Utilizing Channel State Information in Space-Time Coding: Performance Limits and Transmission Techniques

George Jöngren

TRITA–S3–SB–0329
ISSN 1103-8039
ISRN KTH/SB/R - - 03/29 - - SE

Signal Processing
Department of Signals, Sensors and Systems
Royal Institute of Technology (KTH)
Stockholm, Sweden, 2003

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

Utilizing Channel State Information in Space-Time Coding:
Performance Limits and Transmission Techniques

Signal Processing Group
Department of Signals, Sensors and Systems
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

This thesis deals with performance limits and transmission techniques for a wireless communication link where at least the transmitter is equipped with an antenna array and moreover has access to possibly imperfect channel state information.

An antenna array on the transmit side provides the system with an extra spatial dimension that can be utilized for coding both in the spatial as well as the temporal domain. The recent development of such space-time codes shows that there are ways of exploiting multiple transmit antennas while completely avoiding traditional beamforming techniques' need of accurate channel state information. In practice however, the transmitter usually has access to some information about the current state of the channel. The available channel side information can then be used to improve the performance beyond what is possible using only conventional space-time codes. This, together with the need for reliable and fast communication, provides motivation for the work herein which shows how previous space-time coding concepts can be extended to take advantage of even non-perfect channel knowledge at the transmitter.

Performance limits are investigated using tools from information theory. An expression for the channel capacity for the wireless link under consideration is presented. One important result is that adjusting the output of a conventional space-time encoder by means of a transmit weighting matrix that only depends on the channel side information constitutes a capacity achieving transmitter structure. Computational procedures for evaluating the capacity expression are considered and used to obtain numerical results illustrating the gains due to channel knowledge.

The other parts of the thesis are devoted to devising practical methods for exploiting channel knowledge in conjunction with space-time coding. A new performance criterion is developed that takes the quality of the channel side information into account. Motivated by the optimality of

separate space-time coding and transmit weighting, the performance criterion is used for determining a suitable transmit weighting matrix that adapts a predetermined orthogonal space-time block code (OSTBC) to the available channel side information. The result is a low-complexity *weighted OSTBC* transmission scheme providing a seamless combination of the normally complementary strengths offered by conventional beamforming and OSTBC.

Scenarios in which the channel side information takes the form of quantized channel estimates obtained from a feedback link are also considered. The channel feedback is assumed to suffer from quantization errors, feedback delay and bit-errors introduced by a noisy feedback channel. Methods to design the quantizer in the feedback link so as to mitigate all these errors are investigated. By introducing heuristic modifications of our previously developed transmission technique, it is shown how robustness against all three types of channel feedback impairments may be achieved.

To avoid the use of heuristics in case of quantized channel side information, yet another new performance criterion is developed specifically for the problem at hand. Based on the performance criterion, a procedure for utilizing the available side information in the design of unstructured space-time block codes is proposed. These codes offer maximal design freedom at the expense of an increased decoding complexity. Properties of the resulting codes are investigated both analytically and experimentally. The codes outperform corresponding OSTBC schemes even when no channel knowledge is available at the transmitter.

In addition to unstructured codes, closely related techniques based on the same performance criterion are used for designing some linear dispersive space-time block codes as well as designing suitable transmit weighting matrices for weighted OSTBC. An interesting observation that deserves further study is that the design procedure for linear dispersive codes in case of no channel knowledge at the transmitter appears to automatically produce orthogonal space-time block codes, if the parameters under consideration allow it.

Acknowledgements

I would like to express my sincerest gratitude to my advisors Professor Björn Ottersten and Associate Professor Mikael Skoglund for their guidance, inspiration and support during the course of this work. They have always taken the time to come with encouragement and assistance, even up to the very last minute as I have invented new ways of meeting a deadline with an infinitesimal margin. It has been a pleasure working with them and I feel privileged to have experienced their enthusiasm for research.

Probably no acknowledgement here at the Signal Processing Group is complete without thanking Dr. Mats Bengtsson, our guru on the LaTeX type setting system. Mats has been really helpful when I have struggled to get my documents to look the way they should. Another person I feel indebted to regarding LaTeX support is my office neighbor Tech. Lic. Tomas Andersson who has always been quick to come at my rescue.

I am grateful to my colleagues M.Sc. Joakim Jaldén, M.Sc. David Samuelsson and M.Sc. Xi Zhang for their generous assistance in proofreading this thesis. Thanks must also go to the other members of the Signal Processing Group for making my time here more enjoyable and for never hesitating to lend me "lunch coupons" whenever I needed – I hope I have settled all my debts.

Our computer administrators Nina Unkuri and Andreas Stenhall deserve to be acknowledged for the excellent service they provide. I still remember the days of nonexistent computer support. Thanks to you, this memory is quickly fading away.

I also thank Anna for her love and support, not to mention patience, all of which I have depended upon during this entire time.

This work was supported in part by the Swedish Foundation for Strategic Research (SSF) through the project "Adaptive Antennas in Wideband Radio Access Networks" in the Personal Computing and Communication Program (PCC).

Contents

Chapter 1

Introduction

The use of wireless communication has literally exploded during recent years. Not long ago was a mobile phone seen as a luxury item and status symbol affordable by only a few. Nowadays, wireless communication is taken for granted and a mobile phone has become a natural accessory for many. Driven by the demand for land-mobile communication, wireless networks have been deployed around the world. So far, voice communication has been the major application. Current second generation networks such as the widespread GSM system [RWO95] have been designed with this primarily in mind. In the future, it is envisioned that data services providing, for example, Internet access will be another popular application. If the predictions come true, it is likely that there will be a strong demand for data rates dramatically higher than the rather limited communication speeds provided by present second generation equipment. Infrastructure for WCDMA [HT02] and other third generation networks has therefore recently started to be deployed with the hope of offering significantly higher data rates than what has been previously possible.

In wireless networks for land-mobile communication, the geographical area over which service is offered is usually divided into cells. Each cell contains a base station which handles the communication with the mobile user terminals assigned to that cell. Thus, a single base station handles several communication links at the same time. Data is transmitted in both directions, from the base station to the user terminal and vice versa. This is called duplex communication. The communication link from the base station to the terminal is referred to as the downlink while the uplink corresponds to the reverse direction.

In order for a wireless network to accommodate many users and provide high data rates within the typically limited radio spectrum available, it is important that the system is spectrally efficient. In essence, the system should provide as high data rates as possible using the least amount of bandwidth with a minimum of errors in the communication. The imperfections of the wireless communication channel, not to mention constraints on cost and size of equipment, make achieving this a challenging task.

A promising method for increasing the spectral efficiency of the system is to use multiple antennas, also known as antenna arrays, for the transmission and reception of the radio signals. This adds a spatial dimension to the system which can be exploited for reducing many of the problems associated with wireless communication. Techniques for utilizing an antenna array on the receiving side have been known for many years [Jak94] and a wealth of efficient methods exist in the literature.

In current, and conceivably also in future, networks for land-mobile communication, requirements on the size and cost of user terminals make antenna arrays with many elements impractical on the user side. At the base station, however, such arrays are much easier to motivate, this since the constraints on size are not as strict and the cost is shared by all the users that the base station handles. Hence, antenna array reception techniques are likely to improve the performance of the uplink more than what is possible in the downlink. The resulting imbalance in performance is particularly problematic in view of the asymmetric traffic pattern typical of many future data service applications. Internet browsing and streaming multimedia are just two examples of applications where higher data rates are needed in the downlink than in the uplink.

To reduce the performance gap between the uplink and downlink, it is necessary to exploit the spatial dimension also in the downlink by means of a transmit antenna array. Methods for successfully utilizing an antenna array for transmission purposes have traditionally relied on accurate channel state information. Since such information is harder to obtain at the transmitter than at the receiver, an antenna array at the transmitter has been viewed as comparatively more difficult to utilize than an array placed at the receiver. Recent advances in the area of multiple antenna transmission without the use of channel state information help in overcoming this difficulty. In particular, the development of efficient space-time codes [GFBK99, TSC98] that utilize both the spatial and the temporal dimensions for coding the message to be communicated shows that there is hope of achieving high data rate communication also in the

downlink, even if no channel knowledge is available. In case both the transmitter as well as the receiver are equipped with multiple antennas, properly designed space-time codes offer, in both directions, dramatically higher data rates than when only a single antenna array is used [Tel95, FG98].

Conventional space-time codes are designed under the assumption of no channel knowledge at the transmitter. Although this is motivated by the difficulties in obtaining accurate information about the channel at the transmitter, partial or imperfect channel information is often available. Assuming the quality of the channel knowledge is properly taken into account, such channel information can be used to improve the performance beyond what is possible with conventional space-time coding or other more traditional techniques.

The need for reliable and fast communication combined with the potential inherent even in non-perfect channel knowledge at the transmitter motivates the work presented herein. The work in this thesis aims at investigating performance limits and developing efficient transmission schemes and space-time codes for a certain wireless digital communication link where at least the transmitter is equipped with an antenna array and has access to possibly imperfect channel information. The presentation primarily focuses on applications in land-mobile wireless networks. Other possible applications of the work include scenarios that may arise in wireless local area networks and wireless local loops.

1.1 Wireless Communication

Before proceeding with the development, let us review some basic concepts related to wireless digital communication. The focus here is on an individual single-input-single-output (SISO) link in a wireless network for *land-mobile* communication where the transmitter and the receiver each are equipped with a single antenna. However, many of the issues discussed in this section are relevant also for digital communication systems in general, wireless or not.

The actions of a wireless communication system can roughly be summarized as follows [Pro95]. Bits representing the message to be communicated are first coded into a sequence of symbols. These information bearing symbols are modulated into a time-continuous waveform which is upconverted to carrier frequency and transmitted over the wireless channel. At the receiver, the information carrying signal is picked up by the

antenna, downconverted to baseband and demodulated into a sequence of samples which are decoded into bits. If everything has gone well, the communication is error-free and the obtained bits hence coincide with the transmitted ones

1.1.1 Distinctive Properties of the Wireless Channel

Unfortunately, the wireless propagation medium is far from ideal. Additive thermal noise disturbs the information carrying signals. Interference from other wireless users may also plague the transmission. If the noise and interference are sufficiently strong compared to the information carrying signal, it becomes difficult for the receiver to correctly detect the transmitted message. Hence, the signal-to-noise-ratio (SNR) and the signal-to-interference-plus-noise-ratio (SINR) are two relevant parameters. These power ratios are important since they give an indication of the performance of the system and are often relatively easy to measure.

Wireless systems are especially prone to errors in the communication since the signal attenuation incurred by the channel may be very large. This problem is made worse by the fact that the transmitted radio signal interacts with objects in the physical environment [BD91, Bur96]. As a result, the signal usually propagates along several different paths before it arrives at the receiver. The phenomenon is termed *multipath* and is illustrated in Figure 1.1, where only two propagation paths are indicated. Each propagation path affects the signal differently which means that the received signal is a superposition of different, possibly delayed, versions of the original signal. These multipath components add constructively or destructively, depending on the surrounding terrain and the positions of the transmitter and the receiver. The signal level at the receiver may therefore fluctuate wildly over time due to changes in the environment and movement of the transmitter/receiver. In the worst case, such channel fading makes the attenuation so large that the receiver is unable to obtain a useable signal.

If the *delay spread* [Pro95, p. 763] of the multipath is small relative to the inverse bandwidth of the transmitted signal, the individual multipath components are not resolvable and the effective communication channel is therefore essentially frequency-nonselective or flat fading. Consequently, the different frequency components of the information bearing signal undergo the same attenuation and phase shift when propagating through the channel. The channel[1], including the up- and downconversion in fre-

[1] In this work, the notion of a "channel" is used in a rather sloppy manner in the

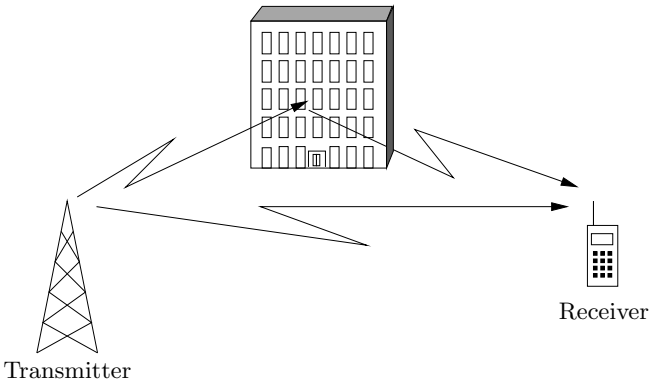

Figure 1.1: The radio signal arrives at the receiver along several different paths, so-called multipath propagation.

quency as well as transmit and receive filtering, may then be modeled by a filter with only one complex-valued tap, or coefficient. A common assumption is that the single channel coefficient that determines the attenuation and phase shift fades according to a complex Gaussian[2] random process [Pro95, p. 761]. Such fading is also known as flat *Rayleigh* fading, since the magnitude of the channel coefficient is Rayleigh distributed. High data rate communication usually requires such a large bandwidth that at least some of the multipath is resolvable. The result is a frequency-selective fading channel that can be modeled by a finite impulse response filter with several complex-valued taps.

This way of absorbing the up- and downconversion into an effective channel is common practice in the field of communication theory and leads to a so-called complex baseband equivalent model of the system [ZM92, Pro95]. When using such a model, both the transmitted and received signal, as well as the channel itself, are potentially complex-valued. Also the additive noise may be complex-valued and is often modeled as

sense that it may or may not include additive noise (and interference). Hopefully, the exact meaning will be clear from context. The present case is an example of when the noise is excluded. Consequently, the term "channel" here solely refers to the mentioned filter. In the following, noise will be considered to be part of the "channel" primarily in connection with information theoretic discussions so as to comply with common conventions within that field.

[2]The theory of *complex* Gaussian distributions and random processes is described in e.g. [Kay93].

a wide sense stationary complex Gaussian random process. Much of the processing of signals in the system can therefore be thought of as being carried out in the complex number field. Moreover, since sampling is typically part of the demodulation at the receiver, sampled versions of the signals in the complex baseband equivalent model are usually considered when analyzing the system. The subsequent presentation generally makes use of such sampled complex baseband representations of signals and channels, unless explicitly stated otherwise.

1.1.2 Fighting Channel Fading

The presence of channel fading is one of the major difficulties associated with wireless communication. An often used strategy for dealing with the fading problem is to employ so-called diversity techniques. The basic idea behind diversity methods is to provide the receiver with several versions of the same information bearing signal where the various versions have been affected by different, preferably independently fading, channels. Hopefully, at least one of the received signals has experienced a channel with little attenuation, thereby increasing the chance that the message can be correctly detected at the receiver. It can be shown that the probability of an error in the communication generally decreases with an increasing number of signal replicas (assuming that the signal replicas have undergone reasonably independent fading). Two examples of common diversity techniques for single antenna systems are listed below.

- *Time diversity*: The same information is transmitted on different time slots where the time slots are sufficiently separated in time so that the fading has changed the channel significantly from one slot to another. Independently fading channels is ensured by letting the time separation of successive slots be large compared with the *coherence time* [Pro95, p. 765] of the channel.

- *Frequency diversity*: In the case of a frequency-selective fading channel, diversity can be obtained by transmitting the same information on different carrier frequencies. As long as the carrier separation is large compared with the *coherence bandwidth* [Pro95, p. 764] of the channel, the signals experience roughly independent fading. A more direct but less obvious diversity approach is to transmit on a single carrier but with a bandwidth large enough for some of the multipath components to be resolvable at the receiver. The resulting distortion of the information carrying signal can be

handled by appropriate processing at the receiver. In any case, the frequency-selectivity of the channel serves to protect against fading and should hence not only be seen as a problem.

1.1.3 Reliable Communication by Means of Coding

To limit the detrimental consequences of noise, interference and fading, the data can be coded prior to the transmission so that the information carrying signals intentionally contain redundant information. The redundancies introduced by the channel code help the receiver to detect the transmitted message without errors, albeit at the cost of a reduced data rate if the communication bandwidth is fixed. Channel coding is a very active area of research, see for example [FU98] for an overview of a part of this field. Indeed, any mapping from the data to be communicated to the information carrying signal can be thought of as a sort of channel coding. If this very wide definition of coding is adopted, all communication setups utilize some sort of coding.

In the above time diversity technique, redundant information is introduced by repeatedly transmitting the same signal. This can be thought of as a simple form of channel code that lowers the probability of a detection error. The reduction in data rate resulting from the use of such a repetition code is unfortunately substantial. Many other channel codes, with perhaps a more suitable data rate versus error probability tradeoff, exist in the literature.

1.1.4 Channel Capacity

No matter how sophisticated channel codes or signal processing techniques the system has at its disposal, the channel itself sets a fundamental limit on the speed of reliable communication within a certain fixed bandwidth and power. Pioneered by Shannon in his famous work in [Sha48], the field of information theory provides an important such performance limit called channel capacity. The channel capacity is a measure that quantifies the highest data rate that a communication channel[3] can maintain while keeping the transfer of information essentially error-free. Loosely speaking, as long as the data rate is below the channel capacity, there exist channel codes that lead to a vanishingly small error probability if the processing time is allowed to tend to infinity. For codes with

[3]The term "channel" now refers to a model describing the relation between transmitted and received signals, i.e., noise is included if present.

data rates higher than the channel capacity, this is not possible. A more precise definition of channel capacity is given in [Gal68, CT91].

It is important to keep in mind that channel capacity is an asymptotic measure in the sense that the coding of the message is over an infinite time period and the decoding is based on the received signal during this entire time period. This approximates a situation where the receiver waits until the end of the transmission before starting to decode the message. There are hence no constraints on the processing delay, which is a drawback from a practical point of view since processing delay is a critical parameter in many applications. Another drawback is that no restrictions on the computational complexity of the encoder and the decoder are imposed. Consequently, coding schemes that achieve the data rate promised by the channel capacity are typically notoriously computationally demanding and are therefore unsuitable for implementation in practice. Nevertheless, the properties of these capacity achieving codes often provide guidance on how good, more practically oriented, codes or transmission schemes should be structured.

The concept of channel capacity is also useful for comparing the potential performance of different communication setups. In the case of wireless communication it can be used to study the impact of the problems discussed previously. This information theoretic tool provides an ultimate limit on the speed of communication for a certain model of the channel. The only way to increase the potential data rate is to, in some sense, improve the actual channel. One way to accomplish that is discussed in the following section.

1.2 Antenna Arrays

The use of antenna arrays is seen as a promising approach for coping with many of the problems associated with wireless communication [PP97]. An array of multiple antennas may be placed at the receiver, the transmitter, or at both sides of the communication link. The antennas in the antenna array are placed at different physical positions in space. Alternatively, the polarization may vary among the antennas. In any case, an antenna array gives the system access to an extra spatial dimension that can be utilized in conjunction with its temporal counterpart for increasing the performance beyond what is possible with pure single antenna transmission and reception. With proper such spatio-temporal processing, antenna arrays can mitigate noise and interference as well as provide

protection against fading.

1.2.1 Exploiting a Receive Antenna Array

The classic use of antenna arrays is on the receiver side. The resulting single-input-multi-output (SIMO) channel provides the receiver with several versions of the transmitted signal. Each signal version passes through a corresponding SISO channel, typically modeled as a finite impulse response filter, and is thereafter disturbed by additive noise before before being available for receiver processing. A schematic model of the whole setup is illustrated in Figure 1.2.

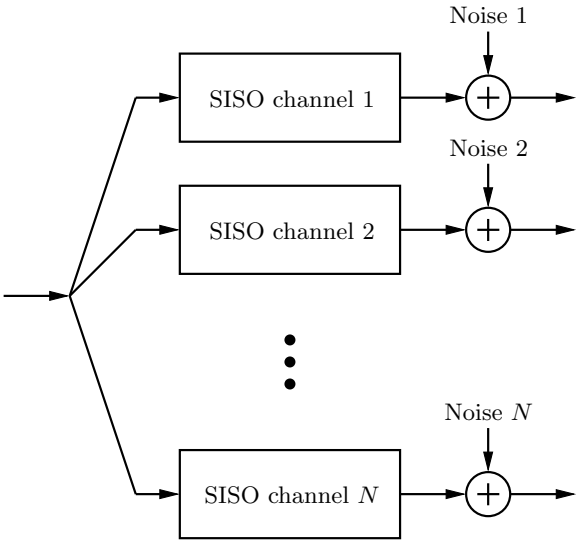

Figure 1.2: Model of a SIMO link corresponding to a system with one transmit antenna and N receive antennas.

The signals at the receiver can be combined in a way so as to suppress noise and increase the SNR. This *array gain* is however not the only benefit. If the antennas are spaced sufficiently far apart for the fading of the individual channels to be reasonably uncorrelated, the antenna outputs may be used for obtaining *spatial diversity*. Well-known such techniques that provide maximum possible spatial diversity include antenna selection and maximum ratio combining [Jak94]. In the former method, the

antenna output with the strongest signal is selected while in the latter both diversity as well as array gain is obtained by adjusting the phase and amplitude of each signal so that the antenna outputs add coherently and maximize the SNR. Maximum ratio combining is designed for a flat fading scenario and can be seen as implementing a simple spatial filter that is matched to the SIMO channel's coefficients. After the combining, a one-dimensional signal is input to the detector. Thus, an equivalent SISO channel is created with properties better than those of the individual SISO channels. In a frequency-selective scenario, the coefficients of a more general spatio-temporal filter structure can be optimized to increase the SNR while equalizing the distortion caused by the channel [BS91].

If the radio signal arrives at the receiver from a distinct direction, the corresponding propagating wavefront will reach the constituent antennas in the array at different points of time in a predictable manner. Assuming that the inverse of the signal bandwidth is sufficiently large compared with the time it takes for the wavefront to pass all antennas, the resulting time differences can be modeled as phase shifts. Based on the effects of these time differences, or phase shifts, the direction of arrival (DOA) of the incoming signal [KV96] can be estimated. Such directional information may be used for designing spatial filters that amplify signals with a certain DOA and suppress signals impinging from other directions [AMVW91]. This is useful for suppressing interference since interfering signals often arrive from another direction than the signal of interest.

In scenarios where there are no distinct directions of arrival, such as when the signals emanate from diffuse scattering on objects in the environment, it becomes difficult to separate the signal of interest from the interference based on physical DOA parameters. However, even if this is the case, the spatial and/or temporal signatures of the various signals usually still differ significantly. This can be exploited for interference suppression without directly relying on physical parameters. In [Asz95, BJ95, May97, JAO00], the spatial correlation of the interference is utilized in the metric of a maximum likelihood sequence estimator [Pro95] to suppress interference while at the same time dealing with the distortion of a frequency-selective channel. By taking the joint spatio-temporal correlation of the interferers into account, the ability to suppress interference with similar approaches can be improved even further [ME86, AO98, BMC99]. Interference can also be suppressed using spatio-temporal filtering techniques [CGS94].

1.2.2 Exploiting a Transmit Antenna Array

Placing an antenna array at the transmitter and using a single receive antenna creates a multi-input-single-output (MISO) channel. This is clearly the dual of the previous SIMO case. Multiple signals are now transmitted, instead of received. Each transmitted signal can be thought of propagating through a separate SISO channel and the outputs of all the SISO channels and noise are added to form the received signal, as illustrated in Figure 1.3.

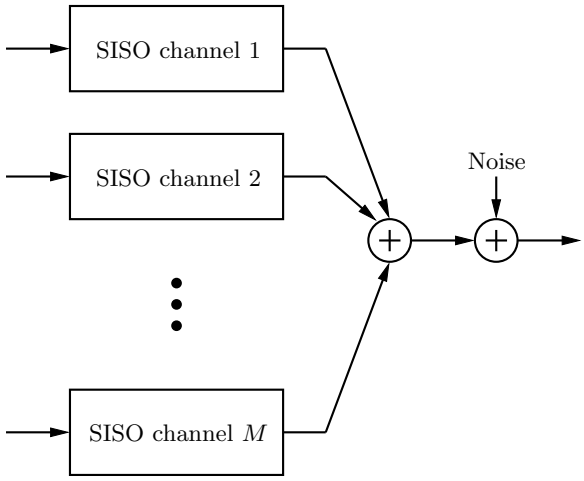

Figure 1.3: A model of a MISO link corresponding to a system equipped with M transmit antennas and one receive antenna. The received signal is obtained as a superposition of all the transmitted signals plus noise.

Because the signals are combined before they are available for receiver processing, schemes for exploiting the spatial domain must be placed on the transmit side. Compared with when the receiver is equipped with an array, it may seem considerably more difficult to achieve performance gains since the many emitted signals are no longer separated and, in fact, interfere with each other at the receiver. Whether this is a real problem or not has traditionally been believed to depend on the degree of channel knowledge at the transmitter.

If the transmitter knows the parameters describing the channel accurately, it can, at least in principle, easily compensate for the influence of the channel on the signals so that the signals combine coherently at

the receiver, thus avoiding interference. Hence, assuming the transmitter has accurate channel knowledge, transmit antenna arrays are relatively straightforward to exploit.

However, in the absence of channel state information, the transmitter has no way of knowing how the channel has affected the relative phases and time delays of the signals when they combine at the receiver. Channel knowledge at the transmitter is typically more difficult to obtain than at the receiver. Because of these two factors, exploiting the potential inherent in transmit antenna arrays has been viewed as difficult and has consequently received rather limited interest. A classic solution that avoids the use of channel information is to transmit the same signal on all antennas but with different carriers so that the signals do not overlap in frequency [Jak94, p. 512]. The signals can then be separated at the receiver by means of bandpass filtering and thereafter be combined to obtain transmit diversity. Unfortunately, this comes at the price of significant bandwidth expansion. It is only recently that more efficient techniques coping with no channel knowledge have been developed.

Channel Knowledge at the Transmitter

When there is significant channel knowledge at the transmitter, some of the methods applicable when the array is placed at the receiver have direct counterparts on the transmit side. For example, the equivalent of maximum ratio combining on the transmit side is transmit beamforming [MM80]. In beamforming, scaled and phase shifted versions of the same information carrying signal is transmitted from the different antennas in such a manner that they add constructively at the receiver with the aim of maximizing the SNR. Beamforming in its simplest form corresponds to a purely spatial filter with the same coefficients as in maximum ratio combining, i.e., the filter is matched to the channel coefficients. This converts the original MISO channel into an artificial SISO channel. Assuming a flat fading scenario, the properties of the SISO channel are such that both the diversity and the array gain are of the same order as what maximum ratio combining with the same number of antennas provides. In other words, beamforming achieves all the spatial diversity the system has to offer. More general spatio-temporal filtering on the transmit side is also possible [Koh98].

Beamforming changes the antenna pattern so that the radiated energy is mostly confined to a narrow beam pointing in a certain direction. There is leakage of energy in other directions as well but it is small when

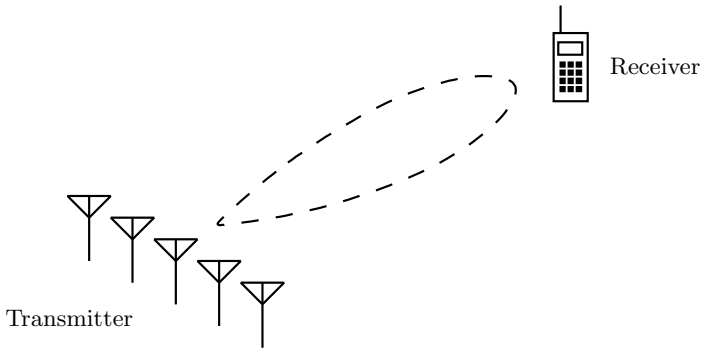

Figure 1.4: Beamforming uses the antenna array to focus the radiated energy in the direction of the receiver.

there are many antennas in the array. If there are no dominant multipath scatterers other than in the immediate vicinity of the receiver, the main beam is directed towards the receiver, as illustrated in Figure 1.4. Physical information about the angular position of the receiver may in this case be used for directly determining appropriate, albeit not optimal, weights of the spatial filter.

In a multiple-access scenario, steering the beam towards the intended receiver has the additional benefit of reducing the impact of the transmission on other user's communication, thus lowering the interference level in the system. This also opens up the possibility for increasing the number of users that can be served in the same frequency band and time slot by spatially separating the different communication links. Several techniques for performing spatial division multiple-access based on directional information and channel statistics have been proposed in the literature [SBEM90, ZO95, XL95, AHD+99, KTT+99]. The basic philosophy behind these schemes is to create a peak in the antenna radiation pattern in the direction of the intended receiver and essentially place nulls in the directions of the other users. Similar methods not directly based on physical channel parameters have also been developed [GP94b, RDJP95, Zet99].

No Channel Knowledge at the Transmitter

When there is no channel knowledge at the transmitter, coherent combining by means of beamforming is not possible. The transmitter is

effectively blind since it cannot predict how the channel will affect the transmission. Despite this fact, an antenna array may provide spatial diversity of the same order as when there is perfect channel knowledge or when the array is placed at the receiver.

Early attempts at exploiting multiple antennas for blind transmitters are focused on achieving spatial diversity. A common strategy is to convert the MISO channel into a synthetic SISO channel upon which standard scalar error-correcting codes can be used. Many ways to accomplish that have been proposed in the literature.

In [HAN92, Wee93, KF97], phase shifted versions of the same information carrying signal are multiplexed to the different antenna elements. By making the phase shifts time-varying, a possibly static MISO channel is transformed into a fast fading SISO channel. The artificially created time-varying channel is used together with conventional time diversity methods such as coding combined with interleaving [Pro95, p. 468]. Thus, spatial diversity is transformed into time diversity.

Another way of producing similar time-variations is to transmit on only one antenna at a time and let the antennas take turn to transmit. The effective SISO channel coefficient thus alternates among the coefficients in the MISO channel. Such a time division approach was proposed in [SW93], where a simple repetition code was used to exploit the resulting temporal variations of the channel. Although the method provides diversity, it comes at the expense of reduced data rate due to the repetition code. A more bandwidth efficient technique called delay diversity was therefore also considered in the same paper.

Several other papers deal with delay diversity [Wit91, Wit93, Mog93, Win94, Win98]. The technique was originally proposed in [Wit91] and offers diversity by multiplexing time-delayed versions of the same information carrying signal onto the different antennas. The time delay increases linearly from no delay at the first antenna to some maximum delay at the last antenna. Usually, the time delay differs with one symbol period between two consecutive antennas. The result is a tapped delay line SISO channel or, in other words, a frequency-selective channel. Decoding the received signal through the use of a maximum likelihood sequence estimator captures the frequency diversity of the synthetic SISO system. A benefit of this strategy is that it does not waste waste bandwidth as in the time division repetition code based approach in [Win98]. Other standard frequency diversity methods are also applicable if bandwidth expansion is tolerated.

In the above blind diversity schemes, a coded stream of information

bearing scalar symbols is spread over the antennas using various multiplexing methods. The multiplexing techniques can all be seen as implementing special cases of a SIMO, possibly time-varying, linear spatio-temporal transmit filter. The resulting synthetic SISO link is represented by the compound channel from the input of the filter to the output of the single receive antenna. These ways of multiplexing are also commonly referred to as linear precoding. Clearly, delay diversity is a special case of linear precoding with a simple time-invariant spatio-temporal filter whose constituent single tap SISO filters implement the delays. The time division method corresponds to a strictly spatial time-varying filter with the only non-zero coefficient placed at the currently transmitting antenna. Similarly, the frequency shifting approach corresponds to a purely spatial filter with time-varying complex exponentials as coefficients.

Since also beamforming is equivalent to a spatial filter with complex-valued taps, frequency shifting implements a form of time-varying beamforming where the direction of the beam thus varies with time and only by chance points towards the intended receiver. Consequently, the instantaneous SNR, and hence the overall performance, is typically much lower than the maximum provided by conventional beamforming.

Information theoretic performance limits of all the mentioned blind linear precoding schemes have been investigated in [NTW99] and compared with the performance of optimal non-linear vector coding. In addition, a random time-varying beamformer was proposed as a low-complexity method for exploiting some of the capacity of the system. The schemes mentioned so far all represent rather rudimentary and specialized examples of spatio-temporal filtering. A broader class of linear precoding techniques has been considered in [WT97].

It should be noted that encoding the data into a stream of scalar symbols and thereafter applying any of the previously mentioned linear precoding schemes is typically not optimal in a mutual information sense when there is no channel knowledge at the transmitter [NTW99]. Although such transmit structures permit simple decoding, they are not general enough to achieve the maximum possible data rate that the system can offer. Since the MISO channel takes a vector-valued input, the channel encoder needs to be designed specifically for vector (as opposed to scalar) symbols as output in order not to limit the potential data rate. Thus, for maximum performance, specifically developed vector codes need to be used that spread the information both in time and space.

Vector codes in transmit antenna array applications are more popularly known as space-time codes and have recently received considerable

attention because of the high data rates and reliable communication they may provide. Most of the literature on the design of space-time codes focuses on a flat fading scenario in which the receiver is assumed to know the channel state parameters perfectly. The same code can usually be used regardless of the number of antennas at the receiver. The benefits of deploying antenna arrays at both ends of the communication link will be further discussed in the following section.

A systematic approach to designing space-time codes was pioneered in [GFK96, GFBK99] where a design criterion was derived and used for verifying the performance of some simple space-time codes. Essentially the same design criterion was later derived in [TSC98] and shown to hold for a number of different channel models. In the latter work, the now popular notion of space-time coding was introduced. Motivated by the design criterion, some examples of trellis codes realizing the full spatial diversity potential of the system were proposed. In addition to diversity, the proposed space-time trellis codes give coding gain. The coding gain increases with the number of states in the trellis, at the expense of higher decoding complexity. These codes were basically handcrafted. Automatic design procedures for space-time trellis codes have been developed in e.g. [BBH00, Blu02] and used for obtaining more efficient codes than the ones found in [TSC98].

An extremely simple yet novel space-time block code for two transmit antennas was given in [Ala98]. The code is commonly known as the Alamouti code after its inventor and has the appealing property that the two antenna signals are orthogonal in time (as well as in space) without the bandwidth expansion normally associated with orthogonal signaling. The code is orthogonal in other ways too. Due to the orthogonal properties of the code, the signals are easily separated at the receiver and maximum possible spatial diversity gain is achieved. An additional and important advantage of this *orthogonal space-time block* (OSTB) code is that optimal low-complexity decoding is possible. Corresponding codes for up to eight transmit antennas was later given in [TJC99], where also further theory was developed.

Other examples of OSTB codes are given in e.g. [GS01, SX03]. As pointed out already in [TJC99], an OSTB code is a special case of a so-called *linear dispersive space-time block code*. That is, the code is linear in terms of some constituent information carrying symbols. The individual codewords can hence be obtained as a linear combination of these symbols where each symbol is weighted by a corresponding weighting matrix. By a clever choice of weighting matrices, OSTB codes arise. However,

other weights are also certainly possible. In [HH02b], the weighting matrices were designed to maximize a channel capacity based performance criterion. The resulting linear dispersive codes provide high reliable data rates if concatenated with suitable outer codes. Thanks to the linear structure, it is possible to apply near optimal decoding techniques that are more computationally efficient than the worst case of a full search [VH02, JMO03].

The space-time coding references mentioned so far all assume perfect channel state information at the receiver. Non-coherent detection scenarios corresponding to situations in which there is no channel knowledge at the receiver have also been considered in several papers. In [TAP98], it was shown that OSTB codes can be decoded without any channel knowledge at the price of some reduction in performance. Also, the classic concept of differential coding [Pro95] for non-coherent detection has been extended to the problem at hand, see e.g. [HS00, Hug00, TJ00, HH02a, GS02].

Another strategy for exploiting multiple transmit antennas is to use conventional scalar codes together with spatial interleaving. The idea is basically to introduce spatial redundancy by spreading the output of a scalar encoder "evenly" over the different antennas. Transmission schemes in this category include the well-known BLAST architecture [Fos96], turbo [SD01] and trellis based techniques [RJ99]. In the latter work, orthogonal frequency division multiplexing (OFDM) [WE71, Cim85, Bin90] is used to handle a frequency-selective channel by dividing the bandwidth into narrow subbands creating a set of parallel flat fading channels. This is a well-known technique for dealing with frequency-selective channels and has also been used together with classical space-time trellis codes in [ATNS98]. By utilizing time reversal techniques, the concept of OSTB codes can be extended for this scenario as well [LP00]. Space-time block codes for frequency-selective channels have also been designed specifically for non-coherent receivers [GS03].

1.2.3 The Potential of Dual Antenna Arrays

Obviously, both sides of the communication link may be equipped with antenna arrays. The resulting multi-input-multi-output (MIMO) channel represents a natural extension of the previously described MISO case. Such dual antenna array systems offer more degrees of spatial freedom for the spatio-temporal processing to exploit. In sufficiently rich multipath scattering environments, these extra degrees of freedom lead to a channel

capacity substantially higher than when only a single antenna array is used [Tel95, FG98, RC98, RJ99], regardless of whether the transmitter knows the channel parameters or not.

The use of dual antenna arrays in rich scattering environments gives rise to a multiplicative effect that makes the channel capacity increase essentially a constant integer factor faster with respect to the SNR than comparable SISO, MISO or SIMO systems [FG98, RJ99]. The numerical value of the factor is given by the minimum of the number of antennas at the transmitter and receiver, respectively. Intuitively speaking, the MIMO channel broadens the channel in the sense that many parallel "data pipes" are available for the communication. The number of data pipes corresponds to the multiplicative factor mentioned above. This explains the improvement in capacity compared with systems that do not use dual antenna arrays.

The encouraging capacity results exhibited by MIMO systems suggest that reliable and high data rate communication may be accomplished in ways that do not incur significant bandwidth expansion. Transmission schemes for multiple transmit antennas generally work with dual antenna array systems as well without modification, but with better performance. In fact, many of the space-time codes described above are designed with an arbitrary number of receive antennas in mind. Some transmission schemes can however be greatly improved if they are specifically adapted to the MIMO channel. For example, if the channel knowledge at the transmitter is perfect, conventional beamforming unnecessarily limits the possible data rates. Instead a related multi-dimensional type of beamforming as in [RC98] is necessary in order not to lose capacity.

Investigations regarding MIMO channel capacity have been conducted for several different scenarios. The classic MIMO capacity formula for flat fading channels with no channel knowledge at the transmitter and perfect channel knowledge at the receiver was presented in [Tel95]. Therein, the capacity gains possible in a Rayleigh fading scenario were illustrated and analytical results provided. The potential for high data rates was further demonstrated in [FG98] using the concept of outage capacity. The case of perfect channel knowledge at the transmitter under the assumption of constant channel parameters was briefly considered in [Tel95] where it was pointed out that traditional water-filling [CT91, p. 253] techniques may be employed. Works on frequency-selective channels treating the cases of perfect and no channel knowledge at the transmitter have been reported in [RC98] and [RJ99], respectively. The MIMO channel capacity when both the transmitter as well as the receiver completely lack channel

state information (corresponding to a non-coherent detection scenario) was derived in [MH99] under the assumption of a flat fading scenario.

Most works on MIMO capacity and space-time coding primarily considers channel models that correspond to direct generalizations of standard SISO [BD91, Bur96] and MISO/SIMO channel models [ECS+98]. More advanced and perhaps more realistic channel models specifically developed for MIMO systems are found in e.g. [SFGK00, GBGP02, KSP+02, WJ02].

1.3 Obtaining Channel State Information

As indicated by the previous discussion about antenna arrays, many techniques both on the transmit and receive side rely on channel state information. Due to the random and time-varying characteristic of the wireless propagation medium, the parameters constituting the channel state information need to be continuously estimated.

Channel estimation may be performed by exciting the system with suitably chosen signals and studying the resulting output to deduce the parameters in the channel model. The model of the channel may be formulated in terms of physical parameters such as DOA, propagation path delays and gains [ECS+98] or in a more abstract form where for example coefficients of finite impulse response filters constitute the channel state.

The most common way to estimate the channel is through the use of training based methods where signals known to the receiver excite the channel. Standard methods in estimation theory [Kay93] may then be utilized for determining the channel parameters. When there are multiple antennas on the transmit side, the choice of training sequences on the different antennas becomes particularly critical in order to obtain good channel estimates [GFBK99].

Another approach is to use channel estimation methods which do not rely on knowing the transmitted signals. Such blind techniques identify the channel solely based on received signals by utilizing the rich structure typical of communication signals and channel models. A survey of some recent blind methods is given in [TP98].

Regardless of whether blind or training based methods are used, the resulting channel estimates are directly available for receiver processing. Obviously, the estimates will not be perfect. Both noise and mismatch between the assumed model and the actual system introduce errors in the channel state information.

There are several ways in which channel estimates obtained at the receiver can be used also for transmission purposes. When the system employs duplex communication, the same device is both a transmitter and a receiver depending on the communication direction. Channel estimates based on reverse link data, as obtained in receive mode, can then be utilized also in the forward link, i.e., in the transmit mode. The two links in duplex communication are however usually separated either in time or frequency and this may severely degrade the quality of the channel information in the transmit mode compared with the receive mode.

In time division duplex (TDD) systems, the forward and reverse links are active on different non-overlapping time slots. Both communication directions use the same carrier frequency. As a result, assuming the duplex time is short compared to the coherence time of the channel, the forward and reverse radio channels are the same. Due to this reciprocity, a channel estimate obtained in receive mode can be directly used in transmit mode, after taking into account that the transmit and receive filters may be different. On the other hand, if the duplex time is not short enough, the channel may have changed considerable compared with the state it was in when the channel estimate was obtained in the receive mode. In other words, the channel estimate is outdated, making the channel state information available in the transmit mode much worse than what the receive mode has access to. This discrepancy may to some extent be lessened by utilizing the temporal correlation of the channel for estimating the current channel based on the outdated channel information [DHSH00].

The same strategy of utilizing a channel estimate obtained in the reverse link for forward link processing is applicable also in frequency division duplex (FDD) systems, under certain conditions. Different carrier frequencies are now used in the two communication directions. If the frequency separation is sufficiently small compared with the coherence bandwidth of the channel, the forward and reverse link channels are essentially the same and just as for the above TDD case, the same channel estimate can be used directly in both the transmit and the receive mode if the differences in transmit and receive filters are compensated for. Frequency correlation properties of the channel may now be used for improving the estimation accuracy in case the frequency separation is not sufficiently small.

An alternative and perhaps better approach when the frequency separation in FDD is too large is to estimate the state of the forward channel based on physical parameters such as DOA that may be assumed to be

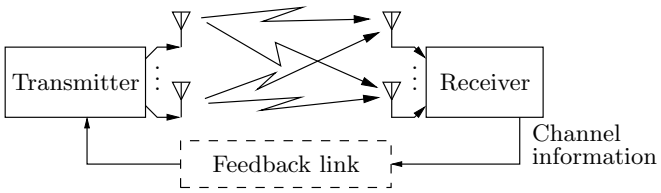

Figure 1.5: The receiver informs the transmitter about the channel via a dedicated feedback link.

invariant to the carrier frequency. Another completely different strategy which is suitable for FDD systems is to equip the system with a dedicated wireless feedback link that conveys estimates of channel parameters from the receiver to the transmitter [GP94a, HP98]. The setup is illustrated in Figure 1.5.

There are however problems also with a feedback approach. Obviously, it takes time to transport the channel estimates over the feedback link. In the mean time, the time-varying channel may have changed so much that the feedback output has become outdated. To avoid such difficulties, the feedback delay must be small in comparison with the coherence time of the channel. In typical land-mobile applications, this means that the channel information at the transmitter needs to be updated at a high rate.

The requirement of a high update rate and the need for a spectrally efficient system in which not too much bandwidth is wasted on the feedback link unfortunately represent two conflicting goals. One way to obtain a reasonable comprise between these two goals is to limit the data rate necessary to maintain a high update rate by quantizing the channel parameters heavily before they are fed back to the transmit side. A drawback with this solution is of course that quantization errors now plague the channel information at the transmitter. Note also that the communication over the feedback link potentially suffers from bit-errors due to the non-perfect nature of the wireless feedback channel. To make matters worse, the requirement on short feedback delay typically prevents the use of advanced channel coding in the feedback link. The error rate can consequently be quite high, adding another significant source of errors. Thus, estimation noise at the receiver, feedback delay, coarse quantization and bit-errors introduced by the feedback channel are all likely to contribute in degrading the quality of the channel information

at the transmitter. Despite these shortcomings, providing channel feedback is often worthwhile as evidenced by the use of a feedback link in the closed-loop transmit diversity mode of the WCDMA system [3GP02b].

1.4 Imperfect Channel Knowledge

Most transmission methods and performance analyses for multiple transmit antennas are developed under either of the two extreme assumptions of perfect channel knowledge or no channel knowledge at the transmitter. Classical beamforming by means of a spatial filter matched to the channel falls in the former category. So does the transmission method and capacity analysis for MIMO systems in [RC98]. Basically the entire field of space-time coding belongs to the latter category where channel knowledge at the transmitter is totally excluded from the development. The well-known MIMO channel capacity works in [Tel95, FG98] also focus on the same extreme case.

However, as should be clear from the previous section, it is often possible to obtain channel information at the transmitter. The problem is that it is typically far from being perfect. Thus, in practice, neither of the two extreme views are correct. This issue has to some extent been addressed in [Wit95] where beamforming weights are determined in such a manner so as to reduce the devastating impact of defective channel information. In another work [HP98], a method for determining beamforming weights based on heavily quantized channel feedback is proposed.

Information theoretic investigations of how the quality of the available channel information influences the properties of optimal signaling show that, if the quality is sufficiently low, conventional beamforming is not optimal in a mutual information sense [NLTW98]. A later and related work gives an analytical expression on the threshold where beamforming is no longer optimal in terms of channel capacity [VM01].

The reason why beamforming performs badly in situations of low channel knowledge quality is because of the one-dimensional nature of the resulting transmission – the emitted energy is focused in only one direction determined by defective channel information. When the quality of the channel information is low, such a single beam transmission scheme often emits energy in the wrong direction. In the extreme case of no channel knowledge, it should be clear that a better strategy is to spread the energy uniformly in all directions. This is what conventional space-time codes aim for.

One way to take channel knowledge into account is to use a transmit weighting matrix for adapting the output of a fixed conventional space-time encoder. The transmit weighting is determined solely from the available channel information while the output of the fixed space-time encoder depends only on the data to be communicated. In contrast to beamforming, such a transmission structure allows energy to be emitted in several directions at the same time, which is necessary if the channel state information is poor. Concentrating all the emitted energy in a single direction is also possible if the circumstances should require it. Separate space-time coding and transmit weighting in this manner represents an appealing and flexible transmission structure. The structure is even more interesting in view of the fact that under certain conditions it is optimal in the sense of preserving the capacity of the MIMO system [SJ03].

By using OSTB codes in the fixed space-time encoder, simple and efficient transmit weighting schemes were developed in [JO99, JSO00, JSO02a]. These transmission methods/structures will in the present work be referred to as *weighted OSTBC*, where OSTBC stands for orthogonal space-time block coding. The same weighted OSTBC transmission schemes have later been proposed and investigated in [ZG02a, ZG02b] for a special case of the setup considered in [JO99, JSO00, JSO02a]. Weighted OSTBC may also be extended to handle quantized channel information obtained from a non-ideal feedback link [JS00, JS01]. An analysis on the diversity order potentially provided by separate OSTB coding and transmit weighting in case of quantized feedback is found in [LGSW02].

Another way of utilizing available channel knowledge is to completely avoid the use of conventional space-time codes and instead consider the design of entirely new space-time codes that are allowed to depend on the channel information. Such an approach has been explored in [JSO02b] for the design of *unstructured space-time block codes* which exploit quantized channel feedback. Compared to the above transmit weighting methods, the unstructured nature of the code gives more degrees of freedom in determining appropriate signals to transmit. Consequently, the performance increases but at the expense of substantially higher decoding complexity.

1.5 Outline and Contributions

This thesis investigates performance limits and develops transmission methods and code designs for a flat fading MIMO wireless communication link where the transmitter has access to possibly imperfect channel side information while the receiver knows the channel perfectly. An outline of the remaining chapters in the thesis is presented below, where also some of the contributions are mentioned.

Chapter 2

In this chapter, an expression for the channel capacity of the system under study is presented and subsequently analyzed. The capacity expression applies to a scenario in which the channel and the channel side information may be modeled as jointly stationary and ergodic random processes, subject to certain additional assumptions. A block fading scenario where these processes are piecewise constant in time is shown to also be handled by the same capacity expression. The corresponding block fading system model forms the basis of the models used in later chapters.

There are several contributions in the chapter in addition to presenting an expression for the channel capacity. It is pointed out that the capacity expression, originally derived for a scenario where power control is allowed, reduces to previously used expected mutual information based performance measures, if power control is prohibited. This is important since it establishes the validity of these widely used performance measures. Another important result is that separate space-time coding and transmit weighting is a capacity achieving transmitter structure. Weighted OSTBC is a simple special case of this structure and shown to be optimal in the case of two transmit antennas and one receive antenna. The latter result provides motivation for the low-complexity weighted OSTBC transmission scheme proposed in later chapters. The capacity expression turns out to often be difficult to evaluate. Procedures for numerically computing the capacity are therefore considered and a theory concerned with symmetrical properties of quantized channel feedback is introduced. Numerical examples illustrate the capacity gains due to channel side information at the transmitter in a spatially uncorrelated Rayleigh fading scenario.

Parts of this material have been published as

> M. Skoglund and G. Jöngren. On the capacity of a multiple-antenna communication link with channel side information. *IEEE Journal*

on Selected Areas in Communications, 21(3):395–405, April 2003.

This paper contains, among other things, the proof of the capacity formula, which the development in the chapter is based on. Even though the proof is important we have chosen to omit it in the development to follow and instead focus on the *consequences* of the capacity formula.

Chapter 3

In this chapter, the system model used for the remainder of the thesis is introduced. Like in parts of the previous chapter, a block fading scenario is considered where the channel and the channel side information are modeled as constant during a block of samples and then vary from one block to another. An important difference compared with our earlier information theoretic investigations is that the channel coding is henceforth assumed to be performed for one fading block at a time, i.e., the transmitted codewords do not extend over possibly different channel realizations.

In addition to a system model, three different classes of space-time block codes that can take channel side information into account are described for later reference. These three classes, mentioned in order of increasingly restrictive structure, correspond to unstructured codes, linear dispersive codes and weighted OSTBC. Whereas unstructured and linear dispersive codes are previously known code types, weighted OSTBC was originally proposed in the works that form the basis of this thesis. Hence, the weighted OSTBC structure represents an important contribution.

Each code is represented by a set of codeword matrices. The three code classes differ in the kind of structural constraints that are imposed on the matrices. The structure of weighted OSTBC is so restrictive that it is perhaps better labeled as a transmission structure rather than a code structure. Both notions will be used interchangeably wherever appropriate. However, thinking in terms of a code makes weighted OSTBC, together with the other structures, fit naturally within a common framework of channel side information dependent space-time block codes. The framework serves to emphasize that the code design (or transmission) procedures developed in subsequent chapters are easily modified to handle the construction of any of these code classes.

Chapter 4

This chapter deals with code design in the presence of possibly imperfect channel side information. The non-ideal nature of the latter is modeled by assuming that the channel conditioned on the side information is complex Gaussian [Kay93, p. 507] distributed. A major contribution is the development of a new performance criterion for space-time codes that takes channel knowledge at the transmitter into account. The performance criterion is based on an upper bound on the pairwise codeword error probability conditioned on the available side information.

Another major contribution is the use and analysis of the performance criterion for code design. An outline of code design procedures for all of the three code classes is given. Particular attention is thereafter directed to our proposed weighted OSTBC structure and how to reduce the computational complexity when designing the transmit weighting matrix. It turns out that the resulting transmission scheme can be seen as a seamless combination of conventional OSTB coding and beamforming, where the quality of the channel side information determines whether the transmitted output is more like that of the former or latter method. Numerical results for a spatially uncorrelated Rayleigh fading scenario illustrate that significant gains compared with both beamforming and OSTBC are possible. In particular, the scheme provides robustness against impairments in the channel side information in contrast to conventional beamforming which suffers severely from such errors.

The parts about the pairwise error probability and weighted OSTBC have been published as

> G. Jöngren, M. Skoglund, and B. Ottersten. Combining beamforming and orthogonal space-time block coding. *IEEE Transactions on Information Theory*, 48(3):611–627, March 2002.

and have also appeared in

> G. Jöngren and B. Ottersten. Combining transmit antenna weights and orthogonal space-time block codes by utilizing side information. In *Proc. 33th Asilomar Conference on Signals, Systems and Computers*, October 1999.
>
> G. Jöngren, M. Skoglund, and B. Ottersten. Combining transmit beamforming and orthogonal space-time block codes by utilizing side information. In *Proc. First IEEE Sensor Array and Multichannel Signal Processing Workshop*, March 2000.

Chapter 5

In this chapter, the use of weighted OSTBC based on quantized channel side information is considered. The channel side information is obtained from the receiver via a dedicated feedback link. Quantization errors, feedback delay and feedback channel bit-errors are all assumed to plague the side information. Techniques related to the field of channel optimized vector quantization (COVQ) [FV91] are used for determining how the channel information can be quantized into a small number of bits while taking the error sources into account. The bits are conveyed to the transmitter where they are decoded into an estimate of the current channel realization.

Two different ways of measuring the error between the true channel and the channel estimate are used in the design of the quantizer. In the first approach, the error is defined based on the differences between the true and estimated channel coefficients, similarly to as in standard minimum mean-square error (MMSE) estimation. Such an error measure is easy to analyze but requires a rather large number of bits to perform well. This motivates the second more efficient approach of separately measuring relative phase and amplitude errors.

The two types of feedback links can be used with any of the three code classes or with other transmission schemes. The presentation in the chapter is however entirely focused on the use of weighted OSTBC. The transmit weight design procedure described in the preceding chapter is directly applicable if the first feedback link type is used. For the case of the second feedback link type, the design procedure needs to be modified. A heuristic method for doing this is presented. Simulation results show that the resulting transmission scheme compares favorably with both beamforming as well as conventional OSTB coding. In particular, robustness against quantization errors, feedback delay and bit-errors is achieved.

Parts of this chapter have been published in

> G. Jöngren and M. Skoglund. Utilizing quantized feedback information in orthogonal space-time block coding. In *Proceedings IEEE Global Telecommunications Conference*, November 2000.
>
> G. Jöngren and M. Skoglund. Improving orthogonal space-time block codes by utilizing quantized feedback information. In *Proceedings International Symposium on Information Theory*, June 2001.

and the whole chapter, with some minor modifications, has been submitted as

> G. Jöngren and M. Skoglund. Quantized feedback information in orthogonal space-time block coding. Submitted to IEEE Transactions on Information Theory.

Chapter 6

This chapter presents another approach to tackle the problem of designing space-time codes that take advantage of quantized channel side information. Instead of relying on heuristic modifications of a design procedure originally derived for non-quantized side information, a new performance criterion specifically tailored to quantized feedback is derived. The obtained performance criterion turns out to be quite similar in structure to the previous criterion. Hence, it can be used together with the code design procedures described in earlier chapters.

To illustrate the use of the new performance criterion, the design of unstructured codes, linear dispersive codes and weighted OSTBC is considered, with the primary focus on the first type of code. Attention is now limited to a scenario in which the feedback link does not suffer from bit-errors and where the channel information is quantized using a specific feedback scheme known as partial phase combining [HP98]. However, the design methods we propose are easily applicable to other types of feedback schemes and can be generalized in a straightforward manner to deal with the effects of feedback channel bit-errors.

Properties that optimal unstructured codes must possess are derived and tables of code design results are provided. In particular, an example of an unstructured four codeword code that outperforms, even without any channel knowledge, the well-known Alamouti OSTB code is provided.

Some linear dispersive codes are also constructed. An interesting observation is that the resulting codes are essentially equivalent to OSTB codes, in the case of no channel knowledge. In addition, a design procedure for weighted OSTBC is implemented.

The concept of symmetric feedback again turns out to be useful for reducing computational complexity. Simulation results show that the constructed codes outperform both OSTB coding and beamforming even in the extreme cases of no or perfect channel knowledge.

Contributions in the chapter include the development of the performance criterion, the code design methods and the subsequent analysis. Some parts of this material have been published in

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*, October 2002.

while the parts pertaining to unstructured codes have been submitted as

G. Jöngren, M. Skoglund, and B. Ottersten. Design of channel estimate dependent space-time block codes. Submitted to IEEE Transactions on Communications.

1.6 Future Work

The work in this thesis leave several issues open for future research. Both the information theoretic analysis and the code design methods can be extended in a number of ways. Some extensions are relatively straightforward while other require substantial effort.

The information theoretic investigations in Chapter 2 are based on a single capacity expression that is more versatile than what the present study may indicate. It is for example possible to broaden the scope of the work to include more realistic fading scenarios and/or study the effects of channel side information impairments other than quantization errors. Both numerical and analytical results would be useful in such studies for providing further insight into the problem. Feedback delay may be modeled as in [NLTW98] and the impact on performance examined. Bit-errors introduced by a noisy feedback channel is another area that is interesting to investigate. The gains provided by the use of power control are also worthwhile to examine in more detail. A potential problem with these extensions is however that they typically further complicate numerical evaluation of the capacity expression. Thus there is a need for more efficient optimization procedures than those presented herein. Another area to explore is the use of the capacity expression as a criterion for the design of codes. Similar techniques as in [HH02b] can then be used for obtaining encoder structures that maximize the capacity expression.

The work concerning methods for code design leaves plenty of topics for future research. Perhaps the most obvious extension is to develop design methods specifically for a frequency-selective channel. One way to handle this essentially within the present framework is to use OFDM for converting a frequency-selective channel into a set of parallel flat fading MIMO channels. The methods developed herein can then

be applied separately for each MIMO channel. However, for maximum performance, techniques for jointly dealing with all the parallel channels need to be developed. An alternative approach which avoids OFDM is to use single-carrier transmission and instead more directly deal with a non-flat channel in both the performance criterion as well as the design procedures, similarly to as in [GS03].

The code design methods all rely on a power constraint that corresponds to a scenario without power control. It is however possible to include power control into the design procedure by relaxing this constraint, albeit at the expense of increased computational complexity. In scenarios with quantized side information, the design is performed off-line and the additional complexity may be tolerable. Power control can even be combined with variable rate coding so that the data rate is adapted to the channel conditions via the available channel side information.

In this thesis, the code design is conducted separately from the design of the feedback link. However, since they both depend on each other, the code and the feedback link should be designed jointly for maximum performance. An outline on how to perform such a joint design is given in our work in [JSO02b]. Numerical procedures remain to be implemented and the properties of the resulting codes and feedback links need to be further investigated.

Another area which receives limited attention in this thesis is that of multiple-access and how to mitigate the total interference in a wireless network. In the literature, there are methods for performing global optimization of beamforming weights so that the coupling between different communication links is kept as small as possible while meeting certain quality of service constraints [RFLT98, BO02]. Similar techniques may be possible to apply for transmitter adaption based on the more general weighted OSTBC structure. By modeling the interference from other users as Gaussian with some spatial color, the performance criteria developed in the present work can easily be extended to take also this kind of disturbance into account. It is however still an open question if it is possible to develop sufficiently efficient numerical procedures for jointly determining the transmit beamforming weights of all users in the system.

Finally, the methods need to be evaluated under more realistic conditions. Channel estimation errors also at the receiver should for example be considered. The impact of a continuously time-varying channel (as opposed to the present block fading assumption) should also be examined. Preliminary investigations indicate that the methods can be made to cope quite well with such common deviations from the idealized system

model considered herein. Nevertheless, much further work is required to fully investigate all aspects of this issue.

Chapter 2

Capacity Results

This chapter considers a MIMO wireless communication system with possibly imperfect channel side information at the transmitter and investigates performance limits from an information theoretic perspective. The channel and the channel side information are modeled as jointly stationary and ergodic random processes. A formula for the channel capacity, valid under certain additional assumptions, is given and used for gaining useful insights about the problem by applying it to different scenarios. The proof of the capacity formula is not included herein, but can be found in our work in [SJ03].

It turns out that separate space-time coding and transmit weighting is an optimal capacity achieving transmitter structure. In the case of two transmit antennas and one receive antenna, an OSTB encoder followed by a transmit weighting is also shown to achieve capacity. This provides motivation for the weighted OSTBC transmission scheme developed in later chapters. Attention is in particular devoted to a scenario in which the side information consists of quantized channel information obtained from the receiver via a feedback link. This setup is closely related to the closed-loop transmit diversity mode in the WCDMA system, making the corresponding capacity results interesting from a practical point of view. Numerical examples for a spatially uncorrelated flat Rayleigh fading situation illustrate that the use of channel knowledge at the transmitter results in considerable gains compared with when channel knowledge is not utilized.

2.1 Introduction

Information theoretic studies show the high data rates that wireless MIMO systems may offer [Tel95, FG98, RC98, RJ99]. The recent development of space-time codes [GFBK99, TSC98, TJC99] provides the means for exploiting the potential of the MIMO channel in practice.

In the literature, different capacity measures have been used depending on the scenario under consideration. The classical MIMO capacity formula for a scenario in which the channel fades in an ergodic manner is given in [Tel95]. With a slow block-wise channel fading scenario in mind, the concept of outage capacity is used in [FG98] to illustrate the extraordinarily high data rates that are possible when both the transmitter and the receiver are equipped with antenna arrays. Both these works rely on the assumption of perfect channel state information at the receiver while the transmitter does not know the channel. In contrast, a non-coherent detection scenario is considered in [MH99], where the capacity is derived under the assumption that neither the transmitter nor the receiver know the channel. The case of perfect channel knowledge at both the transmitter and the receiver has also been treated. Assuming a constant channel, the corresponding capacity is investigated in [RC98]. In [BCT01], the focus is instead on block-wise channel fading and outage capacity as a performance measure.

The works mentioned so far all take on the extreme view that the transmitter either knows the channel perfectly or not at all. However, as discussed in Section 1.3, channel side information at the transmitter, if available, is typically far from being perfect in practice. Even if the side information about the channel is not perfect, it may still help in increasing the performance of the system. This provides motivation for the work presented in this chapter that introduces and analyzes a capacity formula that is applicable also to the more realistic scenario of a transmitter having access to imperfect channel state information.

