = \mathbf{0}, \mathbf{A}_2 = \mathbf{0}, \mathbf{A}_2' = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{4.23}$$

then, at the relays, the codewords of the distributed code are given as

$$\mathbf{\Phi} = (\ \mathbf{A}_1\mathbf{S} \quad \mathbf{A}_2'\mathbf{S}^*\ ). \tag{4.24}$$

We notice that, in the first part of the transmission, the zeros can be omitted. This corresponds to $T_1 \neq T_2$. It is important to have this on mind in the normalization of the transmitted signals and for the calculation of the power allocation. At the transmitter, we send

$$\mathbf{S} = \begin{pmatrix} \cos\alpha \\ \sin\alpha \cdot s_1 \\ \sin\alpha \cdot s_2 \end{pmatrix}. \tag{4.25}$$

Now, the relay matrices should be chosen as

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \mathbf{A}_1' = \mathbf{0}, \mathbf{A}_2 = \mathbf{0}, \mathbf{A}_2' = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{4.26}$$

4.5.2 Construction from Sp(2) Code

Another code of the form (3.9) can be constructed from Sp(2) codes [23] which arise from the symplectic Lie group Sp(2). This code can be seen as a generalization of Alamouti code to dimension four. Its symbol rate is one. However, we note that this code is not a orthogonal code. The code has the following structure

$$\mathcal{B} = \left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{V}_1\mathbf{V}_2 & \mathbf{V}_1\mathbf{V}_2^{\mathrm{H}} \\ -\mathbf{V}_1^{\mathrm{H}}\mathbf{V}_2 & (\mathbf{V}_1\mathbf{V}_2)^{\mathrm{H}} \end{bmatrix} \right\}, \tag{4.27}$$

where, for i=1,2,

$$\mathbf{V}_i = \frac{1}{\sqrt{|\epsilon_i|^2 + |\kappa_i|^2}} \begin{bmatrix} \epsilon_i & -\kappa_i^* \\ \kappa_i & \epsilon_i^* \end{bmatrix}. \tag{4.28}$$

and $\epsilon_i \in \mathcal{F}_i, \kappa_i \in \mathcal{G}_i$ are information symbols from the finite sets $\mathcal{F}_i$ and $\mathcal{G}_i$. The choice of $\mathcal{F}_i$ and $\mathcal{G}_i$ are arbitrary. Sufficient and necessary condition for full diversity of Sp(2) code with PSK signals was provided in [23]. If we define

$$\begin{aligned} u_1 &= \frac{\epsilon_1\epsilon_2 - \kappa_1\kappa_2^*}{\sqrt{2}\prod_{i=1}^{2}\sqrt{|\epsilon_i|^2 + |\kappa_i|^2}}, \\ u_2 &= -\frac{\epsilon_1^*\kappa_2^* + \kappa_1^*\epsilon_2}{\sqrt{2}\prod_{i=1}^{2}\sqrt{|\epsilon_i|^2 + |\kappa_i|^2}}, \\ u_3 &= -\frac{\epsilon_1^*\epsilon_2 - \kappa_1^*\kappa_2^*}{\sqrt{2}\prod_{i=1}^{2}\sqrt{|\epsilon_i|^2 + |\kappa_i|^2}}, \end{aligned}$$

and

$$u_4 = \frac{\epsilon_1 \kappa_2^* + \kappa_1 \epsilon_2}{\sqrt{2}\prod_{i=1}^{2}\sqrt{|\epsilon_i|^2 + |\kappa_i|^2}}, \tag{4.29}$$

from (4.28), the codewords of $\mathcal{B}$ (4.27) can be written as

$$\mathbf{B} = \begin{bmatrix} u_1 & -u_2^* & -u_3^* & u_4 \\ u_2 & u_1^* & -u_4^* & -u_3 \\ u_3 & -u_4^* & u_1^* & -u_2 \\ u_4 & u_3^* & u_2^* & u_1 \end{bmatrix}. \tag{4.30}$$

Therefore, the Sp(2) code is actually a special kind of a quasi-orthogonal space-time block code [21]. By special choices of the information symbols u_i, the Sp(2) code can be unitary [23]. So, the non-coherent code obtained from Sp(2) using exponential map (2.23), can be written in the form (3.9).

This code can be used in a network where we have 1 transmit antenna and 4 relay nodes, or in a network where we have 2 transmit antennas and 2 relay nodes, each equipped with 2 antennas. In the following, we present code constructions for these network configurations.

M=1, K=4

For the case of $M = 1$ and $K = 4$, in the first part of the transmission we send

$$\mathbf{S} = \begin{pmatrix} a & bu_1 & bu_2 & bu_3 & bu_4 \end{pmatrix}^{\mathrm{T}}. \tag{4.31}$$

where u_1, u_2, u_3 and u_4 are built from the information symbols $\epsilon_1, \epsilon_2, \kappa_1$ and κ_2 according to (4.29).

We define the relay matrices as follows:

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \mathbf{A}_1' = \mathbf{0}, \mathbf{A}_2' = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \mathbf{A}_2 = \mathbf{0},$$

$$\mathbf{A}_3' = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \mathbf{A}_3 = \mathbf{0}, \mathbf{A}_4 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \mathbf{A}_4' = \mathbf{0}. \tag{4.32}$$

The distributed code built at the relays is given as

$$\mathbf{\Phi} \;=\; (\; \mathbf{A}_1\mathbf{S} \quad \mathbf{A}_2'\mathbf{S}^* \quad \mathbf{A}_3'\mathbf{S}^* \quad \mathbf{A}_4\mathbf{S} \;). \tag{4.33}$$

M=2, K=2

Here we address the case when the transmitter has two antennas and there are two relays in the network, each equipped with two antennas. We note that this is necessary in order that the relays are able to receive the signal from the transmitter. In the first part of the transmission we send

$$\mathbf{S} = \begin{pmatrix} a & 0 & bu_1 & bu_2 & bu_3 & bu_4 \\ 0 & a & -bu_2^* & bu_1^* & -bu_4^* & bu_3^* \end{pmatrix}^{\mathrm{T}}. \tag{4.34}$$

We define the relay matrices as follows:

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \mathbf{A}_1' = \mathbf{0}, \mathbf{A}_2' = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \mathbf{A}_2 = \mathbf{0}.$$

The distributed code is built at the relays as:

$$\mathbf{\Phi} = (\; \mathbf{A}_1\mathbf{S} \quad \mathbf{A}_2'\mathbf{S}^* \;) \tag{4.35}$$

It is easy to show that this is a Grassmann code.

4.5.3 Construction from Recursive Grassmann Codes

A subclass of the codes of the form (3.25) obtained by the recursive procedure introduced in (3.5), are also applicable as distributed space-time codes. Indeed, if $\mathcal{B}_0$ and the codes used for multiplication $\mathcal{U}_0, \mathcal{U}_1, \ldots, \mathcal{U}_{k-1}$ are Alamouti or Sp(2) code, then the codewords $\mathbf{\Phi}$ of the distributed code can always be written in the form

$$\mathbf{\Phi} = (\ \hat{\mathbf{A}}_1\hat{\mathbf{S}} \quad \hat{\mathbf{A}}_2\hat{\mathbf{S}} \quad \ldots \quad \hat{\mathbf{A}}_K\hat{\mathbf{S}}\ ). \tag{4.36}$$

We note that the distributed codes constructed in this chapter can all be represented in this form. They correspond to recursive codes where only one recursion is performed.

4.6 Decoding

According to the system model (4.4), the code at the relays is a non-coherent code. The ML-decoding rule is then

$$\hat{\mathbf{\Phi}} = \arg\max_{\mathbf{\Phi}\in\mathcal{C}} \|\mathbf{Y}^{\mathrm{H}}\mathbf{\Phi}\|_{\mathrm{F}}^2 = \arg\max_{\mathbf{\Phi}\in\mathcal{C}} \left[\mathrm{tr}(\mathbf{\Phi}^{\mathrm{H}}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\mathbf{\Phi})\right], \tag{4.37}$$

where $\mathbf{Y}$ is the received signal at receiver. Let us take the Grassmann codes in $G_{T,M}^{\mathbb{C}}$ which can be written in the form

$$\mathbf{\Phi} = \begin{pmatrix} a\mathbf{I}_M \\ b\mathbf{B} \end{pmatrix}, \tag{4.38}$$

where $a = \cos(\alpha), b = \sin(\alpha)$ and $\mathbf{B} \in \mathbb{C}^{(T-M)\times M}$. We decompose the received signal as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \tag{4.39}$$

where $\mathbf{Y}_1$ is a $M \times M$ matrix and $\mathbf{Y}_2$ is a $(T-M)\times M$ matrix. For the special case when $\mathbf{B}^{\mathrm{H}}\mathbf{B} = \mathbf{I}_M$, we have

$$\begin{aligned} \hat{\mathbf{\Phi}} &= \arg\max_{\mathbf{\Phi}}\{\mathrm{tr}[(a\mathbf{Y}_1 + b\mathbf{B}^{\mathrm{H}}\mathbf{Y}_2)(a\mathbf{Y}_1^{\mathrm{H}} + b\mathbf{Y}_2^{\mathrm{H}}\mathbf{B})]\} \\ &= \arg\max_{\mathbf{\Phi}}\{2ab\cdot\Re[\mathrm{tr}(\mathbf{B}^{\mathrm{H}}\mathbf{Y}_2\mathbf{Y}_1^{\mathrm{H}})]\}. \end{aligned} \tag{4.40}$$

If $\mathbf{B}$ comes from an Alamouti code, the decoding procedure can be further simplified.

4.6.1 Decoding of NDSTCs based on Alamouti Codes

If $\mathbf{B}$ comes from an Alamouti code, the decoding procedure can be further simplified. When $\mathbf{B}$ resembles a codeword from an Alamouti code, it has the form

$$\mathbf{B} = \begin{pmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{pmatrix}.$$

We denote the received signal as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \\ y_{41} & y_{42} \end{pmatrix}.$$

Then, the decoding rule (4.40) becomes

$$\begin{aligned} \hat{\mathbf{\Phi}} &= \arg\max_{\mathbf{\Phi}\in\mathcal{C}} \{\Re[\mathrm{tr}(\mathbf{B}^{\mathrm{H}}\mathbf{Y}_2\mathbf{Y}_1^{\mathrm{H}})]\} \\ &= \arg\max_{\mathbf{\Phi}\in\mathcal{C}} \left\{ \Re\left(\mathrm{tr}\left[\begin{pmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{pmatrix}^{\mathrm{H}} \cdot \begin{pmatrix} y_{31} & y_{32} \\ y_{41} & y_{42} \end{pmatrix} \cdot \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}^{\mathrm{H}} \right] \right) \right\} \\ &= \arg\max_{\mathbf{\Phi}\in\mathcal{C}} \left\{ \Re\left[\mathrm{tr}\begin{pmatrix} \varepsilon_1 s_1^* + \varepsilon_2 s_2^*, & \varepsilon_3 s_1^* + \varepsilon_4 s_2^* \\ \varepsilon_2 s_1 - \varepsilon_1 s_2, & \varepsilon_4 s_1 - \varepsilon_3 s_2 \end{pmatrix} \right] \right\} \\ &= \arg\max_{\mathbf{\Phi}\in\mathcal{C}} \left\{ \Re\left(\varepsilon_1 s_1^* + \varepsilon_2 s_2^* + \varepsilon_4 s_1 + \varepsilon_3 s_2 \right) \right\}, \end{aligned} \tag{4.41}$$

where

$$\begin{aligned} \varepsilon_1 &= y_{11}y_{31} + y_{21}y_{32}, \\ \varepsilon_2 &= y_{11}y_{41} + y_{21}y_{42}, \\ \varepsilon_3 &= y_{12}y_{31} + y_{22}y_{32}, \\ \varepsilon_4 &= y_{12}y_{41} + y_{22}y_{42}. \end{aligned}$$

From the above equation, we can see that the information symbols can be decoded independently.

4.7 Examples and Simulation Results

Several distributed codes of the form (3.25) have been constructed. The performance is compared with distributed differential codes, which are proposed in [25]. For larger block lengths the non-coherent DSTCs slightly outperform the differential DSTCs. For lower block lengths, the differential codes perform better. This is in part due to

the hidden correlation between the elements of the matrix $\mathbf{H}$ in (4.8). Namely, the elements of $\mathbf{H}$ are not exactly independently distributed, since they arise from the product of the channel coefficients from the transmitter to the relays and the channel coefficients from the relays to the receiver. The differential codes, in a way, estimate the channel and compensate for this effect.

M=1, K=2 and N=2

Here we use the code designed in 4.5.1, and a recursive construction. In the first part of transmission, we do not send the unnecessary zeros. That means we transmit the signal as defined in (4.25) at the transmitter. The starting code $\mathcal{B}_0$ is an Alamouti code, with $s_1, s_2 \in$ QPSK. The other parameters are summarized in Tab. 4.1

α	T_1	T_2	$\mathcal{U}$	$\eta[b.p.c.u]$
$\alpha_0 = 0.9$	3	4	-	0.5714
$\alpha_0 = 0.9$, $\alpha_1 = 1.0$	5	6	$\mathcal{U}_0 -$ Alamouti $u_1, u_2 \in$ BPSK	0.5455
$\alpha_0 = 0.9$, $\alpha_1 = 1.0$ $\alpha_2 = 1.1$	7	8	$\mathcal{U}_0, \mathcal{U}_1 -$ Alamouti $u_1, u_2 \in$ BPSK	0.5383
$\alpha_0 = 0.9$, $\alpha_1 = 1.0$ $\alpha_2, \alpha_3 = 1.1$ $\alpha_4 = 1.2$	11	12	$\mathcal{U}_0, \mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3 -$ Alamouti $u_1, u_2 \in$ BPSK	0.5217

Table 4.1: Parameters of the distributed code based on the Alamouti code.

The simulation results are given in Fig. 4.1. We can see that for $T_1 = 11$ and $T_2 = 12$, the performance of non-coherent DSTCs outperform the differential DSTCs, and at a slightly higher transmission rate (0.522 bits/c.u. compared to 0.5 bits/c.u.).

M=1, K=4 and N=4

For this case, we use the code designed in 4.5.2, and the recursive procedure introduced in 3.5. The starting code $\mathcal{B}_0$ is an Sp(2) code, with $\epsilon_1, b\kappa_1 \in$ QPSK and $\epsilon_2, \kappa_2 \in$ 3PSK. The other parameters are summarized in Tab. 4.2

The simulation results are given in Fig. 4.2. Since finding a good bit-to-symbol mapping in this case is not trivial, we compared the performance on block error basis. For block lengths $T_1 = 9$ and $T_2 = 12$, the non-coherent codes outperform the differential codes, at a slightly higher data rate (0.6829 bits/c.u. versus 0.6462

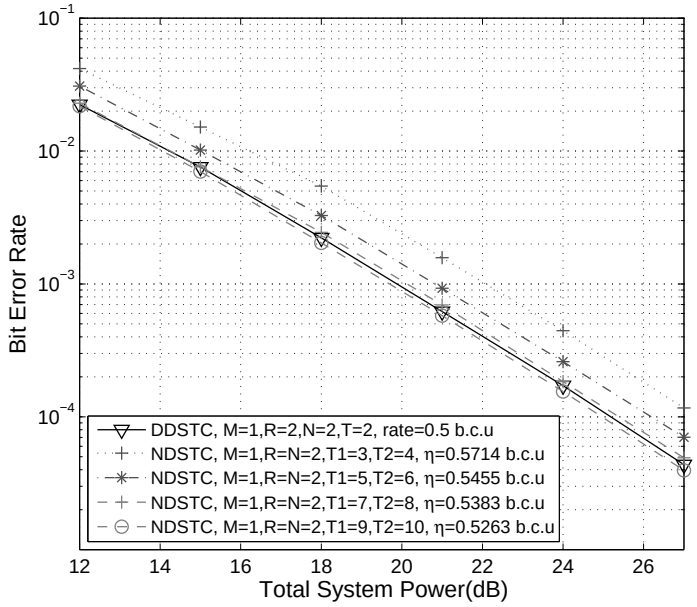

Figure 4.1: Comparison of non-coherent and differential distributed space-time codes, M=1, K=2 and N=2.

α	T_1	T_2	$\mathcal{U}$	$\eta[b.p.c.u]$
$\alpha_0 = 0.8$, $\alpha_1 = 1.0$	9	12	$\mathcal{U}_0 = \mathrm{Sp}(2)$ $\epsilon_1, \kappa_1 \in$ QPSK $\epsilon_2, \kappa_2 \in$ 3PSK	0.6829

Table 4.2: Parameters of distributed code based on the Sp(2) code.

bits/c.u.). With an appropriate bit-to-symbol mapping, the difference would be more significant, since the non-coherent DSTCs have larger block lengths compared to the differential DSTCs, and thus less bits on average will be in error.

M=2, K=2 and N=4

For this case, we use the code designed in 4.5.2, and the recursive procedure introduced. The starting code $\mathcal{B}_0$ is an Sp(2) code with $\epsilon_1, \kappa_1 \in$ QPSK and $\epsilon_2, \kappa_2 \in$ 3PSK. The other parameters are summarized in Tab. 4.3 The simulation results are given in Fig. 4.2. For block lengths $T_1 = 10$ and $T_2 = 12$, the non-coherent codes outperform the differential codes, at a higher data rate (0.6518 b.c.u versus 0.6462 b.c.u). Again, the comparison is made on a block basis.

α	T_1	T_2	$\mathcal{U}$	$\eta[b.p.c.u]$
$\alpha_0 = 0.8$, $\alpha_1 = 1.0$	10	12	$\mathcal{U}_0 = \mathrm{Sp}(2)$ $\epsilon_1, \kappa_1 \in$ QPSK $\epsilon_2, \kappa_2 \in$ 3PSK	0.6518

Table 4.3: Parameters of the distributed code based on the Sp(2) code.

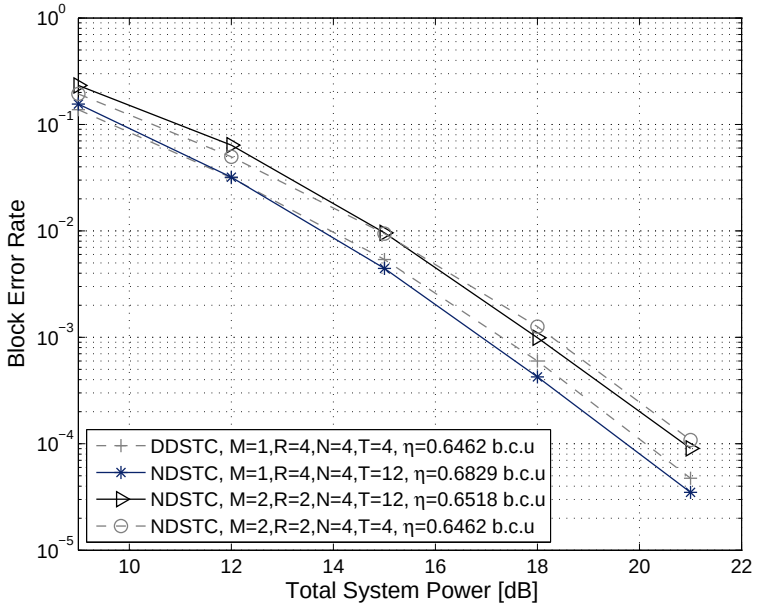

Figure 4.2: Comparison of non-coherent DSTCs and differential DSTCs, M=2, K=2, N=4.

4.8 Chapter Summary

- We introduce non-coherent communication in one-way wireless relay networks, where the relays and the receiver do not require channel information.

- The focus is on generalizing the concept of distributed space-time coding to the non-coherent case.

Distributed Non-Coherent Space-Time Coding in One-Way Relaying Networks

- The equivalent channel model between the transmitter and the receiver can be written as

$$\mathbf{Y} = \sqrt{\frac{P_1 P_2 T_2}{(P_1+1)M}}\mathbf{\Phi}\mathbf{H} + \mathbf{W},$$

- The distributed code constructed at the relays is given as

$$\mathbf{\Phi} = \left(\begin{array}{cccc} \hat{\mathbf{A}}_1\hat{\mathbf{\Phi}} & \hat{\mathbf{A}}_2\hat{\mathbf{\Phi}} & \cdots & \hat{\mathbf{A}}_K\hat{\mathbf{\Phi}} \end{array} \right).$$

- The optimum power allocation for a given average system power P is

$$P_1 = P_2 K \cdot \sqrt{\frac{T_2}{T_1}}.$$

Codes Design

- The problem of non-coherent distributed space-time coding is the problem of choosing suitable codes $\hat{\mathbf{\Phi}}$ for the first part of the transmission as well as relay matrices $\hat{\mathbf{A}}_i$ to make the distributed space-time code to be a Grassmann code.

$$\mathbf{\Phi} = \left(\begin{array}{c} \cos\alpha \cdot \mathbf{I}_M \\ \sin\alpha \cdot \mathbf{B} \end{array} \right) = \left(\begin{array}{c} a \cdot \mathbf{I}_M \\ b \cdot \mathbf{B} \end{array} \right).$$

  where $a = \cos\alpha$ and $b = \sin\alpha$.

- Construction where $\mathbf{B}$ comes from codes such as Alamouti, Sp(2) or recursive Grassmann code.

