Constellation Design under Channel Uncertainty

JOCHEN GIESE

KTH Electrical Engineering

Doctoral Thesis
Stockholm, Sweden 2005

TRITA–S3–KT–0518
ISSN 1653–3860
ISRN KTH/KT/R - - 05/18 - - SE

ROYAL INSTITUTE
OF TECHNOLOGY

Constellation Design under Channel Uncertainty

Jochen Giese

TRITA–S3–KT–0518
ISSN 1653-3860
ISRN KTH/KT/R - - 05/18 - - SE
ISBN 91-7178-179-X

Communication Theory
Department of Signals, Sensors and Systems
Royal Institute of Technology
Stockholm, Sweden, 2005

Submitted to the School of Electrical Engineering, Royal Institute of Technology, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Constellation Design under Channel Uncertainty

Communication Theory Group
Department of Signals, Sensors and Systems
School of Electrical Engineering
Royal Institute of Technology (KTH)
SE-100 44 Stockholm, Sweden

Tel. +46 8 790 6000, Fax. +46 8 790 7260
http://www.s3.kth.se

Abstract

The topic of this thesis is signaling design for data transmission through wireless channels between a transmitter and a receiver that can both be equipped with one or more antennas. In particular, the focus is on channels where the propagation coefficients between each transmitter–receiver antenna pair are only partially known or completetly unknown to the receiver and unknown to the transmitter.

A standard signal design approach for this scenario is based on separate training for the acquisition of channel knowledge at the receiver and subsequent error-control coding for data detection over channels that are known or at least approximately known at the receiver. If the number of parameters to estimate in the acquisition phase is high as, e.g., in a frequency-selective multiple-input multiple-output channel, the required amount of training symbols can be substantial. It is therefore of interest to study signaling schemes that minimize the overhead of training or avoid a training sequence altogether.

Several approaches for the design of such schemes are considered in this thesis. Two different design methods are investigated based on a signal representation in the time domain. In the first approach, the symbol alphabet is preselected, the design problem is formulated as an integer optimization problem and solutions are found using simulated annealing. The second design method is targeted towards general complex-valued signaling and applies a constrained gradient-search algorithm. Both approaches result in signaling schemes with excellent detection performance, albeit at the cost of significant complexity requirements.

A third approach is based on a signal representation in the frequency domain. A low-complexity signaling scheme performing differential space–frequency modulation and detection is described, analyzed in detail and evaluated by simulation examples.

The mentioned design approaches assumed that the receiver has no

knowledge about the value of the channel coefficients. However, we also investigate a scenario where the receiver has access to an estimate of the channel coefficients with known error statistics. In the case of a frequency-flat fading channel, a design criterion allowing for a smooth transition between the corresponding criteria for known and unknown channel is derived and used to design signaling schemes matched to the quality of the channel estimate. In particular, a constellation design is proposed that offers a high level of flexibility to accomodate various levels of channel knowledge at the receiver.

Acknowledgments

Looking back at the years working towards this thesis, I feel a deep gratitude towards many people that have contributed in one or more ways.

First and foremost, my advisor Prof. Mikael Skoglund has been the major source of input, constructive criticism and encouragement. His interest, sharp mind, patience and great support have been very helpful during the entire work on this thesis. I feel honoured of having become his first graduate student and am very grateful for the past five years.

Moreover, I feel deep gratitude towards everybody in the Communication Theory and Signal Processing groups for creating the great research environment that I had the privilege to work in. Starting as a newly arrived graduate student from a foreign country, I could hardly have felt more comfortable and welcome. The exciting learning atmosphere created by my colleagues and coworkers together with the efficiency of administrative routines and by now impeccable computer support have not stopped to impress me.

I would also like to thank Prof. Bölcskei from the Swiss Federal Institute of Technology, Zurich, Switzerland, for the opportunity to visit his research group in 2003. Financial support of the AWSI and WINNER projects as well as the Graduate School of Telecommunications is gratefully acknowledged. My special thanks go to Dr Bertrand Hochwald for acting as opponent on this thesis.

Some other experiences are worth mentioning here because they were very helpful for the preparation of this thesis even though they had been planned without any relation to my research work. Being originally from densely populated Germany, I had the opportunity to discover the Swedish way of life together with some special outdoor activities and sports which cannot be enjoyed in this way in my country of origin. My series of discoveries started after two months in Sweden when I was talked

into participating in the cross-country skiing event "Vasaloppet-Öppet Spår" and has not stopped with running through a rather sunny summer night in Lapland this year. One of the key requirements for activities of this kind is perseverance. Training this character trait in spare-time turned out to be extremely helpful in my research work. A very big thank you goes to all my friends, in particular from Stockholm City Triathlon, who supported me during these trainings. It's a great feeling to see the finish line!

Jochen Giese
Stockholm, October 2005

Contents

Chapter 1

Introduction

The presence of telecommunication services in every day life has seen an enormous growth during the last ten to 15 years. The Internet, originally conceived as a military computer network allowing data communication even in the case of a nuclear attack, has by now grown to a world-wide network allowing such diverse services as personal communication, information retrieval, financial services (e.g., banking), shopping, gaming, video-on-demand, TV and radio, flirting and dating, interaction with public authorities (e.g., tax declarations) and many more.

The amazing expansion of the Internet was paralleled in the 1990's by a massive growth of the market for wireless telecommunication services. Caused by sharply decreasing consumer cost, the global system for mobile communication (GSM), the dominating standard for second generation mobile telephony in Europe, was no longer considered a luxury for the general population but became a part of everyday life. By 2004, user penetration in terms of assigned mobile phone numbers per inhabitant has surpassed 100% in Sweden [WÖst05].

The convergence of these two developments, i.e., the combination of wireless reachability with Internet data services "anytime anywhere" is no longer in its infancy. Huge efforts were made to develop modern wireless telecommunication systems such as the universal mobile telecommunication system (UMTS) or various wireless local area network (WLAN) standards which support higher data rates than the GSM system and thereby allow more advanced multimedia communication services. It is a common belief that this trend to ever increasing data rates in mobile communication prevails. Thus, the enabling technologies allowing

this development continue to be the subject of intensive research efforts. This thesis constitutes a contribution to this field, more specifically in the area of signal design for wireless communication. In this introductory chapter, the challenges imposed by the physical reality in wireless communication are summarized in Section 1.1 and a potential technique for increasing data rates is introduced in Section 1.2. A categorization of existing works relating to this technique in Section 1.3 then leads in Section 1.4 to the description of the research problem discussed in this thesis. The contributions and possible solution approaches are then outlined in Section 1.5 followed by a summary of the notation and acronyms used throughout this text.

1.1 The Wireless Radio Channel

In a wireless radio communication system, the transmitter maps the data to be transmitted onto electromagnetic waves that propagate through the space between the transmitter and the receiver. The wave propagation is affected in a number of ways [PP97] imposed by the physical conditions of transmission, or, in more technical terms, the *channel* between the transmitter and the receiver. A communication system designed to recover the data at the receiver therefore requires a sufficiently accurate description of the channel properties.

One main characteristic of wireless communication channels is the possibility of *multipath*, i.e., the signal can propagate from the transmitter to the receiver along a number of different paths. These different paths are caused by signal reflection or refraction of the radio wave by objects in the environment, usually causing time and phase shifts between the waves propagating along different paths. The incoming signals can thus interfere constructively (the different signal powers add up) or destructively (the incoming waves cancel each other). When the environment changes or the transmitter or the receiver are mobile, the multipath environment changes, leading to a change of the shift pattern between different paths. This in turn causes a variation in received signal power called *fading*. Moreover, the channels can become *frequency-selective* if the spread of time shifts between incoming waves is significant in relation to the received signal bandwidth and the receiver may be able to resolve several distinct copies of the incoming signals with different time shifts.

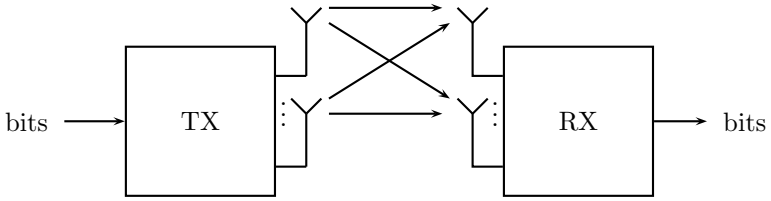


Figure 1.1: Diversity in MIMO systems

1.2 Multiple Antennas

Because of the difficult transmission conditions in the wireless radio communication channel, the supported data rates for reliable communication are rather low in standard single-input single-output (SISO) systems. An immediate conclusion in Shannon's pioneering work on information theory [Sha48] is that reliable communication at higher rates can be possible at the cost of higher signal bandwidth or power. Unfortunately, these two options do not have the potential to meet the need for cheap and easy-to-use mobile communication services. Radio wave spectrum has become an extremely expensive resource. The radio wave spectrum licenses for UMTS covering about 85MHz of bandwidth were sold in Germany for more than € 50 billion [Age00]. Transmitter power radiation is limited by the battery life in mobile terminals; moreover, regulatory authorities impose strict rules for power emission that must not be violated.

A potential way out of this dilemma was illustrated by, e.g., Foschini and Gans [FG98], Telatar [Tel99] and earlier by Winters et al. [WSG94]. They showed that the use of multiple antennas at the transmitter and the receiver side allows for a potential increase in data rates for reliable communication without requiring more spectrum or larger output power. Exploiting the potential of "the spatial dimension" in so-called multiple-input multiple-output (MIMO) systems is still a topic for extensive research.

One of the reasons for the capability to support much higher data rates compared to a single antenna system is *diversity*, see Figure 1.1. A signal transmitted via several antennas reaches the receiver antennas via a number of different propagation paths. Each receiver antenna thereby picks up the signal transmitted by each transmitter antenna. If the an-

tennas are located sufficiently far apart from each other such that the corresponding propagation coefficients are independent, the probability that all paths are bad is significantly reduced. In other words, the probability of having a useless channel is much smaller, which in turn leads to a lower probability of error at the receiver compared to the case with only a single antenna at both the transmitter and the receiver.

Whereas the diversity benefit leading to higher robustness against channel fading is available in systems with multiple antennas at either the transmitter or the receiver (or both, of course), the advantage of *spatial multiplexing* requires multiple antennas on both sides of the communication link. In essence, the multiple antenna channel can then be viewed as a set of parallel single-input single-output *spatial* channels, that together support a much higher data rate compared to single channel. An information-theoretic discussion concerning the exploitation of both spatial multiplexing and diversity and the tradeoff between these effects is available in [ZT03].

1.3 Signaling Schemes for Multiple Antennas

In order to realize the mentioned performance gains of multiple transmitter antennas, a definition of the signaling scheme is required, i.e., a description of the mapping of data bits to time-continuous waveforms that can be radiated via the antennas as well as a description of the receiver structure based on the received waveforms.

This definition is usually done with the help of a time-discrete channel model that summarizes the impact of the analog frontend at the transmitter (i.e., pulse-shaping, upconversion to radio frequency and radiation), wave propagation through the wireless radio channel and the analog frontend at the receiver (reception, downconversion to baseband and filtering together with sampling), see Figure 1.2. The remaining part in the definition of a MIMO communication system is then the description of the *encoder*, i.e., the mapping between data bits and symbols appropriate for the transmission via the time-discrete channel, and the *decoder*, i.e., the mapping between outputs of the discrete channel and bits that were transmitted.

In the following, we present two ways of categorizing existing work about encoder and decoder design for MIMO communication systems and introduce some terminology that will be used throughout this thesis.

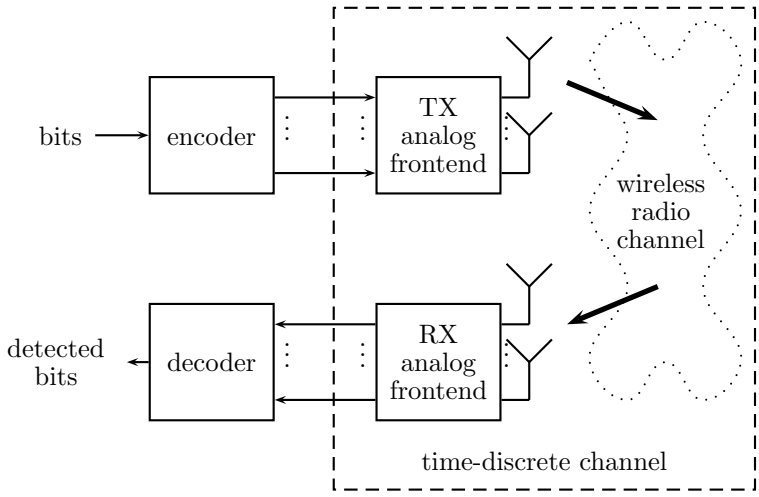


Figure 1.2: Schematic View of Multiple Antenna Communication System

The first way is oriented towards some typical subblocks of the encoder and the second categorization is based on the assumptions concerning the channel model.

1.3.1 Encoder Blocks

The design of the encoder / decoder pair involves trading off a multitude of different requirements that cannot be optimized independently, e.g., the support of high bit rates at very low or vanishing probability of bit error with small decoding delay, low transmission power, small bandwidth and low implementation cost. Defining this tradeoff for a large variety of different application scenarios is in general a tremendous task (see, e.g., the the WINNER project [D2.05,IR204]) and it is by far beyond the scope of this thesis in its totality.

Therefore, the study of the design of the encoder and decoder is usually partitioned into subblocks that are easier to characterize and design. Any such partitioning implies assumptions about the system design and restricts the generality of possible design approaches. Therefore, regrouping separate subblocks together and optimizing them jointly can be beneficial from a performance point of view. As an example, consider the

partitioning of an encoder design in shown in Figure 1.3. In this exam-

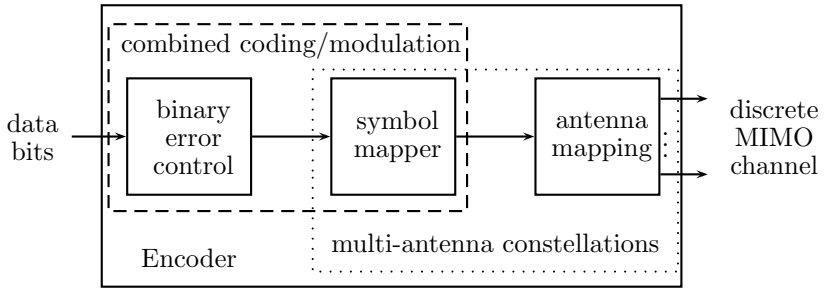


Figure 1.3: An example of encoder design that is split up into three subblocks

ple, the bits that are to be transmitted are first encoded using a binary error-control code. The result is a redundant representation of the data bits, i.e., a representation of the input data with an increased number of bits. The symbol mapper then maps blocks of bits to complex-valued symbols. These symbols are then transmitted via one or more antennas using design rules defining the mapping from complex symbols to several antennas.

The first two parts can also be found in a traditional system with a single transmitter antenna. Frequently, they are designed and optimized jointly, which is sometimes described as a combination of coding and modulation [Ung82, FGL+84, Bos98, CHIW98] leading in general to significant improvements in error protection compared to separate designs. The third block can then be understood as extension of a single-antenna design to a system accommodating multiple transmitter antennas. Often, this mapping is also described using a compound of "coding" as in "space-time coding" [Ala98, TJC99, LS03], "linear dispersion codes" [HH02] or "space-frequency coding" [BP00].

The symbol mapper and the antenna mapping can also be designed jointly. The resulting symbol sequences for each antenna as a direct function of the input bits are termed "constellations" [ARU01, HMR+00] or "codes" as in [JSO02, GS02c] for multiple antennas. The word "coding" is here understood in the general sense of mapping a set of input symbols (bits) to output symbols similar as in the general scheme of Figure 1.2. Depending on the dimensionality of the resulting output symbols in comparison with the input symbols, these multiple antenna constellations

contain significant redundancy for error protection and can therefore also be considered as codes for error protection, taking the output bits of the first block in Figure 1.3 as input bits. Still, these constellations resp. codes are usually designed for transmission during a single fading interval of the channel, i.e., during a period where the characteristics of the channel are (at least approximately) constant. Then, an outer coder as the first block in Figure 1.3 can still be beneficial in the communication system by spreading the data to be transmitted over several fading intervals. We will throughout this thesis consider the general mapping of some input symbols or bits to output symbols as a "code". If this mapping is restricted to symbol sequences transmitted over the time-discrete MIMO channel in a single fading interval, the signal mapping is also referred to as a "constellation" as in Fig. 1.3. We emphasize here that this term refers to a mapping of bits to a potentially large number of symbols. Each of them can itself be restricted to a finite alphabet (as, e.g., the BPSK "constellation") and compared to schemes traditionally termed with "coding," albeit in a single fading interval, see Section 3.2.

1.3.2 Assumptions on the Channel

The second way of categorizing existing literature on MIMO communication systems is related to the assumptions on the underlying channel model and the knowledge both the transmitter and the receiver have about the parameters describing this model. Knowledge about the values of these parameters is frequently termed *channel state information (CSI)*. Standard examples for these parameters are e.g., propagation coefficients associated with the signal electromagnetic wave propagation between each transmitter and receiver antenna pair.

In general, obtaining accurate CSI is more difficult at the transmitter than at the receiver because some kind of data feedback from the receiver to the transmitter is required to obtain CSI at the transmitter. This feedback can be transmitted either over a dedicated control channel or in the form of data transmission in duplex mode. However, this is not always possible. If such a feedback is available, it is usually assumed that the receiver has accurate CSI. Therefore, methods exploiting CSI at the transmitter usually expect some CSI at the receiver. Two possibilities of exploiting perfect transmitter CSI are beamforming and spatial multiplexing. These methods have also been adjusted to the case that the available CSI is imperfect (or "partial"), possibly caused by a feedback link from the receiver with very tight throughput constraints [JS04]. If

no CSI is available at the transmitter, traditional space-time code design methods (see, e.g., the work by Guey et al. [GFBK99], the popular Alamouti scheme described in [Ala98] and the space-time coding papers by Tarokh et al., see e.g., [TSC98, TJC99]) can be used. In these standard references, frequency-flat fading MIMO channels with perfect CSI at the receiver are assumed. Generalizations to frequency-selective channels can be found in, e.g., [LP00, ZG03, LS03] using a single-carrier approach and for designs based on orthogonal frequency division multiplexing (OFDM) in [BP00, LW00, ATNS98].

It is apparent that with an increasing number of antennas, the number of parameters describing the current channel state grows accordingly (see Figure 1.1). This number grows even further if a frequency-selective channel requiring distinct parameters in different frequency bands is used. The high number of parameters then leads to an increased effort in estimating the CSI, which possibly reduces system throughput. Moreover, in a communication system with fast moving mobile users, the coherence time of the channel might be so short that it becomes impractical to acquire CSI [HM00]. Therefore, communication schemes not relying on the assumption of perfect CSI have been of interest in the research on wireless communication systems and they are also the topic of this thesis.

1.4 Signal Design for Receivers without Accurate CSI

Many coding approaches, such as, e.g., Turbo coding [BGT93, BG96] were originally conceived with a frequency-flat additive white Gaussian noise (AWGN) SISO channel in mind. Such a channel can be characterized by a single parameter, the signal to noise ratio (SNR). If such a coding scheme has to work over frequency-selective channels with possibly multiple antennas where the individual path gains for the different antenna pairs are unknown to the receiver over the entire frequency band, the much more complicated nature of the channel has to be taken into account. An immediate conceptual approach to do so is to try to estimate the characteristics of the channel (i.e., the CSI) by means of pilot sequences [ATV02] and then to cancel the effects of the channel using an equalizer [And99] such that the channel behaves similarly to an AWGN channel from the encoder and decoder point of view. A schematic view of such a system is given in Figure 1.4.

With the switch in position 1, a sequence of symbols known to trans-

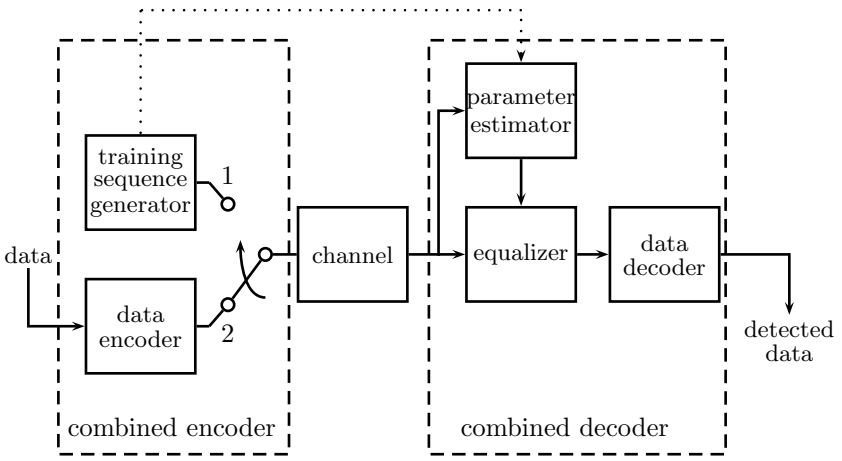


Figure 1.4: Separate coding and detection for channel estimation and data transmission vs. combined coding and decoding

mitter and receiver (training symbols) is transmitted to the receiver via the channel. The receiver then performs an estimation of the unknown channel parameters (resulting phase and amplitude of the incoming waves). Those coefficients are then used in the data transmission phase (switch in position 2) when standard error-control coding is applied on the transmitter side. The receiver uses the knowledge of the channel coefficients in the decoder, possibly by means of an equalizer [Pro95] or directly in the decoding phase [CAC01]. Alternatives to the use of pilot sequences are *blind* approaches [TP98] for unknown-input-known-output system identification or *semi-blind* methods [GL97] which use a combination of pilot symbols and data in order to produce channel estimates.

The previous approaches are based on the assumption that the signal design to transmit data is fixed and channel estimation is a necessary step to allow the adaption of existing schemes to the more complicated communication channel. In sharp contrast, the approach pursued in this thesis is to consider the combination of training data (i.e., pure redundancy) and error-control coded data (i.e., data and redundancy) as one *joint* signal set which is optimized for communication over an unknown channel. In other words, the redundancy for error protection is not selected independently of the training redundancy, but both are optimized jointly for the problem at hand. The training sequence design and the

error control code in Figure 1.4 thus joined in a *combined encoder*. Similarly, the different steps in standard decoder design, namely channel estimation, equalization and decoding are joined into a *combined decoder*.

We note here that information-theoretic justification for the system design based on training sequences has been provided recently by [ATV02] with similar results in [VHHK01, MGO02, HH03] where it is stated that the usage of optimized pilot sequences implies no or only small loss in channel capacity, at least in the high signal-to-noise ratio (SNR) domain. It should be emphasized, however, that an argument based on capacity results essentially requires that no delay constraints are imposed on the system, because infinite block-lengths are needed to achieve capacity. Thus, in real-time applications with strict delay constraints, the use of separate coding and training is not necessarily optimal in a general sense. It is therefore of interest to investigate communication systems that take the uncertain nature of the channel characteristics into account without restricting the design to the transmission of training sequences.

For frequency-flat fading MIMO channels, constellation design where the need for explicit CSI at the receiver is removed has been investigated in a line of papers [MH99, HM00, HMR$^+$00]. Extensions to frequency-selective channels are part of this thesis and were also considered in concurrent work [BB04].

1.5 Outline of the Thesis and Contributions

The areas to which this thesis contributes are illustrated in Figure 1.5. On the horizontal axis, we roughly characterize the assumptions on the channel model as either frequency-flat or frequency-selective (even though the former could be understood as a special case of the latter). On the vertical axis, we categorize the assumption on available receiver CSI, with the extreme cases of perfectly known and unknown channel. In between these extremes is the case of partial CSI, i.e., when the receiver has access to an estimate of the channel together with its error statistics. The case of zero-error thereby corresponds to perfect receiver CSI and a useless estimate to unknown CSI.

After a description and analysis of our applied system model in Chapter 2, we consider in Chapters 3 and 4 signaling design approaches for receivers that have no access to CSI prior to data transmission. These schemes are explicitly developed for the frequency-selective channel where uncertainty about the level of CSI is even more relevant than in the

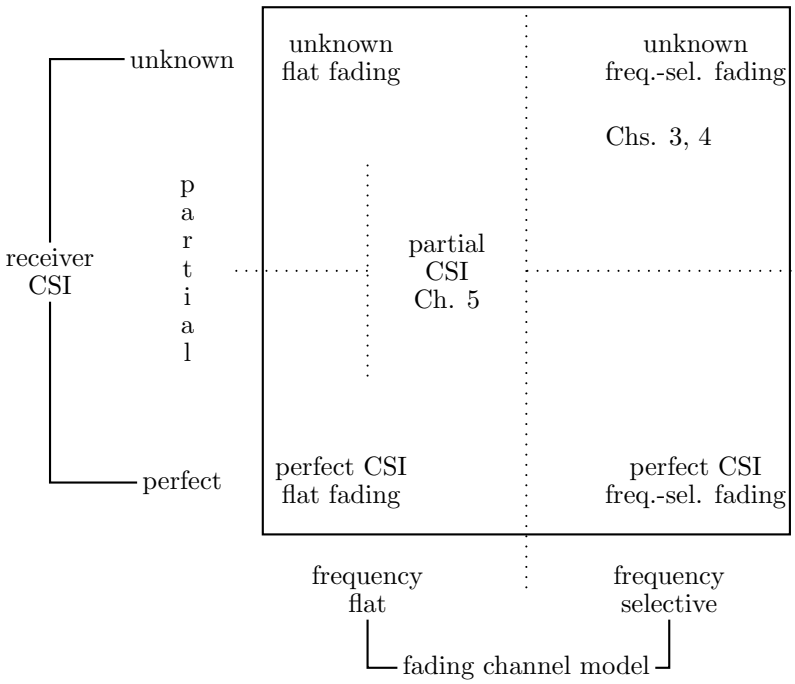


Figure 1.5: An overview of the contributions of this thesis

frequency-flat fading case because even more parameters are unknown.

In Chapter 5, we generalize from the assumption that the decoder has no knowledge about the channel realization and assume that an estimate (possibly due to a training sequence) is available together with information about its quality. This assumption includes the case of no CSI if the quality of the estimate is poor and therefore useless. We derive a design criterion for constellations that encompasses standard criteria used for a perfectly known channel and unknown channel at the receiver, thereby allowing a smooth transition between these two extreme cases. Example applications of the design criterion illustrate the value of a design matched to partial CSI at the receiver.

The order of the Chapters 3 through 5 reflects the chronological development of this thesis. Several different design approaches are described which depend on analytical results presented in Chapter 2. These results

were not all available from the start of the work leading to this thesis but they are anyhow summarized in Chapter 2 for better reference.

In more detail, the outline of this thesis is as follows.

1.5.1 Chapter 2, System Model and Analysis

This chapter describes some of the fundamental problems in detection over continuous-time channels that are unknown to the receiver. Several modeling assumptions as well as approaches to approximately optimal front-end processing in continuous time are presented. A detailed description of the data models that will be used in the remaining parts of the thesis along with two possible receiver architectures is presented. Moreover, the performance of these receiver operations is analyzed in various ways. The results of this chapter form the basis for the analysis and design of signaling schemes in Chapters 3, 4 and 5.

1.5.2 Chapter 3, Design in the Time Domain

Based on the work [SP00], a constellation design using signaling from a preselected discrete symbol alphabet was extended to longer codes that allow comparison with a larger variety of relevant benchmark codes. This work was targeted towards SISO systems and was published in

[SGP02] M. Skoglund, J. Giese, and S. Parkvall. Code design for combined channel estimation and error protection. *IEEE Transactions on Information Theory*, 48(5):1162–1171, May 2002.

Moreover, the above work was extended to MIMO systems and published in

[GS02c] J. Giese and M. Skoglund. Space–time code design for unknown frequency-selective channels. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, Florida, USA, May 2002.

[GS02a] J. Giese and M. Skoglund. Space–time code design for combined channel estimation and error protection. In *Proc. IEEE International Symposium on Information Theory*, Lausanne, Switzerland, June 2002.

[GS02b] J. Giese and M. Skoglund. Space-time code design for combined channel estimation and error protection. In *Proc.*

RadioVetenskap och Kommunikation (RVK), Stockholm, Sweden, June 2002.

The work using binary signaling relied on the optimization of a criterion based on exact pairwise error probabilities. During the final stages of this work, the paper [BV01] was published, providing a framework for detailed analysis of approximating the exact formulas used in [SGP02, GS02c, GS02b, GS02a] by asymptotic expressions that were much easier to analyze. Extending the results of [BV01] to the assumptions in our setup, it was now possible to formulate and solve the optimization problem on a parameter space in continuous variables and thus codes corresponding to general complex-valued signaling could be obtained. The results of this work were published in

[GS03a] J. Giese and M. Skoglund. Combined coding and modulation design for unknown frequency-selective channels. In *Proc. IEEE International Symposium on Information Theory*, Yokohama, Japan, 2003.

for SISO systems and in

[GS03c] J. Giese and M. Skoglund. Space–time constellations for unknown frequency-selective channels. In *Proc. IEEE International Conference on Communications*, Anchorage, AK, USA, 2003.

for MIMO systems. In addition, a journal paper extending the results in [GS03a] and [GS03c] was submitted as

[GS03b] J. Giese and M. Skoglund. Single and multi-antenna constellations for communication over unknown frequency-selective fading channels. Submitted to *IEEE Transactions on Information Theory*, May 2003. Revised October 2005.

The chapter concludes with a summary of both design approaches.

1.5.3 Chapter 4, Design in the Frequency Domain

The orthogonality of the subcarriers in an OFDM system simplifies the formulation of two signal designs for unknown channel operating in the frequency domain. In the first part of Chapter 4, a simple formulation of differential space-frequency modulation in analogy to space-time differential modulation is described. Using an analysis similar to the method

described in Chapter 2, a criterion is derived which guarantees the exploitation of full space-frequency diversity for a class of diagonal codes which were conceived for space-time differential transmission. Simulation results illustrate the application of this criterion. The results of this section were published in

[GS04] J. Giese and M. Skoglund. Performance of unitary differential space-frequency modulation. In *International Symposium on Information Theory and its Applications*, Parma, Italy, October 2004.

The signal model used in this chapter was also extended to a multiuser uplink scenario and investigated in

[D2.05] WINNER D2.7. Assessment of advanced beamforming and MIMO technologies. Technical report, Wireless World Initiative New Radio, February 2005. Available online (October 2005) at `https://www.ist-winner.org/DeliverableDocuments/D2-7.pdf`.

1.5.4 Chapter 5, Design for Partial CSI at the Receiver

A framework for the analysis of signaling schemes designed for the operation over frequency-flat fading channels is developed where it is assumed that the receiver has access to a channel estimate with known error statistics. The framework thereby includes the extreme cases of perfectly known or completely unknown channel at the receiver. Design examples illustrate the value of the proposed framework. The content of this chapter appears in

[GS05a] J. Giese and M. Skoglund. Space–time constellation design for partial CSI at the receiver. In *Proc. IEEE International Symposium on Information Theory*, Adelaide, Australia, September 2005.

[GS05b] J. Giese and M. Skoglund. Space–time constellation design for partial CSI based on code combination. In *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, October 2005.

as well as in the form of a journal paper in

[GS05c] J. Giese and M. Skoglund. Space–time constellation design for partial CSI at the receiver. October 2005. In Preparation.

1.5.5 Chapter 6, Summary and Future Work

This chapter summarizes the results of this thesis and concludes about the possible improvements obtained by using joint code design. Several open questions that could not yet be answered are outlined as a proposal for future work.

1.5.6 Appendix A, Some Useful Lemmas and Rules

Some standard results of linear algebra together with a useful lemma on the multivariate complex Gaussian probability density is presented for easier reference to the reader.

1.6 Acronyms

Some of the acronyms used in this thesis are explained below.

AWGN additive white Gaussian noise

BER bit error rate

BPSK binary phase shift keying

CSI channel state information

CT continuous time

GLRT generalized likelihood-ratio test

GSM global system for mobile communication, earlier: group spéciale mobile

ISI intersymbol interference

LLR log-likelihood ratio

LOS line of sight

MIMO multiple-input multiple-output

MISO multiple-input single-output

ML maximum-likelihood

MMSE minimum mean square error

OFDM orthogonal frequency division multiplexing

PAM pulse amplitude modulation

PEP pairwise error probability

PSK phase-shift keying

PSWF prolate spheroidal wave function

QAM quadrature amplitude modulation

QPSK quaternary phase shift keying

RX receiver

SIMO single-input multiple-output

SISO single-input single-output

SNR signal to noise ratio

TDL tapped delay line

TX transmitter

UMTS universal mobile telecommunication system

WER word error rate

WINNER Wireless Initiative New Radio

WLAN wireless local area network

w.l.o.g. without loss of generality

w.r.t. with respect to

1.7 Notation

Symbol	meaning
$\mathbf{v}$, $\{\mathbf{v}\}_k$	a column vector and its k.th element
$\mathbf{A}$, $\{\mathbf{A}\}_{kl}$	a matrix and its element in row k and column l
$\mathbf{I}_M$	the $M \times M$ identity matrix
$\mathbf{I}$	the identity matrix. The dimension is clear from the context and the dimension index is omitted to simplify notation.
$\mathbf{A}^H$,$\mathbf{A}^T$	Hermitian transpose and transpose of $\mathbf{A}$
$\mathrm{tr}\,(\mathbf{A})$	the trace of a square matrix $\mathbf{A}$
$\|\mathbf{A}\| = \sqrt{\mathrm{tr}\,(\mathbf{A}^H\mathbf{A})}$	the Frobenius norm of $\mathbf{A}$
$\|\mathbf{A}\|_S$	spectral norm of $\mathbf{A}$, i.e., square root of the largest eigenvalue of $\mathbf{A}^H\mathbf{A}$
$\|\mathbf{A}\|$	the determinant of a square matrix $\mathbf{A}$
$\mathbf{A}^\dagger$	the Moore-Penrose pseudo inverse of $\mathbf{A}$. If $\mathbf{A}$ is full column rank, then $\mathbf{A}^\dagger = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$
$\mathbf{P}_\mathbf{A}$	the projection matrix on the column space of $\mathbf{A}$. It holds that $\mathbf{P}_\mathbf{A} = \mathbf{A}\mathbf{A}^\dagger$.
$\mathbf{P}_\mathbf{A}^\perp$	the projection matrix on the orthogonal complement of the column space of $\mathbf{A}$. It holds that $\mathbf{P}_\mathbf{A}^\perp = \mathbf{I} - \mathbf{P}_\mathbf{A}$.
$\mathbf{A} \otimes \mathbf{B}$	the Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathbb{S}$	a set
$\mathbb{A} \setminus \mathbb{B}$	set of all elements that are in $\mathbb{A}$ but not in $\mathbb{B}$
$\mathbf{a} \sim \mathcal{CN}(\mu, \mathbf{C})$	the elements of $\mathbf{a}$ are circular symmetric complex Gaussian random variables with mean μ and covariance matrix $\mathbf{C}$
$\mathrm{diag}\,\{d_1, \ldots, d_D\}$	a diagonal $D \times D$ matrix with elements $d_1, \ldots, d_D$ on the main diagonal
$[x]^+$	the positive part of the real number x, i.e., $[x]^+ = x$ if $x > 0$ and $[x]^+ = 0$ if $x \leq 0$
$\binom{n}{k}$	binomial coefficient, i.e. $\binom{n}{k} \triangleq \frac{n!}{k!(n-k)!}$
$y(t) = f(t) \star x(t)$	continuous-time convolution of $x(t)$ and $f(t)$, i.e. $y(t) = \int_{-\infty}^{\infty} x(\tau)y(t-\tau)d\tau$
$\Re(z)$	real part of the complex scalar z
$\Im(z)$	imaginary part of the complex scalar z
z^*	complex conjugate of the complex scalar z
$a \lesssim b$	b is an approximation of an upper bound on a

1.8 Common Identifiers

quantity	meaning
j	imaginary unit, $j^2 = -1$
	j is sometimes also used as index. The distinction between these two meanings is clear from the context.
M_T	number of transmitter antennas
M_R	number of receiver antennas
K	number of different users
L	number of resolvable paths
Z	size of the signal set
T	number of symbols transmitted experiencing the same channel realization
$\mathbb{C}$,$\mathbb{R}$	the set of complex and real numbers

Chapter 2

System Model and Analysis

This thesis deals with the transmission of digital data (usually in the form of bits) between a transmitter and a receiver. The bits signifying the message to be transmitted determine the way in which the transmitter manipulates a physical medium which is accessible to both transmitter and receiver. In other words, the bits determine the signal transmitted to the receiver. The fundamental problem in data communication is the fact that the receiver in general does not have access to the exact signal that the transmitter has sent but rather to a signal which is a (more or less accurate) representation of the transmitted signal. The impact of this modification of the transmitted signal is modeled and summarized as the impact of the so-called "channel" in between transmitter and receiver, see Figure 2.1. A large variety of example channels with different

Figure 2.1: A schematic view of the data transmission problem

characterizations exist, among others underwater, data storage, wire line and and wireless communication channels [Pro95].

In order to design a communication system transmitting data through the channel, it is of utmost importance to characterize the channel's be-

havior and impact. This characterization is usually done by means of a mathematical model that describes the essential features of the underlying physical medium while still being simple enough to allow for mathematical analysis. It is not uncommon that a compromise between realism and mathematical tractability is necessary to arrive at a useful channel model. The considerations in this chapter are an example of this tradeoff.

We review in Section 2.1 the general concept of linear time-time waveform channels. Two alternatives for the conception of a data communication system and in particular for the design of signals used for transmission over such channels are within the focus of this thesis: single carrier systems and multicarrier systems. We first comment on a typical comparison between these two approaches in Section 2.2 and discuss our models of a single carrier system and of a multicarrier system in Sections 2.3 and 2.4, respectively. Both models lead to similar mathematical formulations which allow a unified discussion of receiver strategies and performance analysis in Sections 2.5 and 2.6, respectively. Section 2.7 concludes this chapter.

2.1 Linear Waveform Channels in Radio Communications

An important class of channels are so-called *linear* waveform channels: The continuous time (CT) transmitted signal $s(t)$ is distorted by a linear system with impulse response $f(t)$. Moreover, the signal is disturbed by additive noise $w(t)$ which is usually assumed to be white (caused by, e.g., thermal agitation of electrons in a conductor), see Figure 2.2. This results in the model

$$y(t) = s(t) \star f(t) + w(t) = \int_{-\infty}^{\infty} s(t-\tau) f(\tau) d\tau + w(t) \tag{2.1}$$

where $y(t)$ is the signal at the receiver. Depending on the physical layer conditions, different characterizations of $f(t)$ exist which are reflected in various ways to model $f(t)$. Usually, the model summarizes upconversion to radio frequency, radio transmission, reception and downconversion in the so-called complex baseband model of f, which can therefore be a complex-valued function. In radio communication channels with mobile transmitter or receiver, f can be a time-varying function $f = f(t, \tau)$ where $f(t, \tau)$ represents the response of the channel at time t when an

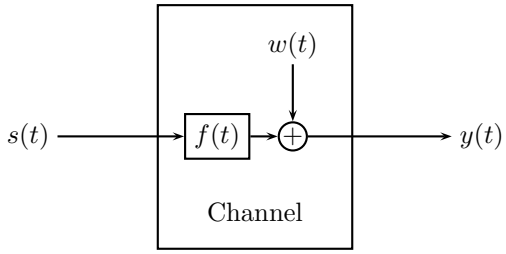


Figure 2.2: A channel with linear distortion

impulse enters the system at time $t-\tau$. This dependency on time reflects the possibility that the propagation conditions are subject to change, in particular in radio communication channels between mobile transmitters and/or receivers.

In general, radio waves do not necessarily reach the transmitter in the direct line of sight (LOS) between transmitter and receiver. Due to scattering, reflection, refraction or diffraction [PP97] in the propagation medium, the outgoing wave can reach the receiver via a number of different paths. When a large number of different rays with negligible time but significant phase shifts arrive at the receiver, the resulting wave is the sum of all impinging waves. If their number is large, a statistical modeling of the incoming waves appears favorable and the central limit theorem can be applied. $f(t;\tau)$ can then often be considered a complex-valued Gaussian random process [Pro95]. If $f(t;\tau)$ is zero-mean, its envelope $|f(t;\tau)|$ is Rayleigh distributed and therefore the fading is said to be *Rayleigh* fading. If the mean of $f(t;\tau)$ is non-zero (i.e. there are some fixed scatterers or there is a LOS component), the fading is called *Rician*.

An important special case is the *frequency-nonselective* channel in slow fading. The channel gain is constant over the relevant signal bandwidth and constant during the transmission of the complete signaling interval. This can be modeled as $f(t;\tau) = f_0\delta(\tau)$ leading to

$$y(t) = f_0 s(t) + w(t). \tag{2.2}$$

which is a significant simplification compared to the more general model in (2.1).

The goal of transmitter design is to find a function $s(t)$ dependent on the data bits. The goal of receiver design is to be able to detect the

information embedded in $s(t)$ by observing $y(t)$.

2.2 Single Carrier vs. Multicarrier

The design of a data communication system and in particular the design of $s(t)$ is in general a highly complex task involving a multitude of parameters that have to be compared and traded off under the design assumptions as, e.g., data and error rate, delay or implementation cost. The assumptions required for a deliberate judgment of this tradeoff are the properties of the communication channel that the system is operating on. If the available frequency band is sufficiently large such that the channel can no longer be modeled as narrow-band (see (2.2)), the system design should take this into account. Two design alternatives for this kind of channel are considered in this thesis: Single carrier systems and multicarrier systems [Bin90], the latter in the form of orthogonal frequency division multiplexing (OFDM). Whereas a single carrier system employs only a single carrier frequency and takes the character of the channel explicitly into account in receiver or transmitter design (or both), a multicarrier system can be understood as subdividing the available frequency band into smaller frequency bands where the channel is (at least approximately) constant, thus leading to a similar setup as a for frequency-nonselective channel.

There is a vast number of different studies concerning the benefits of these design approaches [TdPE+04]. OFDM systems are in general attractive due to low-complexity channel equalization but can have problematic requirements concerning the linearity of the employed amplifiers because the time-domain signal that is radiated over the channel can have large variations in output power. Single carrier systems are more robust to synchronization errors in the receiver but require more complex channel equalization methods. These characteristics are just examples and in general several additional properties have to be taken into account related to the implementation of a complete communication system

In this thesis, we do not attempt to contribute in detail to this discussion and focus our attention to a particular difference between single carrier and multicarrier systems. Whereas in OFDM, due to orthogonality of the carriers, we obtain observables which are a priori independent of symbols transmitted on other carriers, such a separation of observables is a priori not available for a single carrier system. The orthogonality property of OFDM requires however the transmission of a cyclic prefix,

i.e., a short signal carrying redundant data which reduces the available power per data symbol.

Both system and data models will be described in turn, leaving out all implementation aspects mentioned before and using idealized system components. We will however comment in detail on the orthogonality property available in OFDM in contrast to a single carrier system.

2.3 Single Carrier Systems

We describe here our model for the transmitter and receiver front-end in a single carrier system for a single antenna at both transmitter and receiver. The generalization to multiple antennas is then straightforward.

A typical transmitter design using an encoder and memoryless linear modulation is depicted in Figure 2.3. We refer throughout this thesis to the equivalent complex baseband description. The incoming bits $b_\nu \in$

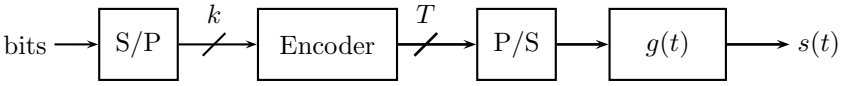


Figure 2.3: Typical transmitter design

$\{0,1\}$, $\nu = 0,1,2,\ldots$ are grouped into disjoint blocks of k bits. In the encoder, these k bits determine a sequence of T complex symbols $s_n \in \mathcal{A} \subset \mathbb{C}$, $n = 0,...,T-1$ where $\mathcal{A}$ is the set of possible symbols (also called the symbol alphabet). The symbol sequence is then fed into a pulse-shaper with waveform $g(t)$ to result in the transmitted signal

$$s(t) = \sum_{n=0}^{T-1} s_n g(t - nT_s) \tag{2.3}$$

where $1/T_s$ is the symbol rate. Standard choices for the symbols s_n that are based on a fixed alphabet $\mathcal{A}$ include

pulse amplitude modulation (PAM), where

$$\mathcal{A} = \{s | s = (2m - 1 - M)d, \quad m = 1,\ldots,M\}$$

where M is the number of symbols and d the distance between adjacent symbols. An important special case is binary phase shift keying (BPSK) for $M = 2$ and $d = 0.5$ leading to $\mathcal{A} = \{-1, 1\}$,

phase-shift keying (PSK), where

$$\mathcal{A} = \{s | s = e^{j2\pi(m-1)/M}, \quad m = 1, \ldots, M\}$$

with the important special case $M = 4$ known as quaternary phase shift keying (QPSK), and

quadrature amplitude modulation (QAM), where both quadrature carriers (real and imaginary part of s_n) are modulated, possibly using a combination of PAM and PSK [Pro95].

The format of the symbols used for transmission is subject to discussion in Chapter 3. In particular, we will investigate two cases:

1. As an example of a preselected alphabet, we will restrict our code search to codes with BPSK symbols.

2. We will not use any preselected symbol alphabet and design values s_n that can take on any complex number (subject to a power constraint) as an example for a general form of QAM.

It is the topic of this thesis to determine for both cases good (in a sense to be described later) mappings from the source bits to the symbols $s_0, \ldots, s_{T-1}$, or, in other words, to determine the encoder. In order to assess the characteristics and performances of different codes, the receiver operation must be known during the design. This is the topic of the following sections.