Most of the presentation is devoted to the case when the output power is allowed to fluctuate with the channel side information, corresponding to a system with power control. However, scenarios without power control are also treated to some extent. The capacity formula is based on the assumption of an ergodic flat fading scenario in which the channel side information at the transmitter may be non-perfect while the receiver knows both the channel as well as the channel side information available at the transmitter. The assumptions about the receiver cover situations when the receiver can estimate the channel parameters accurately and

uses a feedback link, with a corresponding ideal feedback channel that does not introduce bit-errors, to convey heavily quantized versions of the channel parameters to the transmitter. When the receiver's knowledge is more restrictive, our capacity results represents an upper bound on the achievable performance. Although the focus is on flat fading scenarios, a frequency-selective channel may be handled within the present framework by using multicarrier techniques such as OFDM [WE71, Cim85, Bin90] to convert the channel into a set of parallel flat fading channels.

Previous information theoretic studies concerned with communication links with possibly imperfect channel side information and multiple transmit antennas are surprisingly scarce.

In [NLTW98], a system with one receive antenna in a Rayleigh fading scenario is considered. The channel side information at the transmitter and the channel are assumed to be jointly Gaussian distributed. Mutual information averaged over the ensemble of channel and side information realizations is used as performance measure. It is *assumed* that this measure gives the maximum rate of reliable communication but it is not clear whether it actually corresponds to the capacity of some channel. This is in contrast to the present work which establishes an expression for the true channel capacity under the assumption of a more general fading scenario. As discussed later in this chapter, when the scenario is specialized into the one used in [NLTW98], our capacity expression reduces to the same expected mutual information based performance measure as in that paper. Consequently, the performance measure in [NLTW98] constitutes a channel capacity and is hence well motivated.

Another work dealing with wireless MISO systems and channel side information is presented in [VM00]. In this work, so-called mean or covariance feedback is considered, referring to the fact that the second order statistics of the channel now play the role of the side information at the transmitter. With an underlying assumption of Gaussian channel fading, this in turn determines the stationary marginal distribution of the channel fading. Based on the classical MIMO capacity expression in [Tel95], optimal transmission strategies are derived. The use of the classical MIMO capacity expression implies a scenario in which the side information is static and the channel is time-varying in an ergodic fashion. In other words, the transmitted signals should experience all possible channel realizations while the side information remains the same. This may be a reasonable assumption in scenarios were the side information is updated very slowly compared with the coherence time of the channel fading. However, it is not an accurate model for the important case when

a feedback link is used for real-time tracking of channel variations. Our work, on the other hand, includes both scenarios as special cases since the side information is allowed to be time-varying.

The capacity of a communication link with possibly imperfect channel side information has been studied in other previous works as well but then with the focus on SISO systems. For example, in [CS99], the capacity of a SISO link was derived for the case when there is channel side information at the transmitter while the receiver knows a quantity corresponding to the instantaneous SNR. The channel gain was assumed to be positive and assumed to fade in an ergodic fashion. A class of general discrete systems was also investigated. In [DN02], a related multiple-access scenario was considered. Other examples of works dealing with SISO links and channel side information, perfect as well as non-perfect, include [Sha58, Sal92, GV97, Vis99].

The main contribution of the present work is the development and study of an expression for the capacity of a MIMO system in which the transmitter has access to channel side information. A result of particular importance is the conclusion that separate space-time coding and transmit weighting is optimal in the sense that it is a capacity achieving encoder structure. In this structure, the transmit weighting depends only on the side information while the space-time encoder uses a fixed codebook which is independent of the side information. It is also shown that for the case of two transmit and one receiver antenna, such a capacity achieving structure can be implemented using an OSTB code, giving rise to the previously mentioned weighted OSTBC transmission structure. This optimality in a capacity sense provides motivation for the weighted OSTBC transmission scheme proposed in later chapters. Numerical evaluation of the capacity formula is in the general case non-trivial since it involves solving a difficult optimization problem. Therefore, optimization procedures are discussed and a theory about so-called symmetric feedback is developed. By utilizing the symmetric feedback theory, computing the capacity becomes feasible in some cases when quantized side information is available. Quantized side information typically arises in situations where a feedback link is used for providing channel information to the transmitter. Assuming spatially uncorrelated flat Rayleigh fading and one receive antenna, numerical results on the capacity are presented. Motivation for the setup under consideration is provided by the similarity with the closed-loop transmit diversity mode of the WCDMA system.

The chapter is organized as follows. The system model is described in Section 2.2. An expression for the capacity of the described system

is given in Section 2.3, where also the structure of a capacity achieving encoder is discussed. Furthermore, extensions to block-wise channel fading scenarios are also presented. In Section 2.4, the extreme cases of no and perfect channel knowledge at the transmitter are considered and the capacity expression is shown to then reduce to previously known expressions. An alternative capacity achieving structure is presented in Section 2.5, where it is proved that weighted OSTBC is optimal for the case of two transmit antennas and one receive antenna. Computation of the capacity is treated in Section 2.6, where the symmetric feedback concept is used to significantly simplify the problem. Finally, numerical results on the capacity are given in Section 2.7.

2.2 System Model

Consider the symbol sampled and complex baseband equivalent model of a wireless narrow-band MIMO communication system depicted in Figure 2.1. There are M transmit antennas, N receive antennas and n denotes the integer-valued sample index. The frequency-nonselective fading MIMO channel at discrete time instant n is represented by the $N \times M$ matrix $\boldsymbol{H}(n)$ with complex-valued elements $H_{kl}(n)$, $k = 1, \cdots, N$, $l = 1, \cdots, M$, where $H_{kl}(n)$ represents the channel between the lth transmit and the kth receive antenna.

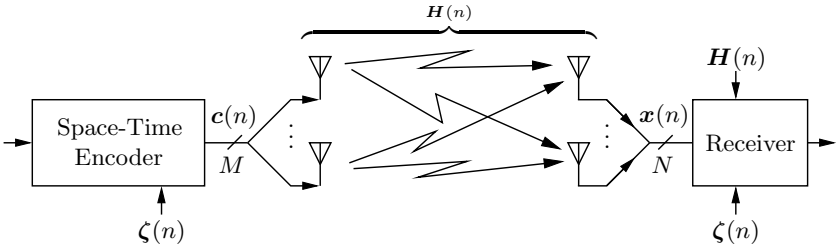

Figure 2.1: System model.

Channel side information is available at the transmitter. The side information comes in the form of a time-varying vector $\boldsymbol{\zeta}(n)$ which is statistically related to $\boldsymbol{H}(n)$ in the sense that $\{\boldsymbol{\zeta}(n)\}$ and $\{\boldsymbol{H}(n)\}$ are assumed to be jointly stationary and ergodic random processes. A space-time encoder maps the message to be transmitted and the side information into M parallel streams of channel symbols by choosing a codeword

to transmit out of a set of codewords in a codebook. The codebook corresponds to the channel code. The channel symbols at time instant n are represented by the output vector

$$\boldsymbol{c}(n) \triangleq \begin{bmatrix} c_1(n) & c_2(n) & \cdots & c_M(n) \end{bmatrix}^{\mathrm{T}} , \tag{2.1}$$

which is assumed to satisfy the power constraint

$$\mathrm{E}[\|\boldsymbol{c}(n)\|^2] = P , \tag{2.2}$$

where $\|\cdot\|$ denotes the usual vector or spectral matrix norm [HJ96, p. 295] and $\mathrm{E}[\cdot]$ is the expectation operator. Here, $(\cdot)^{\mathrm{T}}$ represents the transpose operator, $\triangleq$ denotes "equality per definition" and $c_l(n)$ is a potentially complex-valued channel symbol transmitted on the lth antenna at time instant n.

Adaptation of the transmission to variations in the channel conditions is facilitated by the assumption of an encoder output that is allowed to depend on past and present side information. In other words, $\boldsymbol{c}(n)$ is a function of $\{\boldsymbol{\zeta}(k)\}_{k=-\infty}^{n}$ (causal side information), in addition to its dependence on the message to be communicated. The resulting information carrying signals are transmitted over the wireless channel, picked up by the receiver's antenna array and thereafter filtered and symbol sampled to produce the received signals. The complex baseband equivalent received signals are represented by the vector

$$\boldsymbol{x}(n) \triangleq \begin{bmatrix} x_1(n) & x_2(n) & \cdots & x_N(n) \end{bmatrix}^{\mathrm{T}} , \tag{2.3}$$

where $x_k(n)$ is the output from the kth antenna at time n. Decoding at the receiver is performed under the assumption that both the channel as well as the side information are known at the receiver, as indicated in the figure. Channel state information at the receiver may in practice be obtained by estimating the channel based on, for example, a training sequence. The side information is in some cases implicitly known at the receiver while in other cases it also needs to be estimated. An example of the former case is when the side information $\boldsymbol{\zeta}(n)$ is a deterministic function of the channel $\boldsymbol{H}(n)$, such as in a system where $\boldsymbol{H}(n)$ is quantized at the receiver and then conveyed to the transmitter via a feedback link that does not introduce bit-errors.

The signal output from each receive antenna is a weighted superposition of the transmitted signals, corrupted by additive noise. Hence, the

output from the kth receive antenna can be written as

$$x_k(n) = \sum_{l=1}^{M} H_{kl}(n)c_l(n) + e_k(n)\,, \tag{2.4}$$

where $e_k(n)$ represents the additive noise. By combining (2.1), (2.3), and (2.4), the received signal vector can be written as

$$\boldsymbol{x}(n) = \boldsymbol{H}(n)\boldsymbol{c}(n) + \boldsymbol{e}(n)\,, \tag{2.5}$$

where the noise term

$$\boldsymbol{e}(n) \triangleq \begin{bmatrix} e_1(n) & e_2(n) & \cdots & e_N(n) \end{bmatrix}^{\mathrm{T}}$$

is assumed to be generated from a zero-mean, temporally white, complex Gaussian[1] random process $\{\boldsymbol{e}(n)\}$. The processes $\{\boldsymbol{e}(n)\}$, $\{(\boldsymbol{H}(n), \boldsymbol{\zeta}(n))\}$ and the message to be transmitted are all assumed to be mutually independent. The spatial color of the noise is given by the covariance matrix $\mathrm{E}[\boldsymbol{e}(n)\boldsymbol{e}(n)^*] \triangleq \boldsymbol{R}_{\boldsymbol{ee}}$, where $(\cdot)^*$ denotes the complex conjugate transpose operator. Spatially white noise is usually a good model for thermal noise while colored noise can be used to model interference from other users in a cellular network.

2.2.1 An Additional Assumption

For the above system, an upper bound on the data rate possible for essentially error-free communication (as the processing block length tends to infinity) is derived in our work in [SJ03]. However, as pointed out therein, to proceed and obtain a compact capacity expression, which shows that the upper bound is in fact attainable, makes it necessary to impose an additional constraint regarding the statistical relation between the side information and the channel. Similarly to as in the related SISO work in [CS99], the required constraint is the assumption that $\{\boldsymbol{\zeta}(k)\}_{k=-\infty}^{n-1}$, $\boldsymbol{\zeta}(n)$, $\boldsymbol{H}(n)$ form a Markov chain $\{\boldsymbol{\zeta}(k)\}_{k=-\infty}^{n-1}$ – $\boldsymbol{\zeta}(n)$ – $\boldsymbol{H}(n)$. In other words,

$$p\big(\{\boldsymbol{\zeta}(k)\}_{k=-\infty}^{n-1} | \boldsymbol{\zeta}(n), \boldsymbol{H}(n)\big) = p\big(\{\boldsymbol{\zeta}(k)\}_{k=-\infty}^{n-1} | \boldsymbol{\zeta}(n)\big)\,, \tag{2.6}$$

[1] If desired, the *complex* Gaussian assumptions found throughout this work (and indeed the whole thesis) may all be generalized to the corresponding Gaussian assumptions, without resulting in nothing but straightforward modifications of the development.

where $p(\cdot)$ denotes a probability mass/density function (PMF or PDF, respectively). The above condition means that $\boldsymbol{H}(n)$ gives no further knowledge about the past side information $\{\boldsymbol{\zeta}(k)\}_{k=-\infty}^{n-1}$ when $\boldsymbol{\zeta}(n)$ is known.

2.2.2 Scenarios Satisfying the Assumptions

It should be clear that the major part of the system model described above is reasonable in many scenarios. However, the additional assumption in (2.6) seems rather arcane and does in fact significantly limit the field of application. Despite this, there are a number of scenarios encountered in practice in which the entire system model is applicable. Below, examples of such scenarios are given.

No Channel Knowledge

The extreme case of no channel knowledge at the transmitter can be modeled by assuming that $\{\boldsymbol{\zeta}(n)\}$ and $\{\boldsymbol{H}(n)\}$ are mutually independent. Clearly, (2.6) is satisfied and hence the present work includes this common scenario as a special case.

Perfect Channel Knowledge

For the other extreme of perfect channel knowledge, the side information is equal to the channel, or more precisely, $\boldsymbol{\zeta}(n) = \boldsymbol{\gamma}(\boldsymbol{H}(n))$, where $\boldsymbol{\gamma}(\cdot)$ defines an arbitrary one to one mapping. Again, it is clear that (2.6) is satisfied and that perfect channel knowledge may be handled.

Memoryless Channel Fading

A third example of when (2.6) holds is in scenarios in which the channel process $\{\boldsymbol{H}(n)\}$ can be modeled as memoryless and $\boldsymbol{\zeta}(n)$ is a stochastic or deterministic function of the *current* channel matrix $\boldsymbol{H}(n)$. To see that this means that the additional assumption is satisfied, observe that $\{\boldsymbol{\zeta}(k)\}_{k=-\infty}^{n-1}$ is independent of both $\boldsymbol{H}(n)$ and $\boldsymbol{\zeta}(n)$. It follows that

$$\begin{aligned} p\big(\{\boldsymbol{\zeta}(k)\}_{k=-\infty}^{n-1} | \boldsymbol{\zeta}(n), \boldsymbol{H}(n)\big) &= p\big(\{\boldsymbol{\zeta}(k)\}_{k=-\infty}^{n-1}\big) \\ &= p\big(\{\boldsymbol{\zeta}(k)\}_{k=-\infty}^{n-1} | \boldsymbol{\zeta}(n)\big) \end{aligned}$$

and the condition in (2.6) is thus fulfilled.

In practice, a memoryless channel may model scenarios where the fading is fast or where fast fading has been artificially created through the use of long interleaving or frequency hopping. Side information often consists of channel estimates obtained either in the reverse link of a TDD system or from the receiver via a feedback link when a FDD system is used. Obviously, the side information for both these cases is a function of the channel and, as long as several channel estimates are not combined to form the side information, a function of only one current or past channel.

Feedback delay may be modeled within the present context by assuming that $\boldsymbol{H}(n)$ and $\boldsymbol{\zeta}(n)$ are correlated and jointly Gaussian. If in addition $\{\boldsymbol{H}(n)\}$ and $\{\boldsymbol{\zeta}(n)\}$ are memoryless, $\boldsymbol{\zeta}(n)$ is a stochastic function of $\boldsymbol{H}(n)$ and the condition in (2.6) is hence satisfied. Such a model formed the basis of the information theoretic performance investigations in [NLTW98].

Another commonly encountered feedback situation covered by the system model is when the channel feedback is quantized and potentially suffers from bit-errors introduced by deficiencies in the feedback channel. In Section 2.6.2 we will spend a great deal of effort on showing how to compute the capacity in the case of quantized feedback and an ideal feedback channel. Such feedback scenarios will also be extensively studied from a code design point of view in later chapters of this thesis.

2.3 Capacity of a MIMO System with Side Information

A limit on the maximum data rate possible for essentially error-free communication is given by the information theoretic channel capacity. Assuming all the previous system assumptions are true, the capacity C for the considered MIMO system is, expressed in bits per channel use, given by

$$C = \max_{\substack{\boldsymbol{W}(\cdot) \\ \mathrm{E}[\|\boldsymbol{W}(\boldsymbol{\zeta})\|_{\mathrm{F}}^2]=P}} \mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H}\boldsymbol{W}(\boldsymbol{\zeta})\boldsymbol{W}(\boldsymbol{\zeta})^* \boldsymbol{H}^*)] \,, \qquad (2.7)$$

where $\boldsymbol{W}(\cdot)$ is a deterministic time-invariant $M \times M$ matrix-valued function with elements in the complex number field and where the time index n has been dropped to emphasize that the expectation is with respect to the stationary marginal distribution of the random process $\{(\boldsymbol{H}(n), \boldsymbol{\zeta}(n))\}$. Here, $\log_2(\cdot)$ denotes the base 2 logarithm.

The proof of (2.7) is found in our work in [SJ03] when the noise is spatially white, i.e., $\boldsymbol{R_{ee}}$ is a scaled identity matrix. However, it is straightforward to extend the proof to also include the above case of spatially colored noise.

An alternative representation of (2.7) is obtained by re-parameterizing the problem and instead maximizing with respect to $\boldsymbol{Z}(\boldsymbol{\zeta}) \triangleq \boldsymbol{W}(\boldsymbol{\zeta})\boldsymbol{W}(\boldsymbol{\zeta})^*$, thus obtaining the capacity as

$$C = \max_{\substack{\boldsymbol{Z}(\cdot) \\ \boldsymbol{Z}(\cdot)=\boldsymbol{Z}(\cdot)^* \succeq \boldsymbol{0} \\ \mathrm{E}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\zeta}))]=P}} \mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\zeta})\boldsymbol{H}^*)]\,, \tag{2.8}$$

where $\boldsymbol{A} \succeq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is a positive semi definite matrix and where the trace operator $\mathrm{tr}(\cdot)$ and the relation $\|\boldsymbol{A}\|_{\mathrm{F}}^2 = \mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^*)$ have been used to rewrite the power constraint as

$$\begin{aligned}\mathrm{E}[\|\boldsymbol{W}(\boldsymbol{\zeta})\|_{\mathrm{F}}^2] &= \mathrm{E}[\mathrm{tr}(\boldsymbol{W}(\boldsymbol{\zeta})\boldsymbol{W}(\boldsymbol{\zeta})^*)] \\ &= \mathrm{E}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\zeta}))] = P\,.\end{aligned}$$

Moreover, $\boldsymbol{\zeta}$ has been dropped in the constraint $\boldsymbol{Z}(\cdot) = \boldsymbol{Z}(\cdot)^* \succeq \boldsymbol{0}$ to emphasize that the constraint applies for all $\boldsymbol{\zeta}$. Obvious variations of this notation will be used in the following.

As will be apparent from the development in later sections, the first capacity expression is useful for providing intuition about the structure of the system while the second expression is convenient to use in analytical derivations. Both will be used interchangeably in the following. An important property of (2.8) is that it represents a convex optimization problem, making the task of solving it easier, as will be further discussed in Section 2.6. To see that the problem is convex, utilize the determinant relation (C.5), found in Appendix C, for writing

$$\begin{aligned}&\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\zeta})\boldsymbol{H}^*) \\ &\qquad = \log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\zeta})\boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^*)\,,\end{aligned} \tag{2.9}$$

where $\boldsymbol{R}_{\boldsymbol{ee}}^{1/2}$ is a (possibly Hermitian) matrix square-root of $\boldsymbol{R}_{\boldsymbol{ee}}$ such that $\boldsymbol{R}_{\boldsymbol{ee}} = \boldsymbol{R}_{\boldsymbol{ee}}^{1/2}(\boldsymbol{R}_{\boldsymbol{ee}}^{1/2})^*$. Since $\log\det(\cdot)$ is a concave function over the set of positive semi definite matrices [BV99, VB96], the argument of the determinant in (2.9) is positive semi definite and linear in $\boldsymbol{Z}(\boldsymbol{\zeta})$, and $\mathrm{E}[\cdot]$ is a linear operator, it follows that the criterion function in (2.8) is concave over all functions $\boldsymbol{Z}(\cdot)$ satisfying the constraints. The linearity

of E[·] moreover means that the power constraint is convex. Finally, it is easily verified that also the constraint $\boldsymbol{Z}(\cdot) = \boldsymbol{Z}(\cdot)^* \succeq \mathbf{0}$ is convex [BV99, VB96]. Thus, it can be concluded that the entire optimization problem is convex[2].

Note that in the present work the focus is on a scenario in which the output power is allowed to vary with the realizations of the side information, as evident from the constraint on the average output power in (2.2). This means that power control is part of the scenario. However, power control is not always desirable in practice because of the resulting increased demands on the linearity of transmission amplifiers. Scenarios with no power control would instead be better modeled by the constraint

$$\mathrm{E}[\|\boldsymbol{c}(n)\|^2 | \{\boldsymbol{\zeta}(k)\}_{k=-\infty}^{n}] = P\,, \tag{2.10}$$

where the expectation is taken with respect to the message to be communicated. The reason why (2.10) is more appropriate in non-power-controlled scenarios is because it implies that the average output power does not depend on the side information available at the transmitter[3]. In other words, the side information can no longer control the average output power.

The proof leading to the present capacity expressions needs only minor modifications to handle the power constraint in (2.10). The resulting channel capacity turns out to still be given by (2.7), or (2.8), if the constraint on the average power is replaced with $\|\boldsymbol{W}(\boldsymbol{\zeta})\|_{\mathrm{F}}^2 = \mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\zeta})) = P, \forall \boldsymbol{\zeta}$. Clearly, this is more restrictive than $\mathrm{E}[\|\boldsymbol{W}(\boldsymbol{\zeta})\|_{\mathrm{F}}^2] = \mathrm{E}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\zeta}))] = P$ and the capacity for the non-power-controlled scenario can therefore not be higher than the corresponding capacity when power control is allowed. In a typical case, the capacity is lower.

For the case of one receive antenna and the above scenario which prohibits power control, our capacity formula reduces to the expected mutual information based performance measure in [NLTW98]. As pointed out in Section 2.2.2, the statistical relation between the channel and the side information assumed in that paper corresponds to a special case of the

[2] A minimization problem is convex if the criterion function and all the constraints are convex. Similarly, a maximization problem is convex if the criterion function is concave and all the constrains are convex. The notion of a convex problem in the latter case refers to the fact that the maximization problem is easily turned into an equivalent minimization problem by negating the criterion function, making the originally concave criterion function convex.

[3] Note that the constraint in (2.10) does not fix the *instantaneous* power of the transmitted signals. Hence, the instantaneous power may still vary in time and with the realizations of the message to be communicated.

present system model. Thus, we come to the important conclusion that the measure in [NLTW98] does in fact correspond to the channel capacity for the scenario considered therein.

2.3.1 Structure of a Capacity Achieving Transmitter

Although the capacity expressions give an upper limit on the speed of communication, they do not say how to reach the limit in a practical system. However, as pointed out in our work in [SJ03], from the proof of the capacity formula it is possible to draw some conclusions about the structure of a capacity achieving transmitter. More specifically, our development therein makes it clear that capacity can be achieved using a space-time encoder constrained to have the structure

$$\boldsymbol{c}(n) = \boldsymbol{W}\bar{\boldsymbol{c}}(n)\,, \tag{2.11}$$

where the dependence on the side information is only through $\boldsymbol{W} \triangleq \boldsymbol{W}(\boldsymbol{\zeta}(n))$ and where $\bar{\boldsymbol{c}}(n)$ is produced based on a fixed codebook and depends only on the message to be transmitted and not on the side information. The matrix-valued function $\boldsymbol{W}(\cdot)$ corresponds to the maximizing argument of (2.7) and the codebook can be drawn from a zero-mean complex Gaussian distribution with independent and identically distributed (IID) unit variance components.

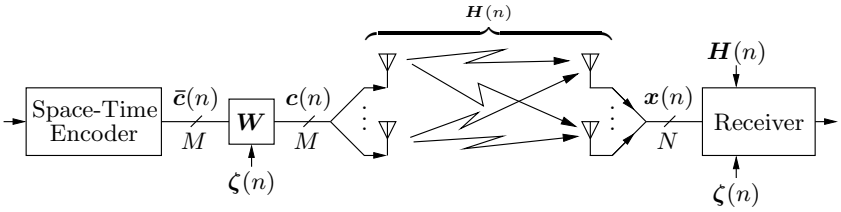

Figure 2.2: Structure of a capacity achieving transmitter.

From (2.11) it is realized that the channel side information dependent space-time encoder can be divided into two separate parts without capacity loss. As illustrated in Figure 2.2, the message to be transmitted may first be mapped into channel vectors $\bar{\boldsymbol{c}}(n)$ by means of an outer space-time encoder that does not take the side information into account. Since the corresponding codebook is independent of the side information, it makes sense to have a codebook that strives to distribute the power isotropically in space. This is ensured by the symmetrical nature of the

previously mentioned IID Gaussian distribution from which the codebook may be drawn. After space-time coding, the output $\bar{\boldsymbol{c}}(n)$ is multiplied/weighted by $\boldsymbol{W}$ in order to adapt the transmission to the channel knowledge available at the transmitter. The weighting performed by $\boldsymbol{W}$ modifies the isotropic power distribution created by the outer space-time encoder so that some "transmit-directions"[4] are allocated more power than others. Since typically several directions are involved, the weighting may be viewed as implementing multi-dimensional beamforming, as opposed to conventional beamforming which transmits only along a single direction. Such multidimensional beamforming is closely related to the standard technique of allocating power along the right singular eigenvectors of the channel by means of water-filling based on the corresponding singular values [Tel95]. We will get back to the similarities with water-filling in Section 2.4.2, where the perfect channel knowledge case is considered.

The described capacity-achieving encoder structure implies that separate space-time coding and transmit weighting is optimal in the sense of preserving channel capacity. In essence, the weighting is used to "color" the spatially white output of the outer space-time encoder, whose codebook is designed without regard to the side information. A related separation result for a SISO real-valued channel with fading and additive white Gaussian noise was derived in [CS99]. It is interesting to note that both these separation results are in many ways similar to the classical theorem in information theory stating that separate source and channel coding is optimal [CT91, p. 215].

Note that, in the case of no channel knowledge at the transmitter, a space-time encoder with a codebook drawn from an IID Gaussian distribution achieves capacity [Tel95]. Further note that the same codebook also achieves capacity in the above optimal structure when there is channel side information. This is appealing from a system design point of view since it suggests that the same conventional space-time code designed for a no channel knowledge scenario is suitable also when channel side information is available, if the code is weighted appropriately. Indeed, the space-time encoder may even be based on classical codes designed for a SISO additive white Gaussian noise channel, since also in this case does an IID Gaussian generated codebook achieve capacity [CT91]. However,

[4] Strictly speaking, the notion of directions here refers to directions in an M dimensional vector space, rather than to physical directions. However, such non-physical directions may in some cases be mapped into their physical counterparts in a straightforward manner.

keep in mind that these are information theoretic based arguments that rely on the usual assumption of an infinite time horizon. Hence, caution should be observed so as to not draw too far-reaching conclusions about the optimality of different code types. In practice, when the processing block length is limited, the performance may vary considerably depending on which type of code is used.

2.3.2 Capacity in a Block Fading Scenario

In practice, it is common to design wireless systems so that the channel parameters can be seen as constant during a block of samples and only change from one block to another. An important advantage of such an approach is that it becomes easier to obtain accurate channel estimates since only a single channel realization needs to be estimated per block, thereby making it possible to improve the accuracy by averaging the noise over an entire block of samples. Block fading scenarios are also common in the wireless MIMO literature, see for example [FG98, TSC98].

Naturally, the popularity of the scenario makes it interesting to compute the corresponding channel capacity. Recall that the capacity formulas in (2.7) and (2.8) are derived under the assumption that $\{\boldsymbol{H}(n)\}$ and $\{\boldsymbol{\zeta}(n)\}$ are jointly stationary and ergodic processes. Clearly, the channel process in a block fading situation is, except in degenerate cases, not stationary as assumed in the system model. It is hence not immediately clear whether the capacity formulas are valid in the present situation of a block fading scenario. In this section we will however show that they are in fact applicable also to such scenarios.

Consider a block fading scenario in which the received signal vector $\boldsymbol{x}_{\mathrm{bf}}(n)$ can be written as

$$\boldsymbol{x}_{\mathrm{bf}}(n) = \boldsymbol{H}_{\mathrm{bf}}(n)\boldsymbol{c}_{\mathrm{bf}}(n) + \boldsymbol{e}_{\mathrm{bf}}(n)\,, \tag{2.12}$$

corresponding to (2.5) in the previous symbol sampled system model but with an additional index 'bf' to indicate a block fading model. The assumptions in Section 2.2 are taken to hold for the above signals, except for the requirement regarding the stationarity and ergodicity of the channel and side information processes. These processes are instead assumed to be piecewise constant so as to model a block fading scenario. Hence, the channel is described by

$$\boldsymbol{H}_{\mathrm{bf}}(n) \triangleq \boldsymbol{H}(k), \quad kL \leq n \leq kL + L - 1\,,$$

where L denotes the block length and $\boldsymbol{H}(k)$ represents the channel during the kth block. Similarly, for the side information,

$$\boldsymbol{\zeta}_{\mathrm{bf}}(n) \triangleq \boldsymbol{\zeta}(k), \quad kL \leq n \leq kL + L - 1\,,$$

where $\boldsymbol{\zeta}(k)$ represents the side information in the kth block. It is here assumed that $\{\boldsymbol{H}(k)\}$ and $\{\boldsymbol{\zeta}(k)\}$ are jointly stationary and ergodic random processes. By stacking $\boldsymbol{x}_{\mathrm{bf}}(n)$ for $n = kL, \cdots, kL + L - 1$ into a vector $\boldsymbol{x}(k)$, it follows from (2.12) that the received data corresponding to the kth block can be written as

$$\boldsymbol{x}(k) = \boldsymbol{\mathcal{H}}(k)\boldsymbol{c}(k) + \boldsymbol{e}(k)\,,$$

where

$$\begin{aligned}
\boldsymbol{x}(k) &\triangleq \begin{bmatrix}\boldsymbol{x}_{\mathrm{bf}}^{\mathrm{T}}(kL) & \boldsymbol{x}_{\mathrm{bf}}^{\mathrm{T}}(kL+1) & \cdots & \boldsymbol{x}_{\mathrm{bf}}^{\mathrm{T}}(kL+L-1)\end{bmatrix}^{\mathrm{T}} \\
\boldsymbol{c}(k) &\triangleq \begin{bmatrix}\boldsymbol{c}_{\mathrm{bf}}^{\mathrm{T}}(kL) & \boldsymbol{c}_{\mathrm{bf}}^{\mathrm{T}}(kL+1) & \cdots & \boldsymbol{c}_{\mathrm{bf}}^{\mathrm{T}}(kL+L-1)\end{bmatrix}^{\mathrm{T}} \\
\boldsymbol{e}(k) &\triangleq \begin{bmatrix}\boldsymbol{e}_{\mathrm{bf}}^{\mathrm{T}}(kL) & \boldsymbol{e}_{\mathrm{bf}}^{\mathrm{T}}(kL+1) & \cdots & \boldsymbol{e}_{\mathrm{bf}}^{\mathrm{T}}(kL+L-1)\end{bmatrix}^{\mathrm{T}} \\
\boldsymbol{\mathcal{H}}(k) &\triangleq \boldsymbol{I}_L \otimes \boldsymbol{H}(k)\,,
\end{aligned}$$

with $\otimes$ denoting the Kronecker product. Recall that $\boldsymbol{c}_{\mathrm{bf}}(n)$ is a function of past and present side information. Consequently, the output vector $\boldsymbol{c}(k)$ in the above block model is a function of $\{\boldsymbol{\zeta}(l)\}_{l=-\infty}^{k}$. Furthermore, since the equivalent channel $\boldsymbol{\mathcal{H}}(k)$ in the block model is related to $\boldsymbol{H}(k)$ via a one to one mapping, $\{\boldsymbol{\mathcal{H}}(k)\}$ and $\{\boldsymbol{\zeta}(k)\}$ are jointly stationary and ergodic. The whiteness of the underlying symbol sampled noise process $\{\boldsymbol{e}_{\mathrm{bf}}(n)\}$ means that also $\{\boldsymbol{e}(k)\}$ is white with an accompanying spatial covariance matrix given by $\boldsymbol{\mathcal{R}}_{\boldsymbol{ee}} \triangleq \mathrm{E}[\boldsymbol{e}(k)\boldsymbol{e}(k)^*] = \boldsymbol{I}_L \otimes \boldsymbol{R}_{\boldsymbol{ee}}$, where $\boldsymbol{R}_{\boldsymbol{ee}}$ denotes the corresponding covariance matrix of $\boldsymbol{e}_{\mathrm{bf}}(n)$.

If k is considered to be the sample index, we see that the above block model corresponds directly to the symbol sampled model of Section 2.2, with $\boldsymbol{\mathcal{H}}(k)$ now playing the role of the channel. From the above, it is clear that all the conditions needed for our capacity formulas to hold are satisfied. Hence, applying (2.8) to the block model shows that the capacity C_{bm}, expressed in bits per block, is given by

$$C_{\mathrm{bm}} = \max_{\substack{\boldsymbol{\mathcal{Z}}(\cdot) \\ \boldsymbol{\mathcal{Z}}(\cdot)=\boldsymbol{\mathcal{Z}}(\cdot)^* \succeq \boldsymbol{0} \\ \mathrm{E}[\mathrm{tr}(\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta}))]=P_{\mathrm{bm}}}} \mathrm{E}[\log_2 \det(\boldsymbol{I}_{LN} + \boldsymbol{\mathcal{R}}_{\boldsymbol{ee}}^{-1}\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})\boldsymbol{\mathcal{H}}^*)]$$

$$= \max_{\substack{\boldsymbol{\mathcal{Z}}(\cdot) \\ \boldsymbol{\mathcal{Z}}(\cdot)=\boldsymbol{\mathcal{Z}}(\cdot)^* \succeq \mathbf{0} \\ \mathrm{E}[\mathrm{tr}(\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta}))]=P_{\mathrm{bm}}}} \mathrm{E}[\log_2 \det(\boldsymbol{I}_{LN} + (\boldsymbol{I}_L \otimes \boldsymbol{R}_{\boldsymbol{ee}}^{-1})(\boldsymbol{I}_L \otimes \boldsymbol{H})\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})(\boldsymbol{I}_L \otimes \boldsymbol{H}^*))]\,, \tag{2.13}$$

where the maximization is now over all $ML \times ML$ matrix-valued functions $\boldsymbol{\mathcal{Z}}(\cdot)$ and where P_{bm} denotes the average output power taken over the *entire* block. Keep in mind that maximization may alternatively be performed in terms of $\boldsymbol{\mathcal{W}}(\cdot)$, where $\boldsymbol{\mathcal{W}}(\boldsymbol{\zeta})\boldsymbol{\mathcal{W}}(\boldsymbol{\zeta})^* \triangleq \boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})$. In Appendix 2.A it is proved that an optimal $\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})$ can be taken to be block diagonal in the sense that

$$\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta}) = \boldsymbol{I}_L \otimes \boldsymbol{Z}(\boldsymbol{\zeta})\,,$$

for some $M \times M$ matrix-valued function $\boldsymbol{Z}(\cdot)$, without loss of optimality. Obviously, since $\boldsymbol{\mathcal{W}}(\boldsymbol{\zeta})\boldsymbol{\mathcal{W}}(\boldsymbol{\zeta})^* = \boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})$, the corresponding weighting may similarly be chosen as $\boldsymbol{\mathcal{W}}(\boldsymbol{\zeta}) = \boldsymbol{I}_L \otimes \boldsymbol{W}(\boldsymbol{\zeta})$, for some $M \times M$ matrix-valued function $\boldsymbol{W}(\cdot)$. By utilizing the block diagonal property, the power constraint can be rewritten as

$$\mathrm{E}[\mathrm{tr}(\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta}))] = L\,\mathrm{E}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\zeta}))] = P_{\mathrm{bm}}\,,$$

while, using the Kronecker relation (C.3) in Appendix C, the determinant in the criterion function may be expressed as

$$\begin{aligned}
&\det(\boldsymbol{I}_{LN} + (\boldsymbol{I}_L \otimes \boldsymbol{R}_{\boldsymbol{ee}}^{-1})(\boldsymbol{I}_L \otimes \boldsymbol{H})(\boldsymbol{I}_L \otimes \boldsymbol{Z}(\boldsymbol{\zeta}))(\boldsymbol{I}_L \otimes \boldsymbol{H}^*)) \\
&\qquad = \det(\boldsymbol{I}_{LN} + \boldsymbol{I}_L \otimes (\boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\zeta})\boldsymbol{H}^*)) \\
&\qquad = \det(\boldsymbol{I}_L \otimes (\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\zeta})\boldsymbol{H}^*)) \\
&\qquad\qquad = \big(\det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\zeta})\boldsymbol{H}^*)\big)^L\,.
\end{aligned}$$

Inserting the above into the expression for C_{bm} gives

$$C_{\mathrm{bm}} = \max_{\substack{\boldsymbol{Z}(\cdot) \\ \boldsymbol{Z}(\cdot)=\boldsymbol{Z}(\cdot)^* \succeq \mathbf{0} \\ L\,\mathrm{E}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\zeta}))]=P_{\mathrm{bm}}}} L\,\mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\zeta})\boldsymbol{H}^*)]\,. \tag{2.14}$$

In order to compare (2.14) with the original capacity formula in (2.8), the block based units of measure currently used first need to be converted into their sample based counterparts. Since C_{bm} is measured in bits per block, it should be divided by L to correspond to bits per *symbol sampled* channel use. Similarly, introduce $P \triangleq P_{\mathrm{bm}}/L$ so as to formulate the power

constraint in terms of the power corresponding to a single sample rather than an entire block of samples. The resulting capacity C_{bf} in the block fading scenario, measured in bits per symbol sampled channel use, can then be expressed as

$$C_{\mathrm{bf}} \triangleq C_{\mathrm{bm}}/L = \max_{\substack{\boldsymbol{Z}(\cdot) \\ \boldsymbol{Z}(\cdot)=\boldsymbol{Z}(\cdot)^* \succeq \boldsymbol{0} \\ \mathrm{E}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\zeta}))]=P}} \mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H} \boldsymbol{Z}(\boldsymbol{\zeta}) \boldsymbol{H}^*)] \,.$$

Clearly, this is the same as (2.8) and it can be concluded that the original capacity formulas are applicable also to a block fading scenario.

Regarding an optimal encoder structure, observe that from Section 2.3.1 it follows that

$$\boldsymbol{c}(k) = \boldsymbol{\mathcal{W}}(\boldsymbol{\zeta}(k)) \bar{\boldsymbol{c}}(k) \,, \tag{2.15}$$

with

$$\bar{\boldsymbol{c}}(k) \triangleq \begin{bmatrix} \bar{\boldsymbol{c}}_{\mathrm{bf}}^{\mathrm{T}}(kL) & \bar{\boldsymbol{c}}_{\mathrm{bf}}^{\mathrm{T}}(kL+1) & \cdots & \bar{\boldsymbol{c}}_{\mathrm{bf}}^{\mathrm{T}}(kL+L-1) \end{bmatrix}^{\mathrm{T}} \,,$$

corresponding to $\bar{\boldsymbol{c}}(\cdot)$ in (2.11), is a capacity achieving structure. Hence, separate space-time coding and transmit weighting on a *block basis* is optimal. Due to the block diagonal structure of an optimal $\boldsymbol{\mathcal{W}}(\cdot)$, the expression in (2.15) can be reformulated to obtain

$$\begin{aligned} \boldsymbol{c}_{\mathrm{bf}}(n) &= \boldsymbol{W}(\boldsymbol{\zeta}(k)) \bar{\boldsymbol{c}}_{\mathrm{bf}}(n), \quad kL \leq n \leq kL+L-1 \\ &= \boldsymbol{W}(\boldsymbol{\zeta}_{\mathrm{bf}}(n)) \bar{\boldsymbol{c}}_{\mathrm{bf}}(n) \,, \end{aligned} \tag{2.16}$$

representing the symbol sampled encoder output for time index n. We therefore come to the important conclusion that separate space-time coding and transmit weighting is optimal also for the symbol sampled block fading model in (2.12). It is interesting to note that the transmit weighting is piecewise constant in time in a similar manner as the channel and the side information. In other words, the transmit weighting is constant for the entire duration of a block and is only updated at the start of a new block. Compared with the optimal structure for a non-block fading scenario, the transmit weighting hence varies more slowly – from block to block instead of from sample to sample.

2.4 Specializing the Capacity Formula

To better understand the our capacity formula, it is instructive to consider the extreme cases of no and perfect channel knowledge at the transmitter. It will in the following be shown that in the former case the

classical MIMO capacity expression is obtained while in the latter case the capacity is given by the well-known water-filling type of power allocation.

2.4.1 No Channel Knowledge

Consider first a scenario in which the transmitter has no knowledge about the channel. As previously mentioned, this may be modeled by assuming that the side information process $\{\boldsymbol{\zeta}(n)\}$ is statistically independent of $\{\boldsymbol{H}(n)\}$.

From the discussion immediately after (2.9), it is realized that $\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{X})$ is concave over the set of positive semi definite matrices $\boldsymbol{X}$. This permits the use of Jensen's inequality [CT91, p. 25] to obtain an upper bound on the expectation in (2.8), i.e.,

$$\begin{aligned}
&\mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\zeta})\boldsymbol{H}^*)] \\
&\qquad = \mathrm{E}_{\boldsymbol{H}}\left[\mathrm{E}_{\boldsymbol{\zeta}}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\zeta})\boldsymbol{H}^*)|\boldsymbol{H}]\right] \\
&\qquad \le \mathrm{E}_{\boldsymbol{H}}\left[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\,\mathrm{E}_{\boldsymbol{\zeta}}[\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\zeta})\boldsymbol{H}^*|\boldsymbol{H}])\right] \\
&\qquad = \mathrm{E}_{\boldsymbol{H}}\left[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\,\mathrm{E}_{\boldsymbol{\zeta}}[\boldsymbol{Z}(\boldsymbol{\zeta})|\boldsymbol{H}]\boldsymbol{H}^*)\right] \\
&\qquad\qquad = \mathrm{E}_{\boldsymbol{H}}\left[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\,\mathrm{E}_{\boldsymbol{\zeta}}[\boldsymbol{Z}(\boldsymbol{\zeta})]\boldsymbol{H}^*)\right], \quad (2.17)
\end{aligned}$$

which holds with equality if $\boldsymbol{Z}(\cdot)$ is a constant function and where the last equality is due to $\{\boldsymbol{\zeta}(n)\}$ and $\{\boldsymbol{H}(n)\}$ being mutually independent. The power constraint in (2.8) can be rewritten as

$$\mathrm{E}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\zeta}))] = \mathrm{tr}(\mathrm{E}_{\boldsymbol{\zeta}}[\boldsymbol{Z}(\boldsymbol{\zeta})]) = P\,. \quad (2.18)$$

By combining (2.17), (2.18) and (2.8), an upper bound on the capacity follows as

$$\begin{aligned}
C &\le \max_{\substack{\boldsymbol{Z}(\cdot) \\ \boldsymbol{Z}(\cdot)=\boldsymbol{Z}(\cdot)^* \succeq \boldsymbol{0} \\ \mathrm{tr}(\mathrm{E}_{\boldsymbol{\zeta}}[\boldsymbol{Z}(\boldsymbol{\zeta})])=P}} \mathrm{E}_{\boldsymbol{H}}\left[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\,\mathrm{E}_{\boldsymbol{\zeta}}[\boldsymbol{Z}(\boldsymbol{\zeta})]\boldsymbol{H}^*)\right] \quad (2.19) \\
&\le \max_{\substack{\boldsymbol{Z} \\ \boldsymbol{Z}=\boldsymbol{Z}^* \succeq \boldsymbol{0} \\ \mathrm{tr}(\boldsymbol{Z})=P}} \mathrm{E}_{\boldsymbol{H}}\left[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}\boldsymbol{H}^*)\right]. \quad (2.20)
\end{aligned}$$

However, the above inequalities can be replaced with equalities. To see this, note that the upper bound in (2.17) is attainable if $\boldsymbol{Z}(\cdot)$ is a constant function. Hence, the inequality in (2.19) holds with equality. Let the

$M \times M$ matrix $\boldsymbol{Z}$ represent the value of the constant function, i.e., $\boldsymbol{Z}(\boldsymbol{\zeta}) = \boldsymbol{Z}$. Since then $\mathrm{E}_{\boldsymbol{\zeta}}[\boldsymbol{Z}(\boldsymbol{\zeta})] = \boldsymbol{Z}$, it is seen that (2.19) reduces to (2.20) which means that also the second inequality holds with equality. Thus, the capacity in the no channel knowledge case is given by

$$C = \max_{\substack{\boldsymbol{Z} \\ \boldsymbol{Z}=\boldsymbol{Z}^* \succeq \boldsymbol{0} \\ \mathrm{tr}(\boldsymbol{Z})=P}} \mathrm{E}_{\boldsymbol{H}}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H} \boldsymbol{Z} \boldsymbol{H}^*)] \,. \tag{2.21}$$

Consider the present no channel knowledge case and further specialize by in addition assume spatially uncorrelated Rayleigh fading, meaning that the elements of $\boldsymbol{H}$ are zero-mean IID complex Gaussian random variables. Moreover, assume that the noise is spatially white so that $\boldsymbol{R}_{\boldsymbol{ee}} = \sigma^2 \boldsymbol{I}_N$, where σ^2 denotes the noise variance. The expression in (2.21) then corresponds to the same optimization problem as in [Tel95], where the optimal solution was shown to be $\boldsymbol{Z} = P/M \boldsymbol{I}_N$. Since $\boldsymbol{Z}(\boldsymbol{\zeta}) = \boldsymbol{W}(\boldsymbol{\zeta}) \boldsymbol{W}(\boldsymbol{\zeta})^*$, the corresponding optimal transmit weighting $\boldsymbol{W}$ in (2.11) can be taken as a constant scaled unitary matrix. This choice of $\boldsymbol{W}$ makes sense for the present case of no channel knowledge since a unitary $\boldsymbol{W}$, together with a zero-mean IID Gaussian generated codebook, distributes the power isotropically in space and hence does not favor one direction over another. After inserting the optimum $\boldsymbol{Z} = P/M \boldsymbol{I}_N$ in (2.21) the classical MIMO capacity formula [Tel95]

$$C = \mathrm{E}_{\boldsymbol{H}}[\log_2 \det(\boldsymbol{I}_N + \frac{P}{M\sigma^2} \boldsymbol{H} \boldsymbol{H}^*)] \tag{2.22}$$

is obtained as a special case of our more general capacity expression.

2.4.2 Perfect Channel Knowledge

The second case we consider is the other extreme of perfect channel knowledge at the transmitter. This is modeled by assuming that the side information equals the channel, or more precisely, that there is a one to one deterministic mapping between the channel $\boldsymbol{H}(n)$ and the side information $\boldsymbol{\zeta}(n)$. In the following, let $\boldsymbol{\gamma}(\cdot)$ represent such a deterministic mapping and hence assume that $\boldsymbol{\zeta}(n) = \boldsymbol{\gamma}(\boldsymbol{H}(n))$.

For the problem at hand, the capacity formula in (2.8) can be rewritten as

$$C = \max_{\substack{\boldsymbol{Z}(\cdot) \\ \boldsymbol{Z}(\cdot)=\boldsymbol{Z}(\cdot)^* \succeq \boldsymbol{0} \\ \mathrm{E}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H})))]=P}} \mathrm{E}_{\boldsymbol{H}}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H} \boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H})) \boldsymbol{H}^*)] \,. \tag{2.23}$$

To see how perfect channel knowledge affects the capacity, utilize the determinant relation (C.5) in Appendix C for writing the determinant in (2.23) as

$$\begin{aligned}\det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H}))\boldsymbol{H}^*) &= \det(\boldsymbol{I}_M + \boldsymbol{H}^*\boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H})))\\ &= \det(\boldsymbol{I}_M + \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H})))\\ &= \det(\boldsymbol{I}_M + \boldsymbol{\Lambda}\boldsymbol{U}^*\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H}))\boldsymbol{U})\,, \quad (2.24)\end{aligned}$$

where $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$ denotes the eigenvalue decomposition (EVD) of $\boldsymbol{H}^*\boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}$. In the EVD, $\boldsymbol{U}$ is an $M \times M$ unitary matrix containing the eigenvectors and $\boldsymbol{\Lambda}$ is an $M \times M$ diagonal matrix with ordered elements $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq \lambda_{r+1} = \cdots = \lambda_M = 0$ corresponding to the eigenvalues. Here, $r \triangleq \min\{M, N\}$ denotes the maximum rank of $\boldsymbol{H}^*\boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}$. Since $\boldsymbol{U}$ is unitary, the power constraint in (2.23) may be written as

$$\begin{aligned}\mathrm{E}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H})))] &= \mathrm{E}[\mathrm{tr}(\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H})))]\\ &= \mathrm{E}[\mathrm{tr}(\boldsymbol{U}^*\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H}))\boldsymbol{U})] = P\,. \quad (2.25)\end{aligned}$$

Due to the one to one mapping $\boldsymbol{\gamma}(\cdot)$, $\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H}))$ is an arbitrary deterministic function of the channel. From this and (2.24), (2.25), it follows that the capacity formula may be written in terms of the positive semi definite matrix-valued deterministic function $\tilde{\boldsymbol{Z}}(\boldsymbol{H}) \triangleq \boldsymbol{U}^*\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H}))\boldsymbol{U}$. The resulting capacity formula is

$$C = \max_{\substack{\tilde{\boldsymbol{Z}}(\cdot)\\ \tilde{\boldsymbol{Z}}(\cdot)=\tilde{\boldsymbol{Z}}(\cdot)^* \succeq \boldsymbol{0}\\ \mathrm{E}_{\boldsymbol{H}}[\mathrm{tr}(\tilde{\boldsymbol{Z}}(\boldsymbol{H}))]=P}} \mathrm{E}_{\boldsymbol{H}}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{\Lambda}(\boldsymbol{H})\tilde{\boldsymbol{Z}}(\boldsymbol{H}))]\,, \quad (2.26)$$

where, with a slight abuse of notation, the eigenvalues' functional dependence on the channel has been made explicit. Applying Hadamard's inequality [HJ96, p. 477] to the determinant shows that $\tilde{\boldsymbol{Z}}(\boldsymbol{H})$ can be taken to be diagonal without loss of optimality. A detailed proof is found in Appendix 2.B. Hence, let

$$\tilde{\boldsymbol{Z}}(\boldsymbol{H}) = \mathrm{diag}(\tilde{Z}_{11}(\boldsymbol{H}), \tilde{Z}_{22}(\boldsymbol{H}), \cdots, \tilde{Z}_{MM}(\boldsymbol{H}))\,, \quad (2.27)$$

where $\mathrm{diag}(x_1, x_2, \cdots)$ denotes a diagonal matrix with elements $x_1, x_2, \cdots$ on the main diagonal, and combine this with (2.26) to obtain

$$C = \max_{\substack{\{\tilde{Z}_{kk}(\cdot)\}_{k=1}^r\\ \tilde{Z}_{kk}(\cdot)\geq 0,\, k=1,\cdots,r\\ \mathrm{E}[\sum_{k=1}^r \tilde{Z}_{kk}(\boldsymbol{H})]=P}} \mathrm{E}_{\boldsymbol{H}}\left[\sum_{k=1}^{r} \log_2(1 + \lambda_k(\boldsymbol{H})\tilde{Z}_{kk}(\boldsymbol{H}))\right]. \quad (2.28)$$

Here, the $M-r$ last $\tilde{Z}_{kk}(\boldsymbol{H})$ have been set to zero without loss of optimality since they do not affect the criterion function. Let

$$f(\{\tilde{Z}_{kk}(\boldsymbol{H})\}_{k=1}^{r}) \triangleq \sum_{k=1}^{r} \log_2(1+\lambda_k(\boldsymbol{H})\tilde{Z}_{kk}(\boldsymbol{H}))$$

and note that the criterion function $\mathrm{E}[f(\{\tilde{Z}_{kk}(\boldsymbol{H})\}_{k=1}^{r})]$ is concave, since $\log_2(1+x)$ is concave and the expectation operator is linear. Necessary and sufficient conditions on $\{\tilde{Z}_{kk}(\cdot)\}_{k=1}^{r}$ to maximize $\mathrm{E}[f(\{\tilde{Z}_{kk}(\boldsymbol{H})\}_{k=1}^{r})]$ can then be expressed in terms of the derivatives

$$\begin{aligned}
&\frac{\partial f(\{\tilde{Z}_{kk}(\boldsymbol{H})\}_{k=1}^{r})}{\partial \tilde{Z}_{kk}} = \tilde{\mu}, \quad \text{all } k \text{ and } \boldsymbol{H} \text{ such that } \tilde{Z}_{kk}(\boldsymbol{H}) > 0 \\
&\frac{\partial f(\{\tilde{Z}_{kk}(\boldsymbol{H})\}_{k=1}^{r})}{\partial \tilde{Z}_{kk}} \leq \tilde{\mu}, \quad \text{all } k \text{ and } \boldsymbol{H} \text{ such that } \tilde{Z}_{kk}(\boldsymbol{H}) = 0\,,
\end{aligned}$$

where $\tilde{\mu}$ is a Lagrange multiplier. This can be shown by extending the proof of Theorem 4.4.1 in [Gal68, p. 87] to cover the present case of optimization with respect to *functions* instead of a finite number of parameters. By differentiating $f(\cdot)$ it is seen that $\{\tilde{Z}_{kk}(\cdot)\}_{k=1}^{r}$ solves the maximization problem if (and only if) they satisfy the conditions

$$\frac{\lambda_k(\boldsymbol{H})}{1+\lambda_k(\boldsymbol{H})\tilde{Z}_{kk}(\boldsymbol{H})} = \mu, \quad \text{all } k \text{ and } \boldsymbol{H} \text{ such that } \tilde{Z}_{kk}(\boldsymbol{H}) > 0 \quad (2.29)$$

$$\frac{\lambda_k(\boldsymbol{H})}{1+\lambda_k(\boldsymbol{H})\tilde{Z}_{kk}(\boldsymbol{H})} \leq \mu, \quad \text{all } k \text{ and } \boldsymbol{H} \text{ such that } \tilde{Z}_{kk}(\boldsymbol{H}) = 0\,. \quad (2.30)$$

Here, $\mu \triangleq \log(2)\tilde{\mu}$ has been introduced for notational convenience. These conditions reduce to the condition stated in [CS99] for the SISO channel case when $r=1$. To obtain a solution, solve for $\tilde{Z}_{kk}(\boldsymbol{H})$ in (2.29) and (2.30), leading to the equivalent conditions

$$\begin{aligned}
&\tilde{Z}_{kk}(\boldsymbol{H}) = 1/\mu - 1/\lambda_k(\boldsymbol{H}), \quad \text{all } k \text{ and } \boldsymbol{H} \text{ such that } \tilde{Z}_{kk}(\boldsymbol{H}) > 0 \\
&\tilde{Z}_{kk}(\boldsymbol{H}) \geq 1/\mu - 1/\lambda_k(\boldsymbol{H}), \quad \text{all } k \text{ and } \boldsymbol{H} \text{ such that } \tilde{Z}_{kk}(\boldsymbol{H}) = 0\,.
\end{aligned}$$

It is now easily verified that

$$\tilde{Z}_{kk}(\boldsymbol{H}) = \max\{1/\mu - 1/\lambda_k(\boldsymbol{H}), 0\}, \quad k = 1, \cdots, r\,, \quad (2.31)$$

with $1/\lambda_k(\boldsymbol{H})$ replaced by ∞ if $\lambda_k(\boldsymbol{H}) = 0$, satisfies both conditions and hence represents the optimal solution to the convex problem in (2.28).

The Lagrange multiplier μ may be determined from the power constraint by solving the equation

$$\sum_{k=1}^{r} \mathrm{E}[\max\{1/\mu - 1/\lambda_k(\boldsymbol{H}), 0\}] = P\,. \tag{2.32}$$

Once μ is determined, a simplified capacity formula is finally obtained by utilizing the expression for $\tilde{Z}_{kk}(\boldsymbol{H})$ to arrive at

$$C = \sum_{k=1}^{r} \mathrm{E}_{\boldsymbol{H}}[\log_2(1 + \lambda_k(\boldsymbol{H})\tilde{Z}_{kk}(\boldsymbol{H}))]\,. \tag{2.33}$$

Interpretations and Comparisons

Because of the relations $\tilde{\boldsymbol{Z}}(\boldsymbol{H}) = \boldsymbol{U}^* \boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H}))\boldsymbol{U}$, $\boldsymbol{Z}(\boldsymbol{\zeta}) = \boldsymbol{W}(\boldsymbol{\zeta})\boldsymbol{W}(\boldsymbol{\zeta})^*$ and the diagonal nature of $\tilde{\boldsymbol{Z}}(\boldsymbol{H})$, it follows from the above that the transmit weighting may be chosen as

$$\begin{aligned}\boldsymbol{W} &= \boldsymbol{U}\,\mathrm{diag}(\tilde{Z}_{11}^{1/2}, \cdots, \tilde{Z}_{rr}^{1/2}, 0, \cdots, 0) \\ &= \sum_{k=1}^{r} \boldsymbol{u}_k \tilde{Z}_{kk}^{1/2}\,,\end{aligned} \tag{2.34}$$

where the dependence on the channel and the side information has been suppressed and where $\boldsymbol{u}_k$ denotes the kth column of $\boldsymbol{U}$. In view of the capacity achieving encoder structure in Section 2.3.1, we see that the transmission is along up to r different orthogonal directions, corresponding to the eigenvectors of $\boldsymbol{H}^* \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H}$, with the output power distributed according to (2.31). In the special case of spatially white noise, $\boldsymbol{R}_{\boldsymbol{ee}}$ is a scaled identity matrix and the eigenvectors of $\boldsymbol{H}^* \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H}$ also correspond to the right singular eigenvectors in the singular value decomposition (SVD) of $\boldsymbol{H}$ [HJ96, p. 414].

Another special case is that of $N = 1$ receive antenna. The channel $\boldsymbol{H}$ then becomes a row vector and hence $\boldsymbol{H}^* \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H} = \boldsymbol{H}^* \boldsymbol{H}/\sigma^2$ is a rank $r = 1$ matrix, where σ^2 denotes the noise variance. Consequently, only its first eigenvalue $\lambda_1(\boldsymbol{H})$ is non-zero and, since $(\boldsymbol{H}^* \boldsymbol{H})\boldsymbol{H}^* = \boldsymbol{H}^*(\boldsymbol{H}\boldsymbol{H}^*) = \|\boldsymbol{H}\|^2 \boldsymbol{H}^*$, that eigenvalue is given by

$$\lambda_1(\boldsymbol{H}) = \|\boldsymbol{H}\|^2/\sigma^2 = \sum_{l=1}^{M} |H_{kl}|^2/\sigma^2\,, \tag{2.35}$$

with $\boldsymbol{u}_1 = \boldsymbol{H}^*/\|\boldsymbol{H}^*\| = \boldsymbol{H}^*/\|\boldsymbol{H}\|$ representing the corresponding (right singular) eigenvector. The remaining eigenvalues $\lambda_k(\boldsymbol{H})$, $k = 2, \cdots, M$ are all seen to be zero, implying that $\tilde{Z}_{kk}(\boldsymbol{H}) = 0$, $k = 2, \cdots, M$. Consequently, the capacity expression in (2.33) now reduces to

$$C = \mathrm{E}_{\boldsymbol{H}}[\log_2(1 + \frac{\|\boldsymbol{H}\|^2}{\sigma^2}\tilde{Z}_{11}(\boldsymbol{H}))]\,, \tag{2.36}$$

where

$$\tilde{Z}_{11}(\boldsymbol{H}) = \max\{1/\mu - \sigma^2/\|\boldsymbol{H}\|^2, 0\}\,, \tag{2.37}$$

with μ determined from the simplified power constraint

$$\mathrm{E}[\max\{1/\mu - \sigma^2/\|\boldsymbol{H}\|^2, 0\}] = P\,. \tag{2.38}$$

From the above and (2.34) it directly follows that the transmit weighting $\boldsymbol{W}$ is of rank one with all power transmitted in a single direction $\boldsymbol{u}_1 = \boldsymbol{H}^*/\|\boldsymbol{H}\|$. This corresponds to beamforming in the direction of the channel, thus establishing the optimality of such beamforming.

The power distribution given by (2.31) is in general of the traditional water-filling type [CT91]. This means that, for each channel realization, more power is allocated on strong eigen-directions, i.e., directions where $\lambda_k(\boldsymbol{H})$ is large, and less on weak ones. Obviously, both the power distribution and the corresponding transmit-directions vary over the ensemble of channel realizations. Achieving capacity by distributing the power along some eigen-directions through the use of water-filling is a classical result, see for example [Tel95, RC98]. However, most previous such MIMO capacity results rely on the assumption that the channel is *constant and non-random*. One notable exception to this is the work in [BCT01] which briefly discusses ergodic capacity and states, with reference to [Tel95], water-filling expressions corresponding to (2.31) - (2.33). However, the given motivation for the water-filling expressions is somewhat misleading since [Tel95] deals with an entirely different, constant channel, scenario. The development in the present section can therefore be seen as providing a more rigorous proof for that water-filling result.

The capacity for a constant channel scenario may be obtained as a special case of the development herein, since a constant channel is clearly included in the system model of the present work. Toward this end, let $\boldsymbol{Z} \triangleq \boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H}))$ and observe that the expectation in (2.23) disappears. The resulting constant channel capacity $C_{\mathrm{c}}(\boldsymbol{H})$ is therefore given by

$$C_{\mathrm{c}}(\boldsymbol{H}) = \max_{\substack{\boldsymbol{Z} \\ \boldsymbol{Z}=\boldsymbol{Z}^* \succeq \boldsymbol{0} \\ \mathrm{tr}(\boldsymbol{Z})=P}} \log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}\boldsymbol{H}^*)\,, \tag{2.39}$$

which, after utilizing (2.31) - (2.33), obviously reduces to

$$C_{\mathrm{c}}(\boldsymbol{H}) = \sum_{k=1}^{r} \log_2(1 + \lambda_k \tilde{Z}_{kk}) \,, \tag{2.40}$$

where $\tilde{Z}_{kk} \triangleq \tilde{Z}_{kk}(\boldsymbol{H})$ is obtained from the water-filling solution (2.31) and μ is determined from the power constraint

$$\sum_{k=1}^{r} \max\{1/\mu - 1/\lambda_k, 0\} = P \,. \tag{2.41}$$

Since the expectation operator has vanished, it is easy to devise an algorithm that solves for μ. Hence, computing the capacity is straightforward.