4.8.1 Decoding

For the special case when $\mathbf{B}$ is a unitary matrix, we have

$$\begin{aligned} \hat{\mathbf{\Phi}} &= \arg\max_{\mathbf{\Phi}} \{\mathrm{tr}[(a\mathbf{Y}_1 + b\mathbf{B}^{\mathrm{H}}\mathbf{Y}_2)(a\mathbf{Y}_1^{\mathrm{H}} + b\mathbf{Y}_2^{\mathrm{H}}\mathbf{B})]\} \\ &= \arg\max_{\mathbf{\Phi}} \{2ab \cdot \Re[\mathrm{tr}(\mathbf{B}^{\mathrm{H}}\mathbf{Y}_2\mathbf{Y}_1^{\mathrm{H}})]\} \end{aligned}$$

If $\mathbf{B}$ is an Alamouti code, the decoding procedure can be made on symbol-by-symbol basis.

Chapter 5

Wireless Relay Networks: Non-Coherent Two Way Relaying

In this chapter we analyze the non-coherent two-way relay channel. We first derive bounds on the achievable two-way rate in the high SNR regime. Additionally, we derive the degree of freedom of this network and show that amplify-and-forward is optimal in the high SNR regime. Besides the derivation of the bounds for the achievable rate, we present two schemes which are appropriate for communication over the two-way relaying channel. The first scheme is a genuine non-coherent scheme where the information is carried by subspaces. The second one is a differential scheme, where the information is carried by the difference between the transmit matrices. Parts of this chapter were published in [64], [57], [58], [59], [50]

5.1 Introduction

The one-way relaying protocols suffer from a loss in spectral efficiency due to the half-duplex constraint of the terminals. In order to increase the spectral efficiency of such a relay network, a bi-directional (two-way) communication between two terminals where the relay assists in the two-way communication was introduced in [39].

The connection between two-way relaying and network coding was established in [35], where protocols for two-way (bi-directional) relaying based on network coding on symbol level were introduced. The main idea is that the relays combine the information from the terminals and broadcast it in the next stage. Each terminal then subtracts its own contribution and decides about the signal transmitted from the other terminal. The two-way relaying schemes promise throughput gain compared to one-way relaying, by saving time slots in the broadcast stage.

The two-way channel was first studied by Shannon[43], where he found inner and outer bounds on the capacity. In [39], the authors investigate the achievable two-way rate for different relaying techniques in two-way wireless relay channels.

The schemes described in [39],[35] require channel knowledge at the terminals and/or the relays. Here, we focus on two-way relaying without channel knowledge requirements at the terminals and the relays. Note that an analysis where the relays do not require channel knowledge is performed in [62]. However, the authors there assume channel knowledge at the receivers.

First, we will derive bounds on the achievable two-way rate in the relay network. The analysis is related to the capacity results for the point-to-point MIMO block Rayleigh fading channel without channel knowledge assumption at the receiver (neither at the transmitter), derived in [65].

Second, we will present two schemes which are appropriate for communication over the two-way relaying channel. The first one is a genuine non-coherent scheme, where the information is carried by subspaces. The scheme is motivated by the constructions for the point-to-point channel. The second scheme is differential scheme, where the information is carried by the difference between the transmit matrices.

5.2 System Model

We consider a wireless network with two terminals and K relay nodes (in parallel), each equipped with M antennas, as shown in Fig. 5.1. It is important to make the remark that in the two-way relaying setup we usually assume that the terminals have the same number of antennas. This is due to the fact that both terminals exchange information simultaneously and are symmetric in any aspect regarding transmission, reception, decoding etc. We know that for a point-to-point channel, under the non-coherent assumption, having different number of transmit and receive antennas does not bring advantage in the high SNR regime. Additionally, it is also not useful to built a distributed space-time code at the relays. We recall that in the distributed space-time coding setup the relays effectively act as transmit antennas which support the transmission and bring advantage from a diversity perspective. However, in that case a receiver with $N = MK$ receive antennas is required, which is not the case here.

A more precise explanation will be given in the further text where we derive an upper bound on the achievable two-way rate.

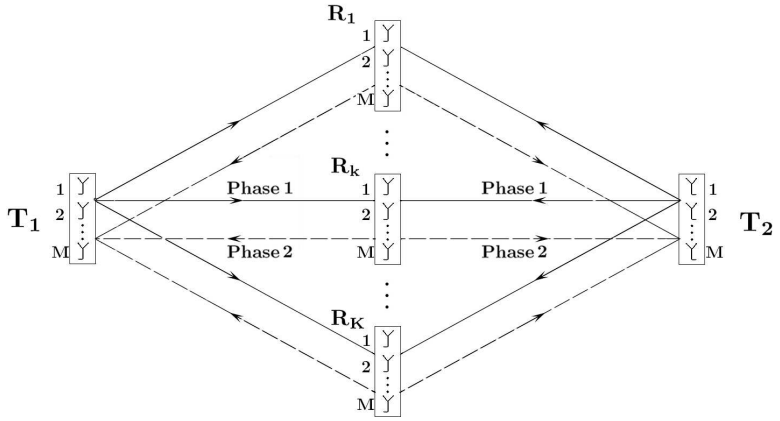

Figure 5.1: Two-way relay network

The terminals and the relays operate in a half-duplex mode, i.e. they do not transmit and receive simultaneously. We assume block Rayleigh model where the channel is constant in a certain time block. We denote by P_1 and P_2 the average transmit power for one transmission of Terminal 1 and Terminal 2, respectively. With this, the signal transmitted from Terminal 1 is $\sqrt{P_1}\mathbf{\Phi}$ and the signal transmitted from Terminal 2 is $\sqrt{P_2}\mathbf{\Psi}$. The matrix $\mathbf{\Phi}$ is a $T \times M$ matrix normalized such that $\mathrm{E}[\mathrm{tr}(\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi})] = T$. Accordingly, $\mathbf{\Psi}$ is a $T \times M$ matrix, normalized such that $\mathrm{E}[\mathrm{tr}(\mathbf{\Psi}^{\mathrm{H}}\mathbf{\Psi})] = T$. The codebooks of Terminal 1 and Terminal 2 are denoted as $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. Further, we denote the average power for one transmission for the relay k as γ_k and a power constraint for the total power of the K relays as $\sum_{k=1}^{K} \gamma_k = P_R$. Additionally, we have the constraint on the total network power, $P_1 + P_2 + P_R = P_{tot}$. The total power constraint serves for fair comparison in the case when the number of relays takes different values. It is important to note that under this assumption, the sum power of the relays, P_R, as well as the power of the terminals, P_1 and P_2, are always a fraction of the total power P_{tot}.

Further, the channel matrix between Terminal 1 and the k-th relay (multicast stage) is denoted as $\mathbf{H}_k$ and the channel between the relays and Terminal 1 in the broadcast stage as $\mathbf{H}_k^{(r)}$. The channel matrix between Terminal 2 and the k-th relay is denoted as $\mathbf{G}_k$ and the channel in the broadcast stage as $\mathbf{G}_k^{(r)}$. The channel matrices are assumed to have entries which are i.i.d $CN(0,1)$. We make a notice that the notation here slightly differs from the notation in the one-way relaying setup. In the one-way relaying setup we use distributed space-time coding and therefore we adapted the

notation to the specifics of that scheme. Additionally, in the distributed space-time coding setup we assumed that each relay has a single antenna. We showed that we can do this without loss of generality. In the two-way relaying setup we do not assume a distributed scheme and we also assume that the relays have multiple antennas in general. However, when possible we will try to keep the same notation.

We will assume an AF (Amplify-and-Forward) scenario. It has not been proved that the AF scheme is optimal in the non-coherent two-way relaying setup, however we conjecture that the AF scheme exploits the degrees of freedom offered by the network. We are going to show this later when we present the bounds on the achievable two-way rate. In the AF scenario the relay k receives

$$\mathbf{R}_k = \sqrt{P_1}\mathbf{\Phi}\mathbf{H}_k + \sqrt{P_2}\mathbf{\Psi}\mathbf{G}_k + \mathbf{V}_k, \tag{5.1}$$

where $\mathbf{V}_k$ is the noise contribution at the relay with entries which are i.i.d $CN(0,\sigma^2)$. In the second step (broadcast stage) the relay k sends

$$\sqrt{\gamma_k}\mathbf{T}_k = \sqrt{\frac{\gamma_k}{P_1+P_2+\sigma^2}}\mathbf{R}_k, \tag{5.2}$$

where a normalization is performed such that $\mathrm{E}[\mathrm{tr}(\mathbf{T}_k^{\mathrm{H}}\mathbf{T}_k)] = T$. The received signal of Terminal 2 is given as

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{H}' + \mathbf{\Psi}\mathbf{G}' + \mathbf{W}, \tag{5.3}$$

where

$$\begin{aligned}
\mathbf{H}' &= \sum_{k=1}^{K} \sqrt{\frac{P_1\gamma_k}{P_1+P_2+\sigma^2}}\mathbf{H}_k\mathbf{G}_k^{(r)} \\
\mathbf{G}' &= \sum_{k=1}^{K} \sqrt{\frac{P_2\gamma_k}{P_1+P_2+\sigma^2}}\mathbf{G}_k\mathbf{G}_k^{(r)} \\
\mathbf{W} &= \sum_{k=1}^{K} \sqrt{\frac{\gamma_k}{P_1+P_2+\sigma^2}}\mathbf{V}_k\mathbf{G}_k^{(r)} + \mathbf{Z},
\end{aligned} \tag{5.4}$$

and noise matrix $\mathbf{Z}$ has i.i.d. $CN(0,\sigma^2)$ entries. To be consistent, throughout this chapter we will use that the high SNR assumption is fulfilled when $\sigma^2 \to 0$.

For the received signal at Terminal 1 we have

$$\mathbf{U} = \mathbf{\Phi}\mathbf{F}' + \mathbf{\Psi}\mathbf{E}' + \mathbf{N}, \tag{5.5}$$

where $\mathbf{F}', \mathbf{E}'$ and $\mathbf{N}$ have the same form as in (5.4). In the following we will only address the decoding at Terminal 2. Due to symmetry, the same conclusions will hold for Terminal 1.

5.3 Bounds on the achievable Two-Way Rate

Here we will derive bounds on the achievable two-way rate in the relay network. The analysis is related to the capacity results for the point-to-point MIMO block Rayleigh fading channel without channel knowledge assumption at the receiver (neither at the transmitter), derived by Zheng and Tse in [65]. The non-coherent point-to-point MIMO channel is the building block in the derivation of the capacity bounds of the non-coherent two-way relaying channel. The high SNR capacity of this channel has been derived in [65] and presented in Chapter 3.

5.3.1 Upper bound on the achievable two-way rate

Here we derive the upper bound on the achievable rates of our non-coherent two-way channel. The upper bound can be obtained from the multiple access cut-set bound, which in the case of the two-way relaying network gives

$$R_{12} \leq \frac{1}{2} I(\mathbf{T}_1, \dots \mathbf{T}_K; \mathbf{Y}), \tag{5.6}$$

where $I(Y;X)$ denotes the mutual information between X and Y. Similarly

$$R_{21} \leq \frac{1}{2} I(\mathbf{T}_1, \dots \mathbf{T}_K; \mathbf{U}). \tag{5.7}$$

The factor $\frac{1}{2}$ is due to the half-duplex constraint, i.e. the fact that the terminals transmit one half of the time. The derivation of the upper bound is as follows. Let us recall that the received signal $\mathbf{Y}$ is given as

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{\gamma_k} \mathbf{T}_k \mathbf{G}_k^{(r)} + \mathbf{Z}, \tag{5.8}$$

which can be rewritten as

$$\mathbf{Y} = \left(\sqrt{\gamma_1}\mathbf{T}_1 \;\; \sqrt{\gamma_2}\mathbf{T}_2 \;\; \cdots \;\; \sqrt{\gamma_K}\mathbf{T}_K \right) \begin{pmatrix} \mathbf{G}_1^{(r)} \\ \mathbf{G}_2^{(r)} \\ \vdots \\ \mathbf{G}_K^{(r)} \end{pmatrix} + \mathbf{Z}, \tag{5.9}$$

or equivalently

$$\mathbf{Y} = \sqrt{P_R} \mathbf{T} \mathbf{G}^{(r)} + \mathbf{Z}, \tag{5.10}$$

where

$$\mathbf{T} = \sqrt{\frac{1}{P_R}} \left(\sqrt{\gamma_1}\mathbf{T}_1 \;\; \sqrt{\gamma_2}\mathbf{T}_2 \;\; \cdots \;\; \sqrt{\gamma_K}\mathbf{T}_K \right), \tag{5.11}$$

and

$$\mathbf{G}^{(r)} = \begin{pmatrix} \mathbf{G}_1^{(r)} \\ \mathbf{G}_2^{(r)} \\ \vdots \\ \mathbf{G}_K^{(r)} \end{pmatrix}. \tag{5.12}$$

The power constraint is $\mathrm{E}[\mathrm{tr}(\mathbf{T}^{\mathrm{H}}\mathbf{T})] = T$ and the elements of $\mathbf{G}^{(r)}$ are i.i.d $CN(0,1)$.

The form in which the equation is written is the same as the system model for the point-to-point MIMO block Rayleigh fading channel (2.1) with MK transmit and M receive antennas and coherence time (block length) T. The high SNR capacity of this channel has also been derived in [65]. Further, it has been shown that having more transmit then receive antennas does not increase capacity in the case without channel knowledge. Hence, as limit for the high SNR capacity of this channel we have the case with M transmit and M receive antennas, which is given as

$$C = M\left(1 - \frac{M}{T}\right)\log_2 \frac{P_R}{\sigma^2} + c + o(1), \tag{5.13}$$

where c is given by (2.4) and the term $\frac{P_R}{\sigma^2}$ plays the role of SNR. Hence, for the achievable rate R_{12} we have

$$R_{12} \leq \frac{M}{2}\left(1 - \frac{M}{T}\right)\log_2 \frac{P_R}{\sigma^2} + \frac{1}{2}c + o(1). \tag{5.14}$$

Equivalently, for the rate R_{21} we have

$$R_{21} \leq \frac{M}{2}\left(1 - \frac{M}{T}\right)\log_2 \frac{P_R}{\sigma^2} + \frac{1}{2}c + o(1). \tag{5.15}$$

5.3.2 Lower Bound on the achievable Two-Way Rate

The derivation of the lower bound is more complicated, at least in the general case. In the derivation we will assume that $K \to \infty$. This assumption, as we will see facilitates the derivation and is often used in the distributed space-time coding setup. Later we will argue on how critical this assumption is in the derivation of the rate bounds and the degrees of freedom.

We will take a gradual approach and discuss two different cases. In the first case we assume that $\mathbf{H}_k^{(r)} = \mathbf{H}_k^{\mathrm{H}}$ and $\mathbf{G}_k^{(r)} = \mathbf{G}_k^{\mathrm{H}}$, i.e. the channels in the broadcast and the multicast stage are reciprocal. When the channels are reciprocal we will show that the self-interference from the users can be ideally subtracted. The system model then is equivalent to the point-to-point MIMO channel (2.1). In the second case we assume independent channels in the multicast and the broadcast stage.

Reciprocal channels

We remember that according to (5.3) the received signal at terminal 2 can be written as

$$\mathbf{Y} = \mathbf{\Phi H}' + \mathbf{\Psi G}' + \mathbf{W}, \tag{5.16}$$

where $\mathbf{H}', \mathbf{G}'$ and $\mathbf{W}$ are given in (5.4).

We observe that when $K \to \infty$, the entries of $\mathbf{H}'$ and $\mathbf{W}$ are i.i.d Gaussian, $h'_{i,j} \sim CN\left(0, \mathcal{P}_1\right), n_{i,j} \sim CN\left(0, \nu^2\right)$, where $\mathcal{P}_1 = \frac{P_1 P_R}{P_1+P_2+\sigma^2}$ and $\nu^2 = \frac{P_R \sigma^2}{P_1+P_2+\sigma^2} + \sigma^2$. We recall that the matrix $\mathbf{G}'$ is given as

$$\mathbf{G}' = \sum_{k=1}^{K} \sqrt{\frac{P_2 \gamma_k}{P_1+P_2+\sigma^2}} \mathbf{G}_k \mathbf{G}_k^{\mathrm{H}}. \tag{5.17}$$

Hence, from the law of large numbers we have

$$\mathbf{G}' \to \mathcal{P}_2 \mathbf{I}, \tag{5.18}$$

where $\mathcal{P}_2 = \frac{P_2 P_R}{(P_1+P_2+\sigma^2)}$. The system model can be rewritten as

$$\mathbf{Y} = \sqrt{\mathcal{P}_1} \mathbf{\Phi H} + \sqrt{\mathcal{P}_2} \mathbf{\Psi I} + \mathbf{W}, \tag{5.19}$$

where for the elements of $\mathbf{H}$ and $\mathbf{W}$ we have $h_{i,j} \sim CN\left(0,1\right), g_{i,j} \sim CN\left(0,1\right)$ and $n_{i,j} \sim CN\left(0, \nu^2\right)$ In other words, when the channels are reciprocal (although unknown), the self-interference is constant in the limit and can be subtracted from the received signal. The signal after the subtraction of the self-interference is

$$\mathbf{Y}' = \sqrt{\mathcal{P}_1} \mathbf{\Phi H} + \mathbf{W}. \tag{5.20}$$

The high SNR capacity (b/s/Hz) of this channel is [65]

$$C = M\left(1 - \frac{M}{T}\right) \log_2 \frac{\mathcal{P}_1}{\nu^2} + c + o(1), \tag{5.21}$$

where c is given by (2.4).

If we want to represent the capacity as a function of $\frac{P_R}{\sigma^2}$, and thus directly compare to the upper bound, we should have on mind the normalization

$$\frac{\mathcal{P}_1}{\nu^2} = \frac{P_R}{\sigma^2} \frac{P_1}{P_1 + P_2 + P_R + \sigma^2}. \tag{5.22}$$

With this for the achievable rate when AF is used as relaying strategy, we have

$$R_{12}^{AF} = \frac{M}{2}\left(1 - \frac{M}{T}\right) \log_2 \frac{P_R}{\sigma^2} + \frac{1}{2} r_{12} + o(1), \tag{5.23}$$

where $r_{12} = c + \frac{M}{2}\left(1 - \frac{M}{T}\right)\log_2 \frac{P_1}{P_1+P_2+P_R+\sigma^2}$. Equivalently

$$R_{21}^{AF} = \frac{M}{2}\left(1 - \frac{M}{T}\right)\log_2 \frac{P_R}{\sigma^2} + \frac{1}{2}r_{21} + o(1), \tag{5.24}$$

where $r_{21} = c + \frac{M}{2}\left(1 - \frac{M}{T}\right)\log_2 \frac{P_2}{P_1+P_2+P_R+\sigma^2}$. The terms $\frac{P_1}{P_1+P_2+P_R+\sigma^2}$ and $\frac{P_2}{P_1+P_2+P_R+\sigma^2}$ are constant under the high SNR assumption $\sigma^2 \to 0$, and depend only on the power distribution between the terminals and the relays. In general we have $P_1 = \alpha P_{tot}$, $P_2 = \beta P_{tot}$ and $P_R = (1-\alpha-\beta)P_{tot}$, where $P_{tot} = P_1 + P_2 + P_R$ is the total power in the network. It is easy to show that under the total power constraint, the sum rate is maximized when $P_1 = P_2 = \frac{P_R}{2} = \frac{P}{4}$

Non-reciprocal channels

When the channels are non-reciprocal, the self interference can not be subtracted and has to be taken into account in the derivation. We remember that the signal received at Terminal 2 is

$$\mathbf{Y} = \sqrt{\mathcal{P}_1}\mathbf{\Phi}\mathbf{H} + \sqrt{\mathcal{P}_2}\mathbf{\Psi}\mathbf{G} + \mathbf{W}, \tag{5.25}$$

where the distributions of the elements of $\mathbf{H}$, $\mathbf{G}$ and $\mathbf{W}$ are i.i.d Gaussian with $h_{i,j} \sim CN(0,1)$, $g_{i,j} \sim CN(0,1)$ and $n_{i,j} \sim CN\left(0,\nu^2\right)$ Since terminal 2 knows $\mathbf{\Psi}$, the mutual information is

$$I(\mathbf{Y};\mathbf{\Phi}|\mathbf{\Psi}) = h(\mathbf{Y}|\mathbf{\Psi}) - h(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi}). \tag{5.26}$$

We observe that given $\mathbf{\Phi}$ and $\mathbf{\Psi}$, the column vectors $\mathbf{y}_j$, $j = 1,\ldots,M$ of $\mathbf{Y}$ are independent Gaussian vectors with identical covariance matrix

$$\mathbf{R} = \mathcal{P}_1\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}} + \mathcal{P}_2\mathbf{\Psi}\mathbf{\Psi}^{\mathrm{H}} + \nu^2\mathbf{I}. \tag{5.27}$$

This follows directly from the fact that

$$\mathbf{y}_j = \sum_{i=1}^{M}\sqrt{\mathcal{P}_1}\mathbf{x}_i h_{ij} + \sum_{i=1}^{M}\sqrt{\mathcal{P}_2}\mathbf{u}_i g_{ij} + \mathbf{n}_j. \tag{5.28}$$

The covariance matrix $\mathbf{R}$ can be rewritten as

$$\begin{aligned}\mathbf{R} &= \mathcal{P}_1\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}} + \mathcal{P}_2\mathbf{\Psi}\mathbf{\Psi}^{\mathrm{H}} + \nu^2\mathbf{I}_T \\ &= \mathbf{\Theta}\mathbf{\Lambda}\mathbf{\Theta}^{\mathrm{H}} + \nu^2\mathbf{I}_T,\end{aligned} \tag{5.29}$$

where $\mathbf{\Theta}\mathbf{\Lambda}\mathbf{\Theta}^{\mathrm{H}}$ is a decomposition of $\mathcal{P}_1\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}} + \mathcal{P}_2\mathbf{\Psi}\mathbf{\Psi}^{\mathrm{H}}$ which includes the non-zero eigenvalues, whose number is at most M. Here, we will assume that there are exactly