2.3.1 Known Channel at the Receiver

Given the received time time signal $y(t)$ in (2.1), the receiver has to decide which sequence $s_0, \ldots, s_{T-1}$ has been transmitted. If the channel $f(t)$ is perfectly known at the receiver and of finite length and energy, Forney [For72] showed that the complex baseband model in (2.1) is statistically equivalent to the discrete-time model

$$y_n = \sum_{m=0}^{L-1} s_{n-m} h_m + w_n, n = 0, \ldots, T + L - 2 \tag{2.4}$$

where the discrete-time channel coefficients h_m are related to the overall channel response $c(t) = f(t) \star g(t)$ by

$$\int_{-\infty}^{\infty} c(\theta)c^*(\theta - kT_s)d\theta = \sum_{\nu=0}^{L-1} h_\nu h^*_{\nu-k}$$

(see also [CP96]) and the noise w_k is white. Statistical equivalence means here that no information relevant to maximum likelihood sequence detection on $s_0, \ldots, s_{T-1}$ based on a limited time-interval is lost when reducing a continuous-time observation of $y(t)$ to the set of observables $\{y_n\}$. A subsequent decoder algorithm can thus operate on a finite number of observables $\{y_n\}$ with independent noise which is a great simplification in comparison to any potential decoding algorithm operating on the continuous-time observation. The model (2.4) results from filtering matched to $c(t)$ and subsequent noise-whitening along with symbol-rate sampling. An alternative receiver without noise-whitening was described by Ungerboeck [Ung74] taking into account colored noise in the data detection algorithm.

The model (2.4) has the nice property that it is analogous to (2.1) in discrete time. The intuitive appeal of this solution might be the reason that the necessary assumptions for (2.4) to result from an optimal front-end processing are sometimes neglected. In particular, it is imperative that the overall channel response $c(t)$ and therefore the channel response $f(t)$ are *perfectly known* at the receiver. If $f(t)$ is unknown, the model in (2.4) cannot be applied without modification, additional assumptions or further justification.

2.3.2 Unknown Channel at the Receiver

Chugg et al. presented in [CP96] detailed considerations of the front-end processing for joint maximum likelihood sequence estimation and data detection when $f(t)$ is *unknown* to the receiver. Under the assumption that the channel can be modeled as a tapped delay line (TDL), i.e.,

$$f(t) = \sum_{l=0}^{N_r L_c} f_l \delta(t - lT_r) \tag{2.5}$$

corresponding to $N_r L_c + 1$ equispaced "taps" (or resolvable paths at the receiver) and under the assumption that $T_s = N_r T_r$, i.e., the symbol time T_s is an integer multiple of the channel resolution time T_r, a set

of sufficient statistics for the proposed sequence detection of s_k can be obtained using pulse-matched filtering and subsequent sampling at rate $1/T_r$, possibly followed by a discrete-time whitening filter $1/V^*(z)$, see Figure 2.4. In the case of a general continuous-time channel model, i.e.,

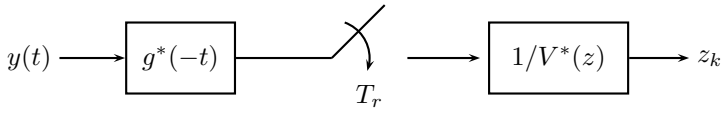


Figure 2.4: Optimal front-end processing for the channel model in (2.5)

without imposing any further structure on f apart from being of finite length and energy, it is shown in [CP96] that there is no counterpart to the model in [For72] that does not imply a loss of information with respect to ML sequence detection of the symbols when f in unknown at the receiver. The corresponding metric based on the received waveform for the detection of the sequence of symbols s_n does in general not exist in the mean-square sense. Thus, the standard approach of deriving a finite set of discrete observables that form a sufficient statistic breaks down. Several suboptimal front-ends have been proposed which are either based on approximating the continuous-time channel response $f(t)$ using a TDL model or based on approximating the detection metric. Hansson [Han03] proposed an alternative approximation by representing the received waveform $y(t)$ by a finite number of projections onto prolate spheroidal wave function (PSWF)s [SP61a, SP61b, SP62] which have the appealing property that the representation error is minimized in the least square sense. Moreover, the number of necessary distinct base functions can be quantified for a given allowable error threshold. Still, an approximation is made which is in principle information lossy.

We can therefore conclude that under the assumptions made we do not know of any description of the transition from continuous time into a finite discrete-time representation which is optimal in the sense that no information is lost in the general case. We therefore have to accept a suboptimal approach. Partly for the sake of mathematical tractability in subsequent derivations on the discrete-time model, we follow the approach proposed by Chugg [CP96] (see also Figure 2.4) with symbol-based sampling, taking into account that the generalization to an oversampled description is possible within the framework we are deriving in the next section. A similar approach was used in [CC00].

2.3.3 Single Carrier Data Model

Motivated by the previous section, we introduce now a discrete-time data model for a single carrier system operating over a frequency-selective channel. The model is first derived for a single-antenna system. The generalization to a system using M_T transmitter and M_R receiver antennas is then straightforward.

We consider the transmission of T symbols $\{s_t\}$, $t = 0, \ldots, T-1$, $s_t \in \mathbb{C}$ through a frequency-selective channel described by the channel coefficients $\{h_l\}$, $l = 0, \ldots, L-1$, $h_l \in \mathbb{C}$ with additive white complex Gaussian noise of variance σ^2 denoted by w_t. Motivated by our discussion of the continuous-time front-end in Section 2.3.2, we model the received complex baseband discrete time signal y_t as

$$y_t = \sum_{l=0}^{L-1} h_l s_{t-l} + w_t, \qquad t = 0, \ldots, T+L-2 \tag{2.6}$$

where we define $s_t \triangleq 0$ for $t < 0$ and $t > T-1$.

A generalization to multiple antennas at transmitter and receiver leads to

$$y_{t,n} = \sum_{m=1}^{M_T} \sum_{l=0}^{L-1} h_{l,m,n} s_{t-l,m} + w_{t,n}$$

where $\{h_{l,m,n}\}_{l=0}^{L-1}$, $h_{l,m,n} \in \mathbb{C}$ describes the frequency-selective channel from transmitter antenna m to receiver antenna n, $s_{t,m} \in \mathbb{C}$ is the t-th symbol transmitted on antenna m, $y_{t,n}$ is the t-th received symbol on antenna n and $w_{t,n}$ is zero-mean circular symmetric complex Gaussian noise with variance σ^2.

In order to simplify notation, we summarize the transmitted symbols in the $T \times M_T$ matrix $\mathbf{S}$ with elements $\{\mathbf{S}\}_{ij} = s_{i-1,j}$ and write the received symbols on antenna n as

$$\mathbf{y}_n = \bar{\mathbf{S}}\mathbf{h}_n + \mathbf{w}_n,$$

where $\bar{\mathbf{S}}$ is a $(T+L-1) \times LM_T$ matrix that contains L copies of $\mathbf{S}$ as in

$$\bar{\mathbf{S}} \triangleq \begin{bmatrix} \mathbf{S} & & & & \mathbf{0} \\ & \mathbf{S} & & & \\ & & \mathbf{S} & \ddots & \\ & & & \ddots & \mathbf{S} \\ & \mathbf{0} & & & \end{bmatrix}, \tag{2.7}$$

$\mathbf{w}_n$ is the vector of stacked noise components $w_{0,n}, \ldots, w_{T+L-1,n}$ and we define

$$\mathbf{h}_n \triangleq [h_{0,1,n} \ldots h_{0,M_T,n}\ h_{1,1,n} \ldots h_{1,M_T,n}\ \ldots\ h_{L-1,1,n} \ldots h_{L-1,M_T,n}]^T.$$

Rearranging the vectors $\mathbf{y}_1, \ldots, \mathbf{y}_{M_R}$ and $\mathbf{h}_1, \ldots, \mathbf{h}_{M_R}$ into the matrices $\mathbf{Y}$ and $\mathbf{H}$, respectively, we obtain

$$\underbrace{[\mathbf{y}_1 \ \ldots \ \mathbf{y}_{M_R}]}_{\mathbf{Y}} = \bar{\mathbf{S}} \underbrace{[\mathbf{h}_1 \ \ldots \ \mathbf{h}_{M_R}]}_{\mathbf{H}} + \underbrace{[\mathbf{w}_1 \ \ldots \ \mathbf{w}_{M_R}]}_{\mathbf{W}}. \tag{2.8}$$

The extension from $\mathbf{S}$ to $\bar{\mathbf{S}}$ in (2.7) represents an extension from a system with frequency-flat to frequency-selective fading where the description of the transmitted symbols is given in both $\bar{\mathbf{S}}$ as well as $\mathbf{S}$. For $L = 1$, i.e., for a frequency-flat channel, we have $\mathbf{S} = \bar{\mathbf{S}}$.

We note here also that the system model with M_T transmitter antennas and a channel of length L is equivalent to a model with $M_T L$ antennas over a frequency-flat fading channel where the signals on $M_T(L-1)$ antennas are constrained to be temporally shifted copies of the signals transmitted on the remaining M_T antennas. This observation is sometimes explained using the term "virtual antennas" [LFT01]. We can therefore consider the given setup to be equivalent to a scenario in flat fading with the additional constraint that (2.7) holds. This mathematical equivalence is further investigated in Chapter 3 where we compare constellations optimized based on constraint (2.7) with design rules optimized without (2.7).

2.4 Orthogonal Frequency Division Multiplexing

In this section we will introduce our system model for a multicarrier modulation system when OFDM is applied. In a similar format as in the previous section, we will first discuss a continuous time model and describe then the reduction to a finite set of observables. The derivation is mainly inspired by [San96].

2.4.1 Continuous-Time Model

In contrast to forming the transmitted signal as a sequence of elementary pulses, we assume here that the signal is constructed as linear combination of waveforms $\phi_n(t)$ in

$$s(t) = \sum_{n=0}^{N-1} s_n \phi_n(t), \quad 0 \le t \le T$$

where the design of these waveforms is largely motivated by the desire to transform the frequency-selective channel $f(t)$ into a frequency-flat channel such that the receiver can compute observables y_n that depend only on s_n and not on other symbols $s_k, k \neq n$. One set of suitable waveforms for transmission through a channel of finite support of length T_L

$$\phi_k(t) = \begin{cases} \frac{1}{\sqrt{T-T_L}} \exp\left(jk\frac{2\pi}{T-T_L}(t-T_L)\right) & \text{if } 0 \le t \le T \\ 0 & \text{elsewhere} \end{cases}$$

which are truncated harmonic oscillations of frequency $k/(T-T_L)$ in the interval $[T_L, T]$. These functions satisfy the orthonormality condition

$$\int_{T_L}^{T} \phi_k(t)\phi_n^*(t)dt = \begin{cases} 1 \text{ if } k = n \\ 0 \text{ otherwise.} \end{cases}$$

The interval $[0, T_L]$ is the so-called *cyclic prefix* which assures orthogonality of the waveforms ϕ_k even in the presence of a temporal shift τ, $0 \le \tau < T_L$ because

$$\begin{aligned} \int_{T_L}^{T} \phi_k(t-\tau)\phi_n^*(t)dt &= \frac{1}{T-T_L}\int_{T_L}^{T} e^{j\frac{2\pi}{T-T_L}(k-n)(t-T_L)} e^{-j\frac{2\pi}{T-T_L}k\tau} dt \\ &= \begin{cases} \exp(-j\frac{2\pi}{T-T_L}k\tau) \text{ if } k = n \\ 0 \text{ if } k \neq n \end{cases} \end{aligned}$$

The benefit of this orthogonality is a simplification of the receiver. After filtering the received signal $y(t)$ with the pulse $\psi_n(t) = \phi_n^*(T-t), 0 \leq t \leq T - T_L$ and subsequent sampling at time instant T we obtain

$$\begin{aligned} y_k &= \int_{-\infty}^{\infty} y(t)\psi_k(T-t)dt \\ &= \int_{T_L}^{T}\int_{0}^{T_L} \left(f(\tau)\sum_{n=0}^{N-1} s_n\phi_n(t-\tau) + w(t) \right) d\tau\phi_k^*(t)dt \\ &= \int_{0}^{T_L} f(\tau)\sum_{n=0}^{N-1} s_n \int_{T_L}^{T} \phi_n(t-\tau)\phi_k^*(t)dtd\tau + \int_{T_L}^{T} w(t)\phi_k^*(t)dt \\ &= s_k \int_{0}^{T_L} f(\tau)e^{-j\frac{2\pi}{T-T_L}k\tau}d\tau + \int_{T_L}^{T} w(t)\phi_k(t) \\ &= s_k f_k + w_k \end{aligned}$$

where $f_k \triangleq \int_0^{T_L} f(\tau)\exp(-j\frac{2\pi}{T-T_L}k\tau)$ is the Fourier transform of $f(t)$ evaluated at the frequency $k/(T-T_l)$ and the w_k are Gaussian random variables which are independent in k because the applied function $\phi(t)$ are orthogonal.

This simplicity in the dependence of the observables y_k on the transmitted symbols s_k is an important feature of an OFDM system, effectively removing interference between symbols on different carriers. However, similar to the single carrier system, the receiver front-end cannot be considered optimal in the general sense of providing sufficient statistics in the discrete variables y_k for detecting of the entire sequence s_k. A simple intuitive reason for this is the fact that the cyclic prefix which contains information about the transmitted symbols is effectively discarded at the receiver (the receiver filter ψ_k has an impulse response of length $T - T_L$ (not T) which is required to assure orthogonality). We can therefore conclude that also for an OFDM system, we do not know of any simple front-end which is not information lossy and therefore we must accept suboptimal approaches.

2.4.2 Discrete-Time OFDM Data Model

The data model in the previous subsection $y_k = s_k h_k + w_k$ also results from a discrete-time model [San96] of the OFDM system which assumes the data symbols to be defined in the frequency domain. Before transmission through a discrete-time channel, the symbols are transformed to

the time domain using inverse discrete time Fourier transformation and a cyclic prefix is prepended in the time domain. The channel is then modeled as a discrete-time convolution similar to (2.4). The receiver discards the cyclic prefix and computes the received symbols in the frequency domain using a discrete Fourier transformation. This model implies N symbols during a time interval $T - T_L$ which corresponds to a sampling time $T_s = (T - T_L)/N$.

If the channel is modeled as before with a tapped delay line in

$$f(t) = \sum_{l=0}^{L-1} h_l \delta(t - lT_s)$$

we obtain

$$\begin{aligned} f_k &= \sum_{l=0}^{L-1} \int_0^{T_L} h_l \delta(t - lT_s) e^{-j\frac{2\pi}{T-T_L}k\tau} d\tau \\ &= \sum_{l=0}^{L-1} h_l e^{-j\frac{2\pi}{T-T_L}klT_s} \\ &= \sum_{l=0}^{L-1} h_l e^{-j\frac{2\pi}{N}kl} \end{aligned}$$

Using the matrix

$$\mathbf{D}_N \triangleq \text{diag}\{e^{-j\frac{2\pi}{N}0}, e^{-j\frac{2\pi}{N}1}, \ldots, e^{-j\frac{2\pi}{N}(N-1)}\}$$

and the vector $\mathbf{s} = [s_0, \ldots, s_{N-1}]^T$ we can now write

$$\mathbf{y} = [\mathbf{s} \quad \mathbf{D}_N\mathbf{s} \quad \ldots \quad \mathbf{D}_N^{L-1}\mathbf{s}] \begin{bmatrix} h_0 \\ \vdots \\ h_{L-1} \end{bmatrix} + \begin{bmatrix} w_0 \\ \vdots \\ w_{N-1} \end{bmatrix}.$$

The generalization to multiple transmitter antennas is straightforward. Let $\mathbf{S}$ be an $N \times M_T$ matrix with element $\{\mathbf{S}\}_{nm}$ being the symbol using waveform n on transmitter antenna m. We can then write

$$\mathbf{Y} = [\mathbf{y}_1 \quad \ldots \quad \mathbf{y}_{M_T}] = [\mathbf{S} \quad \mathbf{DS} \quad \ldots \quad \mathbf{D}^{L-1}\mathbf{S}]\mathbf{H} + \mathbf{W} \tag{2.9}$$

where column r of $\mathbf{H}$ contains the weights of the $(L-1)M_T$ taps from the transmitter to receiver antenna r and $\mathbf{y}_r$ are the symbols received on antenna r.

2.4.3 Comparison of Data Models

It is apparent that the models in (2.8) and (2.9) are similar. Both the data models in (2.9) and (2.8) with the constraint (2.7) essentially represent a discrete-time convolution. A significant difference however is the substructure imposed on the signaling matrix in (2.7) and (2.9). Whereas in the single carrier model, the symbols are shifted through the rows, they are multiplied in the OFDM model using the matrix $\mathbf{D}$ reflecting a *cyclic* shift of symbols in the time domain. Moreover, a cyclic prefix is needed in the OFDM model which is not required in the single carrier model.

Moreover, the structure of $\bar{\mathbf{S}}$ in (2.7) implies non-overlapping reception of symbol blocks. One example where this assumption holds is in a frequency-hopping system where each block is transmitted in a different frequency range with independent channel realizations. In the completely general case, the assumption of non-overlapping received blocks may require some kind of "guard intervals" before and after the transmission of each block. In this general case, we assume that the guard intervals consist of zero-valued symbols but we could consider alternative solutions such as the cyclic prefix as well (see, e.g., [TdPE$^+$04]). In that sense, the single carrier model has a higher degree of design freedom, which is achieved at the cost of non-orthogonality of the observables at the receiver. Since we are investigating mostly sequence detection over the entire block, i.e., we consider all symbols on all carriers jointly, this disadvantage is not of major importance. A notable exception here is differential space-frequency coding, see Section 4.

2.5 Receiver Design

The data models in Sections 2.3.3 and 2.4.2 can be summarized in the following linear model

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{w}$$

where the signaling matrix $\mathbf{X}$ is of dimension $T' \times M$, the channel $\mathbf{h}$ is of dimension $M \times 1$ and noise vector $\mathbf{w}$ as well as the received vector $\mathbf{y}$ are of dimension $T' \times 1$. For the single carrier system of (2.8), this reformulation can be obtained by stacking the columns of $\mathbf{Y}$, $\mathbf{H}$ and $\mathbf{W}$ into $\mathbf{y}$, $\mathbf{h}$ and $\mathbf{w}$, respectively as well as replacing $\mathbf{X}$ with $\mathbf{I}_{M_R} \otimes \bar{\mathbf{S}}$. Similarly, the reformulation for the multicarrier system in (2.9) replaces $\mathbf{X}$ with $\mathbf{I}_{M_R} \otimes [\mathbf{S} \quad \mathbf{DS} \quad \dots \quad \mathbf{D}^{L-1}\mathbf{S}]$ and the other matrices as before. The dimension M of the vector $\mathbf{h}$ therefore is $M_T M_R L$.

The task of the receiver is to form a decision on the transmitted signal matrix $\mathbf{X}$ which was chosen from a predetermined set $\{\mathbf{X}_0, \ldots, \mathbf{X}_{Z-1}\}$. The formulation of a rule to make this decision requires a consideration of the involved quantities. Apart from noise, the channel vector $\mathbf{h}$ is unknown. Whereas the noise is usually modeled as a stochastic quantity described by a probability density, different approaches on the modeling of $\mathbf{h}$ exist. We will discuss and compare two of these approaches.

2.5.1 ML

Usually, the noise vector $\mathbf{w}$ is modeled as a random variable implying that the vector of observables $\mathbf{y}$ is also random. The receiver can then be defined as the mapping from the received signal $\mathbf{y}$ to the detected signal $\hat{\mathbf{X}} = \hat{\mathbf{X}}(\mathbf{y})$. The objective thereby is to choose this mapping such that the probability of false detection

$$\Pr(\hat{\mathbf{X}}(\mathbf{y}) \neq \mathbf{X}|\mathbf{y}) = 1 - \Pr(\hat{\mathbf{X}}(\mathbf{y}) = \mathbf{X}|\mathbf{y})$$

is minimized. Thus, we should choose

$$\hat{\mathbf{X}}(\mathbf{y}) = \arg\max_{\mathbf{X}} \Pr(\mathbf{X}|\mathbf{y}).$$

Using Bayes theorem, we can reformulate this to

$$\begin{aligned}\hat{\mathbf{X}} &= \arg\max_{\mathbf{X}} \frac{p(\mathbf{y}|\mathbf{X}) \Pr(\mathbf{X})}{p(\mathbf{y})} \\ &= \arg\max_{\mathbf{X}} p(\mathbf{y}|\mathbf{X}) \Pr(\mathbf{X})\end{aligned}$$

If the signals $\mathbf{X}$ are equally likely (as we assume throughout this thesis), we obtain the maximum-likelihood (ML) detector

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}} p(\mathbf{y}|\mathbf{X}). \tag{2.10}$$

This detector obviously requires knowledge of the density of the received signal conditioned on an assumed transmitted signal $\mathbf{X}$. This density depends on the model used for the noise $\mathbf{w}$ and the channel $\mathbf{h}$. For Gaussian noise $\mathbf{w}$ with variance $\sigma^2\mathbf{I}$ and deterministic $\mathbf{h}$, we have the density as function of $\mathbf{h}$ in

$$p(\mathbf{y}|\mathbf{X};\mathbf{h}) = \frac{1}{|\pi\sigma^2\mathbf{I}|} \exp(-\sigma^{-2}\|\mathbf{y} - \mathbf{X}\mathbf{h}\|^2)$$

which leads to useful detectors if $\mathbf{h}$ is perfectly known. If $\mathbf{h}$ is unknown, it is frequently modeled as a Gaussian random variable with $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C_h})$ which is independent of the receiver noise. Conditioned on $\mathbf{X}$, the received vector $\mathbf{y}$ is then also Gaussian with density

$$p(\mathbf{y}|\mathbf{X}) = \frac{1}{|\pi(\sigma^2\mathbf{I} + \mathbf{X}\mathbf{C_h}\mathbf{X}^H)|} \exp(-\mathbf{y}^H(\sigma^2\mathbf{I} + \mathbf{X}\mathbf{C_h}\mathbf{X}^H)^{-1}\mathbf{y}) \quad (2.11)$$

which can then be inserted into (2.10).

2.5.2 GLRT

If the channel $\mathbf{h}$ is unknown and deterministic, i.e., $\mathbf{h}$ is not characterized by a given probability density function, a general optimal receiver architecture in the sense of minimizing the detection error probability for arbitrary $\mathbf{h}$ is unknown. The procedure of the generalized likelihood-ratio test (GLRT) is then applicable [LS03]. For each hypothesis, i.e., for each signal that could possibly have been transmitted, we estimate the channel assuming the hypothesis is correct and use this estimate in the computation of a likelihood value. Thus, assuming $\mathbf{X}$ had been transmitted, we can compute the estimate

$$\begin{aligned} \hat{\mathbf{h}} &= \arg\max_{\mathbf{h}} p(\mathbf{y}|\mathbf{X};\mathbf{h}) \\ &= (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\mathbf{y} = \mathbf{X}^\dagger\mathbf{y} \end{aligned} \quad (2.12)$$

and detect $\mathbf{X}$ using

$$\begin{aligned} \hat{\mathbf{X}} &= \arg\max_{\mathbf{X}} p(\mathbf{y}|\mathbf{X};\hat{\mathbf{h}}(\mathbf{X})) \\ &= \arg\max_{\mathbf{X}} \frac{1}{|\pi\sigma^2\mathbf{I}|} \exp(-\sigma^{-2}\|\mathbf{y} - \mathbf{X}\mathbf{h}\|^2)\big|_{\mathbf{h}=\hat{\mathbf{h}}} \\ &= \arg\max_{\mathbf{X}} \exp(-\|\mathbf{y} - \mathbf{X}\mathbf{X}^\dagger\mathbf{y}\|^2) \\ &= \arg\max_{\mathbf{X}} \exp(-\|\mathbf{P}_{\mathbf{X}}^{\perp}\mathbf{y}\|^2) \\ &= \arg\max_{\mathbf{X}} \exp(\mathbf{y}^H\mathbf{P}_{\mathbf{X}}\mathbf{y}) \end{aligned} \quad (2.13)$$

where the idempotent matrices $\mathbf{P}_{\mathbf{X}} = \mathbf{X}\mathbf{X}^\dagger$ and $\mathbf{P}_{\mathbf{X}}^{\perp} = \mathbf{I} - \mathbf{P}_{\mathbf{X}}$ are the projection matrix on the column space of $\mathbf{X}$ and on its orthogonal complement, respectively. The two maximizations in (2.12) and (2.13) can of course also be considered as a joint maximization in

$$(\hat{\mathbf{X}}, \mathbf{h}) = \arg\max_{\mathbf{X},\mathbf{h}} p(\mathbf{y}|\mathbf{X};\mathbf{h}).$$

2.5.3 Comparison

Both the GLRT and the ML detector can be summarized as

$$\hat{\mathbf{X}} = \arg\max l(\mathbf{y}, \mathbf{X})$$

where the function $l(\mathbf{y}, \mathbf{X})$ for the ML receiver is set to

$$l(\mathbf{y}, \mathbf{X}) \triangleq \frac{1}{|\pi(\sigma^2\mathbf{I} + \mathbf{X}\mathbf{C}_\mathbf{h}\mathbf{X}^H)|} \exp(-\mathbf{y}^H(\sigma^2\mathbf{I} + \mathbf{X}\mathbf{C}_\mathbf{h}\mathbf{X}^H)^{-1}\mathbf{y})$$

and for the GLRT receiver to

$$l \triangleq \exp(\mathbf{y}^H\mathbf{P}_\mathbf{X}\mathbf{y}).$$

Frequently, these functions can be simplified if additional assumptions such as orthogonal signal matrices or i.i.d. fading are made. We will consider these assumptions later and keep the general form for now to apply it in the performance analysis.

A comparison between both detectors involves the comparison between two different modeling assumptions: Assuming that the unknown channel $\mathbf{h}$ is random with a known probability density (ML) or it is deterministic without further information (GLRT). Clearly, the ML receiver makes use of additional knowledge concerning the received signal: Apart from the channel distribution, the noise variance is required to be known. If the wireless channel is properly modeled using random $\mathbf{h}$ and the ML receiver has accurate information concerning its statistical parameters, it will outperform the GLRT receiver in terms of error probability. The validity of this assumption however is not necessarily always guaranteed. The GLRT receiver that makes no use of this information is apparently robust with respect to variations in these parameters.

We also want to point out that both the GLRT and the ML receiver can coincide, i.e., produce equivalent results. If the channel is i.i.d. fading, i.e., the channel covariance matrix is a scaled identity matrix, and the signals have orthonormal columns, the ML and GLRT receiver produce equivalent decisions. Such design assumptions are not uncommon in the literature on constellation design and therefore the distinction between ML and GLRT receiver is not always apparent. However, the different philosophies behind the two receiver strategies are important if these design assumptions cannot be met. We will come back to this point in our performance comparison of both the GLRT and ML receiver in Section 2.6.4.

2.6 Performance Analysis

A common criterion to assess the performance of different constellation designs is the pairwise error probability (PEP)

$$\Pr(\mathbf{X}_i \to \mathbf{X}_j) \triangleq \Pr\left(l(\mathbf{y}, \mathbf{X}_i) < l(\mathbf{y}, \mathbf{X}_j) \quad | \mathbf{X}_i\right) \tag{2.14}$$

i.e., the probability that a binary test between two signaling matrices $\mathbf{X}_j$ and $\mathbf{X}_i$ decides in favor of $\mathbf{X}_j$ when $\mathbf{X}_i$ was transmitted. For future reference, we define

$$\Lambda(\mathbf{y}, \mathbf{X}_i, \mathbf{X}_j) \triangleq \ln \frac{l(\mathbf{y}, \mathbf{X}_j)}{l(\mathbf{y}, \mathbf{X}_i)}$$

and reformulate (2.14) as

$$\Pr\left(\Lambda(\mathbf{y}, \mathbf{X}_i, \mathbf{X}_j) > 0 \quad | \quad \mathbf{X}_i\right). \tag{2.15}$$

Whereas for the ML receiver the evaluation of (2.15) involves averaging over the Gaussian distribution assumed for the noise and the channel, the corresponding calculation for the GLRT receiver is less obvious. The probability in (2.14) can be computed for any given channel $\mathbf{h}$, thereby averaging over the noise distribution only. Since our ultimate goal is the design of signal matrices $\mathbf{X}$ that work for an unknown channel, a dependency on the specific channel vector is clearly undesirable. Therefore, we assume a distribution for $\mathbf{h}$ also in the case of the GLRT receiver in order to determine average performance. In that sense, the evaluation of (2.15) is similar for both ML and GLRT with the difference being in the computation of Λ, but not in the assumptions on channel and noise.

In the following, we will discuss three approaches to compute the PEP. The first approach based on numerical integration provides the exact solution. The second approach computes an approximation of the PEP if the SNR becomes large ($\sigma^2 \to 0$). The third approach is an upper bound on the PEP using the Chernoff bound which in turn can be approximated for high SNR. The discussion of these analysis approaches involves some derivations with technical detail. A summary is provided in Section 2.6.4.

2.6.1 Numerical Integration

A numerical solution to compute (2.15) has a common structure for both ML and GLRT receiver. We will discuss both receiver types in turn.

- **ML receiver.** Defining

$$\mathbf{C}_i \triangleq \sigma^2\mathbf{I} + \mathbf{X}_i\mathbf{C}_\mathbf{h}\mathbf{X}_i^H$$
$$\mathbf{C}_j \triangleq \sigma^2\mathbf{I} + \mathbf{X}_j\mathbf{C}_\mathbf{h}\mathbf{X}_j^H$$

we obtain

$$\Lambda(\mathbf{y}, \mathbf{X}_i, \mathbf{X}_j) = \mathbf{y}^H(\mathbf{C}_i^{-1} - \mathbf{C}_j^{-1})\mathbf{y} - \ln\frac{|\mathbf{C}_j|}{|\mathbf{C}_i|} \tag{2.16}$$

where conditioned on the transmission of $\mathbf{X}_i$ the received vector $\mathbf{y}$ is Gaussian with zero mean and covariance $\mathbf{C}_i$. Since $\mathbf{C}_i$ is positive definite, it can be factorized into $\mathbf{C}_i^{\frac{1}{2}}\mathbf{C}_i^{\frac{1}{2}}$ with Hermitian invertible $\mathbf{C}_i^{\frac{1}{2}}$. The vector $\boldsymbol{\eta} \triangleq \mathbf{C}_i^{-\frac{1}{2}}\mathbf{y}$ has elements η_k which are i.i.d. complex Gaussian variables with variance 1. We can then rewrite

$$\begin{aligned}\Lambda(\mathbf{y}, \mathbf{X}_i, \mathbf{X}_j) &= \underbrace{\mathbf{y}^H\mathbf{C}_i^{-\frac{1}{2}H}}_{\boldsymbol{\eta}^H}\mathbf{C}_i^{\frac{1}{2}}(\mathbf{C}_i^{-1} - \mathbf{C}_j^{-1})\mathbf{C}_i^{\frac{1}{2}}\underbrace{\mathbf{C}_i^{-\frac{1}{2}}\mathbf{y}}_{\boldsymbol{\eta}} - \ln\frac{|\mathbf{C}_j|}{|\mathbf{C}_i|} \\ &= \sum_k \lambda_k^{\mathrm{ML}}\|\eta_k\|^2 - c_{ij}\end{aligned}$$

where λ_k^{ML} are the eigenvalues of the matrix $\mathbf{C}_i^{\frac{1}{2}}(\mathbf{C}_i^{-1} - \mathbf{C}_j^{-1})\mathbf{C}_i^{\frac{1}{2}}$ and $c_{ij} \triangleq \ln\frac{|\mathbf{C}_j|}{|\mathbf{C}_i|}$.

Thus, the PEP can be reformulated as the probability that a weighted sum of independent central χ^2 variables with two degrees of freedom is larger than a given constant. The same structure (even simpler) results for the

- **GLRT receiver**. We obtain

$$\begin{aligned}\Lambda(\mathbf{y}, \mathbf{X}_i, \mathbf{X}_j) &= \mathbf{y}^H(\mathbf{P}_{\mathbf{X}_j} - \mathbf{P}_{\mathbf{X}_i})\mathbf{y} \\ &= \boldsymbol{\eta}^H\mathbf{C}_i^{\frac{1}{2}}(\mathbf{P}_{\mathbf{X}_j} - \mathbf{P}_{\mathbf{X}_i})\mathbf{C}_i^{\frac{1}{2}}\boldsymbol{\eta} \\ &= \sum_k \lambda_k^{\mathrm{GLRT}}\|\eta_k\|^2\end{aligned}$$

where $\lambda_k^{\mathrm{GLRT}}$ are the eigenvalues of the matrix $\mathbf{C}_i^{\frac{1}{2}}(\mathbf{P}_{\mathbf{X}_j} - \mathbf{P}_{\mathbf{X}_i})\mathbf{C}_i^{\frac{1}{2}}$.

Thus, for both the GLRT and ML receiver, we have to solve a problem of the type

$$\Pr\left(\sum_k \lambda_k\chi_k^2 > x\right)$$

where the λ_k and x are known ($x = 0$ for the GLRT and $x = c_{ij}$ for the ML receiver) and the central χ_k^2 variables have two degrees of freedom. This is a recurring problem in the literature on error analysis for digital communication systems with both known and unknown channel at the receiver, see, e.g., [TB02, SFG02, Pro95]. We focus here on the numerical solution to the generic problem which is mentioned in [MP92] pointing to [Imh61] where

$$\Pr\left(\sum_k \chi_k^2 \|\eta_k\|^2 > x\right) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin\theta(u)}{u\rho(u)} du \tag{2.17}$$

with

$$\theta(u) \triangleq \frac{1}{2}\left(2\sum_{k=1}^{T'} \tan^{-1}(\lambda_k u)\right) - \frac{1}{2}xu$$

$$\rho(u) \triangleq \prod_{k=1}^{T'} (1+\lambda_k^2 u^2)^{\frac{1}{2}}$$

and the integrand in (2.17) is well defined for $u \to 0$. The advantage of this result is that it is exact and numerically stable for computation of the PEP for SNR values of $1/\sigma^2$ as high as 30dB. Moreover, it can easily be extended for the case that $\mathbf{h} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{h}, \mathbf{C}_\mathbf{h})$, i.e., a situation where the channel realization has a non-zero mean which is a standard way of modeling line-of-sight components. However, it does not allow much insight into the behavior of the PEP dependent on the structure of the signals. Therefore, approximations and bounds are useful.

2.6.2 Asymptotic Analysis

The behavior of the PEP in the case of high SNR, i.e. $\sigma^2 \to 0$ is of particular interest because design criteria based on the PEP such as the union bound on the error probability lead to meaningful results for high SNR. Moreover, a communication system should perform well in favorable conditions, i.e., high SNR, and therefore consideration of this area is important. Brehler and Varanasi investigated in [BV01] the limiting behavior of the PEP for both the ML and the GLRT receiver. Their approach is based on an approximation of the non-zero eigenvalues λ_k in the case $\sigma^2 \to 0$. The results are cited here in the notation of this thesis.

- **ML receiver.** Assuming full rank of the matrix $[\mathbf{X}_i \quad \mathbf{X}_j]^H[\mathbf{X}_i \quad \mathbf{X}_j]$, it is shown in [BV01, Proposition 4] that

$$\lim_{\sigma^2\to 0}(\sigma^2)^{-M}\Pr(\mathbf{X}_i \to \mathbf{X}_j) = \frac{K_{ij}}{|\mathbf{C_h}||\mathbf{X}_i^H\mathbf{P}^{\perp}_{\mathbf{X}_j}\mathbf{X}_i|}$$

where K_{ij} is

$$K_{ij} = \frac{|\mathbf{X}_i^H\mathbf{X}_i|}{|\mathbf{X}_j^H\mathbf{X}_j|}\sum_{k=0}^{M-1}\begin{pmatrix} 2M-1-k \\ M \end{pmatrix}\frac{1}{k!}\left(\ln\frac{|\mathbf{X}_j^H\mathbf{X}_j|}{|\mathbf{X}_i^H\mathbf{X}_i|}\right)^k \tag{2.18}$$

if $|\mathbf{X}_i^H\mathbf{X}_i| \le |\mathbf{X}_j^H\mathbf{X}_j|$. Otherwise we have

$$K_{ij} = \sum_{k=0}^{M}\binom{2M-1-k}{M-1}\frac{1}{k!}\left(\ln\frac{|\mathbf{X}_i^H\mathbf{X}_i|}{|\mathbf{X}_j^H\mathbf{X}_j|}\right)^k. \tag{2.19}$$

If $|\mathbf{X}_i^H\mathbf{X}_i| = |\mathbf{X}_j^H\mathbf{X}_j|$, the convention $0^0 = 1$ is used and the sums in (2.18) and (2.19) reduce to a single term.

- **GLRT receiver**. From [BV01, Proposition 6], we obtain

$$\lim_{\sigma^2\to 0}(\sigma^2)^{-M}\Pr(\mathbf{X}_i \to \mathbf{X}_j) = \frac{\begin{pmatrix} 2M-1 \\ M \end{pmatrix}}{|\mathbf{C_h}||\mathbf{X}_i^H\mathbf{P}^{\perp}_{\mathbf{X}_j}\mathbf{X}_i|}. \tag{2.20}$$

2.6.3 Chernoff Bound

The third method of analyzing the PEP is based on the Chernoff bound. In the case for signal matrices with orthogonal columns, a Chernoff bound for ML detection was presented in [HM00]. Later, Dogandzic published much more general results avoiding this restriction in [Dog03]. A bound for the GLRT receiver was also presented in [LS03]. In the following derivations, we present the general bound for the ML and GLRT receiver, approximate this bound for high SNR and optimize a remaining parameter to minimize the high SNR approximation.

The quantity

$$\Lambda(\mathbf{y}) = \ln\frac{l(\mathbf{y},\mathbf{X}_j)}{l(\mathbf{y},\mathbf{X}_i)}$$

is a random variable because it is a function of the random variable $\mathbf{y}$. The PEP can then be expressed as

$$\begin{aligned}\Pr(\mathbf{X}_i \to \mathbf{X}_j) &= \Pr(\Lambda(\mathbf{y}) > 0 \quad | \quad \mathbf{X}_i \text{ transmitted}) \\ &= \int_{\Lambda(\mathbf{y})>0} p_{\mathbf{y}|\mathbf{X}_i}(\mathbf{y}|\mathbf{X}_i) d\mathbf{y}\end{aligned}$$

Now for $s \geq 0$ we can upperbound the PEP as

$$\begin{aligned}\Pr(\mathbf{X}_i \to \mathbf{X}_j) &\leq \int_{\Lambda(\mathbf{y})>0} e^{s\Lambda(\mathbf{y})} p_{\mathbf{y},\mathbf{X}_i}(\mathbf{y}|\mathbf{X}_i) d\mathbf{y} \\ &\leq \int e^{s\Lambda(\mathbf{y})} p_{\mathbf{y}|\mathbf{X}_i}(\mathbf{y}|\mathbf{X}_i) d\mathbf{y} \qquad (2.21)\end{aligned}$$

where the last integral is over the entire definition space of $\mathbf{y}$.

- **ML receiver.** The expression for Λ_{ML} was given in (2.16). Inserting this result into (2.21) and using $p_{\mathbf{y}|\mathbf{X}_i}$ from (2.11), we obtain the upper bound

$$\begin{aligned}&\Pr(\mathbf{X}_i \to \mathbf{X}_j) \qquad (2.22) \\ &\quad \leq \int \frac{|\pi\mathbf{C}_i|^s}{|\pi\mathbf{C}_j|^s} \frac{1}{|\pi\mathbf{C}_i|} \exp\left(-\mathbf{y}^H (s\mathbf{C}_j^{-1} - s\mathbf{C}_i^{-1} + \mathbf{C}_i^{-1})\mathbf{y}\right) d\mathbf{y} \\ &\quad = \frac{1}{|\pi\mathbf{C}_j|^s |\pi\mathbf{C}_i|^{1-s}} \int \exp\left(-\mathbf{y}^H (s\mathbf{C}_j^{-1} + (1-s)\mathbf{C}_i^{-1})\mathbf{y}\right) d\mathbf{y}\end{aligned}$$

  Since integrating the multivariate Gaussian density over the entire definition space results in unity, it is known that for any positive definite matrix $\mathbf{B}$ it holds that

$$\int \exp\left(-\mathbf{y}^H \mathbf{B}\mathbf{y}\right) d\mathbf{y} = |\pi\mathbf{B}^{-1}| \qquad (2.23)$$

  Now since $\mathbf{C}_j$ and $\mathbf{C}_i$ are positive definite, it holds that the matrix $s\mathbf{C}_j^{-1} + (1-s)\mathbf{C}_i^{-1}$ is positive definite for $0 \leq s \leq 1$ and thus we obtain by using (2.23) in (2.22)

$$\Pr(\mathbf{X}_i \to \mathbf{X}_j) \leq \frac{|\pi(s\mathbf{C}_j^{-1} + (1-s)\mathbf{C}_i^{-1})^{-1}|}{|\pi\mathbf{C}_j|^s |\pi\mathbf{C}_i|^{1-s}} \qquad (2.24)$$

  After some algebraic manipulations (see Appendix 2.A), we obtain

$$\Pr(\mathbf{X}_i \to \mathbf{X}_j) \leq \sigma^{2M} Q(\mathbf{X}_i, \mathbf{X}_j, s, \sigma^2)$$

where

$$Q(\mathbf{X}_i, \mathbf{X}_j, s, \sigma^2) \triangleq \frac{|\sigma^2\mathbf{I} + \mathbf{C_h}\mathbf{X}_i^H\mathbf{X}_i|^s |\sigma^2\mathbf{I} + \mathbf{C_h}\mathbf{X}_j^H\mathbf{X}_j|^{1-s}}{\left|\sigma^2\mathbf{I} + \begin{pmatrix} s\mathbf{C_h} & \\ & (1-s)\mathbf{C_h} \end{pmatrix} \begin{pmatrix} \mathbf{X}_i^H\mathbf{X}_i & \mathbf{X}_i^H\mathbf{X}_j \\ \mathbf{X}_j^H\mathbf{X}_i & \mathbf{X}_j^H\mathbf{X}_j \end{pmatrix}\right|} \tag{2.25}$$

For high SNR $\sigma^2 \to 0$, we get

$$\begin{aligned} \lim_{\sigma^2\to 0} Q(\mathbf{X}_i, \mathbf{X}_j, s, \sigma^2) &= \frac{|\mathbf{X}_i^H\mathbf{X}_i|^s |\mathbf{X}_j^H\mathbf{X}_j|^{1-s}}{s^M(1-s)^M|\mathbf{C_h}| \begin{vmatrix} \mathbf{X}_i^H\mathbf{X}_i & \mathbf{X}_i^H\mathbf{X}_j \\ \mathbf{X}_j^H\mathbf{X}_i & \mathbf{X}_j^H\mathbf{X}_j \end{vmatrix}} \\ &= \frac{s^{-M}(1-s)^{-M}\left(\frac{|\mathbf{X}_i^H\mathbf{X}_i|}{|\mathbf{X}_j^H\mathbf{X}_j|}\right)^s}{|\mathbf{C_h}||\mathbf{X}_i^H\mathbf{P}_{\mathbf{X}_j}^{\perp}\mathbf{X}_i|}. \end{aligned}$$

where we used (A.9) and assumed that the matrix

$$[\mathbf{X}_i \quad \mathbf{X}_j]^H[\mathbf{X}_i \quad \mathbf{X}_j] = \begin{pmatrix} \mathbf{X}_i^H\mathbf{X}_i & \mathbf{X}_i^H\mathbf{X}_j \\ \mathbf{X}_j^H\mathbf{X}_i & \mathbf{X}_j^H\mathbf{X}_j \end{pmatrix}$$

has full column rank. We can now choose the parameter $0 \le s < 1$ such that the bound is minimized. Setting the derivative with respect to s to zero, we obtain for the optimal s^* that

$$\frac{1-2s^*}{s^*(1-s^*)} = \frac{1}{M}\ln\frac{|\mathbf{X}_i^H\mathbf{X}_i|}{|\mathbf{X}_j^H\mathbf{X}_j|} \triangleq Z_{ij}$$

Apparently, for $|\mathbf{X}_i^H\mathbf{X}_i| = |\mathbf{X}_j^H\mathbf{X}_j|$ implying $Z_{ij} = 0$ we obtain as only solution $s^* = 1/2$. This is an important special case, in particular for signals with orthogonal columns (see Section 2.6.4) The solution for general Z_{ij} in the allowed interval $0 \le s^* < 1$ is

$$s^* = \frac{1}{2}\left(1 - \frac{Z_{ij}}{2 + \sqrt{Z_{ij}^2 + 4}}\right)$$

and consequently

$$\lim_{\sigma^2\to 0} Q(\mathbf{X}_i, \mathbf{X}_j, s^*, \sigma^2) = \frac{\left(2 + \sqrt{Z_{ij}^2 + 4}\right)^M e^{MZ_{ij}s^*}}{|\mathbf{C_h}||\mathbf{X}_i^H\mathbf{P}_{\mathbf{X}_j}^{\perp}\mathbf{X}_i|}$$

which is smaller than the choice for $s^* = \frac{1}{2}$ which leads to

$$\lim_{\sigma^2 \to 0} Q\left(\mathbf{X}_i, \mathbf{X}_j, s = \frac{1}{2}, \sigma^2\right) = \frac{2^{2M} e^{MZ_{ij}/2}}{|\mathbf{C_h}||\mathbf{X}_i^H \mathbf{P}_{\mathbf{X}_j}^{\perp} \mathbf{X}_i|}.$$

- **GLRT receiver.** Here we have

$$\Lambda = \ln \frac{e^{\mathbf{y}^H \mathbf{P}_{\mathbf{X}_j} \mathbf{y}}}{e^{\mathbf{y}^H \mathbf{P}_{\mathbf{X}_i} \mathbf{y}}}$$

and therefore obtain from (2.21)

$$\Pr(\mathbf{X}_i \to \mathbf{X}_j) \leq \int_{-\infty}^{\infty} \frac{1}{|\pi \mathbf{C}_i|} \exp\left(-\mathbf{y}^H (\mathbf{C}_i^{-1} - s\mathbf{P}_{\mathbf{X}_j} + s\mathbf{P}_{\mathbf{X}_i})\mathbf{y}\right) d\mathbf{y}.$$

After some algebraic manipulations that are presented in Appendix 2.B, we obtain

$$\Pr(\mathbf{X}_i \to \mathbf{X}_j) \leq \sigma^{2M} Q(\mathbf{X}_i, \mathbf{X}_j, s, \sigma^2)$$

where

$$Q(\mathbf{X}_i, \mathbf{X}_j, s, \sigma^2) \triangleq \frac{1}{|s(1-s)\mathbf{X}_i^H \mathbf{P}_{\mathbf{X}_j}^{\perp} \mathbf{X}_i \mathbf{C_h} + \sigma^2 (\mathbf{I} - s^2 \mathbf{X}_i^H \mathbf{P}_{\mathbf{X}_j}^{\perp} \mathbf{X}_i^{H\dagger})|} \tag{2.26}$$

and $0 \leq s < 1$ as before. In the high SNR region $\sigma^2 \to 0$ we obtain

$$Q(\mathbf{X}_i, \mathbf{X}_j, s, 0) = \frac{1}{|s(1-s)\mathbf{X}_i^H \mathbf{P}_{\mathbf{X}_j}^{\perp} \mathbf{X}_i||\mathbf{C_h}|}$$

which is minimized for $s = 1/2$.