Using Constant Channel Capacity in Fading Scenarios

It is interesting to further compare a random versus a constant channel and the connection to scenarios with and without power control. For this purpose, consider, as in Section 2.3.2, a block fading scenario in which the channel is constant for a block of samples but fades from one block to another in a random fashion. The expression for the constant channel capacity $C_{\mathrm{c}}(\boldsymbol{H})$ in (2.39) then corresponds to a random variable. If we assume that the blocks are long enough for each realization of $C_{\mathrm{c}}(\boldsymbol{H})$ to be an accurate reflection of the maximum possible data rate in that block, it makes sense to measure the performance in terms of the *average capacity*, mathematically defined as

$$\bar{C}_{\mathrm{c}} \triangleq \mathrm{E}_{\boldsymbol{H}}[C_{\mathrm{c}}(\boldsymbol{H})] = \mathrm{E}_{\boldsymbol{H}} \Big[\max_{\substack{\boldsymbol{Z} \\ \boldsymbol{Z}=\boldsymbol{Z}^* \succeq \boldsymbol{0} \\ \mathrm{tr}(\boldsymbol{Z})=P}} \log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H} \boldsymbol{Z} \boldsymbol{H}^*) \Big] \,. \tag{2.42}$$

It is here implicitly assumed that the average output power (over the ensemble of messages to be communicated) remains constant regardless of the channel realization, i.e., a scenario without power control is considered. The average capacity as a performance measure was used, for example, in [BFAOHY02] to investigate the performance limits of MIMO systems. An appealing feature of the average capacity is that it is straightforward to compute by means of Monte-Carlo integration using conventional water-filling techniques for evaluating (2.40) and (2.41) for each channel realization.

From the above development, it is not yet clear whether averaging the constant channel capacity as in (2.42) gives the capacity for the present block fading scenario. To see that the capacity is indeed obtained, note first that the true channel capacity, assuming the original output power constraint (2.2) and assuming that the channel realizations in the block fading model are generated from a stationary and ergodic process, is given by (2.23). In fact, (2.23) gives the channel capacity regardless of the block length, as proved in Section 2.3.2. Further note that the channel capacity in the corresponding non-power-controlled scenario is possibly lower and is given by (2.23) if the expectation in the power constraint is dropped. In other words,

$$\begin{aligned} C_{\mathrm{pc}} &= \max_{\substack{\boldsymbol{Z}(\cdot) \\ \boldsymbol{Z}(\cdot)=\boldsymbol{Z}(\cdot)^* \succeq \boldsymbol{0} \\ \mathrm{E}_{\boldsymbol{H}}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H})))]=P}} \mathrm{E}_{\boldsymbol{H}}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H} \boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H})) \boldsymbol{H}^*)] \\ &\geq \max_{\substack{\boldsymbol{Z}(\cdot) \\ \boldsymbol{Z}(\cdot)=\boldsymbol{Z}(\cdot)^* \succeq \boldsymbol{0} \\ \mathrm{tr}(\boldsymbol{Z}(\cdot))=P}} \mathrm{E}_{\boldsymbol{H}}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H} \boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H})) \boldsymbol{H}^*)] \\ &= C_{\mathrm{npc}} \,, \end{aligned} \tag{2.43}$$

where C_{pc} and C_{npc} denote the capacity for the scenarios with and without power control, respectively. For the latter, there is no coupling between $\boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H}))$'s corresponding to different outcomes of $\boldsymbol{H}$. Hence, the maximization operator may be moved inside the expectation. After letting $\boldsymbol{Z} \triangleq \boldsymbol{Z}(\boldsymbol{\gamma}(\boldsymbol{H}))$, the result can be written on the form

$$C_{\mathrm{npc}} = \mathrm{E}_{\boldsymbol{H}} \left[\max_{\substack{\boldsymbol{Z} \\ \boldsymbol{Z}=\boldsymbol{Z}^* \succeq \boldsymbol{0} \\ \mathrm{tr}(\boldsymbol{Z})=P}} \log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H} \boldsymbol{Z} \boldsymbol{H}^*) \right]. \tag{2.44}$$

Comparing this with (2.42) shows that $C_{\mathrm{npc}} = \bar{C}_{\mathrm{c}}$ and hence it is concluded that the average capacity indeed equals the true capacity in the case of a non-power-controlled scenario.

Note that, even if power control is considered, the corresponding non-power-controlled scenario is useful since it can be utilized for obtaining an easily computable lower bound on the channel capacity of interest. This is important as the expectations in (2.32) are typically difficult to acquire in closed-form, making it challenging to solve for the needed Lagrange multiplier. Numerical computation of the channel capacity when power control is used is in general a non-trivial problem and is therefore further explored later in the chapter.

2.5 Optimality of Weighted OSTBC

The design of space-time codes [GFBK99, TSC98, TJC99] is an active area of research and numerous codes have been proposed for realizing some of the high data rates promised by previous MIMO capacity results as reported in e.g. [Tel95, FG98]. Traditionally, the design methods have in general relied on the assumption that the transmitter does not know the channel. Such conventional space-time codes are therefore unable to exploit channel knowledge at the transmitter, if available, for improving the performance.

However, the capacity result in Section 2.3.1 concerning the optimality of separate space-time coding and transmit weighting suggests that even for scenarios in which channel information is available, the use of the existing vast body of conventional space-time codes is reasonable if the output of the space-time encoder is weighted based on the channel information at the transmitter prior to the transmission. With this approach, some space-time codes are in practice easier to use than others. For block fading scenarios, a particularly appealing choice is the class of OSTB codes, leading to the previously mentioned weighted OSTBC transmission structure. Ways to determine the transmit weighting with the goal of minimizing the error probability will be considered in detail in later chapters. In this section, the focus is on the information theoretic aspects of weighted OSTBC. It will be shown that weighted OSTBC is optimal from a capacity perspective in the special case of two transmit antennas and one receive antenna.

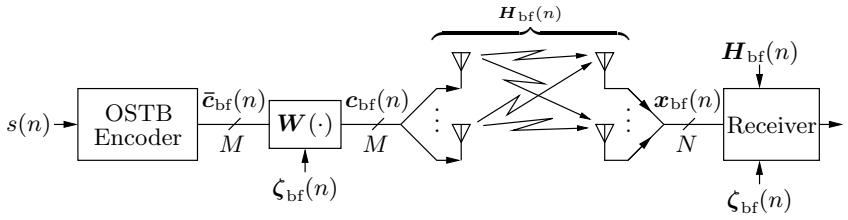

Figure 2.3: The setup in weighted OSTBC.

Consider the block fading scenario in Section 2.3.2 for $M = 2$ transmit antennas and $N = 1$ receive antenna. The structure of weighted OSTBC is closely related to the capacity achieving block based structure in (2.16). The setup is illustrated in Figure 2.3. A sequence of symbols $\{s(n)\}$, representing the data to be transmitted, is mapped into codewords of

length $L = 2$ by an OSTB encoder using the Alamouti code [Ala98]. Based on the channel side information $\boldsymbol{\zeta}_{\mathrm{bf}}(n)$, the encoder output $\bar{\boldsymbol{c}}_{\mathrm{bf}}(n)$ is weighted using $\boldsymbol{W}(\boldsymbol{\zeta}_{\mathrm{bf}}(n))$ producing the vector

$$\boldsymbol{c}_{\mathrm{bf}}(n) = \boldsymbol{W}(\boldsymbol{\zeta}_{\mathrm{bf}}(n))\bar{\boldsymbol{c}}_{\mathrm{bf}}(n)\,, \tag{2.45}$$

where the weighting function $\boldsymbol{W}(\cdot)$ may be chosen arbitrarily subject to a power constraint that will be introduced later. Clearly, this is similar to the capacity achieving encoder structure in (2.16), except that $\bar{\boldsymbol{c}}_{\mathrm{bf}}(n)$ is now constrained to be the output of an OSTB encoder, thus potentially limiting the possible data rates of the system.

The information carrying signals are transmitted over the wireless channel resulting in the received signal

$$x_{\mathrm{bf}}(n) = \boldsymbol{H}_{\mathrm{bf}}(n)\boldsymbol{W}(\boldsymbol{\zeta}_{\mathrm{bf}}(n))\bar{\boldsymbol{c}}_{\mathrm{bf}}(n) + e_{\mathrm{bf}}(n)\,, \tag{2.46}$$

where the variance of the noise term $e_{\mathrm{bf}}(n)$ is σ^2 and some of the boldface in the notation has been removed to reflect the fact that $x_{\mathrm{bf}}(n)$ and $e_{\mathrm{bf}}(n)$ are scalars because of the one receive antenna assumption. Now, collect the received signal corresponding to the kth block into a row vector

$$\begin{bmatrix} x_{\mathrm{bf}}(2k) & x_{\mathrm{bf}}(2k+1) \end{bmatrix}\,,$$

and combine (2.45) and (2.46) to obtain

$$\begin{aligned} \begin{bmatrix} x_{\mathrm{bf}}(2k) & x_{\mathrm{bf}}(2k+1) \end{bmatrix} &= \boldsymbol{H}(k)\boldsymbol{C}(k) + \boldsymbol{E}(k) \\ &= \boldsymbol{H}(k)\boldsymbol{W}(\boldsymbol{\zeta}(k))\bar{\boldsymbol{C}}(k) + \boldsymbol{E}(k)\,, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{C}(k) &\triangleq \begin{bmatrix} \boldsymbol{c}_{\mathrm{bf}}(2k) & \boldsymbol{c}_{\mathrm{bf}}(2k+1) \end{bmatrix} = \boldsymbol{W}(\boldsymbol{\zeta}(k))\bar{\boldsymbol{C}}(k) \\ \bar{\boldsymbol{C}}(k) &\triangleq \begin{bmatrix} \bar{\boldsymbol{c}}_{\mathrm{bf}}(2k) & \bar{\boldsymbol{c}}_{\mathrm{bf}}(2k+1) \end{bmatrix} \\ \boldsymbol{E}(k) &\triangleq \begin{bmatrix} e_{\mathrm{bf}}(2k) & e_{\mathrm{bf}}(2k+1) \end{bmatrix}\,, \end{aligned}$$

with the 2×2 matrix $\bar{\boldsymbol{C}}(k)$ representing the codeword output from the OSTB encoder in the kth block. Here, $\boldsymbol{\zeta}_{\mathrm{bf}}(n) = \boldsymbol{\zeta}(k)$, $kL \leq n \leq kL + L - 1$, as before. The Alamouti code takes two consecutive symbols $s(n)$ and constructs the output according to

$$\bar{\boldsymbol{C}}(k) = \begin{bmatrix} s(2k) & s^*(2k+1) \\ s(2k+1) & -s^*(2k) \end{bmatrix}\,. \tag{2.47}$$

Assume that the total output power over the entire block is P_{bm}, i.e., $\mathrm{E}[\|\boldsymbol{C}(k)\|_{\mathrm{F}}^2] = \mathrm{E}[\|\boldsymbol{c}_{\mathrm{bf}}(2k)\|^2 + \|\boldsymbol{c}_{\mathrm{bf}}(2k+1)\|^2] = P_{\mathrm{bm}}$, and that the corresponding symbols are normalized such that $\mathrm{E}[|s(2k)|^2 + |s(2k+1)|^2] = 2$. Combining these two assumptions and utilizing the orthogonality property $\bar{\boldsymbol{C}}(k)\bar{\boldsymbol{C}}(k)^* = (|s(2k)|^2 + |s(2k+1)|^2)\boldsymbol{I}_M$ of the Alamouti code gives

$$\begin{aligned}
\mathrm{E}[\|\boldsymbol{C}(k)\|_{\mathrm{F}}^2] &= \mathrm{E}[\|\boldsymbol{W}(\boldsymbol{\zeta}(k))\bar{\boldsymbol{C}}(k)\|_{\mathrm{F}}^2] \\
&= \mathrm{E}\left[\,\mathrm{tr}\left(\boldsymbol{W}(\boldsymbol{\zeta}(k))\bar{\boldsymbol{C}}(k)\bar{\boldsymbol{C}}(k)^*\boldsymbol{W}(\boldsymbol{\zeta}(k))^*\right)\right] \\
&= \mathrm{E}\left[\,\mathrm{tr}\left(\boldsymbol{W}(\boldsymbol{\zeta}(k))\boldsymbol{W}(\boldsymbol{\zeta}(k))^*\right)(|s(2k)|^2 + |s(2k+1)|^2)\right] \\
&= \mathrm{E}\left[\,\mathrm{tr}\left(\boldsymbol{W}(\boldsymbol{\zeta}(k))\boldsymbol{W}(\boldsymbol{\zeta}(k))^*\right)\right]\mathrm{E}\left[|s(2k)|^2 + |s(2k+1)|^2\right] \\
&= \mathrm{E}\left[\,\mathrm{tr}\left(\boldsymbol{W}(\boldsymbol{\zeta}(k))\boldsymbol{W}(\boldsymbol{\zeta}(k))^*\right)\right]2 = P_{\mathrm{bm}}\,, \qquad (2.48)
\end{aligned}$$

which is seen to correspond to a constraint on the weighting function.

An alternative representation of the setup more suitable for capacity calculations is to, similarly to e.g. in [HH02b], take the complex conjugate of $x_{\mathrm{bf}}(2k+1)$ and instead consider the received vector

$$\boldsymbol{x}(k) \triangleq \begin{bmatrix} x_{\mathrm{bf}}(2k) & x_{\mathrm{bf}}^*(2k+1) \end{bmatrix}^{\mathrm{T}} = \boldsymbol{\mathcal{H}}(k)\boldsymbol{s}(k) + \boldsymbol{e}(k)\,, \qquad (2.49)$$

where

$$\begin{aligned}
\boldsymbol{\mathcal{H}}(k) &\triangleq \begin{bmatrix} \bar{H}_{11}(k) & \bar{H}_{12}(k) \\ -\bar{H}_{12}^*(k) & \bar{H}_{11}^*(k) \end{bmatrix}, \quad \bar{H}_{kl}(k) \triangleq [\boldsymbol{H}(k)\boldsymbol{W}(\boldsymbol{\zeta}(k))]_{kl} \\
\boldsymbol{s}(k) &\triangleq \begin{bmatrix} s(2k) & s(2k+1) \end{bmatrix}^{\mathrm{T}}, \quad \boldsymbol{e}(k) \triangleq \begin{bmatrix} e(2k) & e^*(2k+1) \end{bmatrix}^{\mathrm{T}}
\end{aligned}$$

and where $[\boldsymbol{A}]_{kl}$ denotes element (k,l) of the matrix $\boldsymbol{A}$. The complex conjugate has been introduced purely for mathematical convenience in later expressions and does not affect the channel capacity since $\boldsymbol{x}(k)$ is a deterministic invertible function of the original received signal. Furthermore, note that the weighting $\boldsymbol{W}(\boldsymbol{\zeta}(k))$ has now been absorbed into an equivalent channel matrix $\boldsymbol{\mathcal{H}}(k)$.

To prove the optimality of weighted OSTBC, we will work on a block basis and show that the capacity of the above setup with the constrained encoder, assuming the weighting function has been chosen appropriately, is equal to the capacity for the general block fading setup in Section 2.3.2. The capacity for the latter is given by (2.14), which in this case reduces to

$$C_{\mathrm{bm}} = \max_{\substack{\boldsymbol{Z}(\cdot) \\ \boldsymbol{Z}(\boldsymbol{\zeta}) = \boldsymbol{Z}(\boldsymbol{\zeta})^* \succeq \boldsymbol{0} \\ 2\,\mathrm{E}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\zeta}))] = P_{\mathrm{bm}}}} 2\,\mathrm{E}[\log_2(1 + \sigma^{-2}\boldsymbol{H}\boldsymbol{Z}(\boldsymbol{\zeta})\boldsymbol{H}^*)]\,. \qquad (2.50)$$

Some additional notation needs to be introduced in order to compare this with the capacity achieved if the encoder is restricted to weighted OSTBC. Let C_{wo} denote the capacity when the encoder is constrained to the weighted OSTBC *structure* in (2.45) and let $C_{\text{wo}}(\boldsymbol{W}(\cdot))$ represent the capacity for a *fixed* weighting function $\boldsymbol{W}(\cdot)$. Since the encoders corresponding to C_{bm}, C_{wo} and $C_{\text{wo}}(\boldsymbol{W}(\cdot))$ are increasingly restrictive, it is clear that $C_{\text{bm}} \geq C_{\text{wo}} \geq C_{\text{wo}}(\boldsymbol{W}(\cdot))$. All these capacities are based on a channel model with $(\boldsymbol{c}_{\text{bf}}(2k), \boldsymbol{c}_{\text{bf}}(2k+1))$ as input and $(x_{\text{bf}}(2k), x_{\text{bf}}(2k+1), \boldsymbol{H}(k), \boldsymbol{\zeta}(k))$ as output.

A potential lower bound on $C_{\text{wo}}(\boldsymbol{W}(\cdot))$ is obtained by constraining the setup even further and, with reference to (2.49), investigating the capacity of a related *fictitious* channel model taking $\boldsymbol{s}(k)$ as input and $(\boldsymbol{x}(k), \boldsymbol{\mathcal{H}}(k))$ as output, with the power constraint $\mathrm{E}[|s(2k)|^2 + |s(2k+1)|^2] = \mathrm{E}[\|\boldsymbol{s}(k)\|^2] = 2$. Assume that the weighting function $\boldsymbol{W}(\cdot)$ is fixed and let $C_{\text{fch}}(\boldsymbol{W}(\cdot))$ denote the capacity of the fictitious channel model. The reason why $C_{\text{fch}}(\boldsymbol{W}(\cdot))$ may constitute a lower bound on $C_{\text{wo}}(\boldsymbol{W}(\cdot))$ is because the original output $(\boldsymbol{x}(k), \boldsymbol{H}(k), \boldsymbol{\zeta}(k))$ has here been replaced with $(\boldsymbol{x}(k), \boldsymbol{\mathcal{H}}(k))$, which is more restrictive since it is a function of the former.

In this new channel model both the weighting $\boldsymbol{W}(\boldsymbol{\zeta}(k))$ and the OSTB encoder are viewed as part of the channel and $\boldsymbol{s}(k)$ is seen as the output of a fictitious transmitter that does not rely on any side information. In other words, $\boldsymbol{\zeta}(k)$ has been absorbed into the channel model creating a system that seemingly is without side information at the transmitter.

To compute the capacity for the fictitious channel model, observe that $\boldsymbol{\mathcal{H}}(k)$ now plays the role of the channel matrix and is a deterministic function of both the original channel matrix $\boldsymbol{H}(k)$ as well as the side information $\boldsymbol{\zeta}(k)$. Hence, $\{\boldsymbol{\mathcal{H}}(k)\}$ is a stationary and ergodic random process. It is easily verified that $\{\boldsymbol{e}(k)\}$ is a zero-mean complex Gaussian random process with a covariance matrix $\boldsymbol{\mathcal{R}}_{\boldsymbol{ee}} = \sigma^2 \boldsymbol{I}_2$. From this it follows that the fictitious channel model corresponds to a MIMO system in which the receiver knows the channel matrix $\boldsymbol{\mathcal{H}}(k)$ perfectly while there is no channel knowledge at the transmitter. Consequently, the capacity formula in (2.21) for the no channel knowledge case is applicable and results in

$$C_{\text{fch}}(\boldsymbol{W}(\cdot)) = \max_{\substack{\boldsymbol{Z} \\ \boldsymbol{Z}=\boldsymbol{Z}^* \succeq \boldsymbol{0} \\ \mathrm{tr}(\boldsymbol{Z})=2}} \mathrm{E}_{\boldsymbol{\mathcal{H}}}[\log_2 \det(\boldsymbol{I}_2 + \sigma^{-2} \boldsymbol{\mathcal{H}} \boldsymbol{Z} \boldsymbol{\mathcal{H}}^*)] \,. \tag{2.51}$$

Note that $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^* = (|\bar{H}_{11}|^2 + |\bar{H}_{12}|^2)\boldsymbol{I}_2$. Dropping the maximization in

(2.51) and inserting $\boldsymbol{Z} = \boldsymbol{I}_2$ then gives a lower bound on the capacity as

$$\begin{aligned} C_{\text{fch}}(\boldsymbol{W}(\cdot)) &\geq \mathrm{E}_{\boldsymbol{\mathcal{H}}}[\log_2 \det(\boldsymbol{I}_2 + \sigma^{-2}\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^*)] \\ &= \mathrm{E}_{\boldsymbol{\mathcal{H}}}[\log_2 \det(\boldsymbol{I}_2 + \sigma^{-2}(|\bar{H}_{11}|^2 + |\bar{H}_{12}|^2)\boldsymbol{I}_2)] \\ &= \mathrm{E}_{\boldsymbol{H},\boldsymbol{\zeta}}[\log_2 \det(\boldsymbol{I}_2 + \sigma^{-2}\|\boldsymbol{H}\boldsymbol{W}(\boldsymbol{\zeta})\|_{\text{F}}^2\boldsymbol{I}_2)] \\ &= 2\,\mathrm{E}_{\boldsymbol{H},\boldsymbol{\zeta}}[\log_2 \det(1 + \sigma^{-2}\|\boldsymbol{H}\boldsymbol{W}(\boldsymbol{\zeta})\|_{\text{F}}^2)] \\ &= 2\,\mathrm{E}_{\boldsymbol{H},\boldsymbol{\zeta}}[\log_2 \det(1 + \sigma^{-2}\boldsymbol{H}\boldsymbol{W}(\boldsymbol{\zeta})\boldsymbol{W}(\boldsymbol{\zeta})^*\boldsymbol{H}^*)]\,. \end{aligned}$$

By maximizing this lower bound over all weighting functions $\boldsymbol{W}(\cdot)$ that satisfy the constraint in (2.48), and utilizing the fact that $C_{\text{wo}} \geq C_{\text{wo}}(\boldsymbol{W}(\cdot)) \geq C_{\text{fch}}(\boldsymbol{W}(\cdot))$, it is realized that the capacity for a weighted OSTBC constrained encoder is lower bounded as

$$\begin{aligned} C_{\text{wo}} &\geq \max_{\substack{\boldsymbol{W}(\cdot) \\ 2\,\mathrm{E}[\mathrm{tr}(\boldsymbol{W}(\boldsymbol{\zeta})\boldsymbol{W}^*(\boldsymbol{\zeta}))] = P_{\text{bm}}}} 2\,\mathrm{E}_{\boldsymbol{H},\boldsymbol{\zeta}}[\log_2 \det(1 + \sigma^{-2}\boldsymbol{H}\boldsymbol{W}(\boldsymbol{\zeta})\boldsymbol{W}(\boldsymbol{\zeta})^*\boldsymbol{H}^*)] \\ &= C_{\text{bm}}\,, \end{aligned}$$

where the equality is due to the fact that the above optimization problem corresponds to (2.50). However, since C_{wo} cannot be larger than C_{bm}, it follows that the inequality holds with equality, i.e., $C_{\text{wo}} = C_{\text{bm}}$.

Thus, we conclude that the weighted OSTBC structure in the case of two transmit antennas and one receive antenna is optimal in the sense that it can achieve the (unconstrained encoder) channel capacity. Unfortunately, this result does not hold for other antenna configurations. In such cases, the use of weighted OSTBC will incur a non-negligible capacity loss that will be particularly severe if the number of elements in the transmit antenna array is such that an OSTB code with real-valued symbols must be used, see [TJC99] for examples of the latter type of codes. An idea of the magnitude of the loss can be inferred from [HH02b], which provides a capacity study of OSTB codes in the special case of no channel knowledge at the transmitter.

2.6 Numerical Computation of Capacity

Unfortunately, computing the channel capacity by solving the accompanying optimization problem is in general a non-trivial task. This is primarily because the capacity formula relies on expectations which typically are not available in closed-form. A possible remedy is to perform a

Monte-Carlo simulation of the system and replace the expectations with their corresponding sample estimates. The present section explores such strategies in more detail when the channel knowledge is perfect as well as when it is quantized.

2.6.1 Perfect Channel Knowledge

Consider the case of perfect channel knowledge at the transmitter and the resulting capacity expressions in (2.31) - (2.33). In this case, the problem simplifies considerably since the maximization has already been performed analytically leaving only μ to be determined from the power constraint in (2.32). Solving (2.32) is however complicated for fading scenarios in which a closed-form expression of $\mathrm{E}[\max\{1/\mu - 1/\lambda_k(\boldsymbol{H}), 0\}]$ is unavailable. A Monte-Carlo approach is certainly possible but computationally demanding since the number of samples required to obtain good estimates of $\mathrm{E}[\max\{1/\mu - 1/\lambda_k(\boldsymbol{H}), 0\}]$, $k = 1, 2 \cdots, r$, means that an equation containing an enormous number of terms has to be solved numerically. However, after determining μ and hence $\tilde{Z}_{kk}(\boldsymbol{H})$, the capacity expression in (2.33) is readily computed based on Monte-Carlo integration.

As will be shown next, determining μ is relatively straightforward in the case of one receive antenna and a channel matrix $\boldsymbol{H}$ with elements $\{H_{kl}\}$ drawn from a zero-mean IID complex Gaussian distribution. The latter corresponds to a spatially uncorrelated Rayleigh fading scenario. As previously mentioned, the capacity for a single receive antenna is determined by (2.36) - (2.38). The constant μ needs to be computed by solving the power constraint equation in (2.38). Let σ^2 and σ_h^2 denote the noise variance and the variance of H_{kl}, respectively. Note that, because the H_{kl}'s are zero-mean IID complex Gaussian, $2\|\boldsymbol{H}\|^2/\sigma_h^2$ is a chi-squared variable with $2M$ degrees of freedom [Kay98, p. 24]. The corresponding PDF is sufficiently simple so as to permit closed-form evaluation of $\mathrm{E}[\max\{1/\mu - \sigma^2/\|\boldsymbol{H}\|^2, 0\}]$, giving rise to the power constraint equation

$$\frac{\exp(-\alpha\mu)}{\mu} \sum_{k=0}^{M-1} \frac{(\alpha\mu)^k}{k!} - \alpha \frac{\exp(-\alpha\mu)}{M-1} \sum_{k=0}^{M-2} \frac{(\alpha\mu)^k}{k!} = P\,, \tag{2.52}$$

where $\alpha \triangleq \sigma^2/\sigma_h^2$. The proof is presented in Appendix 2.C. Standard methods for numerical solution of equations can now be used to solve for

μ in (2.52) and hence avoid the most difficult part of the Monte-Carlo simulation approach.

2.6.2 Memoryless Quantized Deterministic Feedback

In FDD type of systems, quantized channel information is usually conveyed from the receiver to the transmitter via a dedicated feedback link. This is for example the case in the closed-loop transmit diversity mode of the WCDMA system [3GP02b]. Consider Figure 2.4 in which a similar feedback scenario is illustrated. Assume that the channel process $\{\boldsymbol{H}(n)\}$ is memoryless and that the feedback quantizer/encoder is described by a deterministic function $\varepsilon(\cdot)$ which maps the channel realization $\boldsymbol{H}(n)$ into a Q-valued integer $i(n) \in \{0, 1, \cdots, Q-1\}$. The output of the quantizer is conveyed over an ideal non-distorting feedback channel to the transmitter for which $i(n)$ constitutes the available side information, i.e., $\boldsymbol{\zeta}(n) = i(n)$.

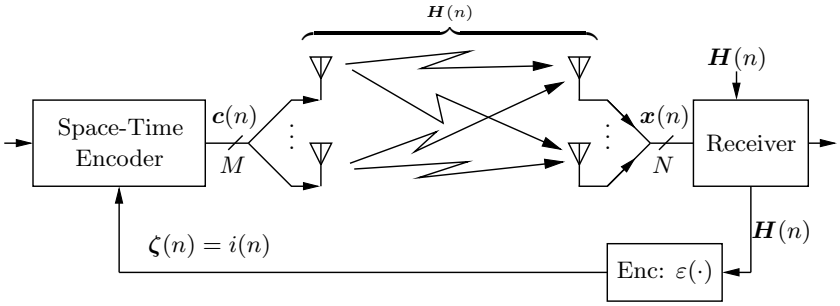

Figure 2.4: Quantized side information via a feedback link.

The encoder function $\varepsilon(\cdot)$ defines a partition $\{\mathcal{S}_k : k \in \{0, 1, \cdots, Q-1\}\}$ of $\mathbb{C}^{N\times M}$ such that $\boldsymbol{H}(n) \in \mathcal{S}_k \Rightarrow \boldsymbol{\zeta}(n) = i(n) = \varepsilon(\boldsymbol{H}(n)) = k$. Here, $\mathbb{C}$ denotes the complex number field and $\mathcal{S}_k$ represents the kth *encoder region* corresponding to a quantizer output value of k. Define the probability that the quantizer outputs a value k as $p_k \triangleq \Pr[\boldsymbol{H} \in \mathcal{S}_k] = \Pr[\varepsilon(\boldsymbol{H}) = k]$ and let $\boldsymbol{Z}_k \triangleq \boldsymbol{Z}(k) = \boldsymbol{W}(k)\boldsymbol{W}(k)^*$, where we have dropped the time index. By conditioning on the outputs of the quantizer,

the capacity formula in (2.8) can be rewritten to arrive at

$$C = \max_{\substack{\{\boldsymbol{Z}_k\}_{k=0}^{Q-1} \\ \boldsymbol{Z}_k = \boldsymbol{Z}_k^* \succeq \boldsymbol{0}, \forall k \\ \sum_{k=0}^{Q-1} p_k \operatorname{tr} \boldsymbol{Z}_k = P}} \sum_{k=0}^{Q-1} p_k \, \mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H} \boldsymbol{Z}_k \boldsymbol{H}^*) | \boldsymbol{H} \in \mathcal{S}_k] \,. \tag{2.53}$$

Similarly to (2.8), the above optimization problem is convex and hence all local optima are also globally optimal, making it considerably easier to find a global optimum. Constrained convex optimization problems are often efficiently solved using interior point methods [NN94]. The use of such methods to obtain a numerical value on the capacity will therefore be considered next.

Background on Interior Point Methods

The philosophy behind interior point methods is to convert a constrained optimization problem into an unconstrained problem, which can be easily solved using standard numerical optimization techniques like the Newton method. Obtaining an unconstrained problem is accomplished by incorporating the constraints by means of *barrier functions*.

To illustrate, assume a criterion function $f(\boldsymbol{x})$ is to be maximized with respect to $\boldsymbol{x}$, subject to some constraints described by $f_k(\boldsymbol{x}) \leq 0$, $k = 1, 2, \cdots$. Let $\bar{\boldsymbol{x}}$ denote the optimal solution and introduce the barrier function $\tilde{f}_k(\boldsymbol{x})$ corresponding to the kth constraint. An unconstrained optimization problem is obtained by instead solving

$$\max_{\boldsymbol{x}} \left(t f(\boldsymbol{x}) + \sum_k \tilde{f}_k(\boldsymbol{x}) \right) , \tag{2.54}$$

producing the solution $\bar{\boldsymbol{x}}(t)$. Here, $t > 0$ is a parameter which is used to obtain a tradeoff between the original criterion function and the barrier functions. Ideally, the barrier functions should not affect the criterion function in (2.54) for values of $\boldsymbol{x}$ that satisfy the constraints and, conversely, should prevent violations of the constraints. In other words, ideal barrier functions are given by

$$\tilde{f}_k(\boldsymbol{x}) = \begin{cases} 0, & f_k(\boldsymbol{x}) \leq 0 \\ -\infty, & f_k(\boldsymbol{x}) > 0 \,, \end{cases} \tag{2.55}$$

meaning that the solution $\bar{\boldsymbol{x}}(t)$ of the unconstrained problem is also the desired solution $\bar{\boldsymbol{x}}$ of the original problem, i.e., $\bar{\boldsymbol{x}}(t) = \bar{\boldsymbol{x}}$ regardless of the choice of t.

However, in practice it is attractive to work with smooth differentiable functions. Hence, the property in (2.55) is approximated using a function which is non-zero, smooth and differentiable when $f_k(\boldsymbol{x}) \leq 0$ and which tends to $-\infty$ as $\boldsymbol{x}$ approaches the boundary described by $f_k(\boldsymbol{x}) = 0$. When $\boldsymbol{x}$ is close to a boundary, we say that the corresponding barrier function operates close to its singular domain. The optimal solution $\bar{\boldsymbol{x}}(t)$ now becomes a function of t since the barrier functions are non-zero within the allowed set of $\boldsymbol{x}$. By intuition, it is realized that $\bar{\boldsymbol{x}}(t)$ should be close to the optimum $\bar{\boldsymbol{x}}$ if t is large so that the influence from the non-zero barrier functions is negligible. To speed up the convergence, (2.55) is however better solved using successively larger values $t = t_l$, $l = 1, 2, \cdots$, with $\bar{\boldsymbol{x}}(t_l)$ serving as the starting point in the numerical solution of (2.54) in iteration $l + 1$. Under certain weak assumptions, it can be shown that $\bar{\boldsymbol{x}}(t_l)$ converges to the optimum $\bar{\boldsymbol{x}}$ of the original problem [BV99].

Utilizing Interior Point Methods for Computing the Capacity

At least in principle, it seems straightforward to compute the capacity by solving (2.53) using interior point methods. As shown in [BV99], semi definite constraints like $\{\boldsymbol{Z}_k \succeq \boldsymbol{0}\}$ may be taken into account by the use of Q barrier functions

$$\tilde{f}(\boldsymbol{Z}_k) = \log\det(\boldsymbol{Z}_k), \quad k = 0, \cdots, Q-1\,.$$

Since the remaining power constraint is linear in the optimization parameters, it is possible to re-parameterize the problem so as to directly satisfy the constraint. Both the expectations as well as the PMF p_k of the quantizer output can be replaced by appropriate sample estimates taken from a Monte-Carlo simulation.

However, there are drawbacks with an interior point approach based on Monte-Carlo simulation. While the p_k's are constant and can hence be computed once and for all prior to the optimization, the conditional expectations need to be re-estimated every time the criterion function or any of its derivatives are needed. The huge number of resulting terms in the criterion function means that obtaining a solution is a computationally demanding process. Moreover, the maximization in (2.53) is to be performed jointly[5] over all the $\boldsymbol{Z}_k$'s, which means that M^2Q real-valued

[5]For the non-power-controlled scenario, the power constraint in (2.53) is replaced

parameters needs to be optimized ($\boldsymbol{Z}_k$ contains M^2 real-valued parameters since it is a Hermitian matrix). This may be a large number, even for moderate transmit antenna array sizes and fairly coarse quantization, thus further slowing down the convergence of the numerical procedure. If in addition the optimal solution is close to beamforming, in the sense that many or all of the optimal $\boldsymbol{Z}_k$'s are close to rank one, many of the constraints $\{\boldsymbol{Z}_k \succeq \boldsymbol{0}\}$ are tight and the corresponding barrier functions operate close to their singular domains. This creates severe numerical difficulties, making it even more difficult to reach convergence and hence solving the optimization problem within a reasonable time frame. Scenarios in which the receiver only has one antenna are particularly vulnerable to the latter kind of problem. However, there are in some cases remedies, as will be discussed in the following section.

2.6.3 Symmetric Feedback

To avoid numerical difficulties in computing the capacity, there is a need for reducing the problem size and in particular reducing the number of tight constraints. In this section we will consider the previous memoryless quantized deterministic feedback scenario and introduce the notion of symmetric feedback. It will be shown that in case of symmetric feedback, the optimization problem simplifies considerably. The rather theoretical concept of symmetric feedback will moreover be applied to a scenario of practical interest.

Definition of Symmetric Feedback

It often makes sense to design the quantizer so that the encoder regions $\{\mathcal{S}_k\}$ exhibit certain rotational symmetries. Such rotational symmetries form the basis of our definition of symmetric feedback. Toward this end, let $\mathcal{S}_{\mathrm{s}}^{(q)} = \{\mathcal{S}_k : k \in \mathcal{I}_{\mathrm{s}}^{(q)}\}$ be a subset of the encoder regions $\{\mathcal{S}_k\}_{k=0}^{Q-1}$ with the corresponding index set $\mathcal{I}_{\mathrm{s}}^{(q)}$ hence representing a subset of $\{0, 1, \cdots, Q-1\}$. Renumber the index set using the function $\kappa^{(q)}(k)$, defined such that $\mathcal{I}_{\mathrm{s}}^{(q)} = \{\kappa^{(q)}(k)\}_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1}$, where $|\mathcal{I}_{\mathrm{s}}^{(q)}|$ denotes the number of elements in $\mathcal{I}_{\mathrm{s}}^{(q)}$. Let there be Q' different $\mathcal{S}_{\mathrm{s}}^{(q)}$ such that $\{\mathcal{S}_{\mathrm{s}}^{(q)}\}_{q=0}^{Q'-1}$

by Q constraints, each containing only one $\boldsymbol{Z}_k$. The optimization problem therefore decouples and each term in the criterion function can be separately maximized. This considerably simplifies the optimization problem and in fact makes the process of determining the capacity rather painless.

partitions the elements of $\{\mathcal{S}_k\}_{k=0}^{Q-1}$ into Q' subsets.

Recall that a unitary matrix may correspond to a rotation. Guided by this and the above, we say that the feedback is symmetric with respect to the encoder regions in $\mathcal{S}_{\mathrm{s}}^{(q)}$ if the following two conditions are true:

1. For any k such that $\kappa^{(q)}(k) \in \mathcal{I}_{\mathrm{s}}^{(q)}$ there exists a unitary matrix $\boldsymbol{Q}_k^{(q)}$ such that $\boldsymbol{H} \in \mathcal{S}_{\kappa^{(q)}(0)} \Leftrightarrow \boldsymbol{H}\boldsymbol{Q}_k^{(q)} \in \mathcal{S}_{\kappa^{(q)}(k)}$.

2. The PDF $p_{\boldsymbol{H}}(\boldsymbol{H})$ of the channel satisfies the relation $p_{\boldsymbol{H}}(\boldsymbol{H}) = p_{\boldsymbol{H}}(\boldsymbol{H}\boldsymbol{U})$ for all unitary $M \times M$ matrices $\boldsymbol{U}$.

In other words, any encoder region $\mathcal{S}_{\kappa^{(q)}(k)}$ in $\mathcal{S}_{\mathrm{s}}^{(q)}$ can be obtained from $\mathcal{S}_{\kappa^{(q)}(0)}$ via a "rotation" described by $\boldsymbol{Q}_k^{(q)}$ and the distribution of the channel fading is invariant to all unitary linear transformations $\boldsymbol{U}$.

If the two conditions are true for the entire set of encoder regions, i.e., only $Q' = 1$ subset $\mathcal{S}_{\mathrm{s}}^{(0)} = \{\mathcal{S}_k\}_{k=0}^{Q-1}$ is needed, we have a *symmetric feedback scenario*. To stress that it may be necessary to introduce several subsets $\mathcal{S}_{\mathrm{s}}^{(q)}$, the scenario is termed *partially* symmetric if $Q' < Q$, meaning that there is at least one subset $\mathcal{S}_{\mathrm{s}}^{(q)}$ with more than one element satisfying the conditions. Note that the first condition is automatically satisfied for the degenerate case of a single element subset $\mathcal{S}_{\mathrm{s}}^{(q)}$. Hence there exists at least one partition $\{\mathcal{S}_{\mathrm{s}}^{(q)}\}_{q=0}^{Q'-1}$ (choose $\mathcal{S}_{\mathrm{s}}^{(q)} = \mathcal{S}_q$, $q = 0, \cdots, Q-1$) such that the first condition is true for all $\mathcal{S}_{\mathrm{s}}^{(q)}$. It is therefore always possible to reformulate the capacity formula using the terminology of the present section, as will be shown next.

Exploiting Symmetric Feedback for Reducing Problem Size

For a partially symmetric feedback scenario, the size of the optimization problem in (2.53) can be decreased. Loosely speaking, the symmetries implied by the two conditions translate into corresponding symmetries in the criterion function and the power constraint, from which it follows that both the number of optimization parameters and the number of terms can be reduced. To be more specific, it is shown in Appendix 2.D.1 that if the feedback is symmetric with respect to some subset $\mathcal{S}_{\mathrm{s}}^{(q)}$, then the corresponding encoder output probabilities are all equal, i.e.,

$$p_{\kappa^{(q)}(0)} = p_{\kappa^{(q)}(1)} = \cdots = p_{\kappa^{(q)}(|\mathcal{I}_{\mathrm{s}}^{(q)}|-1)}\,, \tag{2.56}$$

and

$$\begin{aligned}&\mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{X}\boldsymbol{H}^*)|i=\kappa^{(q)}(k)]\\&\quad=\mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Q}_k^{(q)}\boldsymbol{X}(\boldsymbol{Q}_k^{(q)})^*\boldsymbol{H}^*)|i=\kappa^{(q)}(0)]\,,\end{aligned}\quad(2.57)$$

for some non-random matrix $\boldsymbol{X}$. By grouping terms that correspond to the same subset $\mathcal{S}_{\mathrm{s}}^{(q)}$ together, utilizing (2.56) and (2.57), reparameterizing the problem using $\tilde{\boldsymbol{Z}}_k^{(q)} \triangleq \boldsymbol{Q}_k^{(q)}\boldsymbol{Z}_{\kappa^{(q)}(k)}(\boldsymbol{Q}_k^{(q)})^*$ and finally noticing that all $\tilde{\boldsymbol{Z}}_k^{(q)}$ corresponding to a certain subset can be taken to be equal, or more precisely, $\tilde{\boldsymbol{Z}}_k^{(q)} = \tilde{\boldsymbol{Z}}^{(q)}$, $\forall k$, it is shown in Appendix 2.D.2 that the capacity formula in (2.53) reduces to

$$C = \max_{\substack{\{\tilde{\boldsymbol{Z}}^{(q)}\}_{q=0}^{Q'-1}\\ \tilde{\boldsymbol{Z}}^{(q)}=(\tilde{\boldsymbol{Z}}^{(q)})^*\succeq\boldsymbol{0},\forall q\\ \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)}|\mathcal{I}_{\mathrm{s}}^{(q)}|\,\mathrm{tr}(\tilde{\boldsymbol{Z}}^{(q)})=P}} \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)}|\mathcal{I}_{\mathrm{s}}^{(q)}| f^{(q)}\big(\tilde{\boldsymbol{Z}}^{(q)}\big)\,, \quad (2.58)$$

where

$$f^{(q)}(\boldsymbol{X}) \triangleq \mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{X}\boldsymbol{H}^*)|i=\kappa^{(q)}(0)]\,.$$

Compared with (2.53), the optimization problem in (2.58) is simpler if the scenario is at least partially symmetric since the maximization is then only over the $Q' < Q$ different $\tilde{\boldsymbol{Z}}^{(q)}$'s. The solution in terms of the original parameters $\{\boldsymbol{Z}_k\}$ is obtained by solving for $\boldsymbol{Z}_{\kappa^{(q)}(k)}$ in $\tilde{\boldsymbol{Z}}^{(q)} = \tilde{\boldsymbol{Z}}_k^{(q)} = \boldsymbol{Q}_k^{(q)}\boldsymbol{Z}_{\kappa^{(q)}(k)}(\boldsymbol{Q}_k^{(q)})^*$, giving $\boldsymbol{Z}_{\kappa^{(q)}(k)} = (\boldsymbol{Q}_k^{(q)})^*\tilde{\boldsymbol{Z}}^{(q)}\boldsymbol{Q}_k^{(q)}$, where $\tilde{\boldsymbol{Z}}^{(q)}$ now is assumed to be part of an optimal solution of (2.58). Since $\boldsymbol{Z}_k = \boldsymbol{W}(k)\boldsymbol{W}(k)^*$, optimal transmit weights are given by

$$\boldsymbol{W}(\kappa^{(q)}(k)) = (\boldsymbol{Q}_k^{(q)})^*\boldsymbol{W}^{(q)}\,, \quad (2.59)$$

where $\boldsymbol{W}^{(q)}$ is such that $\boldsymbol{W}^{(q)}(\boldsymbol{W}^{(q)})^* = \tilde{\boldsymbol{Z}}^{(q)}$. In other words, the symmetry of the quantization regions leads to a corresponding symmetry of the transmit weights and only one weight $\boldsymbol{W}^{(q)}$ per subset needs to be optimized.

Example 2.6.1 (A symmetric feedback scenario)
To exemplify the symmetric feedback results, consider the previously described memoryless quantized feedback scenario with a receiver that has

only $N = 1$ receive antenna. Furthermore, assume spatially uncorrelated Rayleigh fading, meaning that the elements of

$$\boldsymbol{H}(n) = \begin{bmatrix} H_{11}(n) & H_{12}(n) & \cdots H_{1M}(n) \end{bmatrix}$$

are drawn independently according to a zero-mean complex Gaussian distribution. Consider a quantizer that employs a form of uniform phase quantization to map the channel $\boldsymbol{H}(n)$ into a b-bit integer $i(n)$. There are hence $Q = 2^b$ different quantization regions. Such a quantization strategy is reasonable in view of the fact that the phases of the channel coefficients are uniformly distributed [Pro95, p. 767]. It is common practice to divide the channel by the first channel coefficient [NLTW98] and instead uniformly quantize the phase of the elements in $\boldsymbol{H}(n)/H_{11}(n)$. This is the case both in the so-called partial phase combining scheme in [HP98] as well as in the closed-loop mode of the WCDMA system [3GP02b]. It is therefore interesting to study similar such quantization schemes and we hence assume the feedback encoder in our example quantizes according to

$$i(n) = \varepsilon(\boldsymbol{H}(n)) = \arg \min_{k \in \{0,1,\cdots,2^b-1\}} \|\boldsymbol{H}(n)^{\mathrm{T}}/H_{11}(n) - \boldsymbol{\varepsilon}_k\|^2 , \quad (2.60)$$

with the *codebook vectors* $\{\boldsymbol{\varepsilon}_k\}_{k=0}^{2^b-1}$ given by

$$\boldsymbol{\varepsilon}_k = \begin{bmatrix} 1 & e^{j\phi_{i_0(k)}} & \cdots & e^{j\phi_{i_{M-2}(k)}} \end{bmatrix}^{\mathrm{T}} , \quad (2.61)$$

where $j \triangleq \sqrt{-1}$ and $\phi_{i_l(k)} = 2\pi i_l(k)/2^{\bar{b}}$. Here, $\bar{b} = b/(M-1)$ represents the number of bits per complex-valued dimension and $i_l(k) \in \{0, 1, \cdots, 2^{\bar{b}} - 1\}$ is implicitly defined through the relation $k = \sum_{l=0}^{M-2} i_l(k) 2^{\bar{b}l}$. The latter relation means that $i_0(k)$ represents the decimal number corresponding to the $\bar{b}$ least significant bits of a natural binary representation of k. Similarly, $i_1(k)$ corresponds to the next $\bar{b}$ bits and so on up to $i_{M-2}(k)$ which corresponds to the $\bar{b}$ most significant bits. It is hence seen that (2.60) and (2.61) together implement $\bar{b}$ bit uniform *scalar* quantization of the phases of $H_{1l}(n)/H_{11}(n)$, $l = 2, \cdots, M$.

For the problem at hand, experiments show that the optimal solution of (2.53) often corresponds to beamforming, even for a modest number of bits b. This may be of no surprise considering the optimality of beamforming established for the perfect channel knowledge case in Section 2.4.2. A related result is also found in [NLTW98] where based on a similar quantization scheme it is proved that beamforming is optimal for

the case of two transmit antennas and a non-power-controlled scenario. Unfortunately, the optimality of beamforming means that the optimal $\boldsymbol{Z}_k$'s are essentially rank one creating numerical difficulties as previously explained. These difficulties can however in this case be overcome by utilizing the theory about symmetric feedback. To see this, first note that the codebook vectors can all be expressed in terms of the first codebook vector as $\boldsymbol{\varepsilon}_k = \boldsymbol{Q}_k \boldsymbol{\varepsilon}_0$, where

$$\boldsymbol{Q}_k \triangleq \operatorname{diag}(1, e^{j\phi_{i_0(k)}}, e^{j\phi_{i_1(k)}}, \cdots, e^{j\phi_{i_{M-2}(k)}}) \tag{2.62}$$

is a diagonal unitary matrix that rotates $\boldsymbol{\varepsilon}_0$ to the desired position. It is reasonable to expect that these rotational symmetries carry over both to the encoder regions as well as to the conditional expectations in the criterion function, the latter because of the symmetry inherent in the IID complex Gaussian fading distribution. In Appendix 2.D.3 it is shown that this is indeed the case by proving that the feedback is symmetric with respect to *all* encoder regions $\{\mathcal{S}_k\}_{k=0}^{Q-1}$, with the diagonal $\boldsymbol{Q}_k$'s representing the unitary matrices mentioned in the symmetric feedback conditions. Hence, the two feedback conditions are satisfied for $\mathcal{S}_{\mathrm{s}}^{(0)} = \{\mathcal{S}_k\}_{k=0}^{Q-1}$ with $\boldsymbol{Q}_k^{(0)} = \boldsymbol{Q}_k$, where the renumbering function $\kappa^{(0)}(k)$ is taken as $\kappa^{(0)}(k) = k$. Consequently, there is only $Q' = 1$ subset and the number of elements in the corresponding index set $\mathcal{I}_{\mathrm{s}}^{(0)}$ is $Q = 2^b$. This constitutes a symmetric feedback scenario and (2.58) can thus be used to obtain the capacity as

$$\begin{aligned} C &= \max_{\substack{\tilde{\boldsymbol{Z}}^{(0)} \\ \tilde{\boldsymbol{Z}}^{(0)} = (\tilde{\boldsymbol{Z}}^{(0)})^* \succeq \boldsymbol{0} \\ p_0 Q \operatorname{tr}(\tilde{\boldsymbol{Z}}^{(0)}) = P}} p_0 Q \, \mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1} \boldsymbol{H} \tilde{\boldsymbol{Z}}^{(0)} \boldsymbol{H}^*) | i = 0] \\ &= \max_{\substack{\tilde{\boldsymbol{Z}}^{(0)} \\ \tilde{\boldsymbol{Z}}^{(0)} = (\tilde{\boldsymbol{Z}}^{(0)})^* \succeq \boldsymbol{0} \\ \operatorname{tr}(\tilde{\boldsymbol{Z}}^{(0)}) = P}} \mathrm{E}[\log_2(1 + \sigma^{-2} \boldsymbol{H} \tilde{\boldsymbol{Z}}^{(0)} \boldsymbol{H}^*) | i = 0] \,, \end{aligned} \tag{2.63}$$

where σ^2 denotes the noise variance and where $p_0 Q = 1$, or more precisely $p_k = 1/Q$, follows from (2.56) and the fact that $\sum_{k=0}^{Q-1} p_k = 1$.

Clearly, the task of solving (2.63) is much easier than solving the original problem in (2.53). For example, the number of real-valued parameters in the optimization problem has been reduced from $M^2 Q$ to only M^2, the criterion function consists of only one conditional expectation and only one barrier function corresponding to $\tilde{\boldsymbol{Z}}_0^{(0)} \succeq \boldsymbol{0}$ needs to be used.

Note that since *all* transmit weighting matrices are now related through unitary transformations as in (2.59), the output power does not vary with the outcome of the side information $\boldsymbol{\zeta}(n) = i(n)$. Because of the quantization scheme and the channel fading, power control is not used even though it is allowed by the power constraint. Thus, (2.63) gives the capacity for the non-power-controlled scenario as well. □

2.7 Numerical Examples

In this section, results from numerical evaluations of some of the capacity expressions are presented with the main goal of illustrating the gains due to channel information at the transmitter. The high data rates offered by MIMO systems are also illustrated as a bi-product of the comparisons to follow. The results provide motivation for the development in later chapters of transmission schemes that take channel information into account.

Quantized channel information as well as the extreme cases of no and perfect channel knowledge are considered. Numerical values are obtained using the Monte-Carlo based computation methods described in Section 2.6. All of the following examples are concerned with a Rayleigh fading scenario in which the channel process $\{\boldsymbol{H}(n)\}$ is memoryless with the elements of $\boldsymbol{H}(n)$ drawn from a zero-mean IID complex Gaussian distribution with the common variance $\sigma_h^2 = 1$. The noise is spatially white with variance σ^2 and the output power is $P = 1$. Furthermore, SNR is defined as

$$\text{SNR} \triangleq \frac{P\sigma_h^2}{\sigma^2},$$

corresponding to the power ratio $\mathrm{E}[\|\boldsymbol{H}(n)\boldsymbol{c}(n)\|_{\mathrm{F}}^2] / \mathrm{E}[\|\boldsymbol{e}(n)\|^2]$ when the transmitter does not know the channel. Additional assumptions are given for each specific example below.

2.7.1 Perfect versus No Channel Knowledge

To give an idea of the maximum gains to expect from the use of channel information, the two extreme cases of no and perfect channel knowledge will be compared for various antenna configurations. A scenario without power control is considered for the sake of simplifying the capacity computation in the latter case. Recall from Section 2.4 that no channel knowledge is modeled by assuming that the side information process $\{\boldsymbol{\zeta}(n)\}$ is

statistically independent of the channel matrix process $\{\boldsymbol{H}(n)\}$, resulting in the classical capacity formula in (2.22). Perfect channel knowledge, on the other hand, is modeled by the assumption that the side information is equal to the channel or more precisely that $\boldsymbol{\zeta}(n) = \boldsymbol{\gamma}(\boldsymbol{H}(n))$, where $\boldsymbol{\gamma}(\cdot)$ defines a one to one mapping. In the present case of no power control, the capacity is given by (2.44) which also corresponds to the average capacity. Keep in mind that this constitutes a lower bound on the capacity when power control is allowed. Hence, the gains due to channel knowledge may for a power-controlled scenario be even larger than what will be shown next.

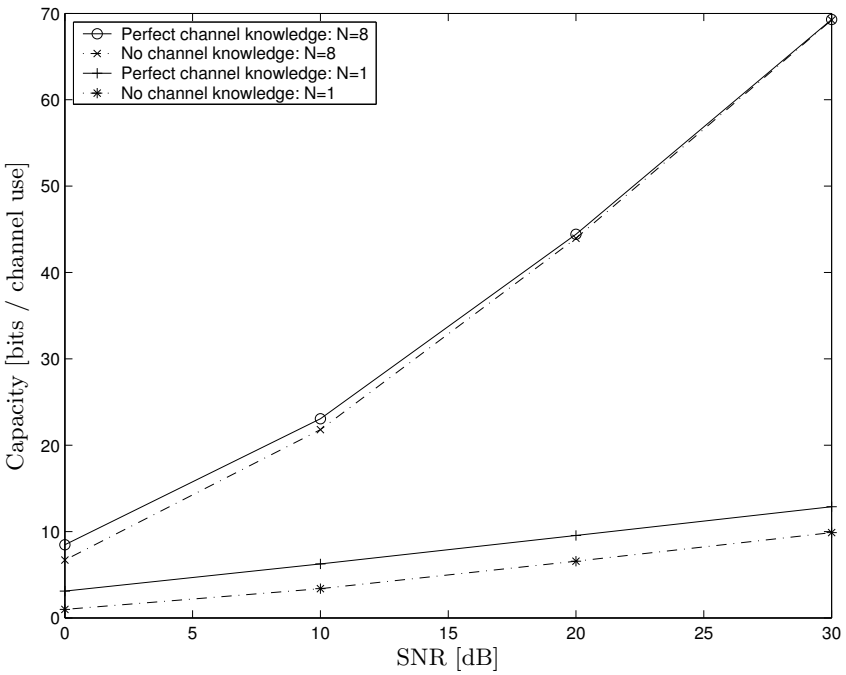

Figure 2.5: Comparing the no channel knowledge capacity in (2.22) with the perfect channel knowledge capacity in (2.44) as a function of SNR. A Rayleigh fading non-power-controlled scenario with $M = 8$ transmit antennas, $\sigma_h^2 = 1$ and $P = 1$ is assumed.

In Figure 2.5, the capacity as a function of the SNR is depicted. The number of transmit antennas is eight and the number of receive

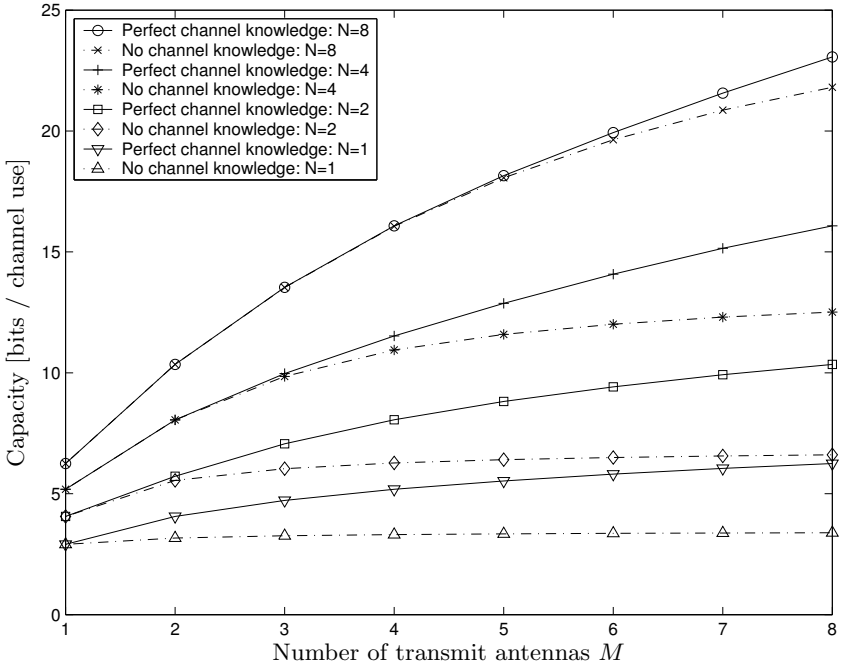

Figure 2.6: Comparing the no channel knowledge capacity in (2.22) with the perfect channel knowledge capacity in (2.44) as a function of the number of transmit antennas M. The SNR is set at 10 dB and a Rayleigh fading non-power-controlled scenario with $\sigma_h^2 = 1$ and $P = 1$ is assumed.

antennas is either one or eight. In the former case, the difference in performance between the no and perfect channel knowledge cases is seen to be substantial. This is to be expected since the rank of the channel is one and hence for each channel realization, only the power along the first eigen-direction, corresponding to the single non-zero eigenvalue, will reach the receiver. A significant fraction of the transmitted power will therefore always be wasted using the isotropic power distribution of the no channel knowledge transmitter. In contrast, the water-filling procedure used by the perfect channel knowledge transmitter ensures that all power is allocated to this single eigen-direction, as is clear from (2.34) - (2.38), and therefore wastes no power on directions which do not contribute to

the received signal.

The eight receive antenna curves demonstrate the high data rates offered by a MIMO structure with an equal number of transmit and receive antennas. Typically, all eigen-directions now contribute, more or less, to the received signal. It is therefore not surprising that the non-informed transmitter with its spatially isotropic power distribution performs almost as well as the water-filling transmitter. The performance gap is no longer constant and is even seen to vanish at high SNR values. Such a decreasing performance gap is reasonable in view of the water-filling power distribution which becomes more uniform, or more spatially isotropic, as the SNR is increased. The utility of channel knowledge is hence larger when the SNR is low than when it is high. This is a rather general phenomenon that will also be observed in later chapters in connection with performance evaluation of codes designed to take channel knowledge into account.

The impact of different antenna configurations on the performance is further illustrated in Figure 2.6. The SNR is now fixed at 10 dB and the number of transmit and receive antennas is varied. Again, it is seen that the gains due to channel knowledge are more pronounced when the difference between the number of transmit and receive antennas is large. By connecting the data points corresponding to an equal number of antennas at both sides, it is also possible to discern the linear increase in capacity with respect to $M = N$ that is typical of MIMO systems and which was reported in e.g. [FG98].

2.7.2 Quantized Channel Information

As previously mentioned, quantized channel information obtained from the receiver via a feedback link is frequently encountered in practice. Obviously, such partial channel knowledge is better than none at all while, on the other hand, the performance cannot be expected to be as good as when the channel is known perfectly. This tradeoff is to some extent investigated below for the memoryless deterministic quantized feedback scenario described in Section 2.6.2, assuming a scenario with power control. Hence, each channel realization is independently quantized into a Q-valued integer $i(n)$ by a deterministic feedback encoder function $\varepsilon(\cdot)$. The encoder output constitutes the side information, i.e., $\boldsymbol{\zeta}(n) = i(n)$. To facilitate an easy computation of the capacity, the study is limited to one receive antenna and quantization schemes that exhibit a high degree of symmetry, permitting the use of the theory of symmetric feedback. In

practice, the assumption of one receive antenna is not as restrictive as it may first seem since the terminals in many wireless systems are typically not equipped with more than one antenna. The study therefore covers the important case of communication in the downlink from for example a base station with several antennas to a single antenna terminal. In particular, the setup is closely related to the one in the closed-loop mode of the WCDMA system [3GP02b].