M eigenvalues which are not zero. This, of course depends on the choice of the codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$, but we will consider that the codebooks are chosen such that this is fulfilled. Thus, $\mathbf{\Theta} \in \mathbb{C}^{T\times M}$. The conditional entropy $h(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi})$ is given by

$$\begin{aligned} h(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi}) &= M\mathrm{E}\left[\log_2(\pi e)^T \det\mathbf{R}\right] \\ &= MT\log_2 \pi e + M\mathrm{E}\left[\log_2 \det(\mathbf{\Theta\Lambda\Theta}^{\mathrm{H}} + \nu^2\mathbf{I}_T)\right]. \end{aligned} \tag{5.30}$$

From $\det(\mathbf{I}_T + \mathbf{AB}) = \det(\mathbf{I}_M + \mathbf{BA})$, where $\mathbf{A} \in \mathbb{C}^{T\times M}$, $\mathbf{B} \in \mathbb{C}^{M\times T}$, we have

$$\begin{aligned} \det(\mathbf{\Theta\Lambda\Theta}^{\mathrm{H}} + \nu^2\mathbf{I}_T) &= \nu^{2T}\det\left(\frac{1}{\nu^2}\mathbf{\Theta}^{\mathrm{H}}\mathbf{\Lambda\Theta} + \mathbf{I}_M\right) \\ &= \nu^{2(T-M)}\prod_{i=1}^{M}\left(\lambda_i + \nu^2\right). \end{aligned} \tag{5.31}$$

After some rewriting, for the differential entropy we have

$$\begin{aligned} h(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi}) = M\mathrm{E}\left[\log_2\prod_{i=1}^{M}(\lambda_i+\nu^2)\right] + M^2\log_2\pi e \\ + M(T-M)\log_2\pi e\nu^2, \end{aligned} \tag{5.32}$$

We recall that we have the following power constraint

$$\mathrm{E}\left[\mathrm{tr}\left(\mathbf{\Lambda}\right)\right] \leq (\mathcal{P}_1 + \mathcal{P}_2)T.$$

Hence, it holds

$$\mathrm{E}\left[\prod_{i=1}^{M}\left(\lambda_i + \nu^2\right)\right] \leq \left(\frac{(\mathcal{P}_1+\mathcal{P}_2)T + \nu^2}{M}\right)^M, \tag{5.33}$$

where the equality is achieved when all eigenvalues are equal. Thus, for the differential entropy we have

$$\begin{aligned} h(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi}) \leq M^2\log_2\left[(\mathcal{P}_1+\mathcal{P}_2)T + \nu^2\right] + M^2\log_2\frac{\pi e}{M} \\ + M(T-M)\log_2\pi e\nu^2, \end{aligned} \tag{5.34}$$

The calculation of $h(\mathbf{Y}|\mathbf{\Psi})$ is more complicated. A bound can be obtained by conditioning on $\mathbf{G}$, since conditioning does not increase the entropy [8]

$$h(\mathbf{Y}|\mathbf{\Psi}) \geq h(\mathbf{Y}|\mathbf{\Psi},\mathbf{G}) = h(\mathbf{Y}'), \tag{5.35}$$

where

$$\mathbf{Y}' = \sqrt{\mathcal{P}_1}\mathbf{\Phi H} + \mathbf{W}. \tag{5.36}$$

Since we are interested in the high SNR region, we approximate $h(\mathbf{Y}') \approx h(\sqrt{\mathcal{P}_1}\mathbf{\Phi H})$. The approximation is valid and was also performed in [65] in the high SNR capacity derivation. We introduce $\mathbf{\Phi}' = \sqrt{\mathcal{P}_1}\mathbf{\Phi}$ for the further derivation. We follow the approach in [65] and introduce the coordinate change

$$\mathbf{\Phi' H} \rightarrow (\mathbf{C}_{\mathbf{\Phi' H}}, \mathbf{\Omega}_{\mathbf{\Phi' H}}), \tag{5.37}$$

where $\mathbf{C}_{\mathbf{\Phi' H}} \in \mathbb{C}^{(M \times M)}$ and $\mathbf{\Omega}_{\mathbf{\Phi' H}}$ is the subspace that the columns of $\mathbf{\Phi' H}$ span. Note that $\mathbf{\Phi' H}$, $\mathbf{\Phi H}$ and $\mathbf{\Phi}$ span the same subspace $\mathbf{\Omega}_{\Phi}$. We observe that $\mathbf{\Phi' H}$ is a random matrix which is isotropically distributed (i.d.) in the subspace $\mathbf{\Omega}_{\mathbf{\Phi}}$. From [65] we know that for the i.d. random matrix $\mathbf{\Phi' H} \in \mathbb{C}^{T \times M}$ the differential entropy can be calculated as

$$\begin{aligned} h(\mathbf{\Phi' H}) = {} & h(\mathbf{C}_{\mathbf{\Phi' H}}) + \log_2 |G(T, M)| \\ & + (T - M)\mathrm{E}\left[\log_2 \det(\mathbf{\Phi' H}^{\mathrm{H}} \mathbf{\Phi' H})\right]. \end{aligned} \tag{5.38}$$

Additionally, the distribution of $\mathbf{C}_{\Phi' H}$ is the same as the distribution of $\mathbf{QAH}$ where $\mathbf{Q}$ is a $M \times M$ unitary matrix and $\mathbf{A} = \mathrm{diag}\|\mathbf{x}_i'\|$, $i = 1, 2, \ldots, M$ is a $M \times M$ diagonal matrix. With this, for the differential entropy $h(\mathbf{\Phi' H})$ we have

$$\begin{aligned} h(\mathbf{\Phi' H}) = {} & h(\mathbf{C}_{\Phi' H}) + \log_2 |G^{\mathbb{C}}_{T,M}| \\ & + (T - M)\mathrm{E}\left[\log_2 \det(\mathbf{H}^{\mathrm{H}} \mathbf{\Phi'}^{\mathrm{H}} \mathbf{\Phi' H})\right] \\ = {} & h(\mathbf{QAH}) + \log_2 |G^{\mathbb{C}}_{T,M}| \\ & + (T - M)\mathrm{E}\left[\log_2 \det(\mathbf{A}^2)\right] \\ & + (T - M)\mathrm{E}\left[\log_2 \det(\mathbf{H}^{\mathrm{H}} \mathbf{H})\right]. \end{aligned} \tag{5.39}$$

We observe that

$$h(\mathbf{QAH}) \leq M^2 \log_2 \mathcal{P}_1 \pi e T, \tag{5.40}$$

with equality achieved for Gaussian entries. Similarly,

$$\begin{aligned} \mathrm{E}\left[\log_2 det(\mathbf{A^2})\right] &\leq \log_2 \left(\frac{\mathcal{P}_1 T}{M}\right)^M \\ &= M \log_2 \mathcal{P}_1 + M \log_2 \frac{T}{M}, \end{aligned} \tag{5.41}$$

with equality achieved when the diagonal entries of $\mathbf{A}$ are all equal. Finally, for the differential entropy $h(\mathbf{\Phi}'\mathbf{H})$ we obtain

$$\begin{aligned} h(\mathbf{\Phi}'\mathbf{H}) \leq & \log_2 |G_{T,M}^{\mathbb{C}}| + M^2 \log_2 \pi e T \\ & + M(T-M) \log_2 \mathcal{P}_1 + M(T-M) \log_2 \frac{T}{M} \\ & + (T-M) \mathrm{E}\left[\log_2 \det(\mathbf{H}\mathbf{H}^{\mathrm{H}})\right]. \end{aligned} \tag{5.42}$$

Hence, for the mutual information we have

$$\begin{aligned} I(\mathbf{Y}, \mathbf{\Phi}|\mathbf{\Psi}) \approx & h(\mathbf{\Phi}'\mathbf{H}) - h(\mathbf{Y}|\mathbf{\Phi}, \mathbf{\Psi}) \\ & \leq \log_2 |G_{T,M}^{\mathbb{C}}| + M(T-M) \log_2 \frac{\mathcal{P}_1}{\nu^2} \\ & + M(T-M) \log_2 \frac{T}{M\pi e} \\ & + (T-M) \mathrm{E}\left[\log_2 \det(\mathbf{H}\mathbf{H}^{\mathrm{H}})\right] \\ & + M^2 \log_2 \mathcal{P}_1 - M^2 \log_2(\mathcal{P}_1 + \mathcal{P}_2). \end{aligned} \tag{5.43}$$

With normalization over the blocklength T and having on mind the half-duplex constraints, for the achievable rate R_{12}^{AF} (in bits/channel use) we have

$$\begin{aligned} R_{12}^{AF} = & \frac{M}{2}\left(1 - \frac{M}{T}\right) \log_2 \frac{\mathcal{P}_1}{\nu^2} + \frac{1}{T} \log_2 |G_{T,M}^{\mathbb{C}}| \\ & + \frac{M}{2}\left(1 - \frac{M}{T}\right) \log_2 \frac{T}{M\pi e} \\ & + \frac{1}{2}\left(1 - \frac{M}{T}\right) \mathrm{E}\left[\log_2 \det(\mathbf{H}\mathbf{H}^{\mathrm{H}})\right] \\ & + \frac{M^2}{2} \log_2 \mathcal{P}_1 - M^2 \log_2(\mathcal{P}_1 + \mathcal{P}_2). \end{aligned} \tag{5.44}$$

We can rewrite the term as

$$R_{12}^{AF} = \frac{M}{2}\left(1 - \frac{M}{T}\right) \log_2 \left(\frac{P_R}{\sigma^2}\right) + \frac{1}{2} r_{12} + o(1), \tag{5.45}$$

where

$$\begin{aligned} r_{12} = & \frac{M}{2}\left(1-\frac{M}{T}\right)\log_2\left(\frac{P1}{P_1+P_2+P_R+\sigma^2}\right) \\ & + \frac{1}{T}\log_2|G^{\mathbb{C}}_{T,M}| + M\left(1-\frac{M}{T}\right)\log_2\frac{T}{M\pi e} \\ & + \left(1-\frac{M}{T}\right)\mathrm{E}\left[\log_2\det(\mathbf{H}\mathbf{H}^{\mathrm{H}})\right] \\ & + \frac{M^2}{T}\log_2\left(\frac{\mathcal{P}_1}{\mathcal{P}_1+\mathcal{P}_2}\right). \end{aligned} \tag{5.46}$$

Similarly,

$$R_{21}^{AF} = \frac{M}{2}\left(1-\frac{M}{T}\right)\log_2\left(\frac{P_R}{\sigma^2}\right) + \frac{1}{2}r_{21} + o(1). \tag{5.47}$$

Again, it is easy to show that under the total power constraint $P_1 + P_2 + P_R = P_{tot}$, the sum rate is maximized when $P_1 = P_2 = \frac{P_R}{2} = \frac{P_{tot}}{4}$.

We see that at high SNR the lower and the upper bound in both cases (reciprocal and non-reciprocal channels) differ only up to a constant. Further, the lower bound was computed under the AF assumption. However, no AF assumption was made for the calculation of the upper bound, which was only done based on the multiple access cut-set bound. Hence, we can make two conclusions. First, the non-coherent two-way relay channel achieves $M(1-\frac{M}{T})$ degrees of freedom in total (two-way), although there is no channel knowledge at neither the relays nor at the transceivers. Second, these degrees of freedom are achieved by an AF scheme.

5.3.3 Finite K

In this part we will discuss the role of the assumption $K \to \infty$ in the derivation of the bounds of the achievable rate. The assumption was used twice in the derivation. First, it was critical in the Gaussian assumption for the equivalent noise term

$$\mathbf{W} = \sum_{k=1}^{K} \sqrt{\tfrac{\gamma_k}{P_1+P_2+\sigma^2}}\mathbf{V}_k\mathbf{G}_k^{\mathrm{H}} + \mathbf{Z}, \tag{5.48}$$

For finite K, this assumption is not true. Hence, the elements of the received matrix $\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi}$ are not Gaussian. Nevertheless, the conditional entropy $h(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi})$ is upper bounded by the entropy of a matrix with Gaussian entries (the Gaussian distribution has the maximal differential entropy among the distributions with same first and second moment). For the mutual information it means that

$$\begin{aligned} I(\mathbf{Y},\mathbf{\Phi}|\mathbf{\Psi}) &= h(\mathbf{Y}|\mathbf{\Psi}) - h(\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi}) \\ &\geq h(\mathbf{Y}|\mathbf{\Psi}) - h(\mathbf{Y}_G|\mathbf{\Phi},\mathbf{\Psi}), \end{aligned} \tag{5.49}$$

where $h(\mathbf{Y}_G|\mathbf{\Phi},\mathbf{\Psi})$ is the entropy with Gaussian assumption for the effective noise at the receiver. What remains is to address the influence of the finite K assumption on the first term in the mutual information, $h(\mathbf{Y}|\mathbf{\Psi})$. We remember that we obtained a lower bound on this term by taking the conditional entropy

$$h(\mathbf{Y}|\mathbf{\Psi}) \geq h(\mathbf{Y}|\mathbf{\Psi},\mathbf{G}) = h(\mathbf{Y}'), \tag{5.50}$$

where

$$\mathbf{Y}' = \mathbf{\Phi}\mathbf{H}' + \mathbf{W}. \tag{5.51}$$

In the calculation of $h(\mathbf{Y})' \approx h(\mathbf{\Phi}\mathbf{H}')$ we used the Gaussian assumption for the equivalent channel

$$\mathbf{H}' = \sum_{k=1}^{K} \sqrt{\frac{P_1\gamma_k}{P_1+P_2+\sigma^2}}\mathbf{H}_k\mathbf{G}_k^{(r)}.$$

Under this assumption, $\mathbf{\Phi}\mathbf{H}'$ is isotropically distributed (i.d.) in the subspace $\mathbf{\Omega}_\Phi$ and this was used in the exact calculation of the achievable rate. Without the Gaussian assumption, the derivation becomes intractable since it includes the distribution of product of random matrices with Gaussian entries. In a simplified form we can see this system as a system where the channel induces a certain correlation. Under these circumstances the input distribution which maximizes the mutual information does not have the form 5.41. Nevertheless, the achievable rate in this case will still differ from the one with the Gaussian assumption within a constant. This is due to the fact that the degrees of freedom available are still $M(1-\frac{M}{T})$. The reason for this is that the information- carrying object is still a random subspace $\mathbf{\Omega}_\Phi$, which is a random point in the Grassmann manifold. Thus, the number of degrees of freedom is the dimension of the set of all column spaces of $T \times M$ matrices, i.e. the dimension of the Grassmann manifold $G_{T,M}^{\mathbb{C}}$ which is exactly $M(1-\frac{M}{T})$.

5.4 Codes for Genuine Non-Coherent Two-Way Relaying

A hint of how we can transmit information in the non-coherent setup comes from the code constructions for the point-to-point MIMO non-coherent block fading channels.

We recall that the signal received at Terminal 2 can be written as

$$\mathbf{Y} = \mathbf{\Phi}'\mathbf{H} + \mathbf{\Psi}'\mathbf{G} + \mathbf{W}. \tag{5.52}$$

Before discussing the transmission strategy, let us see what the system model tells us about the possible decoding of the signal $\mathbf{\Phi}'$ (or equivalently) $\mathbf{\Phi}$ at Terminal 2. We can see that apart from the desired signal, we have the term $\mathbf{\Psi}'\mathbf{G}$ which is basically the self-interference. However, due to the unknown channel matrix $\mathbf{G}$ we can not

subtract this term from the received signal. The problem looks somewhat similar to the known two-user interference channel. Nevertheless, there is a subtle difference: Although we do not know the channel matrix $\mathbf{G}$, we can use the knowledge of $\mathbf{\Psi}'$ in the decoding process.

5.4.1 Code Construction

We propose a code construction which is appropriate in the two-way relaying setup. Note that the system model (5.5) can be written as

$$\mathbf{Y} = \mathcal{P}\left(\begin{array}{cc} \mathbf{\Phi}' & \mathbf{\Psi}' \end{array} \right) \left(\begin{array}{c} \mathbf{H} \\ \mathbf{G} \end{array} \right) + \mathbf{W}. \tag{5.53}$$

Now, we can think of $\left(\begin{array}{cc} \mathbf{\Phi}' & \mathbf{\Psi}' \end{array} \right)$ as an equivalent tranmit matrix and of $\left(\begin{array}{c} \mathbf{H} \\ \mathbf{G} \end{array} \right)$ as an equivalent channel. This is similar to the system model for the non-coherent MIMO point-to-point block fading channel with $2M$ transmit, M receive antennas and coherence time T. We should note that $T \geq 4M$ is required in this context[65]. Let us denote by $\Omega_{\mathbf{Q}}$ the $2M$-dimensional subspace spanned by the columns of $\left(\begin{array}{cc} \mathbf{\Phi}' & \mathbf{\Psi}' \end{array} \right)$, where the columns of $\mathbf{Q}$ represent an orthonormal basis for $\Omega_{\mathbf{Q}}$, i.e. $\mathbf{Q}^{\mathrm{H}}\mathbf{Q} = \mathbf{I}_{2M}$. We denote by $\mathcal{Q}$ the codebook obtained by the above concatenation. The cardinality of the codebook is $|\mathcal{Q}| = |\mathcal{C}_1||\mathcal{C}_2|$. A $2M$-dimensional subspace $\Omega_{\mathbf{Q}}$ collapses into a M-dimensional subspace after the channel action. However, we can still perform the decoding by looking for the most likely transmitted subspace, having the received matrix $\mathbf{Y}$, as in [31]

$$\mathbf{Q} = \arg \max_{\mathbf{Q}_i \in \mathcal{Q}} \|\mathbf{Y}^{\mathrm{H}}\mathbf{Q}_i\|_F^2. \tag{5.54}$$

Having $\Omega_{\mathbf{Q}}$, we get the pair $(\mathbf{\Phi}, \mathbf{\Psi})$. When looking for the most likely subspace $\Omega_{\mathbf{Q}}$ we can use the fact that we know $\mathbf{\Psi}$ in advance, which limits the number of the subspaces we have to search. If we compare (5.54) and (5.70) we can see that this is not the ML decoding. However, we will show with the simulations that the difference is negligible, when we compare the two decoding methods.

We can see that the performance of the code depends on the properties of the codebook $\mathcal{Q}$, in the context of the well established criteria for construction of codes for the non-coherent channel [31], such as chordal distance and diversity product. However, the subtle difference (which makes the problem more difficult) is the fact that the codewords of the codebook $\mathcal{Q}$ are obtained by concatenation of the transmit matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$. The question is up to which degree the properties of a code obtained in this way can be controlled.

In the following we propose a construction which in a way takes care of the specifics of the coding problem. The codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$ are not chosen independently, but

rather as two "well separated" subsets of a larger codebook. This construction originates in [31] and [55] and represents an example of a code design for the non-coherent channel with a certain structure. The codewords are given as

$$\mathbf{\Phi} = \begin{pmatrix} \cos(\alpha_1)\mathbf{A}_\Phi \\ \sin(\alpha_1)\mathbf{B}_\Phi \end{pmatrix} \tag{5.55}$$

$$\mathbf{\Psi} = \begin{pmatrix} \cos(\alpha_2)\mathbf{A}_\Psi \\ \sin(\alpha_2)\mathbf{B}_\Psi \end{pmatrix}, \tag{5.56}$$

where $\mathbf{A}_\Phi = \mathbf{I}_{\frac{T}{2},M}$, $\mathbf{A}_\Psi = \mathbf{I}^{\perp}_{\frac{T}{2},M}$and $\mathbf{B}_\Phi$, $\mathbf{B}_\Psi$ are codewords which belong to space-time codes constructed for the known channel. In the following we give worked-out examples of codes which are appropriate for the cases with one, respectively two antennas

One Antenna case (M = 1)

The signal transmitted from Terminal 1 is a $T \times 1$ vector ϕ. Similarly, Terminal 2 transmits a $T \times 1$ vector ψ. For the case $T = 4$, codes for both terminals are

$$\phi = \begin{pmatrix} \cos(\alpha_1) \\ 0 \\ \sin(\alpha_1)\phi_1 \\ \sin(\alpha_1)\phi_2 \end{pmatrix}, \ \psi = \begin{pmatrix} 0 \\ \cos(\alpha_2) \\ \sin(\alpha_2)\psi_1 \\ \sin(\alpha_2)\psi_2 \end{pmatrix}, \tag{5.57}$$

with symbols from Terminal 1 $\phi_1, \phi_2 \in$ QPSK, symbols from Terminal 2 $\psi_1, \psi_2 \in$ QPSK. This construction is motivated from the non-coherent codes constructed in **??**.

Two Antennas case (M = 2)

The signal transmitted from Terminal 1 is a $T \times 2$ matrix $\mathbf{\Phi}$. Similarly, Terminal 2 transmits a $T \times 2$ matrix $\mathbf{\Psi}$. For case that $T = 8$, as in (5.55) and (5.56), $\mathbf{A}_\Phi = \mathbf{I}_{4,2}$, $\mathbf{A}_\Psi = \mathbf{I}^{\perp}_{4,2}$, $\mathbf{B}_\Phi$, $\mathbf{B}_\Psi$ are

$$\mathbf{B}_\Phi = \begin{pmatrix} \phi_1 & \phi_2 \\ -\phi_2^* & \phi_1^* \\ -\phi_3^* & -\phi_4^* \\ \phi_4 & -\phi_3 \end{pmatrix} \qquad \mathbf{B}_\Psi = \begin{pmatrix} \psi_3 & \psi_4 \\ -\psi_4^* & \psi_3^* \\ \psi_1^* & \psi_2^* \\ -\psi_2 & \psi_1 \end{pmatrix}$$

with symbols from Terminal 1 $\phi_1, \phi_2, \phi_3, \phi_4 \in$ QPSK, symbols from Terminal 2 $\psi_1, \psi_2, \psi_3, \psi_4 \in$ QPSK.