2.6.4 Comparison of the Analysis Approaches

It is very useful to point out the similarities between the results presented in this section so far. In order to obtain a meaningful statement about the design of the two matrices $\mathbf{X}_i$ and $\mathbf{X}_j$, we consider the behavior of

$$\mathrm{PEP}(\mathbf{X}_i, \mathbf{X}_j) \triangleq \Pr(\mathbf{X}_i \to \mathbf{X}_j) + \Pr(\mathbf{X}_j \to \mathbf{X}_i)$$

and illustrate the presented analysis approaches in Figure 2.5 for the GLRT and in Figure 2.6 for the ML receiver, respectively. The two

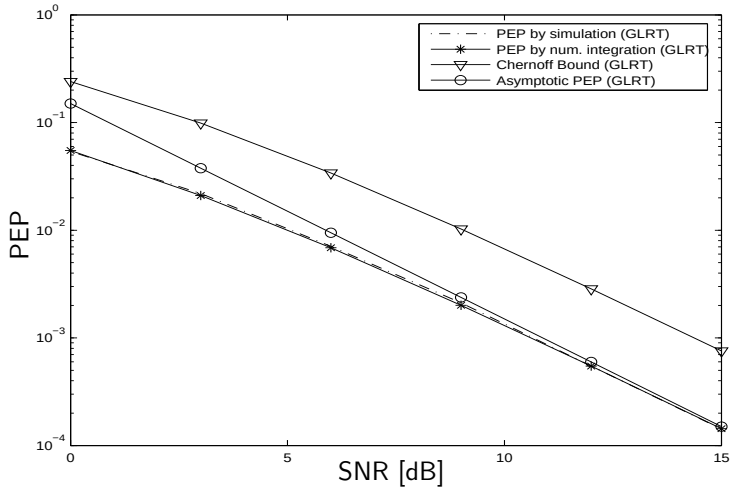


Figure 2.5: Illustration of the PEP together with its Chernoff upper bound and its approximation using asymptotic analysis for the GLRT receiver

signal matrices with $M = 2$ and $T' = 8$ were picked at random with i.i.d. elements and normalized to unit power per channel use. The channel correlation profile was generated as the square of a random matrix.

Comparing Figures 2.5 and 2.6, we note that the performance results for the ML and the GLRT receiver are very similar. In both cases, the curves for PEP obtained by Monte Carlo simulation (dashdotted) match the values computed with numerical integration very well. At an SNR of about 12 to 15 dB and higher, the value of the PEP is very well approximated by the asymptotic PEP. In contrast, the values of the Chernoff bound and of the PEP differ by a significant factor. This factor appears to be constant over the investigated SNR range. For high SNR, the slope of all curves is the same. In this regime, the asymptotic analysis established that

$$\mathrm{PEP}(\mathbf{X}_i, \mathbf{X}_j) \approx \sigma^{2^M} \frac{Q^{\mathrm{As}}(\mathbf{X}_i, \mathbf{X}_j)}{|\mathbf{C_h}|\,|[\mathbf{X}_i \quad \mathbf{X}_j]^H [\mathbf{X}_i \quad \mathbf{X}_j]|} \tag{2.27}$$

and the high SNR approximation on the Chernoff upper bound resulted

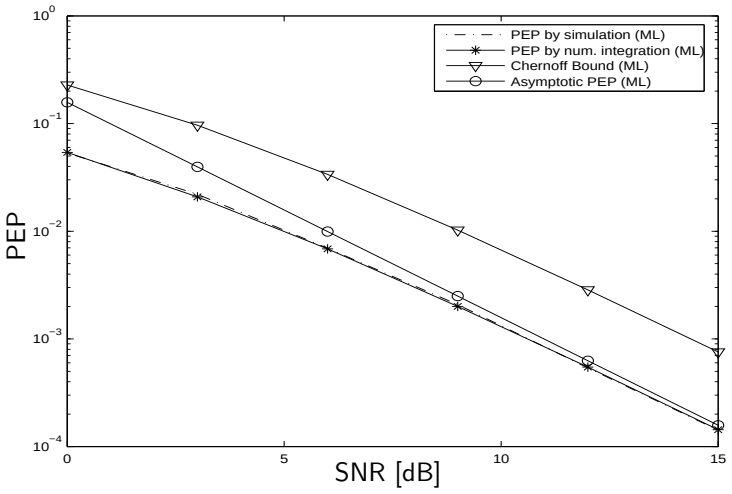


Figure 2.6: Illustration of the PEP together with its Chernoff upper bound and its approximation using asymptotic analysis for the ML receiver

in

$$\mathrm{PEP}(\mathbf{X}_i, \mathbf{X}_j) \lesssim (\sigma^2)^M \frac{Q^{\mathrm{CB}}(\mathbf{X}_i, \mathbf{X}_j)}{|\mathbf{C_h}|\,|[\mathbf{X}_i \quad \mathbf{X}_j]^H [\mathbf{X}_i \quad \mathbf{X}_j]|} \tag{2.28}$$

where the values of $Q(\mathbf{X}_i, \mathbf{X}_j)$ for the applied receiver are given in Table 2.1 for both the ML and GLRT receiver. These results rely on the assumption that

$$|[\mathbf{X}_i \quad \mathbf{X}_j]^H [\mathbf{X}_i \quad \mathbf{X}_j]| > 0, \tag{2.29}$$

i.e., that $[\mathbf{X}_i \quad \mathbf{X}_j]$ has full column rank. Under this condition it is guaranteed by the exponent of σ^2 in (2.27) and (2.28) that the slope of the resulting PEP curve as a function of the SNR $1/\sigma^2$ on a double logarithmic plot is M, i.e., a diversity level of M is achieved. In other words, a 10 dB increase in SNR will lead to a reduction in PEP by a factor of 10^M. If the signals $\mathbf{X}_i$ and $\mathbf{X}_j$ are designed for low error probability in the high SNR regime, the highest possible diversity order is usually required and can be fulfilled by satisfying the rank criterion (2.29). Moreover, in the high SNR region, the influence of the channel correlation profile represented in $\mathbf{C_h}$ is decoupled from the signal design as long as $\mathbf{C_h}$ is full

receiver type	analysis	$Q(\mathbf{X}_i, \mathbf{X}_j)$ in (2.27) and (2.28)
ML	asymptotic analysis	$\lvert\mathbf{X}_j^H\mathbf{X}_j\rvert \sum_{k=0}^{T'} \binom{2M-k}{M} \frac{(MZ_{ij})^k}{k!}, \; Z_{ij} \geq 0$
ML	Chernoff bound	$2\frac{\left(2+\sqrt{Z_{ij}^2+4}\right)^M \sqrt{\lvert\mathbf{X}_i^H\mathbf{X}_i\rvert\lvert\mathbf{X}_j^H\mathbf{X}_j\rvert}}{\exp\left(\frac{Z_{ij}^2 M/2}{2+\sqrt{Z_{ij}^2+4}}\right)}$
GLRT	asymptotic analysis	$\left(\begin{array}{c} 2M-1 \\ M \end{array}\right) (\lvert\mathbf{X}_i^H\mathbf{X}_i\rvert + \lvert\mathbf{X}_j^H\mathbf{X}_j\rvert)$
GLRT	Chernoff bound	$2^{2M}(\lvert\mathbf{X}_i^H\mathbf{X}_i\rvert + \lvert\mathbf{X}_j^H\mathbf{X}_j\rvert)$

Table 2.1: Overview of the analysis approaches

rank (as is generally assumed in this thesis). Further design rules are then targeted towards minimization of $Q(\mathbf{X}_i, \mathbf{X}_j)$ and can be considered independent on the channel correlation profile.

Using Table 2.1, we can point out the mentioned factor between the Chernoff upper bound and the asymptotic expression in the high SNR regime. Assuming $|\mathbf{X}_i^H\mathbf{X}_i| = |\mathbf{X}_j^H\mathbf{X}_j|$, the Chernoff bound is off by a factor of $2^{2M}/\binom{2M-1}{M}$ for both the ML and the GLRT receiver (note that $\binom{2M}{M} = 2\binom{2M-1}{M}$), see also [BV01].

As mentioned in Section 2.6.3, the case that $|\mathbf{X}_i^H\mathbf{X}_i| = |\mathbf{X}_j^H\mathbf{X}_j|$ implying $c_{ij} = Z_{ij} = 0$ deserves additional attention. Then, asymptotic analysis leads to the same resulting PEP for both GLRT and ML receiver, i.e. their performance is asymptotically equivalent. This additional condition is met in the special case of signal with orthonormal columns

$$\mathbf{X}_i^H\mathbf{X}_i = \mathbf{X}_j^H\mathbf{X}_j = \frac{T'}{M}\mathbf{I} \tag{2.30}$$

which is of particular interest since for high SNR or $T' \gg M$ and

i.i.d. channel fading, orthonormal signals are optimal in terms of capacity [MH99] in the absence of receiver CSI. Therefore, condition (2.30) has been imposed in several approaches on designing signals, e.g., [HMR+00, MBV02, ARU01, BB02]. Another advantage of signals satisfying (2.30) is the simplicity of the resulting PEP expression for the ML receiver (the values in Table 2.1 reduce to a constant) which is important for the purpose of obtaining a simple design criterion that can easily be evaluated. If condition (2.30) cannot be imposed (if the signals have specific substructure, see Chapter 3, such that the optimality statement does not necessarily hold), the PEP expressions for the GLRT receiver are much simpler than for the ML receiver. Apart from the different philosophies behind the receiver choice, the design criterion is easier to handle for GLRT, providing an additional motivation for its usage.

2.6.5 Conclusion on Performance Analysis

We have presented in this section three different approaches to analyze the PEP dependent on the signal design. Whereas the first approach based on numerical integration provides exact results, it does not allow much insight on the design rules that should be imposed on the signaling matrices. Such rules are easily derived from both the second and third approach which approximate the PEP for high SNR or upperbound it.

In the course of this thesis, not all results concerning the PEP were available from the start. We became aware of the numerical solution from Section 2.6.1 first and designed the optimization method accordingly, i.e., based on a preselected symbol alphabet (see Chapter 3). Later, the asymptotic analysis in [BV01] allowed the application of a design method based on gradient search. The Chernoff bounds were first derived for signal matrices with orthogonal columns (see [HM00]) but could not be used for signaling matrices not constrained on unitary columns. Such a bound was presented in [LS03] for the GLRT receiver and in a very general form for the ML receiver in [Dog03]. Even though the asymptotic analysis appears to lead to a much tighter approximations, the Chernoff bound as such is still helpful to guarantee the upper bound on the PEP.

2.7 Summary

We presented in this chapter two standard models for both a single carrier as well as a multicarrier system. The resulting data models were found to

be very similar and allowed a unified description of two receiver strategies together with their subsequent PEP analysis. The results derived in this chapter form the basis for the criteria and methods for constellation design in subsequent chapters.

Appendix 2.A Chernoff Bound for the ML Receiver.

Continuing from (2.24), we have

$$\begin{aligned}
&\frac{|\pi(s\mathbf{C}_j^{-1}+(1-s)\mathbf{C}_i^{-1})^{-1}|}{|\pi\mathbf{C}_j|^s|\pi\mathbf{C}_i|^{1-s}} \\
&\quad= \frac{|\mathbf{C}_j(s\mathbf{C}_i+(1-s)\mathbf{C}_j)^{-1}\mathbf{C}_i|}{|\mathbf{C}_j|^s|\mathbf{C}_i|^{1-s}} \\
&\quad= \frac{|\mathbf{C}_i|^s|\mathbf{C}_j|^{1-s}}{|s\mathbf{C}_i+(1-s)\mathbf{C}_j|} \\
&\quad= \frac{|\sigma^2\mathbf{I}+\mathbf{X}_i\mathbf{C}_h\mathbf{X}_i^H|^s|\sigma^2\mathbf{I}+\mathbf{X}_j\mathbf{C}_\mathbf{h}\mathbf{X}_j^H|^{1-s}}{\left|\sigma^2\mathbf{I}+\begin{pmatrix}\mathbf{X}_i & \mathbf{X}_j\end{pmatrix}\begin{pmatrix}s\mathbf{C}_\mathbf{h} & \\ & (1-s)\mathbf{C}_\mathbf{h}\end{pmatrix}\begin{pmatrix}\mathbf{X}_i^H \\ \mathbf{X}_j^H\end{pmatrix}\right|}
\end{aligned}$$

After applying relation (A.8), we obtain

$$\begin{aligned}
&= \frac{(\sigma^2)^{T'-M}|\sigma^2\mathbf{I}+\mathbf{C}_h\mathbf{X}_i^H\mathbf{X}_i|^s|\sigma^2\mathbf{I}+\mathbf{C}_h\mathbf{X}_j^H\mathbf{X}_j|^{1-s}}{(\sigma^2)^{T'-2M}\left|\sigma^2\mathbf{I}+\begin{pmatrix}s\mathbf{C}_\mathbf{h} & \\ & (1-s)\mathbf{C}_\mathbf{h}\end{pmatrix}\begin{pmatrix}\mathbf{X}_i^H\mathbf{X}_i & \mathbf{X}_i^H\mathbf{X}_j \\ \mathbf{X}_j^H\mathbf{X}_i & \mathbf{X}_j^H\mathbf{X}_j\end{pmatrix}\right|} \\
&= (\sigma^2)^M Q_{\mathrm{ML}}(\mathbf{X}_i,\mathbf{X}_j,s,\sigma^2)
\end{aligned}$$

where $Q_{\mathrm{ML}}(\mathbf{X}_i,\mathbf{X}_j,s,\sigma^2)$ is defined in (2.25).

Appendix 2.B Chernoff Bound for the GLRT Receiver

We start from

$$\Pr(\mathbf{X}_i\to\mathbf{X}_j)\le\frac{1}{|\pi\mathbf{C}_i|}\int_{-\infty}^{\infty}\exp\left(-\mathbf{y}^H(\mathbf{C}_i^{-1}+s\mathbf{P}_{\mathbf{X}_j}^{\perp}-s\mathbf{P}_{\mathbf{X}_i}^{\perp})\mathbf{y}\right)d\mathbf{y}$$

and rewrite $\mathbf{C}_i^{-1}$ by applying the matrix inversion lemma 6 twice as

$$\begin{aligned}
\mathbf{C}_i^{-1} &= \frac{1}{\sigma^2}\mathbf{I}-\frac{1}{\sigma^4}\mathbf{X}_i\left(\frac{1}{\sigma^2}\mathbf{X}_i^H\mathbf{X}_i+\mathbf{C}_\mathbf{h}^{-1}\right)^{-1}\mathbf{X}_i^H \\
&= \frac{1}{\sigma^2}\left(\mathbf{I}-\mathbf{X}_i(\mathbf{X}_i^H\mathbf{X}_i+\sigma^2\mathbf{C}_\mathbf{h}^{-1})^{-1}\mathbf{X}_i^H\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sigma^2}\left(\mathbf{I} - \mathbf{X}_i\right. \\
&\quad \left. \cdot \left((\mathbf{X}_i^H\mathbf{X}_i)^{-1} - (\mathbf{X}_i^H\mathbf{X}_i)^{-1}((\mathbf{X}_i^H\mathbf{X}_i)^{-1} + \sigma^{-2}\mathbf{C}_{\mathbf{h}}^{-1})^{-1}(\mathbf{X}_i^H\mathbf{X}_i)^{-1}\right)\mathbf{X}_i^H\right) \\
&= \frac{1}{\sigma^2}\mathbf{P}_{\mathbf{X}_i}^{\perp} + \mathbf{X}_i(\sigma^2\mathbf{X}_i^H\mathbf{X}_i + \mathbf{X}_i^H\mathbf{X}_i\mathbf{C}_{\mathbf{h}}^{-1}\mathbf{X}_i^H\mathbf{X}_i)^{-1}\mathbf{X}_i^H
\end{aligned}
$$

Therefore, the matrix

$$
\mathbf{C}_i^{-1} + s\mathbf{P}_{\mathbf{X}_i} - s\mathbf{P}_{\mathbf{X}_j} = \mathbf{C}_i^{-1} - s\mathbf{P}_{\mathbf{X}_i}^{\perp} + s\mathbf{P}_{\mathbf{X}_j}^{\perp}
$$

is positive definite for $0 \leq s < \sigma^{-2}$. Setting $s = \tilde{s}\sigma^{-2}$, $0 \leq \tilde{s} < 1$ we obtain

$$
\begin{aligned}
&\Pr(\mathbf{X}_i \to \mathbf{X}_j) \\
&\quad \leq \frac{1}{|\pi\mathbf{C}_i|}\int \exp\left(-\mathbf{y}^H\left(\mathbf{C}_i^{-1} - s\mathbf{P}_{\mathbf{X}_i}^{\perp} + s\mathbf{P}_{\mathbf{X}_j}^{\perp}\right)\mathbf{y}\right) d\mathbf{y} \\
&\quad = \frac{|\pi(\mathbf{C}_i^{-1} + \tilde{s}\sigma^{-2}(\mathbf{P}_{\mathbf{X}_j}^{\perp} - \mathbf{P}_{\mathbf{X}_i}^{\perp}))^{-1}|}{|\pi\mathbf{C}_i|} \\
&\quad = \frac{1}{\left|\mathbf{I} + \tilde{s}\sigma^{-2}(\mathbf{P}_{\mathbf{X}_i} - \mathbf{P}_{\mathbf{X}_j})\mathbf{C}_i\right|} \\
&\quad = \frac{1}{\left|\mathbf{I} + \tilde{s}(\mathbf{P}_{\mathbf{X}_i} - \mathbf{P}_{\mathbf{X}_j}) + \tilde{s}\sigma^{-2}\mathbf{P}_{\mathbf{X}_j}^{\perp}\mathbf{X}_i\mathbf{C}_{\mathbf{h}}\mathbf{X}_i^H\right|} \\
&\quad = \frac{1}{|\mathbf{I} - \tilde{s}\mathbf{P}_{\mathbf{X}_j}||\mathbf{I} + (\mathbf{I} - \tilde{s}\mathbf{P}_{\mathbf{X}_j})^{-1}(\tilde{s}\mathbf{P}_{\mathbf{X}_i} + \tilde{s}\sigma^{-2}\mathbf{P}_{\mathbf{X}_j}^{\perp}\mathbf{X}_i\mathbf{C}_{\mathbf{h}}\mathbf{X}_i^H)|}
\end{aligned}
$$

Since $\mathbf{P}_{\mathbf{X}_j}$ has eigenvalue 1 with multiplicity M and eigenvalue 0 with multiplicity $T' - M$, it is clear that $|\mathbf{I} - \tilde{s}\mathbf{P}_{\mathbf{X}_j}| = (1-\tilde{s})^M$. Moreover, it is not hard to show that $(\mathbf{I} - \tilde{s}\mathbf{P}_{\mathbf{X}_j})^{-1} = \frac{1}{1-\tilde{s}}\mathbf{I} - \frac{\tilde{s}}{1-\tilde{s}}\mathbf{P}_{\mathbf{X}_j}^{\perp}$. We therefore obtain after replacing $\mathbf{P}_{\mathbf{X}_i} = \mathbf{X}_i^{H\dagger}\mathbf{X}_i^H$

$$
\begin{aligned}
&\Pr(\mathbf{X}_i \to \mathbf{X}_j) \\
&\quad \leq \frac{1}{(1-\tilde{s})^M|\mathbf{I} + (\frac{1}{1-\tilde{s}}\mathbf{I} - \frac{\tilde{s}}{1-\tilde{s}}\mathbf{P}_{\mathbf{X}_j}^{\perp})(\tilde{s}\mathbf{X}_i^{H\dagger} + \tilde{s}\sigma^{-2}\mathbf{P}_{\mathbf{X}_j}^{\perp}\mathbf{X}_i\mathbf{C}_{\mathbf{h}})\mathbf{X}_i^H|} \\
&\quad = \frac{1}{(1-\tilde{s})^M|\mathbf{I} + \mathbf{X}_i^H(\frac{1}{1-\tilde{s}}\mathbf{I} - \frac{\tilde{s}}{1-\tilde{s}}\mathbf{P}_{\mathbf{X}_j}^{\perp})(\tilde{s}\mathbf{X}_i^{H\dagger} + \tilde{s}\sigma^{-2}\mathbf{P}_{\mathbf{X}_j}^{\perp}\mathbf{X}_i\mathbf{C}_{\mathbf{h}})|} \\
&\quad = \frac{1}{|(1-\tilde{s})\mathbf{I} + \mathbf{X}_i^H(\mathbf{I} - \tilde{s}\mathbf{P}_{\mathbf{X}_j}^{\perp})(\tilde{s}\mathbf{X}_i^{H\dagger} + \tilde{s}\sigma^{-2}\mathbf{P}_{\mathbf{X}_j}^{\perp}\mathbf{X}_i\mathbf{C}_{\mathbf{h}})|}
\end{aligned}
$$

$$= \frac{1}{|\mathbf{I} - \tilde{s}^2 \mathbf{X}_i^H \mathbf{P}_{\mathbf{X}_j}^{\perp} \mathbf{X}_i^{H\dagger} + \sigma^{-2}\tilde{s}(1-\tilde{s})\mathbf{X}_i^H \mathbf{P}_{\mathbf{X}_j}^{\perp} \mathbf{X}_i \mathbf{C}_{\mathbf{h}}|}$$
$$= (\sigma^2)^M Q_{\mathrm{GLRT}}(\mathbf{X}_i, \mathbf{X}_j, \tilde{s}, \sigma^2)$$

where $Q_{\mathrm{GLRT}}(\mathbf{X}_i, \mathbf{X}_j, \tilde{s}, \sigma^2)$ is defined in (2.26). We can now just as well replace s with $\tilde{s}$ as used in the main text.

Chapter 3

Design in the Time Domain

We now start attacking the problem of designing signal matrices that are optimized for data transmission over unknown channels. As for any optimization problem, a criterion is needed in order to distinguish "good" designs from "bad" ones. Moreover, the search space, i.e., the range of parameters that can be varied, needs to be defined. In general, it is not uncommon that analytical solutions for a given optimization problem are unknown. In that case, algorithms searching for (at least approximately or locally) optimal solutions have to be applied. Usually, the complexity requirements of the optimization algorithm limit the size of problems that can be solved using computer optimization. Sometimes, structural properties or insight into the choice of parameters for the given specific problem can be exploited to adapt a standard algorithm in order to reduce the required computer resources. The size of problems that can be solved can then be extended.

The optimization problems discussed in this chapter are examples of this general methodology. Two different choices of the parameter space are considered. In the first case, the symbols are drawn from a discrete set which is restricted in all examples to BPSK symbols. In the second case, the symbols are unconstrained complex-valued up to a power normalization over the entire constellation. Due to the different parameter spaces, the applied optimization methods are significantly different. Moreover, our focus of comparison is different for both methods. In the first case,

we compare the obtained constellations with standard constellations in BPSK symbols whose design contains redundancy optimized either for channel estimation or error-correction. The second approach is targeting directly the conditions for the exploitation of maximum diversity gain and optimal coding gain in the high SNR region which is particularly important for multiple transmitter antennas.

A design criterion that is common to both approaches is defined in Section 3.1. The detailed descriptions of both approaches in Sections 3.2 and 3.3, respectively follow a similar format. After the presentation of the optimization method, some standard schemes "off the shelf" are reviewed and their performance evaluated in comparison with the resulting constellations of this chapter. A comparison of both approaches together with a summary is given in Section 3.4.

3.1 Design Criterion and Objective Function

We have derived in Section 2.3.3 the data model for a single carrier system

$$\mathbf{Y} = \bar{\mathbf{S}}\mathbf{H} + \mathbf{W}$$

where $\mathbf{Y}$ represents the received symbols, $\mathbf{H}$ the unknown channel coefficients and $\mathbf{W}$ additive noise. The equivalent signal matrix $\bar{\mathbf{S}}$ is related to the transmitted symbols in $\mathbf{S}$ via the constraint (2.7). Our goal is to determine a constellation of Z signal matrices $\mathbb{S} = \{\mathbf{S}_0, \ldots, \mathbf{S}_{Z-1}\}$ that can be used for data transmission through the unknown channel. As a design criterion, we attempt to design the signals in such a way that the probability $\Pr(\hat{\mathbf{S}} \neq \mathbf{S})$ of false detection of the transmitted signal matrix in the presence of noise is minimized. Since an exact expression of this error probability appears to be mathematically intractable, we resort to using the union bound on error probability. Assuming equally likely signal matrices, this results in the bound

$$\begin{aligned}\Pr(\hat{\mathbf{S}} \neq \mathbf{S}) &= \frac{1}{Z} \sum_{\mathbf{S}_p \in \mathbb{S}} \Pr(\hat{\mathbf{S}} \neq \mathbf{S}_p | \mathbf{S}_p \text{ was transmitted}) \\ &\leq \frac{1}{Z} \sum_{\mathbf{S}_p \in \mathbb{S}} \sum_{\substack{\mathbf{S}_q \in \mathbb{S} \\ \mathbf{S}_q \neq \mathbf{S}_p}} \Pr(\mathbf{S}_p \to \mathbf{S}_q), \end{aligned} \tag{3.1}$$

with the PEP $\Pr(\mathbf{S}_p \to \mathbf{S}_q)$ defined in Section 2.6. The criterion used to design the constellations in this chapter is the minimization of the union bound as objective function, either in scaled exact form in Section 3.2 or in approximative form for high SNR in Section 3.3. As motivated in Section 2.6, the PEP depends on the receiver structure. Due to its robustness with respect to mismatch in assumptions on the channel correlation matrix and due to the simpler analytical expressions in the high SNR region (see Section 2.6.4), the design in this chapter is based on the assumption that the GLRT receiver is used.

Since the PEP can be made arbitrarily small for signals of infinite power, some power constraint has to be imposed on the signal matrices. Assuming as above equally likely signals, the average transmitted power is limited by the constraint

$$\frac{1}{Z}\sum_{k=0}^{Z-1}\|\mathbf{S}_k\|_F^2 = T \tag{3.2}$$

to unit power per channel use.

3.2 Design Based on a Preselected Symbol Alphabet

It is not uncommon in the design of wireless communication systems to restrict the symbol alphabet to a small number of possible symbols. This restriction can be motivated by, e.g., conceptional simplicity in representing binary data or implementational aspects if amplifiers are designed for constant power output.

In order to model a system with these constraints, we deliberately restrict in our first signal design approach the search space to symbols that belong to a finite and preselected alphabet. We choose BPSK, keeping in mind that generalizations to higher order modulation are possible within the proposed framework.

Following the notation in Section 2.3.3, we denote with $\mathbf{S} \in \mathbb{S}$ a $T \times M_T$ matrix where the (t, m) element $s_{t,m}$ is equal to the symbol transmitted on antenna m in time instant $t-1$ (remember that $t = 0, \ldots, T-1$, but matrix indices start with 1 by convention). The symbol $s_{t,m}$ belongs to a finite alphabet that we restrict here to a scaled BPSK alphabet. Using this restriction, we have limited the search space for each $T \times M_T$ matrix $\mathbf{S}$ with binary parameters to a space of exactly 2^{TM_T} possible

configurations. Thus, any possible $\mathbf{S}$ can be characterized by the integer $i \in \mathbb{I}_{\mathbb{S}} \triangleq \{0, 1, \ldots, 2^{TM_T} - 1\}$. This characterization can be described by, e.g., the representation of the integer i as a binary word of length TM_T, i.e.,

$$i = \sum_{r=1}^{M_T T} d_r 2^{M_T T - r}, \quad d_r \in \{0, 1\} \tag{3.3}$$

where the symbol $s_{t,m}$ transmitted on antenna m at time instant $t-1$ is

$$s_{t,m} = \frac{1}{\sqrt{M_T}}(1 - 2d_{(m-1)T+t}), \quad s_{t,m} \in \left\{-\frac{1}{\sqrt{M_T}}, \frac{1}{\sqrt{M_T}}\right\}$$

in agreement with the power constraint (3.2). We will use the given mapping to characterize any *possible* $\mathbf{S}$ with the notation $\mathbf{S} = \mathbf{S}(i)$ where the argument $i \in \mathbb{I}_{\mathbb{S}}$ denotes the representation of $\mathbf{S}$ in integer format. The problem of constellation design can then be reformulated as finding the set of *allowed* signal matrices $\mathbf{S} \in \mathbb{S}$ as subset of all signal matrices that are *possible* given the preselected alphabet. Equivalently, we need to determine a subset $\mathbb{I} \subset \mathbb{I}_{\mathbb{S}}$ describing the indices of the allowed signal matrices, i.e., we need to determine Z distinct indices that describe the set of allowed signal matrices as a subset of $\mathbb{I}_{\mathbb{S}}$ which denotes the index set of possible signal matrices given the preselected alphabet. Arguments of the form $\mathbf{S}(k)$ refer therefore to possible signal matrices whereas subscripts as in $\mathbf{S}_p$ refer to allowed signal matrices. With these notational conventions, the union bound in (3.1) can be reformulated as

$$\Pr(\hat{\mathbf{S}} \neq \mathbf{S}) \leq \frac{1}{Z} \sum_{p \in \mathbb{I}} \sum_{\substack{q \in \mathbb{I} \\ q \neq p}} \Pr(\mathbf{S}(q) \to \mathbf{S}(p)). \tag{3.4}$$

3.2.1 Optimization: Simulated Annealing

With the explicit formulas for $\Pr(\mathbf{S}(q) \to \mathbf{S}(p))$ in (2.17), we can precompute a table of possible signal matrices and evaluate the objective function (3.4) for any desired index set $\mathbb{I}$ matched to additional assumptions on SNR and the channel correlation profile (as required by (2.17)). The constellation design problem can then be formulated as finding this index set. For a given number Z of signal matrices, we therefore need to determine Z integers $i_1, \ldots, i_Z$ with $i_k \in \mathbb{I}_{\mathbb{S}}$ that constitute $\mathbb{I}$. An exhaustive search over all possible subsets is immediately seen to be too

complex since

$$\left(\begin{array}{c} 2^{TM_T} \\ Z \end{array} \right) = \frac{(2^{TM_T})!}{Z!(2^{TM_T} - Z)!}$$

possible subsets need to be checked which is not feasible even for moderate constellation parameters. As an example, choose $T = 6$, $M_T = 1$ and $Z = 8$ (implying a bit rate of $R = 1/2$ per channel use) which leads to 4426165368 different subsets! We are therefore in need of a simplified optimization method. The method of *simulated annealing* has shown very good results in similar problems of code design [GHSW87] and will therefore be applied here as well. In simulated annealing, the objective function to be minimized is interpreted as the energy ε of a physical system with many degrees of freedom in thermal equilibrium. Slowly cooling down (annealing) the system will drive the system to the most probable state of lowest energy when the annealing process is sufficiently slow.

Applied to constellation design, the state is represented by the constellation description $\mathbb{I}$ and its energy ε is described by the objective function (3.4)

$$\varepsilon(\mathbb{I}) = \sum_{p \in \mathbb{I}} \sum_{q \in \mathbb{I}} \Pr(\mathbf{S}(p) \to \mathbf{S}(q))$$

where we have defined $\Pr(\mathbf{S}(p) \to \mathbf{S}(p)) \triangleq 0$ for simplicity. The initial state of the system is an arbitrary constellation described by $\mathbb{I} = \mathbb{I}_1 \subset \mathbb{I}_{\mathbb{S}}$ chosen at a high temperature τ. To simulate thermal movement at time instant n, we first pick two elements

$$u \in \mathbb{I}_n \quad \text{and} \quad v \in \mathbb{I}_n^c = \mathbb{I}_{\mathbb{S}} \setminus \mathbb{I}_n$$

at random (with uniform probability for all elements) and define the set

$$\tilde{\mathbb{I}}_{n+1} = (\mathbb{I}_n \setminus \{u\}) \cup \{v\}.$$

This operation thus exchanges two randomly picked elements $u \in \mathbb{I}_n$ and $v \in \mathbb{I}_n^c$ with each other. If the resulting energy change

$$\Delta\varepsilon = \varepsilon(\tilde{\mathbb{I}}_{n+1}) - \varepsilon(\mathbb{I}_n)$$

is negative, the change is always accepted, i.e., $\mathbb{I}_{n+1} = \tilde{\mathbb{I}}_{n+1}$ whereas changes $\Delta\varepsilon > 0$ are also sometimes accepted, with probability $p = \exp(-\Delta\varepsilon/\tau)$. This probability refers to a Boltzmann factor $\exp(-E/(k_b T))$ (notation see [KGV83]) in statistical mechanics to allow state degradation with a temperature-dependent probability. The

Choose initial code $\mathbb{I}_1$ and temperature τ. Set $n = 0$.
REPEAT
  REPEAT
    $n = n + 1$
    choose $u \in \mathbb{I}_n, v \in \mathbb{I}_n^c$
    set $\tilde{\mathbb{I}}_n = \mathbb{I}_n \setminus \{u\} \cup \{v\}$
    set $\Delta\varepsilon = \varepsilon(\tilde{\mathbb{I}}_n) - \varepsilon(\mathbb{I}_n)$
    IF $(\Delta\varepsilon < 0)$ THEN set $\mathbb{I}_{n+1} = \tilde{\mathbb{I}}_n$
    ELSE
      with probability $p = \exp(-\Delta\varepsilon/\tau)$
        set $\mathbb{I}_{n+1} = \tilde{\mathbb{I}}_n$
      otherwise
        $\mathbb{I}_{n+1} = \mathbb{I}_n$
  UNTIL (several energy drops or too many iterations)
  lower temperature: set $\tau = \alpha\,\tau$, where $0 < \alpha < 1$
UNTIL (freezing temperature reached)

Figure 3.1: High-level description of code design algorithm.

possibility of accepting changes increasing the energy allows the system to escape from local energy minima configurations. If the energy change is rejected, we have of course $\mathbb{I}_{n+1} = \mathbb{I}_n$.

The system is annealed by slowly decreasing τ (or, equivalently, the probability of code degradation) in order to drive the system to a stable state with minimum energy. Note that the initial code $\mathbb{I}_1$ does not notably influence the final result because the code $\mathbb{I}$ gets randomized quickly at high τ.

A high-level description of the design algorithm can be stated as in Figure 3.1. Note that the cooling of the problem involves randomness in the sense that the perturbation of the signal set is random. Furthermore, a perturbation that makes the signal set *worse* is accepted with a certain probability. This probability decreases as a function of the decreasing temperature. Hence, bad perturbations are accepted quite often at the beginning, giving the algorithm the possibility to escape from local minima, while almost only better signal sets are accepted towards the end of the design, providing convergence to a resulting signal set (see

also [KGV83, GHSW87]).

In our simulations we initiate the temperature to $\tau_0 = 10^7$, and stop when the freezing temperature $\tau_f = 10^{-7}$ is reached (this number is chosen low enough such that the algorithm ends with a stable code). We use $\alpha = 0.995$ and, furthermore, the inner loop ends if more than 5 perturbations in a row are accepted, or if more than 500 perturbations in a row are not accepted.

A remarkable advantage of this design method is its generality in the sense that it works for any $\Pr(\mathbf{S}(p) \to \mathbf{S}(q))$. That is, a constellation can be designed for *any* given decoder structure and for any channel model, as long as the corresponding pairwise error probabilities can be computed or approximated. If the approaches of Section 2.6 should be non-applicable for the computation of the PEP, such an approximation might also be obtained using Monte Carlo simulation if necessary.

3.2.2 Design Complexity Reduction

In the search for good signal sets in Section 3.2.1, we are in need of accessing $\Pr(\mathbf{S}(q) \to \mathbf{S}(p))$ for all $(p, q) \in \mathbb{I}_{\mathbb{S}} \times \mathbb{I}_{\mathbb{S}}$. Precomputing all PEP values demands a storage of $2^{2M_T T}$ real scalars, which is only feasible for small T and M_T. Inspecting the resulting table for specific examples, it is apparent that a large number of elements have the same numerical value. Thus, we expect a potential for reduction of memory requirement. This section describes a structural property of the PEP table which will be exploited to reduce storage requirements. Instead of accessing a table directly, p and q determine a matrix whose elements are used for accessing a table for storage of $\Pr(\mathbf{S}(p) \to \mathbf{S}(q))$ values. In certain cases, this leads to a significant reduction of needed memory space.

Following the analysis on the GLRT receiver in Section 2.6.1, the PEP is determined by the nonzero eigenvalues of

$$\begin{aligned}
&(\sigma^2\mathbf{I} + \bar{\mathbf{S}}_p\mathbf{C}_\mathbf{h}\bar{\mathbf{S}}_p^H)(\mathbf{P}_{\bar{\mathbf{S}}_q} - \mathbf{P}_{\bar{\mathbf{S}}_p}) \\
&= \sigma^2(\bar{\mathbf{S}}_q(\bar{\mathbf{S}}_q^H\bar{\mathbf{S}}_q)^{-1}\bar{\mathbf{S}}_q^H - \bar{\mathbf{S}}_p(\bar{\mathbf{S}}_p^H\bar{\mathbf{S}}_p)^{-1}\bar{\mathbf{S}}_p^H) - \bar{\mathbf{S}}_p\mathbf{C}_\mathbf{h}\bar{\mathbf{S}}_p^H\mathbf{P}^{\perp}_{\bar{\mathbf{S}}_q} \\
&= [\bar{\mathbf{S}}_p \quad \bar{\mathbf{S}}_q] \begin{bmatrix} -\sigma^2(\bar{\mathbf{S}}_p^H\bar{\mathbf{S}}_p)^{-1} - \mathbf{C}_\mathbf{h} & \mathbf{C}_\mathbf{h}\bar{\mathbf{S}}_p^H\bar{\mathbf{S}}_q(\bar{\mathbf{S}}_q^H\bar{\mathbf{S}}_q)^{-1} \\ \mathbf{0} & \sigma^2(\bar{\mathbf{S}}_q^H\bar{\mathbf{S}}_q)^{-1} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{S}}_p^H \\ \bar{\mathbf{S}}_q^H \end{bmatrix}
\end{aligned}$$

which in turn are equivalent to the nonzero eigenvalues of

$$\begin{bmatrix} -\sigma^2(\bar{\mathbf{S}}_p^H\bar{\mathbf{S}}_p)^{-1} - \mathbf{C}_\mathbf{h} & \mathbf{C}_\mathbf{h}\bar{\mathbf{S}}_p^H\bar{\mathbf{S}}_q(\bar{\mathbf{S}}_q^H\bar{\mathbf{S}}_q)^{-1} \\ \mathbf{0} & \sigma^2(\bar{\mathbf{S}}_q^H\bar{\mathbf{S}}_q)^{-1} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{S}}_p^H\bar{\mathbf{S}}_p & \bar{\mathbf{S}}_p^H\bar{\mathbf{S}}_q \\ \bar{\mathbf{S}}_q^H\bar{\mathbf{S}}_p & \bar{\mathbf{S}}_q^H\bar{\mathbf{S}}_q \end{bmatrix}. \tag{3.5}$$

Here, we used the fact that for any two matrices $\mathbf{A}$ and $\mathbf{B}$ of compatible dimension, the non-zero eigenvalues of $\mathbf{AB}$ are the same as for $\mathbf{BA}$, see [HJ99, Theorem 1.3.20]. Now since the matrix in (3.5) depends on the signal matrices only via $\bar{\mathbf{S}}_p^H \bar{\mathbf{S}}_p$, $\bar{\mathbf{S}}_q^H \bar{\mathbf{S}}_q$ and $\bar{\mathbf{S}}_p^H \bar{\mathbf{S}}_q$, it is apparent that its eigenvalues and therefore the PEP depend only on the elements of $\bar{\mathbf{S}}_p^H \bar{\mathbf{S}}_p$, $\bar{\mathbf{S}}_q^H \bar{\mathbf{S}}_q$ and $\bar{\mathbf{S}}_p^H \bar{\mathbf{S}}_q$. Now since the symbols are taken from a predetermined finite set, the elements of the mentioned matrices are elements in a finite set. We can therefore address a table of PEP values either directly via the mapping in (3.3) or indirectly using the distinct values of the elements of $\bar{\mathbf{S}}_p^H \bar{\mathbf{S}}_p$, $\bar{\mathbf{S}}_q^H \bar{\mathbf{S}}_q$ and $\bar{\mathbf{S}}_p^H \bar{\mathbf{S}}_q$. These matrices have a block-Toeplitz structure, and we therefore need less parameters than their number of elements $(LM_T)^2$ to describe them. Both the matrices $\bar{\mathbf{S}}_p^H \bar{\mathbf{S}}_p$ and $\bar{\mathbf{S}}_q^H \bar{\mathbf{S}}_q$ are Hermitian, i.e., we have a block on the main diagonal and $L-1$ distinct blocks off the diagonal. The block on the main diagonal is characterized by $(M_T^2 - M_T)/2$ parameters, the off-diagonal blocks by M_T^2 parameters. Thus, $\bar{\mathbf{S}}_p^H \bar{\mathbf{S}}_p$ and $\bar{\mathbf{S}}_q^H \bar{\mathbf{S}}_q$ can be described by

$$(M_T^2 - M_T)/2 + (L-1)M_T$$

parameters each. The matrix $\bar{\mathbf{S}}_p^H \bar{\mathbf{S}}_q$ has $2L-1$ distinct blocks of dimension $M_T \times M_T$. Thus, $(2L-1)M_T^2$ parameters are required for the description of this matrix. Adding up the number of required parameters, we conclude that

$$2\left(\frac{M_T^2 - M_T}{2} + (L-1)M_T\right) + (2L-1)M_T^2 = (4L-2)M_T^2 - M_T$$

distinct parameters are sufficient to describe the matrices $\bar{\mathbf{S}}_p^H \bar{\mathbf{S}}_p$, $\bar{\mathbf{S}}_q^H \bar{\mathbf{S}}_q$ and $\bar{\mathbf{S}}_p^H \bar{\mathbf{S}}_q$ and therefore the PEP. Each of these parameters is a sum of at most T elements $1/\sqrt{M_T}$ or $-1/\sqrt{M_T}$ and can thus take on a limited number of distinct values, typically in the order of T (in a rough approximation). Thus a memory for a number of entries in the order

$$T^{(4L-2)M_T^2 - M_T}$$

is sufficient to store all error probabilities. We will call this strategy of indexing the values for the PEP the indirect approach in contrast to the direct approach which stores the full PEP table. The direct approach requires a memory with

$$\left(2^{M_T T}\right)^2 = 2^{2M_T T}$$

elements which is exponential in T. The indirect approach has removed the exponential dependence on T but has introduced an exponential dependence on L. The direct approach demands a memory that grows exponentially in M_T whereas the memory size of the indirect approach grows exponentially in M_T^2! It should be emphasized here that the number of memory elements that is actually *used* in the indirect approach can of course never exceed the number of elements in the full table. However, the indexing of the table can become increasingly complex for larger values of M_T and L. The indirect method is therefore more suited for constellation design in systems with a single transmitter antenna whereas the direct approach is preferable for MIMO systems when the problem size grows large.

Two numerical examples highlight the different properties of these methods: For a design with the parameters $M_T = 1$, $L = 2, T = 10$, the direct approach requires a memory with $2^{20} = 2^{10} \cdot 2^{10} \approx 10^6$ entries, the indirect approach a memory with 10^5 elements (where by far not all spaces are filled, so a sparse memory organization can be used). For $M_T = 2$, $L = 2, T = 6$, the direct approach demands a memory for $2^{24} \approx 1.6 \times 10^7$ entries that can be accessed directly, but the memory organized using the indirect approach is accessed using $(4L-2)M_T^2 - M_T = 22$ parameters, leading to a potential address space of about $6^{22} \approx 1.3 \times 10^{17}$ elements when the rough approximation above is used. Clearly, the memory organization becomes increasingly difficult in the design for MIMO systems using the indirect approach!

3.2.3 Training-Based Schemes

A standard way of enabling the receiver to acquire channel state information is via the use of a training block, i.e., the transmission of known symbols via each antenna. Each signal matrix $\mathbf{S}$ is thereby partitioned into a $q \times M_T$ training block $\mathbf{S}_t$ common to all signal matrices and a $(T-q) \times M_T$ data block $\mathbf{S}_d$ as in

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_t \\ \mathbf{S}_d \end{bmatrix}.$$

The resulting signal matrix $\bar{\mathbf{S}}$ according to (2.7) can then be partitioned into

$$\bar{\mathbf{S}} = \begin{bmatrix} \bar{\mathbf{S}}_t \\ \bar{\mathbf{S}}_d \end{bmatrix} \tag{3.6}$$

where the training matrix $\bar{\mathbf{S}}_t$ with dimension $q \times LM_T$ is completely known to the receiver and the $(T+L-1-q) \times LM_T$ matrix $\bar{\mathbf{S}}_d$ contains the unknown data symbols as well as some training symbols.