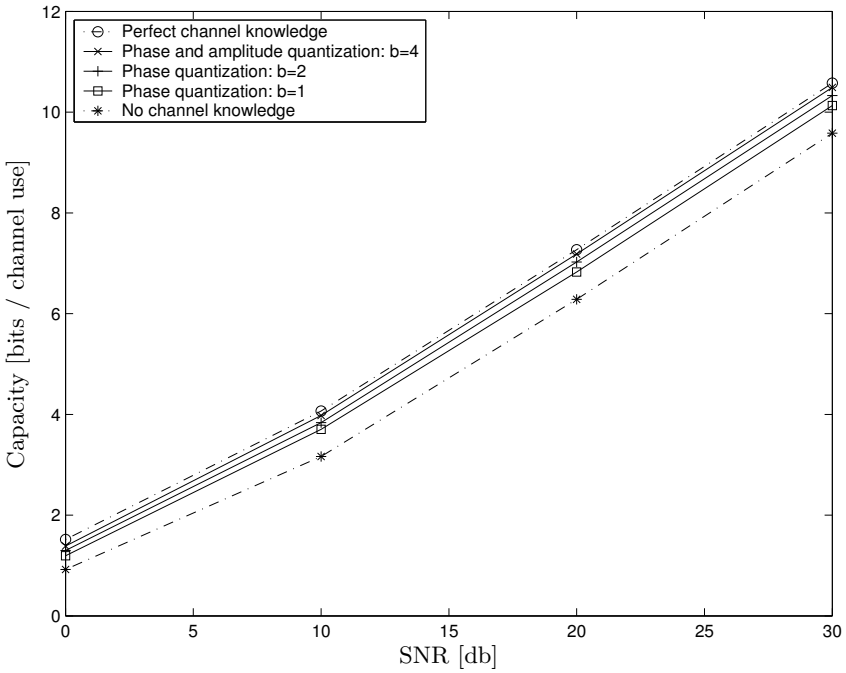

Figure 2.7: Capacity as a function of SNR for different levels of quantization. A Rayleigh fading power-controlled scenario with $M = 2$ transmit antennas, $N = 1$ receive antenna, $\sigma_h^2 = 1$ and $P = 1$ is assumed.

Consider a system with $M = 2$ transmit antennas and $N = 1$ receive antenna system that employs uniform phase quantization of $\boldsymbol{H}(n)/H_{11}(n) = H_{12}(n)/H_{11}(n)$ as defined by (2.60) and (2.61). The feedback scenario is symmetric and the capacity is therefore given by (2.63). In Figure 2.7, numerical results on the capacity as a function of the SNR are depicted for no channel knowledge, $b = 1$ and $b = 2$ bit quan-

tization. In addition, the expression for the power constraint in (2.52) was used to provide numerical values on the capacity when the channel is known perfectly. One bit quantization is seen to give a significant improvement over the no channel knowledge case. Adding another bit gives only a modest improvement which however takes the performance quite close to the limiting perfect channel knowledge case. To come even closer to this limit, also the amplitude needs to be quantized. Consider therefore a second type of quantization scheme in which the encoder is still given by (2.60) but with the encoder codebook

$$\boldsymbol{\varepsilon}_k = \begin{cases} \begin{bmatrix} 1 & 0.8e^{j\pi k/4} \end{bmatrix}^{\mathrm{T}}, & k = 0, \cdots, 7 \\ \begin{bmatrix} 1 & 1.6e^{j\pi(k-8)/4} \end{bmatrix}^{\mathrm{T}}, & k = 8, \cdots, 15\,, \end{cases}$$

which means that the relative phase is quantized using three bits while a fourth bit is used to quantize the relative amplitude. It is easily verified that the feedback is now symmetric with respect to $\{\mathcal{S}_k\}_{k=0}^{7}$ or $\{\mathcal{S}_k\}_{k=8}^{15}$, respectively. Computing the capacity is hence feasible since (2.58) shows that only two different matrices need to be optimized. The resulting $b = 4$ curve in Figure 2.7 is seen to almost coincide with the performance of the perfect channel knowledge transmitter.

Finally, phase-only quantization as described above is further investigated in Figure 2.8 illustrating how the performance is affected by increasing the number of transmit antennas M while keeping the quantization level at $\bar{b} = b/(M-1) = 2$ bits per complex dimension. Again, the no and perfect knowledge cases serve as benchmarks. Increasing M above two basically does not improve the performance when there is no channel information. On the other hand, when using the phase quantization, the capacity increases substantially with the number of transmit antennas. This agrees well with the previous results displayed in Figure 2.6. We conclude that it is possible to come quite close to the limiting perfect channel knowledge curves using only phase quantization.

2.8 Conclusions

This work considered a MIMO wireless communication link in which the transmitter has access to possibly imperfect channel side information. A formula for the channel capacity was presented and subsequently analyzed for a number of different cases.

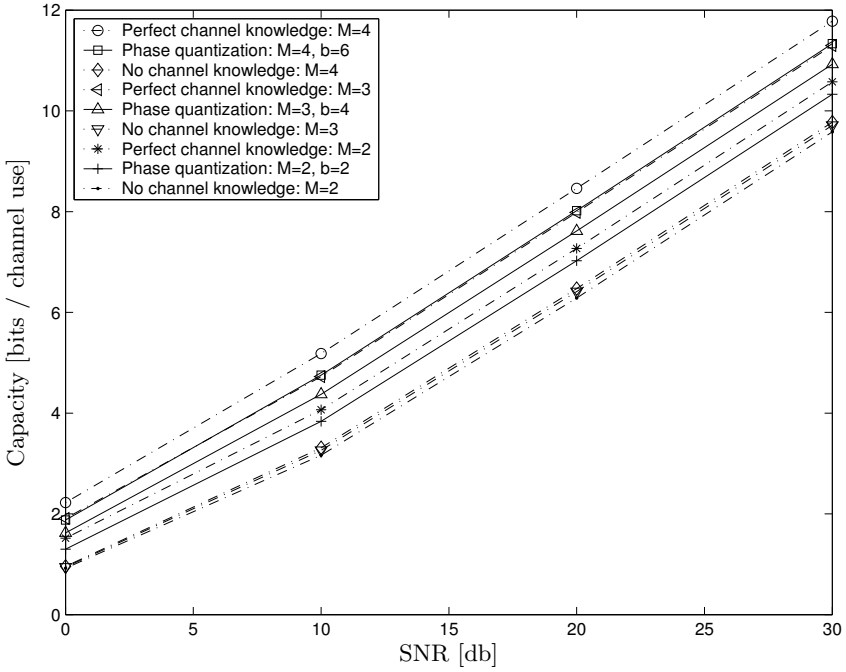

Figure 2.8: Capacity as a function of SNR for different levels of phase quantization and number of transmit antennas. A Rayleigh fading power-controlled scenario with $N = 1$ receive antenna, $\sigma_h^2 = 1$ and $P = 1$ is assumed.

One important result is that separate space-time coding and transmit weighting is an optimal transmission structure. In the case of a system with two transmit antennas and one receive antenna, it was shown that such a structure may utilize the well-known Alamouti space-time code without loss of optimality.

Although the focus was on a power-controlled scenario, it was also pointed out that the presented capacity expression is easy to modify to encompass also the case of a power constraint preventing the use of power control. An interesting side effect of this is that previously used average capacity or expected mutual information based performance measures indeed correspond to true channel capacities.

Methods for evaluating the often numerically challenging capacity ex-

pression were developed. In particular, a theory concerned with symmetrical properties of typical channel feedback schemes was introduced and used for simplifying numerical computation of the capacity. Numerical results on the channel capacity demonstrated the importance of exploiting channel side information.

Appendix 2.A Proving Optimality of Block Diagonal Structure

In this appendix, it is proved that $\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})$ in the capacity formula (2.13) may be taken to be block diagonal without loss of optimality.

The proof proceeds as follows. Partition $\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})$ into L^2 blocks according to

$$\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta}) = \begin{bmatrix} \boldsymbol{\mathcal{Z}}_{11}(\boldsymbol{\zeta}) & \cdots & \boldsymbol{\mathcal{Z}}_{1L}(\boldsymbol{\zeta}) \\ \vdots & & \vdots \\ \boldsymbol{\mathcal{Z}}_{L1}(\boldsymbol{\zeta}) & \cdots & \boldsymbol{\mathcal{Z}}_{LL}(\boldsymbol{\zeta}) \end{bmatrix},$$

where each block is represented by the $M \times M$ sub-matrix $\boldsymbol{\mathcal{Z}}_{kl}(\boldsymbol{\zeta})$. Based on the above partition of $\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})$, it is possible to write

$$\begin{aligned} &(\boldsymbol{I}_L \otimes \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2}\boldsymbol{H})\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})(\boldsymbol{I}_L \otimes \boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^*) \\ &= \begin{bmatrix} \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2}\boldsymbol{H}\boldsymbol{\mathcal{Z}}_{11}(\boldsymbol{\zeta})\boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^* & \cdots & \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2}\boldsymbol{H}\boldsymbol{\mathcal{Z}}_{1L}(\boldsymbol{\zeta})\boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^* \\ \vdots & & \vdots \\ \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2}\boldsymbol{H}\boldsymbol{\mathcal{Z}}_{L1}(\boldsymbol{\zeta})\boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^* & \cdots & \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2}\boldsymbol{H}\boldsymbol{\mathcal{Z}}_{LL}(\boldsymbol{\zeta})\boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^* \end{bmatrix}. \end{aligned}$$

Consider the determinant in (2.13) and use the relation (C.5) in Appendix C and Fisher's inequality [HJ96, p. 478] to obtain the upper bound

$$\begin{aligned} &\det(\boldsymbol{I}_{LN} + (\boldsymbol{I}_L \otimes \boldsymbol{R}_{\boldsymbol{ee}}^{-1})(\boldsymbol{I}_L \otimes \boldsymbol{H})\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})(\boldsymbol{I}_L \otimes \boldsymbol{H}^*)) \\ &\quad = \det(\boldsymbol{I}_{LN} + (\boldsymbol{I}_L \otimes \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2}\boldsymbol{H})\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})(\boldsymbol{I}_L \otimes \boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^*)) \\ &\quad \leq \prod_{k=1}^{L} \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2}\boldsymbol{H}\boldsymbol{\mathcal{Z}}_{kk}(\boldsymbol{\zeta})\boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^*), \end{aligned} \tag{2.64}$$

which holds with equality if all sub-matrices $\boldsymbol{\mathcal{Z}}_{kl}(\boldsymbol{\zeta})$ that are not on the main diagonal are zero. Utilize (2.64) and the fact that $\log_2 \det(\cdot)$ is a concave function to write

$$\begin{aligned} &\mathrm{E}[\det(\boldsymbol{I}_{LN} + (\boldsymbol{I}_L \otimes \boldsymbol{R}_{\boldsymbol{ee}}^{-1})(\boldsymbol{I}_L \otimes \boldsymbol{H})\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})(\boldsymbol{I}_L \otimes \boldsymbol{H}^*))] \\ &\quad \leq \mathrm{E}\Big[\sum_{k=1}^{L} \log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2}\boldsymbol{H}\boldsymbol{\mathcal{Z}}_{kk}(\boldsymbol{\zeta})\boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^*)\Big] \\ &\quad \leq L\,\mathrm{E}\Big[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2}\boldsymbol{H}\frac{\sum_{k=1}^{L}\boldsymbol{\mathcal{Z}}_{kk}(\boldsymbol{\zeta})}{L}\boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^*)\Big], \end{aligned} \tag{2.65}$$

where last step follows from the definition of a concave function. The second inequality holds with equality if the $\boldsymbol{\mathcal{Z}}_{kk}(\boldsymbol{\zeta})$'s are all equal. Furthermore, note that the power constraint in (2.13) can be written as

$$\mathrm{E}[\mathrm{tr}(\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta}))] = \mathrm{E}\left[\mathrm{tr}\left(\sum_{k=1}^{L} \boldsymbol{\mathcal{Z}}_{kk}(\boldsymbol{\zeta})\right)\right] = P_{\mathrm{bm}}\,. \tag{2.66}$$

From (2.65) and (2.66) it is now seen that an upper bound on the capacity in (2.13) is given by

$$\begin{aligned} C \leq & \max_{\substack{\{\boldsymbol{\mathcal{Z}}_{kk}(\cdot)\} \\ \boldsymbol{\mathcal{Z}}_{kk}(\cdot)=\boldsymbol{\mathcal{Z}}_{kk}(\cdot)^* \succeq \mathbf{0},\ \forall k \\ \mathrm{E}\left[\mathrm{tr}\left(\sum_{k=1}^{L} \boldsymbol{\mathcal{Z}}_{kk}(\boldsymbol{\zeta})\right)\right]=P_{\mathrm{bm}}}} L\,\mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2} \boldsymbol{H} \frac{\sum_k \boldsymbol{\mathcal{Z}}_{kk}(\boldsymbol{\zeta})}{L} \boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^*)] \\ = & \max_{\substack{\boldsymbol{Z}(\cdot) \\ \boldsymbol{Z}(\cdot)=\boldsymbol{Z}(\cdot)^* \succeq \mathbf{0} \\ L\,\mathrm{E}[\mathrm{tr}(\boldsymbol{Z}(\boldsymbol{\zeta}))]=P_{\mathrm{bm}}}} L\,\mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1/2} \boldsymbol{H} \boldsymbol{Z}(\boldsymbol{\zeta}) \boldsymbol{H}^*(\boldsymbol{R}_{\boldsymbol{ee}}^{-1/2})^*)]\,, \end{aligned} \tag{2.67}$$

where (2.67) follows from reformulating the optimization problem in terms of the parametrization $\boldsymbol{Z}(\boldsymbol{\zeta}) \triangleq \sum_k \boldsymbol{\mathcal{Z}}_{kk}(\boldsymbol{\zeta})/L$. However, the above inequality holds with equality since (2.67) may be obtained directly from the capacity formula in (2.13) if $\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})$ is constrained as $\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta}) = \boldsymbol{I}_L \otimes \boldsymbol{Z}(\boldsymbol{\zeta})$. Hence, (2.67) gives the capacity and we have shown that $\boldsymbol{\mathcal{Z}}(\boldsymbol{\zeta})$ may be taken to be block diagonal without loss of optimality, as asserted.

Appendix 2.B Proving Optimality of Diagonal Structure

In this appendix, the optimality of the diagonal structure in (2.27) will be proved. Although somewhat shortened, the development proceeds in a similar manner as for the proof concerned with the block diagonal structure in Appendix 2.A.

Let $[\boldsymbol{A}]_{kl}$ denote the element on the kth row and lth column of the matrix $\boldsymbol{A}$ and use this to define $\tilde{Z}_{kk}(\boldsymbol{H}) \triangleq [\tilde{\boldsymbol{Z}}(\boldsymbol{H})]_{kk}$. Applying Hadamard's inequality [HJ96, p. 477] to the determinant gives

$$\det(\boldsymbol{I}_N + \boldsymbol{\Lambda}(\boldsymbol{H})\tilde{\boldsymbol{Z}}(\boldsymbol{H})) \leq \prod_{k=1}^{r} (1 + [\boldsymbol{\Lambda}(\boldsymbol{H})\tilde{\boldsymbol{Z}}(\boldsymbol{H})]_{kk})$$

$$= \prod_{k=1}^{r} (1 + [\boldsymbol{\Lambda}(\boldsymbol{H})]_{kk} [\tilde{\boldsymbol{Z}}(\boldsymbol{H})]_{kk})$$
$$= \prod_{k=1}^{r} (1 + \lambda_k(\boldsymbol{H}) \tilde{Z}_{kk}(\boldsymbol{H})) \,, \tag{2.68}$$

with obvious notation. The above upper bound is attained if $\tilde{\boldsymbol{Z}}(\boldsymbol{H})$ is diagonal. Furthermore, note that

$$\mathrm{E}[\mathrm{tr}(\tilde{\boldsymbol{Z}}(\boldsymbol{H}))] = \sum_{k=1}^{M} \mathrm{E}[\tilde{Z}_{kk}(\boldsymbol{H})] = P \,,$$

which means that also the power constraint can be written in terms of the diagonal elements of $\tilde{\boldsymbol{Z}}(\boldsymbol{H})$. Since (2.68) is attained if $\tilde{\boldsymbol{Z}}(\boldsymbol{H})$ is diagonal and both the upper bound and the power constraint can be written in terms of the corresponding diagonal elements, $\tilde{\boldsymbol{Z}}(\boldsymbol{H})$ may be assumed to be diagonal without loss of optimality. This concludes the proof.

Appendix 2.C Power Constraint in Closed-Form

In this appendix, we prove that the power constraint in (2.38) can, assuming $N = 1$ receive antenna and spatially IID Rayleigh fading as in Section 2.6.1, be written in closed-form to arrive at (2.52).

Note first that $x \triangleq 2\|\boldsymbol{H}\|^2/\sigma_h^2$ is a chi-squared random variable with $2M$ degrees of freedom. Hence, its PDF is given by [Kay98, p. 24]

$$p_x(x) = \begin{cases} \frac{x^{M-1} \exp(-x/2)}{2^M (M-1)!}, & x \geq 0 \\ 0, & x < 0 \,. \end{cases}$$

Let $\alpha \triangleq \sigma^2/\sigma_h^2$ and evaluate the expectation in the power constraint as

$$\begin{aligned} \mathrm{E}_{\boldsymbol{H}}[\max\{1/\mu - \sigma^2/\|\boldsymbol{H}\|^2, 0\}] &= \mathrm{E}_x[\max\{1/\mu - 2\sigma^2/(\sigma_h^2 x), 0\}] \\ &= \int \max\{1/\mu - 2\alpha/x, 0\} p_x(x)\, dx \\ &= \int_{2\alpha\mu}^{\infty} (1/\mu - 2\alpha/x)\, p_x(x)\, dx \\ &= \frac{1}{\mu} \int_{2\alpha\mu}^{\infty} p_x(x)\, dx - 2\alpha \int_{2\alpha\mu}^{\infty} p_x(x)/x\, dx \,. \end{aligned}$$

Since

$$\int x^n e^{ax}\,dx = \frac{(-1)^n \exp(ax) n!}{a^{n+1}} \sum_{k=0}^{n} \frac{(ax)^k (-1)^k}{k!}$$

it follows that

$$\begin{aligned}\int_{2\alpha\mu}^{\infty} p_x(x)\,dx &= \frac{1}{2^M (M-1)!} \int_{2\alpha\mu}^{\infty} x^{M-1} \exp(-x/2)\,dx \\ &= \left[\frac{(-1)^{M-1} \exp(-x/2)(M-1)!}{2^M (M-1)!(-1/2)^M} \sum_{k=0}^{M-1} \frac{(-x/2)^k (-1)^k}{k!} \right]_{2\alpha\mu}^{\infty} \\ &= \exp(-\alpha\mu) \sum_{k=0}^{M-1} \frac{(\alpha\mu)^k}{k!}\end{aligned}$$

and

$$\begin{aligned}\int_{2\alpha\mu}^{\infty} p_x(x)/x\,dx &= \frac{1}{2^M (M-1)!} \int_{2\alpha\mu}^{\infty} x^{M-2} \exp(-x/2)\,dx \\ &= \left[\frac{(-1)^{M-2} \exp(-x/2)(M-2)!}{2^M (M-1)!(-1/2)^{M-1}} \sum_{k=0}^{M-2} \frac{(-x/2)^k (-1)^k}{k!} \right]_{2\alpha\mu}^{\infty} \\ &= \frac{\exp(-\alpha\mu)}{2(M-1)} \sum_{k=0}^{M-2} \frac{(\alpha\mu)^k}{k!}\,.\end{aligned}$$

Hence, the desired expression in (2.52) follows as

$$\begin{aligned}&\mathrm{E}_{\boldsymbol{H}}[\max\{1/\mu - \sigma^2/\|\boldsymbol{H}\|^2, 0\}] \\ &\qquad = \frac{\exp(-\alpha\mu)}{\mu} \sum_{k=0}^{M-1} \frac{(\alpha\mu)^k}{k!} - \alpha \frac{\exp(-\alpha\mu)}{M-1} \sum_{k=0}^{M-2} \frac{(\alpha\mu)^k}{k!}\,.\end{aligned}$$

Appendix 2.D Symmetric Feedback

In this appendix, we derive how the symmetric feedback conditions in Section 2.6.3 affect the capacity formula. Hence, both the first and second condition are tacitly assumed to be satisfied for the remainder of this section.

2.D.1 Encoder Probabilities and Conditional Expectation

The expressions in (2.56), (2.57) for the feedback encoder probabilities and the conditional expectation, respectively, will be proved below. The proof relies on symmetries induced by the symmetric feedback conditions. To start with, note that

$$\begin{aligned} p_{\kappa^{(q)}(0)} &= \Pr[\boldsymbol{H} \in \mathcal{S}_{\kappa^{(q)}(0)}] \\ &= \int_{\boldsymbol{H} \in \mathcal{S}_{\kappa^{(q)}(0)}} p_{\boldsymbol{H}}(\boldsymbol{H})\, d\boldsymbol{H}\,, \end{aligned}$$

where the notation "$d\boldsymbol{H}$" means that the integration is to be carried out with respect to the $2MN$ real-valued parameters of $\boldsymbol{H}$. Now utilize the first condition and make a change of variables to obtain

$$\begin{aligned} p_{\kappa^{(q)}(0)} &= \int_{\boldsymbol{H}\boldsymbol{Q}_k^{(q)} \in \mathcal{S}_{\kappa^{(q)}(k)}} p_{\boldsymbol{H}}(\boldsymbol{H})\, d\boldsymbol{H} \\ &= \int_{\tilde{\boldsymbol{H}} \in \mathcal{S}_{\kappa^{(q)}(k)}} p_{\boldsymbol{H}}\big(\tilde{\boldsymbol{H}}(\boldsymbol{Q}_k^{(q)})^*\big)\, d\tilde{\boldsymbol{H}}\,, \end{aligned}$$

where the second equality follows from the fact that the determinant of the Jacobian of a unitary linear transformation is one. The second condition finally gives

$$\begin{aligned} p_{\kappa^{(q)}(0)} &= \int_{\tilde{\boldsymbol{H}} \in \mathcal{S}_{\kappa^{(q)}(k)}} p_{\boldsymbol{H}}(\tilde{\boldsymbol{H}})\, d\tilde{\boldsymbol{H}} \\ &= \Pr[\boldsymbol{H} \in \mathcal{S}_{\kappa^{(q)}(k)}] \\ &= p_{\kappa^{(q)}(k)}\,, \end{aligned}$$

which shows that all encoder output probabilities belonging to the qth subset are equal, as claimed in (2.56).

Let $p_{\boldsymbol{H}|i}(\boldsymbol{H}|i = \kappa^{(q)}(k))$ denote the PDF of $\boldsymbol{H}$ conditioned on the event that $i = \kappa^{(q)}(k)$, which corresponds to $\boldsymbol{H} \in \mathcal{S}_{\kappa^{(q)}(k)}$. To prove the result on the conditional expectation, first note that

$$p_{\boldsymbol{H}|i}(\boldsymbol{H}|i = \kappa^{(q)}(k)) = \begin{cases} \frac{p_{\boldsymbol{H}}(\boldsymbol{H})}{p_{\kappa^{(q)}(k)}}, & \boldsymbol{H} \in \mathcal{S}_{\kappa^{(q)}(k)} \\ 0, & \text{otherwise} \end{cases}$$

and hence, because of the unitary invariance $p_{\boldsymbol{H}}(\boldsymbol{HU}) = p_{\boldsymbol{H}}(\boldsymbol{H})$, the first condition and $p_{\kappa^{(q)}(0)} = p_{\kappa^{(q)}(k)}$,

$$\begin{aligned} p_{\boldsymbol{H}|i}(\boldsymbol{H}|i = \kappa^{(q)}(0)) &= \begin{cases} \frac{p_{\boldsymbol{H}}(\boldsymbol{HQ}_k^{(q)})}{p_{\kappa^{(q)}(0)}}, & \boldsymbol{H} \in \mathcal{S}_{\kappa^{(q)}(0)} \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{p_{\boldsymbol{H}}(\boldsymbol{HQ}_k^{(q)})}{p_{\kappa^{(q)}(k)}}, & \boldsymbol{HQ}_k^{(q)} \in \mathcal{S}_{\kappa^{(q)}(k)} \\ 0, & \text{otherwise} \end{cases} \\ &= p_{\boldsymbol{H}|i}(\boldsymbol{HQ}_k^{(q)}|i = \kappa^{(q)}(k))\,. \end{aligned}$$

From this the desired relation in (2.57) follows as

$$\begin{aligned} &\mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{HXH}^*)|i = \kappa^{(q)}(k)] \\ &\quad = \int \log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{HXH}^*) p_{\boldsymbol{H}|i}(\boldsymbol{H}|i = \kappa^{(q)}(k))\, d\boldsymbol{H} \\ &= \int \log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\tilde{\boldsymbol{H}}\boldsymbol{Q}_k^{(q)}\boldsymbol{X}(\boldsymbol{Q}_k^{(q)})^*\tilde{\boldsymbol{H}}^*) p_{\boldsymbol{H}|i}(\tilde{\boldsymbol{H}}\boldsymbol{Q}_k^{(q)}|i = \kappa^{(q)}(k))\, d\tilde{\boldsymbol{H}} \\ &= \int \log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\tilde{\boldsymbol{H}}\boldsymbol{Q}_k^{(q)}\boldsymbol{X}(\boldsymbol{Q}_k^{(q)})^*\tilde{\boldsymbol{H}}^*) p_{\boldsymbol{H}|i}(\tilde{\boldsymbol{H}}|i = \kappa^{(q)}(0))\, d\tilde{\boldsymbol{H}} \\ &\quad = \mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{HQ}_k^{(q)}\boldsymbol{X}(\boldsymbol{Q}_k^{(q)})^*\boldsymbol{H}^*)|i = \kappa^{(q)}(0)]\,. \end{aligned}$$

2.D.2 Simplifying the Optimization Problem

This section presents a proof that the capacity formula in (2.53) simplifies to (2.58). By grouping terms that correspond to the same subset $\mathcal{S}_{\mathrm{s}}^{(q)}$ together and utilizing (2.56) and (2.57), the power constraint can be rewritten as

$$\begin{aligned} \sum_{k=0}^{Q-1} p_k \operatorname{tr}(\boldsymbol{Z}_k) &= \sum_{q=0}^{Q'-1} \sum_{k \in \mathcal{I}_{\mathrm{s}}^{(q)}} p_k \operatorname{tr}(\boldsymbol{Z}_k) \\ &= \sum_{q=0}^{Q'-1} \sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1} p_{\kappa^{(q)}(k)} \operatorname{tr}(\boldsymbol{Z}_{\kappa^{(q)}(k)}) \\ &= \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)} \sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1} \operatorname{tr}(\boldsymbol{Z}_{\kappa^{(q)}(k)}) = P \end{aligned}$$

and the criterion function as

$$\begin{aligned}
&\sum_{k=0}^{Q-1} p_k \,\mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}_k\boldsymbol{H}^*)|\boldsymbol{H}\in\mathcal{S}_k] \\
&\qquad = \sum_{q=0}^{Q'-1} \sum_{k\in\mathcal{I}_{\mathrm{s}}^{(q)}} p_k \,\mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}_k\boldsymbol{H}^*)|i=k] \\
&= \sum_{q=0}^{Q'-1} \sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1} p_{\kappa^{(q)}(k)} \,\mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{Z}_{\kappa^{(q)}(k)}\boldsymbol{H}^*)|i=\kappa^{(q)}(k)] \\
&\qquad = \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)} \sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1} f^{(q)}\big(\boldsymbol{Q}_k^{(q)}\boldsymbol{Z}_{\kappa^{(q)}(k)}(\boldsymbol{Q}_k^{(q)})^*\big)\,,
\end{aligned}$$

where

$$f^{(q)}(\boldsymbol{X}) \triangleq \mathrm{E}[\log_2 \det(\boldsymbol{I}_N + \boldsymbol{R}_{\boldsymbol{ee}}^{-1}\boldsymbol{H}\boldsymbol{X}\boldsymbol{H}^*)|i=\kappa^{(q)}(0)]\,.$$

Now re-parameterize the problem using $\tilde{\boldsymbol{Z}}_k^{(q)} \triangleq \boldsymbol{Q}_k^{(q)}\boldsymbol{Z}_{\kappa^{(q)}(k)}(\boldsymbol{Q}_k^{(q)})^*$. Since $\boldsymbol{Q}_k^{(q)}$ is unitary, $\mathrm{tr}(\boldsymbol{Z}_{\kappa^{(q)}(k)}) = \mathrm{tr}((\boldsymbol{Q}_k^{(q)})^*\tilde{\boldsymbol{Z}}_k^{(q)}\boldsymbol{Q}_k^{(q)}) = \mathrm{tr}(\tilde{\boldsymbol{Z}}_k^{(q)})$ and the power constraint becomes

$$\sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)} \sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1} \mathrm{tr}(\tilde{\boldsymbol{Z}}_k^{(q)}) = P\,.$$

The resulting capacity is hence

$$C = \max_{\substack{\{\tilde{\boldsymbol{Z}}_k^{(q)}\} \\ \tilde{\boldsymbol{Z}}_k^{(q)}=(\tilde{\boldsymbol{Z}}_k^{(q)})^*\succeq\boldsymbol{0},\ \forall k,q \\ \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)} \sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1} \mathrm{tr}(\tilde{\boldsymbol{Z}}_k^{(q)})=P}} \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)} \sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1} f^{(q)}\big(\tilde{\boldsymbol{Z}}_k^{(q)}\big)\,. \quad (2.69)$$

Note that $f^{(q)}(\boldsymbol{X})$ is a concave function of the set of positive semi definite matrices. Hence, we have

$$\sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1} f^{(q)}\big(\tilde{\boldsymbol{Z}}_k^{(q)}\big) \le |\mathcal{I}_{\mathrm{s}}^{(q)}| f^{(q)}\left(\frac{\sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1}\tilde{\boldsymbol{Z}}_k^{(q)}}{|\mathcal{I}_{\mathrm{s}}^{(q)}|}\right),$$

which holds with equality if the $\tilde{\boldsymbol{Z}}_k^{(q)}$'s are equal for all $k \in \mathcal{I}_{\mathrm{s}}^{(q)}$. This means that an upper bound on the capacity in (2.69) is given by

$$\begin{aligned} C \leq & \max_{\substack{\{\tilde{\boldsymbol{Z}}_k^{(q)}\} \\ \tilde{\boldsymbol{Z}}_k^{(q)}=(\tilde{\boldsymbol{Z}}_k^{(q)})^* \succeq \mathbf{0},\ \forall k,q \\ \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)} \sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1} \mathrm{tr}(\tilde{\boldsymbol{Z}}_k^{(q)})=P}} \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)} |\mathcal{I}_{\mathrm{s}}^{(q)}| f^{(q)}\left(\frac{\sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1} \tilde{\boldsymbol{Z}}_k^{(q)}}{|\mathcal{I}_{\mathrm{s}}^{(q)}|} \right) \\ = & \max_{\substack{\{\tilde{\boldsymbol{Z}}^{(q)}\}_{q=0}^{Q'-1} \\ \tilde{\boldsymbol{Z}}^{(q)}=(\tilde{\boldsymbol{Z}}^{(q)})^* \succeq \mathbf{0},\ \forall q \\ \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)} |\mathcal{I}_{\mathrm{s}}^{(q)}| \,\mathrm{tr}(\tilde{\boldsymbol{Z}}^{(q)})=P}} \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)} |\mathcal{I}_{\mathrm{s}}^{(q)}| f^{(q)}\left(\tilde{\boldsymbol{Z}}^{(q)}\right), \end{aligned} \tag{2.70}$$

where the equality is due to a straightforward re-parametrization based on $\tilde{\boldsymbol{Z}}^{(q)} \triangleq \sum_{k=0}^{|\mathcal{I}_{\mathrm{s}}^{(q)}|-1} \tilde{\boldsymbol{Z}}_k^{(q)} / |\mathcal{I}_{\mathrm{s}}^{(q)}|$. Clearly, if we for all $k \in \mathcal{I}_{\mathrm{s}}^{(q)}$ constrain $\tilde{\boldsymbol{Z}}_k^{(q)}$ in the capacity formula in (2.69) to be equal to $\tilde{\boldsymbol{Z}}^{(q)}$, the result is (2.70). In other words, the upper bound is attainable and the capacity can therefore be computed according to

$$C = \max_{\substack{\{\tilde{\boldsymbol{Z}}^{(q)}\}_{q=0}^{Q'-1} \\ \tilde{\boldsymbol{Z}}^{(q)}=(\tilde{\boldsymbol{Z}}^{(q)})^* \succeq \mathbf{0},\ \forall q \\ \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)} |\mathcal{I}_{\mathrm{s}}^{(q)}| \,\mathrm{tr}(\tilde{\boldsymbol{Z}}^{(q)})=P}} \sum_{q=0}^{Q'-1} p_{\kappa^{(q)}(0)} |\mathcal{I}_{\mathrm{s}}^{(q)}| f^{(q)}\left(\tilde{\boldsymbol{Z}}^{(q)}\right). \tag{2.71}$$

An optimal solution of the original problem in (2.69) is given by $\tilde{\boldsymbol{Z}}_k^{(q)} = \tilde{\boldsymbol{Z}}^{(q)}$, $\forall k$, where $\tilde{\boldsymbol{Z}}^{(q)}$ is the maximizing argument of (2.71). This concludes the proof.

2.D.3 Proving Symmetry of Feedback in the Example Scenario

In this appendix it is proved that the uniform phase quantization scheme in (2.60), under the assumptions of one receive antenna and spatially IID Rayleigh fading channel coefficients, results in a symmetric feedback scenario. In other words, we will show that the feedback is symmetric with respect to *all* encoder regions. This will be accomplished by showing that the two feedback conditions are satisfied for $\mathcal{S}_{\mathrm{s}}^{(0)} = \{\mathcal{S}_k\}_{k=0}^{Q-1}$, $\boldsymbol{Q}_k^{(0)} =$

$\boldsymbol{Q}_k$, with $\boldsymbol{Q}_k$ defined as in (2.62) and the renumbering function $\kappa^{(0)}(k) = k$.

Since the channel coefficients are zero-mean IID and each channel coefficient is complex Gaussian distributed with variance σ_h^2, the PDF $p_{\boldsymbol{H}}(\boldsymbol{H})$ of the channel is given by [Kay93, p. 507]

$$p_{\boldsymbol{H}}(\boldsymbol{H}) = \frac{\exp\big(-\sum_{k=1}^{M} |H_{1k}|^2/\sigma_h^2\big)}{\pi^M \sigma_h^{2M}} = \frac{\exp(-\boldsymbol{H}\boldsymbol{H}^*/\sigma_h^2)}{\pi^M \sigma_h^{2M}}\,.$$

Note that $(\boldsymbol{H}\boldsymbol{U})(\boldsymbol{H}\boldsymbol{U})^* = \boldsymbol{H}\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{H}^* = \boldsymbol{H}\boldsymbol{H}^*$ for an arbitrary unitary matrix $\boldsymbol{U}$. Hence, $p_{\boldsymbol{H}}(\boldsymbol{H}\boldsymbol{U}) = p_{\boldsymbol{H}}(\boldsymbol{H})$ and the second feedback condition is seen to be satisfied.

To see that also the first feedback condition is satisfied, assume that $\boldsymbol{H} \in \mathcal{S}_{\kappa^{(q)}(0)} = \mathcal{S}_0$. That is, expressed using the encoder function in (2.60), it holds that

$$\varepsilon(\boldsymbol{H}) = \arg\min_l \|\boldsymbol{H}^{\mathrm{T}}/H_{11} - \boldsymbol{Q}_l\boldsymbol{\varepsilon}_0\|^2 = 0\,, \tag{2.72}$$

where $\boldsymbol{Q}_l\boldsymbol{\varepsilon}_0 = \boldsymbol{\varepsilon}_l$. The output of the encoder with $\boldsymbol{H}\boldsymbol{Q}_k^{(0)} = \boldsymbol{H}\boldsymbol{Q}_k$ as input, is then

$$\begin{aligned}
\varepsilon(\boldsymbol{H}\boldsymbol{Q}_k) &= \arg\min_l \|\boldsymbol{Q}_k^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}/H_{11} - \boldsymbol{Q}_l\boldsymbol{\varepsilon}_0\|^2 \\
&= \arg\min_l \|\boldsymbol{Q}_k\boldsymbol{H}^{\mathrm{T}}/H_{11} - \boldsymbol{Q}_l\boldsymbol{\varepsilon}_0\|^2 \\
&= \arg\min_l \|\boldsymbol{Q}_k(\boldsymbol{H}^{\mathrm{T}}/H_{11} - \boldsymbol{Q}_k^*\boldsymbol{Q}_l\boldsymbol{\varepsilon}_0)\|^2 \\
&= \arg\min_l \|\boldsymbol{H}^{\mathrm{T}}/H_{11} - \boldsymbol{Q}_k^*\boldsymbol{Q}_l\boldsymbol{\varepsilon}_0\|^2\,,
\end{aligned}$$

where the last inequality is due to the fact that the Euclidean vector norm is invariant to a unitary linear transformation. Because of the uniform phase quantization, it is realized that $\boldsymbol{Q}_k^*\boldsymbol{Q}_l \in \{\boldsymbol{Q}_m\}_{m=0}^{Q-1}$. Hence, $\boldsymbol{Q}_k^*\boldsymbol{Q}_l$ represents an arbitrary $\boldsymbol{Q}_m$ and, from $\boldsymbol{Q}_k^*\boldsymbol{Q}_k = \boldsymbol{I}_M = \boldsymbol{Q}_0$ in conjunction with (2.72), it follows that

$$\varepsilon(\boldsymbol{H}\boldsymbol{Q}_k) = \arg\min_l \|\boldsymbol{H}^{\mathrm{T}}/H_{11} - \boldsymbol{Q}_k^*\boldsymbol{Q}_l\boldsymbol{\varepsilon}_0\|^2 = k\,.$$

This shows that $\boldsymbol{H}\boldsymbol{Q}_k = \boldsymbol{H}\boldsymbol{Q}_k^{(0)} \in \mathcal{S}_k = \mathcal{S}_{\kappa^{(q)}(k)}$ and consequently that $\boldsymbol{H} \in \mathcal{S}_{\kappa^{(q)}(0)} \Rightarrow \boldsymbol{H}\boldsymbol{Q}_k^{(0)} \in \mathcal{S}_{\kappa^{(q)}(k)}$. To prove the converse, assume $\boldsymbol{H}\boldsymbol{Q}_k^{(0)} = \boldsymbol{H}\boldsymbol{Q}_k \in \mathcal{S}_{\kappa^{(q)}(k)} = \mathcal{S}_k$, i.e.,

$$\varepsilon(\boldsymbol{H}\boldsymbol{Q}_k) = \arg\min_l \|\boldsymbol{Q}_k^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}/H_{11} - \boldsymbol{Q}_l\boldsymbol{\varepsilon}_0\|^2 = k\,. \tag{2.73}$$

The encoder output with $\boldsymbol{H}$ as input is then

$$\begin{aligned}\varepsilon(\boldsymbol{H}) &= \arg\min_l \|\boldsymbol{H}^{\mathrm{T}}/H_{11} - \boldsymbol{Q}_l\boldsymbol{\varepsilon}_0\|^2 \\ &= \arg\min_l \|\boldsymbol{Q}_k(\boldsymbol{H}^{\mathrm{T}}/H_{11} - \boldsymbol{Q}_l\boldsymbol{\varepsilon}_0)\|^2 \\ &= \arg\min_l \|\boldsymbol{Q}_k^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}/H_{11} - \boldsymbol{Q}_k\boldsymbol{Q}_l\boldsymbol{\varepsilon}_0)\|^2 = 0\,,\end{aligned}$$

where the last equality follows from $\boldsymbol{Q}_k\boldsymbol{Q}_l \in \{\boldsymbol{Q}_m\}_{m=0}^{Q-1}$ and $\boldsymbol{Q}_k\boldsymbol{Q}_0 = \boldsymbol{Q}_k$ in conjunction with (2.73). Consequently, $\boldsymbol{H} \in \mathcal{S}_0 = \mathcal{S}_{\kappa^{(q)}(0)}$ and the converse is proved. Thus, $\boldsymbol{H} \in \mathcal{S}_{\kappa^{(q)}(0)} \Leftrightarrow \boldsymbol{H}\boldsymbol{Q}_k^{(0)} \in \mathcal{S}_{\kappa^{(q)}(k)}$, meaning that also the first feedback condition is satisfied.

Chapter 3

System Description and Preliminaries

The remainder of the thesis is concerned with code and transmission scheme design for transmitters with possibly imperfect channel side information. The design methods are to a large extent developed based on the same system model. The present chapter describes the parts of the system model that all the design methods have in common. This *generic* system model will in later chapters be supplemented with additional assumptions for the respective case under study.

In the following, essentially a special case of the flat block fading scenario in Section 2.3.2 is considered. A data model is presented and three different classes of channel side information dependent space-time block codes are described for later reference. The code classes, mentioned in order of increasingly restrictive structure, are unstructured codes, linear dispersive codes and weighted OSTBC. While the first two code types are simple extensions of structures previously known in the literature, weighted OSTBC was originally proposed in the works that form this thesis and hence represents a new code/transmission structure.

3.1 A Generic System Model

Consider a MIMO wireless communication system in a flat block fading scenario with channel side information at the transmitter under assumptions that essentially correspond to a special case of the flat block fading

scenario in Section 2.3.2. Hence, both the channel and the side information is constant during a block of L samples but may vary from one block to another in a statistically stationary fashion. However, in contrast to the previous scenario, the channel coding is now assumed to be performed independently for each block. More specifically, the length of the codewords in the channel code is equal to the block length L and each transmitted codeword experiences only one channel realization.

A typical benefit of such type of channel coding is that the processing delay incurred by the coding scheme is relatively short, since the delay is on the same order of magnitude as the block length L. This suits delay sensitive applications such as voice traffic or video conferencing. On the other hand, a drawback is that the temporal variations of the channel are not exploited for achieving time diversity. If needed however, time diversity can be achieved by interleaving and employing an outer code, at the expense of increased processing delay.

Because the channel coding is performed independently for each block, we may restrict our attention to a single arbitrary block, without loss of generality. In the following, the time-interval $n = 0, 1, \cdots, L-1$ is therefore considered and the block index is dropped.

A symbol sampled complex baseband equivalent model of the system is depicted in Figure 3.1. At the transmitter, a space-time encoder maps bits, representing the message to be transmitted, into a sequence of codeword matrices. The result is a set of parallel and generally different channel symbol streams. Each symbol stream corresponds to the signal for a particular transmit antenna. Side information $\boldsymbol{\zeta}$ about the channel is assumed to be available and utilized for adapting the transmission to the channel characteristics. The space-time encoder has access to a set $\{\mathcal{C}(\boldsymbol{\zeta})\}$ of codes. Out of this set, the side information is utilized for determining the channel code $\mathcal{C}$ currently in use as $\mathcal{C} = \mathcal{C}(\boldsymbol{\zeta})$. At the receiver, the information carrying signals are first decoded into codewords and the result is then mapped back into bits, corresponding to an estimate of the transmitted message. As before, the decoding relies on the assumption that both the channel and the side information is known perfectly.

Recall that there are M transmit and N receive antennas. Let $\boldsymbol{H}$, with complex-valued elements $\{H_{kl}\}$, be an $M \times N$ matrix used for representing the MIMO channel[1] during the block under consideration as

[1] Compared with the system model used for the information theoretic investigations, $\boldsymbol{H}$ is now the complex conjugate of the previous MIMO channel matrix. Hence, each column of $\boldsymbol{H}$ represents the MISO channel between the transmitter's antenna array and the corresponding receive antenna. This will turn out to be convenient.

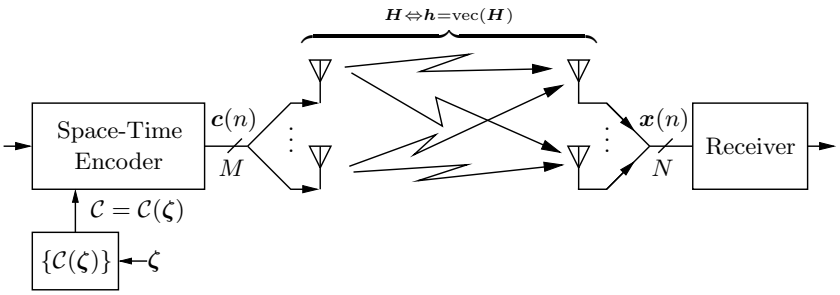

Figure 3.1: The generic system model.

$\boldsymbol{H}^*$. Consequently, the SISO channel coefficient between the kth transmit antenna and the lth receive antenna is H^*_{kl}. Furthermore, let the MIMO channel be alternatively represented by its vectorized counterpart $\boldsymbol{h} \triangleq \mathrm{vec}(\boldsymbol{H})$, with elements $\{h_k\}$. Assumptions determining the exact nature of the channel fading will be introduced in later chapters.

The received signals at sample index $n \in \{0, \cdots, L-1\}$ are collected in a vector $\boldsymbol{x}(n)$, that similarly to (2.12), can be written as

$$\boldsymbol{x}(n) = \boldsymbol{H}^* \boldsymbol{c}(n) + \boldsymbol{e}(n)\,,$$

where the potentially complex-valued symbols, transmitted from the M antennas at time instant n, are represented by

$$\boldsymbol{c}(n) \triangleq \begin{bmatrix} c_1(n) & c_2(n) & \cdots & c_M(n) \end{bmatrix}^{\mathrm{T}}\,.$$

The noise term $\boldsymbol{e}(n)$ is assumed to be generated from a zero-mean temporally white complex Gaussian random process with the spatial covariance matrix $\boldsymbol{R}_{\boldsymbol{ee}} \triangleq \mathrm{E}[\boldsymbol{e}(n)\boldsymbol{e}(n)^*] = \sigma^2 \boldsymbol{I}_N$, where σ^2 denotes the variance of the individual elements in $\boldsymbol{e}(n)$. Thus, the noise is spatially white. By grouping the L consecutive vectors $\boldsymbol{x}(n)$ into a matrix $\boldsymbol{X}$, the received block of vectors can be written as

$$\boldsymbol{X} = \boldsymbol{H}^* \boldsymbol{C} + \boldsymbol{E}\,,$$

where $\boldsymbol{E}$ contains the noise vectors and

$$\boldsymbol{C} \triangleq \begin{bmatrix} \boldsymbol{c}(0) & \boldsymbol{c}(1) & \cdots & \boldsymbol{c}(L-1) \end{bmatrix}$$

corresponds to the codeword output by the space-time encoder. The transmitted codeword $\boldsymbol{C}$ belongs to the currently used code $\mathcal{C} = \mathcal{C}(\boldsymbol{\zeta}) \triangleq$

$\{\boldsymbol{C}_k^{(\boldsymbol{\zeta})}\}_{k=1}^{K}$, where K is the number of codewords and $\boldsymbol{C}_k^{(\boldsymbol{\zeta})}$ represents the kth codeword in $\mathcal{C}(\boldsymbol{\zeta})$. Note that each codeword is a function of the side information. Hence, the number of codes in the set $\{\mathcal{C}(\boldsymbol{\zeta})\}$ is possibly uncountable when $\boldsymbol{\zeta}$ is non-discrete.

The data to be transmitted is modeled as an IID sequence of information bits where both bit-values are equally probable. Since the length of the codewords is L, the coding results in a code rate of $\log_2(K)/L$ information bits per channel use. The choice of L and K allows an appropriate tradeoff between data rate and time redundancy. Conditioned on the use of a certain code $\mathcal{C}(\boldsymbol{\zeta})$, it is assumed that the transmission is such that all the codewords are equally probable and that the average energy/power per information bit is P. In other words, the output power is limited by

$$\frac{\mathrm{E}[\|\boldsymbol{C}\|_{\mathrm{F}}^2|\boldsymbol{\zeta}]}{\log_2(K)} = P\,, \tag{3.1}$$

which corresponds to a scenario without power control since the average output power does not depend on $\boldsymbol{\zeta}$.

The receiver is assumed to recover the transmitted codewords by means of maximum likelihood (ML) decoding. Due to the fact that both the channel and the side information are known, this amounts to decoding the codewords according to

$$\hat{\boldsymbol{C}} = \arg\min_{\boldsymbol{C}\in\mathcal{C}(\boldsymbol{\zeta})} \|\boldsymbol{X} - \boldsymbol{H}^*\boldsymbol{C}\|_{\mathrm{F}}^2\,, \tag{3.2}$$

where $\hat{\boldsymbol{C}}$ denotes the codeword chosen by the receiver. An estimate of the part of the transmitted message corresponding to a block is obtained by mapping the detected codeword into its equivalent bit representation.

3.2 Code Structures

The presence of noise means that the receiver will not always come to the correct decisions. Hopefully, the probability that the receiver makes an error is sufficiently small for the communication to still be meaningful. Channel codes are typically designed with the goal of minimizing the error probability. But low error probability is not the only criterion to consider. Also important is that the codes can be decoded in a computationally efficient manner. In the worst case, the receiver has to perform an exhaustive search over all the K different codewords like indicated in

(3.2). The previously mentioned capacity achieving codes drawn from a Gaussian distribution suffer from such a problem.

To avoid high decoding complexity, it is necessary to introduce structure in the code. Many different code types exist in the literature. Cyclic, convolutional and trellis codes [Pro95] are just a few examples of code types that all have in common that additional constraints on the codewords have been imposed in order to lower the decoding complexity. There is however a tradeoff between performance and structure since structure generally limits the degrees of freedom in the code design. Completely unstructured codes obviously give the best performance, if designed appropriately, but they result in high decoding complexity. Highly structured codes, on the other hand, may be easy to decode but may also perform considerably worse than their unstructured counterparts.

In later chapters of this thesis, the performance versus structure tradeoff will to some extent be investigated by designing codes based on three different structures. The code structures under study are unstructured and linear dispersive space-time block codes as well as weighted OSTBC. The codes have here been mentioned in order of increasingly restrictive structure and are all examples of channel side information dependent codes. That is, they are functions of the channel side information vector $\boldsymbol{\zeta}$. As previously mentioned, the unstructured and linear dispersive codes are simple generalizations of previously known code types aimed for the no channel knowledge case while weighted OSTBC is a new transmission structure proposed herein. Further details are given in the sections below.

3.2.1 Unstructured Space-Time Block Codes

For unstructured space-time block codes, all the elements in all the codeword matrices can be chosen arbitrarily from the complex number field $\mathbb{C}$, subject only to the power constraint. In other words, the search for good codes is over codewords such that $\boldsymbol{C}_k^{(\boldsymbol{\zeta})} \in \mathcal{C}_k \triangleq \mathbb{C}^{M\times L}$, for all k and $\boldsymbol{\zeta}$. Here, $\mathcal{C}_k$ represents the *codeword alphabet* for the kth codeword. As already pointed out, unstructured codes have potentially the best performance but result in high decoding complexity since they require an exhaustive search as in (3.2).

For later reference, note that the power constraint in (3.1) can be

written in terms of the codewords as

$$\mathrm{E}[\|\boldsymbol{C}\|_{\mathrm{F}}^2|\boldsymbol{\zeta}] = \frac{1}{K}\sum_{k=1}^{K}\|\boldsymbol{C}_k^{(\boldsymbol{\zeta})}\|_{\mathrm{F}}^2 = \log_2(K)P\,, \tag{3.3}$$

due to the previously mentioned assumption of equally probable codewords conditioned on $\boldsymbol{\zeta}$.

3.2.2 Linear Dispersive Space-Time Block Codes

When linear dispersive space-time block codes are used, the data to be transmitted corresponding to one codeword is first mapped into symbols s_m, $m = 1, 2, \cdots, L_{\mathrm{d}}$, each taken from some signal constellation alphabet $\mathcal{A}$. The transmitted codeword is thereafter formed as a linear combination of the information bearing symbols according to

$$\boldsymbol{C} = \sum_{m=1}^{L_{\mathrm{d}}}\boldsymbol{B}_m s_m\,, \tag{3.4}$$

where $\boldsymbol{B}_m \triangleq \boldsymbol{B}_m^{(\boldsymbol{\zeta})}$ represents an arbitrary complex-valued $M \times L$ matrix used to weigh the mth symbol. Assume without loss of generality[2] that the signal constellation alphabet $\mathcal{A}$ and hence also the symbols s_m, are real-valued. The symbols depend only on the data to be transmitted while the weighting matrices may vary with the channel side information $\boldsymbol{\zeta}$. Thus, the transmission is adapted to the available channel knowledge by letting the side information determine the current set $\mathcal{B} \triangleq \{\boldsymbol{B}_m\}_{m=1}^{L_{\mathrm{d}}}$ of weighting matrices as $\mathcal{B} = \mathcal{B}(\boldsymbol{\zeta}) \triangleq \{\boldsymbol{B}_m^{(\boldsymbol{\zeta})}\}_{m=1}^{L_{\mathrm{d}}}$. This is illustrated in Figure 3.2 which represents a more specialized form of the setup in the generic system model.

By choosing the weights appropriately, the information may be spread both spatially and temporally to obtain diversity benefits and coding gain while taking the channel knowledge available at the transmitter into account. For the case of no channel side information, such linear dispersive structures have previously been described for example in [TJC99, GS00, HH02b].

A major advantage of the above linear structure is that the decoding complexity may in some cases be drastically reduced compared with

[2]Complex-valued symbols may be handled within the present framework by dividing the symbols into their real and imaginary parts. The codewords can then be written on the same form as in (3.4) but with twice as many terms.

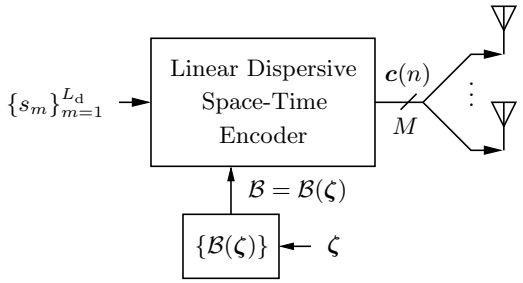

Figure 3.2: Transmitter structure using linear dispersive space-time block codes. The channel side information determines the current set $\mathcal{B}$ of weighting matrices.

the complexity of an exhaustive search. The reduction in complexity is achieved by the use of efficient decoding methods [VH02, JMO03] which typically only incur a minor increase in error probability due to their sub-optimal nature. These decoding methods are applicable if the signal constellation $\mathcal{A}$ corresponds to a regular lattice, such as when the symbols are taken from a pulse or quadrature amplitude modulation (PAM or QAM) constellation. Sphere decoding is one example of a scheme that is relatively efficient when the number of codewords K is large [VH02]. Decoding methods based on optimization techniques for semi definite problems are also applicable assuming the symbols belong to a binary phase shift keying (BPSK) or a four point QAM constellation [JMO03].

In the code design procedures presented in later chapters, the weighting matrices will be optimized based on the available side information while the constellation alphabet $\mathcal{A}$ will be held fixed. Let $s_m^{(k)}$, $m = 1, 2, \cdots, L_\mathrm{d}$ be the particular symbol sequence corresponding to the kth codeword $\boldsymbol{C}_k^{(\boldsymbol{\zeta})}$. Using linear dispersive codes thus means that the codewords are constrained as

$$\boldsymbol{C}_k^{(\boldsymbol{\zeta})} \in \mathcal{C}_k \triangleq \left\{ \sum_{m=1}^{L_\mathrm{d}} \boldsymbol{B}_m s_m^{(k)} : \boldsymbol{B}_m \in \mathbb{C}^{M \times L}, \forall m \right\},$$

for all k and $\boldsymbol{\zeta}$.

The constellation alphabet $\mathcal{A}$ is throughout this thesis assumed to correspond to a PAM constellation of equidistant real-valued signal points such that $\mathrm{E}[s_m] = 0$ and $\mathrm{E}[s_m^2] = 1$. It is further assumed that each symbol represents an integer number of information bits. It follows that

the symbols are IID since they have been mapped from an IID sequence of information bits. Hence, the power constraint in (3.1) can be formulated in terms of the weighting matrices as

$$\begin{aligned}
\mathrm{E}[\|\boldsymbol{C}\|_{\mathrm{F}}^2|\boldsymbol{\zeta}] &= \mathrm{E}[\mathrm{tr}(\boldsymbol{C}\boldsymbol{C}^*)|\boldsymbol{\zeta}] \\
&= \mathrm{E}\Big[\mathrm{tr}\Big(\sum_{m=1}^{L_\mathrm{d}}\sum_{m'=1}^{L_\mathrm{d}}\boldsymbol{B}_l\boldsymbol{B}_{m'}^* s_m s_{m'}\Big)|\boldsymbol{\zeta}\Big] \\
&= \sum_{m=1}^{L_\mathrm{d}}\sum_{m'=1}^{L_\mathrm{d}}\mathrm{tr}\left(\boldsymbol{B}_m^{(\boldsymbol{\zeta})}\big(\boldsymbol{B}_{m'}^{(\boldsymbol{\zeta})}\big)^*\right)\mathrm{E}[s_m s_{m'}] \\
&= \sum_{m=1}^{L_\mathrm{d}}\mathrm{tr}\left(\boldsymbol{B}_m^{(\boldsymbol{\zeta})}\big(\boldsymbol{B}_m^{(\boldsymbol{\zeta})}\big)^*\right) \\
&= \sum_{m=1}^{L_\mathrm{d}}\|\boldsymbol{B}_m^{(\boldsymbol{\zeta})}\|_{\mathrm{F}}^2 = P\log_2(K)\,. \qquad (3.5)
\end{aligned}$$

3.2.3 Weighted OSTBC

We propose the weighted OSTBC transmission structure as a simple but yet effective way of exploiting channel side information in conjunction with space-time coding. In weighted OSTBC, the idea is to use the side information for improving a predetermined OSTB code by means of a linear transformation. This is in line with the previous information theoretic result in Section 2.3.1 concerning the optimality of separate space-time coding and transmit weighting. Indeed, weighted OSTBC has already been briefly described in connection with the information theoretic investigation in Section 2.5, where it was proved that the structure is optimal in a capacity sense if there are two transmit antennas and one receive antenna.

The transmitted codeword in weighted OSTBC can be written on the form

$$\boldsymbol{C} = \boldsymbol{W}\bar{\boldsymbol{C}}\,, \qquad (3.6)$$

where $\boldsymbol{W} \triangleq \boldsymbol{W}^{(\boldsymbol{\zeta})}$ is an $M \times M'$ transmit weighting matrix which depends on the side information $\boldsymbol{\zeta}$ and where $\bar{\boldsymbol{C}}$ is an $M' \times L$ matrix representing the output of a fixed OSTB encoder. Here, $L \geq M'$, which means that $\bar{\boldsymbol{C}}$ has more columns than rows. The OSTB encoder takes the data to be transmitted corresponding to one codeword and first maps it into symbols s_m, $m = 1, 2, \cdots, L_\mathrm{o}$, each taken from some real-valued signal

constellation alphabet $\mathcal{A}$. Again, the symbols are IID and normalized so that $\mathrm{E}[s_m^2] = 1$. Similarly as for the previously described linear dispersive encoder, the symbols are then linearly combined to form the output as

$$\bar{\boldsymbol{C}} = \sum_{m=1}^{L_\mathrm{o}} \bar{\boldsymbol{B}}_m s_m \,,$$

where the $M' \times L$ weighting matrices $\{\bar{\boldsymbol{B}}_m\}$ are chosen so that they correspond to an OSTB code. The exact conditions under which $\{\bar{\boldsymbol{B}}_m\}$ correspond to an OSTB code will be described later in this section. For now it suffices to know that they are fixed and hence independent of the side information. Thus, only the transmit weighting $\boldsymbol{W}$ is affected by the side information, as illustrated in Figure 3.3 where $\bar{\boldsymbol{c}}(n)$ denotes the output of the OSTB encoder.

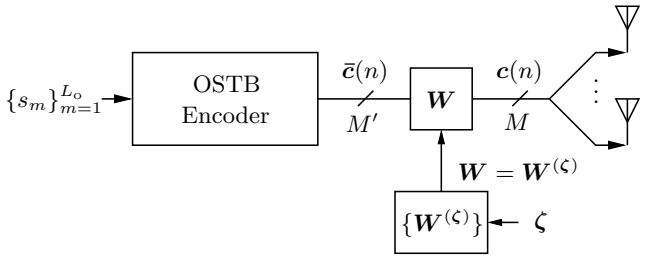

Figure 3.3: Transmitter structure in weighted OSTBC. Only the transmit weighting $\boldsymbol{W}$ is now affected by the channel side information.

In the weighted OSTBC structure, only $\boldsymbol{W}$ is free to be optimized. Hence, the structure is defined through

$$\boldsymbol{C}_k^{(\boldsymbol{\zeta})} \in \mathcal{C}_k \triangleq \left\{ \boldsymbol{W}\bar{\boldsymbol{C}}_k : \boldsymbol{W} \in \mathbb{C}^{M \times M'}, \bar{\boldsymbol{C}}_k = \sum_{m=1}^{L_\mathrm{o}} \bar{\boldsymbol{B}}_m s_m^{(k)} \right\}, \quad \forall k, \forall \boldsymbol{\zeta} \,,$$

where $s_m^{(k)}$, $m = 1, 2, \cdots, L_\mathrm{o}$ denotes the particular symbol sequence corresponding to the kth codeword $\bar{\boldsymbol{C}}_k$ in the OSTB code $\bar{\mathcal{C}} \triangleq \{\bar{\boldsymbol{C}}_k\}_{k=1}^K$. It is seen that weighted OSTBC is a special case of the less restrictive linear dispersive code structure.

Before further discussing weighted OSTBC and OSTB codes in general, it is instructive to consider a specific example illustrating typical properties of the transmission structure.

Example 3.2.1 (Weighted OSTBC using the Alamouti code)
Consider a scenario with $M = 2$ transmit antennas and assume that $M' = M$ so that the transmit weighting matrix $\boldsymbol{W}$ is square. Similarly to (2.47), assume that the OSTB encoder is based on the Alamouti code [Ala98]. Under the assumption that the information symbols are real-valued, such a choice of code leads to the encoder output

$$\bar{\boldsymbol{C}} = \begin{bmatrix} s_1 & -s_2 \\ s_2 & s_1 \end{bmatrix}, \tag{3.7}$$

which can also be written on the linear dispersive form

$$\bar{\boldsymbol{C}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} s_1 + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} s_2 \,. \tag{3.8}$$

Using this code, the transmitted codeword in weighted OSTBC is given by

$$\boldsymbol{C} = \boldsymbol{W}\bar{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 \end{bmatrix} \begin{bmatrix} s_1 & -s_2 \\ s_2 & s_1 \end{bmatrix}, \tag{3.9}$$

where $\boldsymbol{w}_l$ denotes the lth column of $\boldsymbol{W}$. It is seen that the Alamouti code, and hence also the above scheme, has a data rate of one information symbol per channel use.