The acquisition of the construction of $\mathbf{B}_\Phi$ and $\mathbf{B}_\Psi$ comes from the quasi-orthogonal space-time block code introduced in [21]:

$$\mathbf{C} = \begin{pmatrix} c_1 & c_2 & c_3 & c_4 \\ -c_2^* & c_1^* & -c_4^* & c_3^* \\ -c_3^* & -c_4^* & c_1^* & c_2^* \\ c_4 & -c_3 & -c_2 & c_1 \end{pmatrix}.$$

With different combinations of columns of $\mathbf{C}$, variants of $\mathbf{B}_\Phi$ and $\mathbf{B}_\Psi$ can be obtained. Let's denote the above-mentioned $\mathbf{B}_\Phi$ and $\mathbf{B}_\Psi$ as $\mathbf{B}_{\Phi(1,2)}$ and $\mathbf{B}_{\Psi(3,4)}$. There are also

$$\mathbf{B}_{\Phi(1,4)} = \begin{pmatrix} \phi_1 & \phi_4 \\ -\phi_2^* & \phi_3^* \\ -\phi_3^* & \phi_2^* \\ \phi_4 & \phi_1 \end{pmatrix} \qquad \mathbf{B}_{\Psi(2,3)} = \begin{pmatrix} \psi_2 & \psi_3 \\ \psi_1^* & -\psi_4^* \\ -\psi_4^* & \psi_1^* \\ -\psi_3 & -\psi_2 \end{pmatrix}$$

$$\mathbf{B}_{\Phi(1,3)} = \begin{pmatrix} \phi_1 & \phi_3 \\ -\phi_2^* & -\phi_4^* \\ -\phi_3^* & \phi_1^* \\ \phi_4 & -\phi_2 \end{pmatrix} \qquad \mathbf{B}_{\Psi(2,4)} = \begin{pmatrix} \psi_2 & \psi_4 \\ \psi_1^* & \psi_3^* \\ -\psi_4^* & \psi_2^* \\ -\psi_3 & \psi_1 \end{pmatrix}$$

5.4.2 Derivation of the ML-Decoding Rule

Here we derive the ML-decoding rule in the case of non-coherent two-way relaying. The derivation is performed under the assumption $K \to \infty$.

We observe that given $\mathbf{\Phi}'$ and $\mathbf{\Psi}'$, the column vectors $\mathbf{y}_j$, $j = 1, \ldots, M$ of $\mathbf{Y}$ are independent Gaussian vectors with identical covariance matrix

$$\mathbf{R}_{\mathbf{\Phi},\mathbf{\Psi}} = \mathbf{\Phi}'\mathbf{\Phi}'^{\mathrm{H}} + \mathbf{\Psi}'\mathbf{\Psi}'^{\mathrm{H}} + \nu^2\mathbf{I}_T. \tag{5.58}$$

This follows directly from the fact that

$$\mathbf{y}_j = \sum_{i=1}^{M} \sqrt{\mathcal{P}_1}\phi_i h_{ij} + \sum_{i=1}^{M} \sqrt{\mathcal{P}_2}\psi_i g_{ij} + \mathbf{n}_j. \tag{5.59}$$

The pdf of each column vector is thus given as

$$p_{\mathbf{y}_j|\mathbf{\Phi},\mathbf{\Psi}}(\mathbf{s}) = \frac{1}{(2\pi)^{T/2}\sqrt{\det \mathbf{R}_{\Phi,\Psi}}} \exp\Big(-\frac{1}{2}\mathbf{s}^{\mathrm{H}}\mathbf{R}_{\mathbf{\Phi},\mathbf{\Psi}}^{-1}\mathbf{s}\Big), \tag{5.60}$$

where $j = 1, \dots M$. Regarding the ML-decoding, the joint pdf of the received matrix $\mathbf{Y}$ is

$$p_{\mathbf{Y}|\mathbf{\Phi},\mathbf{\Psi}}(\mathbf{S}) = \prod_{j=1}^{M} p_{\mathbf{y}_j|\mathbf{\Phi},\mathbf{\Psi}}(\mathbf{s}). \tag{5.61}$$

We see that the pdf involves the inverse of the covariance matrix $\mathbf{R}_{\mathbf{\Phi},\mathbf{\Psi}}^{-1}$. We will assume that the terminals use same power, $P_1 = P_2 = P$. Additionally, we will assume that the terminals and the relays share the power, i.e. $P = P_R/2$. It can be shown that this is optimal in the non-coherent setup [59]. We introduce $\mathcal{P} = \mathcal{P}_1 = \mathcal{P}_2$. Thus the covariance matrix from (5.58) can be written as

$$\mathbf{R}_{\mathbf{\Phi},\mathbf{\Psi}} = \mathcal{P}\left(\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}} + \mathbf{\Psi}\mathbf{\Psi}^{\mathrm{H}}\right) + \nu^2\mathbf{I}_T. \tag{5.62}$$

We perform the following decomposition

$$\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}} + \mathbf{\Psi}\mathbf{\Psi}^{\mathrm{H}} = \mathbf{\Theta}\mathbf{\Lambda}\mathbf{\Theta}^{\mathrm{H}}, \tag{5.63}$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the non-zero eigenvalues of $\mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}} + \mathbf{\Psi}\mathbf{\Psi}^{\mathrm{H}}$, denoted as λ_j and the columns of $\mathbf{\Theta}$ are the corresponding eigenvectors. The number of non-zero eigenvalues depends on the codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$ and is at most T. Thus, the covariance matrix $\mathbf{R}_{\mathbf{\Phi},\mathbf{\Psi}}$ can now be written as

$$\mathbf{R}_{\mathbf{\Phi},\mathbf{\Psi}} = \nu^2\mathbf{I}_T + \mathcal{P}\mathbf{\Theta}\mathbf{\Lambda}\mathbf{\Theta}^{\mathrm{H}}. \tag{5.64}$$

According to the matrix inversion lemma

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}. \tag{5.65}$$

We denote $\mathbf{A} = \nu^2\mathbf{I}_T$, $\mathbf{B} = \mathbf{\Theta}$, $\mathbf{D} = \mathbf{\Theta}^{\mathrm{H}}$ and $\mathbf{C} = \mathcal{P}\mathbf{\Lambda}$. Thus

$$\begin{aligned}
\mathbf{R}_{\mathbf{\Phi},\mathbf{\Psi}}^{-1} &= \frac{\mathbf{I}_T}{\nu^2} - \frac{\mathbf{I}_T}{\nu^2}\mathbf{\Theta}\Big(\frac{\mathbf{\Lambda}^{-1}}{\mathcal{P}} + \mathbf{\Theta}^{\mathrm{H}}\frac{\mathbf{I}_T}{\nu^2}\mathbf{\Theta}\Big)^{-1}\mathbf{\Theta}^{\mathrm{H}}\frac{\mathbf{I}_T}{\nu^2} \\
&= \frac{\mathbf{I}_T}{\nu^2} - \frac{1}{\nu^2}\mathbf{\Theta}\Big(\frac{\nu^2\mathbf{\Lambda}^{-1}}{\mathcal{P}} + \mathbf{\Theta}^{\mathrm{H}}\mathbf{\Theta}\Big)^{-1}\mathbf{\Theta}^{\mathrm{H}}
\end{aligned} \tag{5.66}$$

Note that $\mathbf{\Theta}^{\mathrm{H}}\mathbf{\Theta} = \mathbf{I}$. Then $\mathbf{R}_{\mathbf{\Phi},\mathbf{\Psi}}^{-1}$ can be finally written as

$$\mathbf{R}_{\mathbf{\Phi},\mathbf{\Psi}}^{-1} = \frac{\mathbf{I}_T}{\nu^2} - \mathbf{\Theta}\mathbf{D}\mathbf{\Theta}^{\mathrm{H}}, \tag{5.67}$$

where $\mathbf{D}$ is a diagonal matrix with elements $d_{jj} = \frac{\mathcal{P}\lambda_j}{\nu^2(\nu^2+\mathcal{P}\lambda_j)}$ on its diagonal.

Note that $\mathbf{\Psi}$ is constant in the decoding, since this is the self-interference. In order to simplify the expressions, in the following text we will use the notation $\mathbf{R}_k$ to denote

the covariance matrix associated with the codeword $\mathbf{\Phi}_k \in \mathcal{C}_1$. Hence, for the ML decoding rule we have

$$\hat{\mathbf{\Phi}} = \arg\max p_{\mathbf{Y}|\mathbf{\Phi}_k,\mathbf{\Psi}}(\mathbf{S}). \tag{5.68}$$

By inserting (5.60) into the decoding rule, we get

$$\begin{aligned}
\hat{\mathbf{\Phi}} &= \arg\max_{\mathbf{\Phi}_k\in\chi}\left\{\prod_{j=1}^{M} p_{\mathbf{y}_j|\mathbf{\Phi}_k,\mathbf{\Psi}}\right\} \\
&= \arg\max_{\mathbf{\Phi}_k\in\chi}\left\{\left(\frac{1}{\sqrt{(2\pi)^T\det(\mathbf{R}_k)}}\right)^M \prod_{j=1}^{M}\exp\left(-\frac{\mathbf{y}_j^{\mathrm{H}}\mathbf{R}_k^{-1}\mathbf{y}_j}{2}\right)\right\} \\
&= \arg\min_{\mathbf{\Phi}_k\in\chi}\left\{M\ln\left(\det\left(\mathbf{R}_k\right)\right)+\sum_{j=1}^{M}\mathbf{y}_j^{\mathrm{H}}\mathbf{R}_k^{-1}\mathbf{y}_j\right\} \\
&= \arg\min_{\mathbf{\Phi}_k\in\chi}\left\{M\ln\left(\det\left(\mathbf{R}_k\right)\right)+\mathrm{tr}\left(\mathbf{Y}^{\mathrm{H}}\mathbf{R}_k^{-1}\mathbf{Y}\right)\right\}.
\end{aligned} \tag{5.69}$$

Now we insert $\mathbf{R}_k^{-1}$ into the decoding rule in (5.69), and we get

$$\begin{aligned}
\hat{\mathbf{\Phi}} &= \arg\min_{\mathbf{\Phi}_k\in\chi}\left\{M\ln\left(\det\left(\mathbf{R}_k\right)\right)+\mathrm{tr}\left(\mathbf{Y}^{\mathrm{H}}\left(\frac{\mathbf{I}_T}{\nu^2}-\mathbf{\Theta}_k\mathbf{D}_k\mathbf{\Theta}_k^{\mathrm{H}}\right)\mathbf{Y}\right)\right\} \\
&= \arg\min_{\mathbf{\Phi}_k\in\chi}\left\{M\ln\left(\det\left(\mathbf{R}_k\right)\right)-\mathrm{tr}\left(\mathbf{Y}^{\mathrm{H}}\mathbf{\Theta}_k\mathbf{D}_k\mathbf{\Theta}_k^{\mathrm{H}}\mathbf{Y}\right)\right\} \\
&= \arg\min_{\mathbf{\Phi}_k\in\chi}\left\{M\ln\left(\det\left(\mathbf{R}_k\right)\right)-\|\mathbf{Y}^{\mathrm{H}}\mathbf{\Theta}_k\mathbf{D}_k^{\frac{1}{2}}\|_F^2\right\}
\end{aligned} \tag{5.70}$$

5.4.3 Examples and Simulation Results

Here we present the simulation results for the cases with $M = 1$ and $M = 2$ antennas. For both cases we assume that the terminals and the relays share the power, $P_1 = P_2 = P_R/2$, under the total power constraint $P = P_1 + P_2 + P_R$. The first aim is to investigate the effect of the subspace decoding compared with the ML-decoding. The second aim is to compare the performance of two-way and one-way relaying under the non-coherent assumption. However, one has to be carefull when comparing non-coherent one-way and two-way relaying schemes. As discussed in Chapter 4, the distributed space-time coding schemes for one-way relaying assume that the receiver has $N = MK$ antennas. In the case of two-way relaying we already argued that, due to the symmetry we can not assume that the terminals use different number of antennas. However, for a fair comparison, we have to fix the number of receive antennas in the one-way relaying protocol to M.

M = 1, T = 4

First, we present a comparison of the performances of the ML decoding rule given by (5.70) and the proposed decoding achieved by searching for the most likely subspace as in (5.54). It is shown in Fig. 5.2 that the ML decoding rule and the decoding based on subspaces yield nearly identical performance. Additionally, we observe the influ-

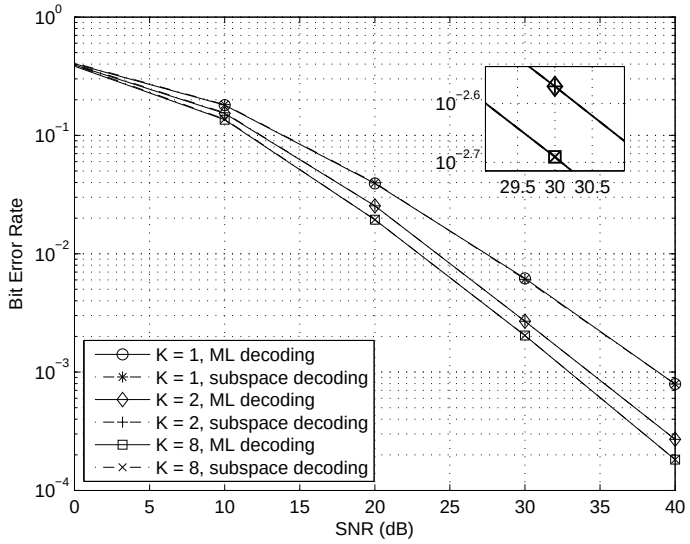

Figure 5.2: Performance comparison for 2 decoding methods of one antennas case $(M = 1)$

ence of the number of relays in the network on the performance. When the number of relays K grows, the noise $\mathbf{W}$ is more white. Moreover, the channel matrices $\mathbf{H}$ and $\mathbf{G}$ are also becoming more Gaussian, since the correlation induced by the channel multiplication is reduced. We fix the total power throughout the simulations. And in the representation of the results we use the power of the terminals $P_1 = P_2 = \frac{P_R}{2}$. As expected, Fig. 5.2 shows the performance is better with an increase of number of relays. For number of relays $K > 4$, the advantage of more relays is not apparent any more. For example, there is only a slight difference between the performance when $K = 4$ and $K = 8$.

In order to support the statement that the two-way relaying is spectral efficient, we compare the performances of the non-coherent two-way relaying scheme proposed in this paper and a non-coherent one-way relaying scheme, normalized to have the same effective rate. For our simulations, we choose the symbols in the signal transmitted by Terminal 1 ϕ, ϕ_1, ϕ_2 from a QPSK constellation. The same holds for the signal ψ transmitted by Terminal 2, $\psi_1, \psi_2 \in$ QPSK. Since the coherence time is $T = 4$,

the effective rate is 0.5 bit/c.u. for each user, which gives $\eta = 1$ bits/c.u. in total for the two users. For the free parameters α_1 and α_2 in the code construction which serve to separate the codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$ such that the new codebook $\mathcal{Q}$ has well distinguishable codewords, we use $\alpha_1 = \alpha_2 = \frac{\pi}{4}$, which provides the best performance according to the simulations. In the case of one-way relaying scheme, we take the code construction as in (3.9)

$$\mathbf{C} = \begin{pmatrix} a_k \mathbf{I}_M \\ b_k a_{k-1} \mathbf{S}_k \\ \vdots \\ b_k \cdots b_2 a_1 \mathbf{S}_2 \times \cdots \times \mathbf{S}_k \\ b_k \cdots b_1 a_0 \mathbf{S}_1 \times \mathbf{S}_2 \times \cdots \times \mathbf{S}_k \\ b_k \cdots b_1 b_0 \mathbf{S}_0 \times \mathbf{S}_1 \times \mathbf{S}_2 \times \cdots \times \mathbf{S}_k \end{pmatrix},$$

where scalars $a_0, a_1, \ldots, a_k$ and $b_0, b_1, \ldots, b_k$ satisfy $a_i = \cos \alpha_i$ and $b_i = \sin \alpha_i$, $\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_k$ are $M \times M$ unitary matrices. This ensures that the effective rate is the same as in the case of the two-way relaying scheme. Additionally, the coherence time T is also kept the same for a fair comparison. Hence, for $M = 1$, the transmit signal ϕ of each terminal is

$$\phi = \begin{pmatrix} a_2 \\ b_2 a_1 s_2 \\ b_2 b_1 a_0 s_1 s_2 \\ b_2 b_1 b_0 s_0 s_1 s_2 \end{pmatrix}$$

with $s_0 \in$ 16PSK, $s_1, s_2 \in$ QPSK, $a_0 = a_1 = a_2 = \cos(\frac{\pi}{4})$ and $b_0 = b_1 = b_2 = \sin(\frac{\pi}{4})$. Fig. 5.3 shows that the non-coherent two-way relaying outperforms the non-coherent one-way relaying, when the effective rate in the network is kept constant.

M = 2, T = 8

For the simulations of two antennas case, the same analysis has been performed. The results in terms of BER curve are given in Fig. 5.4 and Fig. 5.5. Same conclusions as those of the one antenna case can be drawn.

In the case of two-way relaying, for the signal transmitted by Terminal 1 $\mathbf{\Phi}$, we choose the symbols $\phi_1, \phi_2, \phi_3, \phi_4$ from a QPSK constellation. The same holds for the signal $\mathbf{\Psi}$, transmitted by Terminal 2, $\psi_1, \psi_2, \psi_3, \psi_4 \in$ QPSK. Since the coherence time is $T = 8$, the sum rate is $\eta = 1$ bit/c.u. for the two users. We choose $\alpha_1 = \alpha_2 = \frac{\pi}{4}$. In the simulation of the one-way relaying scheme, we use the code construction introduced in (3.9), with the same effective rate $\eta = 1$ bits/c.u. and the same coherent time,

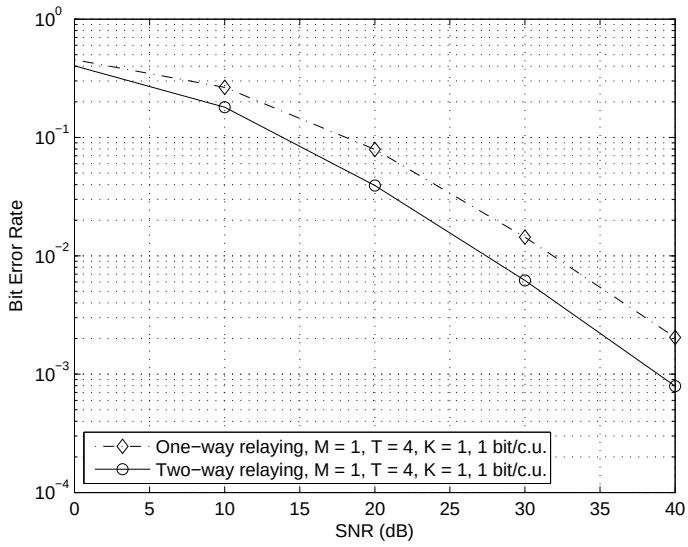

Figure 5.3: Performance comparison for non-coherent two-way relaying scheme and non-coherent one-way relaying scheme $(M = 1)$

$T = 8$. The code construction is as follows

$$\mathcal{C} = \begin{pmatrix} a_2\mathbf{I}_2 \\ b_2a_1\mathbf{S}_2 \\ b_2b_1a_0\mathbf{S}_1\mathbf{S}_2 \\ b_2b_1b_0\mathbf{S}_0\mathbf{S}_1\mathbf{S}_2 \end{pmatrix}$$

where $\mathbf{S}_0, \mathbf{S}_1, \mathbf{S}_2$ are taken from a 2×2 Alamouti code [2]. The symbols in $\mathbf{S}_0$ are from a 16PSK constellation, the symbols in $\mathbf{S}_1$ and $\mathbf{S}_2$ from a QPSK constellation. The scaling factors are chosen such that $a_0 = a_1 = a_2 = \cos(\frac{\pi}{4})$ and $b_0 = b_1 = b_2 = \sin(\frac{\pi}{4})$.

5.5 Differential Scheme for Two-Way Relaying

In the following we present a differential scheme for the two-way relaying setup. The differential transmission does not require channel knowledge at the terminals and at the relays. However, one has to be carefull when speaking about differential and non-coherent transmission. We should have in mind that the system model for the differential scheme differs from the block fading model we use. To be correct, the differential scheme assumes that the channel is constant in *any* two consecutive time blocks, which obviously means that the channel should be constant for the duration of the entire transmission process. However, if the channel is constant for a

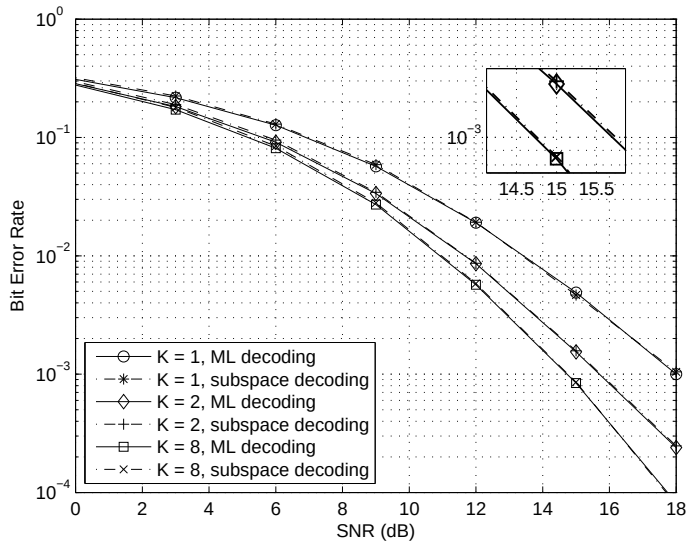

Figure 5.4: Performance comparison for 2 decoding methods of two antennas case $(M = 2)$

fixed block of duration T (coherence time), then the differential scheme should be reinitialized. Another way around this assumption is the introduction of the slowly changing channel, as given in [20]. Here, we will perform the simulations under this assumption.