The design of the training block, i.e., its dimension q, its allocated power and symbols in view of the design of the data matrix can be viewed as a tradeoff of spending resources on either channel acquisition at the receiver or protection of the data against errors. The word resources can here be interpreted in at least two ways, either as "redundancy" or "power." In the interpretation of "redundancy," it is assumed that a standard error-correcting code for the data symbols is applied which uses a redundant representation of the data which is designed such that a distortion of the data, typically in the form of a limited number of bit errors, can be corrected if the channel is known at the receiver. The training block can then be considered as additional pure redundancy used for CSI acquisition at the receiver. A longer training block implies therefore that a shorter error-correcting code with worse error correction properties must be used. In other words, more redundancy allocated to training and CSI acquisition implies reduced redundancy allocated to error protection for the data symbols (and vice versa). Another interpretation of the word "resource" is "power." Too much power spent on training will result in accurate CSI acquisition but not enough power for data in view of the receiver noise. Too much power on data symbols can result in ineffective data detection if the channel is not identified with sufficient accuracy (for instance, two signal matrices $\mathbf{S}_1$ and $-\mathbf{S}_1$ cannot be distinguished at the receiver in the absence of CSI).

The design of the training block is therefore an important issue and has received some attention in the literature (see, e.g., [HH03] for frequency-flat fading MIMO channels and [ATV02, VHHK01] for frequency-selective SISO channels, see also [LS03]). It is a common belief that the resulting training matrix $\bar{\mathbf{S}}$ should have *orthogonal columns* because such a design minimizes the estimation error variance if a channel estimate based on the training matrix only is computed at the receiver. The discussion of the mentioned resource tradeoff understood as power tradeoff is the basis of the capacity considerations as in [HH03] and [ATV02]. Here, we take the viewpoint of trading redundancy optimized for CSI acquisition against redundancy for error protection and compare our constellation design with schemes with large amount of redundancy for training on one side compared to schemes with large amount of redundancy in the representation of the data.

This tradeoff of power or redundancy allocated to data of CSI acqui-

sition can also be studied under the assumptions of simplified receivers compared to the ML or GLRT receivers presented in Section 2.5 and allows more insight into the mentioned tradeoff of resources. We therefore consider several receiver implementations:

- **Scheme 1: CSI acquisition on training only, equalization and hard detection.** Similar to the partitioning of the signal matrix into a training only and signal matrix, we also partition the received signal matrix into $\mathbf{Y}_t$ and $\mathbf{Y}_d$ where $\mathbf{Y}_t$ stems from training only and obtain a channel estimate $\hat{\mathbf{H}}$ in

$$\hat{\mathbf{H}} = (\bar{\mathbf{S}}_t^H \bar{\mathbf{S}}_t)^{-1} \bar{\mathbf{S}}_t^H \mathbf{y}_t = \bar{\mathbf{S}}_t^\dagger \mathbf{Y}_t.$$

  The obtained channel estimate is then considered correct in a subsequent Viterbi equalization of the channel (implying a hard decision on the received symbols) and subsequent detection of the data using binary decoding. The choice of channel estimate is in line with the channel estimate in (2.12). In a different context, the MMSE channel estimate is preferred, see Section 3.3.4.

- **Scheme 2: Detection with perfect CSI, equalization and hard detection.** This receiver works similar as receiver 3.2.3 but it is assumed that it has perfect CSI available. The comparison of receiver 1 and 2 thereby allows the assessment of the impact that erroneous CSI acquisition has on the detection performance.

- **Scheme 3: Joint GLRT decoding.** Here we consider the signal matrix as one joint code and apply the receiver of Section 2.5.

All these receivers suitable for the separate training and data symbol design can be compared to the application of the GLRT receiver in combination with signal matrices explicitly designed for the unknown frequency-selective channel.

3.2.4 Numerical Results

We will illustrate in this section the performance of constellations designed for unknown channels with some simulation examples. The presentation is divided into two subsections: First, we study a standard SISO system and compare our constellations with well-known error-correcting codes extended by a training sequence which we vary in length to illustrate the mentioned tradeoff in the allocation of redundancy. In the

second part, we focus on multi-antenna systems and show among others a comparison of our designed signal set with an extension of the well known Alamouti scheme. Some final remarks summarize this section.

For a SISO system, we choose in our first example with significant redundancy allocated to CSI acquisition the well known $(15, 11)$ Hamming code as standard error-correcting code which is extended with seven pilot bits to obtain a $(22, 11)$ code of rate $R = 1/2$. All three receiver schemes mentioned in the previous section are evaluated. The performance comparison between these three benchmark schemes and a designed $(22, 11)$ constellation using simulated annealing is illustrated in Figure 3.2 in terms of codeword error rate (WER) vs. SNR. The channel is in all cases modeled as a normalized Rayleigh fading channel of length $L = 2$ and with coefficients of equal power, that is, $\mathbf{h} = [h_0\ h_1]$ with $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, 1/L\mathbf{I}_2)$. We define the SNR for these and all following performance simulations as

$$\text{SNR} = \frac{1}{R\sigma^2} \tag{3.7}$$

where the rate factor $R = \frac{\log_2 Z}{T}$ in the denominator of (3.7) assures power normalization on the information bit. The design assumptions for the $(22, 11)$ constellations were matched to the simulation setup and a design SNR of 13 dB was assumed. As can be observed in Figure 3.2, the designed $(22, 11)$ constellation significantly outperforms the other schemes, clearly illustrating the performance benefit of optimizing the signal set. Besides this observation, it is interesting to note that Scheme 2 performs worse than Scheme 3. Note that Schemes 2 and 3 use exactly the same overall code but in Scheme 3 the channel is *not* known at the receiver. However, Scheme 3 uses optimal joint estimation, equalization and soft decoding, while Scheme 2 employs separate equalization and hard decoding. Hence, it appears that the gain of Scheme 2 over Scheme 1 due to the genie aided channel estimation is not as large as the gain of Scheme 3 due to the joint channel estimation and soft decoding.

When comparing the different techniques, it should be noted that the complexity of the schemes that use optimal joint decoding is the highest (Scheme 3 and the detection of the designed constellation), because implementing the GLRT detector requires a search over 2^{11} terms per transmitted signal matrix. We also emphasize that it is in no way obvious how to exploit the structure of the Hamming code or the new signal set to lower the decoding complexity when employing GLRT detection over the investigated unknown channel, as desired here.

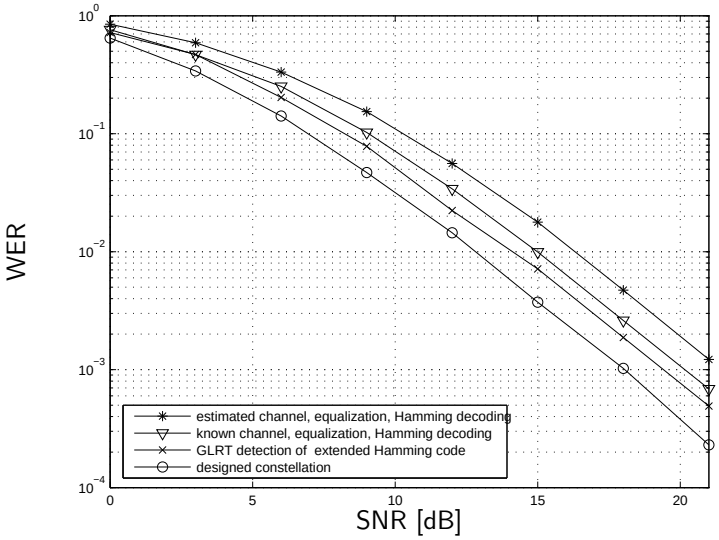


Figure 3.2: Performance comparison of the designed $(22, 11)$ constellation against three benchmark schemes operating on a Hamming code with a pilot sequence extension resulting in a different overall $(22, 11)$ constellation.

In Figure 3.2 both the designed constellation and Scheme 3 use GLRT detection, i.e., joint data detection and channel estimation. Hence the difference in performance between them is due only to the different designs of the signal matrices. Note however that the approach used in Scheme 3 of concatenating separate coding and pilot bits was motivated by Schemes 1 and 2. The performance gap between Scheme 3 and 1 and to thereby illustrates the impact on the performance of separated versus joint decoding.

When using joint decoding it can be argued that the overall code defined by the concatenation of error-control code and training sequence is poor from an error correction point of view because a large part of the available redundancy is spent on the training signal and this redundancy can therefore not be used to improve the error-correcting code, whereas redundancy in the error-correcting code might be useful for channel estimation as well. In an attempt to find a better off-the-shelf code for

use with optimal joint decoding and with insignificant separate training, we utilize a Golay $(23, 12)$ code [Bos98] extended with one known pilot bit. This defines a $(24, 12)$ rate $1/2$ code and therefore an additional benchmark scheme when used with the GLRT detector.

The reason for using a pilot bit, and not the traditional extension (an additional parity bit) of the Golay $(23, 12)$ code to a $(24, 12)$ code, is that in the traditional $(24, 12)$ code (and also in the $(23, 12)$ code) both the all zeros and the all ones blocks are allowed codewords and hence, since the code is linear, for each signal matrix in the code the pattern defined by negating all bits is also a signal matrix. This means that the traditional extension of the $(23, 12)$ Golay code cannot be used with the optimal joint decoder because this decoder cannot tell the difference between $\bar{\mathbf{S}}$ and $-\bar{\mathbf{S}}$ when both matrices correspond to signal matrices resulting in a detectability problem. Adding one known pilot bit to the Golay $(23, 12)$ code is a straightforward means of resolving the ambiguity, and hence make the resulting signaling scheme work over a channel with unknown Rayleigh fading. It can also be interpreted as minimal possible training given the preselected symbol alphabet.

Figure 3.3 illustrates the performance of the extended Golay code and compares it with a $(24, 12)$ design obtained using simulated annealing The channel model is the same as the one used in Figure 3.2 as were the design assumptions for the constellation. As can be seen the new set of signals outperforms the extended Golay code. The gain is about 1.4 dB in SNR at a WER of 10^{-2}, and is hence comparable to the gain using $(22, 11)$ codes over Scheme 3 in Figure 3.2. Note that the computational complexities in the detection of the extended Golay code and the obtained $(24, 12)$ design are the same. There is a significant difference in design complexity however, but since the optimization is carried out off-line, this is of minor practical importance.

Our design approach described in Section 3.2.1 can also be applied to the design of signal sets supporting different rates for more than one transmitter antenna. As mentioned in Section 3.2.2, the required memory resources cannot be reduced as efficiently as in the SISO case and we therefore have to restrict ourselves to rather short example designs.

For $M_T = 2$ transmitter antennas, a performance comparison of designed constellations of word length $T = 6$ with rates $R = 1/3, 1/2, 2/3, 1, 4/3$ is presented in Figure 3.4. In all simulations, we assume a frequency-selective Rayleigh fading channel with $L = 2$ taps between each transmitter and receiver antenna. All signal sets were designed under the assumption of an SNR of 20dB and uncorrelated fading.

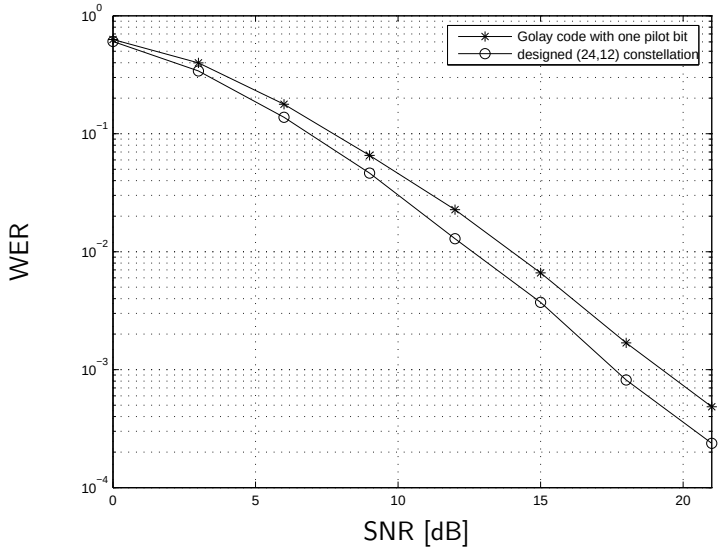


Figure 3.3: Performance comparison of the designed $(24, 12)$ constellation against a Golay $(23, 12)$ code concatenated with a known pilot bit. Both schemes use joint ML decoding.

For simplicity, $M_R = 1$ receiver antenna is assumed. As becomes clear from Figure 3.4, it appears that it is more power-efficient to transmit at lower rates than at higher rates for the given signal length T. For example, the code with $R = 1/2$ has a gain of about 3.6 dB compared to the code with $R = 1$ at a remaining word error rate (WER) of 10^{-3}. Moreover, it is apparent from the slope of the performance curves that not all constellations are able to exploit full diversity. In fact, none of the constellations satisfies the rank criterion (2.29) for full diversity performance in Section 2.6, because $T + L - 1 < 2LM_T$. Depending on the number of signals in each constellation, this does not lead to significant flattening of the performance curves for SNR less than 30 dB as in Figure 3.5.

In order to compare our scheme with a previously known approach, we also construct a signal set based on the Alamouti scheme [Ala98] which was originally conceived for two transmitter antennas over flat fading channels that are known to the receiver. Concatenating two Alamouti

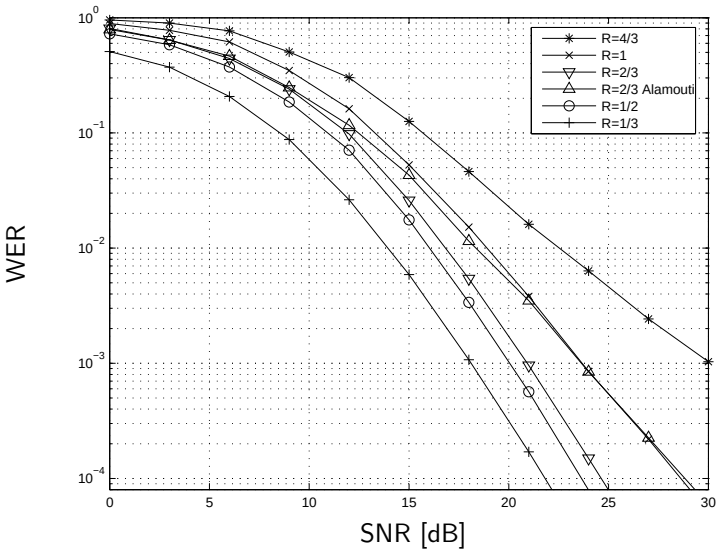


Figure 3.4: Performance comparison for $T = 6$ over $M_T = 2$ transmitter antennas with variable rate R.

blocks of the form

$$\begin{bmatrix} s_1 & s_2 \\ -s_2 & s_1 \end{bmatrix}$$

with a training block of length 2 resulted in a constellation of rate $R = 2/3$. As becomes apparent from Figure 3.4, this constellation has inferior performance compared to the one of the same rate which was designed using simulated annealing.

Additional Remarks

The constellations presented in this section are the result of computer-intensive searches using the simulated annealing optimization algorithm. The constellations with $T = 22$ and $T = 24$ could only be designed using the design complexity reduction method described in Section 3.2.2. A direct approach computing and storing a table with $(2^{22})^2 = 2^{44}$ elements was clearly not feasible whereas the indirect approach using a sparse table organization did not exceed memory requirements on a standard worksta-

tion with 256 MB RAM. In contrast, the indirect approach did not lead to a memory reduction for systems with several transmitter antennas and the design complexity therefore imposed a restriction to constellations spanning only a small number of symbols. It is apparent that constellations of high dimensions may be required for better performance when the channel allows longer signals.

3.3 Constellations Based on General Complex-Valued Symbols

In this section, we present a method to design signal sets that are optimized for data transmission over unknown frequency-selective channels without using any preselected symbol alphabet. The resulting signals are thus built up from arbitrary complex-valued symbols subject to the power constraint (3.2). The design criterion is, similar as in Section 3.2, based on the union bound on decoding error. However, we use an approximation from [BV01] to determine approximate pairwise error probabilities. This results in an objective function for the optimization which is significantly easier to handle analytically, provides insight into conditions on the signal set to allow maximum diversity gains and permits an easy implementation of an optimization algorithm operating on continuous variables.

3.3.1 Design Criterion and Search Space

In Section 2.6, the behavior of the PEP was analyzed for high SNR either by means of the Chernoff upper bound or using asymptotic analysis. It was found that for GLRT detection, the Chernoff upper bound and the asymptotic approximation of the PEP behave essentially as

$$\Pr(\mathbf{S}_p \to \mathbf{S}_q) \approx \Pr{}^a(\mathbf{S}_p \to \mathbf{S}_q) \triangleq \frac{\sigma^{2LM_TM_R}V}{|\bar{\mathbf{S}}_p^H \mathbf{P}_{\bar{\mathbf{S}}_q}^{\perp} \bar{\mathbf{S}}_p|^{M_R}} \tag{3.8}$$

where the constant V summarizes factors independent of the elements of the signal matrices $\mathbf{S}_p$ and $\mathbf{S}_q$ and differs dependent on the type of analysis of the PEP. The expression in (3.8) is easily differentiable in the consitutent symbols and therefore lends itself to numerical optimization based on gradient methods in contrast to the exact expression for the PEP used in Section 3.2. Replacing the PEP in (3.1) with its approximation

for high SNR, we arrive after dropping constant factors independent of the constellation design at the objective function

$$J(\mathbb{S}) = J(\mathbf{S}_0, \ldots, \mathbf{S}_{Z-1}) \triangleq \sum_{p=0}^{Z-1} \sum_{\substack{q=0 \\ p \neq q}}^{Z-1} \frac{1}{|\tilde{\mathbf{S}}_p^H \mathbf{P}_{\tilde{\mathbf{S}}_q}^{\perp} \tilde{\mathbf{S}}_p|^{M_R}} \tag{3.9}$$

The advantage in analytical tractability is achieved at the expense of generality. Whereas the optimization method in Section 3.3.2 can be applied in general for every receiver structure, channel correlation profile and SNR as long as the PEP can be computed or approximated numerically, the objective function in this section is based on the assumption of high SNR, full rank channel correlation matrix (which decouples from the objective function) and the application of the GLRT receiver. Moreover, for signal matrices with $T < 2M_T$, the required rank criterion on $[\tilde{\mathbf{S}}_p\, \tilde{\mathbf{S}}_q]$ cannot be fulfilled which implies that (3.9) cannot be adopted as objective function. Therefore, using (3.9) implies that these additional assumptions must be satisfied.

3.3.2 Optimization

The goal of our constellation design is to find a set $\mathbb{S} = \{\mathbf{S}_0, \ldots, \mathbf{S}_{Z-1}\}$ that minimizes (3.9) under the power constraint (3.2). This problem is non-convex. We therefore apply a generic gradient-search type algorithm to find an approximate solution.

In order to compute gradients with respect to complex-valued parameters, we define the complex differentiation operator for real-valued functions with respect to a complex scalar z as

$$\frac{\partial}{\partial z} \triangleq \frac{1}{2}\left(\frac{\partial}{\partial \Re(z)} + j\frac{\partial}{\partial \Im(z)}\right). \tag{3.10}$$

Moreover, let $s_{tm,k} = [\mathbf{S}_k]_{tm}$ denote the complex symbol sent at time instant $t-1$ on antenna m in the signal matrix with index k. We can then define the derivative $\partial J/\partial \mathbf{S}_k$ of the scalar function J with respect to the complex-valued matrix $\mathbf{S}_k$ as the matrix with elements

$$\left[\frac{\partial J}{\partial \mathbf{S}_k}\right]_{tm} \triangleq \frac{\partial J}{\partial s_{tm,k}}.$$

This gradient on the given cost function J is computed in Appendix 3.A.

The gradient-search method consists now in iteratively approaching a minimum of the objective function by improving a solution at step n

$$\mathbb{S}^n = \{\mathbf{S}_0^n, \ldots, \mathbf{S}_{Z-1}^n\}$$

by "moving" the parameters in $\mathbf{S}_k$ in the direction of maximum decreasing objective function which is $-\partial J/\partial \mathbf{S}_k$. Since the constellation will then no longer satisfy the power constraint (3.2), a normalization is necessary to reach the improved design

$$\mathbb{S}^{n+1} = \{\mathbf{S}_0^{n+1}, \ldots, \mathbf{S}_{Z-1}^{n+1}\}$$

using

$$\mathbf{S}_k^{n+1} = \sqrt{TZ}\frac{\mathbf{S}_k^n - \mu\frac{\partial J}{\partial \mathbf{S}_k^n}}{\sum_{k=0}^{Z-1}\left\|\mathbf{S}_k^n - \mu\frac{\partial J}{\partial \mathbf{S}_k^n}\right\|^2} \tag{3.11}$$

where the step size μ is chosen such that $J(\mathbb{S}^{n+1}) \leq J(\mathbb{S}^n)$.

The signal matrices $\mathbf{S}_k$ are parameterized here directly in the constituent symbols. Alternative parameterizations exist which, e.g., impose the structure of orthonormal columns with equal transmission power on the signal matrices [ARU01]. Such a parameterization is motivated by the result of [MH99, HM00] that for high SNR or $T \gg M_T$, these matrices achieve capacity. Moreover, it was shown in [BV01] that such matrices also minimize the asymptotic union bound on error probability for ML detection if equal energy for all signals is required. The benefit of a parameterization that imposes this structure lies in somewhat more advantageous numerical properties in the optimization since the search space is smaller and the design problem can be cast as an unconstrained optimization problem. However, the capacity result is based on the assumption that the channel is frequency-flat fading and we are not aware of an extension to frequency-selective channels. Thus, we have no stringent motivation to impose a constraint of equal energy signaling. We can however show that optimal matrices with respect to the asymptotic union bound have orthogonal columns, however not necessarily with equal energy.

Lemma 1 *For any* $\mathbb{S} = \{\mathbf{S}_0, \ldots, \mathbf{S}_{Z-1}\}$ *with* $tr\left(\mathbf{S}_k^H\mathbf{S}_k\right) = E_k$, *there exists a set* $\tilde{\mathbb{S}} = \{\tilde{\mathbf{S}}_0, \ldots, \tilde{\mathbf{S}}_{Z-1}\}$ *with* $\tilde{\mathbf{S}}_k^H\tilde{\mathbf{S}}_k = E_k/M_T\,\mathbf{I}_{M_T}$ *and* $J(\tilde{\mathbb{S}}) \leq J(\mathbb{S})$.

The proof is an extension of a similar derivation in [BV01] and presented in Appendix 3.B.

This lemma shows that signals which are scaled unitary, i.e.,

$$\mathbf{S}_k^H \mathbf{S}_k = E_k \, \mathbf{I}_{M_T}$$

are optimal with respect to the objective function (3.9). In other words, it is optimal with respect to asymptotic pairwise error probability to transmit signals which are orthogonal between different antennas. However, we cannot claim that each signal should transmit with the same power (the quantity E_k depends on k). We could therefore just as well use a parameterization as in [ARU01] together with a power parameter for each signal matrix. However, this is again a constrained optimization problem and therefore no significant improvement compared with the direct parameterization in the symbols. Still, it is interesting to note in how far solutions obtained allowing variable transmit power per signal matrix really differ in that respect (see Section 3.3.3).

The recursion in (3.11) is initialized either with a random $\mathbb{S}^0$ or with a set of matrices satisfying $(\mathbf{S}_k^0)^H \mathbf{S}_k^0 = T/M_T \, \mathbf{I}_{M_T}$. The recursion is stopped if the Kuhn-Tucker conditions (see [Fle91])

$$\frac{\partial J}{\partial \mathbf{S}_k^n} + \lambda \mathbf{S}_k^n = 0$$

are approximately satisfied for some $\lambda > 0$ jointly for all k. The convergence can only be guaranteed to a local optimum which is dependent on the initialization, so many optimization runs with different initialization values are performed, picking the best resulting constellation out of the runs. The difference in final objective function value appeared to be insignificant and independent on whether the recursion was initialized with a random matrix constellation or a signal matrices with orthogonal columns. The latter approach lead however to faster convergence.

3.3.3 Constellation Properties

It was pointed out in Section 2.3.3 that a system model with M_T transmitter antennas over a frequency-selective channel with L taps can be considered equivalent to a system model with $M_T L$ transmitter antennas over a flat channel where $M_T(L-1)$ antennas transmit time-shifted copies of the signals on the first M_T antennas. The resulting signal matrix $\bar{\mathbf{S}}$ thus has the structure described in (2.7). As mentioned in Section 3.3.1,

it was shown for flat-fading channels with $L = 1$ where no structure is imposed on $\bar{\mathbf{S}} = \mathbf{S}$ that the signal matrix $\mathbf{S}$ should have orthonormal columns. We consider it interesting to compare in how far the analogy described above influences the constellation design in our setup, i.e., in how far the signal matrices designed for frequency-selective fading follow the design rules for channels with frequency-flat fading. We therefore define

$$\bar{\kappa} = \frac{1}{Z} \sum_{i=0}^{Z-1} \kappa(\bar{\mathbf{S}}_i^H \bar{\mathbf{S}}_i)$$

where $\kappa(\bar{\mathbf{S}}^H \bar{\mathbf{S}})$ denotes the condition number of $\bar{\mathbf{S}}^H \bar{\mathbf{S}}$ with respect to the spectral norm [HJ99]. Equivalently, $\kappa(\bar{\mathbf{S}}^H \bar{\mathbf{S}})$ is the ratio of largest to smallest eigenvalue of $\bar{\mathbf{S}}^H \bar{\mathbf{S}}$. For unitary matrices $\bar{\mathbf{S}}$, all eigenvalues of $\bar{\mathbf{S}}^H \bar{\mathbf{S}} = \mathbf{I}_{T+L-1}$ are equal leading to $\bar{\kappa} = 1$. The difference between $\bar{\kappa}$ and 1 is thus a measure for how "close" the matrices $\bar{\mathbf{S}}$ are to being unitary on average.

Moreover we are interested in the distribution of power between signal matrices. Let

$$E = \max_{\mathbf{S} \in \mathbb{S}} \operatorname{tr}\left(\mathbf{S}^H \mathbf{S}\right) \qquad \text{and} \qquad e = \min_{\mathbf{S} \in \mathbb{S}} \operatorname{tr}\left(\mathbf{S}^H \mathbf{S}\right)$$

denote the maximum and minimum power allocated by a signal matrix of the chosen constellation. The quantity E/e describes in how far the signal power is unevenly spread between signal matrices (which would be signified by $E/e = 1$).

3.3.4 Evaluation

Single-Antenna Schemes with Optimized Training

It was claimed in [ATV02] with similar results in [VHHK01, MGO02, HH03] that optimized training sequences allow coding schemes to capture most of the available channel capacity, at least in the high SNR domain. However, an argumentation based on capacity relies on infinite decoding delay assumptions and is therefore not necessarily relevant in real-time constrained systems. Nevertheless, the design rules given in [ATV02] allow a well motivated choice of benchmark schemes in the comparison with our designed codes.

Referring to the notation introduced in Section 3.2.3 for training-based SISO schemes, we choose as training sequence length $q = L$, i.e., the minimum length if channel estimation is performed exclusively based

on training. The resulting training matrix $\overline{\mathbf{S}}_t$ should have orthonormal columns according to [ATV02] which can be achieved by choosing as only non-zero symbol $s_0 = \sqrt{(2L-1)\rho_t}$ and $s_1 = s_2 = \cdots = s_{q-1} = 0$ where ρ_t is computed according to [ATV02] in order to optimize the power allocation between training and data block. Note that in order to compare a training-based design constructed in this way with signals designed by our scheme, an additional power normalization is necessary since the formulas in [ATV02] are normalized to give average signal matrix power $T+L-1$, not T as in (3.2). Similar to the discussion in Section 3.2.3, we consider a training-only based CSI acquisition followed by a detector using the acquired channel estimate as perfect as well as the joint decoder. However, in agreement with [ATV02], the channel estimate is computed as the MMSE estimate, not as the ML estimate.

- **MMSE channel estimation followed by detection for perfect CSI.** CSI acquisition in [ATV02] is performed using training-only based minimum mean square error (MMSE) channel estimation according to

$$\hat{\mathbf{h}} = \mathbf{C}_{\mathbf{h}}\bar{\mathbf{S}}_t^H(\bar{\mathbf{S}}_t\mathbf{C}_{\mathbf{h}}\bar{\mathbf{S}}_t^H + \sigma^2\mathbf{I})^{-1}\ \mathbf{y}_t.$$

  This estimate is then applied in the detection of $\bar{\mathbf{S}}_d$ as

$$\hat{\bar{\mathbf{S}}}_d = \arg\min_{\bar{\mathbf{S}}_d} \|\mathbf{y}_d - \bar{\mathbf{S}}_d\hat{\mathbf{h}}\|^2.$$

  Note that this receiver requires knowledge of some channel characteristics that are assumed to be $\mathrm{E}[\mathbf{h}] = \mathbf{0}_{L\times 1}$ and $\mathrm{E}\left[\mathbf{h}\mathbf{h}^H\right] = \mathbf{C}_{\mathbf{h}}$. Moreover, knowledge of σ^2 is necessary. Note that the received symbols as well as the channel are vectors because of the single-antenna setup.

Moreover, joint GLRT detection is considered. Given these two receiver schemes, we consider two different benchmark constellations (picked to obtain constellations of rate $R = 1/2$).

- An extension of the well-known $(7, 4)$ Hamming code in BPSK signals with a training block.
- An extension of the $(23, 12)$ Golay code in BPSK signals with a training sequence.

The training block in both schemes is optimized according to [ATV02]and of length L. The simulation results in terms of WER vs. SNR are shown

in Figures 3.5 and 3.6 for $L = 1$ and $L = 3$ in comparison to the extended Hamming and Golay codes together with constellations specifically designed for the scenario. We assume as earlier Rayleigh fading channel coefficients $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, 1/L\mathbf{I})$. Note that the slope of the WER curves in

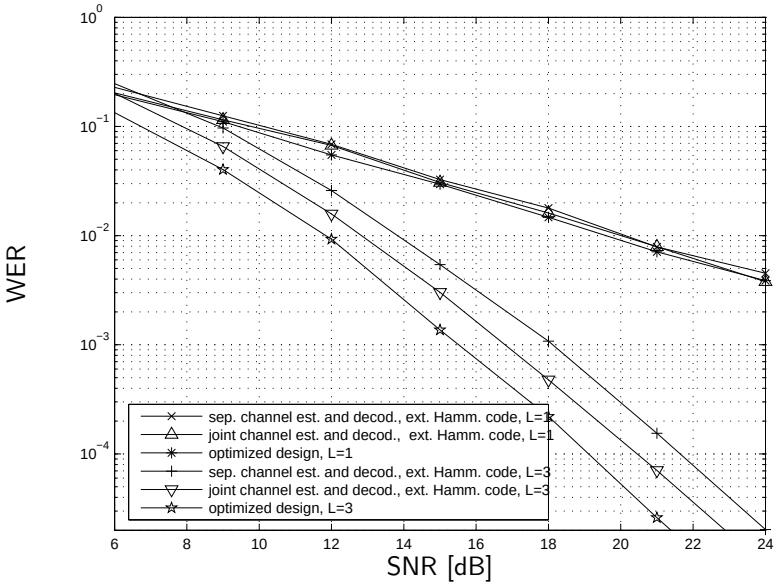


Figure 3.5: Performance comparison of constellations designed for $L = 1$ and $L = 3$ vs. extended Hamming code.

both Figures 3.5 and 3.6 indicates the value of L which shows the benefit of multipath diversity. Our designed constellations clearly outperform both training-based schemes for $L = 3$ whereas the gain appears to be minor for the frequency-flat case $L = 1$. The gain is even more significant when comparing our scheme to the training-based design using separate channel estimation and decoding.

Some more results on SISO systems were obtained when comparing with constellations designed by Hochwald et al. [HMR+00] which are defined for both single-antenna as well as multi-antenna systems. We will present the simulation results in comparison with these signal sets in the following section.

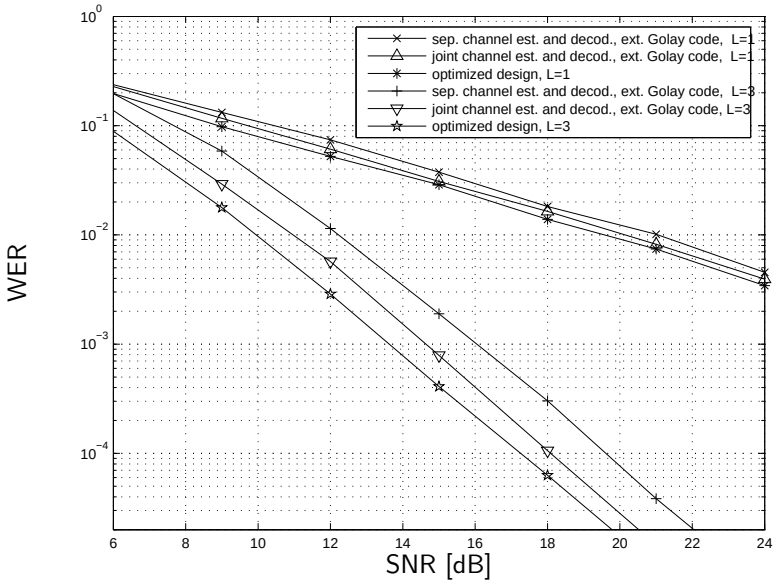


Figure 3.6: Performance comparison of constellations designed for $L = 1$ and $L = 3$ vs. extended Golay code.

Constellations for Unknown Channels

For flat fading Rayleigh channels, a design method to obtain a set of signaling matrices was described by Hochwald et al. in [HMR+00] and we will refer to these constellations as *Hochwald constellations*. All signal matrices are generated by successively rotating an initial signal matrix using diagonal unitary matrices. The initial signal matrix together with the rotation matrix describe the constellation. We compare our design to constellations with $T = 8$ and $Z = 256$ for $M_T = 1$ and $M_T = 2$ that were also used as illustrative examples in [HMR+00]. All simulations assume independently identically Rayleigh distributed channel parameters $\mathbf{h}$ with covariance matrix $\mathbf{C_h} = \mathbf{I}_{LM_T}/L$.

A WER performance comparison of the single antenna Hochwald constellations with signal sets obtained using gradient search is given in Figure 3.7. It turns out that the WER performance difference between our

scheme and the Hochwald constellation is negligible for the flat fading case $L = 1$. This is radically different from the $L = 3$ case where the

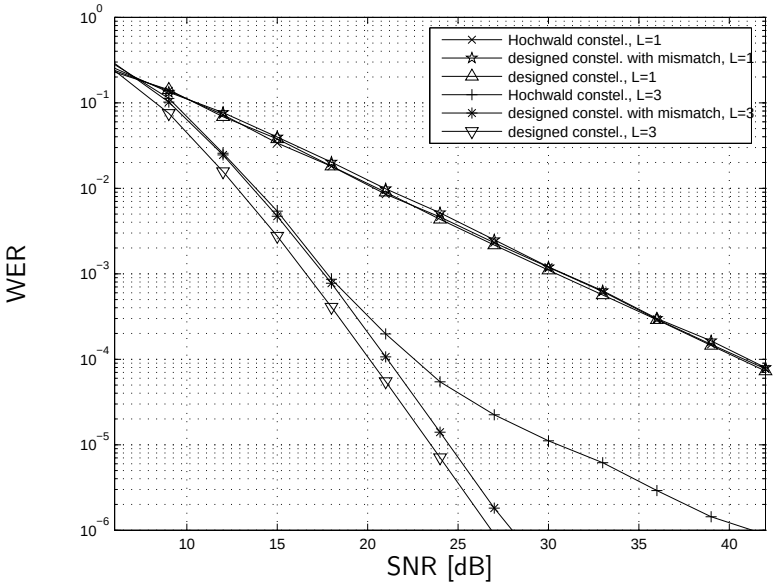


Figure 3.7: Performance comparison with Hochwald constellations for $M_T = 1$ and $L = 1$ and $L = 3$.

slope of the performance curve related to the Hochwald constellation flattens out to have a slope corresponding to $L = 1$ at an SNR of about 25 dB. This behavior can be explained with the fact that the constellation design was not oriented towards frequency-selective channels. In fact, the constellation does not offer full multipath diversity over a frequency-selective channel. The reason for this is that the signal matrices in the Hochwald constellation do not always fulfill the criterion of full rank $\mathbf{C}_{i,j} = [\overline{\mathbf{S}}_i \quad \overline{\mathbf{S}}_j]$, see Section 2.6. As an example, the signals

$$\mathbf{S}_{c_1} = [-1, 1, -1, 1, -1, 1, -1, 1]^T$$

and

$$\mathbf{S}_{c_0} = [1, 1, 1, 1, 1, 1, 1, 1]^T$$

result in a matrix $[\bar{\mathbf{S}}_{c_1} \quad \bar{\mathbf{S}}_{c_0}]$ which is rank-deficient for $L = 3$ preventing the constellation from exploiting full diversity. We also investigated mismatch between the design assumption on L in our scheme. The curves annotated "designed constellations with mismatch" in Figure 3.7 refer to the scenario of applying a constellation designed for $L = 3$ over a channel with $L = 1$ and vice versa. It can be seen in Figure 3.7 that the signal set designed for $L = 1$ offers full diversity gain $L = 3$ in the case of mismatched channel length. There is however a loss of around 1 dB compared to the constellation with matched design. No significant differences can be observed in the frequency-flat case $L = 1$.

The performance results for $M_T = 2$ in terms of WER for the Hochwald constellations as well as the signals set obtained using gradient search are summarized in Figure 3.8. For $L = 1$ channel tap, the

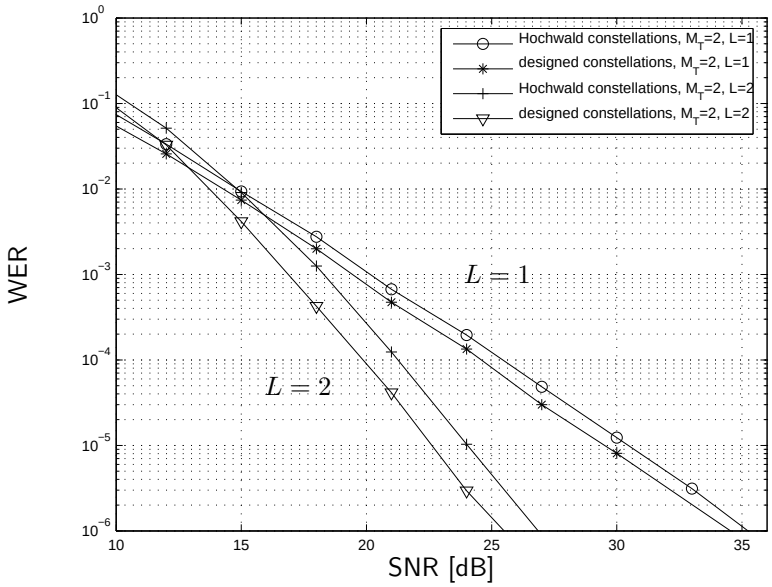


Figure 3.8: Performance comparison with Hochwald codes for $M_T = 2$ over channels with $L = 1$ and $L = 2$.

code designed using gradient search performs ca. 1 dB better than the

M_T	Z	T	L	E/e	$\bar{\kappa}$
1	2^4	8	1	1.0000	1.0
		9	2	1.0025	1.0079
		10	3	1.0067	1.0113
		11	4	1.0083	1.1159
	2^{16}	24	1	1.0031	1.0
		25	2	1.0476	1.0133
		26	3	1.1445	1.0981
		27	4	1.1592	1.1854
	2^8	8	1	1.0031	1.0
		8	3	1.0091	1.0827
2	2^8	8	1	1.0028	1.0
		8	2	1.0236	1.3542
		9	2	1.0116	1.1800

Table 3.1: Some characteristics of the designed constellations

Hochwald constellation. For $L = 2$ taps, the gain is ca. 1.4 dB, illustrating the superior performance obtained when using signal design matched to the given scenario.

3.3.5 Characterization of the Obtained Constellations

We have compared in this section signal sets found using gradient search with several benchmark schemes. The gain between our scheme and the benchmark schemes became particularly apparent for frequency-selective ($L > 1$) channels in both single and multi antenna systems.

We investigated the properties of all designed constellations that we have presented in this chapter using the two characteristic quantities defined in Section 3.3.3. The results are given in Table 3.1. The larger L, the larger the value for $\bar{\kappa}$. We can conclude that the matrices $\overline{\mathbf{S}}$ are close to (but not exactly) unitary, and the stronger the influence of the frequency-selective channel, i.e., the larger L, the less important is orthogonality between columns in $\overline{\mathbf{S}}$. Although this effect is rather small, it provides some insight into the nature of the signal design and can possibly be exploited in future design approaches. We also note that the power is not uniformly distributed among the signals. In particular, the longer the channel, the higher the spread in power between different

signal matrices leading to a similar conclusion that the design criteria move away the further from orthogonal columns in $\overline{\mathbf{S}}$, the stronger the influence of the frequency-selective characteristic of the channel is.

3.4 Comparison between Optimization Approaches

Having presented two design approaches to constellation design for unknown frequency-selective channels, a comparison between the design method in Section 3.2 and 3.3 is illustrative. Whereas the first approach is restricted to a fixed symbol alphabet, the second approach allows arbitrary complex valued symbols (up to power constraint on the entire constellation). The first design criterion is based on the exact union bound which can be evaluated at any SNR, whereas the second approach aims at optimizing an approximation of the union bound which is valid only in the high SNR regime. The second approach imposes design requirements on the signals such that full diversity performance is guaranteed. No such requirements or guarantees apply for the first approach. Due to the different character of the search space, different optimization methods were used (simulated annealing and constrained gradient search). Compared to suitable benchmark schemes, the resulting constellations showed excellent detection error performance.

Both approaches suffered from design complexity problems for large constellations, either in memory requirements for storing large tables (approach 1) or in evaluating the objective function and its gradient (approach 2). The example constellations presented in this chapter with up to 2048 or 4096 distinct signals were testing the limits of the available computer hardware. Moreover, both design approaches resulted in constellations that were unstructured in the sense that an optimal detection required exhaustive search over all possibly transmitted signals. It is therefore a general conclusion that it is highly desirable to investigate constellations that reduce detection as well as design complexity while maintaining excellent detection performance.

In order to compare both approaches in terms of the error performance of the resulting code, we have chosen a scenario with $L = 3$ and $T = 10$ to allow a comparison of both approaches with a standard scheme based on the $(7, 4)$ Hamming code. Since the search space of the first approach is restricted to BPSK symbols and therefore a subset of the search space used in gradient search, we expect the constellation designed using

gradient search to perform at least as good as the constellation based on BPSK, at least for high SNR. The results are shown in Figure 3.9. It

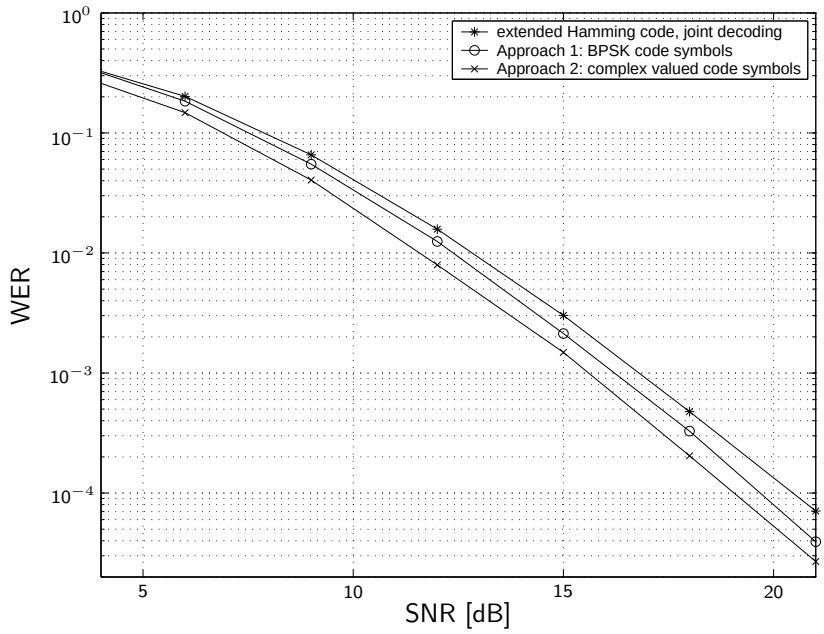


Figure 3.9: Comparison of the presented signal design approaches for $T = 10$, $M_T = 1$ and $L = 3$.

appears that in this chosen scenario, the constellation based on gradient search performs about 0.5dB better than the constellation obtained using simulated annealing and around 1 dB better than the training-based Hamming code (using optimized training according to [ATV02] and using the joint decoder). It is a matter of further research to determine theoretically in how far the preselection of the symbol alphabet necessarily implies losses.