The choice of the transmit weighting $\boldsymbol{W}$ significantly affects the properties of the transmitted signals. Consider first the case when $\boldsymbol{W} = \boldsymbol{I}_2$. The output of the OSTB encoder is then transmitted directly, without modification. As seen from (3.9), the same symbol is transmitted twice at alternating antennas. Switching antennas in this manner serves to protect against fading since if the channels from one of the transmit antennas are poor, there is a good chance that the channels corresponding to the other transmit antenna are better, increasing the likelihood of correctly decoding the corresponding information symbol.

Note that the transmitted signal block is a unitary matrix, i.e., $\boldsymbol{C}\boldsymbol{C}^* = \bar{\boldsymbol{C}}\bar{\boldsymbol{C}}^* = (s_1^2 + s_2^2)\boldsymbol{I}_2$. This means that the distribution of energy, as seen over two sample periods, is isotropic in space. In other words, the energy is distributed evenly in all directions. As already pointed out in Section 2.3.1, such an energy distribution makes sense if the transmitter does not know anything about the channel.

Consider now the case when the second column of $\boldsymbol{W}$ is zero, making the rank of the transmit weighting matrix equal to one. It directly follows that the transmitted codeword

$$\boldsymbol{C} = \boldsymbol{W}\bar{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{w}_1 s_1 & -\boldsymbol{w}_1 s_2 \end{bmatrix}$$

is also of rank one. Each symbol is now transmitted only once and both symbols are weighted by the same vector $\boldsymbol{w}_1$. Thus, classic beamforming in the direction of $\boldsymbol{w}_1$ is performed (the minus sign is irrelevant). Clearly, the isotropic energy distribution of the Alamouti code has been modified by the transmit weighting so that all energy is allocated to a single direction. By an appropriate choice of $\boldsymbol{w}_1$ the transmission can be directed towards the receiver so that the information carrying signals there combine coherently. Naturally, this requires knowledge about the channel at the transmitter.

It should be clear from these two special cases that the transmission, depending on the choice of transmit weighting, may correspond to either conventional OSTBC, beamforming or some combination thereof. Since OSTBC and beamforming are suitable for transmitters with no and perfect channel knowledge, respectively, weighted OSTBC offers a flexible structure that may handle channel side information of varying quality. Thus, the OSTB encoder provides diversity when the quality is low while the transmit weighting gives an additional array gain when the quality is high.

□

As long as $M' = M$, the seamless combination of OSTB coding and beamforming provided by weighted OSTBC holds in the general case as well and not only in the above example. When $M' < M$ beamforming is still possible but unmodified OSTBC is not, since the transmit weighting $\boldsymbol{W}$ has more rows than columns and thus cannot equal the identity matrix. A spatially isotropic energy distribution is therefore impossible. At best, the energy can be spread evenly over M' different directions. However, such a non-square transmit weighting has other merits since it makes it possible to take an OSTB code designed for a small antenna array (here small M') and map it onto a larger array. This is particularly important in view of the fact that OSTB codes designed for antenna arrays with more than two antennas have a much lower maximum symbol rate than the achievable rate when the design is geared towards two antennas [TJC99]. The latter issue will be discussed further in the section below.

Some Properties of OSTB Codes

Probably the most striking characteristic of OSTB codes is the fact that the rows (and also the columns if $M' = L$) in each codeword are orthogonal to each other, or more precisely, form a scaled orthonormal set,

regardless of the symbol sequence. In other words, the weights $\{\bar{\boldsymbol{B}}_m\}$ are such that, for arbitrary real-valued constellation alphabets $\mathcal{A}$,

$$\bar{\boldsymbol{C}}_k \bar{\boldsymbol{C}}_k^* = \sum_{m=1}^{L_\mathrm{o}} \left(s_m^{(k)}\right)^2 \boldsymbol{I}_{M'}, \quad k = 1, 2, \cdots, K\,. \tag{3.10}$$

The very notion of an OSTB code is in fact defined by the orthogonality property in (3.10), together with the linear dispersive structure of the code.

It is straightforward to show that for (3.10) to hold, the weights $\{\bar{\boldsymbol{B}}_m\}$ must satisfy the following conditions [TJC99]

$$\bar{\boldsymbol{B}}_k \bar{\boldsymbol{B}}_k^* = \boldsymbol{I}_{M'}, \quad k = 1, 2, \cdots, L_\mathrm{o} \tag{3.11}$$

$$\bar{\boldsymbol{B}}_k \bar{\boldsymbol{B}}_l^* + \bar{\boldsymbol{B}}_l \bar{\boldsymbol{B}}_k^* = \boldsymbol{0}_{M' \times M'}, \quad 1 \le k < l \le L_\mathrm{o}\,. \tag{3.12}$$

Depending on the symbol rate and the value of M', such weights, or equivalently OSTB codes, may or may not exist [TJC99]. In particular, OSTB codes with the maximum rate of two real-valued symbols per channel use (corresponding to a rate of one complex-valued symbol per channel use) only exist for the case of $M' = 2$. Furthermore, when M' lies between three and eight, the symbol rate cannot be higher than one. For all other M', the rate must be even lower for the conditions in (3.10) or in (3.11) and (3.12) to hold. It is easily verified that the weighting matrices in (3.8) satisfy the above conditions.

A direct consequence of the orthogonal property in (3.10) and the linearity of the code is that

$$\begin{aligned} \boldsymbol{A}(\bar{\boldsymbol{C}}_k, \bar{\boldsymbol{C}}_l) &\triangleq (\bar{\boldsymbol{C}}_k - \bar{\boldsymbol{C}}_l)(\bar{\boldsymbol{C}}_k - \bar{\boldsymbol{C}}_l)^* \\ &= \sum_{m=1}^{L_\mathrm{o}} \left(s_m^{(k)} - s_m^{(l)}\right)^2 \boldsymbol{I}_{M'}, \quad \forall k, l\,. \end{aligned} \tag{3.13}$$

Hence, also the rows in all *pair* of codeword differences form a scaled orthonormal set.

The codeword pair quantity in (3.13) plays a fundamental role in the field of space-time coding since it often appears in expressions for the error probability. To be specific, assume a spatially uncorrelated flat Rayleigh fading scenario. Furthermore, consider the use of an arbitrary fixed space-time code with codewords $\{\boldsymbol{C}_k\}$, i.e., the code does not rely on any channel side information. It can then be shown that the codeword error probability for high SNR values roughly decays as $1/(\mathrm{CG}\cdot\mathrm{SNR})^{N r_{\min}}$,

where CG is the so-called coding gain and where $r_{\min}$ is equal to the minimum rank of $\boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l)$, taken over all codewords pairs [TSC98]. The value of the exponent $Nr_{\min}$ clearly has a large impact on the performance and is referred to as the (spatial) diversity order of the system.

If $\boldsymbol{C}_k = \bar{\boldsymbol{C}}_k$ so that unmodified OSTB codewords are transmitted, it follows from the property in (3.13) that all $\boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l)$, with $k \neq l$, are of full rank. Hence, $Nr_{\min} = NM' = NM$. This shows that OSTB codes provide a diversity order equal to the product of the number of transmit and receive antennas, which is the maximum diversity order the system can offer.

Design methods for conventional space-time codes usually strive to achieve the maximum diversity order while at the same time attempt to increase the coding gain. One common design method is to maximize the determinant of $\boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l)$ over all codeword pairs [TSC98]. Later on in the thesis, $\boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l)$ will be frequently used in performance criteria that take also the channel knowledge into account.

Weighted OSTBC – ML Decoding and Power Constraint

Recall that weighted OSTBC is a special case of the previously described linear dispersive structure. Hence, performance has potentially been sacrificed by means of a more restrictive structure in order to obtain lower decoding complexity. In this case, the decoding complexity is significantly lower thanks to the orthogonality property in (3.10), which makes ML decoding of OSTBC [TSC98], and hence also weighted OSTBC, very efficient. To see why, consider the ML decoding metric in (3.2) and rewrite it for the problem at hand as

$$\begin{aligned}
\|\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{C}_k^{(\boldsymbol{\zeta})}\|_{\mathrm{F}}^2 &= \|\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})} \bar{\boldsymbol{C}}_k\|_{\mathrm{F}}^2 \\
&= \mathrm{tr}((\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})} \bar{\boldsymbol{C}}_k)(\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})} \bar{\boldsymbol{C}}_k)^*) \\
&= \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^*) - \mathrm{tr}(\boldsymbol{X} \bar{\boldsymbol{C}}_k^* (\boldsymbol{W}^{(\boldsymbol{\zeta})})^* \boldsymbol{H} + \boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})} \bar{\boldsymbol{C}}_k \boldsymbol{X}^*) \\
&\quad + \mathrm{tr}(\boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})} \bar{\boldsymbol{C}}_k \bar{\boldsymbol{C}}_k^* (\boldsymbol{W}^{(\boldsymbol{\zeta})})^* \boldsymbol{H}) \\
&= \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^*) - 2\,\mathrm{re}(\mathrm{tr}(\boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})} \bar{\boldsymbol{C}}_k \boldsymbol{X}^*)) \\
&\quad + \mathrm{tr}(\boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})} \bar{\boldsymbol{C}}_k \bar{\boldsymbol{C}}_k^* (\boldsymbol{W}^{(\boldsymbol{\zeta})})^* \boldsymbol{H}) \\
&= -2\,\mathrm{re}\left(\mathrm{tr}\left(\boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})} \Big(\sum_{m=1}^{L_{\mathrm{o}}} \bar{\boldsymbol{B}}_m s_m^{(k)} \Big) \boldsymbol{X}^* \right)\right)
\end{aligned}$$

$$
\begin{aligned}
&+ \operatorname{tr}(\boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})} (\boldsymbol{W}^{(\boldsymbol{\zeta})})^* \boldsymbol{H}) \sum_{m=1}^{L_\mathrm{o}} \left(s_m^{(k)}\right)^2 + \operatorname{tr}(\boldsymbol{X}\boldsymbol{X}^*) \\
&= \operatorname{tr}(\boldsymbol{X}\boldsymbol{X}^*) - \sum_{m=1}^{L_\mathrm{o}} f(\bar{\boldsymbol{B}}_m, s_m^{(k)}) \,,
\end{aligned}
$$

where $\mathrm{re}(x)$ denotes the real part of x and where

$$
f(\bar{\boldsymbol{B}}, s) \triangleq 2\,\mathrm{re}(\operatorname{tr}(\boldsymbol{X}^* \boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})} \bar{\boldsymbol{B}}))s - s^2 \|\boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})}\|_\mathrm{F}^2 \,.
$$

It is seen that minimizing the ML decoding metric with respect to the set of possible codewords can be accomplished by maximizing the terms in the sum separately. Hence, the symbols may be decoded individually as

$$
\hat{s}_m = \arg\max_{s \in \mathcal{A}} 2\,\mathrm{re}(\operatorname{tr}(\boldsymbol{X}^* \boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})} \bar{\boldsymbol{B}}_m))s - s^2 \|\boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\zeta})}\|_\mathrm{F}^2 \,. \tag{3.14}
$$

Such separate decoding is attractive because the number of comparisons has now decreased to $L_\mathrm{o}|\mathcal{A}|$, compared with $|\mathcal{A}|^{L_\mathrm{o}} = K$ as in joint decoding. For signal constellations with many points and/or codes with a large number of symbols, the difference in complexity can be substantial.

Note that by inserting the expression for $\boldsymbol{X}$ in (3.14) and thereafter utilizing (3.12), it can be shown that ML decoding can be performed in an interference-free manner, i.e., when a certain symbol is decoded, the values of the other symbols do not affect the decision in any way. It can also be shown that the symbol error probabilities are all equal.

In addition to a simpler decoding metric, the orthogonality property simplifies the power constraint in (3.1) so that it can be expressed in terms of the transmit weighting $\boldsymbol{W}^{(\boldsymbol{\zeta})}$ as

$$
\begin{aligned}
\mathrm{E}[\|\boldsymbol{C}\|_\mathrm{F}^2 | \boldsymbol{\zeta}] &= \mathrm{E}[\|\boldsymbol{W}\bar{\boldsymbol{C}}\|_\mathrm{F}^2 | \boldsymbol{\zeta}] \\
&= \operatorname{tr}((\boldsymbol{W}^{(\boldsymbol{\zeta})})^* \boldsymbol{W}^{(\boldsymbol{\zeta})} \,\mathrm{E}[\bar{\boldsymbol{C}}\bar{\boldsymbol{C}}^*]) \\
&= \operatorname{tr}((\boldsymbol{W}^{(\boldsymbol{\zeta})})^* \boldsymbol{W}^{(\boldsymbol{\zeta})}) \sum_{m=1}^{L_\mathrm{o}} \mathrm{E}[s_m^2] \\
&= \operatorname{tr}((\boldsymbol{W}^{(\boldsymbol{\zeta})})^* \boldsymbol{W}^{(\boldsymbol{\zeta})}) L_\mathrm{o} \\
&= L_\mathrm{o} \|\boldsymbol{W}^{(\boldsymbol{\zeta})}\|_\mathrm{F}^2 = P \log_2(K) \,.
\end{aligned} \tag{3.15}
$$

Chapter 4

Code Design with Gaussian Side Information in Mind

In this chapter, we consider the case when the transmitter has partial, but not perfect, knowledge about the channel and how to design channel side information dependent space-time block codes so that this fact is taken into account. The development is based on the previously described generic system model which here is specialized by in addition assuming that the channel conditioned on the side information is complex Gaussian distributed. The model covers scenarios in which the channel and the side information are correlated and obey a jointly complex Gaussian distribution. An important example of such a scenario is when the side information corresponds to delayed channel feedback.

Design procedures for the three code structures are developed. However, the focus is primarily on weighted OSTBC and how to efficiently determine the transmit weighting. Toward this end, the optimization problem corresponding to the transmit weight design procedure is shown to be convex when the transmit weighting is a square matrix. Thus, standard numerical methods can be used for efficiently obtaining a solution. In addition, a particularly efficient solution method is developed for the special case of independently fading channel coefficients. The proposed weighted OSTBC transmission scheme combines the benefits of conventional beamforming with those given by OSTB coding. Simulation

results for a narrowband system with multiple transmit antennas and one or more receive antennas demonstrate significant gains over conventional methods in a scenario with non-perfect channel knowledge.

4.1 Introduction

The use of transmit diversity techniques in wireless communication systems has recently received considerable attention [HH78, Wit91, HAN92, Wit93, Mog93, Wee93, SW93, Win94, KF97, Win98, TSC98, GFBK99, TJC99]. By utilizing antenna arrays at both the transmitter as well as the receiver the limitations of the radio channel may be overcome and the data rates increased. The high data rates that these MIMO systems may offer were demonstrated in [Tel95, FG98]. Therein, calculations of the information theoretic capacity assuming a flat Rayleigh fading environment were presented. This triggered the development of space-time codes [GFBK99, TSC98, TJC99] that utilize both the spatial and temporal dimension to achieve a significant portion of the aforementioned capacity.

The present work considers the generic system model described in Section 3.1 and investigates how channel side information at the transmitter can be utilized in the design of space-time block codes. Accurate estimates are in practice not always possible, particularly in environments where the parameters of the channel are rapidly time-varying. This fact is taken into account in the proposed design procedures by modeling the channel knowledge as non-perfect. As a result, the scheme continues to work well also when the quality of the side information is low.

Early attempts at designing transmission schemes for exploiting the potential offered by antenna arrays at the transmit side are generally concerned with increasing the diversity order of the system. Examples of such work include techniques for introducing artificial frequency/phase offsets [HH78, HAN92, Wee93, KF97] and time offsets [Wit91, Wit93, Mog93, SW93, Win94, Win98] between the transmitted signals. The latter technique is commonly referred to as delay diversity and is closely related to the layered space-time architecture in [Fos96]. A more systematic approach to find appropriate codes was pioneered in [GFK96, GFBK99]. A major contribution in [GFK96, GFBK99] was the development of a design criterion involving the rank and eigenvalues of certain matrices. The design criterion was later generalized to multiple receive antennas and to other channel models in [TSC98], where the now

popular notion of space-time coding was introduced. Examples of trellis codes based on the design criterion were also provided. A simple and novel block code for two transmit antennas that leads to a low-complexity receiver was developed in [Ala98]. The code is commonly known as the Alamouti code and is a member of the class of OSTB codes that in [TJC99] was extended to up to eight transmit antennas. In the same paper, it was shown that OSTB codes satisfy the rank constraint of the previously mentioned design criterion. Consequently, these codes provide the maximum spatial diversity order the system has to offer.

Common to the space-time coding schemes mentioned above is that they do not exploit channel knowledge at the transmitter. Information about the channel realization, if it is available, should of course be utilized in order to maximize the performance. As a simple example consider a scenario with perfect channel state information at both sides of the communication link. It is well-known that by appropriate linear processing at the transmitter and the receiver, the system can be transformed into a set of parallel scalar channels. The available transmit power may then be optimally allocated to the individual channels, thereby increasing the performance relative a system with a transmitter that does not exploit channel knowledge. An example of combining such an approach with coding is found in [RC98].

This chapter proposes code design procedures and transmission schemes that combine the two extremes regarding the degree of channel knowledge. Methods for improving the performance by taking the channel side information into account are developed for the three code structures described in Section 3.2. Most of the focus is however on weighted OSTBC and how to reduce the computational requirements of the transmit weight design procedure so that real-time use is feasible. An analysis providing insight into the behavior of the resulting transmission scheme is also presented. The side information is modeled using a purely statistical approach in which the channel conditioned on the side information is assumed to obey a complex Gaussian distribution. An important special case of this model is when the channel and the side information are correlated and jointly complex Gaussian. Specific attention is devoted to the latter as it offers a reasonable way of modeling the common situation of errors introduced by feedback delay.

Previous and related work on the use of non-perfect side information at the transmitter includes [Wit95, NLTW98, NTC99]. While a similar model for the side information was utilized in [Wit95, NLTW98] for determining suitable transmit beamformers in the presence of channel

estimation errors, those papers neither considered space-time codes nor multiple antennas at the receiver. Another possibility for modeling partial channel knowledge is to take on a physical perspective. This is the approach used in [NTC99] for adapting a predetermined space-time code to the particular channel. The assumption in that work is that the signals propagate along a finite number of directions, known at the transmitter, before entering a rich local scattering environment near the receiver.

The present work contributes in several ways. One important contribution is the derivation of a new performance criterion that takes channel knowledge at the transmitter into account. Although the derivation is to some extent similar in spirit to the derivation of the design criterion in [TSC98], that paper did not treat side information and therefore utilized certain approximations which prohibit the resulting design criterion from exploiting any channel knowledge. Another contribution is the development and study of techniques for code design. Of particular importance is our proposed weighted OSTBC structure and the resulting low-complexity transmission scheme which offers a seamless combination of OSTB coding and beamforming. Further motivation for weighted OSTBC is provided by the setup in the WCDMA system, where an OSTB code is used in one of the transmission modes whereas one of the other modes uses transmit beamforming [3GP02a, 3GP02b].

The chapter is organized as follows. In Section 4.2, supplements to the generic system model, including a measure on the quality of the side information, are introduced. Examples of scenarios where the assumptions are reasonable are also given. A performance criterion is derived in Section 4.3 and then utilized in Section 4.4 for developing design procedures for the three code types. The remaining part of the chapter is entirely focused on weighted OSTBC. Section 4.5 describes ways to simplify the originally proposed transmit weight design procedure, leading to a low-complexity transmission scheme. In Section 4.6, an analyses for a number of special cases reveal the highly intuitive behavior of the transmission scheme. A particularly efficient design algorithm is presented in Section 4.7 for a class of simplified fading scenarios. This algorithm is then applied in Section 4.8 to an actual example of a simplified fading scenario. Finally, simulation results are presented in Section 4.9 showing significant gains compared to conventional beamforming as well as conventional OSTB coding.

4.2 System Model

As illustrated in Figure 4.1, a wireless MIMO communication system with potentially imperfect channel side information at the transmitter is considered. The model of the system was described in Chapter 3 and is here briefly reviewed before introducing additional assumptions.

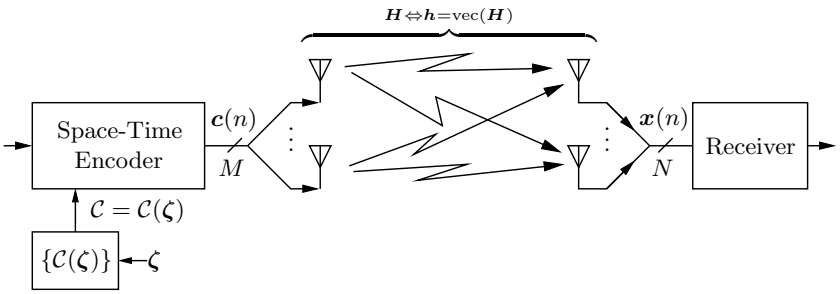

Figure 4.1: System model.

Recall that there are M transmit antennas, N receive antennas and that L consecutive received signal vectors $\boldsymbol{x}(n) = \boldsymbol{H}^* \boldsymbol{c}(n) + \boldsymbol{e}(n)$, corresponding to a fading block, are collected in a matrix

$$\boldsymbol{X} = \boldsymbol{H}^* \boldsymbol{C} + \boldsymbol{E}\,,$$

where the noise elements in $\boldsymbol{E}$ are zero-mean IID complex Gaussian with variance σ^2, the MIMO channel is represented by the $M \times N$ matrix $\boldsymbol{H}$ and where $\boldsymbol{C} \in \mathcal{C} = \{\boldsymbol{C}_k\}_{k=1}^K$ represents the transmitted codeword. Based on the outcome of the side information $\boldsymbol{\zeta}$, the channel code $\mathcal{C}$ currently in use is chosen out of a set $\{\mathcal{C}(\boldsymbol{\zeta})\}$ of codes as $\mathcal{C}(\boldsymbol{\zeta}) = \{\boldsymbol{C}_k^{(\boldsymbol{\zeta})}\}_{k=1}^K$. The transmitted codewords conditioned on $\boldsymbol{\zeta}$ are equally probable and the power is limited by the constraint

$$\frac{\mathrm{E}[\|\boldsymbol{C}\|_{\mathrm{F}}^2 | \boldsymbol{\zeta}]}{\log_2(K)} = P\,.$$

To obtain an estimate $\hat{\boldsymbol{C}}$ of the transmitted codeword, the receiver makes use of ML decoding according to

$$\hat{\boldsymbol{C}} = \arg \min_{\boldsymbol{C} \in \mathcal{C}(\boldsymbol{\zeta})} \|\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{C}\|_{\mathrm{F}}^2\,, \tag{4.1}$$

where we recall that $\boldsymbol{H}$ as well as $\boldsymbol{\zeta}$ are assumed to be known at the receiver.

The statistical relation between the side information and the channel was previously not specified. In this chapter, it is generally assumed that the statistical relation is such that the channel $\boldsymbol{h}$ conditioned on the side information $\boldsymbol{\zeta}$, denoted $\boldsymbol{h}|\boldsymbol{\zeta}$ for short, obeys a complex Gaussian distribution with mean vector $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} \triangleq \mathrm{E}[\boldsymbol{h}|\boldsymbol{\zeta}]$ and covariance matrix $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}} \triangleq \mathrm{E}[(\boldsymbol{h}-\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}})(\boldsymbol{h}-\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}})^*|\boldsymbol{\zeta}]$. As will be apparent from the development to follow, both $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ and $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ are needed in the code design process and hence must be known at the transmitter. However, these statistics can usually be computed from some scenario specific parameters. Examples of scenarios modeled by the Gaussian assumption will be further discussed later in the present section.

Note that $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ is the MMSE estimate of $\boldsymbol{h}$ based on $\boldsymbol{\zeta}$ [Kay93, p. 533]. Since $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ is the corresponding conditional error covariance matrix, $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ quantifies the remaining uncertainty about the channel $\boldsymbol{h}$ when the side information $\boldsymbol{\zeta}$ is known. Hence, the covariance matrix is a measure of the quality of the side information. A salient consequence of this measure is that also the distribution of the true channel is considered as part of the channel knowledge.

Loosely speaking, high quality side information corresponds to a small covariance (measured in a suitable norm) whereas a large covariance corresponds to side information of low quality. With such a measure, it is clear that "perfect side information" (or "perfect channel knowledge") corresponds to[1] $\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\| \to 0$, while "no side information" (or "no channel knowledge") corresponds to $\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1}\| \to 0$. In both these cases, it is assumed that $\boldsymbol{\zeta}$ is fixed so that $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ is a deterministic quantity.

However, not all scenarios where the transmitter is without channel knowledge are covered by the latter definition. This will be more clear in a later section where for illustrative purposes a special case of the system model will be considered, resulting in a need to modify the notion of "no channel knowledge".

[1] Any matrix norm may be used in the definition. The spectral norm used here is only chosen for the sake of convenience in later expressions.

4.2.1 Scenarios Modeled by the Gaussian Assumption

The Gaussian assumption regarding the statistical relation between the channel and the side information may be used to model various practical situations. Examples of such scenarios are given below.

Channel Estimates Obtained from Reverse Link Data

In systems employing TDD or FDD for duplex communication, channel estimates may be obtained in the reverse link. By utilizing the reciprocity of the propagation medium in conjunction with the time or frequency correlation of the fading channel, it is possible to compute estimates of the forward channel based on the reverse channel estimates. These estimates will typically suffer both from the presence of noise in the reverse link as well as from limited time or frequency correlation. This scenario may be modeled by assuming that $\boldsymbol{\zeta}$ is the estimate of the reverse channel. Due to the often Gaussian nature of channel fading and noise, it is reasonable to assume that $\boldsymbol{h}$ and $\boldsymbol{\gamma}$ are jointly complex Gaussian. Consequently, $\boldsymbol{h}|\boldsymbol{\zeta}$ is also complex Gaussian and our model is applicable. The conditional mean $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ would then be the MMSE estimate of the forward channel with the conditional estimation error covariance matrix given by $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$.

Geometrical Properties of the Propagation Environment

In fading scenarios where the time or frequency correlation is low, it may be difficult to obtain a reliable forward channel estimate based on a reverse channel estimate $\boldsymbol{\zeta}$. In the worst case, $\boldsymbol{\zeta}$ and $\boldsymbol{h}$ are zero-mean and completely uncorrelated/independent. The first and second order conditional statistics then reduce to $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} = \mathrm{E}[\boldsymbol{h}] = \boldsymbol{0}$ and $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}} = \mathrm{E}[\boldsymbol{h}\boldsymbol{h}^*]$.

Obviously, the channel estimate $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ does not give any useful information about the forward channel. However, $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ describes the spatial correlation properties of the channel and hence gives some information about the channel realizations in the forward direction. For example, if $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ is not a scaled identity matrix, some types of channel realizations will on the average be more likely than others and this can be used for adapting the transmission so as to obtain approximately coherent combining of the signals at the receiver.

The correlation properties are often heavily influenced by the large scale geometry of the multipath environment such as the angular posi-

tions of scatterers in the propagation paths with respect to the transmitter. Since the large scale geometry typically varies at a much lower rate than the instantaneous channel, it may be easier to track the channel covariance matrix than the channel itself [GP94b, RDJP95, Zet97].

An estimate of the covariance matrix $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}} = \mathrm{E}[\boldsymbol{h}\boldsymbol{h}^*]$ for the forward link channel can be obtained from reverse link data. One idea is to use this data for estimating the covariance of the corresponding channel. In TDD systems with sufficiently short duplex time compared with the coherence time of the channel, the resulting covariance also represent an estimate of the forward channel covariance. This since the reverse and forward links operate on the same carrier frequency. However, when FDD is employed, the reverse channel covariance may need to be modified before it can be used in the forward direction [Zet99]. An alternative approach is to form $\mathrm{E}[\boldsymbol{h}\boldsymbol{h}^*]$ based on reverse link estimates of physical parameters like, for example, the angular positions of scatterers in the propagation paths [ZO95]. In any case, the above Gaussian assumption is a useful model for such scenarios.

Quantized Side Information from a Feedback Link

Consider now a scenario in which the transmitter has access to quantized channel side information obtained from the receiver via a dedicated feedback link. Such a feedback setup is common in FDD systems in order to acquire channel estimates that are more accurate than those that can be obtained based solely on reverse link data. Similarly to as in Section 2.6.2, the side information $\boldsymbol{\zeta}$ is now some integer (or equivalent parameter) representing the quantized output of the feedback link. Consequently, $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ represents the MMSE estimate of the channel based on the quantized information and $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ is a measure of the corresponding quantization error. By modeling the quantization errors as complex Gaussian distributed, it follows that also $\boldsymbol{h}|\boldsymbol{\zeta}$ is complex Gaussian and the assumptions in this chapter thus hold.

Such an approach for dealing with quantized side information will be considered in Chapter 5. However, it should be noted that the assumption of Gaussian quantization errors represents a rather crude approximation. A consequence of this is that to get the corresponding transmission schemes to work really well in the typical case of very coarse quantization requires some heuristic modifications of the design procedures described in the present chapter. In Chapter 6 of this thesis we therefore opt for a different, but closely related, strategy for specifically handling the case of

quantized side information without relying on a Gaussian approximation.

Side Information Suffering from Feedback Delay

Also this example concerns channel side information obtained from a feedback link. In contrast to the previous case, the side information is not quantized but is still potentially imperfect due to feedback delay. The presence of feedback delay, in conjunction with a time-varying channel, may make the output of the feedback link outdated by the time it reaches the transmitter.

Because of the feedback delay, $\boldsymbol{\zeta}$ corresponds to a delayed version of the fading channel vector $\boldsymbol{h}$. The present scenario may be modeled by assuming that $\boldsymbol{h}$ and $\boldsymbol{\zeta}$ are correlated and equally distributed jointly complex Gaussian vectors. Motivation for such an assumption is provided by, for example, the well-known Jakes model [Jak94], which describes the variations of the channel as a function of time due to movement of the receiver.

In the Jakes model, the channel coefficients are samples of stationary zero-mean complex Gaussian processes, each with an autocorrelation function proportional to $J_0(2\pi f_{\mathrm{m}}\tau)$, where $J_0(\cdot)$ is the zero-order Bessel function of the first kind, τ denotes the time lag and f_{m} is the maximum Doppler frequency. If in addition it can be assumed that the channel coefficients fade independently of each other, then the vector of channel coefficients is generated from a stationary and spatially white vector-valued complex Gaussian process where the individual channel coefficients are correlated in time according to the mentioned Bessel function.

Applying this model to the present feedback situation means that the outdated side information $\boldsymbol{\zeta}$ and the current channel $\boldsymbol{h}$ are correlated samples of the same complex Gaussian process with the degree of correlation determined by the feedback delay. It follows that $\boldsymbol{h}|\boldsymbol{\gamma}$ is complex Gaussian and the development in the present chapter is therefore applicable. This scenario will be used to exemplify the transmission schemes to be developed and provide intuitive interpretations.

4.3 An Upper Bound on the Performance

To systematically incorporate channel knowledge into the code design process, the codeword error probability $\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}]$ will be used as the basis for developing a performance criterion. A feasible expression is

obtained by forming an upper bound on $\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}]$. The present section develops this bound.

Assume the codes $\{\mathcal{C}(\boldsymbol{\zeta})\}$ are fixed and consider first the codeword error probability conditioned on the side information and the channel, $\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}|\boldsymbol{\zeta}, \boldsymbol{h}]$. By means of the well-known union bound technique we have,

$$\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}|\boldsymbol{\zeta}, \boldsymbol{h}] \leq \frac{1}{K} \sum_{k \neq l} P(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} \to \boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta}, \boldsymbol{h}) \,, \tag{4.2}$$

where

$$\begin{aligned} &P(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} \to \boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta}, \boldsymbol{h}) \\ &\quad \triangleq \Pr\left[\|\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{C}_k^{(\boldsymbol{\zeta})}\|_{\mathrm{F}}^2 > \|\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{C}_l^{(\boldsymbol{\zeta})}\|_{\mathrm{F}}^2 \,|\, \boldsymbol{C} = \boldsymbol{C}_k^{(\boldsymbol{\zeta})}, \boldsymbol{\zeta}, \boldsymbol{h}\right] \end{aligned} \tag{4.3}$$

is the probability that, given a transmitted codeword $\boldsymbol{C}_k^{(\boldsymbol{\zeta})}$, the ML decoding metric (see (4.1)) corresponding to the codeword $\boldsymbol{C}_l^{(\boldsymbol{\zeta})}$ is smaller. This corresponds to the receiver making an error and choosing $\boldsymbol{C}_l^{(\boldsymbol{\zeta})}$ instead of the transmitted $\boldsymbol{C}_k^{(\boldsymbol{\zeta})}$, in the absence of any other codewords. Similar such pairwise error probabilities will be used frequently in the following. An upper bound on the codeword error probability is obtained by averaging over $\boldsymbol{\zeta}$ and $\boldsymbol{h}$ in (4.2) to arrive at

$$\begin{aligned} \Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}] &\leq \frac{1}{K} \sum_{k \neq l} \mathrm{E}_{\boldsymbol{\zeta}, \boldsymbol{h}}[P(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} \to \boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta}, \boldsymbol{h})] \\ &= \frac{1}{K} \sum_{k \neq l} \mathrm{E}_{\boldsymbol{\zeta}}[\mathrm{E}_{\boldsymbol{h}}[P(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} \to \boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta}, \boldsymbol{h})|\boldsymbol{\zeta}]] \,, \end{aligned} \tag{4.4}$$

where

$$\begin{aligned} &\mathrm{E}_{\boldsymbol{h}}[P(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} \to \boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta}, \boldsymbol{h})|\boldsymbol{\zeta}] \\ &\quad = \Pr\left[\|\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{C}_k^{(\boldsymbol{\zeta})}\|_{\mathrm{F}}^2 > \|\boldsymbol{X} - \boldsymbol{H}^* \boldsymbol{C}_l^{(\boldsymbol{\zeta})}\|_{\mathrm{F}}^2 \,|\, \boldsymbol{C} = \boldsymbol{C}_k^{(\boldsymbol{\zeta})}, \boldsymbol{\zeta}\right] \end{aligned} \tag{4.5}$$

is the pairwise error probability averaged over the channel $\boldsymbol{h}$ and conditioned on $\boldsymbol{\zeta}$ and the use of a certain code $\mathcal{C}(\boldsymbol{\zeta})$.

To obtain a closed-form upper bound on the pairwise error probability in (4.5), we proceed as follows. Since the noise is Gaussian, (4.3) can be expressed in terms of the Gaussian tail function $Q(x) \triangleq \int_x^\infty \exp(-\tilde{x}^2/2)/\sqrt{2\pi}\, d\tilde{x}$ as

$$P(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} \to \boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta}, \boldsymbol{h}) = Q\left(\frac{\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} - \boldsymbol{C}_l^{(\boldsymbol{\zeta})})\|_{\mathrm{F}}}{2\sqrt{\sigma^2/2}}\right) , \tag{4.6}$$

where $\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} - \boldsymbol{C}_l^{(\boldsymbol{\zeta})})\|_{\mathrm{F}}$ is the Euclidean distance between the two codewords after they have been transformed by the channel matrix. Inserting (4.6) into (4.5) and, similarly to [TSC98], utilizing the fact that $Q(x) \leq 0.5\exp(-x^2/2)$ [WJ90, p. 84] gives

$$\begin{aligned}\mathrm{E}_{\boldsymbol{h}}[P(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} \to \boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta},\boldsymbol{h})|\boldsymbol{\zeta}] &= \mathrm{E}_{\boldsymbol{h}}\left[Q\left(\frac{\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} - \boldsymbol{C}_l^{(\boldsymbol{\zeta})})\|_{\mathrm{F}}}{2\sqrt{\sigma^2/2}}\right)\middle|\,\boldsymbol{\zeta}\right] \\ &\leq \frac{1}{2}\,\mathrm{E}_{\boldsymbol{h}}\left[e^{-\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} - \boldsymbol{C}_l^{(\boldsymbol{\zeta})})\|_{\mathrm{F}}^2/(4\sigma^2)}\middle|\boldsymbol{\zeta}\right] \\ &\triangleq V(\boldsymbol{C}_k^{(\boldsymbol{\zeta})}, \boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta})\,. \end{aligned} \tag{4.7}$$

Since the channel $\boldsymbol{h}$ conditioned on the side information $\boldsymbol{\zeta}$ is assumed to be complex Gaussian, the expectation in (4.7) is with respect to the conditional PDF $p_{\boldsymbol{h}|\boldsymbol{\zeta}}(\boldsymbol{h}|\boldsymbol{\zeta})$, given by [Kay93, p. 507]

$$p_{\boldsymbol{h}|\boldsymbol{\zeta}}(\boldsymbol{h}|\boldsymbol{\zeta}) = \frac{e^{-(\boldsymbol{h}-\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}})^* \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}^{-1}(\boldsymbol{h}-\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}})}}{\pi^{MN}\det(\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}})}\,. \tag{4.8}$$

Similarly to [TSC98], let

$$\boldsymbol{A}(\boldsymbol{C}_k^{(\boldsymbol{\zeta})}, \boldsymbol{C}_l^{(\boldsymbol{\zeta})}) \triangleq (\boldsymbol{C}_k^{(\boldsymbol{\zeta})} - \boldsymbol{C}_l^{(\boldsymbol{\zeta})})(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} - \boldsymbol{C}_l^{(\boldsymbol{\zeta})})^*$$

denote a quantity that contains the codeword pair. By utilizing the relations (C.1), (C.2) in Appendix C and the fact that $\|\boldsymbol{X}\|_{\mathrm{F}}^2 = \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^*)$, it is possible to rewrite the Euclidean distance in (4.7) as

$$\begin{aligned}\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(\boldsymbol{\zeta})} - \boldsymbol{C}_l^{(\boldsymbol{\zeta})})\|_{\mathrm{F}}^2 &= \mathrm{tr}\left(\boldsymbol{H}^*\boldsymbol{A}(\boldsymbol{C}_k^{(\boldsymbol{\zeta})}, \boldsymbol{C}_l^{(\boldsymbol{\zeta})})\boldsymbol{H}\right) \\ &= \left(\mathrm{vec}\left(\boldsymbol{A}(\boldsymbol{C}_k^{(\boldsymbol{\zeta})}, \boldsymbol{C}_l^{(\boldsymbol{\zeta})})\boldsymbol{H}\right)\right)^* \mathrm{vec}(\boldsymbol{H}) \\ &= \boldsymbol{h}^*\big(\boldsymbol{I}_N \otimes \boldsymbol{A}(\boldsymbol{C}_k^{(\boldsymbol{\zeta})}, \boldsymbol{C}_l^{(\boldsymbol{\zeta})})\big)\boldsymbol{h}\,. \end{aligned} \tag{4.9}$$

Inserting (4.9) into (4.7) hence gives

$$V(\boldsymbol{C}_k^{(\boldsymbol{\zeta})}, \boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta}) = \int \frac{1}{2} e^{-\boldsymbol{h}^*\left(\boldsymbol{I}_N \otimes \boldsymbol{A}(\boldsymbol{C}_k^{(\boldsymbol{\zeta})}, \boldsymbol{C}_l^{(\boldsymbol{\zeta})})\right)\boldsymbol{h}/(4\sigma^2)} p_{\boldsymbol{h}|\boldsymbol{\zeta}}(\boldsymbol{h}|\boldsymbol{\zeta})\, d\boldsymbol{h}\,. \tag{4.10}$$

For notational convenience, introduce the following expression,

$$\boldsymbol{\Psi}(\boldsymbol{C}) \triangleq \big(\boldsymbol{I}_N \otimes \boldsymbol{C}\boldsymbol{C}^*\big)/(4\sigma^2) + \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}^{-1}\,.$$

After expanding the exponent of the PDF in (4.8) and combining it with (4.10), it is straightforward to verify that the sum of the exponents in the integrand of (4.10) can be written as

$$\begin{aligned}&\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}^{*}\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1}\big(\boldsymbol{\Psi}^{-1}-\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\big)\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}\\&-\big(\boldsymbol{h}-\boldsymbol{\Psi}^{-1}\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}\big)^{*}\boldsymbol{\Psi}\big(\boldsymbol{h}-\boldsymbol{\Psi}^{-1}\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}\big)\,,\end{aligned}$$

where the dependence on the codeword pair has been temporarily omitted in order to simplify the notation. The integral in (4.10) is now easily solved by making use of the fact that

$$\int\frac{e^{-\left(\boldsymbol{h}-\boldsymbol{\Psi}^{-1}\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}\right)^{*}\boldsymbol{\Psi}\left(\boldsymbol{h}-\boldsymbol{\Psi}^{-1}\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}\right)}}{\pi^{MN}\det(\boldsymbol{\Psi}^{-1})}\,d\boldsymbol{h}$$

is the integral of a complex Gaussian PDF and thus equal to one. Consequently, the upper bound on the conditional pairwise error probability can be expressed as

$$V(\boldsymbol{C}_k^{(\boldsymbol{\zeta})},\boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta})=\frac{e^{\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}^{*}\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1}(\boldsymbol{\Psi}(\boldsymbol{C}_k^{(\boldsymbol{\zeta})}-\boldsymbol{C}_l^{(\boldsymbol{\zeta})})^{-1}-\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}})\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}}}{2\det(\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}})\det(\boldsymbol{\Psi}(\boldsymbol{C}_k^{(\boldsymbol{\zeta})}-\boldsymbol{C}_l^{(\boldsymbol{\zeta})}))}\,. \quad (4.11)$$

Using (4.11) and noting that each codeword pair occurs twice in (4.4), an upper bound $P_{\mathrm{UB}}(\{\mathcal{C}(\boldsymbol{\zeta})\})$ on the codeword error probability is thus given by

$$\Pr[\hat{\boldsymbol{C}}\neq\boldsymbol{C}]\leq\frac{2}{K}\sum_{k<l}\mathrm{E}_{\boldsymbol{\zeta}}[V(\boldsymbol{C}_k^{(\boldsymbol{\zeta})},\boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta})]\triangleq P_{\mathrm{UB}}(\{\mathcal{C}(\boldsymbol{\zeta})\})\,. \quad (4.12)$$

4.4 The Code Design Problem

The derived performance bound in (4.12) may be used for designing space-time codes that take the channel side information into account. The idea is to minimize $P_{\mathrm{UB}}(\{\mathcal{C}(\boldsymbol{\zeta})\})$ with respect to the codewords in all the codes $\{\mathcal{C}(\boldsymbol{\zeta})\}$, subject to the power constraint $\mathrm{E}[\|\boldsymbol{C}\|_{\mathrm{F}}^2|\boldsymbol{\zeta}]=\log_2(K)P$ given in (3.1). Since the power constraint acts on each code $\mathcal{C}(\boldsymbol{\zeta})$ separately, it is clear that minimizing

$$\sum_{k<l}V(\boldsymbol{C}_k^{(\boldsymbol{\zeta})},\boldsymbol{C}_l^{(\boldsymbol{\zeta})}|\boldsymbol{\zeta}) \quad (4.13)$$

with respect to $\mathcal{C}(\boldsymbol{\zeta})$ for each $\boldsymbol{\zeta}$ also minimizes $P_{\mathrm{UB}}(\{\mathcal{C}(\boldsymbol{\zeta})\})$. Motivated by this, parameter independent factors in (4.13) are neglected and the code design criterion is defined as

$$W(\{\boldsymbol{C}_k\}|\boldsymbol{\zeta}) \triangleq \sum_{k<l} V(\boldsymbol{C}_k - \boldsymbol{C}_l|\boldsymbol{\zeta})\,, \tag{4.14}$$

where $\{\boldsymbol{C}_k\}$ denotes the codewords in an arbitrary code and where, with a slight abuse of notation[2], $V(\boldsymbol{C}|\boldsymbol{\zeta}) \triangleq e^{\ell(\boldsymbol{C}|\boldsymbol{\zeta})}$ with

$$\ell(\boldsymbol{C}|\boldsymbol{\zeta}) \triangleq \boldsymbol{m}^*_{\boldsymbol{h}|\boldsymbol{\zeta}} \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\boldsymbol{\zeta}} \boldsymbol{\Psi}(\boldsymbol{C})^{-1} \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\boldsymbol{\zeta}} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} - \log\det(\boldsymbol{\Psi}(\boldsymbol{C}))\,. \tag{4.15}$$

Thus, $P_{\mathrm{UB}}(\{\mathcal{C}(\boldsymbol{\zeta})\})$ is minimized by designing the codes as

$$\mathcal{C}(\boldsymbol{\zeta}) = \arg \min_{\substack{\{\boldsymbol{C}_k\} \\ \boldsymbol{C}_k \in \mathcal{C}_k, \forall k \\ \mathrm{E}[\|\boldsymbol{C}\|_{\mathrm{F}}^2|\boldsymbol{\zeta}] = \log_2(K)P}} W(\{\boldsymbol{C}_k\}|\boldsymbol{\zeta}), \quad \forall \boldsymbol{\zeta}\,, \tag{4.16}$$

where the optimization is over a code structure described by the codeword alphabets $\{\mathcal{C}_k\}$. So far, the development is general in the sense that an arbitrary code structure $\{\mathcal{C}_k\}$ may be used. In Section 4.4.2, the code design procedure in (4.16) will be explicitly tailored to the three structures previously described in Section 3.2.

4.4.1 Interpretations of the Performance Criterion

As seen from (4.16), the design criterion is a sum over all codeword pairs. Obviously, $V(\boldsymbol{C}_k - \boldsymbol{C}_l|\boldsymbol{\zeta})$, or $\ell(\boldsymbol{C}_k - \boldsymbol{C}_l|\boldsymbol{\zeta})$, constitutes a performance criterion for a particular codeword pair. A possible alternative to the above code design procedure is therefore to minimize the maximum, taken over all codeword pairs, of $\ell(\boldsymbol{C}_k - \boldsymbol{C}_l|\boldsymbol{\zeta})$. For a similar strategy based on the classic space-time code design criterion, see e.g. [BBH00]. Maximizing the worst codeword pair criterion in this manner will be pursued in Section 4.5 for the weighted OSTBC structure. However, for unstructured or linear dispersive codes, it turns out to be easier to work with a differentiable criterion function as in (4.16).

The two terms in $\ell(\boldsymbol{C}_k - \boldsymbol{C}_l|\boldsymbol{\zeta})$ give rise to some interesting interpretations. The first term mainly deals with the channel knowledge obtained from the actual realization of the side information, as contained in $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$.

[2] We have retained V as a function name, even though the arguments and the function have changed. Obvious variations of this notation will be used in the following.

The second term, on the other hand, does not depend on the realization of the side information and therefore strives for a code design suitable for a system where only the spatial correlation of the channel is known.

This interpretation of how the two terms in $\ell(\boldsymbol{C}_k - \boldsymbol{C}_l|\boldsymbol{\zeta})$ affect the code design is further supported by considering the two special cases of perfect side information ($\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\| \to 0$) and of no side information ($\|\boldsymbol{R}^{-1}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\| \to 0$), respectively. In the first case, the first term is seen to dominate and in the second case, the second term dominates. The codeword pair performance criterion in the second case is equivalent to

$$\frac{1}{\det \boldsymbol{A}(\boldsymbol{C}_k, \boldsymbol{C}_l)}\,,$$

which is basically the same as the criterion used in [GFBK99, TSC98] for designing conventional space-time codes.

As a final remark, it should be emphasized that the proposed performance criterion can also be used for designing conventional space-time codes in various scenarios where there is no side information. To see this, bear in mind that if the side information is statistically independent of the true channel, we have

$$\begin{aligned}\ell(\boldsymbol{C}_k - \boldsymbol{C}_l|\boldsymbol{\zeta}) = \boldsymbol{m}_{\boldsymbol{h}}^* \boldsymbol{R}_{\boldsymbol{hh}}^{-1} \boldsymbol{\Psi}(\boldsymbol{C}_k - \boldsymbol{C}_l)^{-1} \boldsymbol{R}_{\boldsymbol{hh}}^{-1} \boldsymbol{m}_{\boldsymbol{h}} \\ - \log\det(\boldsymbol{\Psi}(\boldsymbol{C}_k - \boldsymbol{C}_l))\,,\end{aligned} \tag{4.17}$$

where now $\boldsymbol{\Psi}(\boldsymbol{C}) = \big(\boldsymbol{I}_N \otimes \boldsymbol{C}\boldsymbol{C}^*\big)/(4\sigma^2) + \boldsymbol{R}_{\boldsymbol{hh}}^{-1}$. Thus, the development in the present work also applies to situations where the transmitter only knows the distribution of the true channel and nothing about the current realization. This version of the codeword pair performance criterion is therefore closely related to the design criterion in [GFBK99, TSC98]. In fact, after some simple manipulations, it can be shown that (4.17) includes several of the results from the various fading scenarios in [GFBK99, TSC98] as special cases.

4.4.2 Code Design Based on the Three Code Structures

Any of the three code structures described in Section 3.2 can be used in conjunction with the above code design procedure. The details for the individual cases are listed below.

1. *Unstructured codes:* In the case of unstructured codes, there are no constraints on the structure, i.e., the codeword alphabets are given

as $\{\mathcal{C}_k = \mathbb{C}^{M\times L}\}$. Since the power constraint can be written as in (3.3), the generic design procedure in (4.16) specializes to

$$\mathcal{C}(\boldsymbol{\zeta}) = \arg \min_{\substack{\{\boldsymbol{C}_k\} \\ \sum_{k=1}^{K} \|\boldsymbol{C}_k\|_{\mathrm{F}}^2 = K\log_2(K)P}} W(\{\boldsymbol{C}_k\}|\boldsymbol{\zeta}), \quad \forall \boldsymbol{\zeta}\,. \tag{4.18}$$

2. *Linear dispersive codes:* The structure of linear dispersive codes is given by the codeword alphabets

$$\mathcal{C}_k = \left\{ \sum_{m=1}^{L_{\mathrm{d}}} \boldsymbol{B}_m s_m^{(k)} : \boldsymbol{B}_m \in \mathbb{C}^{M\times L}, \forall m \right\}, \quad \forall k\,, \tag{4.19}$$

where $s_m^{(k)} \in \mathcal{A}$, $m = 1, 2, \cdots, L_{\mathrm{d}}$ is the symbol sequence corresponding to the kth codeword with $\mathcal{A}$ representing a real-valued PAM signal constellation alphabet. This means that the design procedure simplifies in the sense that the maximization is instead over all possible weighting matrices $\{\boldsymbol{B}_m\}$. Moreover, the power constraint reduces to (3.5) and, consequently, the designed codes may be obtained from

$$\{\boldsymbol{B}_m^{(\boldsymbol{\zeta})}\}_{m=1}^{L_{\mathrm{d}}} = \arg \min_{\substack{\{\boldsymbol{B}_m\} \\ \sum_{m=1}^{L_{\mathrm{d}}} \|\boldsymbol{B}_m\|_{\mathrm{F}}^2 = \log_2(K)P}} W\big(\{\sum_{m=1}^{L_{\mathrm{d}}} \boldsymbol{B}_m s_m^{(k)}\}\big|\boldsymbol{\zeta}\big), \quad \forall \boldsymbol{\zeta}\,. \tag{4.20}$$

3. *Weighted OSTBC:* Codes of type weighted OSTBC are determined by specifying the transmit weighting matrix as a function of the side information. Recall that $\bar{\mathcal{C}} = \{\bar{\boldsymbol{C}}_k\}_{k=1}^{K}$ denotes the constituent OSTB code. From (3.10) and (3.2.3) it is seen that the codeword structure is given by

$$\mathcal{C}_k = \left\{ \boldsymbol{W}\bar{\boldsymbol{C}}_k : \boldsymbol{W} \in \mathbb{C}^{M\times M'}, \bar{\boldsymbol{C}}_k = \sum_{m=1}^{L_{\mathrm{o}}} \bar{\boldsymbol{B}}_m s_m^{(k)} \right\}, \quad \forall k\,,$$

where the fixed $\bar{\boldsymbol{B}}_m$'s describe some OSTB code and are such that

$$\boldsymbol{A}(\bar{\boldsymbol{C}}_k, \bar{\boldsymbol{C}}_l) = (\bar{\boldsymbol{C}}_k - \bar{\boldsymbol{C}}_l)(\bar{\boldsymbol{C}}_k - \bar{\boldsymbol{C}}_l)^* = \mu_{kl}\boldsymbol{I}_{M'}, \quad \forall k, l\,. \tag{4.21}$$

Here, $\mu_{kl} \triangleq \sum_{m=1}^{L_{\mathrm{o}}} \left(s_m^{(k)} - s_m^{(l)}\right)^2$ is a scaling factor which depends on the codeword pair. From the orthogonality property in (4.21) it

follows that $\boldsymbol{A}(\boldsymbol{W}\bar{\boldsymbol{C}}_k, \boldsymbol{W}\bar{\boldsymbol{C}}_l) = \mu_{kl}\boldsymbol{W}\boldsymbol{W}^*$. Inserting this and the power constraint (3.15) into (4.16) leads to

$$\boldsymbol{W}^{(\boldsymbol{\zeta})} = \arg \min_{\substack{\boldsymbol{W} \\ \|\boldsymbol{W}\|_{\mathrm{F}}^2 = P\log_2(K)/L_{\mathrm{o}}}} W(\boldsymbol{W}\boldsymbol{W}^*|\boldsymbol{\zeta}), \quad \forall \boldsymbol{\zeta}\,, \tag{4.22}$$

where

$$W(\boldsymbol{W}\boldsymbol{W}^*|\boldsymbol{\zeta}) \triangleq \sum_{k<l} e^{\ell(\boldsymbol{W}\boldsymbol{W}, \mu_{kl}|\boldsymbol{\zeta})} \tag{4.23}$$

$$\begin{aligned} \ell(\boldsymbol{W}\boldsymbol{W}^*, \mu_{kl}|\boldsymbol{\zeta}) \triangleq\ & \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}^* \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}^{-1} \boldsymbol{\Psi}(\boldsymbol{W}\boldsymbol{W}^*, \mu_{kl})^{-1} \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}^{-1} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} \\ & - \log\det(\boldsymbol{\Psi}(\boldsymbol{W}\boldsymbol{W}^*, \mu_{kl})) \end{aligned} \tag{4.24}$$

$$\boldsymbol{\Psi}(\boldsymbol{W}\boldsymbol{W}^*, \mu_{kl}) \triangleq \big(\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*\big)\mu_{kl}/(4\sigma^2) + \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}^{-1}\,.$$

The design procedures for unstructured and linear dispersive codes both represent complicated optimization problems that are difficult to solve efficiently. The unstructured approach requires $2KML$ real-valued parameters to be optimized. For linear dispersive codes, the corresponding number of parameters is $2MLL_{\mathrm{d}}$, which typically is a considerably smaller number than its unstructured counterpart, for the same number of codewords K. Moreover, the criterion functions in (4.18) and (4.20) each consist of a large number of terms even for moderate K. The large number of parameters and terms effectively makes real-time design of unstructured or linear dispersive codes infeasible in practice. This basically excludes the use of these code structures for the case of non-quantized side information.

On the other hand, when the side information is quantized, these two code structures are perfectly viable due to the finite number of possible outcomes of the side information. The entire channel side information dependent code $\{\mathcal{C}(\boldsymbol{\zeta})\}$ can consequently be precalculated off-line and stored in a lookup table, suitable for real-time use. Such off-line design procedures will be investigated later in this thesis in Chapters 5 and 6.

In contrast to the unstructured and linear dispersive structures, the properties of weighted OSTBC considerably simplifies the design problem. For example, the optimization in (4.22) is just over a single $M \times M'$ matrix $\boldsymbol{W}$. This makes code design based on weighted OSTBC a promising candidate for real-time use. The focus in the remaining part of this chapter is therefore on a transmission scheme based on weighted OSTBC and the development of computationally efficient algorithms for designing the transmit weighting. In particular, it will be argued that suitable

transmit weighting matrices can be designed by considering only one of the terms in the criterion function in (4.23).

4.5 Weighted OSTBC – Simplifying the Design Problem

This section focuses on weighted OSTBC and presents techniques for efficiently designing the transmit weighting matrix. To reach the goal of a computationally efficient transmit weight design procedure, an approximative version of the optimization problem in (4.22) is considered. It is pointed out that the resulting design method minimizes an upper bound on the symbol error probability. Another attractive feature of the design method is that it can be formulated in terms of a convex optimization problem, when the transmit weighting matrix is square.

Reducing the Number of Terms in the Criterion Function

As previously mentioned, the design problem in the case of weighted OSTBC is not as computationally demanding as for the unstructured or linear dispersive codes because of fewer parameters in the optimization problem. However, the number of terms in the criterion function still grows as rapidly with the code size K as for the two other code structures. In order to avoid this problem, an approximate version of the criterion function in (4.23) will instead be considered. Only the largest term(s) corresponding to the worst codeword pair(s) will be retained in the sum. Hence, we take on the common min max approach and minimize the worst pairwise performance criterion, i.e., the transmit weighting is obtained as

$$\boldsymbol{W}^{(\boldsymbol{\zeta})} = \arg \min_{\substack{\boldsymbol{W} \\ \|\boldsymbol{W}\|_{\mathrm{F}}^2 = P \log_2(K)/L_{\mathrm{o}}}} \max_{\substack{k,l \\ k<l}} \ell(\boldsymbol{W}\boldsymbol{W}^*, \mu_{kl}|\boldsymbol{\zeta}), \quad \forall \boldsymbol{\zeta}\,.$$

This min max design problem is considerably easier to solve than the corresponding problem for the other code structures. The reason why is because the maximization part can be performed analytically. Due to the orthogonal property of the constituent fixed space-time code, the dependence on the codeword pair is now only through the scaling factor μ_{kl}. This together with the easily verifiable fact that $\ell(\boldsymbol{W}\boldsymbol{W}^*, \mu_{kl}|\boldsymbol{\zeta})$ is a decreasing function of μ_{kl} means that

$$\max_{\substack{k,l \\ k<l}} \ell(\boldsymbol{W}\boldsymbol{W}^*, \mu_{kl}|\boldsymbol{\zeta}) = \ell(\boldsymbol{W}\boldsymbol{W}^*, \mu_{\min}|\boldsymbol{\zeta})\,,$$

where $\mu_{\min}$ denotes the minimum μ_{kl}. An optimal $\boldsymbol{W}$ (optimal in the sense that it minimizes the criterion function under consideration) can now be determined as

$$\boldsymbol{W}^{(\boldsymbol{\zeta})} = \arg \min_{\|\boldsymbol{W}\|_{\mathrm{F}}^2 = P_{\mathrm{o}}} \ell(\boldsymbol{W}\boldsymbol{W}^*, \mu_{\min}|\boldsymbol{\zeta}), \quad \forall \boldsymbol{\zeta}\,, \tag{4.25}$$

where $P_{\mathrm{o}} \triangleq P \log_2(K)/L_{\mathrm{o}}$ is used to shorten the notation.

Relation to the Symbol Error Probability

Note that, for weighted OSTBC, the above min max approach is actually better than using the original criterion function in (4.23) since, as will be argued below, (4.25) minimizes an upper bound on the *symbol* error probability, as opposed to the codeword error probability.

To see this, recall from the discussion in Section 3.2.3 that each constituent information bearing symbol can be decoded separately and independently of all the others. Hence, the symbol error probability may be studied by keeping only one out of the L_{o} different symbols in the OSTBC output $\bar{\boldsymbol{C}}$. The other symbols may be set to zero. The resulting codewords in such a fictitious code are therefore given by $\{\boldsymbol{W}\bar{\boldsymbol{B}}_m s : s \in \mathcal{A}\}$, where only the mth symbol s_m is kept. As also pointed out in Section 3.2.3, the error probability is the same for all symbols and hence m may here be chosen in an arbitrary manner. Let $s^{(k)}$ denote the kth symbol alternative. Note now that, since $\bar{\boldsymbol{B}}_m \bar{\boldsymbol{B}}_m^* = \boldsymbol{I}_{M'}$, as evident from (3.11), the codeword pair $\boldsymbol{W}\bar{\boldsymbol{B}}_m s^{(k)}$, $\boldsymbol{W}\bar{\boldsymbol{B}}_m s^{(l)}$ satisfies the relation

$$\boldsymbol{A}(\boldsymbol{W}\bar{\boldsymbol{B}}_m s^{(k)}, \boldsymbol{W}\bar{\boldsymbol{B}}_m s^{(l)}) = \boldsymbol{W}\boldsymbol{W}^* \tilde{\mu}_{kl}\,,$$

where $\tilde{\mu}_{kl} \triangleq (s^{(k)} - s^{(l)})^2$. Consequently, the fictitious code can be seen as a degenerate version of weighted OSTBC and the present framework is therefore applicable. After letting $\tilde{\mu}_{\min}$ denote the minimum $\tilde{\mu}_{kl}$ it is realized that $e^{\ell(\boldsymbol{W}\boldsymbol{W}^*, \tilde{\mu}_{\min}|\boldsymbol{\zeta})}$ is, except for a parameter independent scaling factor, an upper bound on the conditional symbol error probability.

Since $\mu_{\min} = \tilde{\mu}_{\min}$, the transmit weight design procedure in (4.25) corresponds to minimizing an upper bound on the symbol error probability. It can be shown that the bound is much tighter than the previously utilized bound on the codeword error probability. In fact, if it were not for the use of the Gaussian tail inequality $Q(x) \leq 0.5 \exp(-x^2/2)$ in the earlier derivation, the design procedure would minimize the exact symbol error probability, since $\mathcal{A}$ is assumed to correspond to a PAM constellation of equidistant points.