A differential scheme for the one-way relaying channel was developed by Jing and Hassibi in [25] and also employed in [27]. Although here we follow a similar framework, the fact that the signals from the terminals are added at the relays makes the problem of differential coding for two way-relaying specific. Namely, without having the knowledge of the channel coefficients, it is difficult (if at all possible) for each terminal to completely subtract its own contribution from the signal broadcasted from the relays. In this sense, the differential scheme will differ from the usual differential scheme for distributed space time coding in one-way (uni-directional) relaying. Thus, we have to handle the problem in a way which allows for recursive representation of the system equations, necessary for differential transmission to take place.

The exchange protocol we consider has two stages. The first stage is a multicast stage when the terminals send their signals to the relays. In the second stage the relays broadcast, after performing some processing. The two terminals, knowing what they sent, can decide on the received data from the other terminal by subtracting their own contribution from the signal broadcasted by the relays. The multicast and the broadcast stage save time slots compared to the case where the information exchange between the terminals requires four stages, two for each direction. This leads

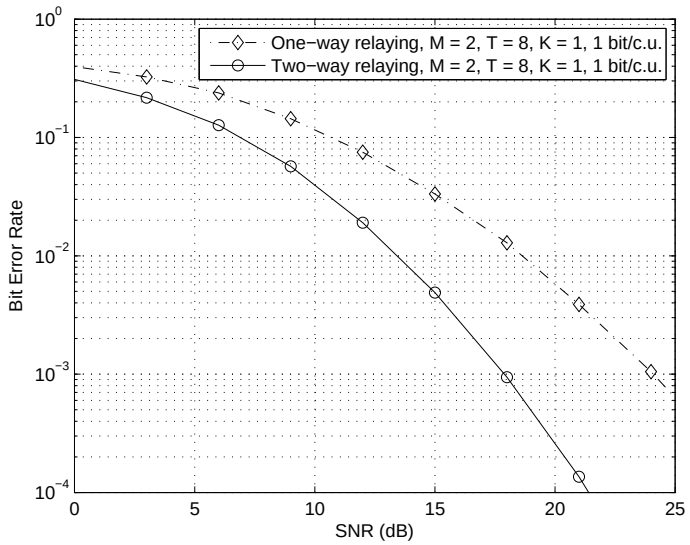

Figure 5.5: Performance comparison for non-coherent two-way relaying scheme and non-coherent one-way relaying scheme ($M = 2$)

to increased throughput in the wireless network. If instead of throughput we like to gain reliability, we can fix the throughput to be the same in both protocols. In the two-stage protocol this will allow for using alphabets of lower cardinality at the terminals compared to the four-stage protocol. Here we propose an idea of applying the distributed differential scheme to two-way relaying. The aim is to represent the transmission in a way which allows for differential encoding and decoding of the information. For that, we first start with a four-stage initialization phase. After the initialization of the protocol, begins the data transmission phase where the information exchange is performed in two stages.

5.5.1 System Model

We use the same system model (5.3) as in Section 5.2, with the difference that we have time dependence due to the differential assumption. Therefore, for all signals we have additional notation which denotes the time block when the transmission takes place. We denote the signal transmitted by T_1 in time block n as $\mathbf{S}^{(n)}$ and the signal transmitted by T_2 as $\mathbf{D}^{(n)}$. We normalize the transmitted signals as $\mathrm{E}[\mathrm{tr}(\mathbf{S}^{(n)\mathrm{H}}\mathbf{S}^{(n)})] = T$ and $\mathrm{E}[\mathrm{tr}(\mathbf{D}^{(n)\mathrm{H}}\mathbf{D}^{(n)})] = T$. Further, we denote the channel between T_1 and the k-th relay as $\mathbf{H}_k^{(n)}$ and the channel between T_2 and the k-th relay as $\mathbf{G}_k^{(n)}$. Additionally, we denote the reverse channels as $\mathbf{H}_k^{(n,r)}$ and $\mathbf{G}_k^{(n,r)}$ respectively. Additionally, when we address the received signals at the relays and at the terminals, we will make a

distinction between the parts of the signal which arise from T_1 and T_2, by using the scripts s and d respectively.

5.5.2 Initialization of the Protocol

In the analysis performed here, without loss of generality, we will observe the signal that T_2 receives. For the received signal of T_1 holds the same analysis, due to the symmetry of the network.

The initialization of the protocol requires four stages. In the first stage T_1 transmits an initialization matrix $\mathbf{S}^{(0)} = \mathbf{I}_M$, which gives the following relay receive vectors

$$\mathbf{R}_{s,k}^{(0)} = \sqrt{P_1 T}\mathbf{S}^{(0)}\mathbf{H}_k^{(0)} + \sqrt{P_2 T}\mathbf{V}_{s,k}^{(0)} \quad k = 1, 2, \ldots, K. \tag{5.71}$$

In the second stage, the relays broadcast to both terminals. The k-th relay broadcasts the vector

$$\mathbf{T}_{s,k}^{(0)} = \sqrt{\tfrac{P_3}{P_1+1}}\mathbf{A}_k\mathbf{R}_{s,k}^{(0)},\ k = 1, 2, \ldots, L, \tag{5.72}$$

and

$$\mathbf{T}_{s,k}^{(0)} = \sqrt{\tfrac{P_3}{P_1+1}}\mathbf{A}_k\mathbf{R}_{s,k}^{*(0)},\ k = L+1, \ldots, K. \tag{5.73}$$

Hence, T_2's received signal which results from T_1's transmitted signal is

$$\mathbf{Y}_s^{(0)} = \sqrt{\tfrac{P_1 P_3 T}{P_1+1}}\mathbf{\Phi}^{(0)}\mathbf{H}^{(0)} + \mathbf{W}_s^{(0)} \tag{5.74}$$

where

$$\begin{aligned}
\mathbf{\Phi}^{(0)} &= \left(\mathbf{A}_1\mathbf{S}^{(0)} \ \ldots\ \mathbf{A}_L\mathbf{S}^{(0)}\ \mathbf{A}_{L+1}\mathbf{S}^{*(0)} \ \ldots\ \mathbf{A}_R\mathbf{S}^{*(0)}\right) \\
\mathbf{H}^{(0)} &= \sum_{k=1}^{K}\mathbf{H}_k^{(0)}\mathbf{G}_k^{(0,r)} \\
\mathbf{W}_s^{(0)} &= \sqrt{\tfrac{P_3}{P_1+1}}\sum_{k=1}^{L}\mathbf{A}_k\mathbf{V}_{s,k}^{(0)}\mathbf{G}_k^{(0,r)} + \sum_{k=L+1}^{R}\mathbf{A}_k\mathbf{V}_{s,k}^{*(0)}\mathbf{G}_k^{(0,r)} + \mathbf{Z}_s^{(0)}.
\end{aligned} \tag{5.75}$$

The same initialization is performed by T_2. In the third stage, T_2 sends its initialization vector to the relays. In the fourth stage the relays broadcast to both terminals. The received signal at T_2 which results from T_2's transmitted signal is given as

$$\mathbf{Y}_d^{(0)} = \sqrt{\tfrac{P_2 P_3 T}{P_2+1}}\mathbf{\Psi}^{(0)}\mathbf{G}^{(0)} + \mathbf{W}_d^{(0)}, \tag{5.76}$$

where

$$\mathbf{\Psi}^{(0)} = \left(\mathbf{A}_1\mathbf{D}^{(0)} \ \ldots \ \mathbf{A}_L\mathbf{D}^{(0)} \ \mathbf{A}_{L+1}\mathbf{D}^{*(0)} \ \ldots \ \mathbf{A}_R\mathbf{D}^{*(0)}\right)$$
$$\mathbf{G}^{(0)} = \sum_{k=1}^{K} \mathbf{G}_k^{(0)}\mathbf{G}_k^{(0,r)}$$

$$\mathbf{W}_d^{(0)} = \sqrt{\tfrac{P_3}{P_1+1}} \sum_{k=1}^{L} \mathbf{A}_k\mathbf{V}_{d,k}^{(0)}\mathbf{G}_k^{(0,r)} + \sum_{k=L+1}^{R} \mathbf{A}_k\mathbf{V}_{d,k}^{*(0)}\mathbf{G}_k^{(0,r)} + \mathbf{Z}_d^{(0)}. \tag{5.77}$$

At the end of the initialization phase, T_1 has received the signals $\mathbf{Y}_{s,T_1}^{(0)}$ and $\mathbf{Y}_{d,T_1}^{(0)}$ and T_2 the signals $\mathbf{Y}_{s,T_2}^{(0)}$ and $\mathbf{Y}_{d,T_2}^{(0)}$ respectively. These signals implicitly contain channel information which is needed for the differential transmission in the data transmission stage.

5.5.3 Data Transmission Phase

After the initialization stage, a two-stage exchange model is applied. In the multicast stage both terminals transmit simultaneously to the relays. The relays perform linear processing of the received sum of both signals, and broadcast in the second stage. The relays received vectors in the first multicast stage are given by

$$\mathbf{R}_k^{(1)} = \sqrt{P_1 T}\mathbf{S}^{(1)}\mathbf{H}_k^{(1)} + \mathbf{D}^{(1)}\mathbf{G}_k^{(1)} + \mathbf{V}_k^{(1)}. \tag{5.78}$$

In the broadcast stage the relays transmit vectors which are linear combinations of the received signals

$$\mathbf{T}_k^{(1)} = \sqrt{\tfrac{P_3}{P_1+P_2+1}}\mathbf{A}_k\mathbf{R}_k^{(1)}, k = 1, 2, \ldots, L \tag{5.79}$$

or

$$\mathbf{T}_k^{(1)} = \sqrt{\tfrac{P_3}{P_1+P_2+1}}\mathbf{A}_k\mathbf{R}_k^{*(1)}, k = L+1, \ldots, R. \tag{5.80}$$

The signal received by T_2 is

$$\mathbf{Y}_1 = \sqrt{\tfrac{P_1P_3T}{P_1+P_2+1}}\mathbf{\Phi}^{(1)}\mathbf{H}^{(1)} + \sqrt{\tfrac{P_2P_3T}{P_1+P_2+1}}\mathbf{\Psi}^{(1)}\mathbf{G}^{(1)} + \mathbf{W}^{(1)}, \tag{5.81}$$

where

$$\begin{aligned}
\mathbf{\Phi}^{(1)} &= \left(\mathbf{A}_1\mathbf{S}^{(1)} \ \dots \ \mathbf{A}_L\mathbf{S}^{(1)} \ \mathbf{A}_{L+1}\mathbf{S}^{*(1)} \ \dots \ \mathbf{A}_R\mathbf{S}^{*(1)}\right) \\
\mathbf{H}^{(1)} &= \sum_{k=1}^{K} \mathbf{H}_k^{(1)}\mathbf{G}_k^{(1,r)} \\
\mathbf{\Psi}^{(1)} &= \left(\mathbf{A}_1\mathbf{D}^{(1)} \ \dots \ \mathbf{A}_L\mathbf{D}^{(1)} \ \mathbf{A}_{L+1}\mathbf{D}^{*(1)} \ \dots \ \mathbf{A}_R\mathbf{D}^{*(1)}\right) \\
\mathbf{G}^{(1)} &= \sum_{k=1}^{K} \mathbf{G}_k^{(1)}\mathbf{G}_k^{(1,r)} \\
\mathbf{W}^{(1)} &= \sqrt{\tfrac{P_3}{P_1+P_2+1}}\left(\sum_{k=1}^{L}\mathbf{A}_k\mathbf{V}_{d,1}^{(1)}\mathbf{G}_k^{(1,r)} + \sum_{k=L+1}^{R}\mathbf{A}_k\mathbf{V}_{d,k}^{*(1)}\mathbf{G}_k^{(1,r)}\right) + \mathbf{Z}_d^{(1)}.
\end{aligned} \tag{5.82}$$

We assume that the channel is slow changing, so we can write $\mathbf{H}^{(1)} = \mathbf{H}^{(0)}$, $\mathbf{H}^{(1,r)} = \mathbf{H}^{(0,r)}$, $\mathbf{G}^{(1)} = \mathbf{G}^{(0)}$and $\mathbf{G}^{(1,r)} = \mathbf{G}^{(0,r)}$. Also, we assume that in general we encode the data of both terminals into scaled unitary matrices, as proposed in [27]. The transmit vectors are given as

$$\mathbf{S}^{(1)} = \tfrac{1}{a^{(0)}}\mathbf{U}^{(1)}\mathbf{S}^{(0)} \quad \text{and} \quad \mathbf{D}^{(1)} = \tfrac{1}{b^{(0)}}\mathbf{V}^{(1)}\mathbf{D}^{(0)}. \tag{5.83}$$

One of the requirements for differential distributed space-time codes is that the relay matrices commute with the matrices of the sets $\mathcal{U}$ and $\mathcal{V}$[25]

$$\begin{aligned}
\mathbf{A}_k\mathbf{U}_i &= \mathbf{U}_i\mathbf{A}_k, \quad \forall\mathbf{U}_i \in \mathcal{U},\ k = 1,\dots,L \\
\mathbf{A}_k\mathbf{U}_i^* &= \mathbf{U}_i\mathbf{A}_k, \quad \forall\mathbf{U}_i \in \mathcal{U},\ k = L+1,\dots,K,
\end{aligned} \tag{5.84}$$

and

$$\begin{aligned}
\mathbf{A}_i\mathbf{V}_k &= \mathbf{V}_k\mathbf{A}_i, \quad \forall\mathbf{V}_k \in \mathcal{V},\ i = 1,\dots,L \\
\mathbf{A}_i\mathbf{V}_k^* &= \mathbf{V}_k\mathbf{A}_i, \quad \forall\mathbf{V}_k \in \mathcal{V},\ i = L+1,\dots,K.
\end{aligned} \tag{5.85}$$

The received signal at T_2 is then

$$\begin{aligned}
\mathbf{Y}_1 &= \sqrt{\tfrac{P_1P_3T}{P_1+P_2+1}}\tfrac{1}{a^{(0)}}\mathbf{U}^{(1)}\mathbf{\Phi}^{(0)}\mathbf{H}^{(0)} + \sqrt{\tfrac{P_2P_3T}{P_1+P_2+1}}\tfrac{1}{b^{(0)}}\mathbf{V}^{(1)}\mathbf{\Psi}^{(0)}\mathbf{G}^{(0)} + \mathbf{W}^{(1)} \\
&= \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{a^{(0)}}\mathbf{U}^{(1)}(\mathbf{Y}_s^{(0)} - \mathbf{W}_s^{(0)}) + \sqrt{\tfrac{P_2+1}{P_1+P_2+1}}\tfrac{1}{b^{(0)}}\mathbf{V}^{(1)}(\mathbf{Y}_d^{(0)} - \mathbf{W}_d^{(0)}) + \mathbf{W}^{(1)} \\
&= \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{a^{(0)}}\mathbf{U}^{(1)}\mathbf{Y}_s^{(0)} + \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{b^{(0)}}\mathbf{V}^{(1)}\mathbf{Y}_d^{(0)} + \tilde{\mathbf{W}}^{(1)},
\end{aligned} \tag{5.86}$$

where

$$\tilde{\mathbf{W}}^{(1)} = -\left(\sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{a^{(0)}}\mathbf{U}^{(1)}\mathbf{W}_s^{(0)} + \sqrt{\tfrac{P_2+1}{P_1+P_2+1}}\tfrac{1}{b^{(0)}}\mathbf{V}^{(1)}\mathbf{W}_d^{(0)}\right) + \mathbf{W}^{(1)}.$$

We observe that T_2 already knows $\mathbf{V}^{(1)}$. Hence, it can decide for $\mathbf{U}^{(1)}$ based on

$$\hat{\mathbf{U}}^{(1)} = \arg\min_{\mathbf{U}^{(1)}\in\mathcal{U}} \left\| \mathbf{Y}_1 - \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{a^{(0)}}\mathbf{U}^{(1)}\mathbf{Y}_s^0 - \sqrt{\tfrac{P_2+1}{P_1+P_2+1}}\tfrac{1}{b^{(0)}}\mathbf{V}^{(1)}\mathbf{Y}_d^{(0)} \right\|_{\mathrm{F}}^2. \tag{5.87}$$

After obtaining $\hat{\mathbf{U}}^{(1)}$, T_2 can obtain an estimate of the noise $\tilde{\mathbf{W}}^{(1)}$ as

$$\hat{\tilde{\mathbf{W}}}^{(1)} = \mathbf{Y}_1 - \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\left(\tfrac{1}{a^{(0)}}\hat{\mathbf{U}}^{(1)}\mathbf{Y}_s^0 - \tfrac{1}{b^{(0)}}\mathbf{V}^{(1)}\mathbf{Y}_d^{(0)}\right). \tag{5.88}$$

In order to be able to proceed with the differential transmission, we need a representation of the received information signal (contribution from T_1) as a function of the previous information signal. Additionally, we want that T_2 subtracts its own contribution to the received signal. The difficulty arises from the fact that both signals are added at the relays in the multicast stage and without channel knowledge it seems difficult (if at all possible) for T_2 to completely subtract its contribution to the received signal.

We propose a possible solution in the following way. First, we divide the expression 5.86 in two parts containing the contributions from each terminal

$$\mathbf{Y}_1 = \mathbf{Y}_s^{(1)} + \mathbf{Y}_d^{(1)}. \tag{5.89}$$

Please note that this representation is virtual, since we can not separate the contributions from the two terminals due to the fact that the signals are added at the relays. However, we can substitute the contributions $\mathbf{Y}_s^{(1)}$ and $\mathbf{Y}_d^{(1)}$ with their estimates $\hat{\mathbf{Y}}_s^{(1)}$ and $\hat{\mathbf{Y}}_d^{(1)}$ respectively

$$\mathbf{Y}_1 = \hat{\mathbf{Y}}_s^{(1)} + \hat{\mathbf{y}}_d^{(1)}, \tag{5.90}$$

where the estimates are given by

$$\begin{aligned}
\hat{\mathbf{Y}}_s^{(1)} &= \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{a^{(0)}}\hat{\mathbf{U}}^{(1)}\mathbf{Y}_s^{(0)} + \frac{\hat{\tilde{\mathbf{W}}}^{(1)}}{2} \\
\hat{\mathbf{Y}}_d^{(1)} &= \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{b^{(0)}}\mathbf{V}^{(1)}\mathbf{Y}_d^{(0)} + \frac{\hat{\tilde{\mathbf{W}}}^{(1)}}{2},
\end{aligned} \tag{5.91}$$

Therefore, the estimated noise is shared among the two vectors $\hat{\mathbf{Y}}_s^{(1)}$ and $\hat{\mathbf{Y}}_d^{(1)}$. These two vectors will be used during the second time block for differential decoding.

We can continue in the same way we described above. At time block n the signal recived at T_2 is given by

$$\mathbf{Y}^{(n)} = \sqrt{\tfrac{P_1P_3T}{P_1+P_2+1}}\mathbf{\Phi}^{(n)}\mathbf{H}^{(n)} + \sqrt{\tfrac{P_2P_3T}{P_1+P_2+1}}\mathbf{\Psi}^{(n)}\mathbf{G}^{(n)} + \mathbf{W}^{(n)}, \tag{5.92}$$

where $\mathbf{\Phi}^{(n)}, \mathbf{H}^{(n)}, \mathbf{\Psi}^{(n)}$ and $\mathbf{G}^{(n)}$ are given as in (5.82). The general recursive system equation for the received signal at T_2 is given by

$$\mathbf{Y}_n = \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{a^{(n-1)}}\mathbf{U}^{(n)}\hat{\mathbf{Y}}_s^{(n-1)} + \sqrt{\tfrac{P_2+1}{P_1+P_2+1}}\tfrac{1}{b^{(n-1)}}\mathbf{V}^{(n)}\hat{\mathbf{Y}}_d^{(n-1)} + \tilde{\mathbf{W}}^{(n)}, \tag{5.93}$$

where

$$\begin{aligned}\hat{\mathbf{Y}}_s^{(n-1)} &= \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{\hat{a}^{(n-2)}}\hat{\mathbf{U}}^{(n-1)}\hat{\mathbf{Y}}_s^{(n-2)} + \frac{\hat{\tilde{\mathbf{W}}}^{(n-1)}}{2}\\ \hat{\mathbf{Y}}_d^{(n-1)} &= \sqrt{\tfrac{P_2+1}{P_1+P_2+1}}\tfrac{1}{\hat{b}^{(n-2)}}\mathbf{V}^{(n-1)}\hat{\mathbf{Y}}_d^{(n-2)} + \frac{\hat{\tilde{\mathbf{W}}}^{(n-1)}}{2}.\end{aligned} \tag{5.94}$$

and

$$\tilde{\mathbf{W}}^{(n)} = -\left(\sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{a^{(n-1)}}\mathbf{U}^{(n)}\mathbf{W}_s^{(n-1)} + \sqrt{\tfrac{P_2+1}{P_1+P_2+1}}\frac{1}{b^{(n-1)}}\mathbf{V}^{(n)}\mathbf{W}_d^{(n-1)}\right) + \mathbf{W}^{(n)}. \tag{5.95}$$

The decoding rule is then

$$\hat{\mathbf{U}}^{(n)} = \arg\min_{\mathbf{U}^{(n)}\in\mathcal{U}} \left\| \mathbf{Y}_n - \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{a^{(n-1)}}\mathbf{U}^{(n)}\hat{\mathbf{Y}}_s^{(n-1)} - \sqrt{\tfrac{P_2+1}{P_1+P_2+1}}\tfrac{1}{b^{(n-1)}}\mathbf{V}^{(n)}\hat{\mathbf{Y}}_d^{(n-1)} \right\|_F^2. \tag{5.96}$$

5.5.4 Power Allocation

The representation of the received signal at T_2 in the form (5.93 makes the problem of optimal power allocation similar to the one in distributed differential space-time coding for one-way relaying [24], with the difference that we have the two-way relay channel. According to the analysis in [9], [24], if we can assume that we can separate the signals originating from the two terminals, after subtracting its own contribution, each terminal sees a one-way relay channel. Thus, following the analogy in [25], [9], [24], the optimal power allocation gives $P_1 = P_2 = P_3 R/2$, i.e. the terminals and the relays share the available power in the network. Please note that if we represent the two-way relay channel as two one-way relay channels, the relays are active in both phases of the transmission and the terminals only in one, which explains the factor of $1/2$. The simulations we performed with power allocations different from the one discussed here support this observation and are in favor of the equal power allocation.