Appendix 3.A Derivation of the Gradient

Let $\mathbf{A}(\theta)$ be a square invertible matrix dependent on a real parameter θ and let $\partial\mathbf{A}/\partial\theta$ denote the matrix obtained from $\mathbf{A}$ by element-wise differentiation. Then the following rules apply (see [Gra81] and e.g. [MN88]):

$$\begin{aligned}
\frac{\partial|\mathbf{A}|}{\partial\theta} &= |\mathbf{A}|\mathrm{tr}\left(\frac{\partial\mathbf{A}}{\partial\theta}\mathbf{A}^{-1}\right)\\
\frac{\partial}{\partial\theta}\frac{1}{|\mathbf{A}|^{M_R}} &= -M_R\frac{1}{|\mathbf{A}|^{M_R}}\mathrm{tr}\left(\frac{\partial\mathbf{A}}{\partial\theta}\mathbf{A}^{-1}\right)\\
\frac{\partial\mathbf{P}_A}{\partial\theta} &= \mathbf{P}_A^{\perp}\frac{\partial\mathbf{A}}{\partial\theta}\mathbf{A}^{\dagger} + (\dots)^H\\
\frac{\partial\mathbf{P}_A^{\perp}}{\partial\theta} &= -\frac{\partial\mathbf{P}_A}{\partial\theta}
\end{aligned}$$

where the $(\dots)^H$ notation indicates the Hermitian transpose of the previous expression. Now let θ_k denote an arbitrary real parameter in the signal matrix $\bar{\mathbf{S}}_k$. Dropping the terms independent of θ_k, the derivative of the cost function in (3.9) is

$$\frac{\partial J}{\partial\theta_k} = \frac{\partial}{\partial\theta_k}\sum_{j\neq k}\left(\frac{1}{|\bar{\mathbf{S}}_k^H\mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp}\bar{\mathbf{S}}_k|^{M_R}} + \frac{1}{|\bar{\mathbf{S}}_j^H\mathbf{P}_{\bar{\mathbf{S}}_k}^{\perp}\bar{\mathbf{S}}_j|^{M_R}}\right)$$

which leads to

$$\begin{aligned}
\frac{\partial J}{\partial\theta_k} =& -M_R\sum_{j\neq k}\frac{\mathrm{tr}\left(\left[\frac{\partial}{\partial\theta_k}\left(\bar{\mathbf{S}}_k^H\mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp}\bar{\mathbf{S}}_k\right)\right](\bar{\mathbf{S}}_k^H\mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp}\bar{\mathbf{S}}_k)^{-1}\right)}{|\bar{\mathbf{S}}_k^H\mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp}\bar{\mathbf{S}}_k|^{M_R}}\\
&-M_R\sum_{j\neq k}\frac{\mathrm{tr}\left(\left[\frac{\partial}{\partial\theta_k}\left(\bar{\mathbf{S}}_j^H\mathbf{P}_{\bar{\mathbf{S}}_k}^{\perp}\bar{\mathbf{S}}_j\right)\right](\bar{\mathbf{S}}_j^H\mathbf{P}_{\bar{\mathbf{S}}_k}^{\perp}\bar{\mathbf{S}}_j)^{-1}\right)}{|\bar{\mathbf{S}}_j^H\mathbf{P}_{\bar{\mathbf{S}}_k}^{\perp}\bar{\mathbf{S}}_j|^{M_R}}
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial}{\partial\theta_k}(\bar{\mathbf{S}}_k^H\mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp}\bar{\mathbf{S}}_k) &= \left(\frac{\partial}{\partial\theta_k}\bar{\mathbf{S}}_k^H\right)\mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp}\bar{\mathbf{S}}_k + \bar{\mathbf{S}}_k^H\mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp}\left(\frac{\partial}{\partial\theta_k}\bar{\mathbf{S}}_k\right)\\
\frac{\partial}{\partial\theta_k}(\bar{\mathbf{S}}_j^H\mathbf{P}_{\bar{\mathbf{S}}_k}^{\perp}\bar{\mathbf{S}}_j) &= \bar{\mathbf{S}}_j^H\left((-\mathbf{P}_{\bar{\mathbf{S}}_k}^{\perp}\left(\frac{\partial}{\partial\theta_k}\bar{\mathbf{S}}_k\right)\bar{\mathbf{S}}_k^{\dagger}) + (\dots)^H\right)\bar{\mathbf{S}}_j
\end{aligned}$$

Now the complex derivative with respect to s_{tmk} defined in (3.10) as

scaled sum of the real derivatives turns out to be

$$\frac{\partial}{\partial s_{tmk}}(\bar{\mathbf{S}}_k^H \mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp} \bar{\mathbf{S}}_k) = \left(\frac{\partial \bar{\mathbf{S}}_k^H}{\partial s_{tmk}}\right) \mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp} \bar{\mathbf{S}}_k$$
$$\frac{\partial}{\partial s_{tmk}}(\bar{\mathbf{S}}_j^H \mathbf{P}_{\bar{\mathbf{S}}_k}^{\perp} \bar{\mathbf{S}}_j) = -\bar{\mathbf{S}}_j^H \bar{\mathbf{S}}_k^{\dagger H} \left(\frac{\partial \bar{\mathbf{S}}_k^H}{\partial s_{tmk}}\right) \mathbf{P}_{\bar{\mathbf{S}}_k}^{\perp} \bar{\mathbf{S}}_j.$$

The derivative of the cost function J can then be expressed as

$$\frac{\partial J}{\partial s_{tmk}} = -M_R \sum_{j \neq k} \operatorname{tr}\left(\left(\frac{\partial \bar{\mathbf{S}}_k^H}{\partial s_{tmk}}\right) \frac{\mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp} \bar{\mathbf{S}}_k (\bar{\mathbf{S}}_k^H \mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp} \bar{\mathbf{S}}_k)^{-1}}{|\bar{\mathbf{S}}_k^H \mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp} \bar{\mathbf{S}}_k|^{M_R}}\right)$$
$$+ M_R \sum_{j \neq k} \operatorname{tr}\left(\left(\frac{\partial \bar{\mathbf{S}}_k^H}{\partial s_{tmk}}\right) \frac{\mathbf{P}_{\bar{\mathbf{S}}_k}^{\perp} \bar{\mathbf{S}}_j (\bar{\mathbf{S}}_j^H \mathbf{P}_{\bar{\mathbf{S}}_k}^{\perp} \bar{\mathbf{S}}_j)^{-1} \bar{\mathbf{S}}_j^H \bar{\mathbf{S}}_k^{H\dagger}}{|\bar{\mathbf{S}}_j^H \mathbf{P}_k^{\perp} \bar{\mathbf{S}}_j|^{M_R}}\right).$$

Defining

$$\mathbf{F}_{jk} = \frac{\mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp} \bar{\mathbf{S}}_k (\bar{\mathbf{S}}_k^H \mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp} \bar{\mathbf{S}}_k)^{-1}}{|\bar{\mathbf{S}}_k^H \mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp} \bar{\mathbf{S}}_k|^{M_R}},$$

this can also be written as

$$\frac{\partial C}{\partial s_{tmk}} = -M_R \operatorname{tr}\left(\left(\frac{\partial \bar{\mathbf{S}}_k^H}{\partial s_{tmk}}\right) \sum_{j \neq k} \left(\mathbf{F}_{jk} - \mathbf{F}_{kj} \bar{\mathbf{S}}_j^H \bar{\mathbf{S}}_k^{\dagger H}\right)\right)$$

Note that $\partial \bar{\mathbf{S}}_k^H / \partial s_{tmk}$ is a selection matrix in the sense that all rows have at most one element equal to one and all other elements equal to zero. In order to compute $\partial C / \partial s_{tmk}$ for all t, m, it is thus sufficient to compute

$$-M_R \sum_{j \neq k} \left(\mathbf{F}_{jk} - \mathbf{F}_{kj} \bar{\mathbf{S}}_j^H \bar{\mathbf{S}}_k^{\dagger H}\right)$$

and add up the appropriate elements of this matrix to arrive at the desired derivative.

Appendix 3.B Proof of Lemma 1

For all $0 \leq k < Z$ let

$$\mathbf{S}_k^H \mathbf{S}_k = \mathbf{\Psi}_k \mathbf{\Lambda}_k \mathbf{\Psi}_k^H$$

with unitary $\mathbf{\Psi}_k$ and diagonal $\mathbf{\Lambda}_k$ with positive diagonal elements. The power of the signal is

$$E_k = \mathrm{tr}\left(\mathbf{S}_k^H \mathbf{S}_k\right) = \mathrm{tr}\left(\mathbf{\Lambda}_k\right).$$

Define the set $\tilde{\mathbb{S}} = \{\tilde{\mathbf{S}}_0, \ldots, \tilde{\mathbf{S}}_{Z-1}\}$ with

$$\tilde{\mathbf{S}}_k = \sqrt{\frac{E_k}{M_T}} \mathbf{S}_k \mathbf{\Psi}_k \mathbf{\Lambda}_k^{-1/2}. \tag{3.12}$$

Note that $\mathbf{S}_k$ and $\tilde{\mathbf{S}}_k$ have the same power since

$$\mathrm{tr}\left(\tilde{\mathbf{S}}_k^H \tilde{\mathbf{S}}_k\right) = \mathrm{tr}\left(\sqrt{E_k/M_T}\mathbf{I}_{M_T}\sqrt{E_k/M_T}\right) = E_k.$$

The $\bar{\tilde{\mathbf{S}}}_k$ associated with $\tilde{\mathbf{S}}_k$ via (2.7) can then be written as

$$\bar{\tilde{\mathbf{S}}}_k = \bar{\mathbf{S}}_k(\mathbf{I}_L \otimes \sqrt{E_k/M_T}\mathbf{\Psi}_k \mathbf{\Lambda}_k^{-1/2})$$

and it easily follows that

$$\mathbf{P}_{\bar{\tilde{\mathbf{S}}}_k}^{\perp} = \mathbf{I}_{T+L-1} - \bar{\tilde{\mathbf{S}}}_k(\bar{\tilde{\mathbf{S}}}_k^H \bar{\tilde{\mathbf{S}}}_k)^{-1}\bar{\tilde{\mathbf{S}}}_k^H = \mathbf{I}_{T+L-1} - \bar{\mathbf{S}}_k(\bar{\mathbf{S}}_k^H \bar{\mathbf{S}}_k)\bar{\mathbf{S}}_k^H = \mathbf{P}_{\bar{\mathbf{S}}_k}^{\perp}.$$

The approximation of the PEP in (3.8) for the detection of $\bar{\tilde{\mathbf{S}}}_j$ when $\bar{\tilde{\mathbf{S}}}_i$ has been transmitted becomes

$$\begin{aligned}
\Pr^a(\tilde{\mathbf{S}}_i \to \tilde{\mathbf{S}}_j) &= \frac{V(\sigma^2)^{LM_TM_R}}{|\bar{\tilde{\mathbf{S}}}_i^H \mathbf{P}_{\bar{\tilde{\mathbf{S}}}_j}^{\perp} \bar{\tilde{\mathbf{S}}}_i|^{M_R}} \\
&= \frac{V(\sigma^2)^{LM_TM_R}}{|\bar{\mathbf{S}}_i^H \mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp} \bar{\mathbf{S}}_i|^{M_R} |\mathbf{I}_L \otimes \frac{E_i}{M_T}\mathbf{\Psi}_i^H \mathbf{\Lambda}_i^{-1}\mathbf{\Psi}_i|^{M_R}} \\
&= \frac{V(\sigma^2)^{LM_TM_R}}{|\bar{\mathbf{S}}_i^H \mathbf{P}_{\bar{\mathbf{S}}_j}^{\perp} \bar{\mathbf{S}}_i|^{M_R} |\frac{E_i}{M_T}\mathbf{\Lambda}_i^{-1}|^{LM_R}} \\
&\leq \Pr^a(\mathbf{S}_i \to \mathbf{S}_j)
\end{aligned}$$

since

$$\left|\frac{E_i}{M_T}\mathbf{\Lambda}_i^{-1}\right| = \frac{\left(\frac{E_i}{M_T}\right)^M}{|\mathbf{\Lambda}_i|} \geq 1$$

because of the inequality of geometric and arithmetic mean. Applying the transformation in (3.12) to all signal matrices $\mathbf{S} \in \mathbb{S}$ will thus result in a cost function value $J(\tilde{\mathbb{S}}) \leq J(\mathbb{S})$. ■

Chapter 4

Design in the Frequency Domain

The approaches to signal design for communication over unknown frequency-selective channels that were presented in Chapter 3 resulted in constellations with excellent performance. However, these constellations were unstructured in the sense that no suitable rule was imposed on the signals that would allow detection with a complexity that is significantly lower than exhaustive search. The signals were designed such that the data influenced all symbols transmitted via all antennas and required a signal set of considerable size as well as detection algorithms of high complexity. One way to decrease the complexity of the signal design and potentially also the detection problem is to partition the transmitted signal into parts that are (in some sense) independent of one another and can therefore be coded independently. Since these partial coding problems are smaller in size, they require smaller signal sets and therefore smaller design complexity.

A subdivision of the signal design problem leading to a simple detection algorithm is difficult to achieve directly in the time domain due to the substructure in the equivalent signal matrix $\bar{\mathbf{S}}$ in (2.7). This substructure implies that most received symbols depend on symbols transmitted in neighboring time intervals (an effect sometimes called "intersymbol interference"), complicating the detection and signal design problem. A standard way of overcoming this difficulty is the use of OFDM which effectively transforms the frequency-selective channel into a set of orthogonal

subchannels which can each be considered frequency-flat. Then, the signals received in each subchannel are not influenced by signals transmitted in neighboring subchannels.

As might be expected, such an advantage does not come without a drawback. Apparently, partitioning the signal design problem into small subproblems which are solved independently implies a loss of design freedom. Therefore, the obtained signal sets have potentially worse performance than signal sets operating over the entire signal block. Moreover, the usage of suboptimal detection algorithms with lower complexity can lead to a further loss in performance.

A main goal of this chapter is the investigation of this complexity-performance tradeoff in the specific application of space–frequency differential transmission. Based on the intuition that the channel gains on neighboring subchannels are highly correlated, constellations designed for differential space–time transmission are applied as differential space–frequency constellations. The usage of these constellations thereby allows the application of a detection algorithm which no longer requires a complexity that grows exponentially with the size of the signal block. The parametric description of the channel model is exploited for a performance analysis and a sufficient criterion for full diversity performance is derived. Numerical examples confirm our theoretical results concerning the exploitation of diversity and provide a comparison with different signaling methods that require higher detection complexity.

The remainder of this chapter is organized as follows. After a short review of the data model in Section 4.1, we describe two concepts required for our presentation of space–frequency transmission. First, the idea in [BB04] of forming equivalent space–frequency signal matrices with orthonormal columns is presented. Second, the standard theory of differential space–time differential transmission is reviewed in Section 4.2. Both concepts are combined in our presentation of space–frequency differential transmission that we discuss in detail in Section 4.3. Then, an error analysis in Section 4.4 results in the derivation of a sufficient criterion for full diversity performance if low-complexity differential detection is applied. Numerical performance results in Section 4.5 illustrate the value of this criterion and provide a comparison to other signaling approaches operating in the frequency domain before we conclude in Section 4.6.

4.1 Data Model

As mentioned in Section 2.4, the data model for the OFDM system with N carriers over a channel with L matrix-valued channel taps $\mathbf{H}_0, \ldots, \mathbf{H}_{L-1}$ can be summarized (see (2.9)) as

$$\mathbf{Y} = [\mathbf{S} \quad \mathbf{D}_N\mathbf{S} \ \ldots \ \mathbf{D}_N^{L-1}\mathbf{S}] \begin{bmatrix} \mathbf{H}_0 \\ \vdots \\ \mathbf{H}_{L-1} \end{bmatrix} + \mathbf{W}$$

where the element $\{\mathbf{S}\}_{n,m}$ describes the symbol transmitted over subcarrier $n-1$ on antenna m, the elements of $\mathbf{Y}$ and $\mathbf{W}$ are the received symbols and noise, respectively on the respective subcarrier and $\mathbf{D}_N = \mathrm{diag}\{e^{-j\frac{2\pi}{N}k}\}_{k=0}^{N-1}$ is the frequency-domain equivalent of a cyclic shift matrix. This model is equivalent to transmitting via $P = LM_T$ "virtual" antennas the $N \times P$ *equivalent signal matrix* (or *pseudocodeword matrix* in [BB04])

$$\mathbf{G}_{N,L}(\mathbf{S}) \triangleq [\mathbf{S} \quad \mathbf{D}_N\mathbf{S} \quad \ldots \quad \mathbf{D}_N^{L-1}\mathbf{S}] \tag{4.1}$$

over the $P \times M_R$ frequency-flat channel

$$\mathbf{H} \triangleq [\mathbf{H}_0^T \ \ldots \mathbf{H}_{L-1}^T]^T$$

resulting in

$$\mathbf{Y} = \mathbf{G}_{N,L}(\mathbf{S})\mathbf{H} + \mathbf{W}. \tag{4.2}$$

The model in (4.2) illustrates the analogy between space–time and space–frequency coding. The pseudocodeword $\mathbf{G}_{N,L}(\mathbf{S})$ in space–frequency coding assumes the role of the signaling matrix $\bar{\mathbf{S}}$ in space–time coding and is its frequency-domain counterpart when using OFDM. Similar to the equivalent signal matrix $\bar{\mathbf{S}}$ in the time domain, there is a structural constraint on $\mathbf{G}_{N,L}(\mathbf{S})$ in (4.1), preventing the direct application of signal design approaches that were conceived for space–time transmission over frequency-flat fading channels.

Still, the analogy of the system model in (4.2) has been exploited to motivate signal design in the frequency-selective case based on constructions for frequency-flat channels. If the channel is frequency-flat and constant during the transmission of the entire signal matrix but changes to an independent realization for each codeword, it was argued in [HM00,MH99] based on capacity considerations that the signaling matrices for space–time transmission limited by a power constraint should

be scaled unitary, i.e., have orthonormal columns. This property was subsequently imposed during the design of signaling matrices in [HMR+00] that can be summarized as

$$\mathbf{S}_k = \mathbf{\Theta}^k \mathbf{S}_0 \tag{4.3}$$

where $\mathbf{\Theta}$ is an $N \times N$ diagonal matrix satisfying $\mathbf{\Theta}^H \mathbf{\Theta} = \mathbf{I}$ and $\mathbf{S}_0$ (and therefore all $\mathbf{S}_k$) satisfy $\mathbf{S}_k^H \mathbf{S}_k = N/M_T \mathbf{I}_{M_T}$. The diagonal elements of $\mathbf{\Theta}$ and the design of $\mathbf{S}_0$ determine the entire constellation. Using these signals in a frequency-selective context, it is apparent that

$$\mathbf{G}_{N,L}(\mathbf{S}_k) = \mathbf{\Theta}^k \mathbf{G}_{N,L}(\mathbf{S}_0)$$

because the diagonal matrices $\mathbf{\Theta}^k$ and $\mathbf{D}_N$ commute when inserting (4.3) into (4.1). The columns of the equivalent signal matrix $\mathbf{G}_{N,L}(\mathbf{S})$ are in general not orthonormal because it is not guaranteed that $\mathbf{G}_{N,L}(\mathbf{S}_0)$ has orthonormal columns if $\mathbf{S}_0$ does. Still, for a specific choice of $\mathbf{S}_0$, this property can be assured. Define the $N \times N$ Fourier transform matrix $\mathbf{F}_N$ with elements

$$\{\mathbf{F}_N\}_{p,q} = \frac{1}{\sqrt{N}} \exp\left(-j\frac{2\pi}{N}(p-1)(q-1)\right) \tag{4.4}$$

and let its k-th column vector be denoted as $\mathbf{f}_k$. The matrix $\mathbf{F}_N$ is unitary, i.e., $\mathbf{F}_N^H \mathbf{F}_N = \mathbf{I}$. Moreover, we have the property

$$\mathbf{D}_N \mathbf{f}_k = \mathbf{f}_{k+1}$$

which is exploited in the following choice of $\mathbf{S}_0$. Setting

$$\mathbf{S}_0 = \sqrt{\frac{T}{M_T}}[\mathbf{f}_1 \ \mathbf{f}_{1+L} \ \ldots \mathbf{f}_{(M_T-1)L}]$$

the matrix $\mathbf{G}_{N,L}(\mathbf{S}_0)$ contains LM_T scaled distinct column vectors of $\mathbf{F}$. Since $\mathbf{F}$ is unitary, $\mathbf{G}_{N,L}(\mathbf{S}_0)$ has orthogonal columns. This construction was exploited in [BB04] to design constellations for frequency-selective channels unknown to the receiver in the form of (4.3), i.e., the matrix $\mathbf{\Theta}$ was determined for some example values of the parameters Z, N, M_T and L. The section of differential space–frequency transmission makes use of a similar construction in order to exploit the analogy to space–time differential transmission over frequency-flat channels.

It is important to note that the restriction to signals with orthonormal columns is, similar as in the design approach in the time domain, not

strictly motivated by capacity considerations. However, it was noted earlier that for signals that satisfy this constraint, the ML and GLRT receiver coincide and can be written in rather simple form. Moreover, performance analysis is somewhat simplified (see [BB04]). Therefore, constructions restricted to this constraint will be considered here.

4.2 A Short Review of Differential Space–Time Transmission

Differential transmission is a standard approach if accurate estimation of the channel at the receiver may require an intolerable amount of training symbols or may be infeasible because the channel fading is too fast. Differential coding [Pro95] is a well-known method for single–input single–output frequency-flat channels in this scenario and has been generalized to multiple–input multiple–output (MIMO) channels in e.g., [HS00, Hug00] allowing the exploitation of space diversity. The general idea of space–time differential transmission in flat Rayleigh fading can be summarized as follows [HS00]: Assume that two $M_T \times M_T$ signal blocks $\mathbf{S}_{k-1}$ and $\mathbf{S}_k$ are transmitted through the same channel $\tilde{\mathbf{H}}$ (which is unknown to the receiver) resulting in the data model

$$\begin{aligned}\mathbf{Y}_{k-1} &= \mathbf{S}_{k-1}\tilde{\mathbf{H}} + \mathbf{W}_{k-1}\\ \mathbf{Y}_k &= \mathbf{S}_k\tilde{\mathbf{H}} + \mathbf{W}_k.\end{aligned}$$

which can just as well be understood as the reception of signals in the combined matrix $\mathbf{Y} = [\mathbf{Y}_{k-1}^T \mathbf{Y}_k^T]^T$ as a result of transmitting the signal $\mathbf{S} = [\mathbf{S}_{k-1}^T \mathbf{S}_k^T]^T$. Note that the subscript here indicates the sequential order of the two matrices, not their index in the signaling alphabet. As mentioned before, in Rayleigh flat fading the signal matrix should have orthonormal columns, i.e., $\mathbf{S}^H\mathbf{S} = 2\mathbf{I}_{M_T}$ which is most easily satisfied if $\mathbf{S}_k^H\mathbf{S}_k = \mathbf{S}_{k-1}^H\mathbf{S}_{k-1} = \mathbf{I}$ (see also [HS00]). The GLRT receiver (as well as the ML receiver if i.i.d. fading is assumed) reduces then to

$$\begin{aligned}(\hat{\mathbf{S}}_{k-1}, \hat{\mathbf{S}}_k) &= \arg\max_{\mathbf{S}} \|\mathbf{P}_{\mathbf{S}}\mathbf{Y}\|_F^2\\ &= \arg\max_{\mathbf{S}_{k-1},\mathbf{S}_k} \operatorname{tr}\left(\mathbf{Y}_{k-1}^H\mathbf{S}_{k-1}\mathbf{S}_k^H\mathbf{Y}_k + \mathbf{Y}_k^H\mathbf{S}_k\mathbf{S}_{k-1}^H\mathbf{Y}_{k-1}\right).\end{aligned} \tag{4.5}$$

The decision variable of this receiver depends on the signaling matrices only via the product $\mathbf{S}_{k-1}\mathbf{S}_k^H$, i.e., the same value is assumed if both

$\mathbf{S}_{k-1}$ and $\mathbf{S}_k^H$ are multiplied by an arbitrary unitary matrix. Therefore, it can just as well be assumed that the matrix

$$\mathbf{S} \cdot \mathbf{S}_{k-1}^H = \begin{bmatrix} \mathbf{S}_{k-1} \\ \mathbf{S}_k \end{bmatrix} \mathbf{S}_{k-1}^H = \begin{bmatrix} \mathbf{I} \\ \mathbf{S}_k \mathbf{S}_{k-1}^H \end{bmatrix}$$

was transmitted. In other words, only data encoded in the product $\mathbf{S}_k \mathbf{S}_{k-1}^H$ can be detected at the receiver. This motivates the "fundamental differential transmission equation" [HS00]

$$\mathbf{S}_k = \mathbf{V}_k \mathbf{S}_{k-1}$$

where the data to be transmitted is encoded in the unitary matrix $\mathbf{V}_k$. After initializing this recursion with some arbitrary unitary $\mathbf{S}_0$ and transmitting the resulting signal matrices $\mathbf{S}_0, \ldots, \mathbf{S}_k$, any two subsequently received matrices $\mathbf{Y}_k$ and $\mathbf{Y}_{k-1}$ can be used to detect $\mathbf{V}_k$.

An alternative derivation starting from

$$\mathbf{Y}_{k-1} = \mathbf{S}_{k-1} \tilde{\mathbf{H}} + \mathbf{W}_{k-1} \tag{4.6}$$

$$\mathbf{Y}_k = \mathbf{V}_k \mathbf{S}_{k-1} \tilde{\mathbf{H}} + \mathbf{W}_k. \tag{4.7}$$

replaces $\mathbf{S}_{k-1}\mathbf{H} = \mathbf{Y}_{k-1} - \mathbf{W}_{k-1}$ from (4.6) in (4.7) leading to

$$\mathbf{Y}_k = \mathbf{V}_k \mathbf{Y}_{k-1} + \mathbf{V}_k \mathbf{W}_{k-1} + \mathbf{W}_k$$

which motivates the detector

$$\begin{aligned} \hat{\mathbf{V}}_k &= \arg \min_{\mathbf{V}_k} \|\mathbf{Y}_k - \mathbf{V}_k \mathbf{Y}_{k-1}\|^2 \\ &= \arg \max_{\mathbf{V}_k} \operatorname{tr} \left(\mathbf{Y}_k^H \mathbf{V}_k \mathbf{Y}_{k-1} + \mathbf{Y}_{k-1}^H \mathbf{V}_k^H \mathbf{Y}_k \right) \end{aligned} \tag{4.8}$$

which coincides with (4.5). Note that the derivation of these receivers is based on the assumption of constant channel $\tilde{\mathbf{H}}$ during the transmission of the signal matrices $\mathbf{S}_{k-1}$ and $\mathbf{S}_k$ if $\mathbf{V}_k$ is detected. For the detection of the data encoded in the following block $\mathbf{V}_{k+1}$, according to (4.8), constant channel during the transmission of $\mathbf{S}_k$ and $\mathbf{S}_{k+1}$ is required. Together with the assumptions used in the detection of $\mathbf{V}_k$, a constant channel during the transmission of $\mathbf{S}_{k-1}$ to $\mathbf{S}_{k+1}$ is required. Continuing this recursion, constant channel during the transmission of all signal matrices is assumed. However, if the channel is constant for a long period of time, it might also be reasonable to estimate it which in turn would imply a different transmission strategy. A very interesting application of

differential transmission is therefore the case where $\mathbf{H}$ is only *approximately* constant during the transmission of two signal matrices. Loosely speaking, the channel should be "constant enough" to allow simple differential detection but not "too constant" such that it would be beneficial to estimate it. This notion is the basis for our derivation of differential space–frequency transmission.

4.3 Differential Space–Frequency Transmission

If the channel is frequency-selective and OFDM is applied, coding over different OFDM subcarriers and antennas was proposed in, e.g., [WY02, MTL03], to allow the exploitation of additional frequency diversity, while differential coding *in time* over consecutive OFDM blocks is used to eliminate the need for channel estimation at the receiver. In contrast, we propose in the present section to use differential modulation *in frequency* within a single OFDM block based on the assumption that the complex-valued channel gains of neighboring OFDM subcarriers are highly correlated. We thereby generalize single-antenna differential modulation in the frequency domain [Lot99] to systems with multiple transmitter antennas. The advantage of this method is that the need for explicit channel estimation is removed within a single OFDM block. It can therefore also be used in the context of noncoherent MIMO-OFDM as in [BB04].

Differential modulation over space and frequency was also investigated in [LS01] in combination with a decision-feedback receiver. Moreover, the idea of coding differentially in frequency was presented in [Mar03]. We note however that neither [LS01] nor [Mar03] provides analytical criteria that can be employed in the design process to guarantee the exploitation of full space and frequency diversity. The introduction of such criteria is one main contribution of the present section. We also demonstrate the existence in some cases of an error floor at high SNR's which can be avoided by satisfying the mentioned criteria for full diversity performance.

4.3.1 Partitioning the Signal Matrix

Similar as space–time differential transmission can be thought of as splitting up a large signal block into smaller pieces that experience piecewise constant channel, we investigate the implications of splitting up the

space–frequency signal matrix into subblocks which experience at least approximately constant channel between two subsequent subblocks such that differential coding can be employed. A main contribution is an analysis concerning the resulting diversity order, explicitly taking into account effects of channel fading in the frequency-selective channel.

Assuming that the number of carriers N is an integer multiple of $P = LM_T$, i.e., $N = BP$, we partition $\mathbf{S}$ into B subblocks $\mathbf{S}_0, \ldots, \mathbf{S}_{B-1}$. Each subblock contains P rows of $\mathbf{S}$, i.e., the a subblock contains the symbols transmitted on P subcarriers via all antennas. Let $\mathbf{s}_n^T$ denote row n of $\mathbf{S}$, i.e., the symbols transmitted on subcarrier n. Let the $P \times M_T$ matrix $\mathbf{S}_k$ consist of equispaced rows of $\mathbf{S}$ spread over the whole frequency band, i.e.,

$$\mathbf{S}_k = [\mathbf{s}_k \quad \mathbf{s}_{k+B} \ \ldots \ \mathbf{s}_{k+(P-1)B}]^T.$$

This grouping of subcarriers effectively distributes the subcarriers within each block over the allocated signal bandwidth, see an example in Figure 4.1.

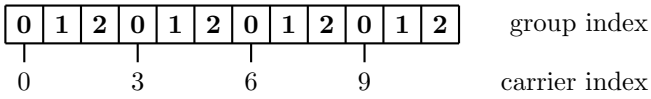


Figure 4.1: Example for the choice of subcarriers within each block for $N = 12$ subcarriers and $P = LM_T = 4$ leading to $B = 3$ groups with four subcarriers each.

Note that subsequent $\mathbf{S}_k, \mathbf{S}_{k+1}$ contain the symbols transmitted on adjacent subcarriers. Any matrix $\mathbf{S}_k$ can now be associated with the corresponding selection of rows in $\mathbf{G}_{N,L}(\mathbf{S})$, i.e.,

$$\mathbf{G}_k = [\mathbf{g}_k \quad \mathbf{g}_{k+B} \ \ldots \ \mathbf{g}_{k+(P-1)B}]^T$$

where

$$\mathbf{g}_k^T = [\mathbf{s}_k^T \quad e^{-j\frac{2\pi}{N}k}\mathbf{s}_k^T \quad e^{-j\frac{2\pi}{N}2k}\mathbf{s}_k^T \ \ldots \ e^{-j\frac{2\pi}{N}(L-1)k}\mathbf{s}_k^T].$$

Now the stacked $P \times P$ matrix $\mathbf{G}_k$ can be written as

$$\mathbf{G}_k = [\mathbf{S}_k \quad \mathbf{D}_P\mathbf{S}_k \ \ldots \ \mathbf{D}_P^{L-1}\mathbf{S}_k] \cdot \left(\operatorname{diag}\left\{e^{-j\frac{2\pi}{N}kl}\right\}_{l=0}^{L-1} \otimes \mathbf{I}_{M_T}\right).$$

For an arbitrary choice of $\mathbf{S}_k$, the resulting matrix $\mathbf{G}_k$ is in general not unitary as desired in [HS00]. However, when choosing

$$\mathbf{S}_k = \sqrt{L}\mathbf{\Phi}_k[\mathbf{f}_1\ \mathbf{f}_{1+L}\ \ldots \mathbf{f}_{(M_T-1)L}]e^{j\frac{\pi(L-1)}{N}k} \tag{4.9}$$

where $\mathbf{\Phi}_k$ is diagonal unitary and $\mathbf{f}_k$ is here the k-th column of a $P \times P$ Fourier transform matrix $\mathbf{F}_P$, we obtain the scaled unitary matrix

$$\mathbf{G}_k = \sqrt{L}\mathbf{\Phi}_k\mathbf{U}_P(\mathrm{diag}\{e^{-j\frac{2\pi}{N}kl}\}_{l=0}^{L-1} \otimes \mathbf{I}_{M_T})e^{j\frac{\pi(L-1)}{N}k}. \tag{4.10}$$

Here, $\mathbf{U}_P$ is column permuted version of $\mathbf{F}_P$ and therefore also unitary. The role of the factor $e^{j\frac{\pi(L-1)}{N}k}$ will become clear in the next section related to encoding and decoding. The factor $\sqrt{L}$ assures energy normalization such that $\mathrm{tr}\left(\mathbf{S}_k^H\mathbf{S}_k\right) = P = LM_T$.

4.3.2 Encoding and Detection

The received symbols on the subcarriers related to $\mathbf{G}_k$ and $\mathbf{G}_{k-1}$ can be written as

$$\mathbf{Y}_{k-1} = \mathbf{G}_{k-1}\mathbf{H} + \mathbf{W}_{k-1} \tag{4.11}$$

$$\mathbf{Y}_k = \mathbf{G}_k\mathbf{H} + \mathbf{W}_k. \tag{4.12}$$

where $\mathbf{Y}_k$ and $\mathbf{W}_k$ are the matrix block-partitions of $\mathbf{Y}$ and $\mathbf{W}$, respectively, which are related to $\mathbf{G}_k$ that in turn depends on $\mathbf{S}_k$ in (4.9) and therefore on the diagonal unitary matrix $\mathbf{\Phi}_k$. The channel $\mathbf{H}$ is the same as in (4.2).

In order to allow decoding without any knowledge of the channel $\mathbf{H}$, we apply differential encoding of the data symbol $\mathbf{V}_k$ in

$$\mathbf{\Phi}_k = \mathbf{V}_k\mathbf{\Phi}_{k-1} \tag{4.13}$$

with initialization $\mathbf{\Phi}_0 = \mathbf{I}_P$.

At the receiver side, replacing $\mathbf{H}$ in (4.12) with $\mathbf{G}_{k-1}^{-1}(\mathbf{Y}_{k-1} - \mathbf{W}_{k-1})$ yields

$$\mathbf{Y}_k = \mathbf{G}_k\mathbf{G}_{k-1}^{-1}(\mathbf{Y}_{k-1} - \mathbf{W}_{k-1}) + \mathbf{W}_k. \tag{4.14}$$

Now using (4.10) we can rewrite

$$\mathbf{G}_k\mathbf{G}_{k-1}^{-1} = \mathbf{\Phi}_k \underbrace{\mathbf{U}_P\left(\mathrm{diag}\left\{e^{-j\frac{2\pi}{N}\left(l-\frac{L-1}{2}\right)}\right\}_{l=0}^{L-1} \otimes \mathbf{I}_{M_T}\right)\mathbf{U}_P^H}_{\triangleq\,\mathbf{E}}\mathbf{\Phi}_{k-1}^H \tag{4.15}$$

and reformulate (4.14) as

$$\mathbf{Y}_k = \mathbf{\Phi}_k \mathbf{E} \mathbf{\Phi}_{k-1}^H \mathbf{Y}_{k-1} + \mathbf{W}_k - \mathbf{\Phi}_k \mathbf{E} \mathbf{\Phi}_{k-1}^H \mathbf{W}_{k-1} \tag{4.16}$$

The appearance of the matrix $\mathbf{E}$ in (4.16) is central to our analysis of the differential modulation scheme in the sense that it captures explicitly the effects of channel correlation on adjacent subcarriers. This is in contrast to the assumption of constant $\mathbf{H}$ between two subblocks in [HS00] in the context of analyzing space–time differential modulation. Note however that for $L = 1$, i.e., a frequency-flat channel, we have $\mathbf{E} = \mathbf{I}_P$ and a setup equivalent to the one in [HS00] results, indicating that our general framework directly includes this special case.

The key observation now is that for $N \gg L$, the arguments of the eigenvalues of $\mathbf{E}$ are small and therefore $\mathbf{E}$ is "close" to the identity matrix. In most practical systems, the number of subcarriers is much larger than the number of channel taps and therefore $N \gg L$ is satisfied. We therefore extend (4.16) to

$$\begin{aligned} \mathbf{Y}_k &= \mathbf{\Phi}_k \mathbf{\Phi}_{k-1}^H \mathbf{Y}_{k-1} + \mathbf{\Phi}_k (\mathbf{E} - \mathbf{I}) \mathbf{\Phi}_{k-1}^H \mathbf{Y}_{k-1} + \mathbf{W}_k \\ &= \mathbf{V}_k \mathbf{Y}_{k-1} + \mathbf{\Phi}_k (\mathbf{E} - \mathbf{I}) \mathbf{\Phi}_{k-1}^H \mathbf{Y}_{k-1} + \mathbf{W}_k, \end{aligned} \tag{4.17}$$

and apply the decoder

$$\hat{\mathbf{V}}_k = \arg \min_{\mathbf{V}} \| \mathbf{Y}_k - \mathbf{V} \mathbf{Y}_{k-1} \|^2 \tag{4.18}$$

similar to the decoder in space–time differential transmission in [HS00]. Note that the detection of the entire sequence of $\mathbf{V}$ as in [LLK03, SL02] is also possible. In the subsequent discussion, we will however use the decoder (4.18).

The observation that a matrix $\mathbf{E}$ which is close to the identity matrix implies a standard setup with approximately constant $\mathbf{H}$ is also the motivation for the factor $e^{j \frac{\pi (L-1)}{N} k}$ in (4.9). This factor minimizes the maximum arguments of the complex-valued eigenvalues of $\mathbf{E}$, i.e., the eigenvalues of $\mathbf{E}$ are as close as possible to one.

4.4 Performance Analysis

In order to study the performance of the detector in (4.18), we investigate the pairwise error probability (PEP) similar to the analysis in Section 2.6.

The PEP computing a value in (4.18) in favor of the wrong signal matrix $\mathbf{V}_e$ when $\mathbf{V}_c$ was transmitted can be written as

$$\Pr(\mathbf{V}_c \to \mathbf{V}_e) = \Pr\left(\|\mathbf{Y}_k - \mathbf{V}_c\mathbf{Y}_{k-1}\|^2 > \|\mathbf{Y}_k - \mathbf{V}_e\mathbf{Y}_{k-1}\|^2\right) \tag{4.19}$$

conditioned on the transmission of $\mathbf{V}_c$. We have the following

Lemma 2 *If the matrix* $\mathbf{A} = (\mathbf{I} - \mathbf{V}_c^H\mathbf{V}_e)\mathbf{E} + \mathbf{E}^H(\mathbf{I} - \mathbf{V}_c^H\mathbf{V}_e)^H$ *is positive definite, full diversity performance is achieved. If* $\mathbf{A}$ *has positive and negative eigenvalues, the PEP is bounded away from zero as* $\sigma^2 \to 0$.

Proof: See Appendix 4.A.

This lemma is the extension to space–frequency differential transmission of the statement that full diversity is exploited in flat Rayleigh fading channels if $|(\mathbf{V}_c - \mathbf{V}_e)^H(\mathbf{V}_c - \mathbf{V}_e)| = |2\mathbf{I} - \mathbf{V}_c^H\mathbf{V}_e - \mathbf{V}_e^H\mathbf{V}_c| > 0$ which is equivalent to the statement in Lemma 2 for $\mathbf{E} = \mathbf{I}$. The properties of $\mathbf{E}$ determine therefore in how far the varying channel influences the diversity performance of the used signal constellations.

An immediate conclusion of Lemma 2 is thus that it is a sufficient criterion to check full diversity performance if $\lambda_{\min}$, defined as the minimum eigenvalue among all $\mathbf{A}$ dependent on different signal matrices, is positive. Conversely, $\lambda_{\min} < 0$ corresponds to the existence of an error floor.

A particular special case is the usage of the constellations designed for flat Rayleigh fading channel. In [HS00], a set of signaling matrices which can be considered as a cyclic group was presented constructing all matrices $\mathbf{V}_k$ and therefore $\mathbf{\Phi}_k$ as matrix multiples of a "generator" matrix $\mathbf{\Theta}$ in

$$\mathbf{\Phi}_k \in \{\mathbf{\Theta}^n\}_{n=0}^{Z-1}$$

where

$$\mathbf{\Theta} = \operatorname{diag}\left\{e^{-j\frac{2\pi}{Z}q_k}\right\}_{k=1}^{P}, \quad q_k \in \{0, \ldots, Z-1\}$$

with constellation parameters $q_1, \ldots, q_P$ and size Z. An important group property here is the fact that for any two matrices within the constellation, the product of these matrices is again within the constellation. In other words, the signal set is closed under matrix multiplication. Thus, the signals generated in the fundamental differential transmission equation (4.13) never leave the set. Moreover, an evaluation of Lemma 2 for all distinct $\mathbf{V}_c$ and $\mathbf{V}_e$ can be reduced to an evaluation in the single matrix $\mathbf{V}_c^H\mathbf{V}_e$ which is again within the set. If the signal matrices are

constructed in this way, the minimum eigenvalue of $\mathbf{A}$ in Lemma 2 in the the frequency-flat case $\mathbf{E} = \mathbf{I}_P$ results in

$$\begin{aligned}\lambda_{\min} &= \min_{k\in\{1,\ldots,P\}} \min_{n\in\{1,\ldots,Z-1\}} 2 - e^{-j\frac{2\pi}{Z}q_k n} - e^{j\frac{2\pi}{Z}q_k n} \\ &= \min_{k\in\{1,\ldots,P\}} \min_{n\in\{1,\ldots,Z-1\}} 4\sin^2\left(\frac{\pi}{Z}q_k n\right) \end{aligned} \tag{4.20}$$

which means that the q_k have to be chosen such that $q_k > 0$ and Z are relatively prime in order to guarantee $\lambda_{\min} > 0$ and therefore full diversity performance. This condition was mentioned in [HS00] and used in the search for good constellation parameters.

It is now of special interest if a constellation designed to obtain full diversity performance in differential space–time transmission is guaranteed to do so when operating in space–frequency transmission over a frequency-selective channel. The following lemma is helpful for this problem relating the length of the channel, the number of carriers and the size of signaling codebook.

Lemma 3 *A constellation with parameters q_k, $k = 1, \ldots, P$ leading to a scheme with full diversity performance in the frequency-flat channel will exploit full space–frequency diversity in a frequency-selective channel if*

$$\sin^2\left(\frac{\pi}{Z}\right) > 2\sin\left(\frac{\pi(L-1)}{2N}\right). \tag{4.21}$$

Proof: We establish the above lemma using Theorem 6.3.2 in [HJ99] on eigenvalue perturbation using the spectral norm [HJ99, Section 5.6.6]

$$\|\mathbf{X}\|_S \triangleq \max\{\sqrt{\lambda} : \lambda \text{ is an eigenvalue of } \mathbf{X}^H\mathbf{X}\}.$$

We know that for any eigenvalue λ_k of

$$\begin{aligned}\mathbf{A} = {} & 2\mathbf{I} - \mathbf{V}_e^H\mathbf{V}_c - \mathbf{V}_c^H\mathbf{V}_e \\ & + (\mathbf{I} - \mathbf{V}_e^H\mathbf{V}_c)(\mathbf{E}-\mathbf{I}) + (\mathbf{E}-\mathbf{I})^H(\mathbf{I} - \mathbf{V}_c^H\mathbf{V}_e)\end{aligned}$$

there exists an eigenvalue $\tilde{\lambda}_i$ of $2\mathbf{I} - \mathbf{V}_e^H\mathbf{V}_c - \mathbf{V}_c^H\mathbf{V}_e$ such that

$$\begin{aligned}|\lambda_k - \tilde{\lambda}_i| &\le \|(\mathbf{I} - \mathbf{V}_\mathbf{e}^\mathbf{H}\mathbf{V}_\mathbf{c})(\mathbf{E}-\mathbf{I}) + (\mathbf{E}-\mathbf{I})^\mathbf{H}(\mathbf{I} - \mathbf{V}_\mathbf{c}^\mathbf{H}\mathbf{V}_\mathbf{e})\|_S \\ &\le 2\|\mathbf{E}-\mathbf{I}\|_S \, \|\mathbf{I} - \mathbf{V}_\mathbf{e}^\mathbf{H}\mathbf{V}_\mathbf{c}\|_S \\ &\le 8\sin\left(\frac{\pi(L-1)}{2N}\right)\end{aligned}$$

because $\|\mathbf{I} - \mathbf{V_e^H V_c}\|_S \leq 2$ and $\|\mathbf{E} - \mathbf{I}\|_S < 2\sin\left(\frac{\pi(L-1)}{2N}\right)$. Now since we assume full diversity performance in the frequency-flat channel, we have that $\tilde{\lambda}_i \geq 4\sin^2(\pi/Z)$ (see (4.20)) and it follows immediately that $\lambda_k > 0$ if

$$4\sin^2\left(\frac{\pi}{Z}\right) > 8\sin\left(\frac{\pi(L-1)}{2N}\right). \qquad \blacksquare$$

The factor $(L-1)/N$ describes the impact of the frequency-selective channel. The smaller the number of taps for a given number of carriers, the less pronounced are the effects of variations in the channel coefficients for adjacent carriers. In other words, the correlation of channel gains of adjacent carriers is higher and therefore the assumption of constant channel when decoding two subsequent codewords is more and more justified. On the other hand, if $(L-1)/N$ becomes large, channel gain correlation decreases and the channel varies significantly between two adjacent carriers. An apparent interpretation of Lemma 3 is that if the constellation elements $e^{-j\frac{2\pi}{Z}q_k n}$ are significantly far apart, i.e., Z is sufficiently small, the constellation design has inherent redundancy to cope with the variation of channel coefficients and therefore full diversity performance can still be achieved.