Turning the Design Procedure into a Convex Problem

Although (4.25) is considerably easier to solve than the original weight design procedure, it still represents a rather complicated optimization problem since the corresponding criterion function has multiple minima[3]. With the goal of obtaining a more efficient solution procedure, we therefore re-parameterize the problem as follows. An inspection of (4.25) suggests the parametrization $\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{W}^*$. A two step procedure can now be used for finding an optimal solution to the design problem in (4.25). Rewriting the criterion function and the constraints in terms of the new parameters give rise to the following optimization problem

$$\boldsymbol{Z}^{(\boldsymbol{\zeta})} = \arg \min_{\substack{\boldsymbol{Z} \\ \boldsymbol{Z}=\boldsymbol{Z}^* \succeq 0 \\ \mathrm{rank}(\boldsymbol{Z}) \leq \min\{M', M\} \\ \mathrm{tr}(\boldsymbol{Z})=P_\mathrm{o}}} \ell(\boldsymbol{Z}|\boldsymbol{\zeta}), \quad \forall \boldsymbol{\zeta}\,, \tag{4.26}$$

where, with a slight abuse of notation, the codeword pair performance criterion is here written as

$$\begin{aligned} \ell(\boldsymbol{Z}|\boldsymbol{\zeta}) &\triangleq \ell(\boldsymbol{Z}, \mu_{\min}|\boldsymbol{\zeta}) \\ &= \boldsymbol{m}^*_{\boldsymbol{h}|\boldsymbol{\zeta}} \boldsymbol{R}^{-1}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}} \big((\boldsymbol{I}_N \otimes \boldsymbol{Z})\eta + \boldsymbol{R}^{-1}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}} \big)^{-1} \boldsymbol{R}^{-1}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} \\ &\quad - \log\det\left((\boldsymbol{I}_N \otimes \boldsymbol{Z})\eta + \boldsymbol{R}^{-1}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}} \right), \end{aligned} \tag{4.27}$$

with $\eta \triangleq \mu_{\min}/(4\sigma^2)$. The matrix $\boldsymbol{Z}$ has here been constrained to have a maximum rank of $\min\{M', M\}$ to take into account that $\boldsymbol{W}$ may be non-square. An optimal transmit weighting is finally obtained as $\boldsymbol{W}^{(\boldsymbol{\zeta})} = (\boldsymbol{Z}^{(\boldsymbol{\zeta})})^{1/2}$, where $(\cdot)^{1/2}$ is a matrix square-root such that $\boldsymbol{Z}^{(\boldsymbol{\zeta})} = \boldsymbol{W}^{(\boldsymbol{\zeta})}(\boldsymbol{W}^{(\boldsymbol{\zeta})})^*$. Note that a square-root always exists since $\boldsymbol{Z}^{(\boldsymbol{\zeta})}$ is a non-negative definite matrix. Clearly, the solution is not unique.

Assuming $M' = M$ so that the transmit weighting is square, the rank constraint vanishes. In this case, the described reformulation is particularly attractive since both the criterion function as well as the remaining constraints in (4.26) are convex, making the entire optimization problem convex. To see that the criterion function is convex, first note that

$$\boldsymbol{\Psi}(\boldsymbol{Z}, \mu_{\min}) = (\boldsymbol{I}_N \otimes \boldsymbol{Z})\eta + \boldsymbol{R}^{-1}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}$$

[3] To see this, consider an optimal solution $\boldsymbol{W}^{(\boldsymbol{\zeta})}$ of (4.25) and note that $\boldsymbol{W}^{(\boldsymbol{\zeta})}\boldsymbol{U}$, where $\boldsymbol{U}$ denotes an arbitrary unitary $M' \times M'$ matrix, is also a solution. Hence the solution is not unique.

defines an affine transformation of $\boldsymbol{Z}$ and that $\boldsymbol{\Psi}(\boldsymbol{Z}, \mu_{\min})$ is positive definite over the set of all positive semi definite $\boldsymbol{Z}$. Since affine transformations preserve convexity [BV99, VB96], the convexity of the criterion function can now be established by proving that

$$\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}^{*} \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1} \boldsymbol{\Psi}^{-1} \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} + \log\det(\boldsymbol{\Psi}^{-1}) \tag{4.28}$$

is convex over the set of positive definite matrices $\boldsymbol{\Psi}$.

The first term in (4.28) is shown to be convex by utilizing a theorem saying that a function is convex over a set $\mathcal{X}$ if it is convex when restricted to any line that intersects $\mathcal{X}$ [BS93, p. 94]. For this purpose, let $\boldsymbol{\Psi} = \lambda\boldsymbol{\Psi}_1 + (1-\lambda)\boldsymbol{\Psi}_2$, where $\boldsymbol{\Psi}_1 = \boldsymbol{\Psi}_1^{*}$, $\boldsymbol{\Psi}_2 = \boldsymbol{\Psi}_2^{*}$ are arbitrary positive definite matrices and $0 \leq \lambda \leq 1$, represent any line in the set of positive definite matrices. By making use of the identity [Kay93, p. 73]

$$\frac{\partial \boldsymbol{X}(\theta)^{-1}}{\partial \theta} = -\boldsymbol{X}(\theta)^{-1} \frac{\partial \boldsymbol{X}(\theta)}{\partial \theta} \boldsymbol{X}(\theta)^{-1} \,,$$

the second order derivative of the first term with respect to λ is easily obtained as

$$2\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}^{*} \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1} \boldsymbol{\Psi}^{-1} (\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2) \boldsymbol{\Psi}^{-1} (\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2) \boldsymbol{\Psi}^{-1} \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} \,.$$

This quadratic form is non-negative since the matrix

$$\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1} \boldsymbol{\Psi}^{-1} (\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2) \boldsymbol{\Psi}^{-1} (\boldsymbol{\Psi}_1 - \boldsymbol{\Psi}_2) \boldsymbol{\Psi}^{-1} \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1}$$

clearly is positive semi definite. Because the second order derivative is non-negative it follows that the first term is convex with respect to λ (see e.g. [BS93, p. 91]) and thus, according to the theorem in [BS93, p. 94], also over the set of positive definite matrices $\boldsymbol{\Psi}$.

The convexity of the second term in (4.28) follows directly from e.g. [HJ96, p. 466]. Consequently, (4.28) is convex over all positive definite $\boldsymbol{\Psi}$. Due to the affine relation between $\boldsymbol{\Psi}$ and $\boldsymbol{Z}$, the criterion function is convex also with respect to $\boldsymbol{Z}$. It is easily verified that the constraints are convex [BV99, VB96]. The entire optimization problem is therefore convex, which implies that all local minima are also global minima.

We will not go into great detail describing an algorithm that solves this particular convex optimization problem, since there are a number of standard techniques that are applicable. For example, interior point methods can be used for efficiently solving this kind of problem [NN94].

Note that when $M' < M$, corresponding to a tall transmit weighting matrix, the presence of the rank constraint makes the optimization

problem in (4.26) non-convex. One possible suboptimal approach is then to first neglect the rank constraint and solve the resulting convex optimization problem. If the obtained solution $\boldsymbol{Z}^{(\boldsymbol{\zeta})}$ is of rank M' or lower, the optimum also of the original design problem is found. On the other hand, if this is not the case, the solution of the convex problem is modified by replacing $\boldsymbol{Z}^{(\boldsymbol{\zeta})}$ with its nearest, in a Frobenius norm sense, rank M' matrix [HJ96, p. 427]. This is done by utilizing the SVD for setting the $M - M'$ smallest singular values of $\boldsymbol{Z}^{(\boldsymbol{\zeta})}$ to zero. Such a strategy is clearly suboptimal but provides an intuitively appealing way of dealing with a transmit weighting matrix which has more rows than columns. As will be shown in Section 4.7, it turns out that an optimal solution is easily obtained even for this case if some additional assumptions on the statistics of the channel and the side information are made.

Another alternative approach is of course to avoid the rank constraint altogether by optimizing with respect to $\boldsymbol{W}$ as in (4.25). Numerical optimization techniques can be employed to "solve" this problem. However, since (4.25) is not convex it is difficult to ensure that a global optimal solution is found.

4.6 Properties of the Designed Transmit Weighting

Although the transmit weight design procedure given by (4.26) must in general be solved numerically, there are a few special cases that permit a closed-form solution. These special cases concern the asymptotic properties of the solution. In particular, attention is here turned toward the behavior of the solution when the channel quality is perfect and when there is no channel information, respectively. In addition, the influence of the SNR level is investigated. The solutions turn out to agree well with intuition and allow for some interesting interpretations. For simplicity it is assumed throughout this section that $M' = M$ so that the rank constraint vanishes. Detailed derivations can be found in Appendix 4.A.

No Channel Knowledge

In the first case, no channel knowledge is considered, i.e., $\boldsymbol{\zeta}$ is assumed to be fixed and $\|\boldsymbol{R}^{-1}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\| \to 0$. For technical reasons, it is also assumed that $\|\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}\|$ is bounded as $\|\boldsymbol{R}^{-1}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\| \to 0$. The criterion function used

in (4.26) is then minimized for $\boldsymbol{Z}_{\text{as}}^{(\boldsymbol{\zeta})} = \boldsymbol{I}_M P_{\text{o}}/M$, where the index "as" emphasizes that this is an asymptotic result. Consequently, the optimal transmit weighting is a scaled unitary matrix, an obvious choice is $\boldsymbol{W}_{\text{as}}^{(\boldsymbol{\zeta})} = \boldsymbol{I}_M \sqrt{P_{\text{o}}/M}$. Thus, the codewords are transmitted without modification. This makes sense considering the assumptions under which the predetermined OSTB code was designed. It also makes sense in view of the fact that the transmitter does not know the channel and therefore has to choose a "neutral" spatially isotropic output power distribution.

High SNR

The second case concerns infinite SNR, in the sense that[4] $\eta = \mu_{\min}/(4\sigma^2) \to \infty$. Similarly to the case of no channel knowledge, the optimal transmit weighting can be chosen to be a scaled unitary matrix. This indicates that the usefulness of channel knowledge diminishes as the SNR increases. Simulation results described in Section 4.9 further support this claim. Moreover, recall that we came to a similar conclusion in connection to the numerical results on channel capacity presented in Figure 2.5.

Perfect Channel Knowledge

In the third case, the channel knowledge is assumed to be perfect. Hence, $\boldsymbol{\zeta}$ is fixed and $\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\| \to 0$. It is further assumed that $\bar{\boldsymbol{m}}_{\boldsymbol{h}|\boldsymbol{\zeta}} \triangleq \lim_{\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\|\to 0} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ exists and is finite. Introduce the $M \times N$ asymptotic MMSE channel estimate matrix $\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}} \triangleq \lim_{\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\|\to 0} \mathrm{E}[\boldsymbol{H}|\boldsymbol{\zeta}]$, also obtained by rearranging $\bar{\boldsymbol{m}}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ so as to satisfy the relation $\bar{\boldsymbol{m}}_{\boldsymbol{h}|\boldsymbol{\zeta}} = \mathrm{vec}(\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}})$. Moreover, let

$$\boldsymbol{\Theta} \triangleq \bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}} \bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}}^{*} \,. \tag{4.29}$$

To simplify the analysis it is assumed that one of the eigenvalues of $\boldsymbol{\Theta}$ is strictly larger than all the others. This assumption is further commented on in the following. Studying the behavior of the solution as $\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\| \to 0$ gives the following asymptotically optimal transmit weighting,

$$\boldsymbol{W}_{\text{as}}^{(\boldsymbol{\zeta})} = \sqrt{P_{\text{o}}} \begin{bmatrix} \boldsymbol{v}_M & \boldsymbol{0} & \cdots & \boldsymbol{0} \end{bmatrix} , \tag{4.30}$$

where $\boldsymbol{v}_M$ is the unit norm eigenvector of $\boldsymbol{\Theta}$ corresponding to the largest eigenvalue.

[4] The analysis can easily be extended to cover a more general definition of the SNR.

Due to the special structure of OSTB codes, and since only one column of $\boldsymbol{W}_{\text{as}}^{(\boldsymbol{\zeta})}$ is non-zero, (4.30) may be interpretated as beamforming in the direction of $\boldsymbol{v}_M$. To see this, consider for example the two transmit antenna case and assume like in (3.7) that the output of the OSTB encoder is given by

$$\bar{\boldsymbol{C}} = \begin{bmatrix} s_1 & -s_2 \\ s_2 & s_1 \end{bmatrix}, \tag{4.31}$$

where s_m represents the mth information bearing symbol in the fading block. The encoder output in (4.31) corresponds to the well-known Alamouti space-time code [Ala98]. By utilizing the asymptotic result in (4.30) and the expression for the space-time code it is seen that the signal transmitted over the two antennas during the two time instants can be written as

$$\boldsymbol{C} = \boldsymbol{W}_{\text{as}}^{(\boldsymbol{\zeta})}\bar{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{c}(0) & \boldsymbol{c}(1) \end{bmatrix} = \sqrt{P_{\text{o}}} \begin{bmatrix} \boldsymbol{v}_M s_1 & -\boldsymbol{v}_M s_2 \end{bmatrix}.$$

Clearly, beamforming in the direction of $\boldsymbol{v}_M$ is performed (the minus sign in the second column is irrelevant). The present development can be generalized to all the OSTB codes found in [TJC99].

Note that because of the perfect channel knowledge, $\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}}$ is essentially the same as $\boldsymbol{H}$. Consequently, for all practical purposes, $\boldsymbol{v}_M$ can be considered equal to the left singular vector of $\boldsymbol{H}$ corresponding to the largest singular value [HJ96, p. 414]. Thus, the transmission is now conducted in much the same way as in schemes which utilize the SVD of the channel matrix to convert the MIMO system into a set of parallel subchannels. Such a strategy was adopted for example in [RC98], where a water-filling procedure was used for allocating transmit power among all the subchannels and determining the corresponding data rates. However, our transmission scheme differs, among other things, in that only the *strongest* sub-channel is used.

The structure of the underlying OSTB code together with the present goal of minimizing error probability, as opposed to maximizing data rate as in [RC98], explain why such a transmission strategy is obtained. To see this, recall that the constituent data symbols are decoded in an interference-free manner, allowing the transmission scheme to be studied by considering the symbols separately from each other. For example, consider again the Alamouti code in (4.31) and observe that the contribution to the transmitted signal from s_1 is

$$\boldsymbol{C} = \boldsymbol{W}_{\text{as}}^{(\boldsymbol{\zeta})} \begin{bmatrix} s_1 & 0 \\ 0 & s_1 \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{c}(0) & \boldsymbol{c}(1) \end{bmatrix} = \sqrt{P_{\mathrm{o}}} \begin{bmatrix} \boldsymbol{w}_1 s_1 & \boldsymbol{w}_2 s_2 \end{bmatrix} ,$$

where $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ represent the columns of $\boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})}$. It is now easily shown that to maximize the SNR for $\boldsymbol{c}(0)$ and $\boldsymbol{c}(1)$, which in this case is equivalent to minimizing the symbol error probability, the two columns of $\boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})}$ should be matched to the channel. In other words, both should be parallel to the strongest left singular vector of the channel. Similar arguments apply to s_2 and also to other OSTB codes.

Also note that the assumption that $\boldsymbol{\Theta}$ has a strictly largest eigenvalue is weak. This is due to the often random nature of $\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}}$ (or $\boldsymbol{H}$, since $\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}} \approx \boldsymbol{H}$) when $\boldsymbol{\zeta}$ varies as in practical fading scenarios. Consequently, also $\boldsymbol{\Theta}$ is a random quantity, which makes it highly unlikely that the assumption is violated except in some degenerate cases. In particular, the probability that $\boldsymbol{\Theta}$ does not have a strictly largest eigenvalue is vanishingly small for the simplified fading scenario described in Section 4.8.

Low SNR

Finally, in the fourth case we consider an SNR value tending to zero, i.e., $\eta \to 0$. It turns out that the result is similar to the one derived in the previous case. Hence, the asymptotically optimal transmit weighting is again given by (4.30). However, $\boldsymbol{\Theta}$ is now defined as

$$\boldsymbol{\Theta} \triangleq \boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\zeta}} \boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\zeta}}^{*} + \boldsymbol{R}_{\boldsymbol{H}\boldsymbol{H}|\boldsymbol{\zeta}} ,$$

where $\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\zeta}} \triangleq \mathrm{E}[\boldsymbol{H}|\boldsymbol{\zeta}]$ and $\boldsymbol{R}_{\boldsymbol{H}\boldsymbol{H}|\boldsymbol{\zeta}} \triangleq \mathrm{E}[(\boldsymbol{H} - \boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\zeta}})(\boldsymbol{H} - \boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\zeta}})^{*}|\boldsymbol{\zeta}]$. As seen, also the spatial correlation properties of the channel influence the transmission through the inclusion of the additional term $\boldsymbol{R}_{\boldsymbol{H}\boldsymbol{H}|\boldsymbol{\zeta}}$. Again, the result relies on the existence of an eigenvalue of $\boldsymbol{\Theta}$ that is strictly larger than all the others.

Note that in the case of one receive antenna, the beamforming strategy proposed in [NLTW98], although derived using a different performance criterion, is seen to also give a beamformer proportional to $\boldsymbol{v}_M$. The approach taken on in [NLTW98] was to minimize the average SNR. As also pointed out therein, such a performance criterion makes sense for small SNR values. Hence, the result for the fourth case in this section provides a generalization of the corresponding result in [NLTW98] to multiple receive antennas when a predetermined OSTB code is used.

4.7 A Weight Design Algorithm for a Simplified Scenario

In this section we consider a simplified fading scenario in order to obtain a semi closed-form solution of the transmit weight design problem given in (4.26). In spite of the existence of fairly efficient numerical optimization techniques for the general case, the complexity of the algorithm described in this section is substantially lower.

Introduce the simplifying assumption that the conditional covariance matrix is diagonal, also expressed as $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}} = \alpha \boldsymbol{I}_{MN}$. Here, α represents the conditional variance of the channel coefficients. A scenario where this assumption is reasonable will be considered in Section 4.8. By letting $\hat{\boldsymbol{\Upsilon}} \triangleq \frac{1}{\alpha}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\zeta}}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\zeta}}^*$ it is now possible to rewrite the performance criterion (4.27) in the following way

$$\begin{aligned}
\ell(\boldsymbol{Z}|\boldsymbol{\zeta}) &= \frac{1}{\alpha}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}^*\big((\boldsymbol{I}_N \otimes \boldsymbol{Z}\alpha\eta) + \boldsymbol{I}_{MN}\big)^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} \\
&\quad - \log\det\big((\boldsymbol{I}_N \otimes \boldsymbol{Z}\alpha\eta) + \boldsymbol{I}_{MN}\big) + MN\log(\alpha) \\
&= \frac{1}{\alpha}\operatorname{tr}\Big(\big(\boldsymbol{I}_N \otimes (\boldsymbol{Z}\alpha\eta + \boldsymbol{I}_M)\big)^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}^*\Big) \\
&\quad - \log\det\big(\boldsymbol{I}_N \otimes (\boldsymbol{Z}\alpha\eta + \boldsymbol{I}_M)\big) + MN\log(\alpha) \\
&= \operatorname{tr}\big((\boldsymbol{Z}\alpha\eta + \boldsymbol{I}_M)^{-1}\hat{\boldsymbol{\Upsilon}}\big) - N\log\det(\boldsymbol{Z}\alpha\eta + \boldsymbol{I}_M) + MN\log(\alpha)\,,
\end{aligned}$$

where the second equality is due to the well-known relation $\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{BA})$. In order to minimize this criterion, let $\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^*$ and $\hat{\boldsymbol{\Upsilon}} = \hat{\boldsymbol{V}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{V}}^*$ represent the EVD of $\boldsymbol{Z}$ and $\hat{\boldsymbol{\Upsilon}}$, respectively. The diagonal elements of $\boldsymbol{\Lambda}$ and $\hat{\boldsymbol{\Lambda}}$, corresponding to the eigenvalues, are here denoted by $\{\lambda_i\}_{i=1}^M$ and $\{\hat{\lambda}_i\}_{i=1}^M$, respectively. In each set the eigenvalues are sorted in ascending order. The $M \times M$ matrices $\boldsymbol{V}$ and $\hat{\boldsymbol{V}}$ are unitary.

Next, the above preliminaries will be used to derive an optimal transmit weighting. In the presentation, it is assumed that $M' = M$ so that the rank constraint is absent. The derivation is however easily generalized to $M' \neq M$ as commented on at the end of the present section.

Substitute the new parametrization into $\ell(\boldsymbol{Z}|\boldsymbol{\zeta})$ and into the constraints and neglect parameter independent terms to arrive at an equivalent minimization problem with the criterion function

$$\begin{aligned}
\ell(\{\lambda_i\}_{i=1}^M, \boldsymbol{V}) \triangleq \operatorname{tr}\big((\boldsymbol{\Lambda}\alpha\eta + \boldsymbol{I}_M)^{-1}\boldsymbol{V}^*\hat{\boldsymbol{V}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{V}}^*\boldsymbol{V}\big) \\
- N\log\det(\boldsymbol{\Lambda}\alpha\eta + \boldsymbol{I}_M)\,, \qquad (4.32)
\end{aligned}$$

and the constraints

$$\sum_{k=1}^{M} \lambda_k = P_{\mathrm{o}} \tag{4.33}$$

$$\lambda_i \geq 0, \quad i = 1, \cdots, M \tag{4.34}$$

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_M \,. \tag{4.35}$$

It is seen that $\boldsymbol{V}$ is independent of the constraints and that it only affects the first term in (4.32). Keeping $\boldsymbol{\Lambda}$ fixed and following the development in [And63, p. 131], the optimal $\boldsymbol{V}$ can then be chosen as $\boldsymbol{V} = \hat{\boldsymbol{V}}$. For this to hold, (4.35) is needed.

The remaining optimization problem is clearly convex. The solution may therefore be obtained by means of the Karush-Kuhn-Tucker (KKT) conditions [BS93, p. 164]. The approach used for deriving the solution is to temporarily relax the problem by omitting the last constraint (4.35), and then find a set of eigenvalues which satisfies the KKT conditions for the relaxed problem. A detailed derivation is provided in Appendix 4.B. The optimal eigenvalues for the relaxed problem turn out to be given by

$$\lambda_i = \max\left\{0, \frac{\alpha\eta N + \sqrt{\alpha^2\eta^2 N^2 + 4\alpha\eta\hat{\lambda}_i\mu}}{2\alpha\eta\mu} - \frac{1}{\alpha\eta}\right\}, \tag{4.36}$$

where μ is the Lagrange multiplier corresponding to the power constraint. Note that this is also the optimum for the original problem since the above solution automatically satisfies (4.35).

The value of μ is obtained by inserting (4.36) into the power constraint (4.33) and solving the resulting equation. One possible procedure for accomplishing this is now described. To start with, assume that the number of eigenvalues equal to zero in the optimum solution is known. Let $l-1$ denote this quantity. Inserting (4.36) into the power constraint (4.33) then gives the equation

$$P_{\mathrm{o}} + \frac{M-l+1}{\alpha\eta} - \sum_{k=l}^{M} \frac{\alpha\eta N + \sqrt{\alpha^2\eta^2 N^2 + 4\alpha\eta\hat{\lambda}_k\mu}}{2\alpha\eta\mu} = 0\,, \tag{4.37}$$

from which μ can be determined. Let $f(\mu, l)$ represent the left hand side of the equation. Since $f(\mu, l)$ is strictly increasing as a function of μ, the solution is unique and may be found numerically. For example, applying

Newton's method gives rapid convergence. In this case, a suitable starting value is

$$\mu = \frac{\alpha\eta(M-l+1)^2}{(M-l+1+\alpha\eta P_{\mathrm{o}})^2}\left(\hat{\lambda}_l + \frac{N(M-l+1+\alpha\eta P_{\mathrm{o}})}{M-l+1}\right),$$

obtained by using equal power on all eigenvectors whose eigenvalues are assumed to be non-zero and solving for μ for $i = l$. In order to arrive at the correct value of l, an iterative approach is used where, starting at $l = 1$, successive values of l are tried. An algorithm similar to the one utilized when computing the well-known water-filling power profile can be used for this purpose [CT91, p. 253]. The optimal transmit weighting is finally obtained by an appropriate matrix square-root of $\boldsymbol{Z}$. Thus, the whole procedure can be summarized as follows

1. Set $l = 1$
2. Solve $f(\mu, l) = 0$ with respect to μ
3. Compute $\lambda_i = \frac{\alpha\eta N + \sqrt{\alpha^2\eta^2N^2 + 4\alpha\eta\hat{\lambda}_i\mu}}{2\alpha\eta\mu} - \frac{1}{\alpha\eta},\ i = l, \cdots, M$
4. If $\lambda_l < 0$ then set $\lambda_l = 0$, $l = l + 1$ and repeat from 2
5. Compute $\boldsymbol{W}^{(\zeta)} = \hat{\boldsymbol{V}}\boldsymbol{\Lambda}^{1/2}$

As previously mentioned, the proof and hence the algorithm for the simplified scenario can easily be adapted to also handle the important case of an $M \times M'$ transmit weighting where $M' \leq M$. The resulting design procedure is the same as above except that the starting value of l should be modified to $M - M' + 1$ and only the columns of $\boldsymbol{W}^{(\zeta)}$ that correspond to $\{\lambda_i\}_{i=M-M'+1}^{M}$ should be retained from step five. In this way, a simple predetermined code, designed for a small number of transmit antennas, can be used in conjunction with a much larger antenna array. Such a transmission scheme is particularly interesting in view of the fact that OSTB codes with high symbol rates exist for only a limited number of transmit antennas, see [TJC99] and the brief discussion in Section 3.2.3.

4.8 A Simplified Fading Scenario

In this section, the generic assumption concerning the statistics of the channel and the side information is specialized to arrive at a so-called

simplified fading scenario. The simplified fading scenario will be used to illustrate properties of weighted OSTBC. It will also provide the setup for the simulation results presented in the section to follow.

In the simplified fading scenario, it is assumed that the antennas at both the transmitter and the receiver are spaced sufficiently far apart so that the fading is independent. A rich scattering environment with non-line-of-sight conditions is also assumed. It is then reasonable to model the elements $\{h_k\}$ of the channel vector $\boldsymbol{h}$ as zero-mean IID complex Gaussian, corresponding to a situation of spatially uncorrelated Rayleigh fading. Let σ_h^2 denote the variance of each individual channel coefficient. An estimate $\hat{\boldsymbol{h}}$ of $\boldsymbol{h}$ now plays the role of the channel side information. Hence, $\boldsymbol{\zeta} = \hat{\boldsymbol{h}}$, where $\hat{\boldsymbol{h}}$ is modeled in exactly the same way as $\boldsymbol{h}$. It is further assumed that $\hat{\boldsymbol{h}}$ and $\boldsymbol{h}$ are jointly complex Gaussian distributed. Each estimated channel coefficient $\hat{h}_k$ is correlated with the corresponding true channel coefficient h_k, and uncorrelated with all others. In order to describe the degree of correlation, introduce the normalized correlation coefficient $\tilde{\rho} \triangleq \mathrm{E}[h_k \hat{h}_k^*]/\sigma_h^2$. For later reference, note that $0 \leq |\tilde{\rho}| \leq 1$. Thus, the distribution of the true channel and the side information is completely characterized by the mean vectors

$$\boldsymbol{m}_{\boldsymbol{h}} \triangleq \mathrm{E}[\boldsymbol{h}] = \boldsymbol{0}, \quad \boldsymbol{m}_{\hat{\boldsymbol{h}}} \triangleq \mathrm{E}[\hat{\boldsymbol{h}}] = \boldsymbol{0}\,,$$

the covariance matrices

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}} &\triangleq \mathrm{E}[(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}})(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}})^*] = \sigma_h^2 \boldsymbol{I}_{MN} \\ \boldsymbol{R}_{\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}} &\triangleq \mathrm{E}[(\hat{\boldsymbol{h}} - \boldsymbol{m}_{\hat{\boldsymbol{h}}})(\hat{\boldsymbol{h}} - \boldsymbol{m}_{\hat{\boldsymbol{h}}})^*] = \sigma_h^2 \boldsymbol{I}_{MN}\,,\end{aligned}$$

and the cross-covariance matrix $\boldsymbol{R}_{\boldsymbol{h}\hat{\boldsymbol{h}}} \triangleq \mathrm{E}[(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}})(\hat{\boldsymbol{h}} - \boldsymbol{m}_{\hat{\boldsymbol{h}}})^*] = \sigma_h^2 \tilde{\rho} \boldsymbol{I}_{MN}$. Since $\boldsymbol{h}$ and $\boldsymbol{\zeta} = \hat{\boldsymbol{h}}$ are jointly complex Gaussian, it follows that the conditional mean and covariance of $\boldsymbol{h}$ needed in the design procedure are given by [Kay93, p. 509]

$$\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} = \boldsymbol{m}_{\boldsymbol{h}} + \boldsymbol{R}_{\boldsymbol{h}\hat{\boldsymbol{h}}} \boldsymbol{R}_{\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}}^{-1} (\hat{\boldsymbol{h}} - \boldsymbol{m}_{\hat{\boldsymbol{h}}}) = \tilde{\rho} \hat{\boldsymbol{h}} \tag{4.38}$$

$$\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}} = \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}\hat{\boldsymbol{h}}} \boldsymbol{R}_{\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}}^{-1} \boldsymbol{R}_{\boldsymbol{h}\hat{\boldsymbol{h}}}^* = \sigma_h^2 (1 - |\tilde{\rho}|^2) \boldsymbol{I}_{MN}\,. \tag{4.39}$$

These assumptions may model a situation in which the channel side information suffers from a feedback delay, as discussed in Section 4.2.1. In other words, $\hat{\boldsymbol{h}}$ can be thought of as an old channel realization and the magnitude of the correlation coefficient $\tilde{\rho}$ is close to one if the feedback delay is short and correspondingly close to zero if the delay is long.

Although the previous measure of channel quality, as represented by $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$, can be retained in this scenario, we now opt for $\rho \triangleq |\tilde{\rho}|$ as the quantity describing the channel quality. Such a measure was also used in [NLTW98]. Perfect channel knowledge corresponds to $\rho \to 1$. As seen from (4.39), this in turn implies that $\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\| \to 0$. Hence, we also have perfect channel quality as defined by the original channel quality measure. On the other hand, no channel knowledge corresponds to $\rho \to 0$. For this case, $\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}^{-1}\|$ cannot tend to zero since σ_h^2 is fixed. The two measures thus disagree. However, what seems like an inconsistency is really not since the corresponding asymptotically optimal transmit weighting can be shown to be $\boldsymbol{W}_{\text{as}}^{(\boldsymbol{\zeta})} = \boldsymbol{I}_M \sqrt{P_{\text{o}}/M}$, regardless of which of the two quality measures is used. The similarity in the asymptotic results is explained by the inherent symmetry in the distribution implied by (4.38) and (4.39) as $\rho \to 0$. Due to the symmetry, the distribution can be considered to be non-informative from the perspective of the transmitter, resulting in a system without any channel knowledge. This is clearly not true in general, since even if $\boldsymbol{h}$ and $\hat{\boldsymbol{h}}$ are uncorrelated, the distribution of the true channel represents a form of channel knowledge on its own.

4.8.1 Applying Weighted OSTBC

This subsection deals with how the transmission scheme that was described in Section 4.7 can be customized for the simplified scenario. In addition, the behavior of the optimal transmit weighting is studied.

In order to use the transmission scheme for the scenario at hand, α and $\hat{\boldsymbol{\Upsilon}}$ need to be determined. Based on (4.38) and (4.39), it is seen that

$$\hat{\boldsymbol{\Upsilon}} = \frac{1}{\alpha} \boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\zeta}} \boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\zeta}}^{*} = \frac{\rho^2}{\alpha} \hat{\boldsymbol{H}} \hat{\boldsymbol{H}}^{*}, \tag{4.40}$$

where $\alpha = \sigma_h^2(1-\rho^2)$ and $\hat{\boldsymbol{H}}$ is obtained from $\hat{\boldsymbol{h}}$ so that it corresponds to an estimate of $\boldsymbol{H}$. It is now straightforward to apply the algorithm described in Section 4.7.

Let us investigate how the transmission scheme distributes the available output power. Assuming perfect side information, i.e., $\rho \to 1$, the asymptotic result for the perfect channel knowledge case, discussed in Section 4.6, is applicable. Hence, all the power is allocated in the direction of the eigenvector $\boldsymbol{v}_M$, corresponding to the largest eigenvalue of $\hat{\boldsymbol{\Upsilon}}$. In addition, only λ_M in (4.36) is non-zero. In view of (4.40), it is clear that $\boldsymbol{v}_M$ is also the strongest left singular vector of $\hat{\boldsymbol{H}}$.

Consider the other extreme case of no channel knowledge in the sense that $\rho \to 0$. From (4.40) it is obvious that $\hat{\boldsymbol{\Upsilon}}$ then tends to zero, which means that the corresponding eigenvalues $\{\hat{\lambda}_i\}_{i=1}^{M}$ of $\hat{\boldsymbol{\Upsilon}}$ also tend to zero. It follows from (4.36) that $\{\lambda_i\}_{i=1}^{M}$ will all be equal. As a result, $\boldsymbol{W}_{\text{as}}^{(\boldsymbol{\zeta})} = \hat{\boldsymbol{V}}\sqrt{P_{\text{o}}/M}$. Such a transmit weighting implies that $\boldsymbol{Z}_{\text{as}}^{(\boldsymbol{\zeta})} = \boldsymbol{I}_M P_{\text{o}}/M$, which constitutes a transmission scheme equivalent[5] to conventional OSTBC.

For the special case when $\min\{M', N\} \leq 2$, the transmission scheme can be simplified further. To see this, first note that if $N \leq 2$, only two of the eigenvalues $\hat{\lambda}_i$ are possibly non-zero, since $\hat{\boldsymbol{\Upsilon}}$ can be seen as the sum of N rank one matrices of size $M \times M$ and hence has a rank which at most is equal to two. Note also that if $M' \leq 2$, only λ_{M-1} and λ_M are potentially non-zero. In any case, this allows the transmission scheme to be simplified by reorganizing the remaining terms in (4.37) and then squaring repeatedly so that a polynomial equation is obtained.

Obviously, the condition $\min\{M', N\} \leq 2$ covers the situation in which the number of antennas at *either* the transmitter or the receiver is two or lower. For example, consider a system with one receive antenna and assume the simplified scenario. It follows that

$$\hat{\boldsymbol{\Upsilon}} = \frac{\rho^2}{\alpha}\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^* .$$

Analytical expressions for the eigenvalues and the eigenvector corresponding to the largest eigenvalue are easily found to be given by

$$\hat{\lambda}_1 = \cdots = \hat{\lambda}_{M-1} = 0, \qquad \hat{\lambda}_M = \frac{\rho^2}{\alpha}\|\hat{\boldsymbol{h}}\|^2$$

and

$$\boldsymbol{v}_M = \frac{\hat{\boldsymbol{h}}}{\|\hat{\boldsymbol{h}}\|}, \tag{4.41}$$

respectively. Tedious but straightforward calculations now show that the procedure for determining the optimal eigenvalues reduces to

1. Let $\kappa = \alpha(M + \alpha\eta P_{\text{o}})$ and compute
   $\mu = \frac{\eta P_{\text{o}}\left(\kappa(2M-1)+\rho^2\|\hat{\boldsymbol{h}}\|^2+\sqrt{2\kappa(2M-1)\rho^2\|\hat{\boldsymbol{h}}\|^2+\rho^4\|\hat{\boldsymbol{h}}\|^4+\kappa^2}\right)}{2(M+\alpha\eta P_{\text{o}})^2}$
2. Compute $\lambda = \frac{1}{\mu} - \frac{1}{\alpha\eta P_{\text{o}}}$

[5] Equivalent in the sense that the performance as measured by $\ell(\boldsymbol{Z}|\boldsymbol{\zeta})$ is the same.

3. If $\lambda > 0$ then set $\lambda_1 = \cdots = \lambda_{M-1} = \lambda$ and compute $\lambda_M = P_{\mathrm{o}} - (M-1)\lambda$

4. If $\lambda \leq 0$ then set $\lambda_1 = \cdots = \lambda_{M-1} = 0, \quad \lambda_M = P_{\mathrm{o}}$

Although we have assumed the simplified fading scenario, the development easily generalizes to all scenarios where $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ is diagonal. One important example of such a situation is when there are line-of-sight conditions so that the mean value of the true channel is non-zero, like in, for example, an environment with Ricean fading.

By analyzing the above procedure it is possible to make some interesting observations regarding the distribution of power among the eigendirections of the channel. The expression for λ in the second step of the procedure is clearly decreasing as a function of $\|\hat{\boldsymbol{h}}\|$. Hence, when $\|\hat{\boldsymbol{h}}\|$ is above some threshold, the expression after the comparison in step four will be executed and all the power is allocated to the direction of the channel estimate $\hat{\boldsymbol{h}}$. On the other hand, falling below the threshold means that a part of the total power is allocated to $\hat{\boldsymbol{h}}$ and the remaining power is divided equally between the $M-1$ directions orthogonal to the channel estimate.

Recall that σ^2 is inversely proportional to η. A slightly more involved analysis then shows that the power distribution behaves similarly with respect to the noise variance σ^2 as well. The opposite behavior is observed for the fading variance σ_h^2. Thus, when σ_h^2 is below a certain threshold all power is allocated to $\hat{\boldsymbol{h}}$, whereas exceeding the same threshold leads to a portion of the total power being allocated to $\hat{\boldsymbol{h}}$ and the remaining power equally divided among the orthogonal directions.

Since the SNR can be taken to be proportional to σ_h^2/σ^2, the above discussion shows that beamforming in the direction of the channel estimate $\hat{\boldsymbol{h}}$ available at the transmitter is optimal as long as the SNR is below some threshold. Similar results based on information theoretic criteria have been reported in [NLTW98, VM01]. This threshold effect is important for understanding the numerical examples to follow.

In the case of only two transmit antennas, a closed-form expression for $\boldsymbol{W}^*(\boldsymbol{\zeta})$ may be formulated. Recall that $\hat{h}_1$ and $\hat{h}_2$ represent the two elements of $\hat{\boldsymbol{h}}$. The eigenvector $\boldsymbol{v}_2$ is obtained from (4.41), whereas the other remaining eigenvector is obtained by forming a vector orthogonal to $\hat{\boldsymbol{h}}$. It follows from step five in the design procedure of Section 4.7 that

the optimal transmit weighting can be written as

$$\boldsymbol{W}^{(\boldsymbol{\zeta})} = \frac{1}{\sqrt{|\hat{h}_1|^2 + |\hat{h}_2|^2}} \begin{bmatrix} -\hat{h}_2^* & \hat{h}_1 \\ \hat{h}_1^* & \hat{h}_2 \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix}.$$

Thus, the proposed transmission scheme basically consists of a threshold test and some simple computations which are easily implemented using a lookup table. The complexity of the algorithm must therefore be considered very low.

Based on the assumptions in the present section and using the corresponding transmission scheme, the transmit weighting may now be efficiently determined. However, in order for the optimization to be carried out, the variances σ^2 and σ_h^2 and the correlation coefficient ρ must be known, in addition to the channel estimate $\hat{\boldsymbol{h}}$. In practice, σ^2, σ_h^2 and ρ may be estimated at the receiver and fed back to the transmitter. Another approach is to treat these as design parameters chosen such that they roughly match the conditions the system is operating in. Nevertheless, in the simulations to follow, the parameters are assumed to be perfectly known at the transmitter.

4.9 Numerical Examples

In order to examine the performance of the proposed weighted OSTBC transmission scheme, and to investigate how it compares with conventional methods, simulations were conducted for several different cases under the assumptions of the simplified fading scenario. The min max based design procedure in Section 4.5 was utilized for determining the transmit weighting. The resulting performance was compared with three other methods – conventional OSTBC, conventional beamforming and, what is here referred to as, ideal beamforming. Conventional beamforming means that the transmitted signal can be written on the form $\boldsymbol{c}(n) = \boldsymbol{v}s(n)$, where $s(n)$ represents the nth data symbol and where $\boldsymbol{v}$ is a transmit weight vector proportional to the strongest left singular eigenvector of the channel estimate matrix obtained by rearranging $\boldsymbol{\zeta} = \hat{\boldsymbol{h}}$ into an $M \times N$ matrix $\hat{\boldsymbol{H}}$. Ideal beamforming is similar to conventional beamforming except that $\boldsymbol{v}$ is based on the true channel matrix $\boldsymbol{H}$.

For all examined cases, perfect knowledge of σ^2, σ_h^2 and ρ was assumed. The variance of the channel coefficients was arbitrarily set to $\sigma_h^2 = 1$ and the output power was fixed at a level of $P = 1$. The channel

was constant during the transmission of a codeword and independently fading from one codeword to another. Furthermore, all the OSTB encoders under consideration used codes found in [TJC99] with a symbol rate of one. Only square transmit weighting matrices were considered, i.e., $M' = M$. The particular code used in each case is therefore directly determined by the number of transmit antennas. In all the transmission methods, the IID information bearing symbols were taken from a BPSK constellation, resulting in a data rate of one bit per channel use. Throughout the simulations, the bit error rate (BER) was used as the performance measure[6]. Finally, the SNR was measured for the benchmark system using conventional OSTBC and defined as

$$\mathrm{SNR} \triangleq \frac{\mathrm{E}[\|\boldsymbol{H}^*\boldsymbol{C}\|_{\mathrm{F}}^2]}{LN\sigma^2},$$

where $\boldsymbol{C} = \sqrt{P_{\mathrm{o}}/M}\bar{\boldsymbol{C}}$ represents the transmitted signals. The expression for the SNR is equal to the total received average signal energy, divided by the total average noise energy. Since the codes under consideration span as many time instants as the number of transmit antennas, L is here equal to M.

Varying the SNR

In the first case, a system with two transmit antennas and one receive antenna was considered. The channel quality was set to $\rho = 0.9$. The BER as a function of the SNR for the various transmission methods is depicted in Figure 4.2.

It is seen that the performance of the proposed weighted OSTBC transmission scheme with $\rho = 0.9$ is for all SNR values better than conventional OSTBC but, as expected, worse than ideal beamforming. As the SNR decreases, the curve for weighted OSTBC approaches the one for ideal beamforming whereas for increasing SNR it approaches the performance of conventional OSTBC. Thus, the proposed scheme efficiently combines beamforming with OSTBC. This is also in good agreement with both the asymptotic results of Section 4.6 as well as the observations in Section 4.8.1 regarding the allocation of power among the transmit-directions.

[6] Normally, the codeword error probability would have to be used in order to provide the grounds for fair comparisons. However, because the individual binary symbols both in beamforming and in (weighted) OSTBC can be decoded independently (c.f. (3.14)), BER is also a fair performance measure.

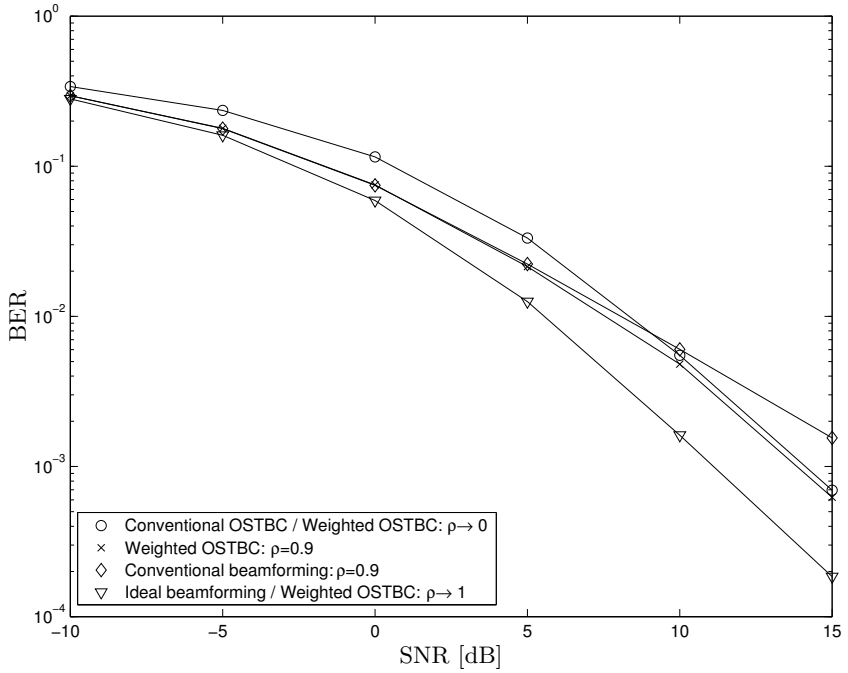

Figure 4.2: The performance of weighted OSTBC compared with OSTBC and beamforming for the case of $M = M' = 2$ transmit antennas and $N = 1$ receive antenna.

Note that the two curves for conventional OSTBC and ideal beamforming also show the performance of weighted OSTBC in the case of $\rho \to 0$ and $\rho \to 1$, respectively. Conventional beamforming is seen to give good performance at low SNR values, but as the SNR increases, the lack of correct channel knowledge leads to a serious performance degradation.

In the second case the number of transmit antennas was increased to eight. This was done in order to illustrate how the number of transmit antennas influences the performance. The channel quality was now set to $\rho = 0.7$. The BER versus the SNR for the four methods are presented in Figure 4.3.

As seen, the potential gains due to channel knowledge are now considerably higher. These gains remain to a large extent even when the number of receive antennas is increased, as illustrated in a comparison between

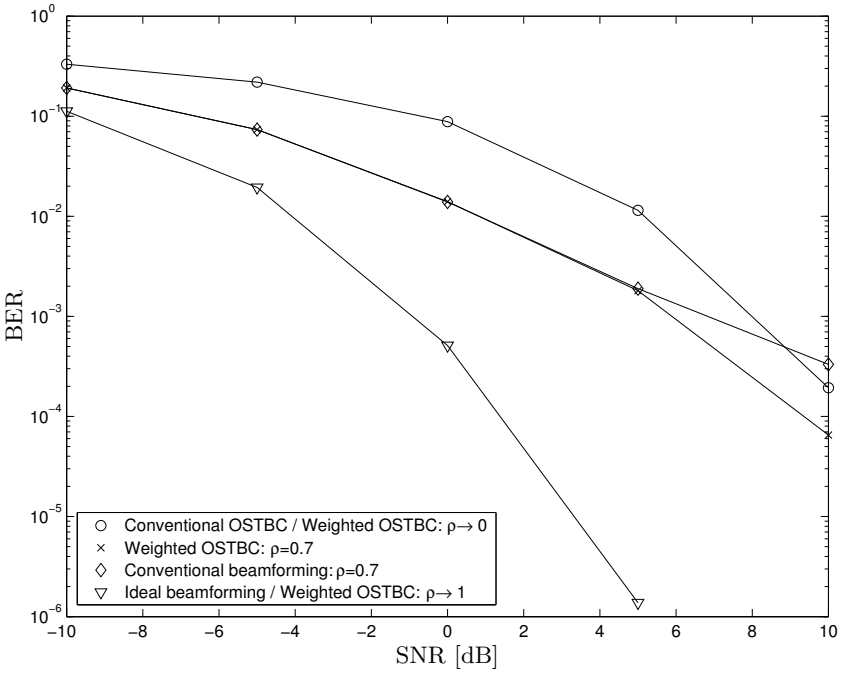

Figure 4.3: The performance of weighted OSTBC compared with OSTBC and beamforming for the case of $M = M' = 8$ transmit antennas and $N = 1$ receive antenna.

the proposed method and conventional OSTBC in Figure 4.4. Although not presented, simulation results demonstrating significant gains were also obtained for scenarios with fewer transmit antennas.

Varying the Channel Quality

The third and last case concerns how the channel quality affects the performance. Again, a system with two transmit antennas and one receive antenna is considered. The SNR was fixed at 10 dB and the BER versus the channel side information quality ρ was plotted. The result is shown in Figure 4.5, which thus provides an illustration of how the proposed scheme adapts to the variations in the channel quality. Hence, when the channel quality is low weighted OSTBC is similar to conventional OS-

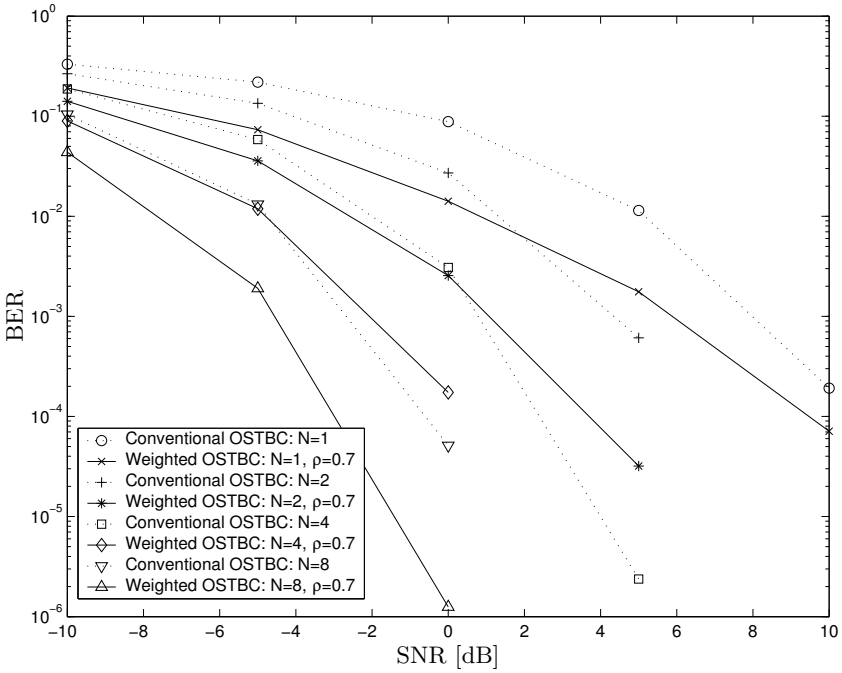

Figure 4.4: The performance of weighted OSTBC compared with OSTBC for different number of receive antennas N while the number of transmit antennas is set to $M = M' = 8$.

TBC and when it is high weighted OSTBC is essentially the same as ideal beamforming.

4.10 Conclusions

In this work, a new performance criterion for the design of space-time codes was derived. The performance criterion takes channel knowledge at the transmitter, if there is any, into account. Design procedures for three types of space-time block codes were proposed. Specific attention was paid to weighted OSTBC and how to design the transmit weighting in an efficient manner. Weighted OSTBC was shown to result in a transmission scheme that can be seen as a seamless combination of OSTB coding and

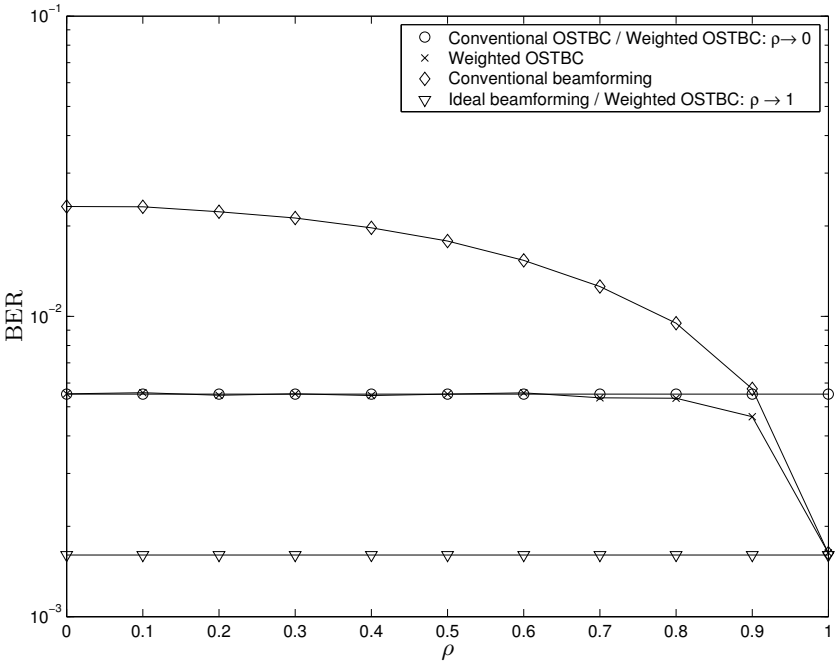

Figure 4.5: Shows how the channel knowledge quality measure ρ affects the performance. Weighted OSTBC is compared with OSTBC and beamforming for a system with $M = M' = 2$ transmit antennas, $N = 1$ receive antenna and a fixed SNR=10 dB.

classical beamforming. The transmit weight design optimization problem was proved to be convex for square transmit weighting matrices, implying that it can be solved in a reasonably efficient manner. Closed-form solutions were derived under certain asymptotic assumptions. Furthermore, the assumption of a simplified fading scenario was shown to result in a particularly efficient optimization algorithm. Numerical results demonstrated significant gains over both an OSTBC system and a system using conventional beamforming.

Appendix 4.A Asymptotic Results

Our strategy for deriving the asymptotic results presented in Section 4.6 is to make use of the fact that, under certain conditions, it is possible to interchange the order of the limit and minimization operator, as exemplified by

$$\begin{aligned}\boldsymbol{x}_{\mathrm{as}} &\triangleq \lim_{\varrho\to a} \arg\min_{\boldsymbol{x}\in\mathcal{X}} V_\varrho(\boldsymbol{x}) \\ &= \arg\min_{\boldsymbol{x}\in\mathcal{X}} \lim_{\varrho\to a} V_\varrho(\boldsymbol{x}) \\ &= \arg\min_{\boldsymbol{x}\in\mathcal{X}} \bar{V}(\boldsymbol{x})\,,\end{aligned}$$

where $\mathcal{X}$ denotes the feasibility set and $\bar{V}(\boldsymbol{x}) \triangleq \lim_{\varrho\to a} V_\varrho(\boldsymbol{x})$. From [SS89, p. 221] it follows that this holds if $V_\varrho(\boldsymbol{x})$ converges uniformly in $\boldsymbol{x}$ over $\mathcal{X}$ to the limit function $\bar{V}(\boldsymbol{x})$, $\mathcal{X}$ is a compact set (i.e., closed and bounded) and $\bar{V}(\boldsymbol{x})$ is continuous and has a unique global minimum.

To apply this theorem for the problem at hand, introduce a criterion function $\ell'(\boldsymbol{Z})$ that is equal to the original criterion function $\ell(\boldsymbol{Z}|\boldsymbol{\zeta})$, except for parameter independent terms and factors. Let $\bar{\ell}_i(\boldsymbol{Z}) \triangleq \lim \ell'(\boldsymbol{Z})$, where the limit will be taken as either $\|\boldsymbol{R}^{-1}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}\| \to 0$, $\eta \to 0$, $\|\boldsymbol{R}_{\boldsymbol{hh}|\hat{\boldsymbol{h}}}\| \to 0$ or $\eta \to \infty$, depending on the asymptotic case under consideration. The index i is here used to number the four different cases. Assume for simplicity that $M' = M$ throughout this appendix so that the rank constraint can be disregarded. Moreover, define the set over which the parameters are optimized as

$$\mathcal{Z}(\varepsilon) \triangleq \{\boldsymbol{Z}|\boldsymbol{Z} = \boldsymbol{Z}^* \succeq \varepsilon\boldsymbol{I}_M, \mathrm{tr}(\boldsymbol{Z}) = P_{\mathrm{o}}\}\,. \tag{4.42}$$

Clearly, the requirement that $\mathcal{Z}(\varepsilon)$ is compact is satisfied. Normally, ε is taken to be zero. The set in (4.42) then corresponds to the feasibility set of the original optimization problem, as described by (4.26). However, in order to satisfy the requirement of a continuous limit function $\bar{\ell}_i(\boldsymbol{Z})$, we will for some of the cases first restrict $\mathcal{Z}(\varepsilon)$ by assuming that ε is small and positive and then argue why we can let $\varepsilon = 0$ without affecting the result. Proving that $\ell'(\boldsymbol{Z})$ converges uniformly to $\bar{\ell}_i(\boldsymbol{Z})$ over $\mathcal{Z}(\varepsilon)$ amounts to showing that

$$\lim \sup_{\boldsymbol{Z}\in\mathcal{Z}(\varepsilon)} |\ell'(\boldsymbol{Z}) - \bar{\ell}_i(\boldsymbol{Z})| = 0\,.$$

In order to simplify the notation, the original criterion function is for

the remaining part of this section written as

$$\ell(\boldsymbol{Z}|\boldsymbol{\zeta}) = \boldsymbol{m}^* \boldsymbol{R}^{-1} \left((\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + \boldsymbol{R}^{-1} \right)^{-1} \boldsymbol{R}^{-1} \boldsymbol{m} - \log\det\left((\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + \boldsymbol{R}^{-1} \right) ,$$

where it is noted that $\eta = \mu_{\min}/(4\sigma^2)$ is a quantity proportional to the SNR. Furthermore, recall that $\|\boldsymbol{X}\| = \sigma_{\max}$, where $\sigma_{\max}$ is the maximum singular value of $\boldsymbol{X}$. Keep also in mind that if the argument is a vector, the result is the usual vector norm.

4.A.1 Case 1: No Channel Knowledge

The first case that is considered is no channel knowledge, i.e., $\|\boldsymbol{R}^{-1}\| \to 0$. Furthermore, it is assumed that $\|\boldsymbol{m}\|$ is bounded as $\|\boldsymbol{R}^{-1}\| \to 0$.

To remove parameter independent terms in the limit function, the equivalent criterion function $\ell'(\boldsymbol{Z}) \triangleq \ell(\boldsymbol{Z}|\boldsymbol{\zeta}) + MN\log(\eta)$ is considered. For now, assume that $\varepsilon > 0$ so as to ensure that $\boldsymbol{Z}$ is invertible. Hence, $\ell'(\boldsymbol{Z})$ can be written as

$$\begin{aligned}\ell'(\boldsymbol{Z}) = {} & \boldsymbol{m}^* \boldsymbol{R}^{-1} \left((\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + \boldsymbol{R}^{-1} \right)^{-1} \boldsymbol{R}^{-1} \boldsymbol{m} \\ & - \log\det\left(\boldsymbol{I}_{MN} + (\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2} \boldsymbol{R}^{-1} (\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2} \right) \\ & - \log\det\left(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta \right) + MN\log(\eta) , \end{aligned} \tag{4.43}$$

where $(\cdot)^{1/2}$ is now a matrix square-root with Hermitian symmetry. As shown below, this function converges uniformly in $\boldsymbol{Z}$ to the limit function

$$\begin{aligned}\bar{\ell}_1(\boldsymbol{Z}) \triangleq \lim_{\|\boldsymbol{R}^{-1}\| \to 0} \ell'(\boldsymbol{Z}) &= -\log\det\left(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta \right) + MN\log\eta \\ &= -N\log\det(\boldsymbol{Z}) . \end{aligned} \tag{4.44}$$

The limit function is obviously continuous. By utilizing Lagrangian multipliers and an EVD of $\boldsymbol{Z}$, it is straightforward to show that $\bar{\ell}_1(\boldsymbol{Z})$, subject to $\boldsymbol{Z} \in \mathcal{Z}(\varepsilon)$, has a unique global minimum $\boldsymbol{Z}_{\text{as}}^{(\boldsymbol{\zeta})} = \boldsymbol{I}_M P_{\text{o}}/M$. This fact, and the uniform convergence in $\boldsymbol{Z}$ over $\mathcal{Z}(\varepsilon)$, implies that, for a fixed positive $\varepsilon \leq P_{\text{o}}/M$,

$$\boldsymbol{Z}_{\text{as}}^{(\boldsymbol{\zeta})} = \lim_{\|\boldsymbol{R}^{-1}\| \to 0} \arg \min_{\boldsymbol{Z} \in \mathcal{Z}(\varepsilon)} \ell'(\boldsymbol{Z}) = \boldsymbol{I}_M P_{\text{o}}/M . \tag{4.45}$$

Note that the above development holds for arbitrarily small $\varepsilon > 0$ and observe that the solution in (4.45) is independent of ε and does not

lie on the boundary defined by the constraint $\boldsymbol{Z} = \boldsymbol{Z}^* \succeq \varepsilon \boldsymbol{I}_M$. Further note that if $\boldsymbol{Z}$ lies on the boundary when $\varepsilon = 0$, i.e., $\boldsymbol{Z}$ is singular, then $\ell'(\boldsymbol{Z})$ grows without bound as $\|\boldsymbol{R}^{-1}\| \to 0$. In contrast, the value of $\ell'(\boldsymbol{Z}_{\text{as}}^{(\boldsymbol{\zeta})})$ is finite also when $\|\boldsymbol{R}^{-1}\| \to 0$. It can therefore be concluded that $\boldsymbol{Z}_{\text{as}}^{(\boldsymbol{\zeta})}$ is the asymptotically optimal solution even if the constraint $\boldsymbol{Z} = \boldsymbol{Z}^* \succeq \varepsilon \boldsymbol{I}_M$ is relaxed by letting $\varepsilon = 0$, so as to correspond to the original optimization problem. Since $\boldsymbol{W}\boldsymbol{W}^* = \boldsymbol{Z}$ and $\boldsymbol{Z}_{\text{as}}^{(\boldsymbol{\zeta})} = \boldsymbol{I}_M P_{\text{o}}/M$, the optimal transmit weighting in the case of no channel knowledge may consequently be chosen as $\boldsymbol{W}_{\text{as}}^{(\boldsymbol{\zeta})} = \boldsymbol{I}_M \sqrt{P_{\text{o}}/M}$.