5.5.5 Codes for Differential Two-Way Relaying

As differential DSTCs which are applicable for two-way relaying we can use the usual codes for differential distributed space time coding. They should have the following characteristics

- All codewords are scaled unitary matrices respecting the transmit power constraint.
- There exist R unitary matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_K$ of size $T \times T$ such that the first L of them satisfy $\mathbf{A}_k\mathbf{C} = \mathbf{C}\mathbf{A}_k, k = 1, \ldots, L, \forall \mathbf{C} \in \mathcal{C}$ and the remaining $K - L$ of them satisfy $\mathbf{A}_k\mathbf{C}^* = \mathbf{C}\mathbf{A}_k, k = L+1, \ldots, R, \forall \mathbf{C} \in \mathcal{C}$.
- There exists an initial matrix $\mathbf{S}^{(0)}$ such that the matrix

$$\mathbf{\Phi}^{(0)} = (\mathbf{A}_1\mathbf{S}^{(0)}\ \mathbf{A}_2\mathbf{S}^{(0)} \ldots \mathbf{A}_L\mathbf{S}^{(0)}\ \mathbf{A}_{L+1}\mathbf{S}^{(0)*}\ \mathbf{A}_K\mathbf{S}^{(0)*}) \tag{5.97}$$

  is unitary.

Codes which satisfy this conditions have been studied in the literature. We will divide them in two major groups, unitary and scaled unitary codes. In the first group we find the Alamouti code [2], $S_p(2)$ codes [23], the circulant codes introduced in [25], etc. The scaled unitary codes were introduced in [27]. The choice of appropriate codes is not the main topic here, but we will summarize some of the codes appropriate as distributed differential codes for two-way relaying and used in our simulations.
Unitary codes - When the distributed differential STCs codes are unitary, we set the scaling factors $a^{(n)}, b^{(n)} = 1$. In this group we find the Alamouti code [2], $S_p(2)$ codes, the circulant codes introduced in [25], etc.
Scaled unitary codes - Distributed differential space-time codes with low decoding complexity (group decodable), based on scaled unitary codes are introduced in [27]. The codes are based on DSTCs from extended Clifford algebras developed in [38] for coherent collocated MIMO communication. The design for 4 transmit antennas is given as

$$\mathcal{C} = \begin{pmatrix} u_1 & u_2 & -u_3^* & -u_4^* \\ u_2 & u_1 & -u_4^* & -u_3^* \\ u_3 & u_4 & u_1^* & u_2^* \\ u_4 & u_3 & u_2^* & u_1^* \end{pmatrix}. \tag{5.98}$$

and the design for larger number of transmit antennas can also be easily constructed. The signal set is chosen as in [38].

5.5.6 Examples and Simulation Results

In this section, we simulate the performance of the proposed distributed differential scheme and compare it with the performance of a 4-time slots (one-way relaying) distributed differential scheme. As a horizontal reference in the shown figures, we use the total power used in one time slot for a one-way transmission. The noises are assumed to be independent complex Gaussian random variables with mean zero and

variance one. As for the channel, we used a GSM channel model with a symbol sampling period of $T_s = 3.693\mu$s and a maximum Doppler shift of 75 Hz. This ensures a slowly changing channel and allows the assumption of a constant channel over two consecutive time blocks. Fig. 5.6depicts the performance of the differential scheme

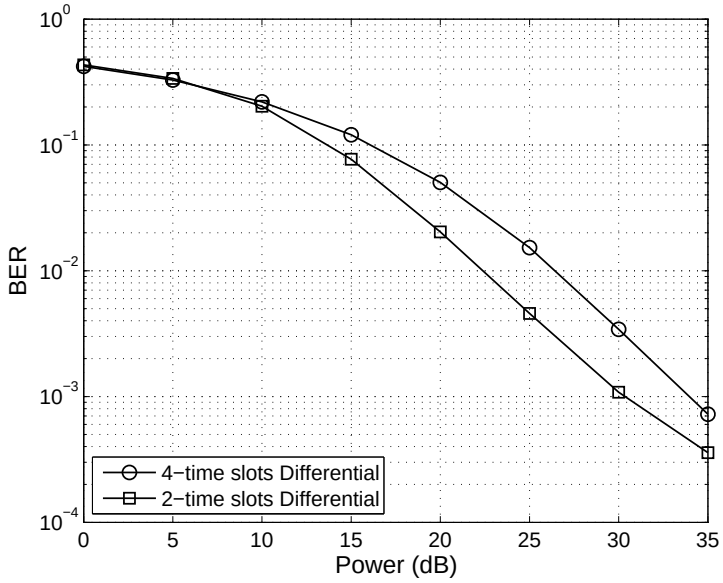

Figure 5.6: Performance of the distributed differential scheme in a network with two relays

using 4 and 2 time slots protocols in a network with two relays. For that, we fixed the transmission rates by choosing our information symbols from a 16-PSK constellation for the 4 time slots protocol, and from a 4-PSK constellation for the 2 time slots protocol. This insures a fixed one-way transmission rate of 1 bit/cu, and allows for a fair reliability comparison. Furthermore, we used the Alamouti space-time code as a unitary matrix in which we embedded the information symbols. As shown in the figure, the 2 time slots protocol outperforms the 4 time slots one by approximately 4dB. In Fig. 5.7, the same simulations have been extended to a network with four relays. The $Sp(2)$ space-time code was used as a unitary matrix to carry the information symbols. Also, we fixed the one-way bitrate to 2.56 bit/cu in order to fairly compare the performance of the schemes. For that, we chose $P = 5$ and $Q = 7$ for the 4 time slots protocol, and $P = 2$ and $Q = 3$ for the 2 time slots protocol. In the middle SNR region, the 2 time slots protocol outperforms the 4 time slots protocol by approximately 2dB. However, as we can see, in the high SNR region, there is an effect of error floor. We suspect that the reason behind this is the way in which we separate the signals at the receiver in order to write the differential system equation and be able to detect

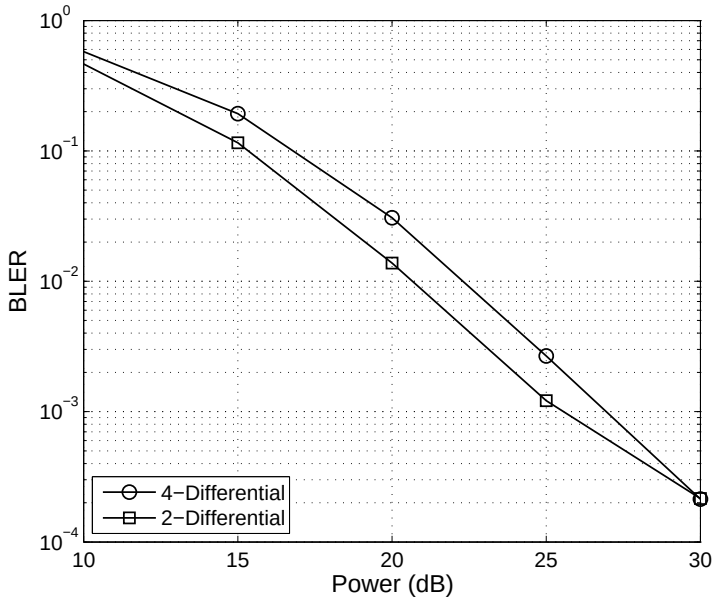

Figure 5.7: Performance of the distributed differential scheme in a network with four relays

the signals in a differential way. As we mentioned before, the main difficulty arises from the fact that without channel knowledge it is difficult to completely avoid the self interference. Another reason might be that the effective channel matrices $\mathbf{H}$ and $\mathbf{G}$ are products of two channel matrices, and thus not Gaussian matrices any more. Additionally, the entries in the matrices $\mathbf{H}$ and $\mathbf{G}$ are not statistically independent, which we think attributes to the performance.

5.6 Chapter Summary

Two-Way Relaying

- The two-way relaying schemes promise throughput gain compared to one-way relaying, by saving time slots in the multicast and the broadcast stage.

Bounds on the achievable two-way rate in the relay network

Upper bound on the achievable rates in the non-coherent MIMO two-way channel

$$R_{12} \leq \frac{M}{2}\left(1-\frac{M}{T}\right)\log_2\frac{P_R}{\sigma^2}+\frac{1}{2}c+o(1).$$

$$R_{21} \leq \frac{M}{2}\left(1-\frac{M}{T}\right)\log_2\frac{P_R}{\sigma^2}+\frac{1}{2}c+o(1).$$

Lower bound on the achievable rates

$$R_{12}^{AF}=\frac{M}{2}\left(1-\frac{M}{T}\right)\log_2\left(\frac{P_R}{\sigma^2}\right)+\frac{1}{2}r_{12}+o(1).$$

$$R_{21}^{AF}=\frac{M}{2}\left(1-\frac{M}{T}\right)\log_2\left(\frac{P_R}{\sigma^2}\right)+\frac{1}{2}r_{21}+o(1).$$

- At high SNR the lower and the upper bound in both cases (reciprocal and non-reciprocal channels) differ only up to a constant.
- The non-coherent two-way relay channel achieves $M(1-\frac{M}{T})$ degrees of freedom, although there is no channel knowledge at neither the relays nor at the transceivers. Second, these degrees of freedom are achieved by an AF scheme.

Codes for genuine non-coherent two-way relaying

- The received signal is given as

$$\mathbf{Y}=\mathbf{\Phi}'\mathbf{H}+\mathbf{\Psi}'\mathbf{G}+\mathbf{W}$$

Derivation of the ML-Decoding Rule

$$\hat{\mathbf{\Phi}}=\arg\min_{\mathbf{\Phi}_k\in\mathcal{X}}\left\{M\ln\left(\det\left(\mathbf{R}_k\right)\right)-\|\mathbf{Y}^{\mathrm{H}}\mathbf{\Theta}_k\mathbf{D}_k^{\frac{1}{2}}\|_F^2\right\}.$$

Code Construction

- The codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$ are not chosen independently, but rather as two "well separated" subsets of a larger codebook. The codewords are given as

$$\mathbf{\Phi}=\begin{pmatrix}\cos(\alpha_1)\mathbf{A}_\Phi\\ \sin(\alpha_1)\mathbf{B}_\Phi\end{pmatrix}$$

$$\mathbf{\Psi}=\begin{pmatrix}\cos(\alpha_2)\mathbf{A}_\Psi\\ \sin(\alpha_2)\mathbf{B}_\Psi\end{pmatrix}.$$

- We perform the decoding by looking for the most likely transmitted subspace, having the received matrix $\mathbf{Y}$, as in [31]

$$\mathbf{Q}=\arg\max_{\mathbf{Q}_i\in\mathcal{Q}}\|\mathbf{Y}^{\mathrm{H}}\mathbf{Q}_i\|_F^2$$

Differential scheme

- The general recursive system equation for the received signal at T_2 is given by

$$\mathbf{Y}^{(n)} = \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{a^{(n-1)}}U^{(n)}\hat{\mathbf{Y}}_s^{(n-1)} + \sqrt{\tfrac{P_2+1}{P_1+P_2+1}}\tfrac{1}{b^{(n-1)}}\mathbf{V}^{(n)}\hat{\mathbf{Y}}_d^{(n-1)} + \tilde{\mathbf{w}}^{(n)},$$

  where

$$\hat{\mathbf{Y}}_s^{(n-1)} = \sqrt{\tfrac{P_1+1}{P_1+P_2+1}}\tfrac{1}{\hat{a}^{(n-2)}}\hat{U}^{(n-1)}\hat{\mathbf{Y}}_s^{(n-2)} + \tfrac{\hat{\tilde{\mathbf{w}}}^{(n-1)}}{2}$$
$$\hat{\mathbf{Y}}_d^{(n-1)} = \sqrt{\tfrac{P_2+1}{P_1+P_2+1}}\tfrac{1}{\hat{b}^{(n-2)}}\mathbf{V}^{(n-1)}\hat{\mathbf{Y}}_d^{(n-2)} + \tfrac{\hat{\tilde{\mathbf{w}}}^{(n-1)}}{2}.$$

Conclusions and Future Work

In this thesis we addressed the problem of non-coherent transmission in block Rayleigh fading channels. We focused on two communication systems. The first system is the point-to-point MIMO channel without channel knowledge assumption at the transmitter and the receiver. The second system is the wireless network with relays and no direct links between the terminals.

The problem of non-coherent space-time coding can be given a geometric interpretation as constrained packing problem in Grassmann manifolds. The geometric interpretation shads new light on the code construction problem and brings intuition to the problem of understanding the fundamental limits of the communication systems.

In this work we concentrated mostly on presenting novel code constructions for the two systems we mentioned. Additionally, for the two-way relaying channel we presented fundamental bounds on the achievable two-way rates.

In Chapter 1 we presented the necessary preliminaries for understanding the problem of space-time coding in general. We introduced the Rayleigh fading and discussed the specifics of slowly fading channels. The concept of diversity was further discussed. We also presented the block fading model which is used throughout the thesis and gave introduction to multiple antenna channels and space-time coding.

In Chapter 2 we introduced the problem of non-coherent communication over block fading channels. We introduced the concept of coordinate change and explained the geometric interpretation of the coding problem in detail. We also gave the necessary introduction to the geometry of Grassmann manifolds. We focused on the differentiable structure of the Grassmann manifold which provided the basis for the code constructions presented in the next chapters.

In Chapter 3 we presented novel constructions of codes for the non-coherent channel based on the geometric interpretation. The constructions are based on mapping from the tangent space to the manifold based on the exponential map. The first construction relies on mapping unitary codes for the coherent channel from the tangent

space to the manifold. This yields codes of special block structure which can be used in the decoding. Additionally, we showed that the codes obtained in this way preserve some of the properties of the coherent codes such as the diversity product. The second construction is based on the observation that non-coherent codes can be obtained from coherent codes and vice versa. This leads to a recursive construction with block length increasing in each step of the construction. The codes obtained in this way benefit from the large block length and have good properties in terms of diversity and coding gain. The third code construction relies on lattice constructions in the tangent space. It can be shown that some special lattices such as the Gosset, Barnes-Wall and Leech lattice provide Grassmann codes for systems of practical relevance. The dense packing properties of these lattices yield superior performance and allow for simplified decoding. In order to improve the diversity product of the codes, additionally a lattice rotation is performed. The last construction is based on the observation that the Grassmann manifold is a $U(T)$ homogeneous space, which is a generalization of spherical symmetry. We propose a construction from a special high-dimensional code originally developed for vector quantization of multi-dimensional Gaussian sources.

In Chapter 4 and Chapter 5 we addressed the problem of non-coherent transmission in wireless relay networks. We focused on one-way and two-way relaying protocols with half-duplex constraints on the terminals. For the one-way relaying network we presented novel distributed space-time codes based on Grassmann codes for the non-coherent point-to-point MIMO channel. The codes are based on codes designed for the coherent channel, such as the Alamouti, Sp(2) and quasi-orthogonal codes. For the two-way relaying channel we first introduced the concept of non-coherent communication, showing that non-coherent communication is possible, although no channel knowledge is present at neither the terminals nor at the relays. Further, we presented bounds on the achievable two-way rate in the non-coherent setup. We showed that the upper and the lower bound meet in the high SNR regime and differ from the capacity within a constant. As a byproduct we derived the degrees of freedom of the two-way network and showed that amplify-and-forward is optimal strategy in the non-coherent setup. Motivated from the results, we presented two communication schemes. The first one was a differential scheme which extends over the known differential schemes for one-way relaying and adapts to the specifics of the two-way channel. The second scheme we introduced is a genuine non-coherent scheme, designed for the block channel model. Again, the scheme is based on the constructions designed for the non-coherent point-to-point MIMO channel. As important conclusion, we showed that the codebooks of the terminals should be chosen jointly and in a way which allows for eliminating of the self interference and successful decoding of the received signal.

The geometric approach looks promising for construction of codes for non-coherent

communication over block fading point-to-point channels and wireless relay networks. In this work we touched some of the topics where the geometric approach offers possible solutions. When it comes to the most general problem of derivation of the fundamental limits of wireless networks, the geometric approach can also offer additional advantage. In principle, there is a more general approach which arises from the geometrical interpretation of some fundamental concepts in information theory. As example we have the Fisher information which is a way of measuring the amount of information that an observable random variable X carries about an unknown parameter θ. In the case of N unknown parameters, the Fisher information takes a form of an $N \times N$ matrix. The Fisher information matrix is a $N \times N$ positive semidefinite symmetric matrix, defining a Riemannian metric on the N-dimensional parameter space, thus connecting Fisher information to differential geometry. In that context, this metric is known as the Fisher information metric (or Fisher-Rao metric), and the topic is called information geometry. Information geometry can be viewed as an equivalent description of information theory, where the algebraic concepts are given a geometric flavor.

When it comes to the derivation of the fundamental limits of wireless networks, some results from statistical physics can also be used in the development of the geometric framework and the capacity characterization. Wireless networks are fundamentally physical systems, governed by the laws of physics. If a communication system with many degrees of freedom (in time, space, and/or frequency) is modeled as a thermodynamic system, the Shannon capacity is a statistical phase transition point, beyond which arbitrarily low error probability is impossible. Statistical physics offers a number of modeling tools for dealing with non-equilibrium systems and large quantities of random variables. In all of the different disciplines where phase transitions play a role, models are characterized by certain sets of parameters. The space of such parameters can be endowed with a metric and geometrical structure. Actually, the geometric connection is given through the Fisher information. For such a metric a scalar curvature can be calculated. The scalar curvature plays a central role in any attempt to look at phase transitions from a geometrical perspective. For all the models that have been considered so far, the curvature diverges at (and only at) a phase transition point for physical ranges of the parameter values. This gives an indirect connection between the capacity characterization of wireless networks and differential geometry. The connection arises, on one side, due to the interpretation of the capacity as statistical phase transition point, and on other side, from the study of phase transitions in the context of parameter statistics.

Besides the full capacity characterization, capacity approximations may be the key to understanding the performance limits of wireless networks. Promising recent venture in this direction include the degrees of freedom approach. Interference alignment is one technique which is based on the degrees of freedom approach and is used for

the famous K user interference channel. It refers to the idea of constructing signals in such a way that they cast overlapping shadows over one half of the signal space observed by each receiver where they constitute interference, leaving the other half of the signal space free of interference for the desired signal. However, closed form solutions have only been found in certain cases. It is shown that by using long symbol extension the degrees of freedom achieved per dimension approach arbitrarily close to the theoretical outerbound, thereby establishing the degrees of freedom of time-varying interference and X networks. However, the extent to which interference can be aligned over a limited number of dimensions remains an open problem. As a consequence, the maximum number of degrees of freedom that can be achieved through alignment of interference signal vectors is not known in general, i.e. the feasibility of interference alignment over a limited number of signalling dimensions is an open problem. We conjecture that the problem of feasibility of interference alignment has a geometric counterpart as a packing problem with certain metric. The geometric frameworks is expected to give a better insight in the feasibility of interference alignment in some particular cases.

As already mentioned, all results which come from the geometric approach can also be obtained by algebraic means. However, the geometric approach brings additional intuition to the problem and very often offers a more natural interpretation of the results. Therefore, we believe that the geometric approach can be a powerful mean for the understanding and the description of complex wireless networks.

Appendix $\mathcal{A}$

Mathematical Derivations

A.1 Geometry and Algebra Preliminaries

Here we present the preliminaries necessary to understand the structure of Grassmann manifolds. The definitions mainly follow [6] and [10].

A.1.1 Manifolds, Differentiable Manifolds

The formal definition of a *manifold* requires introduction of the term *topological space* and *topological basis*. Formally, a topological space is a set $\mathcal{X}$ together with a collection of open subsets $\mathcal{T}$ that satisfies the conditions:

- The empty set $\varnothing$ is in $\mathcal{T}$,
- $\mathcal{X}$ is in $\mathcal{T}$,
- the intersection of a finite number of sets in $\mathcal{T}$ is also in $\mathcal{T}$,
- the union of an arbitrary number of sets in $\mathcal{T}$ is also in $\mathcal{T}$.

A topological basis is a subset $\mathcal{B}$ of a set $\mathcal{T}$ in which all other open sets can be written as unions or finite intersections of $\mathcal{B}$.

A manifold is then a topological space which is second countable (has a countable topological basis) and is locally Euclidean, i.e. every point on the manifold has a neighborhood which is topologically equivalent (homeomorphic) to an open ball in R^n. In this case the manifold is said to have a dimension n.