4.5 Numerical Results

We are interested in the behavior of $\lambda_{\min}$ to predict the constellation performance for different values of N and L. An illustration of $\lambda_{\min}$ is presented in Figure 4.2 when using the corresponding signal sets of [HS00, Table I] with rate one. We expect an error floor because of $\lambda_{\min} < 0$ for $N \leq 12$ and $L = 3$, for $N \leq 20$ and $L = 4$ as well as for $N \leq 50$ for $L = 5$. For $L = 4$ and $M_T = 1$, computed values of the error floor (using the union upper bound) are presented in Figure 4.3 together with simulation results in terms of symbol error rate (SER) vs. SNR $1/\sigma^2$, confirming a very good agreement between numerical and theoretical results.

In particular, no error floor results for $N \geq 24$. An example for the exploitation of full space and frequency diversity is presented in Figure 4.4 for three combinations of taps L and transmitter antennas M_T. The slope of all curves indicates the diversity order of $P = LM_T = 4$. The receiver performance is slightly worse in the case $M_T = 1, L = 4$ because the error implied in the approximation $\mathbf{E} \approx \mathbf{I}$ in the receiver is larger than in the

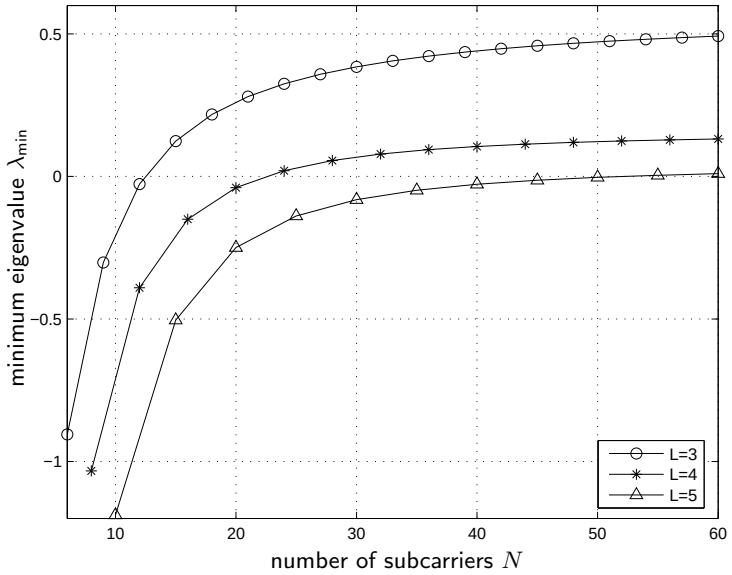


Figure 4.2: Minimum eigenvalue of all $\mathbf{A}$ dependent on the number of subcarriers N for $L = 3$, $L = 4$ and $L = 5$ where $M_T = 1$.

case $M_T = 2, L = 2$ and $M_T = 4, L = 1$. Note that the difference in receiver performance in the latter two cases is insignificant.

One of the main motivations for the application of space–frequency differential transmission was the reduced complexity both in detection as well as in signal design compared to alternative approaches similar to the ones described in Chapter 3. It is therefore beneficial to evaluate "the price" of this reduced complexity in terms of detection performance. Two alternative schemes are considered. In [BB04], a constellation design for comparatively small signal sets and restricted choice of N, L and M_T is presented which is structured in the design of the signal set but requires exhaustive search in the detection. As a second alternative, we applied the constrained gradient search algorithm presented in Chapter 3 on the design of space–frequency constellations. For $Z = 64$, $N = 8$ and $M_T = 1$ and $L = 2$, performance evaluations of the mentioned scheme

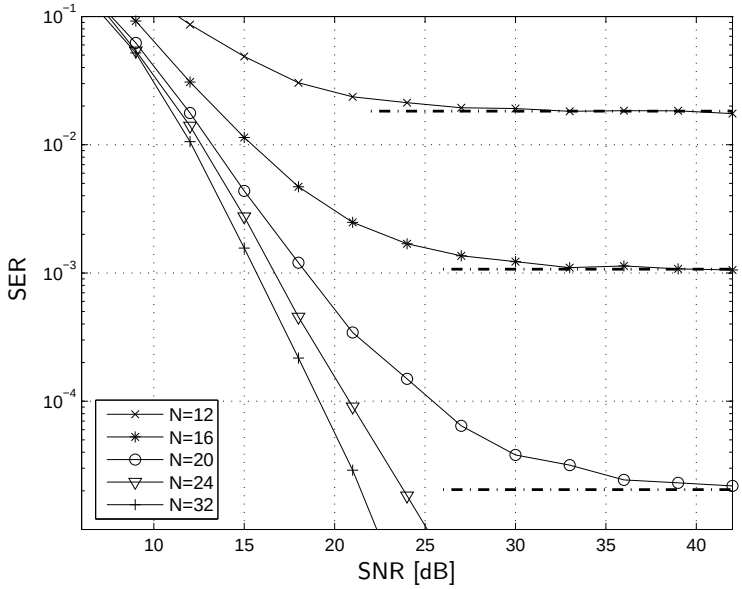


Figure 4.3: Simulation results assuming $M_R = M_T = 1$ and $L = 4$ illustrating the error floor when N is small.

together with the differential transmission are presented in Figure 4.5. Whereas the "noncoherent space-frequency codes" presented in [BB04] and the unstructured constellation designed using gradient search perform practically identical, the differential transmission scheme with the low-complexity detector in (4.18) incurs a loss of about 4.5 dB at a remaining word error rate (i.e., of the entire block) of 10^{-2} when compared to the alternative schemes. Note that with the mentioned parameters, each OFDM block is encoded using four 2×2 diagonal matrices. Using joint detection of all four subblocks instead of using (4.18) reduces the performance gap to about 1.2 dB. This loss is therefore solely due to the highly structured signal design used in differential transmission whereas the additional 3.3 dB performance is caused by the low-complexity detector.

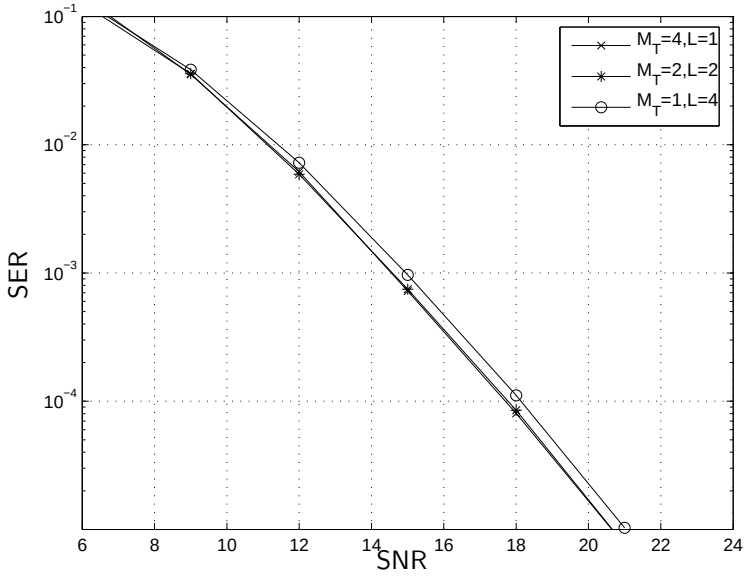


Figure 4.4: Simulation results for three combinations of frequency and transmitter diversity of combined order $P = 4$. $N = 64$ subcarriers and $M_R = 1$ antenna were assumed with $N \gg L$.

4.6 Conclusions

We have investigated space–frequency differential transmission as a low-complexity scheme for communication over unknown frequency-selective channels. Based on the intuition that the unknown channel is approximately constant on adjacent subcarriers, we employed diagonal matrices designed for space–time differential transmission in order to code differentially in space and frequency using scaled unitary equivalent signal matrices. A performance analysis of this transmission scheme revealed the possibility of an error floor if the assumption of approximately constant channel on adjacent subcarriers was not met. A sufficient criterion that guarantees the absence of an error floor as well as full diversity performance for a specific type of constellations was also derived.

In an example where two alternative signaling schemes based on much

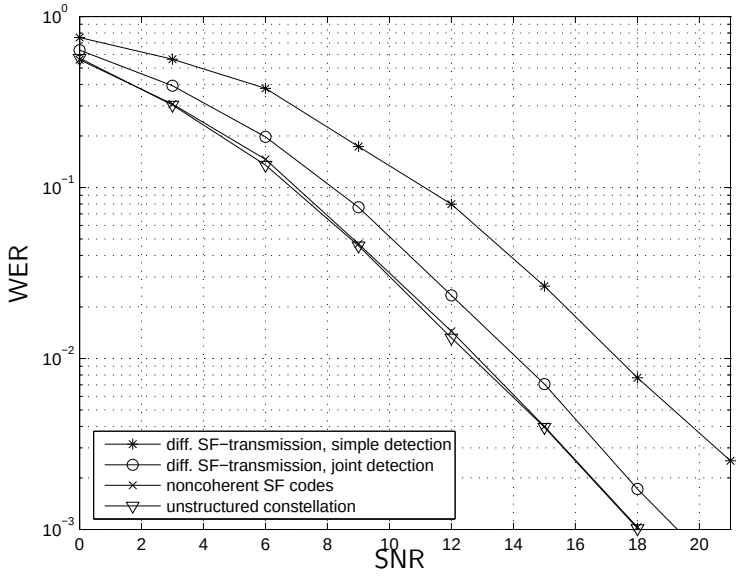


Figure 4.5: Comparison of four different signaling and detection schemes.

higher design as well as required detection complexity were available, the performance penalty incurred by differential space–frequency transmission with low complexity detection was significant. The expected tradeoff complexity vs. performance therefore appears inevitable.

Appendix 4.A Proof of Lemma 2

We first rewrite (4.11) and (4.12) as

$$\mathbf{y}_{k-1} = \tilde{\mathbf{G}}_{k-1}\mathbf{h} + \mathbf{w}_{k-1}$$
$$\mathbf{y}_{k} = \tilde{\mathbf{G}}_{k}\mathbf{h} + \mathbf{w}_{k}$$

where the vectors $\mathbf{y}_{k-1}$, $\mathbf{y}_k$, $\mathbf{h}$, $\mathbf{w}_{k-1}$ and $\mathbf{w}_k$ contain the stacked columns of $\mathbf{Y}_{k-1}$, $\mathbf{Y}_k$, $\mathbf{H}$, $\mathbf{W}_{k-1}$ and $\mathbf{W}_k$, respectively and $\tilde{\mathbf{G}}_{k-1} \triangleq \mathbf{I}_{M_R} \otimes \mathbf{G}_{k-1}$ as well as $\tilde{\mathbf{G}}_k \triangleq \mathbf{I}_{M_R} \otimes \mathbf{G}_k$. The PEP can now be expressed as

$$\Pr(\Lambda(\mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{V}_c, \mathbf{V}_e) > 0)$$

where

$$\begin{aligned}
&\Lambda(\mathbf{y}_k, \mathbf{y}_{k-1}, \mathbf{V}_c, \mathbf{V}_e) \\
&\quad \triangleq \quad \ln \frac{\exp(\|\mathbf{y}_k - (\mathbf{I}_{M_R} \otimes \mathbf{V}_c)\mathbf{y}_{k-1}\|^2)}{\exp(\|\mathbf{y}_k - (\mathbf{I}_{M_R} \otimes \mathbf{V}_e)\mathbf{y}_{k-1}\|^2)} \\
&\quad = \quad \begin{pmatrix} \mathbf{y}_k^H & \mathbf{y}_{k-1}^H \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c) \\ \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c)^H & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{y}_k \\ \mathbf{y}_{k-1} \end{pmatrix}.
\end{aligned}$$

Similar to the analysis in Section 2.6, we can express the PEP as

$$\Pr(\mathbf{V}_c \to \mathbf{V}_e) = \Pr\left(\sum \lambda_k \chi_k^2 > 0\right) \tag{4.22}$$

where λ_k are the eigenvalues of

$$\begin{pmatrix} \mathbf{0} & \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c) \\ \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c)^H & \mathbf{0} \end{pmatrix} \mathbf{C}_{\mathbf{y}} \tag{4.23}$$

with

$$\mathbf{C}_{\mathbf{y}} = \begin{pmatrix} \tilde{\mathbf{G}}_k \\ \tilde{\mathbf{G}}_{k-1} \end{pmatrix} \mathbf{C}_{\mathbf{h}} \begin{pmatrix} \tilde{\mathbf{G}}_k^H & \tilde{\mathbf{G}}_{k-1}^H \end{pmatrix} + \sigma^2 \mathbf{I}. \tag{4.24}$$

These eigenvalues can be computed in order to evaluate the PEP using numerical integration. Since the behavior of these eigenvalues is of particular importance for the high SNR case, i.e., $\sigma^2 \to 0$, we use a modified lemma from [BV01]. Given three matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ of compatible dimension and with invertible $\mathbf{BA}$, the eigenvalues of $\mathbf{AB} + \sigma^2\mathbf{C}$ are arbitrarily close to the eigenvalues of $\mathbf{BA}$ and $\sigma^2\mathbf{C}(\mathbf{I} - \mathbf{A}(\mathbf{BA})^{-1}\mathbf{B})$ for

$\sigma^2 \to 0$. Identifying

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c) \\ \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c)^H & \mathbf{0} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{G}}_k \\ \tilde{\mathbf{G}}_{k-1} \end{pmatrix} \mathbf{C_h}$$
$$\mathbf{B} = \begin{pmatrix} \tilde{\mathbf{G}}_k^H & \tilde{\mathbf{G}}_{k-1}^H \end{pmatrix}$$
$$\mathbf{C} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c) \\ \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c)^H & \mathbf{0} \end{pmatrix}$$

we can therefore conclude that the eigenvalues of (4.23) approach the eigenvalues of

$$\begin{pmatrix} \tilde{\mathbf{G}}_k^H & \tilde{\mathbf{G}}_{k-1}^H \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c) \\ \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c)^H & \mathbf{0} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{G}}_k \\ \tilde{\mathbf{G}}_{k-1} \end{pmatrix} \mathbf{C_h}$$

which can also be rewritten as

$$(\mathbf{I} \otimes (\mathbf{G}_k^H (\mathbf{V}_e - \mathbf{V}_c) \mathbf{G}_{k-1} + \mathbf{G}_{k-1}^H (\mathbf{V}_e - \mathbf{V}_c)^H \mathbf{G}_k)) \mathbf{C_h} \tag{4.25}$$

and $\sigma^2 \mathbf{C}(\mathbf{I} - \mathbf{A}(\mathbf{BA})^{-1}\mathbf{B}$. If (4.25) is indefinite, i.e. has positive and negative eigenvalues, the PEP in (4.22) can be approximated for $\sigma^2 \to 0$ using only the eigenvalues of (4.25) which do not dependend on σ^2. Since both positive and negative eigenvalues are assumed, the high SNR approximation of the PEP results in a constant value larger than zero, i.e., an error floor is obtained.

If the eigenvalues of (4.25) are all negative, it is clear that the eigenvalues of (4.23) can be upperbounded by $\sigma^2 \lambda_{\max}$ where $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{C}(\mathbf{I} - \mathbf{A}(\mathbf{BA})^{-1}\mathbf{B}$. Thus the matrix

$$\mathbf{I} - s \begin{pmatrix} \mathbf{0} & \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c) \\ \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c)^H & \mathbf{0} \end{pmatrix} \mathbf{C_y}$$

is positive definite for $s < \sigma^{-2}/\lambda_{\max}$. Then a Chernoff bound analysis along the lines in Section 2.6.3 shows that

$$\Pr(\mathbf{V}_c \to \mathbf{V}_e) = \frac{1}{\left| \mathbf{I} - s \begin{pmatrix} \mathbf{0} & \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c) \\ \mathbf{I} \otimes (\mathbf{V}_e - \mathbf{V}_c)^H & \mathbf{0} \end{pmatrix} \mathbf{C_y} \right|}$$

Choosing $s = \sigma^{-2}/(2\lambda_{\max})$ we obtain that

$$\Pr(\mathbf{V}_c \to \mathbf{V}_e) = (\sigma^2)^{LM_T M_R} Q(\mathbf{V}_c, \mathbf{V}_e, \sigma^2)$$

where

$$\lim_{\sigma^2 \to 0} Q(\mathbf{V}_c, \mathbf{V}_e, \sigma^2) = \frac{(2\lambda_{\max})^{2LM_T M_R}}{\left|(\mathbf{I} \otimes (\mathbf{G}_k^H(\mathbf{V}_c - \mathbf{V}_e)\mathbf{G}_{k-1} + \mathbf{G}_{k-1}^H(\mathbf{V}_c - \mathbf{V}_e)^H\mathbf{G}_k))\right| |\mathbf{C_h}|}$$

showing that we obtain full diversity performance under the condition that (4.25) is negative definite.

We can simplify this condition by noting that (4.25) is negative definite if $(\mathbf{G}_k^H(\mathbf{V}_e - \mathbf{V}_c)\mathbf{G}_{k-1} + \mathbf{G}_{k-1}^H(\mathbf{V}_e - \mathbf{V}_c)^H\mathbf{G}_k)$ is negative definite which is equivalent to the statement that $(\mathbf{V}_c - \mathbf{V}_e)\mathbf{G}_{k-1}\mathbf{G}_k^H + \mathbf{G}_k\mathbf{G}_{k-1}^H(\mathbf{V}_c - \mathbf{V}_e)^H$ is positive definite. Using (4.15), this is equivalent to the condition that

$$(\mathbf{I} - \mathbf{V}_c^H\mathbf{V}_e)\mathbf{\Phi}_{k-1}\mathbf{E}\mathbf{\Phi}_{k-1}^H + \mathbf{\Phi}_{k-1}^H\mathbf{E}^H\mathbf{\Phi}_{k-1}(\mathbf{I} - \mathbf{V}_e^H\mathbf{V}_c)$$

is positive definite which in turn is equivalent to the condition that

$$(\mathbf{I} - \mathbf{V}_c^H\mathbf{V}_e)\mathbf{E} + \mathbf{E}^H(\mathbf{I} - \mathbf{V}_e^H\mathbf{V}_c) \tag{4.26}$$

is positive definite (where it was exploited that $\mathbf{V}_c, \mathbf{V}_e$ and $\mathbf{\Phi}_{k-1}$ are diagonal and therefore commute). We can therefore summarize: If the matrix in (4.26) is positive definite, full diversity performance is obtained. If the matrix is indefinite, an error floor is obtained for high SNR. ■

Chapter 5

Design for Partial CSI at the Receiver

The assumption in Chapters 3 and 4 that the receiver has no CSI available for detection of the data can be considered as overly pessimistic. If the channel coefficients do not obey a strict block fading model in the sense that the realization of the channel coefficients in each block is independent of the subsequent blocks (in other words, the current channel state provides some information about the subsequent state), some kind of channel estimation or tracking might be useful at the receiver. However, the implementation of the receiver has to take the imperfections of this channel estimation into account. Having an estimate with known error statistics available at the receiver is thereby an intermediate case between the extremes of perfect CSI and absence of CSI at the receiver and can be relevant when the amount of pilot signals to estimate the channel is kept at a minimum (implying a noisy channel estimate) or the channel fades between subsequent channel uses are such that the current channel state is loosely correlated to the previous state (i.e., neither perfectly the same nor completely independent).

It appears surprising that the effort spent on designing constellations when only so-called "partial" CSI is available at the receiver has been rather limited even though related work in a single input single output system dates back as far as about 40 years [Vit65]. In an investigation on MIMO systems assuming training blocks to acquire channel knowledge, the length of a training block and its corresponding power allocation is

optimized using capacity considerations [HH03] but leaves the question open of how to design a practical coding scheme. Recently, Borran et al. [BSAV03] presented a design method based on the Kullback-Leibler distance. Their scheme however does not guarantee that diversity, a major benefit of the MIMO channel, is exploited.

In this chapter, we derive a criterion that allows the assessment and design of constellations for perfect, partial or no CSI at the receiver, respectively with guaranteed diversity gain and quantified coding gain. Moreover, the criterion provides insight into the design of possible training signals if CSI has to be obtained at the receiver. For the frequency-flat fading MIMO channel, we also present two constructions of constellations adapted to the level of CSI. Whereas the first approach is based on numerical optimization using gradient search, the second approach combines signals originally designed for perfect CSI and for no CSI in an appropriate tradeoff adapted to the level of CSI at the receiver. The second construction also allows for receivers of reduced complexity by exploiting structural properties of the signal combination. Both design approaches are compared using numerical simulation results.

5.1 System Model

As in Chapter 2, we assume a communication system transmitting T symbols via M_T antennas to a receiver equipped with M_R antennas. We restrict ourselves to the frequency-flat channel model

$$\mathbf{Y} = \mathbf{SH} + \mathbf{W} \tag{5.1}$$

where $\mathbf{Y}$ is the $T \times M_R$ matrix of received symbols, $\mathbf{S}$ is the $T \times M_T$ matrix of transmitted symbols, the element $\{\mathbf{H}\}_{tr}$ of the $M_T \times M_R$ matrix $\mathbf{H}$ describes the complex channel coefficient between transmitter antenna t and receiver antenna r and the $T \times M_R$ matrix $\mathbf{W}$ contains the additive receiver noise. The signal matrix $\mathbf{S}$ is chosen from a constellation $\{\mathbf{S}_0, \ldots, \mathbf{S}_{Z-1}\}$ with power constraint

$$\frac{1}{Z} \sum_{n=0}^{Z-1} \mathrm{tr}\left(\mathbf{S}_n^H \mathbf{S}_n\right) = T,$$

the elements of $\mathbf{W}$ and $\mathbf{H}$ are modeled as circular-symmetric complex Gaussian variables with $\mathbf{w} \triangleq \mathrm{vec}(\mathbf{W}) \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ and

$\mathbf{h} \triangleq \text{vec}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{C_h})$ with full rank $\mathbf{C_h}$. The channel $\mathbf{H}$ stays constant during the transmission of the signal matrix $\mathbf{S}$.

If the channel does not change too rapidly, some knowledge of the channel coefficients in $\mathbf{H}$ is possibly available at the receiver in an estimate $\hat{\mathbf{H}}$. This knowledge can be obtained by a channel estimation algorithm which tracks the evolution of the channel coefficients or by the transmission of a pilot sequence. The ML detection rule at the receiver is then

$$\hat{\mathbf{S}} = \arg\max_{\mathbf{S}} p(\mathbf{Y}|\mathbf{S}, \hat{\mathbf{H}}),$$

i.e., the receiver is aware that the available CSI is an estimate with possible imperfections.

We now discuss three typical models for the available channel knowledge and define the corresponding likelihood functions $p(\mathbf{Y}|\mathbf{S}, \hat{\mathbf{H}})$.

1. The channel estimate $\hat{\mathbf{H}}$ and the error $\mathbf{E} = \mathbf{H} - \hat{\mathbf{H}}$ are independent in the relation

   $$\mathbf{H} = \hat{\mathbf{H}}_1 + \mathbf{E}_1 \tag{5.2}$$

   with $\mathbf{e}_1 \triangleq \text{vec}(\mathbf{E}_1) \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\mathbf{e}_1})$. We denote the probability density function of $\mathbf{Y}$ conditioned on the data matrix $\mathbf{S}$ and the channel estimate $\hat{\mathbf{H}}_1$ by

   $$p_1(\mathbf{Y}|\mathbf{S}, \hat{\mathbf{H}}_1; \mathbf{C}_{\mathbf{e}_1})$$

   taking into account that it is also a function of $\mathbf{C}_{\mathbf{e}_1}$.

2. In contrast, the second channel estimate in

   $$\hat{\mathbf{H}}_2 = \mathbf{H} + \mathbf{E}_2 \tag{5.3}$$

   is defined under the assumption of independence of the channel $\mathbf{H}$ and the error $\mathbf{E}$ with $\mathbf{e}_2 \triangleq \text{vec}(\mathbf{E}_2) \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\mathbf{e}_2})$ and the corresponding probability density function is denoted

   $$p_2(\mathbf{Y}|\mathbf{S}, \hat{\mathbf{H}}_2; \mathbf{C}_{\mathbf{e}_2}).$$

3. The receiver has no knowledge about the realization of the channel coefficients $\mathbf{H}$. However, a pilot signal $\mathbf{S}_P$ that is known to both transmitter and receiver is transmitted through the same channel as the data as in

   $$\mathbf{Y}_P = \mathbf{S}_P\mathbf{H} + \mathbf{W}_P. \tag{5.4}$$

ML detection can then be performed as

$$\begin{aligned}\hat{\mathbf{S}} &= \arg\max_{\mathbf{S}} p(\mathbf{Y}, \mathbf{Y}_P|\mathbf{S}; \mathbf{S}_P) \\ &= \arg\max_{\mathbf{S}} p(\mathbf{Y}|\mathbf{Y}_P, \mathbf{S}; \mathbf{S}_P) p(\mathbf{Y}_P|\mathbf{S}; \mathbf{S}_P) \\ &= \arg\max_{\mathbf{S}} p(\mathbf{Y}|\mathbf{Y}_P, \mathbf{S}; \mathbf{S}_P)\end{aligned}$$

because $p(\mathbf{Y}_P|\mathbf{S}; \mathbf{S}_P) = p(\mathbf{Y}_P; \mathbf{S}_P)$ since $\mathbf{Y}_P$ is independent of $\mathbf{S}$ and the density $p(\mathbf{Y}_P; \mathbf{S}_P)$ is constant when maximizing over $\mathbf{S}$. We denote the required probability density function as

$$p_3(\mathbf{Y}|\mathbf{Y}_P, \mathbf{S}; \mathbf{S}_P).$$

The following lemma is essential for casting all three cases into a common framework.

Lemma 4 *Assuming full column rank of* $\mathbf{S}_P$ *and independent columns of* $\mathbf{E}_1$, $\mathbf{E}_2$ *and* $\mathbf{H}$ *with identical column covariance matrix,i.e.,*

$$\mathbf{C}_{\mathbf{e}_1} = \mathbf{I}_{M_R} \otimes \mathbf{C}_{\mathbf{E}_1} \tag{5.5}$$
$$\mathbf{C}_{\mathbf{e}_2} = \mathbf{I}_{M_R} \otimes \mathbf{C}_{\mathbf{E}_2} \tag{5.6}$$
$$\mathbf{C}_{\mathbf{h}} = \mathbf{I}_{M_R} \otimes \mathbf{C}_{\mathbf{H}} \tag{5.7}$$

the following equalities

$$p_3(\mathbf{Y}|\mathbf{S}, \mathbf{Y}_P; \mathbf{S}_P) = p_1(\mathbf{Y}|\mathbf{S}, \hat{\mathbf{H}}_1; \mathbf{C}_{\mathbf{e}_1}) \tag{5.8}$$
$$= p_2(\mathbf{Y}|\mathbf{S}, \hat{\mathbf{H}}_2; \mathbf{C}_{\mathbf{e}_2}) \tag{5.9}$$

hold if we replace

$$\hat{\mathbf{H}}_1 = (\mathbf{S}_P^H \mathbf{S}_P + \sigma^2 \mathbf{C}_{\mathbf{H}}^{-1})^{-1} \mathbf{S}_P^H \mathbf{Y}_P \tag{5.10}$$
$$\mathbf{C}_{\mathbf{E}_1} = (\sigma^{-2} \mathbf{S}_P^H \mathbf{S}_P + \mathbf{C}_{\mathbf{H}}^{-1})^{-1} \tag{5.11}$$

in (5.8) *using* (5.5) *and*

$$\hat{\mathbf{H}}_2 = (\mathbf{S}_P^H \mathbf{S}_P)^{-1} \mathbf{S}_P^H \mathbf{Y}_P \tag{5.12}$$
$$\mathbf{C}_{\mathbf{E}_2} = \sigma^2 (\mathbf{S}_P^H \mathbf{S}_P)^{-1} \tag{5.13}$$

in (5.9) *using* (5.6).

The proof is presented in appendix 5.A.

Our main conclusion from Lemma 4 is that the likelihood functions and therefore the analysis of data communication system with available channel state information of the forms 1–3 can be cast into a common framework under the conditions that the replacements (5.10) to (5.13) exist. Then for any channel estimate of the models 1 or 2, we can compute an equivalent training block (as in model 3) leading to the same likelihood function and thus the same detection performance. Performance analysis can therefore focus on one of the proposed models. We choose in the following model 3 in order to apply the results of Section 2.6 to this case.

We note here that the estimate in (5.10) and (5.11) is the MMSE estimate whereas the the estimate in (5.12) and (5.13) is commonly referred to as the ML estimate of the channel coefficients based on the pilot block $\mathbf{S}_P$. In [BT04], it was mentioned that data detection based on joint processing of the pilot and signaling matrix (as in model 3) is equivalent to first computing an ML or MMSE estimate based on the pilot matrix alone and then using a receiver based on model 1 or 2, respectively. However, a restriction to orthogonal pilot matrices was made which appears to be unnecessary as proven in Lemma 4.

In view of the result, the restrictions in (5.5) to (5.7) become apparent: If the columns of the channel estimation errors $\mathbf{E}_1$ and $\mathbf{E}_2$ and the channel $\mathbf{H}$ are independent (implying a block diagonal structure of $\mathbf{C}_{\mathbf{e}_1}$,$\mathbf{C}_{\mathbf{h}_2}$ and $\mathbf{C}_{\mathbf{h}}$ but not necessarily with the same blocks on the main diagonal as in (5.5) to (5.7)), then columns of $\mathbf{Y}$ conditioned on the data and the available CSI are independent. For each column of $\mathbf{Y}$, we can then associate a specific pilot matrix $\mathbf{S}_P$ with the given error covariance matrix. Since we only send one training matrix, the error covariance matrices of each column of $\mathbf{Y}$ must be the same, imposing the Kronecker structures in (5.5) to (5.7).

We also want to point out that it was mentioned in [ATV02] that there is no loss in mutual information with respect to the transmitted data if the receiver has only access to an MMSE channel estimate compared to the case that the received symbols in the training block are available. In that sense, Lemma 4 make this statement explicit in terms of likelihood functions and extends it also to ML channel estimates.

5.2 Pairwise Error Probability Analysis

A common criterion to assess the performance of different signal designs is the PEP

$$\Pr(\mathbf{S}_i \to \mathbf{S}_j) \triangleq \Pr\left(p(\mathbf{Y}|\mathbf{S}_i) < p(\mathbf{Y}|\mathbf{S}_j) \quad |\mathbf{S}_i\right), \tag{5.14}$$

i.e., the probability that a binary ML test between two signaling matrices $\mathbf{S}_j$ and $\mathbf{S}_i$ decides in favor of $\mathbf{S}_j$ when $\mathbf{S}_i$ was transmitted. A priori, the PEP depends on the applied likelihood function in (5.14). However, using the results of the previous section, the likelihood functions p_1, p_2 and p_3 are equivalent under the assumptions of Lemma 4 and therefore, the resulting PEPs are the same. We can therefore take into account available CSI at the receiver in the form of a channel estimate in (5.2) or (5.3) by considering the transmission of an equivalent training block $\mathbf{S}_P$ and analyze the resulting PEP as a function of the likelihood function p_3 which depends on the extended matrices

$$\underline{\mathbf{S}}_i = \begin{bmatrix} \mathbf{S}_P \\ \mathbf{S}_i \end{bmatrix}, \quad \underline{\mathbf{S}}_j = \begin{bmatrix} \mathbf{S}_P \\ \mathbf{S}_j \end{bmatrix} \tag{5.15}$$

and assumes no further CSI at the receiver. We can thereby reuse the results in Section 2.6 indicating that

$$\Pr(\underline{\mathbf{S}}_i \to \underline{\mathbf{S}}_j) + \Pr(\underline{\mathbf{S}}_j \to \underline{\mathbf{S}}_i) \approx \sigma^{2M_R M_T} \frac{Q(\underline{\mathbf{S}}_i, \underline{\mathbf{S}}_j)}{|\mathbf{C}_\mathbf{h}||[\underline{\mathbf{S}}_i \quad \underline{\mathbf{S}}_j]^H[\underline{\mathbf{S}}_i \quad \underline{\mathbf{S}}_j]|^{M_R}} \tag{5.16}$$

in the high SNR regime where $Q(\underline{\mathbf{S}}_i, \underline{\mathbf{S}}_j)$ is given in Table 2.1. It was noted already in Section 2.6.4, that under the assumption

$$|\underline{\mathbf{S}}_j^H \underline{\mathbf{S}}_j| = |\underline{\mathbf{S}}_i^H \underline{\mathbf{S}}_i| \tag{5.17}$$

the function $Q(\underline{\mathbf{S}}_i, \underline{\mathbf{S}}_j)$ in (5.16) is a constant and the sum of the PEPs in (5.16) depends essentially only on $|\underline{\mathbf{S}}_j^H \mathbf{P}_{\underline{\mathbf{S}}_i}^{\perp} \underline{\mathbf{S}}_j| = |\underline{\mathbf{S}}_i^H \mathbf{P}_{\underline{\mathbf{S}}_j}^{\perp} \underline{\mathbf{S}}_i|$ which should be as large as possible for a low PEP. The additional assumption in (5.17) is met in the special case of scaled unitary signaling

$$\mathbf{S}_i^H \mathbf{S}_i = \mathbf{S}_j^H \mathbf{S}_j = \frac{T}{M_T} \mathbf{I} \tag{5.18}$$

which is of particular interest since for high SNR or $T \gg M_T$, unitary signals are optimal in terms of capacity [MH99] in the absence of receiver

CSI and have certain attractive properties with perfect CSI (as orthogonal space-time block codes [TJC99]). We will in the remainder of this chapter restrict ourselves to signal matrices satisfying (5.18) such that the common design rule relates to both the results of the asymptotic analysis as well as the Chernoff bound. We therefore assume in the following

$$\mathbf{S}_k = \sqrt{\frac{T}{M_T}}\mathbf{\Phi}_k, \quad \mathbf{\Phi}_k^H\mathbf{\Phi}_k = \mathbf{I}\,\forall k \tag{5.19}$$

In appendix 5.B, we show that $|\underline{\mathbf{S}}_j^H\mathbf{P}_{\underline{\mathbf{S}}_i}^{\perp}\underline{\mathbf{S}}_j|$ can also be written as $|T/M_T\mathbf{D}(\mathbf{\Phi}_i,\mathbf{\Phi}_j,\mathbf{\Sigma})|$ where we define

$$\mathbf{D}(\mathbf{\Phi}_i,\mathbf{\Phi}_j,\mathbf{\Sigma}) \triangleq (\mathbf{\Phi}_i-\mathbf{\Phi}_j)^H(\mathbf{I}+\mathbf{\Phi}_j\mathbf{\Sigma}\mathbf{\Phi}_j^H)^{-1}(\mathbf{\Phi}_i-\mathbf{\Phi}_j), \tag{5.20}$$

after using (5.15) and replacing

$$\mathbf{\Sigma} \triangleq \frac{T}{M_T}(\mathbf{S}_P^H\mathbf{S}_P)^{-1} \tag{5.21}$$

The quantity $|\mathbf{D}(\mathbf{\Phi}_i,\mathbf{\Phi}_j,\mathbf{\Sigma})|$ can be understood as taking the role of a "distance" between the two signal matrices $\mathbf{\Phi}_i$ and $\mathbf{\Phi}_j$ depending on the available CSI: The larger $|\mathbf{D}(\mathbf{\Phi}_i,\mathbf{\Phi}_j,\mathbf{\Sigma})|$, the lower the PEP for high SNR. We formalize this statement in the following

Design Criterion: *The quantity*

$$|\mathbf{D}(\mathbf{\Phi}_i,\mathbf{\Phi}_j,\mathbf{\Sigma})| \tag{5.22}$$

should be as large as possible.

Note however that the quantity in (5.22) is not a distance in the sense of a metric because, e.g., a vanishing distance here does not imply equality of the signals. The condition $|\mathbf{D}(\mathbf{\Phi}_i,\mathbf{\Phi}_j,\mathbf{\Sigma})| > 0$ guarantees that full diversity can be exploited at the receiver and generalizes similar design rules for both perfect CSI and no CSI at the receiver. In the following, the notation $\mathbf{A} \to \mathbf{B}$ means that the spectral norm (see Section 4.4) of $\mathbf{A}-\mathbf{B}$ goes to zero. We then obtain in the two extreme cases the following:

- If we have perfect CSI at the receiver, i.e. $\mathbf{\Sigma} \to \mathbf{0}$, we obtain

  $$\mathbf{D} = \mathbf{D}_c(\mathbf{\Phi}_i,\mathbf{\Phi}_j) \triangleq (\mathbf{\Phi}_i-\mathbf{\Phi}_j)^H(\mathbf{\Phi}_i-\mathbf{\Phi}_j) \tag{5.23}$$

  which is a standard criterion (see, e.g., [TSC98]). We will refer to this case also as coherent reception. Note that the requirement $|\mathbf{D}(\mathbf{\Phi}_i,\mathbf{\Phi}_j,\mathbf{\Sigma})| > 0$ always implies that $\mathbf{\Phi}_i-\mathbf{\Phi}_j$ must be full column rank. This is impossible for $T < M_T$ and we therefore need to assume $T \geq M_T$.

- For noncoherent reception, i.e., no CSI at the receiver modeled by $\mathbf{\Sigma}^{-1} \to \mathbf{0}$, we obtain

$$\begin{aligned}(\mathbf{I} + \mathbf{\Phi}_j \mathbf{\Sigma} \mathbf{\Phi}_j^H)^{-1} &= \mathbf{I} - \mathbf{\Phi}_j (\mathbf{\Phi}_j^H \mathbf{\Phi}_j + \mathbf{\Sigma}^{-1})^{-1} \mathbf{\Phi}_j^H \\ &\to \mathbf{P}_{\mathbf{\Phi}_j}^{\perp} = \mathbf{I} - \mathbf{\Phi}_j \mathbf{\Phi}_j^H\end{aligned}$$

  leading to

$$\mathbf{D} = \mathbf{D}_{\mathrm{nc}}(\mathbf{\Phi}_i, \mathbf{\Phi}_j) \triangleq \mathbf{\Phi}_i^H \mathbf{P}_{\mathbf{\Phi}_j}^{\perp} \mathbf{\Phi}_i = \mathbf{I} - \mathbf{\Phi}_i^H \mathbf{\Phi}_j \mathbf{\Phi}_j^H \mathbf{\Phi}_i. \tag{5.24}$$

  which is also widely used (see, e.g., [MBV02]) if no CSI is available.

$\mathbf{D}_{\mathrm{c}}$ and $\mathbf{D}_{\mathrm{nc}}$ are related to a coherent and noncoherent "distance" $|\mathbf{D}_{\mathrm{c}}|$ and $|\mathbf{D}_{\mathrm{nc}}|$, respectively. The implications of using $|\mathbf{D}_{\mathrm{c}}|$ or $|\mathbf{D}_{\mathrm{nc}}|$ as basis for a criterion to design $\mathbf{\Phi}_i$ and $\mathbf{\Phi}_j$ are significantly different: Whereas for coherent communication we obtain maximum coherent distance by choosing $\mathbf{\Phi}_i$ and $\mathbf{\Phi}_j$ antipodal, i.e. $\mathbf{\Phi}_i = -\mathbf{\Phi}_j$, such a choice of signals would lead to an ineffective communication system in the noncoherent setup because $|\mathbf{D}_{\mathrm{nc}}(\mathbf{\Phi}_i, -\mathbf{\Phi}_i)| = 0$. In contrast, signals with maximum noncoherent distance are orthogonal, i.e. $\mathbf{\Phi}_i^H \mathbf{\Phi}_j = \mathbf{0}$ leading to $|\mathbf{D}_{\mathrm{nc}}| = 1$ and can also be used in a coherent setting, however with suboptimal performance. Since

$$\begin{aligned}\mathbf{D}_{\mathrm{c}} &= (\mathbf{\Phi}_i - \mathbf{\Phi}_j)^H (\mathbf{P}_{\mathbf{\Phi}_j} + \mathbf{P}_{\mathbf{\Phi}_j}^{\perp})(\mathbf{\Phi}_i - \mathbf{\Phi}_j) \\ &= \mathbf{\Phi}_i^H \mathbf{P}_{\mathbf{\Phi}_j}^{\perp} \mathbf{\Phi}_i + (\mathbf{\Phi}_i - \mathbf{\Phi}_j)^H \mathbf{P}_{\mathbf{\Phi}_j} (\mathbf{\Phi}_i - \mathbf{\Phi}_j) \\ &= \mathbf{D}_{\mathrm{nc}} + (\mathbf{\Phi}_i - \mathbf{\Phi}_j)^H \mathbf{P}_{\mathbf{\Phi}_j} (\mathbf{\Phi}_i - \mathbf{\Phi}_j),\end{aligned}$$

the condition $|\mathbf{D}_{\mathrm{nc}}| > 0$ implies $|\mathbf{D}_{\mathrm{c}}| > 0$ showing that if full diversity can be exploited in the absence of CSI, it can also be exploited with perfect CSI. Moreover, the coherent distance is at least as large as the noncoherent distance. Both results are of course quite intuitive because available CSI should result in improved detection performance, i.e., lower probability of error and therefore in particular allow at least the same level of diversity.

In summary, the constellation design and performance are critically dependent on the available CSI at the receiver. We investigate therefore the behavior and implications of our derived design criterion as a function of the error covariance of the channel estimate in the next two sections. In particular, we address the following questions:

- Assuming we use a predetermined constellation with signals $\mathbf{\Phi}_0, \ldots, \mathbf{\Phi}_{Z-1}$ and the channel coherence time T allows for additional symbols such that we can transmit a training sequence. How should the training matrix be designed? This question is particularly relevant if signals designed for perfect CSI at the receiver are used, but CSI is not available yet which can lead to ineffective communication as pointed out above. A standard answer to this question about training design is to use *orthogonal* training, i.e., a training block $\mathbf{S}_P$ with orthogonal columns which leads to a covariance matrix $\mathbf{\Sigma} = \delta^2\mathbf{I}$. This choice minimizes the channel estimation error covariance and has therefore intuitive appeal. When is this choice really optimal (or: When is it not optimal?)? How much power should be dedicated to data transmission and how much to power to training? We will address this question in Section 5.3.

- Assume that a channel estimate at the receiver with equivalent covariance matrix $\mathbf{\Sigma}$ is given, how should the signaling matrices $\mathbf{\Phi}_k$ be designed? This question is of apparent importance due to the remarkable difference of design rules for the extreme cases $\mathbf{\Sigma} \to \mathbf{0}$ and $\mathbf{\Sigma}^{-1} \to \mathbf{0}$. A discussion concerning this problem is presented in Section 5.4.

In both sections, an analytical investigation requires the use of a criterion function to assess the quality of the design. The criterion (5.22) is a pairwise relation and has to be extended to more than two signals. In general, we are interested in minimizing the word error probability

$$\begin{aligned} J_{\text{err}} &= E_{\mathbf{\Phi}}\left[\Pr\left(\hat{\mathbf{\Phi}} \neq \mathbf{\Phi}| \quad \mathbf{\Phi} \text{ transmitted}\right)\right] \\ &= \frac{1}{N}\sum_{n=0}^{N-1} \Pr\left(\hat{\mathbf{\Phi}} \neq \mathbf{\Phi}_n | \mathbf{\Phi}_n \text{ transmitted}\right) \end{aligned} \tag{5.25}$$

assuming the signals $\mathbf{\Phi} \in \{\mathbf{\Phi}_0, \ldots, \mathbf{\Phi}_{Z-1}\}$ are transmitted with equal probability. A comparison between different constellations can be based on the resulting error probability at the same SNR or on comparing the required SNR to achieve a given error probability. If we assume that all signal pairs allow the exploitation of full diversity and therefore the error probability satisfies

$$\Pr(\hat{\mathbf{\Phi}} \neq \mathbf{\Phi}) \approx G\sigma^{2M_T M_R}$$

for large SNR with some constant G, we can just as well define the objective function

$$J_{\mathrm{err,dB}} \triangleq \frac{10}{M_T M_R} \log_{10} \sigma^{-2M_T M_R} J_{\mathrm{err}} \approx \frac{G[\mathrm{dB}]}{M_T M_R}.$$

A difference in the criterion $J_{\mathrm{err,dB}}$ between two different constellation designs then expresses the equivalent loss or gain in SNR.

The mentioned criterion function is usually intractable for analytical analysis and can be evaluated only approximately using Monte Carlo simulations. Instead, two other criteria which are easier to handle are frequency used in practice. In Chapters 3 and 4, we have used the objective function

$$J_{\mathrm{UB}}(\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_{Z-1}, \boldsymbol{\Sigma}) = \sum_{i=0}^{Z-1} \sum_{j \neq i} \frac{1}{|\mathbf{D}(\boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j, \boldsymbol{\Sigma})|^{M_R}}$$

as a scaled upper bound on the error probability. The advantage of this performance measure is the relatively simple dependence of the objective function on the parameters which usually allows for tractable differentiation and therefore lends itself to optimization based on gradient search.

A different criterion or principle is the maximization of the minimum distance. Motivated by classical constellation design for AWGN channels where the minimum distance determines the constellation's asymptotic error performance [Pro95], we can assess the design using the cost function

$$J_{\mathrm{WC}}(\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_{Z-1}, \boldsymbol{\Sigma}) = \frac{1}{\min_{i \neq j} |\mathbf{D}(\boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j, \boldsymbol{\Sigma})|^{M_R}}.$$

A criterion based on the "worst case" distance between two signal matrices has advantages if the minimum distance can be deduced easily due to some design rule and is therefore frequently used in structural design of the constellations.

5.3 Training Based on Predetermined Constellation Design

In this section, we assume that we extend the signals $\boldsymbol{\Phi} \in \{\boldsymbol{\Phi}_0, \ldots, \boldsymbol{\Phi}_{Z-1}\}$ with a training block $\mathbf{S}_P$ and investigate the design of this training block. Our analysis is divided into two parts: First, we investigate the design

of the training matrix assuming that the maximum power devoted to training is limited. In the second part, we assess the optimum level of power devoted to training.