To see that the convergence is uniform in $\boldsymbol{Z}$ over $\mathcal{Z}(\varepsilon)$ as $\|\boldsymbol{R}^{-1}\| \to 0$, consider the difference

$$\begin{aligned}\ell'(\boldsymbol{Z}) - \bar{\ell}_1(\boldsymbol{Z}) = \boldsymbol{m}^* \boldsymbol{R}^{-1} \big((\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + \boldsymbol{R}^{-1}\big)^{-1} \boldsymbol{R}^{-1} \boldsymbol{m} \\ - \log\det\left(\boldsymbol{I}_{MN} + (\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2} \boldsymbol{R}^{-1} (\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2}\right). \quad (4.46)\end{aligned}$$

From [HJ96, p. 471] it readily follows that

$$(\boldsymbol{A} + \boldsymbol{B})^{-1} \preceq \boldsymbol{B}^{-1}, \qquad \boldsymbol{A} \succeq 0, \quad \boldsymbol{B} \succ 0\,, \tag{4.47}$$

and the first term in (4.46) is therefore upperbounded as

$$|\boldsymbol{m}^* \boldsymbol{R}^{-1} \big((\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + \boldsymbol{R}^{-1}\big)^{-1} \boldsymbol{R}^{-1} \boldsymbol{m}| \le \|\boldsymbol{m}\|^2 \|\boldsymbol{R}^{-1}\|\,.$$

Since the determinant equals the product of the eigenvalues, the second term can be written as

$$\sum_{k=1}^{MN} \log(1 + \lambda_k)\,,$$

where λ_k is the kth eigenvalue of

$$(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2} \boldsymbol{R}^{-1} (\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2}\,.$$

This matrix is Hermitian and positive definite which means that its eigenvalues and its singular values are the same. Thus, the eigenvalues can be upperbounded as

$$\begin{aligned}\lambda_k = \sigma_k \le \sigma_{\max} &= \|(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2} \boldsymbol{R}^{-1} (\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1/2}\| \\ &\le \frac{\|(\boldsymbol{I}_N \otimes \boldsymbol{Z})^{-1/2}\|^2 \|\boldsymbol{R}^{-1}\|}{\eta} \le \frac{\|\boldsymbol{R}^{-1}\|}{\varepsilon\eta}\,,\end{aligned}$$

where σ_k represents the kth singular value and $\sigma_{\max}$ denotes the largest singular value. The second equality is due to the fact that the spectral

norm is equal to the maximum singular value of its argument. An upper bound to the second term in (4.46) may be formed as

$$\sum_{k=1}^{MN} \log(1+\lambda_k) \leq MN \log\left(1 + \frac{\|\boldsymbol{R}^{-1}\|}{\varepsilon\eta}\right) . \tag{4.48}$$

By utilizing the triangle inequality it is now clear that

$$\sup_{\boldsymbol{Z}\in\mathcal{Z}(\varepsilon)} |\ell'(\boldsymbol{Z}) - \bar{\ell}_1(\boldsymbol{Z})| \leq \|\boldsymbol{m}\|^2\|\boldsymbol{R}^{-1}\| + MN \log\left(1 + \frac{\|\boldsymbol{R}^{-1}\|}{\varepsilon\eta}\right) .$$

Since this expression, for a constant $\varepsilon > 0$ and a bounded $\boldsymbol{m}$, obviously tends to zero as $\|\boldsymbol{R}^{-1}\| \to 0$, we have shown that $\ell'(\boldsymbol{Z})$ converges uniformly to $\bar{\ell}_1(\boldsymbol{Z})$ within the parameter set defined by $\mathcal{Z}(\varepsilon)$. This completes the derivation for the no channel knowledge case.

4.A.2 Case 2: Infinite SNR

In the second case it is assumed that the SNR tends to infinity, i.e., $\eta = \mu_{\min}/(4\sigma^2) \to \infty$. Again, we start by assuming $\varepsilon > 0$. Similarly to the previous case, an equivalent criterion function can be written as in (4.43), which also in this case converges uniformly in $\boldsymbol{Z}$ to

$$\bar{\ell}_2(\boldsymbol{Z}) \triangleq \lim_{\eta\to\infty} \ell'(\boldsymbol{Z}) = -N \log\det(\boldsymbol{Z}) = \bar{\ell}_1(\boldsymbol{Z}) .$$

To see that the convergence is uniform, consider the two terms in (4.46). Utilizing (4.47), the first term is now upperbounded by

$$\|\boldsymbol{m}\|^2\|\boldsymbol{R}^{-1}\|^2\|(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)^{-1}\| \leq \frac{\|\boldsymbol{m}\|^2\|\boldsymbol{R}^{-1}\|^2}{\varepsilon\eta} ,$$

whereas the upper bound of the second term is again given by (4.48). Hence, we have

$$\sup_{\boldsymbol{Z}\in\mathcal{Z}(\varepsilon)} |\ell'(\boldsymbol{Z}) - \bar{\ell}_2(\boldsymbol{Z})| \leq \frac{\|\boldsymbol{m}\|^2\|\boldsymbol{R}^{-1}\|^2}{\varepsilon\eta} + MN \log\left(1 + \frac{\|\boldsymbol{R}^{-1}\|}{\varepsilon\eta}\right) ,$$

which obviously tends to zero as the SNR tends to infinity. The convergence is therefore uniform. The arguments following (4.45) then show that the asymptotically optimal transmit weighting may once more be chosen as $\boldsymbol{W}_{\text{as}}^{(\zeta)} = \boldsymbol{I}_M\sqrt{P_{\text{o}}/M}$. This completes the derivation for the case of infinite SNR.

4.A.3 Case 3: Perfect Channel Knowledge

The third case concerns perfect channel knowledge in the sense that $\|\boldsymbol{R}\| \to 0$. In addition, it is assumed that $\bar{\boldsymbol{m}} \triangleq \lim_{\|\boldsymbol{R}\|\to 0} \boldsymbol{m}$ exists and is finite (with respect to some norm). For the present and the next case, it is possible to let $\varepsilon = 0$. The original constraints are therefore assumed.

Parameter independent terms and factors in the limit function are removed by considering the equivalent criterion function

$$\begin{aligned}\ell'(\boldsymbol{Z}) &\triangleq \big(\ell(\boldsymbol{Z}|\boldsymbol{\zeta}) - \log\det(\boldsymbol{R}) - \boldsymbol{m}^*\boldsymbol{R}^{-1}\boldsymbol{m}\big)/\eta \\ &= \boldsymbol{m}^*\big(\boldsymbol{I}_{MN} + (\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)\boldsymbol{R}\big)^{-1}\boldsymbol{R}^{-1}\boldsymbol{m}/\eta \\ &\quad - \log\det\big(\boldsymbol{I}_{MN} + \boldsymbol{R}^{1/2}(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)\boldsymbol{R}^{1/2}\big)/\eta \\ &\quad - \boldsymbol{m}^*\boldsymbol{R}^{-1}\boldsymbol{m}/\eta\,. \end{aligned} \tag{4.49}$$

We start by showing that this function converges uniformly in $\boldsymbol{Z}$ to the obviously continuous limit function

$$\bar{\ell}_3(\boldsymbol{Z}) \triangleq \lim_{\|\boldsymbol{R}\|\to 0} \ell'(\boldsymbol{Z}) = -\bar{\boldsymbol{m}}^*(\boldsymbol{I}_N \otimes \boldsymbol{Z})\bar{\boldsymbol{m}}\,. \tag{4.50}$$

The Taylor series [HJ96, p. 301],

$$(\boldsymbol{I} - \boldsymbol{X})^{-1} = \sum_{k=0}^{\infty} \boldsymbol{X}^k,$$

which is valid if $\|\boldsymbol{X}\| < 1$, is used for writing the first term in (4.49) as

$$\boldsymbol{m}^*\Big(\boldsymbol{R}^{-1} - (\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta) + \sum_{k=2}^{\infty}\big(-(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)\boldsymbol{R}\big)^k\boldsymbol{R}^{-1}\Big)\boldsymbol{m}/\eta\,.$$

By exploiting the triangle inequality and the formula for a geometric series, an upper bound on the infinite sum, for sufficiently small $\eta\|\boldsymbol{R}\|$, is obtained as

$$\begin{aligned}\|\sum_{k=2}^{\infty}(-(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)\boldsymbol{R})^k\boldsymbol{R}^{-1}\| &\le \sum_{k=2}^{\infty}\|\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta\|^k\|\boldsymbol{R}\|^{k-1} \\ &= \frac{\|\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)\|^2\|\boldsymbol{R}\|}{1 - \|\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta\|\|\boldsymbol{R}\|} \\ &= \frac{\eta^2\|\boldsymbol{Z}\|^2\|\boldsymbol{R}\|}{1 - \eta\|\boldsymbol{Z}\|\|\boldsymbol{R}\|} \\ &\le \frac{\eta^2 P_{\mathrm{o}}^2\|\boldsymbol{R}\|}{1 - \eta P_{\mathrm{o}}\|\boldsymbol{R}\|}\,. \end{aligned}$$

For the last inequality, the fact that $\|\boldsymbol{Z}\| \leq P_{\mathrm{o}}$ has been used, which is due to the trace constraint on $\boldsymbol{Z}$. Now, let λ_k represent the kth eigenvalue of

$$\boldsymbol{R}^{1/2}(\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta)\boldsymbol{R}^{1/2}\,.$$

Since it then holds that

$$\lambda_k \leq \|\boldsymbol{R}^{1/2}\|^2\|\boldsymbol{I}_N \otimes \boldsymbol{Z}\eta\| = \eta\|\boldsymbol{R}\|\|\boldsymbol{Z}\| \leq \eta P_{\mathrm{o}}\|\boldsymbol{R}\|\,,$$

the second term in (4.49) is upperbounded by

$$MN\log(1+\eta P_{\mathrm{o}}\|\boldsymbol{R}\|)/\eta\,.$$

Finally, collecting the results for the first two terms yields

$$\sup_{\boldsymbol{Z}\in\mathcal{Z}(0)} |\ell'(\boldsymbol{Z}) - \bar{\ell}_3(\boldsymbol{Z})| \leq \frac{\eta P_{\mathrm{o}}^2\|\boldsymbol{R}\|\|\boldsymbol{m}\|^2}{1-\eta P_{\mathrm{o}}\|\boldsymbol{R}\|} + MN\log(1+\eta P_{\mathrm{o}}\|\boldsymbol{R}\|)/\eta\,.$$

The right hand side clearly tends to zero as $\|\boldsymbol{R}\| \to 0$. Hence, the convergence is uniform.

Changing the sign of the limit function in (4.50) and re-parameterizing using $\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{W}^*$ shows that the optimum of the limit function $\bar{\ell}_3(\boldsymbol{Z})$ is given by

$$\boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})} = \arg \max_{\substack{\boldsymbol{W} \\ \|\boldsymbol{W}\|_{\mathrm{F}}^2 = P_{\mathrm{o}}}} \bar{\boldsymbol{m}}^*(\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\bar{\boldsymbol{m}}\,. \tag{4.51}$$

To solve this, let $\boldsymbol{\Omega} \triangleq \bar{\boldsymbol{m}}\bar{\boldsymbol{m}}^*$ and define

$$\boldsymbol{\Theta} \triangleq \sum_{k=1}^{N} \boldsymbol{\Omega}_k\,, \tag{4.52}$$

where $\boldsymbol{\Omega}_k$ denotes the kth block of size $M \times M$ on the diagonal of $\boldsymbol{\Omega}$. The cost function in the above optimization problem can now be formulated as

$$\begin{aligned}
\bar{\boldsymbol{m}}^*(\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\bar{\boldsymbol{m}} &= \operatorname{tr}\big((\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\boldsymbol{\Omega}\big) \\
&= \operatorname{tr}(\boldsymbol{W}^*\boldsymbol{\Theta}\boldsymbol{W}) \\
&= \big(\operatorname{vec}(\boldsymbol{\Theta}^*\boldsymbol{W})\big)^* \operatorname{vec}(\boldsymbol{W}) \\
&= \big(\operatorname{vec}(\boldsymbol{W})\big)^*(\boldsymbol{I}_M \otimes \boldsymbol{\Theta})\operatorname{vec}(\boldsymbol{W})\,,
\end{aligned} \tag{4.53}$$

where the two last equalities are due to (C.1) and (C.2) in Appendix C, respectively. The power constraint is written on the form $\|\operatorname{vec}(\boldsymbol{W})\| = \sqrt{P_{\mathrm{o}}}$.

The optimization problem is readily solved utilizing the EVD of $\boldsymbol{I}_M \otimes \boldsymbol{\Theta}$. For this purpose, let λ_M denote the largest eigenvalue of $\boldsymbol{\Theta}$ and recall the assumption that it is strictly larger than all the other eigenvalues, i.e., λ_M is unique. It can easily be verified that the eigenvalues of $\boldsymbol{I}_M \otimes \boldsymbol{\Theta}$ are obtained by repeating the eigenvalues of $\boldsymbol{\Theta}$ M times. Hence, λ_M is also the largest eigenvalue of $\boldsymbol{I}_M \otimes \boldsymbol{\Theta}$, with multiplicity M. The set of optimal solutions of (4.51) is therefore given by the eigenspace associated with λ_M. After introducing the complex-valued scalars $\{\mu_k\}_{k=1}^M$, the solution can be written on the form

$$\mathrm{vec}(\boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})}) = \mu_1 \boldsymbol{u}_1 + \mu_2 \boldsymbol{u}_2 + \cdots + \boldsymbol{u}_M \tag{4.54}$$

where

$$\boldsymbol{u}_1 = \begin{bmatrix} \boldsymbol{v}_M \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}, \boldsymbol{u}_2 = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{v}_M \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix}, \cdots, \boldsymbol{u}_M = \begin{bmatrix} \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{v}_M \end{bmatrix}$$

and $\boldsymbol{v}_M$ are the eigenvectors of $\boldsymbol{I}_M \otimes \boldsymbol{\Theta}$ and $\boldsymbol{\Theta}$, respectively, corresponding to λ_M. Here, $\boldsymbol{0}$ is an $M \times 1$ vector with all elements equal to zero. Using (4.54) all the solutions may also be expressed as

$$\boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})} = \begin{bmatrix} \mu_1 \boldsymbol{v}_M & \mu_2 \boldsymbol{v}_M & \cdots & \mu_M \boldsymbol{v}_M \end{bmatrix},$$

implying that $\boldsymbol{Z}_{\mathrm{as}}^{(\boldsymbol{\zeta})} = \boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})} (\boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})})^* = \boldsymbol{v}_M \boldsymbol{v}_M^* \sum_{k=1}^M |\mu_k|^2$. Combining this with the power constraint $\|\boldsymbol{W}\|_{\mathrm{F}}^2 = P_{\mathrm{o}}$ means that $\sum_{k=1}^M |\mu_k|^2 = P_{\mathrm{o}}$, and hence $\boldsymbol{Z}_{\mathrm{as}}^{(\boldsymbol{\zeta})} = \boldsymbol{v}_M \boldsymbol{v}_M^* P_{\mathrm{o}}$. Consequently, regardless of the unit norm vector $\mathrm{vec}(\boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})})$ chosen from the aforementioned eigenspace, it holds that $\boldsymbol{Z}_{\mathrm{as}}^{(\boldsymbol{\zeta})} = \boldsymbol{v}_M \boldsymbol{v}_M^* P_{\mathrm{o}}$, which is thus a unique minimum point of $\bar{\ell}_3(\boldsymbol{Z})$. Accordingly, the use of $\bar{\ell}_3(\boldsymbol{Z})$ in the asymptotic analysis is justified. Letting, for example, $\mu_1 = \sqrt{P_{\mathrm{o}}}, \mu_2 = \cdots = \mu_M = 0$ and utilizing $\boldsymbol{Z}_{\mathrm{as}}^{(\boldsymbol{\zeta})} = \boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})} (\boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})})^*$, an asymptotically optimal solution is given by

$$\boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})} = \sqrt{P_{\mathrm{o}}} \begin{bmatrix} \boldsymbol{v}_M & \boldsymbol{0} & \cdots & \boldsymbol{0} \end{bmatrix}.$$

As previously indicated, the solution is not unique. For example, permuting the columns gives the same value of the cost function.

Recall that $\boldsymbol{m} = \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ and $\boldsymbol{R} = \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}$. Note that (4.52) may also be written on the form $\boldsymbol{\Theta} = \bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}} \bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}}^*$, where $\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}}$ is an $M \times N$ matrix defined by the relation $\mathrm{vec}(\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}}) = \bar{\boldsymbol{m}}_{\boldsymbol{h}|\boldsymbol{\zeta}} \triangleq \bar{\boldsymbol{m}} = \lim_{\|\boldsymbol{R}\| \to 0} \boldsymbol{m}$.

Another equivalent way of defining $\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}}$ is as the asymptotic MMSE channel estimate $\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\zeta}} \triangleq \lim_{\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\|\to 0} \mathrm{E}[\boldsymbol{H}|\boldsymbol{\zeta}]$.

4.A.4 Case 4: Zero SNR

In the fourth case, the SNR is assumed to tend to zero, i.e., $\eta \to 0$. The derivation is to a large extent similar to the previous case. The Taylor expansion,

$$\log\det(\boldsymbol{I}+\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{X}) + \mathrm{O}(\|\boldsymbol{X}\|^2)\,,$$

where $\mathrm{O}(\cdot)$ is the big ordo operator, is used to write the second term of (4.49) as

$$\mathrm{tr}\left(\boldsymbol{R}^{1/2}(\boldsymbol{I}_N \otimes \boldsymbol{Z})\boldsymbol{R}^{1/2}\right) + \mathrm{O}(\eta)\,.$$

Combining this with (4.50) results in the limit function

$$\begin{aligned}\bar{\ell}_4(\boldsymbol{Z}) &\triangleq \lim_{\eta\to 0} \ell'(\boldsymbol{Z}) \\ &= -\boldsymbol{m}^*(\boldsymbol{I}_N \otimes \boldsymbol{Z})\boldsymbol{m} - \mathrm{tr}\left(\boldsymbol{R}^{1/2}(\boldsymbol{I}_N \otimes \boldsymbol{Z})\boldsymbol{R}^{1/2}\right). \end{aligned} \tag{4.55}$$

It is now evident from

$$\sup_{\boldsymbol{Z}\in\mathcal{Z}(0)} |\ell'(\boldsymbol{Z}) - \bar{\ell}_4(\boldsymbol{Z})| \leq \frac{\eta P_{\mathrm{o}}^2 \|\boldsymbol{R}\|\|\boldsymbol{m}\|^2}{1-\eta P_{\mathrm{o}}\|\boldsymbol{R}\|} + |\,\mathrm{O}(\eta)|$$

that the convergence is uniform. Thus, after changing the sign of $\bar{\ell}_4(\boldsymbol{Z})$ and parameterizing in terms of $\boldsymbol{W}$, the cost function can be taken as

$$\boldsymbol{m}^*(\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\boldsymbol{m} + \mathrm{tr}\left(\boldsymbol{R}^{1/2}(\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\boldsymbol{R}^{1/2}\right).$$

Using the relation $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A})$, this expression can be rewritten as

$$\mathrm{tr}\left((\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)(\boldsymbol{m}\boldsymbol{m}^* + \boldsymbol{R})\right) = \mathrm{tr}\left((\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\boldsymbol{\Omega}\right), \tag{4.56}$$

where now $\boldsymbol{\Omega} \triangleq \boldsymbol{m}\boldsymbol{m}^* + \boldsymbol{R}$. Finally, due to the similarity between (4.56) and (4.53), the development from the previous case shows that an asymptotically optimal transmit weighting is given by

$$\boldsymbol{W}_{\mathrm{as}}^{(\boldsymbol{\zeta})} = \sqrt{P_{\mathrm{o}}}\begin{bmatrix}\boldsymbol{v}_M & \boldsymbol{0} & \cdots & \boldsymbol{0}\end{bmatrix},$$

where $\boldsymbol{v}_M$ is the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{\Theta}$. Here, $\boldsymbol{\Theta}$ is again defined as in (4.52). Note that an alternative and perhaps more intuitive formulation of $\boldsymbol{\Theta}$ is given by $\boldsymbol{\Theta} = \mathrm{E}[\boldsymbol{H}|\boldsymbol{\zeta}]\,\mathrm{E}[\boldsymbol{H}|\boldsymbol{\zeta}]^* + \mathrm{E}[(\boldsymbol{H}-\mathrm{E}[\boldsymbol{H}|\boldsymbol{\zeta}])(\boldsymbol{H}-\mathrm{E}[\boldsymbol{H}|\boldsymbol{\zeta}])^*|\boldsymbol{\zeta}]$.

Appendix 4.B An Algorithm for a Simplified Scenario

In this appendix, the solution of the optimization problem defined by (4.32) - (4.35) is derived. It is easily seen that both the cost function and the feasibility set are convex. Thus, the solution is given by the KKT conditions. In order to simplify the development, the optimization problem is temporarily relaxed by omitting (4.35). For the remaining problem, the optimum is given by any $\{\lambda_k\}$ that satisfy the KKT conditions

$$\sum_{k=1}^{M} \lambda_k = P_{\mathrm{o}} \tag{4.57}$$

$$\lambda_i \geq 0 \tag{4.58}$$

$$-\frac{\alpha\eta\hat{\lambda}_i}{(1+\alpha\eta\lambda_i)^2} - \frac{\alpha\eta N}{1+\alpha\eta\lambda_i} + \mu - \nu_i = 0 \tag{4.59}$$

$$\nu_i \geq 0 \tag{4.60}$$

$$\nu_i\lambda_i = 0\,, \tag{4.61}$$

where $i = 1, \cdots, M$ and where μ and ν_i are Lagrange multipliers for the power constraint and the inequality constraints, respectively. We start by solving for ν_i in (4.59) and substituting into (4.60) and (4.61). Thus, the last three conditions reduce to

$$\frac{\alpha\eta\hat{\lambda}_i}{(1+\alpha\eta\lambda_i)^2} + \frac{\alpha\eta N}{1+\alpha\eta\lambda_i} \leq \mu \tag{4.62}$$

$$\lambda_i\left(\mu - \frac{\alpha\eta\hat{\lambda}_i}{(1+\alpha\eta\lambda_i)^2} - \frac{\alpha\eta N}{1+\alpha\eta\lambda_i}\right) = 0\,. \tag{4.63}$$

First, assume that $\lambda_i > 0$. It follows that the second factor in (4.63) must be zero. Rewriting this condition as

$$\mu(1+\alpha\eta\lambda_i)^2 - \alpha\eta N(1+\alpha\eta\lambda_i) - \alpha\eta\hat{\lambda}_i = 0$$

and solving for λ_i gives

$$\lambda_i = \frac{\alpha\eta N + \sqrt{\alpha^2\eta^2N^2 + 4\alpha\eta\hat{\lambda}_i\mu}}{2\alpha\eta\mu} - \frac{1}{\alpha\eta}\,, \tag{4.64}$$

where the positive root was picked due to (4.58). Note that (4.62) is now satisfied by equality. Hence, we have a valid solution as long as (4.64)

gives a positive result and (4.57) is satisfied. On the other hand, for the case of a non-positive result, we let $\lambda_i = 0$. That this indeed satisfies the KKT conditions is seen by verifying that (4.62) is true. Since

$$\frac{\alpha\eta N + \sqrt{\alpha^2\eta^2 N^2 + 4\alpha\eta\hat{\lambda}_i\mu}}{2\alpha\eta\mu} - \frac{1}{\alpha\eta} \leq 0$$

and $\lambda_i = 0$ implies that

$$\mu \geq \alpha\eta\hat{\lambda}_i + \alpha\eta N\,, \tag{4.65}$$

it is obvious that all the KKT conditions, with the possible exception of (4.57), are satisfied. Finally, also (4.57) can be handled by writing the optimal eigenvalues as

$$\lambda_i = \max\left\{0, \frac{\alpha\eta N + \sqrt{\alpha^2\eta^2 N^2 + 4\alpha\eta\hat{\lambda}_i\mu}}{2\alpha\eta\mu} - \frac{1}{\alpha\eta}\right\} \tag{4.66}$$

and then solving for μ in (4.57). It is apparent that (4.66) gives eigenvalues that are sorted in ascending order. Therefore, the constraint that was initially omitted, i.e., (4.35), is automatically satisfied. Thus, (4.66) gives the optimal eigenvalues also for the original problem.

Chapter 5

Quantized Channel Feedback: Design Approach I

This chapter considers a wireless MIMO communication link in which the transmitter has access to quantized channel side information obtained from the receiver via a dedicated feedback link. The focus is on weighted OSTBC, although the methods to be developed herein are applicable to all three code structures mentioned in Section 3.2. Suitable transmit weighting matrices are designed by modifying the previously developed transmit weight design procedure in a heuristic manner to better cope with quantized channel knowledge. An alternative and mathematically cleaner technique for dealing with quantized channel feedback will be described in Chapter 6.

Feedback delay, quantization errors and feedback channel bit-errors are all assumed to impair the quality of the channel side information. To reduce these errors, methods based on vector quantization for noisy channels are used in the design of the feedback link. The proposed transmission scheme and feedback link takes the non-perfect nature of the channel information into account while combining the benefits of conventional beamforming with those given by OSTB codes in a similar manner as in Chapter 4. Simulation results for a spatially uncorrelated flat Rayleigh fading scenario with two transmit antennas and one receive antenna demonstrate significant gains over conventional methods, including

robustness to feedback channel bit-errors.

5.1 Introduction

Recently, space-time codes [GFBK99, TSC98, TJC99] have been developed that utilize both the spatial and temporal dimensions for increasing the performance of a wireless communication link where the transmitter is equipped with multiple antennas. A potential drawback with classical space-time codes is that they are designed under the assumption that the transmitter does not know the channel. Hence, even if channel knowledge is available, it is not used for improving the performance. In some communication systems it is reasonable to assume that such channel side information is in fact available. Both TDD systems and FDD systems equipped with a feedback link belong to this category. In the former case the channel may be estimated in the receive mode and then often assumed to be the same for the transmit mode, whereas for the latter case channel estimates are obtained at the receiver and then transported over a dedicated feedback link to the transmitter.

A common way of utilizing channel side information is to make use of beamforming techniques for maximizing the received energy, essentially steering the transmission in the direction of the receiver. There has recently been a growing interest on the use of feedback information in conjunction with transmission methods related to beamforming [Wit95, NLTW98, HP98, MSA01, VM01, OGDH01, HW01]. The fact that beamforming based on channel information from a feedback link is used in the closed-loop mode of the WCDMA system further illustrates the practical significance of such schemes [3GP02b, DGI$^+$02].

To keep the wireless system spectrally efficient, the channel information is often heavily quantized prior to being transported over the feedback link. Needless to say, the channel knowledge at the transmitter is typically far from perfect. Not only quantization may degrade the quality of the channel information. The quality often suffers from a delay in the feedback link, which means that the channel information is, due to the time-varying nature of the wireless channel, more or less outdated by the time it reaches the transmitter [OGDH01, HW01]. Bit-errors introduced by the channel over which the feedback link operates may also severely impair the performance.

In order to mitigate the detrimental consequences of errors in the channel information, space-time codes can be combined with beamform-

ing type of schemes, as evident from previous chapters in this thesis. The idea is to make use of the complementary strengths of both transmission methodologies. Traditional space-time codes [GFBK99, TSC98, TJC99] are designed to operate without any channel knowledge and hence provide the system with a basic level of performance in the absence of reliable channel state information at the transmitter. Beamforming, on the other hand, is advantageous when the channel knowledge is reliable. Weighted OSTBC in conjunction with the transmit weight design procedure developed in Section 4.5 constitutes a particularly attractive combination of space-time coding and beamforming.

In this work, we consider the use of weighted OSTBC in a MIMO wireless communication link where the transmitter has access to *quantized* channel side information obtained from the receiver via a dedicated feedback link. Suitable transmit weighting matrices are constructed by utilizing the transmit weight design procedure in Section 4.5. However, since that design procedure was developed primarily with Gaussian channel side information in mind, it does not perform well in its original form when the channel information is heavily quantized. To obtain better performance, heuristic modifications to the original design procedure are proposed in this chapter. The design of the feedback link is also addressed and specifically tailored to suit the characteristics of weighted OSTBC.

In addition to feedback delay related errors covered by the methods of the previous chapter, the system under study is assumed to suffer from errors in the channel information due to quantization and feedback channel bit-errors. By designing the feedback link based on vector quantization (VQ) for noisy channels [Far90, FV91], the damaging effects caused by these errors are successfully avoided. Two different types of feedback links are investigated. In the first type, the channel coefficients are quantized directly, like in standard VQ, while in the second type the amplitude and phase are first separated and then quantized. The remaining errors are taken into account by the transmit weight design procedure. Previous work dealing with quantized channel information includes [NLTW98, HP98, MSA01]. However, these papers differ substantially from the present work since they study conventional beamforming and see quantization errors as the only error source[1]. Numerical examples for a spatially uncorrelated flat Rayleigh fading scenario with two transmit

[1]The work in [NLTW98] does in fact also consider errors that can be described by modeling the channel side information as correlated with the true channel. However, that case is studied separately from the case of quantized channel information, not jointly as herein.

antennas and one receive antenna show the robustness against channel information impairments of the proposed setup compared with systems which tentatively assume the channel information at the transmitter to be perfect.

The main contribution in this chapter is the development and study of a transmission method and feedback designs that, used together, are robust against feedback delay, quantization and feedback channel bit-errors. We show how weighted OSTBC can utilize quantized feedback information and also introduce the use of vector quantization for noisy channels into the design of the feedback link. In addition, detection of the transmit weighting matrix at the receiver is treated. The setup considered in this work may be motivated by current standardization proposals for the WCDMA system. For example, an OSTB code is used in the open-loop mode, whereas in the closed-loop mode the receiver informs the transmitter about the appropriate transmit antenna weights based on heavily quantized channel estimates [3GP02a, 3GP02b].

The chapter is organized as follows. Section 5.2 presents the system model including a generic overview of the feedback link. A so-called simplified fading scenario is also described for later reference. In Section 5.3, the transmit weight design procedure of the previous chapter is used to determine a suitable codebook of transmit weights. The design of the two types of feedback links is treated in Section 5.4. In Section 5.5, a method for detecting the transmit weighting at the receiver is developed. Finally, based on the simplified fading scenario, numerical examples are presented in Section 5.6.

5.2 System Model

This work considers a MIMO wireless communication system that utilizes weighted OSTBC based on quantized channel side information $\boldsymbol{\zeta}$ obtained from a feedback link. An illustration of the system under consideration is given in Figure 5.1. As seen, the side information is used for choosing a transmit weighting $\boldsymbol{W} = \boldsymbol{W}^{(\boldsymbol{\zeta})}$ out of a fixed codebook $\{\boldsymbol{W}^{(\boldsymbol{\zeta})}\}$ of weights corresponding to all possible outcomes of $\boldsymbol{\zeta}$. Methods for designing the weighting matrices so as to take the available channel knowledge into account are described later in the chapter.

The generic system model and the assumptions regarding code structures described in Section 3.1 still apply. Hence, there are M transmit and N receive antennas and the received signal vector is given by

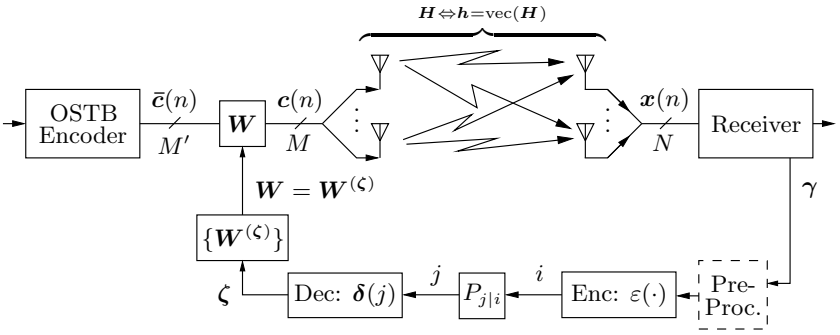

Figure 5.1: Overview of a MIMO system where weighted OSTBC is used in conjunction with quantized channel side information $\boldsymbol{\zeta}$ obtained from a feedback link.

$\boldsymbol{x}(n) = \boldsymbol{H}^* \boldsymbol{c}(n) + \boldsymbol{e}(n)$, where the $M \times N$ matrix $\boldsymbol{H}$ represents the MIMO channel, $\boldsymbol{c}(n)$ the transmitted vector at sample index n and $\boldsymbol{e}(n)$ is a noise term generated from a spatially and temporally white complex Gaussian process. The resulting received block of vectors, corresponding to a transmitted codeword of length L, can be written as $\boldsymbol{X} = \boldsymbol{H}^* \boldsymbol{C} + \boldsymbol{E}$. Since weighted OSTBC is used, it follows from (3.6) that the transmitted codeword $\boldsymbol{C} \in \mathcal{C} = \{\boldsymbol{C}_k\}_{k=1}^K$ is given by $\boldsymbol{C} = \boldsymbol{W}\bar{\boldsymbol{C}}$, where $\bar{\boldsymbol{C}} \in \bar{\mathcal{C}} = \{\bar{\boldsymbol{C}}_k\}_{k=1}^K$ is the codeword output from the OSTB encoder. Hence,

$$\boldsymbol{X} = \boldsymbol{H}^* \boldsymbol{C} + \boldsymbol{E} = \boldsymbol{H}^* \boldsymbol{W} \bar{\boldsymbol{C}} + \boldsymbol{E}\,, \tag{5.1}$$

where the variance of each element of the noise matrix $\boldsymbol{E}$ is equal to σ^2. All codewords in $\bar{\mathcal{C}}$ are assumed equally probable. From (3.15), it follows that the average output power per information bit, conditioned on the channel side information $\boldsymbol{\zeta}$, is limited to P by the power constraint

$$\|\boldsymbol{W}^{(\boldsymbol{\zeta})}\|_{\mathrm{F}}^2 = P_{\mathrm{o}}\,, \tag{5.2}$$

where $P_{\mathrm{o}} \triangleq P \log_2(K)/L_{\mathrm{o}}$.

Although the transmit weighting matrix $\boldsymbol{W}$ may contain more columns than rows so as to map an OSTB code designed for a small antenna array to a larger array, the focus in the present work is on the case when $\boldsymbol{W}$ is square, i.e., $M = M'$. If desired, the methods to be developed herein are easily extended to handle also the case of $M' < M$.

To model the fading, the channel vector $\boldsymbol{h} \triangleq \mathrm{vec}(\boldsymbol{H})$ with elements $\{h_k\}$ is assumed to obey a complex Gaussian distribution described by

the mean vector $\boldsymbol{m}_{\boldsymbol{h}} \triangleq \mathrm{E}[\boldsymbol{h}]$ and the covariance matrix $\boldsymbol{R}_{\boldsymbol{hh}} \triangleq \mathrm{E}[(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}})(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}})^*]$. This is a flexible fading model that can handle both spatially uncorrelated Rayleigh fading as well as more realistic scenarios with correlated fading.

It will be convenient to have a measure of the quality of the channel side information. Since the channel information is quantized, $\boldsymbol{h}|\boldsymbol{\zeta}$ is not complex Gaussian distributed. Hence, unlike in Section 4.2, the conditional covariance $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}} \triangleq \mathrm{E}[(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}})(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}})^*|\boldsymbol{\zeta}]$ does not provide a complete description of the remaining uncertainty in $\boldsymbol{h}$ when the channel information $\boldsymbol{\zeta}$ is known. Nevertheless, it is clear that it constitutes a measure of the reliability of the channel side information. In particular, perfect channel information corresponds to $\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}\| \to 0$, assuming a fixed $\boldsymbol{\zeta}$. The other extreme of no channel knowledge is defined as a scenario in which $\boldsymbol{\zeta}$ is independent of $\boldsymbol{h}$ (so that $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} = \boldsymbol{m}_{\boldsymbol{h}}$ and $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}} = \boldsymbol{R}_{\boldsymbol{hh}}$) and the fading is *non-informative*, i.e., $\boldsymbol{m}_{\boldsymbol{h}}$ is zero and $\boldsymbol{R}_{\boldsymbol{hh}}$ is a scaled identity matrix. This is a reasonable definition, since, as also pointed out in Section 4.8, the symmetrical nature of the statistical distribution of the channel means that all transmit directions are equally good.

5.2.1 An Overview of the Feedback Link

A feedback link is assumed to provide the transmitter with estimates of the current channel realization. In a typical such system, it is desirable to keep the data rate over the feedback link to a minimum in order for the whole system to be spectrally efficient. This often means that the channel estimates must be heavily quantized, which in turn results in a considerable amount of quantization errors. Feedback channel bit-errors and feedback delay may also plague the channel information. To handle these additional problems, we propose the use of techniques related to VQ for noisy channels [Far90, FV91], which take into account that the channel between the encoder and decoder may introduce bit-errors. Based on such procedures, two different types of feedback links are designed in Section 5.4.

In order to model feedback delay, it is assumed that the channel coefficients transmitted over the feedback link are correlated (to an arbitrary degree) with the true channel. As illustrated in Figure 5.1, these channel coefficients are contained in the $MN \times 1$ *initial channel information* vector $\boldsymbol{\gamma}$, with elements $\{\gamma_k\}$. The vectors $\boldsymbol{\gamma}$ and $\boldsymbol{h}$ are assumed to be jointly complex Gaussian distributed. Let $\boldsymbol{m}_{\boldsymbol{\gamma}} \triangleq \mathrm{E}[\boldsymbol{\gamma}]$

represent the mean of $\boldsymbol{\gamma}$ and let $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}} \triangleq \mathrm{E}[(\boldsymbol{h}-\boldsymbol{m}_{\boldsymbol{h}})(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})^*]$ and $\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \triangleq \mathrm{E}[(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})^*]$ denote the relevant cross-covariance matrix and covariance matrix, respectively. Since $\boldsymbol{\gamma}$ and $\boldsymbol{h}$ are correlated, $\boldsymbol{\gamma}$ can be thought of as representing an earlier copy of the true channel $\boldsymbol{h}$, in accordance with the well-known Jakes model [Jak94], which may be used to describe the time-variations of the channel as a temporally correlated stationary complex Gaussian process. Loosely speaking, the correlation properties determine the quality of the initial channel information. Another, more precise, measure is to again use the conditional covariance, in this case $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}} \triangleq \mathrm{E}[(\boldsymbol{h}-\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}})(\boldsymbol{h}-\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}})^*|\boldsymbol{\gamma}]$, where $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}} \triangleq \mathrm{E}[\boldsymbol{h}|\boldsymbol{\gamma}]$, similarly to as when a quality measure for $\boldsymbol{\zeta}$ was defined.

As seen in Figure 5.1, the initial channel information $\boldsymbol{\gamma}$ may, depending on the type of feedback link, be pre-processed[2] prior to being input to the encoder. For the sake of notational brevity, the remaining part of this section assumes that the pre-processing stage is absent. The purpose of the encoder is to map each input source vector $\boldsymbol{\gamma}$ to a corresponding integer i. Since the feedback link is assumed to use b bits for the quantization, the encoder is a mapping $\varepsilon\colon \mathbb{C}^{NM} \to \{0, \cdots, 2^b-1\}$ such that $\varepsilon(\boldsymbol{\gamma}) = i$ for an input vector $\boldsymbol{\gamma}$. The mapping of the encoder is described by $\boldsymbol{\gamma} \in \mathcal{S}_i \Rightarrow \varepsilon(\boldsymbol{\gamma}) = i$, where the set $\{\mathcal{S}_i\}_{i=0}^{2^b-1}$ of encoder regions defines a partition of $\mathbb{C}^{NM}$.

The encoder output i is mapped into bits which are then transmitted over what is here modeled as a memoryless binary symmetric channel with bit-error probability P_{b}. After passing through the feedback channel, the resulting b bits are mapped into an index j representing the output of an equivalent memoryless discrete multilevel input-output channel with transition probabilities $\{P_{j|i} \triangleq \Pr[j|i]\}$, $(i,j) \in \{0, \cdots, 2^b-1\}^2$. Let $d_{\mathrm{h}}(i,j)$ denote the number of bits in which the bit-patterns of i and j differ, the so-called Hamming distance. Since the binary channel is memoryless, it is clear that $P_{j|i} = P_{\mathrm{b}}^{d_{\mathrm{h}}(i,j)}(1-P_{\mathrm{b}})^{b-d_{\mathrm{h}}(i,j)}$.

The output j from the feedback channel is used by the feedback decoder for reconstructing the current channel realization or a related quantity. More precisely, the decoder performs a mapping such that $\boldsymbol{\zeta} = \boldsymbol{\delta}(j)$ for the discrete feedback channel output j. The codebook $\{\boldsymbol{\delta}(j)\}_{j=0}^{2^b-1}$ thus defines the decoder, where $\boldsymbol{\delta}(j)$ represents an estimate of the current channel realization (or related quantity). Based on the outcome of $\boldsymbol{\zeta}$, the transmit weighting $\boldsymbol{W}$ currently in use is chosen, out of a fixed

[2] Although not considered in this work, pre-processing may include predicting the current channel $\boldsymbol{h}$ based on several earlier channel estimates.

codebook $\{\boldsymbol{W}^{(\zeta)}\}$ of 2^b possibly different weights, as $\boldsymbol{W} = \boldsymbol{W}^{(\zeta)}$.

It should now be clear that the feedback decoder and the subsequent block containing the codebook $\{\boldsymbol{W}^{(\zeta)}\}$ may be merged into a single equivalent device with input j and output $\boldsymbol{W}$. The merged device determines the transmit weighting as $\boldsymbol{W} = \tilde{\boldsymbol{W}}^{(j)}$ based on an equivalent codebook $\{\tilde{\boldsymbol{W}}^{(j)}\}_{j=0}^{2^b-1}$, where $\tilde{\boldsymbol{W}}^{(j)} \triangleq \boldsymbol{W}^{(\boldsymbol{\delta}(j))}$. Thus, the codebook is directly indexed by the feedback channel output j. This is actually how the transmission scheme would be implemented in practice. The feedback decoder $\boldsymbol{\delta}(j)$ is here included in the system model only because it will turn out to be useful from a conceptual standpoint when later designing the quantization scheme.

The transmission scheme and the feedback link rely on certain parameters. Because of this, both the transmitter and receiver are assumed to know $\boldsymbol{m}_{\boldsymbol{h}}$, $\boldsymbol{m}_{\boldsymbol{\gamma}}$, $\boldsymbol{R}_{\boldsymbol{hh}}$, $\boldsymbol{R}_{\boldsymbol{h\gamma}}$, $\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}$, σ^2 and P_{b}, which typically vary much slower than the channel matrix. Hence, they can be updated at a relatively low rate, considerably simplifying their distribution over the system.

Note that the transmission scheme and the feedback link are in this work designed separately, i.e., the goal is to first design the feedback link such that a good channel estimate is obtained and then determine a suitable transmit weighting based on this estimate. Another possible approach is of course to design the transmission scheme and the feedback link jointly. The codebook would in this case contain the transmit weights explicitly, like in the equivalent codebook $\{\tilde{\boldsymbol{W}}^{(j)}\}_{j=0}^{2^b-1}$. Preliminary results in this direction may be found in our work in [JSO02b]. However, a major problem with such a joint approach, which motivates the separate design strategy considered in the present work, is that training the quantizer represents a highly computationally demanding optimization problem.

5.2.2 A Simplified Fading Scenario

The present section describes a simple special case of the previously discussed scenario with arbitrary complex Gaussian fading. As will be evident from below, this simplified fading scenario corresponds to the one presented in Section 4.8 but now with a notation that complies with the problem at hand.

In the simplified fading scenario, a rich scattering environment is considered in which the antennas at both the transmitter and the receiver

are spaced sufficiently far apart so that the fading becomes independent. Adhering to the usual assumptions in the space-time coding literature a spatially uncorrelated Rayleigh fading scenario is considered, i.e., the channel coefficients $\{h_k\}$ are modeled as zero-mean IID complex Gaussian with variance equal to σ_h^2. In view of the well-known Jakes fading model [Jak94] and since the initial channel information $\boldsymbol{\gamma}$ is considered to be a delayed copy of the current channel realization, we make the reasonable assumption that $\boldsymbol{h}$ and $\boldsymbol{\gamma}$ are jointly complex Gaussian and equally distributed. See the discussion concerning feedback delay in Section 4.2.1 for a more detailed motivation as to why such an assumption makes sense.

Each initial channel information coefficient γ_k is correlated with the corresponding current channel coefficient h_k and uncorrelated with all others. The degree of correlation is given by the normalized correlation coefficient $\tilde{\rho} \triangleq \mathrm{E}[h_k \gamma_k^*]/\sigma_h^2$. Thus, the distribution of the current channel and the initial channel information is characterized by

$$\begin{aligned} \boldsymbol{m}_{\boldsymbol{h}} &= \boldsymbol{0}, \quad \boldsymbol{m}_{\boldsymbol{\gamma}} = \boldsymbol{0} \\ \boldsymbol{R}_{\boldsymbol{hh}} &= \sigma_h^2 \boldsymbol{I}_{MN}, \quad \boldsymbol{R}_{\boldsymbol{\gamma\gamma}} = \sigma_h^2 \boldsymbol{I}_{MN}, \quad \boldsymbol{R}_{\boldsymbol{h\gamma}} = \sigma_h^2 \tilde{\rho} \boldsymbol{I}_{MN} \,. \end{aligned}$$

From the above and from standard results in estimation theory [Kay93, p. 509] it follows that

$$\begin{aligned} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}} &= \mathrm{E}[\boldsymbol{h}|\boldsymbol{\gamma}] = \tilde{\rho}\boldsymbol{\gamma} \\ \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} &= \mathrm{E}[(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}})(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}})^* | \boldsymbol{\gamma}] = \sigma_h^2 (1 - |\tilde{\rho}|^2) \boldsymbol{I}_{MN} \,. \end{aligned}$$

Let $\rho \triangleq |\tilde{\rho}|$ play the role of a separate quality measure for the initial channel information, in the case of the present simplified fading scenario. Based on such a quality measure, perfect *initial* channel information and no *initial* channel information correspond to $\rho \to 1$ and $\rho \to 0$, respectively. Obviously, the former definition complies with $\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\| \to 0$, while the latter definition corresponds to the previously mentioned case of non-informative fading statistics.

5.3 Determining the Weighting Matrices

Recall that weighted OSTBC utilizes the channel side information for improving a predetermined OSTB code by means of a transmit weighting $\boldsymbol{W} = \boldsymbol{W}^{(\zeta)}$. As previously explained, the weighting is here obtained from a fixed codebook $\{\boldsymbol{W}^{(\zeta)}\}$ of 2^b possibly different weighting matrices. The

strategy in the present chapter for constructing a suitable codebook is to utilize the transmit weight design procedure developed in Section 4.5 for determining the $\boldsymbol{W}^{(\boldsymbol{\zeta})}$'s.

One drawback with such an approach is that the underlying design procedure is not developed specifically for quantized channel side information. Indeed, it is based on the assumption that the channel $\boldsymbol{h}$ conditioned on the channel side information $\boldsymbol{\zeta}$ is complex Gaussian. Such an assumption may in some cases be a good approximation of the statistical relation between $\boldsymbol{h}$ and $\boldsymbol{\zeta}$, particularly if the number of bits used for the quantization is high and the feedback channel does not introduce any bit-errors. This asymptotic argument provides motivation for the strategy considered herein despite the fact that the complex Gaussian assumption is not perfectly valid. Depending on the type of feedback link, the design procedure may need to be modified in a heuristic manner so as to better suit the characteristics of the quantized channel feedback. In any case, the resulting transmission schemes are clearly suboptimal, but still useful as later numerical examples will show.

Applying the above design strategy means that the codebook of transmit weights is obtained based on the design procedure in (4.25), tailored to the problem at hand, to arrive at

$$\boldsymbol{W}^{(\boldsymbol{\zeta})} = \arg \min_{\substack{\boldsymbol{W} \\ \|\boldsymbol{W}\|_{\mathrm{F}}^2 = P_{\mathrm{o}}}} \tilde{\ell}(\boldsymbol{W}; \boldsymbol{m}(\boldsymbol{\zeta}), \boldsymbol{R}(\boldsymbol{\zeta})), \quad \boldsymbol{\zeta} = \boldsymbol{\delta}(j), \quad \forall j\,, \tag{5.3}$$

where

$$\begin{aligned} \tilde{\ell}(\boldsymbol{W}; \boldsymbol{m}, \boldsymbol{R}) \triangleq\ & \boldsymbol{m}^* \boldsymbol{R}^{-1} \big((\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*) \mu_{\min} / (4\sigma^2) + \boldsymbol{R}^{-1} \big)^{-1} \boldsymbol{R}^{-1} \boldsymbol{m} \\ & - \log \det \big((\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*) \mu_{\min} / (4\sigma^2) + \boldsymbol{R}^{-1} \big)\,. \end{aligned} \tag{5.4}$$

As seen, the transmit weight design procedure in (5.3) is formulated in a somewhat different manner than what was done for the original procedure in (4.25). Compared with (4.25), two new parameters $\boldsymbol{m}(\boldsymbol{\zeta})$, $\boldsymbol{R}(\boldsymbol{\zeta})$ now replace the conditional mean $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ and covariance $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}$, respectively. Both $\boldsymbol{m}(\boldsymbol{\zeta})$ and $\boldsymbol{R}(\boldsymbol{\zeta})$ are determined from the output of the feedback link, as explained in the following section where two different types of feedback links will be investigated. This modified notation serves to emphasize that the two parameters will not always represent the conditional mean and covariance.

Techniques for efficiently solving the numerical optimization problem posed by (5.3) were given in Section 4.5. Since the channel feedback is quantized, the transmit weights may be predetermined by designing them

off-line. As a result, the optimization problem in (5.3) is considered solved in the following.

5.4 Feedback Link Design

This section describes the design of two different feedback links. Both provide the transmitter with quantized channel information. The first type employs an encoder that quantizes the channel coefficients directly. It is an uncomplicated design which serves to illustrate the problem and stands on solid theoretical grounds. A drawback is that it wastes bits as simulation results in Section 5.6 will show. To remedy this deficiency, modifications of the basic concept, to some extent based on heuristic arguments, are utilized in the design of the second feedback link. In particular, the channel coefficients are re-mapped prior to the quantization so that the required number of bits is reduced.

Regardless of the type of feedback link, the feedback encoder and decoder are optimized with the aim of minimizing the total distortion due to quantization, feedback delay and feedback channel bit-errors. This is accomplished by using appropriately modified design techniques taken from the field of vector quantization for noisy channels. Although the VQ design is computationally demanding, it may be performed off-line and the result can be stored in lookup tables for quick real-time access. Thus, for a moderate number of bits b, the overall real-time complexity of the feedback link must be considered low.

5.4.1 Feedback Link Type I – Direct Quantization

In the first feedback link that is considered, the channel coefficients are directly quantized. This is illustrated in Figure 5.2, where the initial channel information vector $\boldsymbol{\gamma}$ is input to the encoder. Hence, no pre-processing of $\boldsymbol{\gamma}$ is performed. The decoder is seen to output an estimate $\hat{\boldsymbol{h}}(j) \triangleq \boldsymbol{\delta}(j)$ of the true channel $\boldsymbol{h}$. The estimate, which constitutes the channel information $\boldsymbol{\zeta}$, is not explicitly used by the transmitter. Rather, it is an imaginative output used only for designing a good encoder-decoder pair that in turn produces suitable feedback information. After designing the feedback link, the parameters $\boldsymbol{m}(\boldsymbol{\zeta})$ and $\boldsymbol{R}(\boldsymbol{\zeta})$, needed in the transmit weight design procedure, may be determined based on the feedback link output $\hat{\boldsymbol{h}}(j)$.

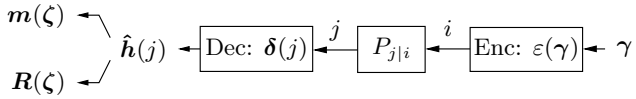

Figure 5.2: Feedback link that quantizes the channel coefficients directly.

Designing the Encoder and the Decoder

Feedback information useful for the transmission scheme is obtained by designing the encoder and the decoder so that the mean-square error between $\hat{\boldsymbol{h}}(j)$ and the true channel $\boldsymbol{h}$ is minimized. We use channel optimized vector quantization (COVQ) [FV91] so that also feedback channel bit-errors are taken into account in the design process. For the problem at hand, the encoder and the decoder are thus considered optimal if

$$D(\{S_i\}, \{\hat{\boldsymbol{h}}(j)\}) \triangleq \mathrm{E}[\|\boldsymbol{h} - \hat{\boldsymbol{h}}(j)\|^2] \tag{5.5}$$

is minimized with respect to the mappings defined by the set of encoder regions $\{S_i\}$ and decoder channel estimates $\{\hat{\boldsymbol{h}}(j)\}$.

The expression in (5.5) is similar to the criterion generally used in the COVQ literature except for the fact that the decoder now attempts to reconstruct $\boldsymbol{h}$ as opposed to reconstructing the source vector $\boldsymbol{\gamma}$. However, with some modifications, (5.5) can be minimized using standard methods [Far90, FV91] for training COVQ:s. In this work, the COVQ is trained using Monte-Carlo simulation of a modified version of the well-known Lloyd algorithm [GG92]. This algorithm alternates between optimizing the encoder while holding the decoder fixed, and optimizing the decoder while holding the encoder fixed, until convergence is achieved. The metrics associated with such a training procedure are now derived.

By conditioning on i and $\boldsymbol{\gamma}$, the distortion measure in (5.5) can be rewritten as $\mathrm{E}\left[\mathrm{E}[\|\boldsymbol{h} - \hat{\boldsymbol{h}}(j)\|^2 | i, \boldsymbol{\gamma}]\right]$. Hence, for the problem at hand, the optimal encoder, assuming the decoder is known and fixed (as defined by $\{\hat{\boldsymbol{h}}(j)\}$), is given by

$$\varepsilon(\boldsymbol{\gamma}) = \arg\min_i \mathrm{E}[\|\boldsymbol{h} - \hat{\boldsymbol{h}}(j)\|^2 | i, \boldsymbol{\gamma}]\,. \tag{5.6}$$

Since $\boldsymbol{h}|\boldsymbol{\gamma}$ is independent of i and j, and $j|i$ is independent of $\boldsymbol{\gamma}$, the

criterion function in (5.6) can be expanded as

$$\begin{aligned}
\mathrm{E}[\|\boldsymbol{h}-\hat{\boldsymbol{h}}(j)\|^2|i,\boldsymbol{\gamma}] &= \sum_{j=0}^{2^b-1} \Pr[j|i,\boldsymbol{\gamma}]\,\mathrm{E}\left[\left(\boldsymbol{h}-\hat{\boldsymbol{h}}(j)\right)^*\left(\boldsymbol{h}-\hat{\boldsymbol{h}}(j)\right)\middle|i,j,\boldsymbol{\gamma}\right] \\
&= \sum_{j=0}^{2^b-1} P_{j|i}\,\mathrm{E}[\boldsymbol{h}^*\boldsymbol{h}-\boldsymbol{h}^*\hat{\boldsymbol{h}}(j)-\hat{\boldsymbol{h}}(j)^*\boldsymbol{h}+\hat{\boldsymbol{h}}(j)^*\hat{\boldsymbol{h}}(j)|i,j,\boldsymbol{\gamma}] \\
&= \sum_{j=0}^{2^b-1} P_{j|i}\big(\mathrm{E}[\boldsymbol{h}^*\boldsymbol{h}|\boldsymbol{\gamma}]-\mathrm{E}[\boldsymbol{h}|\boldsymbol{\gamma}]^*\hat{\boldsymbol{h}}(j)-\hat{\boldsymbol{h}}(j)^*\,\mathrm{E}[\boldsymbol{h}|\boldsymbol{\gamma}]+\hat{\boldsymbol{h}}(j)^*\hat{\boldsymbol{h}}(j)\big) \\
&= \sum_{j=0}^{2^b-1} P_{j|i}\big(\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}-\hat{\boldsymbol{h}}(j)\big)^*\big(\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}-\hat{\boldsymbol{h}}(j)\big)-\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}+\mathrm{E}[\boldsymbol{h}^*\boldsymbol{h}|\boldsymbol{\gamma}]\,,
\end{aligned}$$

where we note that $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}=\mathrm{E}[\boldsymbol{h}|\boldsymbol{\gamma}]$ is the MMSE estimate of the current channel realization $\boldsymbol{h}$ based on $\boldsymbol{\gamma}$. The assumption that $\boldsymbol{h}$ and $\boldsymbol{\gamma}$ are jointly complex Gaussian implies that $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}$ is given by

$$\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}=\boldsymbol{m}_{\boldsymbol{h}}+\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})\,.$$

After omitting terms independent of i, (5.6) simplifies to

$$\varepsilon(\boldsymbol{\gamma})=\arg\min_i \sum_{j=0}^{2^b-1} P_{j|i}\|\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}-\hat{\boldsymbol{h}}(j)\|^2\,. \tag{5.7}$$

Now we turn to the decoder. Since a mean-square error criterion is used, standard results from estimation theory imply that the optimal decoder, given a known encoder (as described by the encoder regions $\{\mathcal{S}_i\}$), is obtained as $\boldsymbol{\delta}(j)=\hat{\boldsymbol{h}}(j)=\mathrm{E}[\boldsymbol{h}|j]$. By again utilizing that $\boldsymbol{h}|\boldsymbol{\gamma}$ is independent of i and j and, moreover, that $\boldsymbol{\gamma}|i$ is independent of j, the decoder is derived as

$$\begin{aligned}
\boldsymbol{\delta}(j)=\hat{\boldsymbol{h}}(j) &= \mathrm{E}[\boldsymbol{h}|j] \\
&= \sum_{i=0}^{2^b-1} \Pr[i|j]\,\mathrm{E}[\boldsymbol{h}|i,j] \\
&= \sum_{i=0}^{2^b-1} P_{i|j}\,\mathrm{E}[\mathrm{E}[\boldsymbol{h}|\boldsymbol{\gamma},i,j]|i,j]
\end{aligned}$$

$$= \sum_{i=0}^{2^b-1} P_{i|j}\, \mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}|i] \tag{5.8}$$

where

$$\mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}|i] = \boldsymbol{m}_{\boldsymbol{h}} + \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}(\mathrm{E}[\boldsymbol{\gamma}|i] - \boldsymbol{m}_{\boldsymbol{\gamma}})$$
$$P_{i|j} \triangleq \frac{P_{j|i}P_i}{\sum_{i=0}^{2^b-1} P_{j|i}P_i},$$

with P_i denoting the probability that the feedback encoder outputs i.

As seen, the channel knowledge contained in $\boldsymbol{\gamma}$ is incorporated into the feedback link design through the MMSE estimate $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}$. This makes sense since the goal is to reproduce the current channel $\boldsymbol{h}$, as opposed to reproducing the source vector $\boldsymbol{\gamma}$ like in standard COVQ. Note that when the MMSE estimate is perfect, so that $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}$ is essentially the same as $\boldsymbol{h}$, the feedback link simplifies to a standard COVQ (as can be verified by setting $\boldsymbol{h}$ in place of $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}$).

Feedback delay, quantization errors and bit-errors in the feedback channel all contribute to the non-perfect nature of the channel information. This is also reflected by the fact that, as shown in Appendix 5.A.1, $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}$, which is a measure of the quality of the channel side information $\boldsymbol{\zeta} = \hat{\boldsymbol{h}}(j)$, can be decomposed into three corresponding terms according to

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}} = \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}} &+ \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\,\mathrm{E}[(\boldsymbol{\gamma} - \mathrm{E}[\boldsymbol{\gamma}|i])(\boldsymbol{\gamma} - \mathrm{E}[\boldsymbol{\gamma}|i])^*|j]\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}^*\\ &+ \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\,\mathrm{E}\left[(\mathrm{E}[\boldsymbol{\gamma}|j] - \mathrm{E}[\boldsymbol{\gamma}|i])(\mathrm{E}[\boldsymbol{\gamma}|j] - \mathrm{E}[\boldsymbol{\gamma}|i])^*\middle|j\right]\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}^*, \end{aligned} \tag{5.9}$$

where the first term can be interpreted as feedback delay distortion (i.e., $\boldsymbol{h}$ and $\boldsymbol{\gamma}$ differ), the second term can be thought of as describing the contribution from quantization errors and the third term corresponds to the feedback channel distortion.

Determining $\boldsymbol{m}(\boldsymbol{\zeta})$ and $\boldsymbol{R}(\boldsymbol{\zeta})$

To construct the codebook of transmit weights, $\boldsymbol{m}(\boldsymbol{\zeta})$ and $\boldsymbol{R}(\boldsymbol{\zeta})$ need to be determined for each of the 2^b possible outcomes of the channel side information vector $\boldsymbol{\zeta} = \boldsymbol{\delta}(j) = \hat{\boldsymbol{h}}(j)$. For the present type of feedback link, we let $\boldsymbol{m}(\boldsymbol{\zeta}) \triangleq \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$, $\boldsymbol{R}(\boldsymbol{\zeta}) \triangleq \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}$, in line with the setup of the original design procedure. In case of an ideal feedback channel, this choice essentially corresponds to modeling the quantization errors as complex

Gaussian. As pointed out in Section 4.2.1, such a model represents a rather crude approximation.

Since it can be assumed that there is a one to one correspondence between $\hat{\boldsymbol{h}}(j)$ and j, the required entities are given by

$$\begin{aligned}\boldsymbol{m}(\boldsymbol{\zeta}) &= \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} = \mathrm{E}[\boldsymbol{h}|j] = \hat{\boldsymbol{h}}(j) \\ \boldsymbol{R}(\boldsymbol{\zeta}) &= \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}} = \mathrm{E}[(\boldsymbol{h} - \mathrm{E}[\boldsymbol{h}|j])(\boldsymbol{h} - \mathrm{E}[\boldsymbol{h}|j])^*|j] \\ &= \mathrm{E}[\boldsymbol{h}\boldsymbol{h}^*|j] - \mathrm{E}[\boldsymbol{h}|j]\,\mathrm{E}[\boldsymbol{h}|j]^* \\ &= \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}} + \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}(\mathrm{E}[\boldsymbol{\gamma}\boldsymbol{\gamma}^*|j] - \mathrm{E}[\boldsymbol{\gamma}|j]\,\mathrm{E}[\boldsymbol{\gamma}|j]^*)\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}^*\,,\end{aligned}$$

where

$$\mathrm{E}[\boldsymbol{\gamma}|j] = \sum_{i=0}^{2^b-1} P_{i|j}\,\mathrm{E}[\boldsymbol{\gamma}|i], \quad \mathrm{E}[\boldsymbol{\gamma}\boldsymbol{\gamma}^*|j] = \sum_{i=0}^{2^b-1} P_{i|j}\,\mathrm{E}[\boldsymbol{\gamma}\boldsymbol{\gamma}^*|i]\,.$$

A more detailed derivation can be found in Appendix 5.A. To implement (5.7) and (5.8) in the VQ design process, and also to compute $\boldsymbol{m}(\boldsymbol{\zeta})$ and $\boldsymbol{R}(\boldsymbol{\zeta})$, the entities P_i, $\mathrm{E}[\boldsymbol{\gamma}|i]$ and $\mathrm{E}[\boldsymbol{\gamma}\boldsymbol{\gamma}^*|i]$ are all replaced by appropriate sample estimates obtained from the Monte-Carlo simulation. The transmit weight design procedure in (5.3) is thereafter used for obtaining the codebook $\{\boldsymbol{W}^{(\boldsymbol{\zeta})}\}$.