Metric can be defined on sets and hence on manifolds. Metric is a nonnegative function $g(x, y)$ describing the "distance" between neighboring points for a given set. A metric satisfies the triangle inequality

$$g(x, y) + g(y, z) \geqslant g(x, z) \tag{A.1}$$

and is symmetric, so $g(x, y) = g(y, x)$. A metric also satisfies $g(x, x) = 0$, as well as the condition that $g(x, y) = 0$ implies $x = y$;

A *coordinate chart* is a way of expressing the points of a small neighborhood of a manifold M as coordinates in Euclidean space. Technically, it is a map $\phi : U \to V$ where U is an open set in M, V is an open set in R^n, where n is the dimension of the manifold. The map ϕ must be one-to-one, and in fact must be a homeomorphism. If there are two neighborhoods U_1 and U_2 with coordinate charts ϕ_1 and ϕ_2, the transition function $\phi_2^{\circ}\phi_1^{-1}$ is well defined since coordinate charts are one-to-one.

An *atlas* is a collection of consistent coordinate charts on a manifold, where "consistent" most commonly means that the transition functions of the charts are smooth. As the name suggests, an atlas corresponds to a collection of maps, each of which shows a piece of a manifold and looks like flat Euclidean space. To use an atlas, one needs to know how the maps overlap. To be useful, the maps must not be too different on these overlapping areas. The overlapping maps from one chart to another are called transition functions. They represent the transition from one chart's point of view to that of another. Let the open unit ball in R^n be denoted B_1. Then if $\phi : U \to B_1$ and $\psi : V \to B_1$ are two coordinate charts, the composition $\phi \circ \psi^{(} - 1)$ is a function defined on $\psi(U \cap V)$. That is, it is a function from an open subset of B_1 to B_1, and given such a function from R^n to R^n, there are conditions for it to be smooth or have k smooth derivatives (i.e., it is a C-k function). A smooth atlas has transition functions that are C-infty smooth (i.e., infinitely differentiable). The consequence is that a smooth function on one chart is smooth in any other chart (by the chain rule for higher derivatives). Similarly, one could have an atlas in class C^k, where the transition functions are in class C-k.

A manifold can be given a differentiable structure locally by using the homeomorphisms in its atlas, combined with the standard differentiable structure on the Euclidean space. In other words, the homeomorphism can be used to give a local coordinate system. To induce a global differentiable structure, the compositions of the homeomorphisms on overlaps between charts in the atlas must be differentiable functions on Euclidean space. In other words, where the domains of charts overlap, the coordinates defined by each chart are required to be differentiable with respect to

the coordinates defined by every other chart. These maps that relate the coordinates defined by the various charts to each other in areas of intersection are called transition maps.

A differentiable manifold is formally a manifold with a globally defined differentiable structure. The manifolds of interest here are differentiable and the notion of differentiability will be exploited later in the construction and analysis of codes in the manifolds. Every differentiable (smooth) manifold M has a tangent bundle T_M, which consists of the tangent space $T_P M$ at all points P in M. Since a tangent space $T_P M$ is the set of all tangent vectors to M at P, the tangent bundle is the collection of all tangent vectors, along with the information of the point to which they are tangent

$$T_M = (P, v) : P \in M,\ v \in T_P M. \tag{A.2}$$

In Riemannian geometry, a Riemannian manifold (M, g) (with Riemannian metric g) is a differentiable manifold M in which each tangent space is equipped with an inner product g in a manner which varies smoothly from point to point. This allows one to define various notions such as angles, lengths of curves, areas (or volumes), curvature, gradients of functions and divergence of vector fields. In other words, a Riemannian manifold is a differentiable manifold in which the tangent space at each point is a finite-dimensional Hilbert space.

A.1.2 Groups, Lie Groups

A *group* G is a finite or infinite set of elements together with a binary operation * (called the group operation) that together satisfy the properties of closure, associativity, the identity property, and the inverse property, defined as:

- Closure: If A and B are two elements in G, then $A * B$ is also in G.
- Associativity: $\forall\ A, B, C \in G,\ (A * B) * C = A * (B * C)$.
- Identity: There is an identity element such that $I * A = A * I = A$, $\forall A \in G$.
- Inverse: There is an inverse of each element, i.e. $\forall A \in G$, $\exists B = A^{-1} \in G$ such that $A * A^{-1} = A^{-1} * A = I$.

A subgroup is a subset H of group elements of a group G that satisfies the group requirements.

Lie group is a group which is also a differentiable manifold, i.e. satisfies the condition that the group operation(s) is(are) differentiable. Due to the differential structure, a *tangent space* can be constructed at every element of the Lie group. Intuitively, the tangent space at a point is the plane tangent to the submanifold at that point. For

d-dimensional manifolds, this plane is a d-dimensional vector space with origin at the point of tangency. The normal space is the orthogonal complement. The tangent space at the identity of a Lie group has the structure of a *Lie algebra*. More formally, a Lie algebra is a nonasociative algebra, i.e a vector space $\mathfrak{g}$ over a field F with a binary operation. The binary operation is the *Lie bracket* which satisfies:

- Bilinearity: $[ax + by, z] = a[x, z] + b[y, z], \quad [z, ax + by] = a[z, x] + b[z, y]$ for all scalars a, b in F and all elements x, y, z in $\mathfrak{g}$.
- Anticommutativity, or skew-symmetry: $[x, y] = -[y, x]$ for all elements x, y in $\mathfrak{g}$. When F is a field of characteristic two, one has to impose the stronger condition $[x, x] = 0$ for all x in $\mathfrak{g}$.
- The Jacobi identity: $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0 \quad$ for all x, y, z in $\mathfrak{g}$.

A.1.3 Exponential Map

The Lie algebra determines the local structure of the Lie group via the *exponential map*. The exponential map is a map from the Lie algebra $\mathfrak{g}$ to the Lie group G

$$\exp\colon \mathfrak{g} \to G \tag{A.3}$$

given by $\exp(X) = \gamma(1)$ where $\gamma\colon \mathbb{R} \to G$ is the unique one-parameter subgroup of G whose tangent vector at the identity is equal to X. It follows easily from the chain rule that $\exp(tX) = \gamma(t)$. The map γ may be constructed as the integral curve of either the right- or left-invariant vector field associated with X. That the integral curve exists for all real parameters follows by right- or left-translating the solution near zero.

The exponential map $\exp\colon \mathfrak{g} \to G$ is a smooth map. Its derivative at the identity, $\exp_*\colon \mathfrak{g} \to \mathfrak{g}$, is the identity map (with the usual identifications). The exponential map, therefore, restricts to a diffeomorphism from some neighborhood of 0 in $\mathfrak{g}$ to a neighborhood of 1 in G.

In Riemannian geometry, the exponential map is a map from a subset of the tangent space at a point P of a Riemannian manifold M, T_PM, to M itself. For a vector $v \in T_PM$, there is a unique geodesic γ_v satisfying $\gamma_v(0) = P$ such that the tangent vector $\gamma_v'(0) = v$. Then the corresponding exponential map is defined by $\exp_P(v) = \gamma_v(1)$. In general, the exponential map really is only locally defined, that is, it only takes a small neighborhood of the origin at T_PM, to a neighborhood of P in the manifold (this is simply due to the fact that it relies on the theorem on existence and uniqueness of ODEs which is local in nature).

If G is a matrix Lie group, then the exponential map coincides with the matrix exponential and is given by the ordinary series expansion:

$$\exp X = \sum_{k=0}^{\infty} \frac{X^k}{k!} = I + X + \frac{1}{2}X^2 + \frac{1}{6}X^3 + \cdots . \tag{A.4}$$

A.2 Introduction to the Geometry of Stiefel and Grassmann Manifolds

A.2.1 The Unitary Group $U(T)$

The unitary group $U(T)$ is the group of unitary matrices

$$U(T) := \{\mathbf{Q} \in \mathbb{C}^{T\times T} | \mathbf{Q}^H\mathbf{Q} = \mathbf{I}_T\}, \tag{A.5}$$

where $\mathbf{I}_T$ is the $T \times T$ identity matrix. The unitary group $U(T)$ is a compact (closed and bounded) Lie group and thus a differentiable manifold of dimension T^2.

An equation defining tangents to the unitary group at a point $\mathbf{Q}$ is easily obtained by differentiating $\mathbf{Q}^{\mathrm{H}}\mathbf{Q} = \mathbf{I}_T$, yielding

$$\mathbf{Q}^{\mathrm{H}}\boldsymbol{\Delta} + \boldsymbol{\Delta}^{\mathrm{H}}\mathbf{Q} = 0, \tag{A.6}$$

i.e., $\mathbf{Q}^{\mathrm{H}}\boldsymbol{\Delta}$ is skew-hermitian. This condition imposes $T + T(T-1) = T^2$ constraints on the matrix, or equivalently, the vector space of all tangent vectors at $\mathbf{Q}$ has dimension (real) $2T^2 - T^2 = T^2$. This is expected, since the dimension of the tangent space has to coincide with the dimension of the unitary group $U(T)$.

Since the Lie algebra of the unitary group, $\mathfrak{u}(T)$, represents the tangent space at the identity element $\mathbf{Q} = \mathbf{I}_T$, it consists of skew-hermittian matrices

$$\mathfrak{u}(T) := \{\boldsymbol{\Delta} \in \mathbb{C}^{T\times T} | \boldsymbol{\Delta} = -\boldsymbol{\Delta}^{\mathrm{H}}\}. \tag{A.7}$$

A.2.2 Stiefel Manifolds

The (complex) Stiefel manifold $V_{M,T}^{\mathbb{C}}$ is the set of M orthonormal vectors in $\mathbb{C}^T$

$$V_{M,T}^{\mathbb{C}} := \{\boldsymbol{\Phi} \in \mathbb{C}^{T\times M} | \boldsymbol{\Phi}^H\boldsymbol{\Phi} = \mathbf{I}_M\}. \tag{A.8}$$

The unitary group $U(T)$ acts transitively on $V_{M,T}^{\mathbb{C}}$, i.e. transforms elements of $V_{M,T}^{\mathbb{C}}$ into each other. The isotropy group (subgroup of $U(T)$ which fixes elements of $V_{M,T}^{\mathbb{C}}$) consists of matrices of the form

$$H = \begin{pmatrix} \mathbf{I}_M & \mathbf{0} \\ \mathbf{0} & U(M) \end{pmatrix}. \tag{A.9}$$

Transitive group action implies that there is only one group orbit, so $V_{M,T}^{\mathbb{C}}$ is isomorphic to the quotient space $U(T)/H$, which justifies the coset representation

$$V_{M,T}^{\mathbb{C}} \cong U(T) / \begin{pmatrix} \mathbf{I}_M & \mathbf{0} \\ \mathbf{0} & U(T-M) \end{pmatrix}. \tag{A.10}$$

With this, the Stiefel manifold $V_{M,T}^{\mathbb{C}}$ carries the structure of a $U(T)$-homogeneous space. Each element of the Stiefel manifold $V_{M,T}^{\mathbb{C}}$ is thus an equivalent class of the unitary group $U(T)$, i.e. a subset of unitary matrices whose first M columns are the same.

Let $\mathfrak{h} \subset \mathfrak{u}(T)$ be the Lie algebra of $H \subset U(T)$. Then $\mathfrak{h}$ consists of matrices of the form

$$\mathbf{X}' = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}, \mathbf{C} \in \mathfrak{u}(T-M) \tag{A.11}$$

$\mathfrak{u}(T)$ can thus be decomposed into a "vertical" and "horizontal" tangent space, $\mathfrak{u}(T) = \mathfrak{h} \cup \mathfrak{h}^{\perp}$. Tangents of $V_{M,T}^{\mathbb{C}}$ at $\mathbf{I}_{T,M}$ are provided by the horizontal space, $\mathfrak{h}^{\perp}$ and have the form

$$\mathbf{X} = \begin{pmatrix} \mathbf{A} & -\mathbf{B}^H \\ \mathbf{B} & \mathbf{0} \end{pmatrix}, \ \mathbf{A} \in \mathfrak{u}(M), \ \mathbf{B} \in \mathbb{C}^{(T-M)\times M}. \tag{A.12}$$

With this identification, $V_{M,T}^{\mathbb{C}}$ is $U(T)$-normal homogeneous.

A canonical distance measure/metric r^V can be defined on $V_{M,T}^{\mathbb{C}}$ by using the space of tangents at the identity. For $\Omega = \exp(X) \cdot I_{T,M}$, $r^V(\Omega, I_{T,M}) = \frac{1}{\sqrt{2}}\|X\|_F$. For $\mathbf{\Phi}, \mathbf{\Psi} \in V_{M,T}^{\mathbb{C}}$, $r^V(\mathbf{\Phi}, \mathbf{\Psi}) = r^V(\tilde{\mathbf{\Psi}}^{-1}\mathbf{\Phi}, I_{T,M})$, because a left multiplication by an unitary matrix is an isometric transformation.

A topological ("chordal") metric on the Stiefel manifold $V_{M,T}^{\mathbb{C}}$ can be defined from the canonical embedding of $V_{M,T}^{\mathbb{C}}$ in the vector space $(\mathbb{C}^{T\times M}, \langle\cdot,\cdot\rangle)$

$$d^V(\mathbf{\Phi}, \mathbf{\Psi}) = \|\mathbf{\Phi} - \mathbf{\Psi}\|_F. \tag{A.13}$$

A.2.3 Grassmann Manifolds

The (complex) Grassmann manifold $G_{n_T,T}^{\mathbb{C}}$ is the set of all n_T-dimensional linear subspaces of $\mathbb{C}^T$. Having introduced the Stiefel manifold, the (complex) Grassmann manifold is formally defined as

$$G_{M,T}^{\mathbb{C}} := \{\langle\mathbf{\Phi}\rangle | \mathbf{\Phi} \in V_{M,T}^{\mathbb{C}}\}. \tag{A.14}$$

In order to distinguish between an element of the Grassmann manifold and the particular matrix representation, the notation $\langle\mathbf{\Phi}\rangle$ should be used to denote the element (subspace) of the Grassmann manifold spanned by the columns of $\mathbf{\Phi}$.

The unitary group $U(T)$ acts transitively on $G^{\mathbb{C}}_{M,T}$ with isotropy group

$$H = \begin{pmatrix} U(M) & \mathbf{0} \\ \mathbf{0} & U(T-M) \end{pmatrix}. \tag{A.15}$$

Each element of the Grassmann manifold can be considered as an equivalent class of the Unitary group $U(T)$, i.e. the subset of unitary matrices whose columns span the same subspace. This is equivalent to the coset representation

$$G^{\mathbb{C}}_{M,T} \cong U(T) / \begin{pmatrix} U(M) & \mathbf{0} \\ \mathbf{0} & U(T-M) \end{pmatrix}. \tag{A.16}$$

The Lie algebra $\mathfrak{h}$ of the isotropy group $H \subset U(T)$ consists of matrices of the form

$$\mathbf{X}' = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & C \end{pmatrix}, \mathbf{A} \in \mathfrak{u}(M), \mathbf{C} \in \mathfrak{u}(T-M) \tag{A.17}$$

$\mathfrak{u}(T)$ can thus be decomposed into a vertical and horizontal tangent space, $\mathfrak{u}(T) = \mathfrak{h} \cup \mathfrak{h}^{\perp}$. Tangents of $G^{\mathbb{C}}_{M,T}$ at $\mathbf{I}_{T,M}$ are provided by the "horizontal" space, $\mathfrak{h}^{\perp}$ and have the form

$$\mathbf{X} = \begin{pmatrix} \mathbf{0} & -\mathbf{B}^{\mathrm{H}} \\ \mathbf{B} & \mathbf{0} \end{pmatrix}, \ \mathbf{A} \in \mathfrak{u}(n_T), \ \mathbf{B} \in \mathbb{C}^{(T-M)\times M}. \tag{A.18}$$

With this identification, $G^{\mathbb{C}}_{M,T}$ is $U(T)$-normal homogeneous.

A canonical distance measure/metric r^G can be defined on $G^{\mathbb{C}}_{M,T}$ by using the space of tangents at the identity. For $\langle \mathbf{\Omega} \rangle = \langle \exp(\mathbf{X}) \cdot \mathbf{I}_{T,M} \rangle$,

$$r^G(\mathbf{\Omega}\rangle, \langle \mathbf{I}_{T,M} >) = \frac{1}{\sqrt{2}} \|\mathbf{X}\|_F. \tag{A.19}$$

For $\langle \mathbf{\Phi} \rangle, \langle \mathbf{\Psi} \rangle \in G^{\mathbb{C}}_{M,T}$, $r^G(\langle \mathbf{\Phi} \rangle, \langle \mathbf{\Psi} \rangle) = r^G(\langle \tilde{\mathbf{\Psi}}^{-1} \mathbf{\Phi}, \langle \mathbf{I}_{T,M} \rangle)$, because a left multiplication by an unitary matrix is an isometric transformation ($G^{\mathbb{C}}_{M,T}$ is $U(T)$-normal homogeneous).

There is an embedding of the Grassmann manifold $G^{\mathbb{C}}_{M,T}$ in Euclidean space which is a result of associating elements of the Grassmann manifold with their projection matrices. For $\langle \mathbf{Phi} \rangle \in G^{\mathbb{C}}_{M,T}$, there is an associate orthogonal projection map from $\mathbb{C}^T$ to $\langle \mathbf{\Phi} \rangle$.

$$P_{\mathbf{\Phi}} := \mathbf{\Phi}\mathbf{\Phi}^{\mathrm{H}}, : \mathbb{C}^T \rightarrow \langle \mathbf{\Phi} \rangle. \tag{A.20}$$

$P_{\mathbf{\Phi}}$ is idempotent and hermittian and $\mathrm{tr}(P_{\mathbf{\Phi}}) = M$. Thus $P_{\mathbf{\Phi}}$ lies in a space of real dimension $T^2 - 1$. Additionally, $\|P_{\mathbf{\Phi}} - \frac{T}{M}\mathbf{I}_T\|_F = \sqrt{M(T-M)/T}$ which justifies the embedding

$$G^{\mathbb{C}}_{M,T} \rightarrow S^{T^2-2}(\sqrt{M(T-M)/T}) \subset R^{T^2-1}, \langle \mathbf{\Phi} \rangle \rightarrow P_{\mathbf{\Phi}} - \frac{T}{M}\mathbf{I}_T. \tag{A.21}$$

This justifies the definition of a topological "chordal" metric on the Grassmann manifold

$$d^G(\langle \mathbf{\Phi} \rangle, \langle \mathbf{\Psi} \rangle) = \frac{1}{\sqrt{2}} \| P_{\mathbf{\Phi}} - P_{\mathbf{\Psi}} \|_F \tag{A.22}$$

A.2.4 Parametrization of Stiefel and Grassmann Manifolds

The exponential map provides connection between the tangent space at the identity I_{T,n_T} and the Grassmann manifold, [6]. Elements of the Grassmann manifold are obtained as [10]

$$\Phi = \exp(\mathbf{X}) \cdot \mathbf{I}_{T,M} \in G^{\mathbb{C}}_{M,T}. \tag{A.23}$$

The exponential map is, however, computationally inefficient. Fortunately, the representation of the tangents in the form (equation tangents), provides efficient computation of the exponential map. Given the tangent (equation tangents), the thin singular value decomposition of $\mathbf{B}, \mathbf{B} \in \mathbb{C}^{(T-M)\times M}$, reads [16], [10]

$$\mathbf{B} = \mathbf{V}\mathbf{\Sigma}\mathbf{W}^{\mathrm{H}}, \tag{A.24}$$

where $V \in \mathbb{C}^{(T-M)\times M}$ and has orthonormal columns, $\mathbf{\Sigma}$ is the matrix of singular values of $\mathbf{B}$ in decreasing order, and $\mathbf{W} \in U(M)$. If we denote $\mathbf{\Phi} = \exp(\mathbf{X})\mathbf{I}_{M,T}$, it can be shown that [16], [10]

$$\mathbf{\Phi} = \begin{pmatrix} \mathbf{W}\mathbf{C}\mathbf{W}^{\mathrm{H}} \\ \mathbf{V}\mathbf{S}\mathbf{W}^{\mathrm{H}} \end{pmatrix}, \tag{A.25}$$

where $\mathbf{C} = \cos(\mathbf{\Sigma})$ and $\mathbf{S} = \sin(\mathbf{\Sigma})$.

A.3 Proof of Optimum Power Allocation

According to (4.15), the PEP is up-bounded (when R is large) by:

$$P(\mathbf{\Phi}, \mathbf{\Psi}) \leq \frac{1}{2} \det \left[\mathbf{I}_{M'} + \varrho(\mathbf{I}_{M'} - \mathbf{\Delta}^{\mathrm{H}}\mathbf{\Delta}) \right]^{-N},$$

which is minimized when

$$\varrho = \frac{(\rho \frac{T_2}{M'})^2}{4(1 + \rho \frac{T_2}{M'})}, \tag{A.26}$$

is maximized.