5.3.1 Structure of the Training Matrix

The problem of designing training matrices has been investigated many times earlier, usually based on the minimization of channel estimation error variance [LS03] or under capacity optimization considerations [HH03]. Intuitively, optimizing the channel estimate such that the estimation error is minimal appears to be a reasonable choice because the purpose of the training block is to provide channel state information to the receiver. The more accurate this information is, the better the receiver should operate. However, we are not aware of any formalization of this intuition in the sense that data detection performance is optimized in the error probability sense. Since our design criterion is related to a tight approximation of the PEP in the high SNR case, we will use the criterion to investigate the optimality of orthogonal training. We can show in the two-signal case that orthogonal training leading to minimized channel estimation error is in general not optimal for every signal design. The difference to the optimal solution however is small and orthogonal training appears to be still an excellent choice for well-designed signal sets with more than two signals.

We will investigate the design of the training block $\mathbf{S}_P$ first under the two signal assumption: Assuming we have only two different signals $\mathbf{\Phi}_0$ and $\mathbf{\Phi}_1$, what training block (equiv. error covariance matrix) maximizes the distance between the two signals under a power constraint on the training block? Formally written, we want to solve the following problem:

$$\begin{array}{ll} \text{maximize} & |\mathbf{D}(\mathbf{\Phi}_0, \mathbf{\Phi}_1, \mathbf{\Sigma})| \quad \text{w.r.t. } \mathbf{\Sigma} \\ \text{subject to} & \operatorname{tr}\left(\mathbf{\Sigma}^{-1}\right) \leq P \end{array} \tag{5.26}$$

From (5.20) it is not difficult to show that

$$|\mathbf{D}(\mathbf{\Phi}_0, \mathbf{\Phi}_1, \mathbf{\Sigma})| = |\mathbf{D}_{\mathrm{c}}||\mathbf{I} - (\mathbf{I} - \mathbf{\Phi}_0^H \mathbf{\Phi}_1)^H \mathbf{D}_{\mathrm{c}}^{-1} (\mathbf{I} - \mathbf{\Phi}_0^H \mathbf{\Phi}_1)(\mathbf{I} + \mathbf{\Sigma}^{-1})^{-1}|$$

where we assume that $\mathbf{D}_{\mathrm{c}}^{-1}$ exists to obtain full diversity performance at least in the case of coherent signaling. Now since $\mathbf{D}_{\mathrm{c}} - \mathbf{D}_{\mathrm{nc}} = (\mathbf{I} - \mathbf{\Phi}_0^H \mathbf{\Phi}_1)(\mathbf{I} - \mathbf{\Phi}_0^H \mathbf{\Phi}_1)^H$, we define

$$(\mathbf{D}_{\mathrm{c}} - \mathbf{D}_{\mathrm{nc}})^{\frac{1}{2}} \triangleq \mathbf{I} - \mathbf{\Phi}_0^H \mathbf{\Phi}_1$$

and use the eigenvalue decomposition of the Hermitian matrix

$$(\mathbf{D}_{\mathrm{c}}-\mathbf{D}_{\mathrm{nc}})^{\frac{1}{2}H}\mathbf{D}_{\mathrm{c}}^{-1}(\mathbf{D}_{\mathrm{c}}-\mathbf{D}_{\mathrm{nc}})^{\frac{1}{2}}=\mathbf{Q}\mathbf{\Lambda}_Q\mathbf{Q}^H \tag{5.27}$$

where $\mathbf{\Lambda}_Q = \mathrm{diag}\{q_1,\ldots,q_{M_T}\}$ to reformulate the problem (5.26) in $\mathbf{X}=\mathbf{\Sigma}^{-1}$ to

$$\begin{array}{ll}\text{maximize} & |\mathbf{I}-\mathbf{\Lambda}_Q(\mathbf{I}+\mathbf{Q}^H\mathbf{X}\mathbf{Q})^{-1}| \ \text{ w.r.t } \ \mathbf{X}\\ \text{subject to} & \mathrm{tr}\,(\mathbf{X}\,)\le P.\end{array} \tag{5.28}$$

Now since for any positive semidefinite matrix $\mathbf{A}$ we have $|\mathbf{A}|\le\prod_k\{\mathbf{A}\}_{kk}$ with equality if and only if $\mathbf{A}$ is diagonal, we have that

$$|\mathbf{I}-\mathbf{\Lambda}_Q(\mathbf{I}+\mathbf{Q}^H\mathbf{X}\mathbf{Q})^{-1}|\le\prod_k(1-q_k\{(\mathbf{I}+\mathbf{Q}^H\mathbf{X}\mathbf{Q})^{-1}\}_{kk})$$

with equality if and only if $\mathbf{Q}^H\mathbf{X}\mathbf{Q}$ is diagonal. We therefore choose

$$\mathbf{X}=\mathbf{Q}\mathrm{diag}\,\{x_1,\ldots,x_{M_T}\}\,\mathbf{Q}^H \tag{5.29}$$

and reformulate the problem to

$$\begin{array}{ll}\text{maximize} & \prod_{k=1}^{M_T}\left(1-\frac{q_k}{1+x_k}\right) \ \text{ w.r.t. } x_1,\ldots,x_{M_T}\\ \text{such that} & \sum_k x_k\le P\end{array}$$

This is a waterfilling type of problem. In Appendix 5.C, we show that we need to set

$$x_k=[q_k\lambda'-1]^+$$

where $[\cdot]^+$ denotes the positive part (see Section 1.7) and we choose λ' such that $\sum_k x_k = P$ is satisfied. We can then choose any $\mathbf{S}_P$ that satisfies (5.29) and (5.21), e.g.,

$$\mathbf{S}_P=\sqrt{\frac{T}{M_T}}\mathbf{Q}\,\mathrm{diag}\{\sqrt{x_1},\ldots,\sqrt{x_{M_T}}\}\mathbf{Q}^H$$

This solution implies that more power of the training block is allocated to directions associated with larger eigenvalues in (5.27). We emphasize that the q_k are not equal in general and therefore some directions obtain more power than others. In other words, the training optimized for the two signals $\mathbf{\Phi}_0$ and $\mathbf{\Phi}_1$ is in general *not* orthogonal!

An intuitive explanation for this can be obtained using the point of view of noncoherent transmission of $\underline{\mathbf{S}}=[\mathbf{S}_P^T \quad \sqrt{T/M_T}\mathbf{\Phi}^T]^T$. For noncoherent transmission, signals with optimal distance should be orthogonal.

Therefore, if $\mathbf{\Phi}_0$ and $\mathbf{\Phi}_1$ are predetermined, $\mathbf{S}_P$ should be chosen such that $\underline{\mathbf{S}}_0$ and $\underline{\mathbf{S}}_1$ are "as orthogonal as possible," which is of course dependent on $\mathbf{\Phi}_0$ and $\mathbf{\Phi}_1$. In how far the optimized training differs from orthogonal training is characterized by the eigenvalues of the matrix (5.27).

In order to analyze the benefit of optimal training in the mentioned sense, we compare the distance $|\mathbf{D}_{\text{opt}}|$ obtained with optimal training with the distance $|\mathbf{D}_{\text{ortho}}|$ achieved with orthogonal training. Assuming P is large enough such that all $x_n > 0$, we obtain

$$\left(\frac{|\mathbf{D}_{\text{opt}}|}{|\mathbf{D}_{\text{ortho}}|}\right)^{\frac{1}{M_T}} = \frac{\frac{1}{M_T}\sum_{n=1}^{M_T}\left(\frac{P}{M_T}+1-q_n\right)}{\prod_{n=1}^{M_T}\left(\frac{P}{M_T}+1-q_n\right)^{\frac{1}{M_T}}} \geq 1$$

with equality if and only if $q_1 = \cdots = q_{M_T}$ (which is the case when orthogonal training is optimal) because of the inequality of geometric and arithmetic mean (see, e.g., [HJ99, p. 535]). As an upperbound for this ratio, we consider the worst case with maximally spread eigenvalues. For an even number of antennas, let $M_T/2$ eigenvalues of $\mathbf{\Lambda}_Q$ approach unity and the other half approach zero, then we obtain

$$\left(\frac{|\mathbf{D}_{\text{opt}}|}{|\mathbf{D}_{\text{ortho}}|}\right)^{\frac{1}{M_T}} \leq \frac{\frac{P}{M_T}+\frac{1}{2}}{\sqrt{\frac{P}{M_T}\left(\frac{P}{M_T}+1\right)}} = \sqrt{1+\frac{1}{4\frac{P}{M_T}\left(\frac{P}{M_T}+1\right)}}$$

which approaches 1 with growing P. Thus, if large training power is used, the difference to optimal training goes to zero (which is also an intuitive result because for high training power P, the channel is estimated with vanishing error covariance for any full rank training matrix).

The generalization of this result to constellations with more than two signals suffers from the dependency of the resulting optimum covariance matrix on the eigenvectors of (5.27). Any two distinct pairs can possibly have distinct eigenvectors of (5.28) and therefore a single training block can not satisfy the waterfilling solution for all signal pairs.

We have tried two approaches using J_{WC} and J_{UB} on a randomly chosen signal set in order to get an insight into the characteristics of the optimum solution for larger signal sets. Both an approach based on sequential waterfilling (i.e., by increasing the power allocated to the training in small steps up to the desired limit where in each step the training

power is allocated with respect to the currently worst-case signal pair) as well as on a gradient search on $\mathbf{\Sigma}^{-1}$ did not result in any solution that was clearly superior to orthogonal training in performance. Our experience that we have obtained from these approaches is that for any "normal" constellation, i.e., with (approximately) isotropically distributed signals, the eigenvectors of the matrix in (5.27) referring to its largest eigenvalue are also approximately isotropically distributed. Thus, a nonorthogonal training which effectively transmits more power in preferred directions will leave other directions with less power, leading overall to worse performance. Therefore, even though orthogonal training is not necessarily optimal in the two-signal case, it still appears to be a reasonable choice for constellations with a large number of signals.

We therefore assume in the following (similar to the standard approaches) that orthogonal training leading to

$$\mathbf{\Sigma} = \delta^2 \mathbf{I} \tag{5.30}$$

has been used. After some standard algebraic manipulations, we can then simplify (5.20) to

$$\mathbf{D} = \rho \mathbf{D}_{\mathrm{c}} + (1-\rho)\mathbf{D}_{\mathrm{nc}} = \mathbf{D}_{\mathrm{nc}} + \rho(\mathbf{D}_{\mathrm{c}} - \mathbf{D}_{\mathrm{nc}}) \tag{5.31}$$

where $\rho = 1/(\delta^2+1)$ describes the quality of the CSI: $\rho = 0$ due to $\delta^2 \to \infty$ implies no CSI at the receiver, $\rho = 1$ due to $\delta^2 \to 0$ means perfect CSI at the receiver and $0 < \rho < 1$ implies imperfect or "partial" CSI. The parameter ρ thereby illustrates the transition of noncoherent distance to coherent distance. For this case of orthogonal training and orthogonal signal matrices, we obtain the important simplification that the matrix relevant to distance with partial CSI is just a linear interpolation of the matrices related to the coherent and noncoherent distances.

5.3.2 Optimized Training Power

A very relevant question in practice is the required quality of the available channel estimates. As mentioned in Section 5.2, a design assuming perfect channel estimation at the receiver can fail completely if no such channel estimate is available. An important question is then the amount of power that should be dedicated to training: Too much training and the channel is estimated well, but the data detection suffers, not enough training and the data symbols can possibly not be distinguished well at the receiver due to the inaccurate channel estimate. Hassibi and

Hochwald [HH03] investigated the required length of the training interval as well as the fraction of power that should be devoted to the training sequence in terms of maximizing capacity, i.e., implicitly assuming the use of an optimized coding scheme for the given scenario operating over infinitely many blocks with statistically independent channel realizations. In contrast, we consider here the case that the signals to be used are given a priori and span only a single channel coherence interval. The coherence time T is long enough that we have the opportunity to extend the data signal with a training sequence, but we have to keep the overall power constant. We can then base the data detection on the joint processing of the training as well as the data sequence as in Section 5.1. We restrict ourselves to unitary training matrices based on the result of the previous subsection and vary the power allocated to training and data in the variable $0 \leq x \leq 1$.

Let $\{\tilde{\mathbf{\Phi}}_0, \ldots, \tilde{\mathbf{\Phi}}_{Z-1}\}$ be the constellation with

$$\tilde{\mathbf{\Phi}}_i = \begin{bmatrix} \sqrt{x}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sqrt{1-x}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{\Phi}_i \end{bmatrix} \tag{5.32}$$

and $\mathbf{\Phi}_i^H \mathbf{\Phi}_i = \mathbf{I}$ as before. Here $0 \leq x \leq 1$ is the ratio of power dedicated to the training sequence which is chosen to be a scaled identity matrix consistent with our assumption of orthogonal training. The power used in transmitting $\tilde{\mathbf{\Phi}}_i$ is the same as transmitting the symbol block $\mathbf{\Phi}_i$, however we have a variable tradeoff of power for training vs. power for data in the variable x. Inserting the definition (5.32) into (5.20) for no prior training $\mathbf{\Sigma}^{-1} \to 0$, we obtain

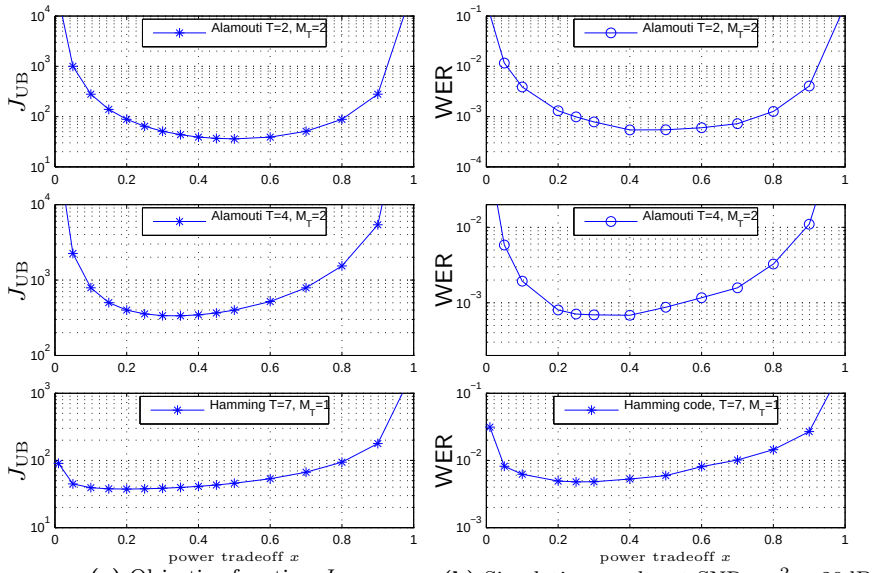
$$\mathbf{D}(\tilde{\mathbf{\Phi}}_i, \tilde{\mathbf{\Phi}}_j) = x(1-x)\mathbf{D}_{\mathrm{c}}(\mathbf{\Phi}_i, \mathbf{\Phi}_j) + (1-x)^2 \mathbf{D}_{\mathrm{nc}}(\mathbf{\Phi}_i, \mathbf{\Phi}_j) \tag{5.33}$$

which is zero for $x = 1$ (all power is used for training) and equal to the noncoherent distance matrix $\mathbf{D}_{\mathrm{nc}}$ for $x = 0$.

We can now optimize x with respect to the mentioned optimization criteria (worst case distance or union bound). These optimizations are problems in one single variable. If this optimization cannot be carried out analytically, numerical optimization using line-search methods is straightforward.

As an example for a single antenna system, we consider the $(7, 4)$ Hamming code in scaled BPSK symbols. For $M_T = 2$ transmitter antennas, the Alamouti scheme in BPSK symbols over one block in $T = 2$ or with two concatenated blocks over $T = 4$ is used. The resulting objective function values J_{UB} for variable data/training tradeoff are presented in

Figure 5.1(a) together with an experimental verification in Figure 5.1(b). In all three example cases, the objective function represents very well the



(a) Objective function J_{UB} **(b)** Simulation results at SNR $\sigma^{-2} = 20\mathrm{dB}$

Figure 5.1: Determining the optimal training power using either objective function or simulation

behavior of the resulting word error probability when using a variable amount of training power. Note however that the objective function is independent of the SNR and therefore the vertical axes in do not represent the same interval for objective function and simulation results.

It appears that 20% of the available power should be dedicated to training for the Hamming code whereas the Alamouti scheme requires half the power on training when detection is performed over a single block and one third of the power when two blocks are detected jointly. It is not surprising that the detection of a single Alamouti block requires exactly half the power since $\mathbf{D}_{\mathrm{nc}} = \mathbf{0}$ for all signal pairs and therefore only the coherent distance is relevant. Its coefficient $x(1-x)$ in (5.33) is maximized for $x = 1/2$.

5.4 Constellation Design Based on Partial CSI

5.4.1 Motivation

We now address the design of signals assuming a certain level of CSI at the receiver. The following simple example motivates the relevance of the amount of channel knowledge in the design of the constellation.

Assume binary communication, i.e., we use a constellation of $Z = 2$ signal matrices $\mathbf{\Phi}_0$ and $\mathbf{\Phi}_1$ and we want to design these matrices assuming channel estimation error variance δ^2 or equivalently $0 \leq \rho \leq 1$ in a single-antenna system. The objective function

$$\begin{aligned}|\mathbf{D}| &= \rho(\mathbf{\Phi}_1 - \mathbf{\Phi}_0)^H(\mathbf{\Phi}_1 - \mathbf{\Phi}_0) + (1-\rho)(\mathbf{I} - \mathbf{\Phi}_1^H\mathbf{\Phi}_0\mathbf{\Phi}_0^H\mathbf{\Phi}_1) \\ &= \rho(1 - \mathbf{\Phi}_0^H\mathbf{\Phi}_1 - \mathbf{\Phi}_1^H\mathbf{\Phi}_0 + \mathbf{\Phi}_1^H\mathbf{\Phi}_0\mathbf{\Phi}_0^H\mathbf{\Phi}_1) + 1 - \mathbf{\Phi}_1^H\mathbf{\Phi}_0\mathbf{\Phi}_0^H\mathbf{\Phi}_1 \\ &= \rho\|1-\theta\|^2 + 1 - \|\theta\|^2\end{aligned}$$

depends only on the scalar $\theta \triangleq \mathbf{\Phi}_0^H\mathbf{\Phi}_1$. Maximizing with respect to θ leads to

$$\theta_{\text{opt}} = \begin{cases} -\frac{\rho}{1-\rho} & \text{if} \quad \rho < 1/2 \\ -1 & \text{if} \quad \rho \geq 1/2 \end{cases}$$

which can be achieved in the case of $T = 2$ by choosing

$$\mathbf{\Phi}_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{\Phi}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} e^{j\cos^{-1}(\theta_{\text{opt}})} \\ e^{-j\cos^{-1}(\theta_{\text{opt}})} \end{pmatrix}.$$

For perfect channel knowledge we have $\theta_{\text{opt}} = -1$ (the signals are antipodal) and for no channel knowledge we have $\theta_{\text{opt}} = 0$, i.e., the signals are orthogonal. In between these extreme cases, the signals assume intermediate values, illustrating our previous conclusion that the constellation design should be matched to the quality level of receiver CSI and choosing a design based on the assumption of either perfect or no CSI at the receiver is in general suboptimal. We can evaluate the benefit of signal design matched to partial CSI at the receiver for different values of ρ. The result is presented in Figure 5.2 in terms of the objective function $J_{\text{UB}} = J_{\text{WC}} = |\mathbf{D}|^{-1}$ when using matched signal design compared to antipodal (coherent) and orthogonal (noncoherent) signals. It is apparent that antipodal signals designed for perfect channel knowledge are ineffective in the absence of channel knowledge because the receiver cannot

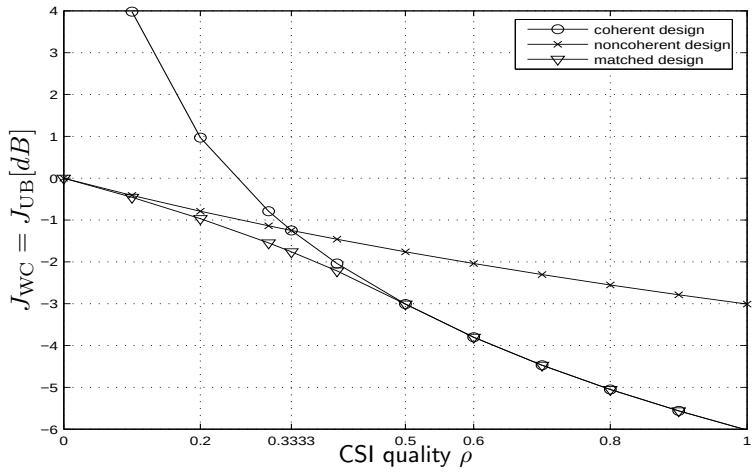


Figure 5.2: Evaluation of signaling matched to the level of receiver CSI compared to coherent and noncoherent signaling for various ρ

distinguish between the different symbols. We also notice the loss of 3dB for noncoherent signaling against coherent signaling when perfect channel information is available, i.e., $\rho = 1$. Both coherent and noncoherent design show equivalent performance for $\rho = 1/3$ where they are inferior to the perfectly matched constellation with a loss of about 0.5dB in SNR. These significant performance losses when using mismatched signaling design motivate the use of signals matched to partial channel knowledge at the receiver.

Moreover, we find it interesting to note that the resulting optimal design for any $\rho \geq 1/2$ is the same as for the coherent case which leads to the conclusion that even in the absence of perfect channel knowledge at the receiver, signals designed for this case can still be appropriate. This is somewhat implicitly assumed in investigations on the impact of CSI errors for a fixed signal set when this error is small (see, e.g., [LS03]), taking for granted that the signal set does not have to be modified.

5.4.2 Design for Partial CSI

We consider two approaches to the design of signals adapted to the level of CSI at the receiver. First, we use a gradient search method to design $\mathbf{\Phi}_0, \ldots, \mathbf{\Phi}_{Z-1}$ by optimizing the objective function J_{UB}. The second approach is based on a combination of constellations previously designed for coherent and noncoherent communication where the more weight is given to the coherent or noncoherent constellations dependent on the available CSI at the receiver.

Design Based on Gradient Search

Similar as in previous chapters, we can use a gradient search method to optimize constellations. Due to our constraint on unitary matrices in (5.18), we have to assure that the resulting solution follows this constraint. Agrawal et al. proposed a method of factorizing unitary matrices in [ARU01] and used this factorization to design constellations for noncoherent communication using a cost function based on the minimum distance (or "worst case") criterion. We make use of this factorization now in the wider context of partially coherent communication. For simplicity, we employ the cost function based on the union bound (similar to [MBV02]).

Structured Design

The intuition behind our structured design method is based on the interpretation of $|\mathbf{D}(\mathbf{S}_i, \mathbf{S}_j, \rho)|$ as a distance and the result that $\mathbf{D}$ is a weighted sum of the corresponding matrices for the coherent and noncoherent distance. Similarly, our proposed construction uses elements of typical designs for the noncoherent and coherent case which are chosen according to the weighting factor between both distances.

Let us first review the geometrical interpretations of both the coherent and noncoherent distance. Following [GvL96, Sec. 12.4.3], we can use the singular value decomposition (SVD) as $\mathbf{\Phi}_i^H \mathbf{\Phi}_j = \mathbf{U} \mathrm{diag}\{r_1, \ldots, r_M\} \mathbf{V}^H$ to compute the principal angles $\theta_1, \ldots, \theta_M$ between the subspaces spanned by $\mathbf{\Phi}_i$ and $\mathbf{\Phi}_j$ using

$$\cos(\theta_k) = r_k.$$

Then, the noncoherent distance in (5.24) can be written as

$$|\mathbf{D}_{\mathrm{nc}}| = \prod_{k=1}^{M}(1 - r_k^2) = \prod_{k=1}^{M} \sin^2(\theta_k),$$

which therefore only depends on the principal angles between the subspaces spanned by $\mathbf{\Phi}_i$ and $\mathbf{\Phi}_j$. In other words,

$$|\mathbf{D}_{\mathrm{nc}}(\mathbf{\Phi}_i, \mathbf{\Phi}_j)| = |\mathbf{D}_{\mathrm{nc}}(\mathbf{\Phi}_i\mathbf{Z}_i, \mathbf{\Phi}_j\mathbf{Z}_j)|$$

for any two unitary $M_T \times M_T$ matrices $\mathbf{Z}_i$ and $\mathbf{Z}_j$. Moreover, two signals in the same subspace have noncoherent distance zero.

In contrast, the coherent distance depends essentially on the singular values of the difference $\mathbf{\Phi}_i - \mathbf{\Phi}_j$. In particular, signals within the same subspace can have nonvanishing coherent distance. In fact, signals with maximum coherent distance are antipodal, i.e., $\mathbf{\Phi}_j = -\mathbf{\Phi}_i$ and thereby lie in the same subspace.

Since $\mathbf{D}_{\mathrm{c}} - \mathbf{D}_{\mathrm{nc}}$ is a positive semidefinite matrix we have

$$|\mathbf{D}| = |\rho(\mathbf{D}_{\mathrm{c}} - \mathbf{D}_{\mathrm{nc}}) + \mathbf{D}_{\mathrm{nc}}| \geq |\mathbf{D}_{\mathrm{nc}}|.$$

Thus, the noncoherent component $|\mathbf{D}_{\mathrm{nc}}|$ always contributes to the distance between two matrices. In other words, a constellation designed for noncoherent communication (and thus advantageous properties of $\mathbf{D}_{\mathrm{nc}}$) is still appropriate if partial or perfect CSI is available. However, the design freedom of having several signals occupy the same subspace cannot be exploited by a pure noncoherent design. For $\rho > 0$, two signals in the same subspace lead to $|\mathbf{D}| = |\rho\mathbf{D}_{\mathrm{c}}|$ which can be a sufficiently large distance for good overall performance if $|\mathbf{D}_{\mathrm{c}}|$ is large. The goal of our design is therefore to produce signal matrices which either occupy distinct subspaces and thereby have good noncoherent distance or have good coherent distance when they lie in the same subspace. The numbers of distinct subspaces and distinct signals within each subspace depend on the quality of the channel estimate: The lower ρ, the more distinct subspaces are needed and less signals occupy each subspace. In the extreme case of $\rho = 0$, only one signal per subspace is admissible.

Inspired by this intuition, we propose here to split up the task of designing the matrices $\{\mathbf{\Phi}_0, \mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_{Z-1}\}$ into two subtasks:

1. the design of distinct subspaces that the signal matrices occupy
2. the "position" of the signal matrices within each subspace.

As before, the design can be evaluated using either the worst-case distance or a union bound based criterion.

The two subtasks are the typical design goals of noncoherent and coherent communication and we therefore make use of available results in literature. We then need to investigate the mentioned tradeoff of either designing a constellation with a high number of distinct subspaces but small number of signals per subspace or choosing signals with a smaller number of subspaces but more signals within each subspace.

To evaluate this tradeoff, we propose the following construction: Let the constellations $\mathcal{X}_{\mathrm{c}} = \{\mathbf{\Phi}_0^{\mathrm{c}}, \ldots, \mathbf{\Phi}_{Z_{\mathrm{c}}-1}^{\mathrm{c}}\}$ and $\mathcal{X}_{\mathrm{nc}} = \{\mathbf{\Phi}_0^{\mathrm{nc}}, \ldots, \mathbf{\Phi}_{Z_{\mathrm{nc}}-1}^{\mathrm{nc}}\}$ be given where $\mathcal{X}_{\mathrm{c}}$ was designed for coherent communication and has elements of dimension $M_T \times M_T$ and $\mathcal{X}_{\mathrm{nc}}$ was designed for noncoherent communication with elements of dimension $T \times M_T$. Then we can construct the new constellation $\mathcal{X}$ with $Z = Z_{\mathrm{c}} Z_{\mathrm{nc}}$ such that for $k = k_1 Z_{\mathrm{c}} + k_2$, $k_1 \in \{0, \ldots, Z_{\mathrm{nc}} - 1\}$, $k_2 \in \{0, \ldots, Z_{\mathrm{c}}\}$ we have

$$\mathbf{\Phi}_k = \mathbf{\Phi}_{k_1}^{\mathrm{nc}} \mathbf{\Phi}_{k_2}^{\mathrm{c}},$$

i.e., we allow every combination of one noncoherent and one coherent signal. The particular choice of the noncoherent signal matrix thereby determines the subspace of the combined signal and the coherent signal determines its "position" within this subspace. It is thereby apparent that the information contained in $\mathbf{\Phi}^{\mathrm{nc}}$ can be recovered without taking into account $\mathbf{\Phi}^{\mathrm{c}}$ in a noncoherent detection. We will investigate the detection of these matrices later in Section 5.5. Note that for $Z_{\mathrm{c}} = 1$ or $Z_{\mathrm{nc}} = 1$, we obtain a construction purely optimized for noncoherent or coherent communication, respectively. The choice of both the coherent and noncoherent constellation depends of course on ρ.

Numerical Evaluation

We will illustrate the performance of this design with several examples. All resulting constellations here have $T = 8$ and $Z = 256$, i.e., a rate of 1 bit per channel use. We compare the performance of a purely noncoherent constellation ($Z_{\mathrm{nc}} = 256$, $Z_{\mathrm{c}} = 1$) with a signal set on combining a $Z_{\mathrm{nc}} = 64$ noncoherent and a $Z_{\mathrm{c}} = 4$ coherent constellation (termed as 6 noncoherent bits, 2 coherent bits), and a constellation combining $Z_{\mathrm{nc}} = 16$ with $Z_{\mathrm{c}} = 16$, i.e. 4 noncoherent and 4 coherent bits. The noncoherent constellations for $M_T = 1, 2$ are taken from [HMR+00], and the coherent signals are Z_{c}-PSK constellations for $M_T = 1$ and a scaled Alamouti code in $\sqrt{Z_{\mathrm{c}}}$-PSK for $M_T = 2$. The channel elements are i.i.d. fading

and the channel estimation error is white such that we can describe the available amount of CSI with the real scalar ρ. A performance characterization in terms of simulation results is provided in Figure 5.3 for $M_T = 1$ transmitter antenna and in Figure 5.4 for $M_T = 2$ antennas. We restrict the simulation results here to the high SNR region with the example point $\sigma^{-2} = 20\text{dB}$ and investigate the performance as function of the CSI quality ρ. We also compare the structured constellations with unstructured design obtained using gradient search. In the $M_T = 1$ case, we obtain

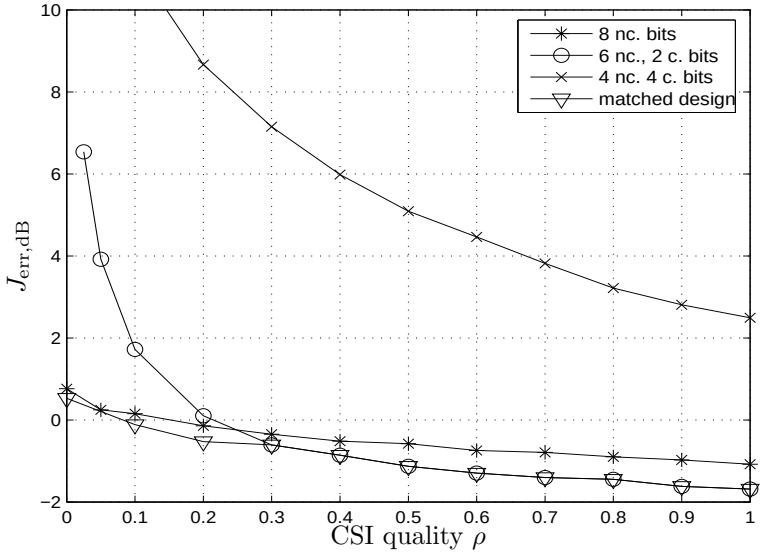


Figure 5.3: Simulation results at $\sigma^{-2} = 20\text{dB}$ with $M_T = 1$ transmitter antenna for various ρ

that the constellation with $Z_\text{c} = 16$, i.e., a large number of signals per subspace, has by far the worst performance among all compared signal sets for any ρ. Both this constellation and the one with 2 coherent bits (i.e., $Z_\text{c} = 4$) are useless for $\rho = 0$. However, performance improves drastically for the $Z_\text{c} = 4$ constellation with growing ρ. At around $\rho \approx 0.25$ the signal set with $Z_\text{c} = 4$ performs just as good as the constellation designed for noncoherent communication and outperforms it by about 0.5dB for $\rho \geq 0.5$. We have tried to improve on this constellation using

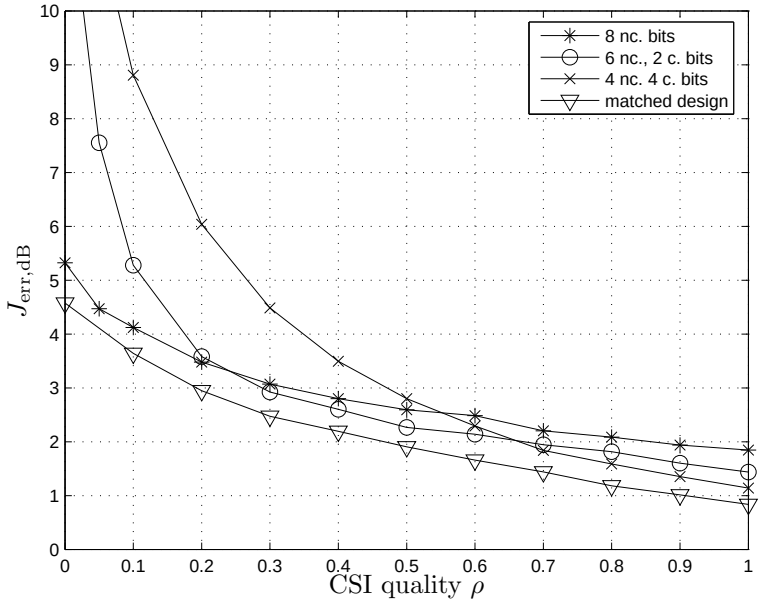


Figure 5.4: Simulation results at $1/\sigma^2 = 20$dB with $M_T = 2$ transmitter antennas for various ρ

our gradient search method taking the given set as starting point but were unable to obtain better results than the given signal set for $\rho > 0.3$. Moreover, for $\rho = 0$, the performance improvement over the noncoherent constellation is visible, but not very large. Thus, the unstructured design for $M_T = 1$ did not lead to significant improvements.

In the case with $M_T = 2$, the numerical optimization using gradient search appeared to be effective in the sense that clear improvements over all structured constellations are obtained for any ρ. As expected, the constellation with $N_{\mathrm{c}} > 1$ is useless for ρ close to zero. It turns out that among the compared constellations, the noncoherent design with $Z_{\mathrm{nc}} = 256$ performs best for $0 \leq \rho < 0.25$, the constellation with $Z_{\mathrm{c}} = 4$ perform best for $0.25 \leq \rho < 0.66$ and the signal set with $N_{\mathrm{c}} = 16$ performs best for $\rho \geq 0.66$. In other words, all of the structured constellations can be associated with values of ρ where they perform best, illustrating the

benefit of designing the signals according to the available CSI at the receiver.

5.5 Detection

We will now focus on the detection or decoding of the symbol matrix $\mathbf{S}$ based on the received symbol matrix $\mathbf{Y}$ and the available channel state estimate. For simplicity, we focus here on CSI available in the form 2 (see Section 5.1), keeping in mind that the equivalence with the other forms of CSI is guaranteed under the assumptions of Lemma 4.

Using the results of appendix 5.A, we know that $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$ conditioned on $\mathbf{S}$ and $\hat{\mathbf{H}}_2$ is Gaussian with mean $\boldsymbol{\mu}_2$ and covariance $\mathbf{C}_2$ (see (5.37) and (5.38))), i.e.,

$$p(\mathbf{y}|\mathbf{S}, \hat{\mathbf{H}}_2, \mathbf{C}_{\mathbf{E}_2}) = \frac{1}{|\pi \mathbf{C}_2|} \exp\left((\mathbf{y}-\boldsymbol{\mu}_2)^H \mathbf{C}_2^{-1}(\mathbf{y}-\boldsymbol{\mu}_2)\right)$$

Now since $\mathbf{S}^H\mathbf{S} = T/M_T\mathbf{I}$ we have that

$$\begin{aligned} |\pi \mathbf{C}_2| &= |\pi(\mathbf{I} \otimes (\sigma^2\mathbf{I} + (\mathbf{C}_{\mathbf{H}}^{-1} + \mathbf{C}_{\mathbf{E}_2}^{-1})^{-1}\mathbf{S}^H\mathbf{S}))| \\ &= \left|\pi\left(\sigma^2\mathbf{I} + \frac{T}{M_T}(\mathbf{C}_{\mathbf{H}}^{-1} + \mathbf{C}_{\mathbf{E}_2}^{-1})^{-1}\right)\right|^{M_R} \end{aligned}$$

which does not depend on $\mathbf{S}$. Therefore, the ML detector reduces to

$$\begin{aligned} \hat{\mathbf{S}} &= \arg\max_{\mathbf{S}} p_2(\mathbf{y}|\mathbf{S}, \hat{\mathbf{H}}_2; \mathbf{C}_{\mathbf{E}_2}) \\ &= \arg\min_{\mathbf{S}} (\mathbf{y}-\boldsymbol{\mu}_2)^H \mathbf{C}_2^{-1}(\mathbf{y}-\boldsymbol{\mu}_2) \\ &= \arg\min_{\mathbf{S}} \mathrm{tr}\left(-\sigma^{-2}\mathbf{Y}^H\mathbf{S}\mathbf{M}\mathbf{S}^H\mathbf{Y} - \mathbf{Y}^H\mathbf{S}\mathbf{M}\mathbf{C}_{\mathbf{E}_2}^{-1}\hat{\mathbf{H}}_2 - \hat{\mathbf{H}}_2^H\mathbf{C}_{\mathbf{E}_2}^{-1}\mathbf{M}\mathbf{S}^H\mathbf{Y}\right) \end{aligned}$$

where

$$\mathbf{M} \triangleq \left(\frac{T}{M_T}\mathbf{I} + \sigma^2\mathbf{C}_{\mathbf{H}}^{-1} + \sigma^2\mathbf{C}_{\mathbf{E}_2}^{-1}\right)^{-1} \tag{5.34}$$

and the algebraic details are provided in appendix 5.D. Assuming orthogonal training or, respectively, a white channel estimation error $\boldsymbol{\Sigma} = \delta^2\mathbf{I}$ as before leading to $\mathbf{C}_{\mathbf{E}_2}^{-1} = \frac{1}{\sigma^2\delta^2}\frac{T}{M_T}\mathbf{I}$ and i.i.d. fading, i.e., $\mathbf{C}_{\mathbf{H}} = \mathbf{I}$, the matrix $\mathbf{M}$ is a scaled identity matrix and the ML detector can be

simplified to

$$\begin{aligned}\hat{\mathbf{S}} &= \arg\min_{\mathbf{S}} \operatorname{tr}\left(\frac{T}{\delta^2 M_T}(-\mathbf{Y}^H\mathbf{S}\hat{\mathbf{H}}_2 - \hat{\mathbf{H}}_2^H\mathbf{S}^H\mathbf{Y}) - \mathbf{Y}^H\mathbf{S}\mathbf{S}^H\mathbf{Y}\right) \\ &= \arg\min_{\mathbf{S}} \left(\rho\|\mathbf{Y} - \mathbf{S}\hat{\mathbf{H}}_2\|^2 + (1-\rho)\|\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Y}\|^2\right)\end{aligned}$$

where the same weighting factor ρ as in (5.31) is used and we have dropped constant factors such as T/M_T (note that $\mathbf{S}\mathbf{S}^H = T/M_T\mathbf{P}_{\mathbf{S}}$). The metric $\|\mathbf{Y} - \mathbf{S}\mathbf{H}\|$ is related to the ML receiver for perfectly coherent communication and we use it here with the estimate $\mathbf{H} = \hat{\mathbf{H}}_2$ as if it were perfect. The decision variable $\|\mathbf{P}_{\mathbf{S}}^{\perp}\mathbf{Y}\|^2$ relates to noncoherent detector. The decision variable computed by a partially coherent detector is therefore a linear weighting between the decision variables for a coherent and a noncoherent detector in analogy with the linear weighting between the matrices $\mathbf{D}_{\mathrm{c}}$ and $\mathbf{D}_{\mathrm{nc}}$ determining the asymptotic error performance for high SNR.

An advantage of this observation is then that we can use properties of the constituent signals described in Section 5.4 for a simplified decoding rule. Using

$$\mathbf{S} = \sqrt{\frac{T}{M_T}}\mathbf{\Phi}^{\mathrm{nc}}\mathbf{\Phi}^{\mathrm{c}}$$

we obtain after some simple algebraic manipulations

$$(\hat{\mathbf{\Phi}}^{\mathrm{nc}}, \hat{\mathbf{\Phi}}^{\mathrm{c}}) = \arg\min_{\mathbf{\Phi}^{\mathrm{nc}}, \mathbf{\Phi}^{\mathrm{c}}} \left\{ \|\mathbf{P}_{\mathbf{\Phi}^{\mathrm{nc}}}^{\perp}\mathbf{Y}\|^2 + \rho \left\| \mathbf{\Phi}^{\mathrm{nc}H}\mathbf{Y} - \sqrt{\frac{T}{M_T}}\mathbf{\Phi}^{\mathrm{c}}\hat{\mathbf{H}}_2 \right\|^2 . \right\} \tag{5.35}$$

The first part constitutes the decision variable for the noncoherent detector (which is always a compound of the decision variable). The second part can be understood as related a coherent detector based on the received matrix $\tilde{\mathbf{Y}} = \mathbf{\Phi}^{\mathrm{nc}H}\mathbf{Y}$ which uses the channel estimate as if it was perfect. Since the columns of $\mathbf{\Phi}^{\mathrm{nc}}$ form an orthogonal basis of the subspace spanned by $\mathbf{\Phi}^{\mathrm{nc}}$, $\tilde{\mathbf{Y}}$ can be understood as representation of $\mathbf{P}_{\mathbf{\Phi}^{\mathrm{nc}}}\mathbf{Y}$ in the basis vectors of the subspace defined by $\mathbf{\Phi}^{\mathrm{nc}}$.

Due to the partitioning of the decision variable into a noncoherent part together with a coherent part applying the channel estimate as if it was perfect, structural design approaches for coherent constellations allowing low detection complexity can be advantageous here as well. As an example, if $\mathbf{\Phi}^{\mathrm{c}}$ is taken from an orthogonal space-time block code

(OSTBC) [TJC99] [Ala98], the computation of the ML decision variable requires Z_{nc} evaluations of the noncoherent decision variable and for each this evaluations, a detection of the coherent signal. Since the complexity of this coherent detection is significantly lower than evaluating all Z_{c} coherent signals (it decouples into scalar detection of the constituent symbols), overall detection complexity is reduced.

We can also decrease the detection complexity using the following simplified decoder:

$$\hat{\mathbf{\Phi}}^{\mathrm{nc}} = \arg\min_{\mathbf{\Phi}^{\mathrm{nc}}} \|\mathbf{P}_{\mathbf{\Phi}^{\mathrm{nc}}}\mathbf{Y}\|^2$$

$$\hat{\mathbf{\Phi}}^{\mathrm{c}} = \arg\min_{\mathbf{\Phi}^{\mathrm{c}}} \left\| \hat{\mathbf{\Phi}}^{\mathrm{nc}}\mathbf{Y} - \sqrt{\frac{T}{M_T}}\mathbf{\Phi}^{\mathrm{c}}\hat{\mathbf{H}} \right\|^2$$

which no longer requires the knowledge of ρ. Note that we are basically performing a hard decision on $\mathbf{\Phi}^{\mathrm{nc}}$ before decoding $\mathbf{\Phi}^{\mathrm{c}}$, thereby possibly losing detection performance.

If we use unstructured noncoherent and coherent constellations with this simplified detector, we require Z_{nc} evaluations of the noncoherent decision variable and Z_{c} evaluations of the coherent variable, which is much lower than performing $Z_{\mathrm{nc}}Z_{\mathrm{c}}$ evaluations of (5.35). Similar as above, using an OSTBC as constituent coherent constellation results in even lower detection complexity.

In Figures 5.5 and 5.6 we compare the performance of the simplified decoding scheme with the optimal ML detection where we used the same constellation combinations as in Section 5.4.2. For $M_T = 1$, we see almost no difference in the performance between the optimal ML and the simplified detector for $\rho \leq 0.5$ and the loss in SNR for the simplified detection is not significant. Only for large ρ, the simplified detector for the constellation with $Z_{\mathrm{nc}} = 64$ and $Z_{\mathrm{c}} = 4$ performs about 0.4dB worse than the optimal detection.

However, the performance difference between the detectors are substantial for $M_T = 2$. The simplified detection of the constellation with 6 noncoherent bits and 2 coherent bits incurs a loss of almost 2dB in SNR at $\sigma^{-2} = 20$dB and detection performance does not improve for $\rho > 0.3$. Apparently, for $\rho > 0.3$, errors in the noncoherent detection dominate (the noncoherent distance between subspaces is significantly smaller than the coherent distance between signals in the same subspace). Therefore, detection can not be improved with better knowledge of the channel. The difference between coherent and noncoherent distances in the constellation with 4 coherent and 4 noncoherent bits however are much smaller.