According to the system model

$$\mathbf{X} = \sqrt{\frac{P_1 P_2 T_2}{(P_1 + 1)M}} \mathbf{\Phi}\mathbf{H} + \mathbf{W},$$

where

$$\mathbf{W} = \begin{pmatrix} \sqrt{\frac{P_2T_2}{(P_1+1)T_1}}\sum_{i=1}^{R} g_{i1}\hat{\mathbf{A}}_i\hat{\mathbf{v}}_i + \mathbf{w}_1 \\ \vdots \\ \sqrt{\frac{P_2T_2}{(P_1+1)T_1}}\sum_{i=1}^{R} g_{iN}\hat{\mathbf{A}}_i\hat{\mathbf{v}}_i + \mathbf{w}_N \end{pmatrix}^{\mathrm{T}},$$

The average signal to noise ratio at each antenna is calculated as

$$\begin{aligned} \rho &= \frac{\frac{P_1P_2T_2}{P_1+1}\cdot R}{\frac{P_2T_2}{(P_1+1)T_1}\cdot T_1R + T_2} \\ &= \frac{P_1P_2R}{P_2R+P_1+1}. \end{aligned}$$

Therefore ϱ becomes

$$\begin{aligned} \varrho &= \frac{(\rho\frac{T_2}{M'})^2}{4(1+\rho\frac{T_2}{M'})} = \frac{(\rho\frac{T_2}{MR})^2}{4(1+\rho\frac{T_2}{MR})} \\ &= \frac{\rho^2T_2^2}{4M^2R^2+4\rho T_2MR} = \frac{T_2^2}{4MR}\cdot\frac{\rho^2}{MR+\rho T_2} \\ &= \frac{T_2^2}{4MR}\cdot\frac{\frac{P_1^2P_2^2R^2}{(P_2R+P_1+1)^2}}{MR+\frac{P_1P_2R}{P_2R+P_1+1}\cdot T_2} \\ &= \frac{T_2^2}{4MR}\cdot\frac{P_1^2P_2^2R}{[P_2R+P_1+1][(P_2R+P_1+1)M+P_1P_2T_2]}. \end{aligned} \tag{A.27}$$

Assume $P_2 = \alpha P_1$, according to the power condition

$$P = \frac{P_1T_1}{T_1+T_2} + R\frac{P_2T_2}{T_1+T_2}.$$

we get

$$P_1 = \frac{T_1+T_2}{T_1+\alpha R\cdot T_2}P, \quad P_2 = \frac{\alpha(T_1+T_2)}{T_1+\alpha R\cdot T_2}P. \tag{A.28}$$

(A.27) becomes

$$\varrho = \frac{T_2^2\cdot P}{4MR}\cdot\frac{\left(\frac{T_1+T_2}{T_1+\alpha RT_2}\right)^4\cdot\alpha^2R}{\left[\frac{\alpha R(T_1+T_2)}{T_1+\alpha RT_2}+\frac{T_1+T_2}{T_1+\alpha RT_2}+\frac{1}{P}\right]\left[\left(\frac{\alpha R(T_1+T_2)}{T_1+\alpha RT_2}+\frac{T_1+T_2}{T_1+\alpha RT_2}+\frac{1}{P}\right)\cdot\frac{M}{P}+\frac{(T_1+T_2)^2}{(T_1+\alpha RT_2)^2}\cdot\alpha T_2\right]}.$$

When P is large, the equation can be simplified

$$\begin{aligned}\varrho &= \frac{T_2^2 \cdot P}{4MR} \cdot \frac{\left(\frac{T_1+T_2}{T_1+\alpha R T_2}\right)^4 \cdot \alpha^2 R}{\left[\frac{\alpha R(T_1+T_2)}{T_1+\alpha R T_2} + \frac{T_1+T_2}{T_1+\alpha R T_2}\right] \cdot \frac{(T_1+T_2)^2}{(T_1+\alpha R T_2)^2} \cdot \alpha T_2} \\ &= \frac{T_2 \cdot P}{4M} \cdot \frac{\alpha(T_1+T_2)}{(\alpha R+1)(T_1+\alpha R T_2)}.\end{aligned}$$

When $\varrho' = 0$, and $\varrho'' < 0$, ϱ is maximized

$$\varrho' = \frac{\partial \varrho}{\partial \alpha} = \frac{T_2 P(T_1+T_2)}{4M(\alpha R+1)^2(T_1+\alpha R T_2)^2} \cdot (T_1 - \alpha^2 R^2 T_2) = 0$$

$$\Rightarrow \alpha = \sqrt{\frac{T_1}{T_2}} \cdot \frac{1}{R} \tag{A.29}$$

Under this condition, we get

$$\varrho'' = \frac{\partial^2 \varrho}{\partial \alpha^2} = -\frac{T_2 P(T_1+T_2)}{2M \cdot (\alpha R+1)^3(T_1+\alpha R T_2)^3} \cdot \left(2RT_1T_2\sqrt{\frac{T_1}{T_2}} + RT_1(T_1+T_2)\right) < 0$$

According to A.28 and A.29 we get:

$$P_1 = \frac{T_1+T_2}{T_1+\sqrt{T_1T_2}}P, \quad P_2 = \frac{T_1+T_2}{(T_2+\sqrt{T_1T_2})R}P \tag{A.30}$$

A.4 Product of Two Alamouti Codes

Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two Alamouti codes,

$$\mathcal{B}_1 = \begin{pmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{pmatrix}, \mathcal{B}_2 = \begin{pmatrix} u_1 & -u_2^* \\ u_2 & u_1^* \end{pmatrix}.$$

The Product of the two codes can be written as:

$$\begin{aligned}\mathcal{B} &= \mathcal{B}_1 \times \mathcal{B}_2 = \begin{pmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{pmatrix} \cdot \begin{pmatrix} u_1 & -u_2^* \\ u_2 & u_1^* \end{pmatrix} \\ &= \begin{pmatrix} s_1u_1 - s_2^*u_2 & -s_1u_2^* - s_2^*u_1^* \\ s_1^*u_2 + s_2u_1 & s_1^*u_1^* - s_2u_2^* \end{pmatrix}.\end{aligned} \tag{A.31}$$

If we denote $\omega_1 = s_1u_1 - s_2^*u_2$, and $\omega_2 = s_1^*u_2 + s_2u_1$, (A.31) can be written as:

$$\mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2 = \begin{pmatrix} \omega_1 & -\omega_2^* \\ \omega_2 & \omega_1^* \end{pmatrix}.$$

So, the product of the two Alamouti codes is still an Alamouti code.

A.5 Product of Two Sp(2) Codes

Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two Sp(2) codes,

$$\mathcal{B}_1 = \begin{pmatrix} s_1 & -s_2^* & -s_3^* & s_4 \\ s_2 & s_1^* & -s_4^* & -s_3 \\ s_3 & -s_4^* & s_1^* & -s_2 \\ s_4 & s_3^* & s_2^* & s_1 \end{pmatrix},$$

$$\mathcal{B}_2 = \begin{pmatrix} u_1 & -u_2^* & -u_3^* & u_4 \\ u_2 & u_1^* & -u_4^* & -u_3 \\ u_3 & -u_4^* & u_1^* & -u_2 \\ u_4 & u_3^* & u_2^* & u_1 \end{pmatrix},$$

where s_i, u_i $(i = 1, 2, 3, 4)$ are in the form of (4.29).

The product of the two code can be written as:

$$\begin{aligned} \mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2 &= \begin{pmatrix} s_1 & -s_2^* & -s_3^* & s_4 \\ s_2 & s_1^* & -s_4^* & -s_3 \\ s_3 & -s_4^* & s_1^* & -s_2 \\ s_4 & s_3^* & s_2^* & s_1 \end{pmatrix} \cdot \begin{pmatrix} u_1 & -u_2^* & -u_3^* & u_4 \\ u_2 & u_1^* & -u_4^* & -u_3 \\ u_3 & -u_4^* & u_1^* & -u_2 \\ u_4 & u_3^* & u_2^* & u_1 \end{pmatrix} \\ &= \begin{pmatrix} \omega_1 & -\omega_2^* & -\omega_3^* & \omega_4 \\ \omega_2 & \omega_1^* & -\omega_4^* & -\omega_3 \\ \omega_3 & -\omega_4^* & \omega_1^* & -\omega_2 \\ \omega_4 & \omega_3^* & \omega_2^* & \omega_1 \end{pmatrix}, \end{aligned}$$

where

$$\begin{aligned} \omega_1 &= s_1u_1 - s_2^*u_2 - s_3^*u_4 + s_4u_4, \\ \omega_2 &= s_2u_1 + s_1^*u_2 - s_4^*u_4 + s_3u_4, \\ \omega_3 &= s_3u_1 - s_4^*u_2 + s_1^*u_4 + s_2u_4, \\ \omega_4 &= s_4u_1 + s_3^*u_2 + s_2^*u_4 + s_1u_4. \end{aligned}$$

So, the product of two Sp(2) codes is still an Sp(2) code.

Appendix $\mathcal{B}$

Notation and Abbreviations

Notation

i	Imaginary unit, $\sqrt{\mathrm{j}} = -1$.
$(\cdot)^*$	Complex conjugate.
$(\cdot)^{\mathrm{H}}$	Complex conjugate transposition.
$(\cdot)^{\mathrm{T}}$	Transposition.
$\arg\{\cdot\}$	Argument of a function.
$\arg\max_x f(x)$	Denotes value of x that maximizes $f(x)$.
$\mathrm{E}[\cdot]$	Expected value.
$\Im(\cdot)$	Imaginary part of a complex value.
$\max\{\cdot,\cdot\}$	Maximum of two values.
$\min\{\cdot,\cdot\}$	Minimum of two values.
$p(\cdot)$	Probability density function.
$\Re(\cdot)$	Real part of a complex value.
$\mathbb{R}$	Set of real numbers.

List of Abbreviations

AF	Amplify-and-Forward
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BLER	Block Error Rate
BPSK	Binary Phase Shift Keying
$CN(m, \sigma^2)$	Complex Gaussian distribution with m mean and σ^2 variance
DF	Decode-and-Forward
DPSK	Differential Phase-Shift Keying
DSTC	Distributed Space-Time Coding
LD	Linear Dispersion
MAC	Medium Access Control
MIMO	Multiple Input Multiple Output
ML	Maximum-Likelihood
ML	Maximum Likelihood
OFDM	Orthogonal Frequency Division Multiplexing
PAM	Pulse Amplitude Modulation
PEP	Pairwise Error Probability
PSK	Phase-Shift Keying
SNR	Signal-to-Noise Ratio
STBC	Space-Time Block Code
STBCs	Space-Time Block Codes
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying

List of Tables

List of Figures

Bibliography

[1] I. Kammoun A. M. Cipriano and J.-C. Belfiore. Simplified decoding for some non-coherent codes over the grassmannian. In *Proc. 2005 IEEE International Conference on Communications, (ICC 2005)*, Istanbul, September 2005.

[2] S. M. Alamouti. A simple transmitter diversity scheme for wireless communications. *IEEE Jour. Selec. Areas Commun.*, pages 1451–1458, October 1998.

[3] K. Azarian, H. El-Gamal, and P. Schniter. On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels. *IEEE Trans. Inform. Theory*, 51(12):4152, December 2005.

[4] A. Barg and D. Yu. Nogin. Bounds on packings of spheres in the grassmannian manifold. ieee trans. inform. theory. *IEEE Trans. Inform. Theory*, 48(9):2450–2454, September 2002.

[5] J.-C. Belfiore, G. Rekaya, and E. Viterbo. The golden code: A 2x2 full-rate spacetime code with nonvanishing determinants. *IEEE Trans. Inform. Theory*, 51(4):1432–1436, April 2005.

[6] W. M. Boothby. *An Introduction to Differential Manifolds and Riemannian Geometry*. CA: Academic, San Diego, 1986.

[7] J. Conway, R. Hardin, and N. Sloane. Packing lines, planes etc. : Packings in grassmannian spaces. *Exp. Mathematics*, 5:139–159, 1996.

[8] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley, 2006.

[9] T. Cui, T. Ho, and A. Nallanathan. Distributed space-time coding for two-way wireless relay networks. In *Proc. IEEE International Conference on Communications (ICC 2008)*, Beijing, 2008.

[10] A. Edelman, T. Arias, and S. Smith. The geometry og algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 1998.

[11] M. Gastpar and M. Vetterli. On the capacity of wireless networks: The relay case. In *Proc. 2002 IEEE InfoCom, (InfoCom 2002)*, 2002.

[12] R. H. Gohary and T. N. Davidson. Non-coherent communication: Grassmannian constellations and efficient detection. *IEEE Trans. Inform. Theory*, June 2005.

[13] R. H. Gohary and T. N. Davidson. On efficient non-coherent detection of grassmann constellations. In *Proc. ISIT IEEE International Symposium on Information Theory, (ISIT 2005)*, Adelaide, September 2005.

[14] J. Hamkins and K. Zeger. Asymptotically dense spherical codes-part 1: wrapped spherical codes. *IEEE Trans. Inform. Theory*, 43(6), November 1997.

[15] O. Henkel. Space frequency codes from spherical codes. In *Proc. 2005 IEEE International Symposium on Information Theory (ISIT 2005)*, pages 1305–1309, Adelaide, Australia, September 2005.

[16] O. Henkel. Sphere packing bounds in the grassmann and stiefel manifolds. *IEEE Trans. Inform. Theory*, 51(10):3445–3456, October 2005.

[17] O. Henkel. Geometrical relations between space-time block code designs and complexity reduction. *IEEE Trans. Inform. Theory*, 52(12):5324–5335, December 2006.

[18] O. Henkel. Space time codes from permutation codes. In *Proc. 2006 Globecom IEEE Globecom (Globecom 2006)*, San Francisco, November-December 2006.

[19] B. M. Hochwald and T-L. Marzetta. Unitary space-time modulation for multiple-antenna communication in rayleigh flat fading. *IEEE Trans. Inform. Theory*, 46(1):543–565, March 2000.

[20] B. Hughes. Differential space-time modulation. *IEEE Trans. Inform. Theory*, 46(11):2567–2578, November 2000.

[21] H. Jafarkhani. A quasi-orthogonal space-time block code. *IEEE Transactions on Communications Letters*, 49(1):1–4, January 2001.

[22] W. C. Jakes. *Microwave Mobile Communications*. IEEE Press, Piscataway, NJ, 1993.

[23] Y. Jing and B. Hassibi. Design of fully diverse multiple-antenna codes based on sp(2). *IEEE Trans. Inform. Theory*, 50:2639–2656, November 2004.

[24] Y. Jing and B. Hassibi. Distributed space-time coding in wireless relay networks. *IEEE Trans. Wireless Commun.*, 5:3524–3536, December 2006.

[25] Y. Jing and H. Jafarkhani. Distributed differential space-time coding for wireless relay networks. *submitted to IEEE Trans. Commun.*, 2008.

[26] I. Kammoun and J.-C. Belfiore. A new family of grassmann space-time codes for non-coherent mimo systems. *IEEE Commun. Letters*, 7(11):528–531, November 2003.

[27] S. Karmakar and B. S. Rajan. Minimum-decoding-complexity maximum-rate space-time block codes from clifford algebras. In *Proc. IEEE International Symposium on Information Theory (ISIT 2006)*, pages 788–792, Seattle, July 2006.

[28] J. N. Laneman and G. W. Wornell. Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks. *IEEE Trans. Inform. Theory*, 49(10):2415–2425, October 2003.

[29] D. J. Love and R. W. Heath. Limited feedback unitary precoding for spatial multiplexing systems. *IEEE Trans. Inform. Theory*, 51(8):2967–2976, August 2005.

[30] A. Tewfik M. O. Damen and J.-C. Belfiore. A construction of a space-time code based on number theory. *IEEE Trans. Inform. Theory*, 48(3):753–760, March 2002.

[31] T. L. Marzetta and B. M. Hochwald. Capacity of a mobile multiple-antenna communication link in rayleigh flat fading environment. *IEEE Trans. Inform. Theory*, 45(1):139–157, January 1999.

[32] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler. Fading relay channels: Performnace limits and space-time signal design. *IEEE JSAC*, 22(6):1099–1109, August 2004.

[33] F. Oggier and B. Hassibi. An algebraic coding scheme for wireless relay networks with multiple antenna nodes. *Submitted IEEE Trans. Sign. Proc.*

[34] C. Pietsch and J. Lindner. On orthogonal space-time codes and packings on the grassmann manifold. In *Proc. 2006 Globecom IEEE Globecom (Globecom 2006)*, San Francisco, November-December 2006.

[35] P. Popovski and H. Yomo. Wireless network coding by amplify-and-forward for bi-directional traffic flows. *IEEE Commun. Letters*, 11(1):16–18, January 2007.

[36] V. V. Prasolov. *Problems and theorems in linear algebra*. American Mathematical Society, 1994.

[37] J. Proakis. *Digital Communications*. McGraw-Hill, 2001.

[38] G. S. Rajan and B. S. Rajan. Signal set design for full-diversity low-decoding-complexity differential scaled-unitary stbcs. In *Proc. IEEE International Symposium on Information Theory (ISIT 2007)*, pages 1616–1620, Nice, June 2007.

[39] B. Rankov and A. Wittneben. Spectral efficient protocols for half-duplex fading relay channels. *IEEE Journal on Selected Areas in Communications*, 25(2):379–389, February 2007.

[40] T. S. Rapaport. *Wireless communications: principle and practice*. Prentice Hall, 2002.

[41] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity-part 1: System description. *IEEE Trans. Commun.*, 51(11):1927–1938, November 2003.

[42] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity-part 2: Implementation aspects and performance analysis. *IEEE Trans. Commun.*, 51(11):1939–1948, November 2003.

[43] C. E. Shannon. Two-way communication channels. In *Proc. Fourth Berkeley Symp. Math. Stat. Prob.*, pages 611–644, 1961.

[44] N. J. A. Sloane. Packing planes in four dimensions and other misteries. *Journal on Algebraic Combinatorics and Related Topics*, 1997.

[45] G. L. Stuber. *Principles of mobile communications*. Kluwer Academic Press, 2000.

[46] V. Tarokh, N. Seshadri, and A. R. Calderbank. Space-time codes for high data rate wireless communication: performance criterion and code construction. *IEEE Trans. Inform. Theory*, 44:744–765, 1998.

[47] J. A. Tropp, I. S. Dhillon, R. W. Heath, and T. Strohmer. Designing structured tight frames via an alternating projection method. *ICES Report 03-50, The University of Texas at Austin*, December 2003.

[48] D. Tse and P. Viswanath. *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[49] Z. Utkovski, P. Chen, and J. Lindner. Some geometric methods for construction of space-time codes in grassmann manifolds. In *Proc. ALLERTON Conference on Communications, Constrol and Computing, (ALLERTON 2008)*, Monticello, September 2008.

[50] Z. Utkovski, Y. Cheng, and J. Lindner. A non-coherent af scheme for two-way wireless relay networks based on packings in grassmann manifolds. In *Accepted at ISITA IEEE International Symposium on Information Theory and Applications, (ISITA 2010)*, 2010.

[51] Z. Utkovski, W. Li, and J. Lindner. Distributed non-coherent grassmann space-time codes for wireles relay networks. In *Proc. ICC IEEE International Conference on Communications, (ICC 2008)*, Beijing, May 2008.

[52] Z. Utkovski, W. Li, and J. Lindner. Non-coherent distributed space-time coding scheme for wireles relay networks. In *Proc. SCC ITG/IEEE Source and Channel Coding Conference, (SCC 2008)*, Ulm, January 2008.

[53] Z. Utkovski and J. Lindner. On the construction of non-coherent space time codes from high-dimensional spherical codes. In *Proc. 2006 IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA 2006)*, Manaus, Brasil, August 2006.

[54] Z. Utkovski and J. Lindner. Construction of high-dimensional space-time codes in the grassmann and stiefel manifolds. In *Proc. ISCTA International Symposium on Coding Theory and Applications, (ISCTA 2007)*, Ambleside, July 2007.

[55] Z. Utkovski and J. Lindner. A recursive construction of high-dimensional grassmann space-time codes for non-coherent mimo systems. In *Proc. 2007 IEEE International Symposium on Information Theory (ISIT 2007)*, Nice, June 2007.

[56] Z. Utkovski and J. Lindner. Spectral-efficient grassmann codes for non-coherent mimo systems. In *Proc. CISS Conference on Information Sciences and Systems, (CISS 2007)*, Baltimore, March 2007.

[57] Z. Utkovski and J. Lindner. A differential scheme for two-way relaying channels. In *Proc. ISIT IEEE International Symposium on Information Theory, (ISIT 2009)*, Seoul, June 2009.

[58] Z. Utkovski and J. Lindner. A non-coherent af scheme for two-way wireless relay networks. In *Proc. WSA ITG/IEEE Workshop on Smart Antennas, (WSA 2010)*, Bremen, February 2010.

[59] Z. Utkovski, A. Sezgin, and J. Lindner. Non-coherent two-way relaying: Rate bounds for the high snr regime. In *Accepted at ISITA IEEE International Symposium on Information Theory and Applications, (ISITA 2010)*, 2010.

[60] Z. Utkovski and A. Utkovski. A spherical quantizer for gaussian sources. Master's thesis, Chalmers Technical University, March 2004.

[61] Z. Utkovski, A. Utkovski, and T. Eriksson. High-dimensional spherical quantization of gaussian sources. In *Proc. CWIT Canadian Workshop on Information Theory, (CWIT 2005)*, Montreal, June 2005.

[62] R. Vaze and R. W. Heath Jr. Capacity scaling for mimo two-way relaying. *submitted to IEEE Trans. Inform. Theory*, 2008.

[63] A. Wittneben. A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation. In *Proc. 1993 IEEE International Conference on Communications, (ICC 1993)*, pages 1630–1634, May 1993.

[64] G. Yammine, Z. Utkovski, and J. Lindner. Performance of sp(n) codes in two-way wireless relay networks. In *Proc. WSA ITG/IEEE Workshop on Smart Antennas, (WSA 2009)*, Berlin, February 2009.

[65] L. Zheng and D. Tse. Communication on the grassmann manifold: a geometric approach to the non-coherent multiple antenna channel. *IEEE Trans. Inform. Theory*, 46:1456–1467, 1999.

[66] L. Zheng and D. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels. *IEEE Trans. Inform. Theory*, 49:1073–1096, May 2003.