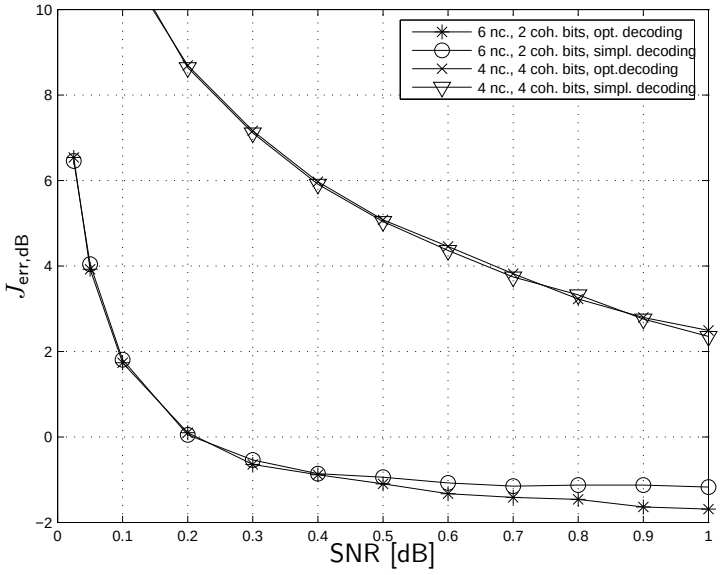


Figure 5.5: Comparison of optimal and simplified decoding for $M_T = 1$ transmitter antennas and $\sigma^{-2} = 20dB$ for various level of CSI quality

Therefore, the same effect of dominating error events in the noncoherent detection has much smaller impact.

It is therefore important to judge for the specific constellations used under which receiver CSI assumptions the simplified detection is applicable without significant performance loss.

5.6 Summary

In this chapter, we investigated the impact of channel state information at the receiver on constellation design. After showing the equivalence of three seemingly different types of receiver CSI, we used a PEP analysis to present a design criterion that unifies previous results for no and perfect channel knowledge at the receiver and allows a continuous transitions between these cases. Two main applications of the design criterion were presented. First, for systems with separate training and data trans-

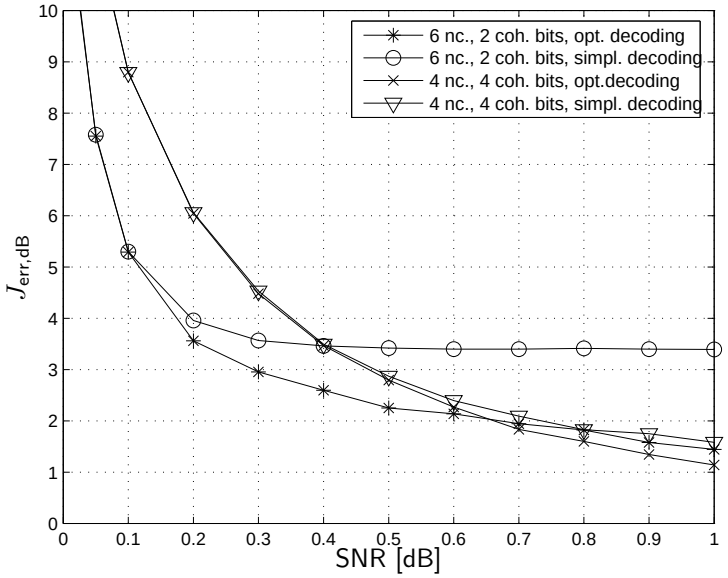


Figure 5.6: Comparison of optimal and simplified decoding for $M_T = 2$ transmitter antennas and $\sigma^{-2} = 20dB$ for various level of CSI quality

mission using a predetermined constellation, we investigated the design of optimal training blocks. It was found that orthogonal training is not optimal in general, but still is usually an excellent solution for reasonably designed signals. The amount of power dedicated to training was also optimized dependent on the design of the constellation. The second application assumed a given level of CSI at the receiver and investigated the implications on the signal design. An unstructured design approach based on numerical optimization together with a structured design combining constellations used for coherent and noncoherent communication were presented. It was found that the combined constellations perform similar to the unstructured constellation for $M_T = 1$ but incur a performance loss for $M_T = 2$. This loss has to be traded off against a possible reduction in receiver complexity due to structure of the consituent coherent signal set. A simplified decoding algorithm for such structured constellations was also presented.

We consider the work presented in this chapter as a starting point for further analysis of signaling schemes under the viewpoint of partially coherent distance. For example, differential detection of two subsequent space-time symbols can be viewed as one training block and one data block, thereby spending exactly half the power (if the transmission power is normalized) on training, half on data. This can be optimal in the sense of maximizing distance, but is not necessarily so if multiple symbol detection with coding between symbols is used. Moreover, the theory in this chapter can possibly be extended to frequency-selective channels which were the subject of Chapters 3 and 4.

Appendix 5.A Equivalence of Receivers

We will now analyze the three different receivers in turn. For notational convenience, we rewrite (5.1) by stacking the columns of $\mathbf{Y}$, $\mathbf{H}$ and $\mathbf{W}$ into their respective vector counterparts $\mathbf{y}$, $\mathbf{h}$ and $\mathbf{w}$ and obtain

$$\mathbf{y} = \tilde{\mathbf{S}}\mathbf{h} + \mathbf{w} \tag{5.36}$$

where $\tilde{\mathbf{S}} = \mathbf{I}_{M_R} \otimes \mathbf{S}$. The vectors $\hat{\mathbf{h}}_1$, $\hat{\mathbf{h}}_2$ and the matrix $\tilde{\mathbf{S}}_P = \mathbf{I}_{M_R} \otimes \mathbf{S}_P$ are similarly defined.

1. Inserting the vectorized form of (5.2) into (5.36), we obtain

   $$\mathbf{y} = \mathbf{S}\hat{\mathbf{h}}_1 + \mathbf{S}\mathbf{e}_1 + \mathbf{w}$$

   with independent Gaussian variables $\hat{\mathbf{h}}_1$, $\mathbf{e}_1$ and $\mathbf{w}$ and therefore $\mathbf{y}$ conditioned on $\mathbf{S}$ and $\hat{\mathbf{h}}_1$ is Gaussian with mean and covariance

   $$\begin{aligned}\boldsymbol{\mu}_1 &= \tilde{\mathbf{S}}\hat{\mathbf{h}}_1 \\ &= \mathrm{vec}(\mathbf{S}\hat{\mathbf{H}}_1) \\ \mathbf{C}_1 &= \sigma^2\mathbf{I} + \tilde{\mathbf{S}}\mathbf{C}_{\mathbf{e}_1}\tilde{\mathbf{S}}^H \\ &= \mathbf{I} \otimes (\sigma^2\mathbf{I} + \mathbf{S}\mathbf{C}_{\mathbf{E}_1}\mathbf{S}^H)\end{aligned}$$

2. $\mathbf{y}$ and $\hat{\mathbf{h}}_2$ in (5.36) and (5.3) are both linear combinations of the independent Gaussian variables $\mathbf{h}$, $\mathbf{e}_2$ and $\mathbf{w}$. They are therefore jointly Gaussian with

   $$\begin{bmatrix} \mathbf{y} \\ \hat{\mathbf{h}}_2 \end{bmatrix} \Bigg| \tilde{\mathbf{S}} \sim \mathcal{CN}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma^2\mathbf{I} + \tilde{\mathbf{S}}\mathbf{C}_{\mathbf{h}}\tilde{\mathbf{S}}^H & \tilde{\mathbf{S}}\mathbf{C}_{\mathbf{h}} \\ \mathbf{C}_{\mathbf{h}}\tilde{\mathbf{S}}^H & \mathbf{C}_{\mathbf{h}} + \mathbf{C}_{\mathbf{e}_2} \end{bmatrix} \right).$$

   Using Lemmas 5 and 6 we conclude after some algebraic manipulations that $\mathbf{y}$ conditioned on $\hat{\mathbf{h}}_2$ and $\tilde{\mathbf{S}}$ is Gaussian with mean and covariance

   $$\begin{aligned}\boldsymbol{\mu}_2 &= \tilde{\mathbf{S}}(\mathbf{C}_{\mathbf{h}}^{-1} + \mathbf{C}_{\mathbf{e}_2}^{-1})^{-1}\mathbf{C}_{\mathbf{e}_2}^{-1}\hat{\mathbf{h}}_2 && (5.37) \\ &= \mathrm{vec}(\mathbf{S}(\mathbf{C}_{\mathbf{H}}^{-1} + \mathbf{C}_{\mathbf{E}_2}^{-1})^{-1}\mathbf{C}_{\mathbf{E}_2}^{-1}\hat{\mathbf{H}}_2) \\ \mathbf{C}_2 &= \sigma^2\mathbf{I} + \tilde{\mathbf{S}}(\mathbf{C}_{\mathbf{h}}^{-1} + \mathbf{C}_{\mathbf{e}_2}^{-1})^{-1}\tilde{\mathbf{S}}^H && (5.38) \\ &= \mathbf{I} \otimes (\sigma^2\mathbf{I} + \mathbf{S}(\mathbf{C}_{\mathbf{H}}^{-1} + \mathbf{C}_{\mathbf{E}_2}^{-1})^{-1}\mathbf{S}^H)\end{aligned}$$

3. Apparently, $\mathbf{y}$ in (5.36) and $\mathbf{y}_P$ in (5.4) are jointly Gaussian with

$$\left[\begin{array}{c} \mathbf{y} \\ \mathbf{y}_P \end{array}\middle|\, \tilde{\mathbf{S}}\right] \sim \mathcal{CN}\left(\left[\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array}\right], \left[\begin{array}{cc} \sigma^2\mathbf{I} + \tilde{\mathbf{S}}\mathbf{C}_{\mathbf{h}}\tilde{\mathbf{S}}^H & \tilde{\mathbf{S}}\mathbf{C}_{\mathbf{h}}\tilde{\mathbf{S}}_P^H \\ \tilde{\mathbf{S}}_P\mathbf{C}_{\mathbf{h}}\tilde{\mathbf{S}}^H & \sigma^2\mathbf{I} + \tilde{\mathbf{S}}_P\mathbf{C}_{\mathbf{h}}\tilde{\mathbf{S}}_P^H \end{array}\right]\right).$$

Again using Lemma 5 and after some algebraic manipulations involving Lemma 6, we obtain that $\mathbf{y}$ conditioned on $\tilde{\mathbf{S}}$ and $\mathbf{y}_P$ is Gaussian with mean and and covariance

$$\begin{aligned} \boldsymbol{\mu}_3 &= \tilde{\mathbf{S}}(\tilde{\mathbf{S}}_P^H\tilde{\mathbf{S}}_P + \sigma^2\mathbf{C}_{\mathbf{h}}^{-1})^{-1}\tilde{\mathbf{S}}_P^H\mathbf{y}_P \qquad (5.39)\\ &= \mathrm{vec}(\mathbf{S}(\mathbf{S}_P^H\mathbf{S}_P + \sigma^2\mathbf{C}_{\mathbf{H}}^{-1})^{-1}\mathbf{S}_P^H\mathbf{Y}_P) \\ \mathbf{C}_3 &= \sigma^2\mathbf{I} + \tilde{\mathbf{S}}(\mathbf{C}_{\mathbf{h}}^{-1} + \sigma^{-2}\tilde{\mathbf{S}}_P^H\tilde{\mathbf{S}}_P)^{-1}\tilde{\mathbf{S}}^H \qquad (5.40)\\ &= \mathbf{I}\otimes(\sigma^2\mathbf{I} + \mathbf{S}(\mathbf{C}_{\mathbf{H}}^{-1} + \sigma^{-2}\mathbf{S}_P^H\mathbf{S}_P)^{-1}\mathbf{S}^H) \end{aligned}$$

Now using the replacements (5.10)–(5.13) we obtain that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$ and $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{C}_3$. Since Gaussian probability densities are completely characterized by their first two moments, the assertion is proven.

Appendix 5.B Some Determinant Calculations

Using Relation A.10, we expand

$$\begin{aligned} |\underline{\mathbf{S}}_j^H\mathbf{P}_{\underline{\mathbf{S}}_i}^{\perp}\underline{\mathbf{S}}_j| &= |\underline{\mathbf{S}}_j^H\underline{\mathbf{S}}_j - \underline{\mathbf{S}}_j^H\underline{\mathbf{S}}_i(\underline{\mathbf{S}}_i^H\underline{\mathbf{S}}_i)^{-1}\underline{\mathbf{S}}_i^H\underline{\mathbf{S}}_j| \\ &= \frac{\left|\begin{array}{cc} \underline{\mathbf{S}}_i^H\underline{\mathbf{S}}_i & \underline{\mathbf{S}}_i^H\underline{\mathbf{S}}_j \\ \underline{\mathbf{S}}_j^H\underline{\mathbf{S}}_i & \underline{\mathbf{S}}_j^H\underline{\mathbf{S}}_j \end{array}\right|}{|\underline{\mathbf{S}}_i^H\underline{\mathbf{S}}_i|} \\ &= \frac{\left|\begin{array}{cc} \mathbf{S}_P^H\mathbf{S}_P + \mathbf{S}_i^H\mathbf{S}_i & \mathbf{S}_P^H\mathbf{S}_P + \mathbf{S}_i^H\mathbf{S}_j \\ \mathbf{S}_P^H\mathbf{S}_P + \mathbf{S}_j^H\mathbf{S}_i & \mathbf{S}_P^H\mathbf{S}_P + \mathbf{S}_j^H\mathbf{S}_j \end{array}\right|}{|\mathbf{S}_P^H\mathbf{S}_P + \mathbf{S}_i^H\mathbf{S}_i|} \end{aligned}$$

where we used Relation (A.9) and the definitions in (5.15). Since adding scalar multiples of columns(rows) to other columns(rows) does not change the value of the determinant, we obtain after some manipulations in the

numerator

$$\begin{aligned}
&\left|\begin{array}{cc} \mathbf{S}_P^H\mathbf{S}_P+\mathbf{S}_i^H\mathbf{S}_i & \mathbf{S}_P^H\mathbf{S}_P+\mathbf{S}_i^H\mathbf{S}_j \\ \mathbf{S}_P^H\mathbf{S}_P+\mathbf{S}_j^H\mathbf{S}_i & \mathbf{S}_P^H\mathbf{S}_P+\mathbf{S}_j^H\mathbf{S}_j \end{array}\right| \\
&\quad=\left|\begin{array}{cc} \mathbf{S}_P^H\mathbf{S}_P+\mathbf{S}_i^H\mathbf{S}_i & \mathbf{S}_P^H\mathbf{S}_P+\mathbf{S}_i^H\mathbf{S}_j \\ (\mathbf{S}_j-\mathbf{S}_i)^H\mathbf{S}_i & (\mathbf{S}_j-\mathbf{S}_i)^H\mathbf{S}_j \end{array}\right| \\
&\quad=\left|\begin{array}{cc} \mathbf{S}_P^H\mathbf{S}_P+\mathbf{S}_i^H\mathbf{S}_i & \mathbf{S}_i^H(\mathbf{S}_j-\mathbf{S}_i) \\ (\mathbf{S}_j-\mathbf{S}_i)^H\mathbf{S}_i & (\mathbf{S}_j-\mathbf{S}_i)^H(\mathbf{S}_j-\mathbf{S}_i) \end{array}\right| \\
&\quad=|\mathbf{S}_P^H\mathbf{S}_P+\mathbf{S}_i^H\mathbf{S}_i| \\
&\qquad|(\mathbf{S}_j-\mathbf{S}_i)^H(\mathbf{I}-\mathbf{S}_i(\mathbf{S}_P^H\mathbf{S}_P+\mathbf{S}_i^H\mathbf{S}_i)^{-1}\mathbf{S}_i)(\mathbf{S}_j-\mathbf{S}_i)|
\end{aligned}$$

where we used the relation (A.9) in the last step. Applying Lemma 6 in the appendix, we conclude that

$$\begin{aligned}
|\underline{\mathbf{S}}_j\mathbf{P}_{\underline{\mathbf{S}}_i}^{\perp}\underline{\mathbf{S}}_j| &= |(\mathbf{S}_j-\mathbf{S}_i)^H(\mathbf{I}+\mathbf{S}_i(\mathbf{S}_P^H\mathbf{S}_P)^{-1}\mathbf{S}_i^H)^{-1}(\mathbf{S}_j-\mathbf{S}_i)| \\
&= |T/M_T(\mathbf{\Phi}_j-\mathbf{\Phi}_i)^H(\mathbf{I}+\mathbf{\Phi}_i\mathbf{\Sigma}\mathbf{\Phi}_i^H)^{-1}(\mathbf{\Phi}_j-\mathbf{\Phi}_i)|
\end{aligned}$$

where used the replacements (5.19) and (5.21).

Appendix 5.C Finding the Optimum Training Matrix

Based on the constraints $P-\sum_n x_n \geq 0$ and $x_n \geq 0$, we construct the Lagrangian

$$\begin{aligned}
&J(x_1,\ldots,x_{M_T}) \\
&\quad=\sum_{n=1}^{M_T}\ln-\left(1-\frac{q_n}{1+x_n}\right)-\lambda(P-\sum_n x_n)-\sum_n \mu_n x_n
\end{aligned}$$

whose derivative with respect to x_n is

$$\frac{\partial J}{\partial x_n}=-\frac{1}{\left(1-\frac{q_n}{1+x_n}\right)}\left(\frac{q_n}{1+x_n}\right)^2+\lambda-\mu_n$$

which, when set to zero after solving yields

$$x_n=q_n\left(\sqrt{\left\|\frac{1}{4}+\frac{1}{\lambda-\mu_n}\right\|}+\frac{1}{2}\right)-1. \tag{5.41}$$

For inactive constraints (i.e., $x_n > 0$) we have $\mu_n = 0$ (inactive constraints have zero multiplier [Fle91]) which implies that

$$x_n = q_n\lambda' - 1$$

where $\lambda' = \sqrt{\frac{1}{4} + \frac{1}{\lambda}} + \frac{1}{2}$ or λ, respectively, needs to be chosen such that $\sum_n x_n = P$ is satisfied. For active constraints, $\mu_n > 0$ is chosen to guarantee $x_n = 0$ in (5.41). As a summary, we obtain

$$x_n = [q_n\lambda' - 1]^+$$

and λ' is chosen such that $\sum_n x_n = P$ is satisfied.

Appendix 5.D Reformulation of the ML Receiver

We evaluate the term $(\mathbf{y}-\boldsymbol{\mu}_2)^H\mathbf{C}_2^{-1}(\mathbf{y}-\boldsymbol{\mu}_2)$ by splitting up this product into four terms that are considered in turn. Since

$$\begin{aligned}\mathbf{C}_2^{-1} &= \mathbf{I}\otimes(\sigma^2\mathbf{I} + \mathbf{S}(\mathbf{C}_{\mathbf{H}}^{-1} + \mathbf{C}_{\mathbf{E}_2}^{-1})^{-1}\mathbf{S}^H)^{-1}\\ &= \mathbf{I}\otimes(\sigma^{-2}\mathbf{I} - \sigma^{-2}\mathbf{S}(\mathbf{S}^H\mathbf{S} + \sigma^2\mathbf{C}_{\mathbf{H}}^{-1} + \sigma^2\mathbf{C}_{\mathbf{E}_2}^{-1})^{-1}\mathbf{S}^H)\\ &= \mathbf{I}\otimes\sigma^{-2}(\mathbf{I} - \mathbf{S}\mathbf{M}\mathbf{S}^H)\end{aligned}$$

with $\mathbf{M}$ defined in (5.34), we have that

$$\begin{aligned}\mathbf{y}^H\mathbf{C}_2^{-1}\mathbf{y} &= \mathrm{tr}(\sigma^{-2}\mathbf{Y}^H(\mathbf{I} - \mathbf{S}\mathbf{M}\mathbf{S}^H)\mathbf{Y})\\ \mathbf{y}^H\mathbf{C}_2^{-1}\boldsymbol{\mu}_2 + \boldsymbol{\mu}_2^H\mathbf{C}_2^{-1}\mathbf{y} &= 2\Re(\mathbf{y}^H\mathbf{C}_2^{-1}\boldsymbol{\mu}_2)\\ &= 2\Re(\mathrm{tr}(\sigma^{-2}\mathbf{Y}^H(\mathbf{I} - \mathbf{S}\mathbf{M}\mathbf{S}^H)\mathbf{S}\\ &\qquad (\mathbf{C}_{\mathbf{H}}^{-1} + \mathbf{C}_{\mathbf{E}_2}^{-1})^{-1}\mathbf{C}_{\mathbf{E}_2}^{-1}\hat{\mathbf{H}}_2))\\ &= 2\Re(\mathrm{tr}(\sigma^{-2}\mathbf{Y}^H\mathbf{S}\mathbf{M}(\mathbf{M}^{-1} - \mathbf{S}^H\mathbf{S})(\mathbf{C}_{\mathbf{H}}^{-1} + \mathbf{C}_{\mathbf{E}_2}^{-1})^{-1}\\ &\qquad \mathbf{C}_{\mathbf{E}_2}^{-1}\hat{\mathbf{H}}_2))\\ &= 2\Re(\mathrm{tr}(\mathbf{Y}^H\mathbf{S}\mathbf{M}\mathbf{C}_{\mathbf{E}_2}^{-1}\hat{\mathbf{H}}_2))\\ \boldsymbol{\mu}_2^H\mathbf{C}_2^{-1}\boldsymbol{\mu}_2 &= \mathrm{tr}(\sigma^{-2}\hat{\mathbf{H}}_2\mathbf{C}_{\mathbf{E}_2}^{-1}(\mathbf{C}_{\mathbf{H}}^{-1} + \mathbf{C}_{\mathbf{E}_2}^{-1})^{-1}\mathbf{S}^H(\mathbf{I} - \mathbf{S}\mathbf{M}\mathbf{S}^H)\mathbf{S}\\ &\qquad (\mathbf{C}_{\mathbf{H}}^{-1} + \mathbf{C}_{\mathbf{E}_2}^{-1})^{-1}\mathbf{C}_{\mathbf{E}_2}^{-1}\hat{\mathbf{H}}_2)\end{aligned}$$

Since $\mathbf{S}^H(\mathbf{I}-\mathbf{SMS}^H)\mathbf{S} = \frac{T}{M_T}(\mathbf{I}-\frac{T}{M_T}\mathbf{M})$ and $\operatorname{tr}(\mathbf{Y}^H\mathbf{Y})$ are constant in a variation of $\mathbf{S}$, we have that

$$\begin{aligned}&\mathbf{y}^H\mathbf{C}_2^{-1}\mathbf{y} - \boldsymbol{\mu}_2^H\mathbf{C}_2^{-1}\mathbf{y} - \mathbf{y}^H\mathbf{C}_2^{-1}\boldsymbol{\mu}_2 + \boldsymbol{\mu}_2\mathbf{C}_2^{-1}\boldsymbol{\mu}_2\\ &\quad = -2\Re(\operatorname{tr}(\mathbf{Y}^H\mathbf{SMC}_{\mathbf{E}_2}^{-1}\hat{\mathbf{H}}_2)) - \sigma^{-2}\operatorname{tr}(\mathbf{Y}^H\mathbf{SMS}^H\mathbf{Y}) + \text{const.}\end{aligned}$$

Chapter 6

Summary and Future Work

6.1 Summary

We have investigated signaling design in a communication system operating over linear time-discrete channels. Our particular focus was on the scenario that the receiver has no or imperfect knowledge of the channel coefficients, i.e., has no or imperfect CSI. Traditional signal design approaches for this scenario were based on separate acquisition of CSI and subsequent data transmission through channels known at the receiver. In contrast, we proposed to adjust the signal design directly to the level of CSI, which resulted in signaling schemes allowing significant performance gain when compared to traditional approaches.

We began our investigation with a description of common models for wireless communication systems where both the transmitter and the receiver are equipped with one or possibly several antennas. Both a single carrier as well as an OFDM signal model were presented for the communication over a frequency-selective block-fading channel. Moreover, two different receiver structures were described that were based on different assumptions on the statistical properties of the channel coefficients. Due to the linear nature of the communication channel and the signal modeling, it was possible to cast the analysis of both system models into a common framework. The two different receiver structures were analyzed in terms of pairwise error probability either by applying the Chernoff

bound, using an asymptotic expression valid in the high SNR region or in exact form. These analysis results laid the groundwork for several design methods applied in subsequent chapters.

In our first approach to signal design suitable for communication over unknown frequency-selective channels, we fixed the symbol alphabet to a given finite set. Assuming knowledge of the system configuration parameters, we applied the results of the preceding chapter to define the minimization of the union upper bound on detection error probability as criterion for our signal design. The corresponding optimization problem was then solved using simulated annealing and the resulting constellations evaluated using Monte Carlo simulation.

The second approach to signal design removed the restriction to a predetermined symbol set and was based on the criterion to minimize a high SNR approximation of the union bound on detection error. The objective function had the advantage of being much easier to evaluate and allowed optimization approaches based on gradient search. Again, the resulting constellations were evaluated numerically and showed superior performance compared to alternative schemes suitable for the unknown channel at the receiver.

The excellent performance of the two mentioned approaches came at the expense of high complexity requirements. The third design method investigated in this thesis allowed the communication through unknown frequency-selective channels with significantly lower demands in complexity, albeit at the price of a penalty in detection error performance. Constellations previously used for space–time differential transmission were applied for differential space–frequency transmission over a frequency-selective channel. A detailed analysis on the error performance of the differential detector resulted in a criterion that guarantees the exploitation of full diversity depending on signal and channel model parameters.

In the last part of this thesis, we extended the assumptions concerning the available channel knowledge at the receiver. Here, we assumed that the receiver has access to an estimate of the channel parameters together with knowledge about its statistical properties. This model included thereby the cases of unknown channel (if the estimate is useless) as before, but also the other extreme case of perfect channel knowledge (i.e., the estimate has zero error) and allowed a smooth transition between these scenarios. Three seemingly different types of channel knowledge were shown to be equivalent under certain circumstances and could be investigated under a common framework. We then motivated that the design of the constellation used for data transmission should take

into account the quality of the available channel estimate at the receiver. Therefore, two different signal constructions that can be matched to the quality of the channel estimate were proposed and compared in performance. Moreover, structural properties of one signal construction were exploited to propose a detection algorithm with lower complexity.

6.2 Future Work

The chapters 3 to 5 reflect the chronological order in which the design approaches were conceived, analyzed and investigated. The extension from no CSI to partial CSI in the receiver in Chapter 5 thereby was developed last and as a first step restricted to frequency-flat fading channels. The application of the theory developed in Chapter 5 to the frequency-selective case is an immediate proposition for future work.

Moreover, the signal design approaches proposed in this thesis were all based on the objective to minimize the detection error probability (or some objective function approximating or bounding it) assuming moderate to high SNR. An interesting topic for future work is therefore an investigation concerning signal design criteria in the low SNR regime. In addition, further research into transmission schemes requiring significantly lower detection complexity than exhaustive search appears promising.

Finally, the system models used in this thesis were idealized in the sense that no further imperfections such as, e.g., synchronization errors in the receiver were taken into account. It is therefore of practical interest to investigate the robustness of the proposed signaling schemes with respect to such effects. Moreover, as a further step in this direction, signal design explicitly taking these effects into account can be an interesting area for future research.

Appendix A

Some Useful Lemmas and Rules

The following standard results are used in various places throughout this thesis.

Lemma 5 *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be two jointly complex Gaussian random vectors with*

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \sim \mathcal{CN}\left(\begin{bmatrix} \bar{\mathbf{a}} \\ \bar{\mathbf{b}} \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{aa} & \mathbf{\Sigma}_{ab} \\ \mathbf{\Sigma}_{ba} & \mathbf{\Sigma}_{bb} \end{bmatrix} \right)$$

then the pdf of $\mathbf{a}$ *conditioned on* $\mathbf{b} = \tilde{\mathbf{b}}$ *is also Gaussian with mean*

$$\bar{\mathbf{a}} + \mathbf{\Sigma}_{ab}\mathbf{\Sigma}_{bb}^{-1}(\tilde{\mathbf{b}} - \bar{\mathbf{b}})$$

and covariance

$$\mathbf{\Sigma}_{aa} - \mathbf{\Sigma}_{ab}\mathbf{\Sigma}_{bb}^{-1}\mathbf{\Sigma}_{ba}$$

Proof: See [Kay93, Appendix 15B]. ■

Lemma 6 *For the four matrices* $\mathbf{A}$,$\mathbf{B}$,$\mathbf{C}$,$\mathbf{D}$ *it holds that*

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1} \qquad \text{(A.1)}$$

if all involved inverse matrices exist.

Proof: The proof can be established by direct verification, i.e., multiplication of the right-hand side of (A.1) with $(\mathbf{A} + \mathbf{BCD})$ with some

subsequent algebraic manipulations. An alternative is to solve the matrix equation

$$(\mathbf{A} + \mathbf{BCD})\mathbf{X} = \mathbf{I} \tag{A.2}$$

in the unknown matrix $\mathbf{X}$. It follows immediately from (A.2) that

$$\mathbf{X} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{BCDX} \tag{A.3}$$

as well as by multiplying (A.2) with $\mathbf{DA}^{-1}$ from the left that

$$(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})\mathbf{CDX} = \mathbf{DA}^{-1}. \tag{A.4}$$

Now isolating $\mathbf{CDX}$ in (A.4) and inserting the result into (A.3) yields the result for $\mathbf{X}$. ■

We also note other standard rules from linear algebra (see, e.g., [Lüt96]). All rules assume that the matrices are of compatible dimension and inverses exist.

$$\begin{aligned}
\mathrm{tr}\left(\mathbf{A}^H\mathbf{B}\right) &= (\mathrm{vec}(\mathbf{A}))^H\mathrm{vec}(\mathbf{B}) && \text{(A.5)}\\
\mathrm{vec}(\mathbf{ABC}) &= (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B}) && \text{(A.6)}\\
(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) &= (\mathbf{AC}) \otimes (\mathbf{BD}) && \text{(A.7)}\\
|k\mathbf{I}_M + \mathbf{AB}| &= k^{M-N}|k\mathbf{I}_N + \mathbf{BA}|, \quad k \neq 0 && \text{(A.8)}\\
\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} &= |\mathbf{A}||\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}| && \text{(A.9)}\\
&= |\mathbf{D}||\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}| && \text{(A.10)}
\end{aligned}$$

Bibliography

[Age00] Federal Network Agency. UMTS-Versteigerungsverfahren abgeschlossen. Press Release, August 2000. Available (October 2005) via `http://www.bundesnetzagentur.de` (in German).

[Ala98] S. M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16(8):1451–1458, October 1998.

[And99] J. B. Anderson. *Digital Transmission Engineering*. Prentice Hall, 1999.

[ARU01] D. Agrawal, T. J. Richardson, and R. Urbanke. Multiple-antenna signal constellations for fading channels. *IEEE Transactions on Information Theory*, 47(6):2618–2626, September 2001.

[ATNS98] D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri. Space–time coded OFDM for high data-rate wireless communication over wideband channels. In *Proc. IEEE Vehicular Technology Conference*, volume 3, pages 2232–2236, May 1998.

[ATV02] S. Adireddy, L. Tong, and H. Viswanathan. Optimal placement of training for frequency-selective block-fading channels. *IEEE Transactions on Information Theory*, 48(8):2338–2353, August 2002.

[BB02] H. Bölcskei and M. Borgmann. Code design for non-coherent MIMO-OFDM systems. In *Proc. Allerton Conference on Communication, Control and Computing*, Monticello, IL, USA, October 2002.

[BB04] Moritz Borgmann and Helmut Bölcskei. Noncoherent space-frequency coded MIMO-OFDM. *IEEE Journal on Selected Areas in Communications*, April 2004. Submitted April 1. Final version April 28, 2005. To appear.

[BG96] C. Berrou and A. Glavieux. Near optimium error correcting coding and decoding: Turbo-codes. *IEEE Transactions on Communications*, 44(10), 1996.

[BGT93] C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon limit error-correcting coding and decoding: Turbo-codes. In *Proc. IEEE International Conf. on Communications*, volume 2, pages 1064–1070, Geneva, May 1993.

[Bin90] J. A. C. Bingham. Multicarrier modulation for data transmission: An idea whose time has come. *IEEE Communications Magazine*, 28(5):5–14, 1990.

[Bos98] M. Bossert. *Kanalkodierung*. B.G. Teubner Stuttgart, 2nd edition, 1998. (in German).

[BP00] H. Bölcskei and A. J. Paulraj. Space-frequency coded broadband OFDM systems. In *IEEE Wireless Communications and Networking Conference*, volume 1, pages 1–6, 2000.

[BSAV03] M. Borran, A. Sabharwal, B. Aazhang, and P. Varshney. Partially coherent constellations for multiple-antenna systems. In *Proc. Asilomar Conference on Signals, Systems, and Computers*, Asilomar Grounds, Pacific Grove, CA, USA, November 2003.

[BT04] E. Biglieri and G. Taricco. Transmission and reception with multiple antennas: Theoretical foundations. *Foundations and Trends in Communications and Information Theory*, 1(2):183–332, 2004.

[BV01] M. Brehler and M. K. Varanasi. Asymptotic error probability analysis of quadratic receivers in Rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space–time receivers. *IEEE Transactions on Information Theory*, 47(6):2383–2399, September 2001.

[CAC01] K. Chugg, A. Anastasopoulos, and X. Chen. *Iterative Detection*. Kluwer Academic Publishers, 2001.

[CC00] O. Coskun and K. M. Chugg. Combined coding and training for unknown ISI channels. Submitted to *IEEE Transactions on Communications*, June 2000.

[CHIW98] D. J. Costello, Jr, J. Hagenauer, H. Imai, and S. B. Wicker. Applications of error-control coding. *IEEE Transactions on Information Theory*, 44(6):2531–2560, October 1998.

[CP96] K. M. Chugg and A. Polydoros. MLSE for an unknown channel–part I: Optimality considerations. *IEEE Transactions on Communications*, 44(7):836–846, July 1996.

[D2.05] WINNER D2.7. Assessment of advanced beamforming and MIMO technologies. Technical report, Wireless World Initiative New Radio, February 2005. Available online (October 2005) at `https://www.ist-winner.org/DeliverableDocuments/D2-7.pdf`.

[Dog03] A. Dogandžić. Chernoff bounds on pairwise error probability of space–time codes. *IEEE Transactions on Information Theory*, 49(5):1327–1336, May 2003.

[FG98] G. J. Foschini and M. J. Gans. On limits of wireles communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6(3):311–335, March 1998.

[FGL+84] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi. Efficient modulation for band-limited channels. *IEEE Journal on Selected Areas in Communications*, SAC-2(5):632–647, September 1984.

[Fle91] R. Fletcher. *Practical Methods of Optimization*. John Wiley and Sons, 2nd edition, 1991.

[For72] G. D. Forney, Jr. Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference. *IEEE Transactions on Information Theory*, 18(3):363–378, May 1972.

[GFBK99] J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo. Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels. *IEEE Transactions on Communications*, 47(4):527–537, April 1999.

[GHSW87] A. A. El Gamal, L. A. Hemachandra, I. Shperling, and V. K. Wei. Using simulated annealing to design good codes. *IEEE Transactions on Information Theory*, IT-33(1):116–123, January 1987.

[GL97] A. Gorokhov and P. Loubaton. Semi-blind second order identification of convolutive channels. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 3905–3908, Seattle,USA, 1997.

[Gra81] A. Graham. *Kronecker Products and Matrix Calculus With Applications*. Ellis Horwood Ltd., 1981.

[GS02a] J. Giese and M. Skoglund. Space–time code design for combined channel estimation and error protection. In *Proc. IEEE International Symposium on Information Theory*, Lausanne, Switzerland, June 2002.

[GS02b] J. Giese and M. Skoglund. Space–time code design for combined channel estimation and error protection. In *Proc. RadioVetenskap och Kommunikation (RVK)*, Stockholm, Sweden, June 2002.

[GS02c] J. Giese and M. Skoglund. Space–time code design for unknown frequency-selective channels. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, Florida, USA, May 2002.

[GS03a] J. Giese and M. Skoglund. Combined coding and modulation design for unknown frequency-selective channels. In *Proc. IEEE International Symposium on Information Theory*, Yokohama, Japan, 2003.

[GS03b] J. Giese and M. Skoglund. Single and multi-antenna constellations for communication over unknown frequency-selective fading channels. Submitted to *IEEE Transactions on Information Theory*, May 2003.

[GS03c] J. Giese and M. Skoglund. Space–time constellations for unknown frequency-selective channels. In *Proc. IEEE International Conf. on Communications*, pages 2583–2587, Anchorage, AK, USA, 2003.

[GS04] J. Giese and M. Skoglund. Performance of unitary differential space-frequency modulation. In *Proc. International Symposium on Information Theory and its Applications*, Parma, Italy, October 2004.

[GS05a] J. Giese and M. Skoglund. Space–time constellation design for partial CSI at the receiver. In *Proc. IEEE International Symposium on Information Theory*, Adelaide, Australia, September 2005.

[GS05b] J. Giese and M. Skoglund. Space–time constellation design for partial CSI based on code combination. In *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, October 2005.

[GS05c] J. Giese and M. Skoglund. Space–time constellation design for partial CSI at the receiver. October 2005. In Preparation.

[GvL96] G. H. Golub and C. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.

[Han03] A. Å. Hansson. *Generalized AAP Detection for Communication over Unknown Time-Dispersive Waveform Channels*. PhD thesis, Chalmers University of Technology, 2003.

[HH02] B. Hassibi and B. M. Hochwald. High-rate codes that are linear in space and time. *IEEE Transactions on Information Theory*, 48(7):1804–1824, July 2002.

[HH03] B. Hassibi and B. M. Hochwald. How much training is needed in multiple-antenna wireless links? *IEEE Transactions on Information Theory*, 49(4):951–963, April 2003.

[HJ99] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1999.

[HM00] B. M. Hochwald and T. L. Marzetta. Unitary space–time modulation for multiple-antenna communications in Rayleigh flat fading. *IEEE Transactions on Information Theory*, 46(2):543–564, March 2000.

[HMR+00] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke. Systematic design of unitary space–time constellations. *IEEE Transactions on Information Theory*, 46(6):1962–1973, September 2000.

[HS00] B. M. Hochwald and W. Sweldens. Differential unitary space–time modulation. *IEEE Transactions on Communications*, pages 2041–2052, December 2000.

[Hug00] B. L. Hughes. Differential space–time modulation. *IEEE Transactions on Information Theory*, 46(7):2567–2578, November 2000.

[Imh61] J. P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3-4):419–426, 1961.

[IR204] WINNER IR2.2. Identification of advanced beamforming and MIMO techniques. Technical report, Wireless World Initiative New Radio, June 2004.

[JS04] G. Jöngren and M. Skoglund. Quantized feedback information in orthogonal space–time block coding. *IEEE Transactions on Information Theory*, 50(10):2473–2482, October 2004.

[JSO02] G. Jöngren, M. Skoglund, and B. Ottersten. Utilizing partial channel information in the design of space–time block codes. In *Proc. 5th International Symposium on Wireless Personal Multimedia Communications*, volume 2, pages 681–685, October 2002.

[Kay93] S. M. Kay. *Fundamentals of Statistical Signal Processing*, volume 1, Estimation Theory. Prentice Hall, 1993.

[KGV83] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, May 1983.

[LFT01] Y. Liu, M. P. Fitz, and O. Y. Takeshita. Space–time codes perfomance criteria and design for frequency-selective fading channels. In *Proc. IEEE International Conf. on Communications*, volume 9, pages 2800–2804, Helsinki, Finland, June 2001.

[LLK03] C. Ling, K. H. Li, and A. C. Kot. Noncoherent sequence detection of differential space–time modulation. *IEEE Transactions on Information Theory*, 49(10):2727–2734, October 2003.

[Lot99] M. Lott. Comparision of frequency and time domain differential modulation in an OFDM system for wireless ATM. In *Proc. 49th IEEE Vehicular Techn. Conf.*, volume 2, pages 877–883, May 1999.

[LP00] E. Lindskog and A. Paulraj. A transmit diversity scheme for channels with intersymbol interference. In *Proc. IEEE International Conf. on Communications*, pages 307–311, 2000.

[LS01] L. H.-J. Lampe and R. Schober. Differential modulation diversity for OFDM. In *Proceedings 6th International OFDM-Workshop*, pages 19–1–19–6, Hamburg, Germany, September 2001.

[LS03] E. G. Larsson and P. Stoica. *Space–Time Block Coding for Wireless Communications*. Cambridge University Press, 2003.

[Lüt96] H. Lütkepohl. *Handbook of Matrices*. John Wiley and Sons, 1996.

[LW00] B. Lu and X. Wang. Space–time code design in OFDM systems. In *Proc. IEEE Global Telecommunications Conference*, volume 2, pages 1000–1004, 2000.

[Mar03] T. L. Marzetta. Differential unitary space-frequency modulation. In *Proc. Allerton Conference on Communication, Control and Computing*, October 2003.

[MBV02] M. L. McCloud, M. Brehler, and M. Varanasi. Signal design and convolutional coding for noncoherent space–time communication on the block-Rayleigh-fading channel. *IEEE Transactions on Information Theory*, 48(5):1186–1194, May 2002.

[MGO02] X. Ma, G. B. Giannakis, and S. Ohno. Optimal training for block transmissions over doubly-selective fading channels. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1509–1512, May 2002.

[MH99] T. L. Marzetta and B. M. Hochwald. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading. *IEEE Transactions on Information Theory*, 45(1):139–157, January 1999.

[MN88] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 1988.

[MP92] A. M. Mathai and S. B. Provost. *Quadratic Forms in Random Variables: Theory and Applications*. Marcel Dekker, Inc., New York, USA, 1992.

[MTL03] Q. Ma, C. Tepedelenlioglu, and Z. Liu. Full diversity block diagonal codes for differential space–time-frequency coded OFDM. In *Proc. IEEE Global Telecommunications Conference*, pages 868–872, San Francisco, CA, USA, December 2003.

[PP97] A. J. Paulraj and B. Papadias. Space–time processing for wireless communications. *IEEE Signal Processing Magazine*, 14(6), November 1997.

[Pro95] J. G. Proakis. *Digital Communications*. McGraw-Hill International Editions, 3rd edition, 1995.

[San96] M. Sandell. *Design and Analysis of Estimators for Multicarrier Modulation and Ultrasonic Imaging*. PhD thesis, Luleå University of Technology, September 1996.

[SFG02] S. Siwamogsatham, M. P. Fitz, and J. H. Grimm. A new view of performance analysis of transmit diversity schemes in correlated Rayleigh fading. *IEEE Transactions on Information Theory*, 48(4):950–956, April 2002.

[SGP02] M. Skoglund, J. Giese, and S. Parkvall. Code design for combined channel estimation and error protection. *IEEE Transactions on Information Theory*, 48(5):1162–1171, May 2002.

[Sha48] C. Shannon. A mathematical theory of communications. *Bell System Technical Journal*, 27, October 1948. Reprint `shannon1948.pdf` available online (October 2005) on `http://cm.bell-labs.com/cm/ms/what/shannonday`.

[SL02] R. Schober and L. H.-J. Lampe. Noncoherent receivers for differential space–time modulation. *IEEE Transactions on Communications*, 50(5):768–777, May 2002.

[SP61a] D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty – I. *Bell System Technical Journal*, 1961.

[SP61b] D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty – II. *Bell System Technical Journal*, 1961.

[SP62] D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty – III: The dimension of the space of essentially time- and band-limited signals. *Bell System Technical Journal*, 41:1295–1336, July 1962.

[SP00] M. Skoglund and S. Parkvall. Code design for combined channel estimation and error correction coding. In *Proc. IEEE International Symposium on Information Theory*, Sorrento, Italy, June 2000.

[TB02] G. Taricco and E. Biglieri. Exact pairwise error probability of space–time codes. *IEEE Transactions on Information Theory*, 48(2):510–513, February 2002.

[TdPE+04] J. Tubbax, L. Van der Perre, M. Engels, H. De Man, and M. Moonen. OFDM versus single carrier: A realistic multi-antenna comparison. *EURASIP Journal on Applied Signal Processing*, (9):1275–1287, 2004.

[Tel99] E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, Nov.-Dec. 1999.

[TJC99] V. Tarokh, H. Jafarkhani, and A. R. Calderbank. Space–time block codes from orthogonal designs. *IEEE Transactions on Information Theory*, 45(5):1456–1467, July 1999.

[TP98] L. Tong and S. Perreau. Multichannel blind identification: From subspace to maximum likelihood methods. *Proceedings of the IEEE*, 86(10):1951–1968, October 1998.

[TSC98] V. Tarokh, N. Seshadri, and A. R. Calderbank. Space–time codes for high data rate wireless communication: Performance criterion and code construction. *IEEE Transactions on Information Theory*, 44(2):744–765, March 1998.

[Ung74] G. Ungerboeck. Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems. *IEEE Transactions on Communications*, pages 624–636, May 1974.

[Ung82] G. Ungerboeck. Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory*, 28(1):55–67, January 1982.

[VHHK01] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath. Optimal training for frequency-selective fading channels. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 2105–2108, 2001.

[Vit65] A. Viterbi. Optimum detection and signal selection for partially coherent binary communication. *IEEE Transactions on Information Theory*, 11(2):239–246, April 1965.

[WÖst05] S. Williamson and F. Öst. Svensk telemarknad 2004. Technical report, Post- och Telestyrelsen, Available (October 2005) online at `http://www.pts.se/Archive/Documents/SE/sv_telemarknad_2004_1.pdf` (in Swedish), 2005.

[WSG94] J. H. Winters, J. Salz, and R. D. Gitlin. The impact of antenna diversity on the capacity of wireless communication systems. *IEEE Transactions on Communications*, 42(2):1740–1751, Feb/Mar/Apr 1994.

[WY02] J. Wang and K. Yao. Differential unitary space–time-frequency coding for MIMO OFDM systems. In *Conference Record of the 36th Asilomar Conf. on Signals, Systems and Computers*, volume 2, pages 1867–1871, Pacific Grove, CA, USA, November 2002.

[ZG03] S. Zhou and G. B. Giannakis. Single-carrier space–time block-coded transmissions over frequency-selective fading channels. *IEEE Transactions on Information Theory*, 49(1):164–179, January 2003.

[ZT03] L. Zheng and D. N. C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels. *IEEE Transactions on Information Theory*, 49(5):1073–1096, May 2003.