Parameter Independent Encoder/Decoder

Clearly, the COVQ depends on a number of parameters related to the joint statistics of $\boldsymbol{h}$ and $\boldsymbol{\gamma}$. From (5.7), (5.8) and the expression for $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}$ it is obvious that $\boldsymbol{m}_{\boldsymbol{h}}$, $\boldsymbol{m}_{\boldsymbol{\gamma}}$, $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}$ and $\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$ directly influence the COVQ. In practice, this may be handled by appropriately quantizing the parameters and designing several COVQ:s off-line, one for each possible realization of the quantized parameters.

Such an approach may be undesirable for complexity reasons. Hence, it would be attractive if the encoder-decoder pair could be held fixed, regardless of the value of these parameters and without loss of performance. Toward this end, note that the dependence upon $\boldsymbol{m}_{\boldsymbol{h}}$ and $\boldsymbol{m}_{\boldsymbol{\gamma}}$ is only superficial and can easily be removed. Because of the expression for $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}$, it is straightforward to show that the COVQ can be trained assuming that $\boldsymbol{h}$ and $\boldsymbol{\gamma}$ are zero-mean vectors, without loss of performance. Consequently, to implement the COVQ, $\boldsymbol{m}_{\boldsymbol{h}}$ is added to the output of the decoder and $\boldsymbol{\gamma} - \boldsymbol{m}_{\boldsymbol{\gamma}}$ is used as input to the encoder.

In general, the COVQ still depends on $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}$ and $\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$. However, when these are equal to scaled identity matrices, it can be shown that also the remaining dependence can be removed. Thus, for the simplified fading scenario the structure of the feedback link reduces to the one illustrated in Figure 5.3. Without loss of performance, the feedback link is trained using $\rho \to 1$ and $\sigma_h^2 = 1$ and is then held fixed. The resulting encoder-decoder pair is used also for other values of ρ and σ_h^2 by normalizing the input to the encoder and computing the channel estimate $\hat{\boldsymbol{h}}(j)$ as the product of the output of the decoder and the true value of $\rho\sigma_h$.

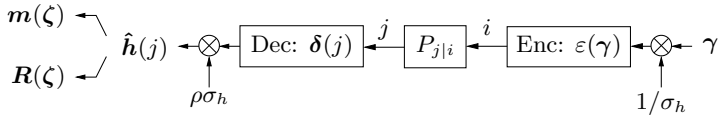

Figure 5.3: Equivalent structure of feedback link for the simplified fading scenario.

Behavior of Weighted OSTBC Based on Quantized Feedback

The transmission scheme and the feedback link cooperate so that robustness against channel information impairments is achieved. To understand what makes the system successful, this section discusses the behavior of the transmission scheme and the feedback link in the two extreme cases of perfect and no channel knowledge, respectively. The discussion is to a large extent also applicable to the second type of feedback link which is presented in Section 5.4.2.

Perfect channel knowledge corresponds to $\|\boldsymbol{R}(\boldsymbol{\zeta})\| = \|\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}\| \to 0$, which means that $\boldsymbol{m}(\boldsymbol{\zeta}) = \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ is essentially equal to the true channel vector $\boldsymbol{h}$. By studying the expression for $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}$ in (5.9) it is realized that in order for the feedback link to output perfect channel information, all of the three error sources must be absent. If this is the case, the first term of the design criterion in (5.4) is seen to dominate. From the previous analysis in Section 4.6, it is clear that this means that the designed transmit weighting is a rank one matrix such that the all the output power is allocated to the transmit-direction given by the strongest left singular eigenvector of the channel matrix $\boldsymbol{H}$. The transmitter is hence seen to perform a sort of beamforming. In particular, in the case of only one receive antenna it holds that $\boldsymbol{H} = \boldsymbol{h}$, which implies that the strongest left singular eigenvector is proportional to $\boldsymbol{h}$. Hence, all the output power

is concentrated in the direction of the current channel realization. It can thus be concluded that when the initial channel information is good, a large number of bits b are used, the feedback channel bit-error probability P_b is small and only one receive antenna is present, the transmission scheme resembles classical beamforming.

Consider now the other extreme of no channel knowledge. In other words, $\boldsymbol{\zeta}$ is independent of $\boldsymbol{h}$ and the channel fading is non-informative so that $\boldsymbol{m}_{\boldsymbol{h}} = \boldsymbol{0}$ and $\boldsymbol{R}_{\boldsymbol{hh}} = \sigma_h^2 \boldsymbol{I}_{MN}$, as is for example the case in the simplified fading scenario in Section 5.2.2. Clearly, $\boldsymbol{\zeta}$ is independent of $\boldsymbol{h}$ if there is no initial channel information (i.e., $\rho = 0$ or $\boldsymbol{\gamma}$ is independent of $\boldsymbol{h}$), $P_\mathrm{b} = 0.5$ and/or $b = 0$. It readily follows that $\boldsymbol{m}(\boldsymbol{\zeta}) = \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} = \boldsymbol{m}_{\boldsymbol{h}} = \boldsymbol{0}$ and $\boldsymbol{R}(\boldsymbol{\zeta}) = \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}} = \boldsymbol{R}_{\boldsymbol{hh}} = \sigma_h^2 \boldsymbol{I}_{MN}$. From the previous development in Section 4.8.1 it is now evident that the designed transmit weighting is a scaled unitary matrix, for example $\boldsymbol{W} = \boldsymbol{I}_M \sqrt{P_\mathrm{o}/M}$. The predetermined codewords in $\bar{\mathcal{C}}$ are hence transmitted without modification, resulting in an open-loop type of system employing conventional OSTBC. This makes sense since the transmitter does not know the channel and therefore has to choose a "neutral" solution. Taking the simplified fading scenario as an example, the transmission scheme thus becomes more and more similar to conventional OSTBC as ρ approaches zero, P_b approaches 0.5 and/or b is decreased.

5.4.2 Feedback Link Type II – Relative Amplitude and Phase

Experimental investigations show that the direct quantization strategy adopted in feedback link type I gives good performance when the number of bits b is large. At least in the case of no bit-errors, this should come as no surprise as the quantization errors are negligible and the scenario hence agrees reasonably well with the assumptions used in the derivation of the performance criterion. However, when b is small, simulation results indicate that there is a need for a distortion measure in the VQ design process that better suits the characteristics of weighted OSTBC.

Accurate estimates of the phases of the channel coefficients are particularly important for obtaining good performance. The signals from the different transmitter antennas might otherwise cancel each other, regardless of the quality of the channel coefficients' amplitude estimates. This agrees well with previous work where primarily the phases of the channel coefficients are taken into account [HP98, MSA01]. In addition, for conventional beamforming it is sufficient for the transmitter to only have

knowledge about the amplitudes and phases relative one of the channel coefficients [NLTW98]. Since the number of degrees of freedom that must be quantized decreases, the number of bits b can be reduced even further without sacrificing performance. To enjoy these benefits also in the transmission methods of the present work, this section presents a feedback link where the design of the VQ is based on a distortion measure that measures the channel information error in terms of relative amplitudes and phases.

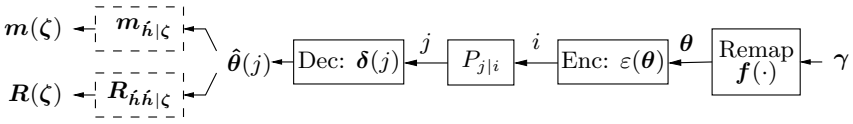

Figure 5.4: Feedback link based on quantizing the relative amplitudes and phases of the channel coefficients.

An illustration of this second type of feedback link is provided in Figure 5.4. Prior to quantization, the initial channel information $\boldsymbol{\gamma}$ is remapped in two steps. In the first step, the elements of $\boldsymbol{\gamma}$ are all divided by the first element γ_1, producing an $MN \times 1$ vector

$$\acute{\boldsymbol{\gamma}} \triangleq \boldsymbol{\gamma}/\gamma_1 = \begin{bmatrix} 1 & \gamma_2/\gamma_1 & \gamma_3/\gamma_1 & \cdots & \gamma_{MN}/\gamma_1 \end{bmatrix}^{\mathrm{T}} \triangleq \begin{bmatrix} 1 & \tilde{\boldsymbol{\gamma}}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} . \quad (5.10)$$

The element corresponding to $\gamma_1/\gamma_1 = 1$ does not need to be quantized and is hence discarded. The second step is to split the remaining elements $\tilde{\boldsymbol{\gamma}}$ into their corresponding amplitudes and phases, resulting in the $(2MN-2) \times 1$ vector

$$\boldsymbol{\theta} \triangleq \begin{bmatrix} \boldsymbol{\theta}_{\mathrm{a}}^{\mathrm{T}} & \boldsymbol{\theta}_{\mathrm{p}}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} , \quad (5.11)$$

where

$$\boldsymbol{\theta}_{\mathrm{a}} \triangleq \begin{bmatrix} |\gamma_2|/|\gamma_1| & \cdots & |\gamma_{MN}|/|\gamma_1| \end{bmatrix}^{\mathrm{T}} \quad (5.12)$$

$$\boldsymbol{\theta}_{\mathrm{p}} \triangleq \begin{bmatrix} \arg(\gamma_2/\gamma_1) & \cdots & \arg(\gamma_{MN}/\gamma_1) \end{bmatrix}^{\mathrm{T}} \quad (5.13)$$

and where $\arg(\cdot)$ now denotes the phase, in the range $-\pi$ to π, of its complex valued argument. Thus, (5.10) - (5.13) together define a function $\boldsymbol{f} : \mathbb{C}^{MN-1} \rightarrow \mathbb{R}^{2MN-2}$, where $\mathbb{R}$ denotes the real number field, such that $\boldsymbol{\theta} = \boldsymbol{f}(\boldsymbol{\gamma})$. The relative amplitudes and phases in $\boldsymbol{\theta}$ are encoded, transmitted over the feedback channel and decoded into an estimate $\hat{\boldsymbol{\theta}}(j) \triangleq \boldsymbol{\delta}(j)$ of $\boldsymbol{\theta}$, where $\hat{\boldsymbol{\theta}}(j)$ represents the channel information

$\boldsymbol{\zeta}$. Based on the channel information, some intermediate entities $\boldsymbol{m}_{\acute{\boldsymbol{h}}|\boldsymbol{\zeta}}$ and $\boldsymbol{R}_{\acute{\boldsymbol{h}}\acute{\boldsymbol{h}}|\boldsymbol{\zeta}}$ are first computed and then transformed into $\boldsymbol{m}(\boldsymbol{\zeta})$ and $\boldsymbol{R}(\boldsymbol{\zeta})$. These will be determined in such a way that they mimic the behavior of the conditional mean and covariance utilized in the first type of feedback link.

Designing the Encoder and the Decoder

The VQ may be designed so as to make the error between the source vector $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}(j)$ as small as possible in the mean-square error sense. A problem with such an approach for the present feedback link is that it does not take into account that the phase wraps around, or in other words, phase values that are equal modulo 2π are equivalent. For example, two vectors that are close to, but on opposite sides of, the negative real axis should not be considered "far apart". However, this is precisely the case when the distance is measured in terms of phase difference. Neglecting the phase wrapping effect is feasible in many cases since experimental studies show that the designs work well anyway. Nevertheless, when bit-errors are introduced by the feedback channel, the performance may suffer due to the resulting mapping of codebook vectors to binary representations. To avoid such a poor mapping, a so-called index assignment algorithm can be used on a previously designed VQ for reordering the vectors in the codebook in a better way. While it seems difficult to directly incorporate phase wrapping into the design of the VQ, it is straightforward to use a distortion measure in the index assignment algorithm that takes the phase wrapping effect into account.

The above discussion motivates us to take on a two-step approach in which the VQ is first designed without regard to bit-errors and then modified using an index assignment algorithm to achieve robustness against bit-errors.

In the first step, an initial codebook is found by training the VQ in a rather conventional manner under the assumption that there are no feedback channel bit-errors, i.e., $P_{\mathrm{b}} = 0$ and hence $j = i$. To obtain an appropriate tradeoff between the importance of amplitude and phase errors, the mean-square error between $\boldsymbol{\theta}$ and the decoder output $\hat{\boldsymbol{\theta}}(i)$ is weighted and the encoder and decoder are considered optimal if

$$D(\{S_i\}, \{\hat{\boldsymbol{\theta}}(i)\}) \triangleq \mathrm{E}[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(i)\|_{\boldsymbol{K}}^2] \tag{5.14}$$

is minimized. The encoder regions $\{S_i\}$ now form a partition of the set of possible $\boldsymbol{\theta}$ and the squared norm $\|\cdot\|_{\boldsymbol{K}}^2$ is defined through $\|\boldsymbol{x}\|_{\boldsymbol{K}}^2 = \boldsymbol{x}^* \boldsymbol{K} \boldsymbol{x}$,

where $\boldsymbol{K}$ constitutes a Hermitian positive definite weighting matrix. An optimal choice of $\boldsymbol{K}$ is beyond the scope of this work and $\boldsymbol{K}$ is instead chosen empirically until satisfactory results are obtained. The encoder and decoder are again found by minimizing (5.14) using Monte-Carlo simulation of the Lloyd algorithm. Under the above conditions, it is straightforward to verify that the required expression for the encoder, assuming a fixed decoder, is given by

$$\varepsilon(\boldsymbol{\theta}) = \arg\min_i \mathrm{E}[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(i)\|^2_{\boldsymbol{K}} | i, \boldsymbol{\theta}] = \arg\min_i \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(i)\|^2_{\boldsymbol{K}}\,, \tag{5.15}$$

whereas the decoder, assuming a fixed encoder, is given by

$$\boldsymbol{\delta}(i) = \hat{\boldsymbol{\theta}}(i) = \mathrm{E}[\boldsymbol{\theta}|i]\,. \tag{5.16}$$

In the second and final step of the design procedure, the probability of bit-error P_b is set at a high level, in our case 5%, and the average distortion $\mathrm{E}[d(\hat{\boldsymbol{\theta}}(i), \hat{\boldsymbol{\theta}}(j))]$ due to bit-errors is reduced by evaluating different permutations of the codebook vectors. For this purpose, an index assignment algorithm presented in [Far90] is used, but with phase wrapping incorporated into a modified distortion measure

$$d\big(\hat{\boldsymbol{\theta}}(i), \hat{\boldsymbol{\theta}}(j)\big) = \|\hat{\boldsymbol{\theta}}_\mathrm{a}(i) - \hat{\boldsymbol{\theta}}_\mathrm{a}(j)\|^2_{\boldsymbol{K}_\mathrm{a}} + \big\|\boldsymbol{q}\big(\hat{\boldsymbol{\theta}}_\mathrm{p}(i) - \hat{\boldsymbol{\theta}}_\mathrm{p}(j)\big)\big\|^2_{\boldsymbol{K}_\mathrm{p}}\,, \tag{5.17}$$

where $\boldsymbol{K}_\mathrm{a}$ and $\boldsymbol{K}_\mathrm{p}$ are the sub-matrices on the main diagonal of $\boldsymbol{K}$ corresponding to $\boldsymbol{\theta}_\mathrm{a}$ and $\boldsymbol{\theta}_\mathrm{p}$, respectively. Hence, $\boldsymbol{K}$ is assumed to be block diagonal. Moreover, $\boldsymbol{q}(\boldsymbol{x})$ is an $(MN-1)\times 1$ vector valued function with its kth element given by

$$q(x_k) \triangleq \begin{cases} x_k, & |x_k| \le \pi \\ 2\pi - x_k, & \pi < x_k < 2\pi \\ -2\pi - x_k, & -2\pi < x_k < -\pi\,. \end{cases} \tag{5.18}$$

As seen from (5.17), the distortion measure is fairly standard except that phase difference is measured according to (5.18) so as to take phase wrapping into account.

Determining $\boldsymbol{m}(\boldsymbol{\zeta})$ and $\boldsymbol{R}(\boldsymbol{\zeta})$ Based on Heuristic Arguments

In order for the transmission scheme to use (5.3) for determining good transmit weighting matrices, appropriate values for $\boldsymbol{m}(\boldsymbol{\zeta})$ and $\boldsymbol{R}(\boldsymbol{\zeta})$ based on the channel information $\hat{\boldsymbol{\theta}}(j)$ need to be found. It is important to

realize that a direct computation of, for example, $\boldsymbol{m}(\boldsymbol{\zeta})$ according to $\boldsymbol{m}(\boldsymbol{\zeta}) \triangleq \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}} = \mathrm{E}[\boldsymbol{h}|\hat{\boldsymbol{\theta}}(j)]$ will not work since information necessary for reconstructing $\boldsymbol{h}$ has been inevitably lost when $\boldsymbol{\gamma}$ was divided by γ_1. Because of the loss of information, the MMSE estimate $\mathrm{E}[\boldsymbol{h}|\hat{\boldsymbol{\theta}}(j)]$ will be poor. Hence, such an estimate is almost useless for conveying channel information to the transmitter. The need for a practical method of determining useful expressions replacing the obvious, but naive, expressions for $\boldsymbol{m}(\boldsymbol{\zeta})$ and $\boldsymbol{R}(\boldsymbol{\zeta})$ naturally arises. The remaining part of this section is therefore devoted to showing how $\hat{\boldsymbol{\theta}}(j)$ can be put into practical use by the proposed transmission scheme.

As indicated by Figure 5.4, the approach in this work is to compute the conditional mean and covariance for $\acute{\boldsymbol{h}} \triangleq \boldsymbol{h}/h_1$, rather than for $\boldsymbol{h}$. Since the number of degrees of freedom in $\acute{\boldsymbol{h}}$ is similarly reduced as in $\hat{\boldsymbol{\theta}}(j)$, such an approach will not suffer from the consequences of information loss as previously described. The use of $\acute{\boldsymbol{h}}$ is justified as follows. Consider the performance criterion in (5.4) with the parameters set to $\boldsymbol{m}(\boldsymbol{\zeta}) \triangleq \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$, $\boldsymbol{R}(\boldsymbol{\zeta}) \triangleq \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}$ and assume perfect channel knowledge at the transmitter, i.e., $\|\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}\| \to 0$ for a fixed $\boldsymbol{\zeta}$. Also, assume that $\bar{\boldsymbol{m}}_{\boldsymbol{h}|\boldsymbol{\zeta}} \triangleq \lim_{\|\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}}\|\to 0} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ exists and is finite. Consequently, the first term in (5.4) dominates and the performance criterion $\tilde{\ell}(\boldsymbol{W}; \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}, \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\zeta}})$ is approximately equivalent to

$$\bar{\ell}(\boldsymbol{W}) \triangleq -\bar{\boldsymbol{m}}^*_{\boldsymbol{h}|\boldsymbol{\zeta}}(\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\bar{\boldsymbol{m}}_{\boldsymbol{h}|\boldsymbol{\zeta}}\,,$$

as evident from (4.50). Due to the perfect channel knowledge assumption, it holds that $\bar{\boldsymbol{m}}_{\boldsymbol{h}|\boldsymbol{\zeta}} \approx \boldsymbol{h}$ and therefore also

$$\bar{\bar{\ell}}(\boldsymbol{W}) \triangleq -\boldsymbol{h}^*(\boldsymbol{I}_N \otimes \boldsymbol{W}\boldsymbol{W}^*)\boldsymbol{h} \approx \bar{\ell}(\boldsymbol{W})\,, \tag{5.19}$$

with good accuracy. Similar to conventional beamforming, if $\acute{\boldsymbol{h}} = \boldsymbol{h}/h_1$ is substituted for $\boldsymbol{h}$ in (5.19) it is clear that neither the amplitude nor the phase of h_1 significantly affects the weighting $\boldsymbol{W}$ that is obtained from the corresponding minimization problem. This roughly provides motivation for using $\acute{\boldsymbol{h}}$ in place of $\boldsymbol{h}$ in the performance criterion, i.e., we wish to replace $\boldsymbol{m}(\boldsymbol{\zeta})$ and $\boldsymbol{R}(\boldsymbol{\zeta})$ with

$$\begin{aligned} \boldsymbol{m}_{\acute{\boldsymbol{h}}|\boldsymbol{\zeta}} &\triangleq \mathrm{E}[\acute{\boldsymbol{h}}|j] = \sum_{i=0}^{2^b-1} P_{i|j}\,\mathrm{E}[\acute{\boldsymbol{h}}|i] \\ \boldsymbol{R}_{\acute{\boldsymbol{h}}\acute{\boldsymbol{h}}|\boldsymbol{\zeta}} &\triangleq \mathrm{E}[(\acute{\boldsymbol{h}} - \mathrm{E}[\acute{\boldsymbol{h}}|j])(\acute{\boldsymbol{h}} - \mathrm{E}[\acute{\boldsymbol{h}}|j])^*|j] \end{aligned} \tag{5.20}$$

$$= \mathrm{E}[\acute{\boldsymbol{h}}\acute{\boldsymbol{h}}^{*}|j] - \mathrm{E}[\acute{\boldsymbol{h}}|j]\,\mathrm{E}[\acute{\boldsymbol{h}}|j]^{*}$$
$$= \sum_{i=0}^{2^b-1} P_{i|j}\,\mathrm{E}[\acute{\boldsymbol{h}}\acute{\boldsymbol{h}}^{*}|i] - \boldsymbol{m}_{\acute{\boldsymbol{h}}|\boldsymbol{\zeta}}\boldsymbol{m}_{\acute{\boldsymbol{h}}|\boldsymbol{\zeta}}^{*}\,, \tag{5.21}$$

respectively. However, these asymptotically motivated entities need to be slightly modified before they can be successfully used in scenarios where the channel knowledge is not perfect. Specifically, three issues need to be addressed:

1. To approximately keep the original proportion between the first and second term in the performance criterion (5.4), the lost amplitude information is partially compensated for by multiplying $\acute{\boldsymbol{h}}$ with the mean $\mathrm{E}[|h_1|]$.

2. Since $\acute{h}_1 \equiv 1$, element $(1,1)$ of $\boldsymbol{R}_{\acute{\boldsymbol{h}}\acute{\boldsymbol{h}}|\boldsymbol{\zeta}}$ will be zero, which means that $\boldsymbol{R}_{\acute{\boldsymbol{h}}\acute{\boldsymbol{h}}|\boldsymbol{\zeta}}$ is singular. A reasonable fix for this problem is to set element $(1,1)$ to the average variance of the other elements. In other words, the element is set to $\mathrm{tr}(\boldsymbol{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}})/(MN-1)$, where $\boldsymbol{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}}$ represents the $(MN-1)\times(MN-1)$ lower right block of $\boldsymbol{R}_{\acute{\boldsymbol{h}}\acute{\boldsymbol{h}}|\boldsymbol{\zeta}}$ and $\tilde{\boldsymbol{h}}$ is defined through $\acute{\boldsymbol{h}} = \begin{bmatrix} 1 & \tilde{\boldsymbol{h}}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$. This modification makes intuitive sense since it means that $\boldsymbol{R}_{\acute{\boldsymbol{h}}\acute{\boldsymbol{h}}|\boldsymbol{\zeta}}$ can be viewed as a parameter which seamlessly tunes between conventional beamforming and OSTBC.

3. The use of (5.20), (5.21) and the two previous modifications in conjunction with the transmission scheme generally gives good performance. However, simulation results indicate that the transmission scheme overcompensates for large quantization errors and thereby steers the weighting $\boldsymbol{W}$ too close toward an open-loop solution. Conventional beamforming is on the other hand relatively robust against pure quantization errors and provides acceptable performance for such scenarios. In cases where only quantization errors affect the quality of the channel information, the performance of weighted OSTBC can be improved by making the conditional covariance small, since the transmission scheme then resembles classical beamforming and thus inherits its for this case good performance. It is of course highly desirable if such a modification also works well when other errors influence the channel information. As is shown next, both these latter aspects are taken into account by

removing a part of the conditional covariance that can be considered to be due to quantization errors.

Recall the result in (5.9) that $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$, assuming a feedback link of type I, can be decomposed into three terms. One of the terms may be attributed to quantization errors. Obviously, the assumptions used when deriving that result do not hold in the present case. Nevertheless, (5.9) is still reasonable for the feedback link type under consideration so inspired by the expression for the term in (5.9) that corresponds to quantization errors, we take

$$\begin{aligned}\boldsymbol{R}^{(q)}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}} &\triangleq \boldsymbol{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{\gamma}}}\boldsymbol{R}^{-1}_{\tilde{\boldsymbol{\gamma}}\tilde{\boldsymbol{\gamma}}}\,\mathrm{E}[(\tilde{\boldsymbol{\gamma}}-\mathrm{E}[\tilde{\boldsymbol{\gamma}}|i])(\tilde{\boldsymbol{\gamma}}-\mathrm{E}[\tilde{\boldsymbol{\gamma}}|i])^*|j]\boldsymbol{R}^{-1}_{\tilde{\boldsymbol{\gamma}}\tilde{\boldsymbol{\gamma}}}\boldsymbol{R}^{*}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{\gamma}}}\\ &= \boldsymbol{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{\gamma}}}\boldsymbol{R}^{-1}_{\tilde{\boldsymbol{\gamma}}\tilde{\boldsymbol{\gamma}}}(\mathrm{E}[\tilde{\boldsymbol{\gamma}}\tilde{\boldsymbol{\gamma}}^*|j]-\sum_{i=0}^{2^b-1}P_{i|j}\,\mathrm{E}[\tilde{\boldsymbol{\gamma}}|i]\,\mathrm{E}[\tilde{\boldsymbol{\gamma}}|i]^*)\boldsymbol{R}^{-1}_{\tilde{\boldsymbol{\gamma}}\tilde{\boldsymbol{\gamma}}}\boldsymbol{R}^{*}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{\gamma}}} \quad (5.22)\end{aligned}$$

as a measure of the contribution of the quantization distortion to $\boldsymbol{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}}$. Motivated by the above discussion, $\boldsymbol{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}}$ is replaced with $\boldsymbol{R}^{(f+c)}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}} \triangleq \boldsymbol{R}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}} - \boldsymbol{R}^{(q)}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}}$. Examining simulation results shows that this pragmatic approach works well and improves the performance in scenarios with coarsely quantized channel information.

Thus, to summarize, incorporating the three heuristic modifications mentioned above into the system means that the parameters passed to the transmission scheme are set to

$$\begin{aligned}\boldsymbol{m}(\boldsymbol{\zeta}) &\triangleq \mathrm{E}[|h_1|]\boldsymbol{m}_{\acute{\boldsymbol{h}}|\boldsymbol{\zeta}}\\ \boldsymbol{R}(\boldsymbol{\zeta}) &\triangleq \mathrm{E}[|h_1|]^2\begin{bmatrix}\mathrm{tr}\left(\boldsymbol{R}^{(f+c)}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}}\right)/(MN-1) & \boldsymbol{0}^{\mathrm{T}}\\ \boldsymbol{0} & \boldsymbol{R}^{(f+c)}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}}\end{bmatrix}.\end{aligned}$$

Because $\boldsymbol{R}^{(f+c)}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}}$ is computed as a difference between two matrices, $\boldsymbol{R}^{(f+c)}_{\tilde{\boldsymbol{h}}\tilde{\boldsymbol{h}}|\boldsymbol{\zeta}}$ and hence $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ is not necessarily positive definite. This problem is avoided by setting negative eigenvalues of $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ to zero and adding a small regularizing term $\epsilon\boldsymbol{I}_{MN}$, $\epsilon > 0$ to the resulting matrix.

5.5 Detecting the Transmit Weighting

For high performance it is desirable to use coherent ML detection at the receiver. To implement such a scheme, the transmit weighting must be

known at the receiver. Because b-bit quantization is assumed, the transmitter may have used any out of the 2^b different weights $\{\boldsymbol{W}^{(\boldsymbol{\delta}(j))}\}_{j=0}^{2^b-1}$. Knowing which one is not a problem as long as the feedback link is free from bit-errors, since then $j = i$ and the receiver can demodulate using $\boldsymbol{W}^{(\boldsymbol{\delta}(i))}$. However, in the presence of bit-errors, i does not necessarily equal j and hence the receiver can no longer be sure that $\boldsymbol{W}^{(\boldsymbol{\delta}(i))}$ was used in the transmission. To mitigate this problem, it is shown next how the receiver can detect the value of j, based on knowledge of i, the received signal and a training sequence.

In the derivation to follow, it is assumed that the transmitted signal can be divided into frames, where each frame contains a block of unknown OSTB codewords, corresponding to the data of interest, and possibly also a known training sequence. Recall from the system model that the channel is assumed known and constant for the duration of a frame. By grouping, similarly to as in (5.1), a block of received signal vectors $\boldsymbol{x}(n)$ into a matrix $\boldsymbol{X}_{\mathrm{t+d}}$, the received signal for one such frame can be written on the form

$$\boldsymbol{X}_{\mathrm{t+d}} = \boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\delta}(j))} \bar{\boldsymbol{C}}_{\mathrm{t+d}} + \boldsymbol{E}_{\mathrm{t+d}} \,,$$

where

$$\bar{\boldsymbol{C}}_{\mathrm{t+d}} = \begin{bmatrix} \bar{\boldsymbol{C}}_{\mathrm{t}} & \bar{\boldsymbol{C}}_{\mathrm{d}} \end{bmatrix}$$

consists of a training part $\bar{\boldsymbol{C}}_{\mathrm{t}}$ and a data part $\bar{\boldsymbol{C}}_{\mathrm{d}}$ of unknown codewords and where $\boldsymbol{X}_{\mathrm{t+d}}$ and $\boldsymbol{E}_{\mathrm{t+d}}$ are similarly defined. The lengths of the training and the data sequence are denoted L_{t} and L_{d}, respectively.

Based on knowledge of i and $\boldsymbol{X}_{\mathrm{t+d}}$, the feedback channel output j is detected possibly jointly with the data $\bar{\boldsymbol{C}}_{\mathrm{d}}$ according to the maximum a posteriori decision rule, i.e.,

$$\begin{aligned}
\{\hat{j}, \hat{\bar{\boldsymbol{C}}}_{\mathrm{d}}\} &= \arg \max_{\{j, \bar{\boldsymbol{C}}_{\mathrm{d}}\}} p(j, \bar{\boldsymbol{C}}_{\mathrm{d}} | i, \boldsymbol{X}_{\mathrm{t+d}}) \\
&= \arg \max_{\{j, \bar{\boldsymbol{C}}_{\mathrm{d}}\}} \frac{p(\boldsymbol{X}_{\mathrm{t+d}} | i, j, \bar{\boldsymbol{C}}_{\mathrm{d}}) p(j, \bar{\boldsymbol{C}}_{\mathrm{d}} | i)}{p(\boldsymbol{X}_{\mathrm{t+d}} | i)} \\
&= \arg \max_{\{j, \bar{\boldsymbol{C}}_{\mathrm{d}}\}} p(\boldsymbol{X}_{\mathrm{t+d}} | j, \bar{\boldsymbol{C}}_{\mathrm{d}}) p(j|i) p(\bar{\boldsymbol{C}}_{\mathrm{d}}) \\
&= \arg \max_{\{j, \bar{\boldsymbol{C}}_{\mathrm{d}}\}} p(\boldsymbol{X}_{\mathrm{t}} | j) p(\boldsymbol{X}_{\mathrm{d}} | j, \bar{\boldsymbol{C}}_{\mathrm{d}}) P_{j|i} \\
&= \arg \min_{\{j, \bar{\boldsymbol{C}}_{\mathrm{d}}\}} \big(\|\boldsymbol{X}_{\mathrm{t}} - \boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\delta}(j))} \bar{\boldsymbol{C}}_{\mathrm{t}}\|_{\mathrm{F}}^2 + \|\boldsymbol{X}_{\mathrm{d}} - \boldsymbol{H}^* \boldsymbol{W}^{(\boldsymbol{\delta}(j))} \bar{\boldsymbol{C}}_{\mathrm{d}}\|_{\mathrm{F}}^2 \\
&\qquad - \sigma^2 \log(P_{j|i}) \big) \,, \qquad (5.23)
\end{aligned}$$

where $p(\boldsymbol{u}|\boldsymbol{v})$ denotes the probability mass/density function of $\boldsymbol{u}$ conditioned on $\boldsymbol{v}$, $p(j, \bar{\boldsymbol{C}}_{\mathrm{d}}|i) = p(j|i, \bar{\boldsymbol{C}}_{\mathrm{d}})p(\bar{\boldsymbol{C}}_{\mathrm{d}}|i) = p(j|i)p(\bar{\boldsymbol{C}}_{\mathrm{d}})$ is due to the assumption that the output $\bar{\boldsymbol{C}}_{\mathrm{d}}$ of the OSTB encoder only depends on the data to be transmitted and is statistically independent of all other random quantities and where the fourth equality follows from discarding parameter independent factors, the whiteness of the noise and the uniform distribution of $\bar{\boldsymbol{C}}_{\mathrm{d}}$. The receiver's knowledge about i is obviously incorporated into the detection process through the last term, which guarantees that $\hat{j} = j = i$ is obtained in the absence of bit-errors.

It is evident from (5.23) that compared with when the transmit weighting is known, the computational complexity in the general case of joint detection increases a factor of 2^b since there are 2^b different transmit weights to be tried. Fortunately, the complexity is far from being as high as when performing an exhaustive search, since for each tentative j the detection of the OSTB code can be efficiently implemented by decoding the constituent information symbols separately as in (3.14).

Nevertheless, such joint detection may still be prohibitively expensive in some cases, making it worthwhile to consider the special case when $L_{\mathrm{d}} = 0$, i.e., the second term in (5.23) is omitted and the detection of the transmit weighting is based on the training part alone. Once the transmit weighting has been established, the data sequence can thereafter be detected as in (3.14), thus avoiding joint detection. Naturally, the requirements on the length of the training sequence might be tougher in this case, compared with when joint detection is used.

Note that the training sequence $\bar{\boldsymbol{C}}_{\mathrm{t}}$ is used for the sole purpose of detecting $\boldsymbol{W}$. This together with the previously mentioned assumption that the channel matrix $\boldsymbol{H}$ is known at the receiver typically means that $\boldsymbol{H}$ needs to be estimated from another training sequence which is not weighted by $\boldsymbol{W}$. In practice, such a training sequence may be transmitted before or after $\bar{\boldsymbol{C}}_{\mathrm{t}}$ or, as in the WCDMA system [3GP02b], in parallel on a so-called common pilot channel.

5.6 Numerical Examples

In order to illustrate the benefits of the proposed transmission schemes, simulations based on the simplified fading scenario described in Section 5.2.2 were performed. The focus is on a system with two transmit antennas and one receive antenna. Two bit quantization in conjunction with a type II feedback link was evaluated for all simulation cases to show

that weighted OSTBC can be made to work well in the common scenario of very coarsely quantized feedback information. Simulation results for the case of more bits and a type I feedback link are also presented. The performance of weighted OSTBC based on quantized feedback information was compared with three other methods – conventional OSTBC, ideal beamforming[3] and conventional beamforming based on circularly quantized feedback. In ideal beamforming it is assumed that the channel is known perfectly so the transmit weighting vector is equal to $\boldsymbol{h}$ (one receive antenna). Conventional beamforming with circularly quantized feedback is similar to standard ways of conveying quantized channel information to a beamformer [HP98, MSA01]. The phase of γ_2/γ_1 is uniformly quantized or, more precisely, the encoder output is given by $i = \varepsilon(\acute{\boldsymbol{\gamma}}) \triangleq \arg\min_k \|\acute{\boldsymbol{\gamma}} - \hat{\acute{\boldsymbol{\gamma}}}(k)\|^2$, with a two bit codebook

$$\hat{\acute{\boldsymbol{\gamma}}}(0) \triangleq \begin{bmatrix} 1 & 1 \end{bmatrix}^{\mathrm{T}}, \qquad \hat{\acute{\boldsymbol{\gamma}}}(1) \triangleq \begin{bmatrix} 1 & \mathrm{j} \end{bmatrix}^{\mathrm{T}}$$
$$\hat{\acute{\boldsymbol{\gamma}}}(2) \triangleq \begin{bmatrix} 1 & -\mathrm{j} \end{bmatrix}^{\mathrm{T}}, \qquad \hat{\acute{\boldsymbol{\gamma}}}(3) \triangleq \begin{bmatrix} 1 & -1 \end{bmatrix}^{\mathrm{T}},$$

where j denotes the imaginary unit. As seen, care has been taken to ensure that the impact of feedback channel bit-errors is reduced by ordering the codebook so that codebook vectors far from each other (e.g. $\hat{\acute{\boldsymbol{\gamma}}}(0)$ and $\hat{\acute{\boldsymbol{\gamma}}}(3)$) differ in two bits ($0 = 00_{\text{binary}}$ versus $3 = 11_{\text{binary}}$) while neighboring vectors differ in only one bit. This serves to minimize the likelihood of large errors, such as when $\hat{\acute{\boldsymbol{\gamma}}}(0)$ is confused with $\hat{\acute{\boldsymbol{\gamma}}}(3)$. Based on the feedback channel output j, the transmit weighting vector is now given by $\hat{\acute{\boldsymbol{\gamma}}}(j)$.

The assumptions in the simulations were as follows. For all the examined cases, the simplified fading scenario was used assuming perfect knowledge of σ^2, σ_h^2, ρ and P_{b} at both the transmitter and receiver. The variance of the channel coefficients and the average energy per information bit were arbitrarily set at $\sigma_h^2 = 1$ and $P = 1$, respectively. The channel was constant during the transmission of a frame of codewords and independently fading from one frame to another. Each frame contained $L_{\mathrm{t}} = 2$ training samples and $L_{\mathrm{d}} = 10$ data samples, as described in Section 5.5. The Alamouti OSTB code [Ala98] was used in the conventional OSTBC system as well as in weighted OSTBC. Data to be transmitted consisted of a sequence of IID information bits that was mapped into a corresponding sequence of data symbols, prior to being input to

[3]Similarly to as in Chapter 4, beamforming is taken to be a system in which the transmitted signal can be written on the form $\boldsymbol{c}(n) = \boldsymbol{v}s(n)$, where $s(n)$ represents the nth data symbol and $\boldsymbol{v}$ is a transmit weighting vector.

the OSTB encoder. Each data symbol was taken from a gray coded, four quadrature phase shift keying (4-QPSK), signal constellation. The type II feedback links were designed using a diagonal weighting matrix $\boldsymbol{K} = \mathrm{diag}(0.2, 0.8)$. This choice was found empirically and emphasizes the importance of primarily conveying phase information. Throughout the simulations, BER was used as the performance measure. The SNR was measured for the system using conventional OSTBC and defined as

$$\mathrm{SNR} \triangleq \frac{\mathrm{E}[\|\boldsymbol{H}^*\boldsymbol{C}\|_{\mathrm{F}}^2]}{LN\sigma^2},$$

where $\boldsymbol{C} = \sqrt{P_{\mathrm{o}}/M}\bar{\boldsymbol{C}}$ represents the transmitted signals. The expression for the SNR is equal to the total received average signal energy, divided by the total average noise energy.

Varying the SNR

In Figure 5.5, the BER as a function of the SNR is plotted. Only quantization errors impair the quality of the channel information. Hence, $\rho \to 1$ and $P_{\mathrm{b}} = 0$, where the latter implies that the receiver automatically knows which transmit weighting that was used.

Using two bit quantization and a type II feedback link, the performance of weighted OSTBC is seen to be similar to the performance of conventional beamforming. As the number of bits is increased, the performance quickly approaches that of ideal beamforming. On the other hand, the performance of weighted OSTBC based on a type I feedback link is significantly poorer for the two bit case. There are simply not enough bits to capture any useful channel information. Because weighted OSTBC takes the quality of the channel information into account, the performance is however not worse than conventional OSTBC. As expected and illustrated by the eight bit curves, the difference between a type I and type II feedback link becomes smaller as the number of bits increases. It is also seen that weighted OSTBC is essentially robust against quantization errors in the sense that the performance is better than conventional OSTBC (with the minor exception of type I eight bit weighted OSTBC at high SNR values). Conventional beamforming with circular quantization is clearly better than a type I system at handling quantization errors.

Robustness against errors in the initial channel information is demonstrated in Figure 5.6. The assumption of $\rho = 0.1$ means that the quality of the channel knowledge is poor, even though $P_{\mathrm{b}} = 0$. The curves corresponding to weighted OSTBC are seen to closely follow the conventional

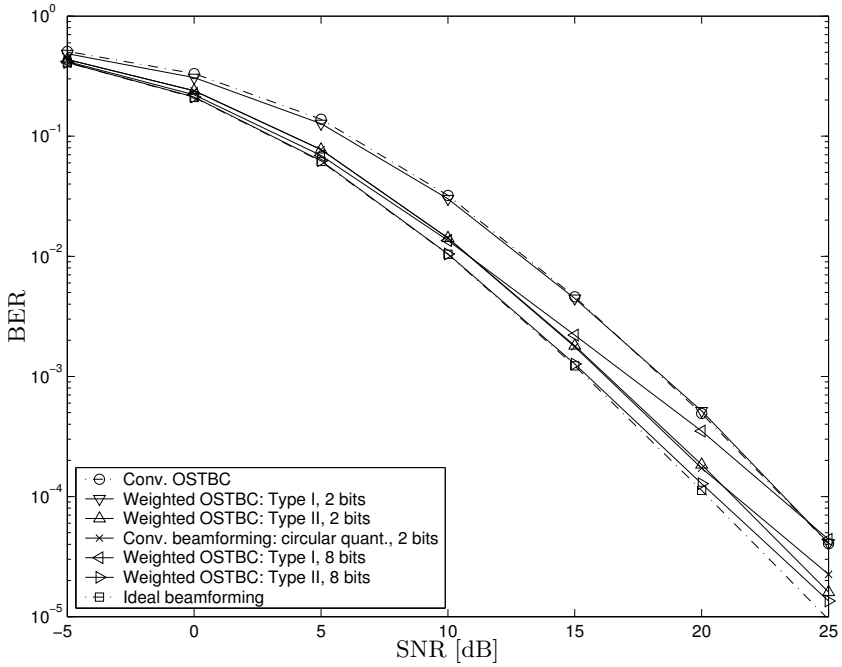

Figure 5.5: Demonstrates how quantization errors affect the performance. Simplified fading scenario with $\rho \to 1$, $M = 2$ transmit antennas, $N = 1$ receive antenna and $P_\mathrm{b} = 0$.

OSTBC curve while conventional beamforming is not robust since the performance rapidly deteriorates as the SNR is increased.

Next, the behavior of the proposed transmission scheme in the presence of feedback channel bit-errors is illustrated in Figures 5.7 and 5.8. The feedback channel BER is set to $P_\mathrm{b} = 0.4$ and results for both when the receiver knows and does not know the weighting matrix $\boldsymbol{W}$ are plotted. For the latter case, the weighting matrix was detected according to (5.23). The same detection method was also used for the beamforming case but now with $\bar{\boldsymbol{C}}_\mathrm{d}$ representing a symbol sequence instead of space-time codewords. Again, weighted OSTBC is robust while conventional beamforming is unable to handle the situation.

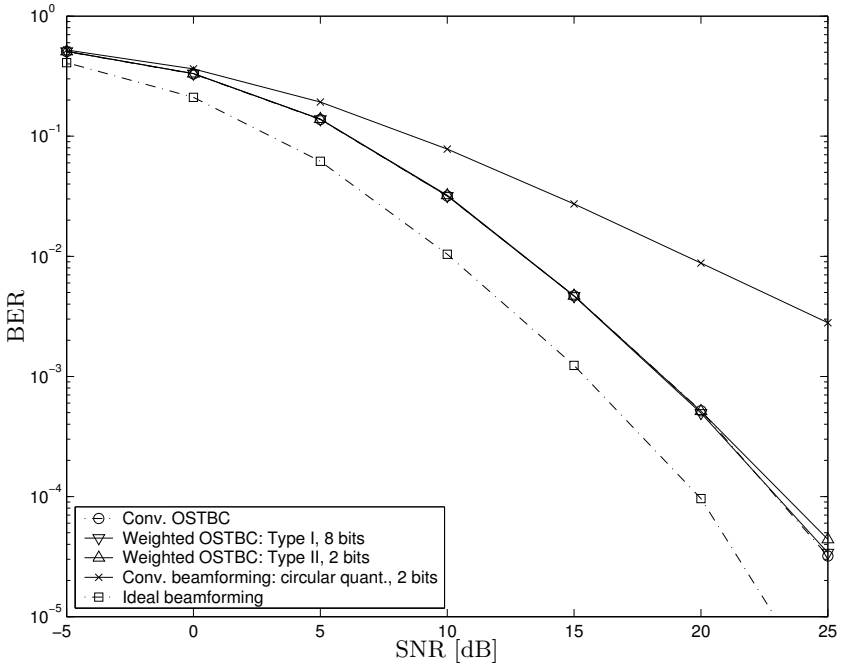

Figure 5.6: Impact of non-perfect initial channel information. Simplified fading scenario with $\rho = 0.1$, $M = 2$ transmit antennas, $N = 1$ receive antenna and $P_\mathrm{b} = 0$.

Varying the BER of the Feedback Channel

To further illustrate the robustness against bit-errors introduced by the feedback channel, the corresponding BER P_b was varied while the SNR was held fixed at 15 dB. Results corresponding to when the receiver knows as well as does not know the weighting matrix are presented in Figures 5.9 and 5.10, respectively.

As expected, conventional beamforming is seen to perform poorly when P_b is large. Weighted OSTBC for known $\boldsymbol{W}$, on the other hand, approaches conventional OSTBC in performance as P_b is increased. Only a modest deterioration is incurred for the type II system when the receiver has to detect the transmit weighting matrix. The type I system suffers more because of higher failure rate in the detection process due to the large number of possible transmit weights to choose from. From these

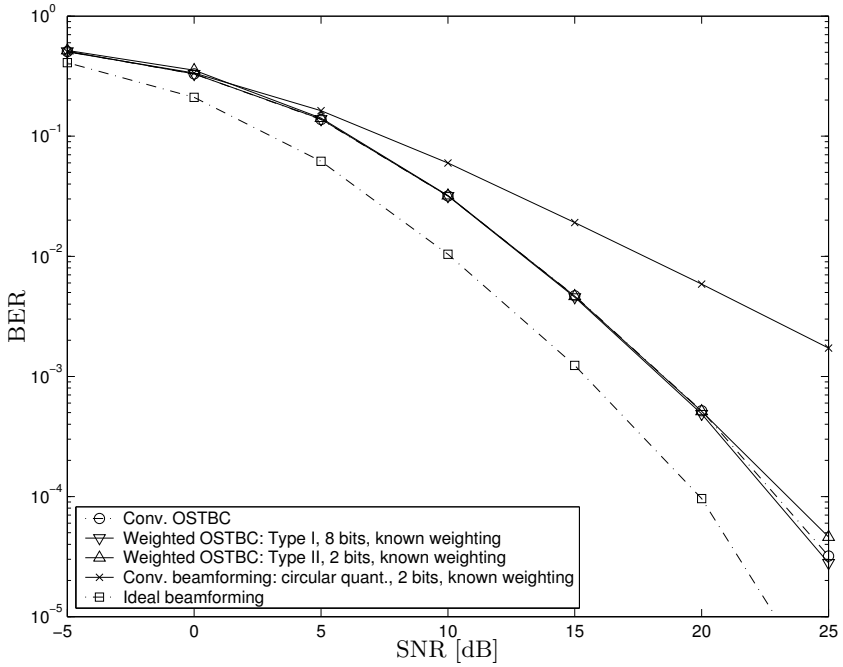

Figure 5.7: Performance in presence of feedback channel bit-errors when receiver knows the transmit weighting. Simplified fading scenario with $\rho \to 1$, $M = 2$ transmit antennas, $N = 1$ receive antenna and $P_\mathrm{b} = 0.4$.

figures, we conclude that the transmission scheme proposed herein is robust against errors due to the feedback channel. However, as mentioned in Section 5.2.1, this comes at the price of estimating and distributing certain necessary parameters.

5.7 Conclusions

This chapter investigated the use of weighted OSTBC in conjunction with quantized channel side information obtained from a feedback link. Techniques from the field of vector quantization for noisy channels were used for designing two different types of feedback links, with the goal of providing the transmitter with useful information about the current channel realization. The channel information, including an accompanying

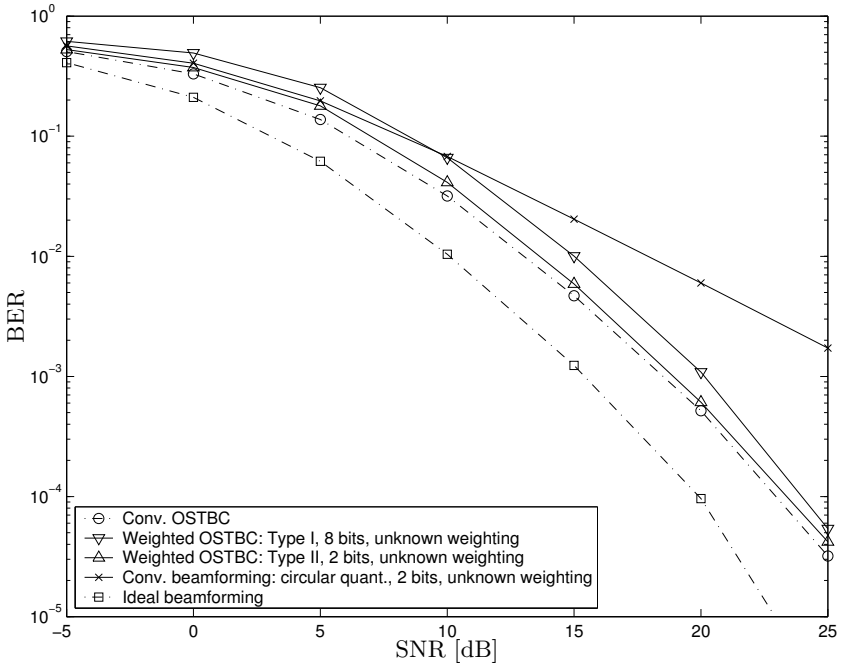

Figure 5.8: Performance in presence of feedback channel bit-errors when receiver has to detect the transmit weighting. Simplified fading scenario with $\rho \to 1$, $M = 2$ transmit antennas, $N = 1$ receive antenna and $P_\mathrm{b} = 0.4$.

reliability measure, was then used for determining a suitable codebook of transmit weighting matrices. The resulting transmission scheme provides a seamless combination of conventional beamforming and OSTB coding, robust against errors in the channel side information due to feedback delay, quantization as well as feedback channel bit-errors. Numerical examples provided support of this claim.

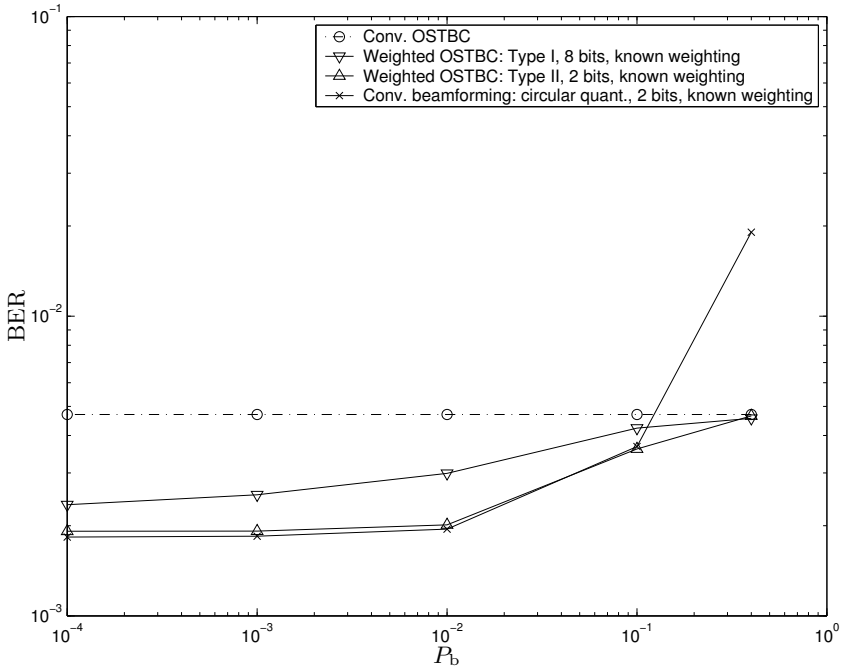

Figure 5.9: Performance in the presence of feedback channel bit-errors when the receiver knows the transmit weighting. Simplified fading scenario with $\rho \to 1$, $M = 2$ transmit antennas, $N = 1$ receive antenna and SNR = 15 dB.

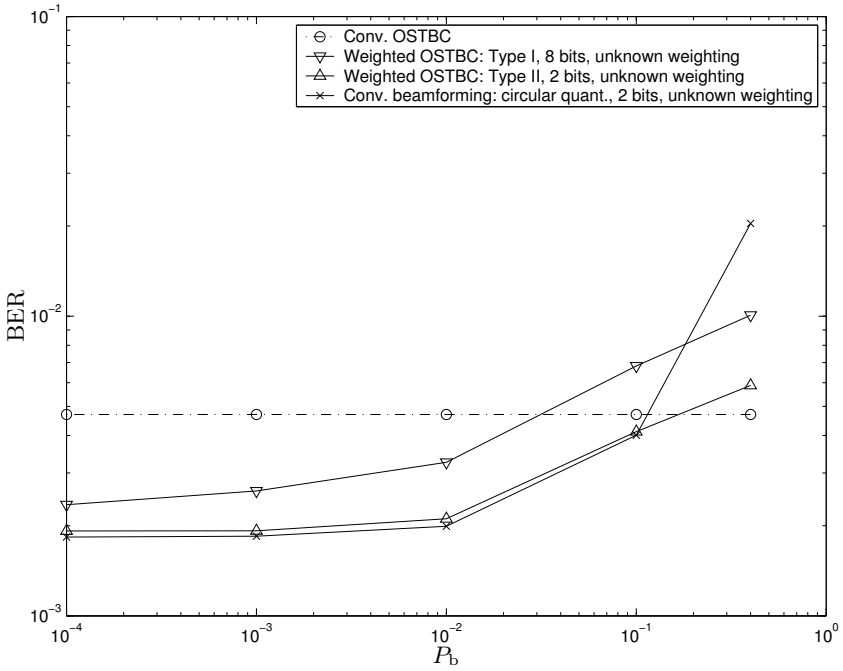

Figure 5.10: Performance in the presence of feedback channel bit-errors when the receiver has to detect the transmit weighting. Simplified fading scenario with $\rho \to 1$, $M = 2$ transmit antennas, $N = 1$ receive antenna and SNR = 15 dB.

Appendix 5.A Conditional Covariance for Feedback Link Type I

The conditional covariance $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\varsigma}}$ for feedback link type I is given by

$$\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\varsigma}} = \mathrm{E}[\boldsymbol{h}\boldsymbol{h}^*|j] - \mathrm{E}[\boldsymbol{h}|j]\,\mathrm{E}[\boldsymbol{h}|j]^* \,. \tag{5.24}$$

Recall that $\boldsymbol{h}|\boldsymbol{\gamma}$ is independent of j. The first term in (5.24) is rewritten as

$$\begin{aligned}
\mathrm{E}[\boldsymbol{h}\boldsymbol{h}^*|j] = \mathrm{E}\left[\,\mathrm{E}[\boldsymbol{h}\boldsymbol{h}^*|\boldsymbol{\gamma},j]\big|j\right] &= \mathrm{E}[\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} + \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*|j] \\
&= \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} \\
&\quad + \mathrm{E}\left[\big(\boldsymbol{m}_{\boldsymbol{h}} + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})\big)\big(\boldsymbol{m}_{\boldsymbol{h}} + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})\big)^*\big|j\right] \\
&= \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} + \boldsymbol{m}_{\boldsymbol{h}}\boldsymbol{m}_{\boldsymbol{h}}^* + \boldsymbol{m}_{\boldsymbol{h}}\,\mathrm{E}[\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}|j]^*\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^* \\
&\quad + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\,\mathrm{E}[\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}|j]\boldsymbol{m}_{\boldsymbol{h}}^* + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\,\mathrm{E}[(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})^*|j]\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^* \,,
\end{aligned}$$

where the third equality follows from the fact that $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} = \boldsymbol{R}_{\boldsymbol{hh}} - \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^*$ is independent of $\boldsymbol{\gamma}$ and hence of j. The second term in (5.24) is expressed as

$$\begin{aligned}
\mathrm{E}[\boldsymbol{h}|j]\,\mathrm{E}[\boldsymbol{h}|j]^* &= \mathrm{E}\left[\,\mathrm{E}[\boldsymbol{h}|\boldsymbol{\gamma},j]\big|j\right]\mathrm{E}\left[\,\mathrm{E}[\boldsymbol{h}|\boldsymbol{\gamma},j]\big|j\right]^* \\
&= \mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}} + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})|j]\,\mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}} + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})|j]^* \\
&= \boldsymbol{m}_{\boldsymbol{h}}\boldsymbol{m}_{\boldsymbol{h}}^* + \boldsymbol{m}_{\boldsymbol{h}}\,\mathrm{E}[\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}|j]^*\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^* + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\,\mathrm{E}[\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}|j]\boldsymbol{m}_{\boldsymbol{h}}^* \\
&\quad + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\,\mathrm{E}[\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}|j]\,\mathrm{E}[\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}|j]^*\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^* \,.
\end{aligned}$$

Hence, combining the first and second term, gives the conditional covariance

$$\begin{aligned}
\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\varsigma}} &= \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} \\
&\quad + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\big(\,\mathrm{E}[(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}})^*|j] - \mathrm{E}[\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}|j]\,\mathrm{E}[\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}|j]^*\big)\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^* \\
&= \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} \\
&\quad + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\,\mathrm{E}\left[(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}-\mathrm{E}[\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}|j])(\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}-\mathrm{E}[\boldsymbol{\gamma}-\boldsymbol{m}_{\boldsymbol{\gamma}}|j])^*\big|j\right]\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^* \\
&= \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} \\
&\quad + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\,\mathrm{E}\left[(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|j])(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|j])^*\big|j\right]\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^* \\
&= \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}(\mathrm{E}[\boldsymbol{\gamma}\boldsymbol{\gamma}^*|j] - \mathrm{E}[\boldsymbol{\gamma}|j]\,\mathrm{E}[\boldsymbol{\gamma}|j]^*)\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^* \,.
\end{aligned} \tag{5.25}$$

5.A.1 Decomposing the Conditional Covariance into Three Terms

Consider

$$\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}} = \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} + \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\,\mathrm{E}\left[(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|j])(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|j])^{*}\big|j\right]\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^{*}\,.$$

in (5.25). Expanding $\mathrm{E}\left[(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|j])(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|j])^{*}\big|j\right]$ by utilizing that $\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|j] = \boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|i]-(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])$ gives

$$\begin{aligned}
&\mathrm{E}\left[(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|j])(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|j])^{*}\big|j\right]\\
&=\mathrm{E}\left[(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|i])(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|i])^{*}\big|j\right]+\mathrm{E}\left[(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])^{*}\big|j\right]\\
&-\mathrm{E}\left[(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|i])^{*}\big|j\right]-\mathrm{E}\left[(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|i])(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])^{*}\big|j\right].
\end{aligned}$$

It is clear that the two last terms are zero since

$$\begin{aligned}
&\mathrm{E}\left[(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|i])^{*}\big|j\right]\\
&\qquad=\mathrm{E}\left[\mathrm{E}\left[(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|i])^{*}\big|i,j\right]\Big|j\right]\\
&\qquad\qquad=\mathrm{E}\left[(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])(\mathrm{E}[\boldsymbol{\gamma}|i]-\mathrm{E}[\boldsymbol{\gamma}|i])^{*}\big|j\right]=\boldsymbol{0}\,.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathrm{E}\left[(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|j])(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|j])^{*}\big|j\right] &= \mathrm{E}\left[(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|i])(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|i])^{*}\big|j\right]\\
&\quad+\mathrm{E}\left[(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])^{*}\big|j\right]
\end{aligned}$$

and thus,

$$\begin{aligned}
\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}} = \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} &+ \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\,\mathrm{E}\left[(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|i])(\boldsymbol{\gamma}-\mathrm{E}[\boldsymbol{\gamma}|i])^{*}\big|j\right]\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^{*}\\
&+ \boldsymbol{R}_{\boldsymbol{h\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\,\mathrm{E}\left[(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])(\mathrm{E}[\boldsymbol{\gamma}|j]-\mathrm{E}[\boldsymbol{\gamma}|i])^{*}\big|j\right]\boldsymbol{R}_{\boldsymbol{\gamma\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h\gamma}}^{*}\,.
\end{aligned}$$

Chapter 6

Quantized Channel Feedback: Design Approach II

Also this chapter considers a wireless MIMO communication system in which the transmitter has access to quantized channel side information obtained from a feedback link. The setup is thus similar to the scenario considered in the previous chapter. This time however, a new space-time code performance criterion is derived that is specifically tailored to quantized channel side information. The performance criterion is similar in structure to the one used earlier and can hence be utilized for constructing any of the three code types described in the thesis. Still, the focus in the present chapter is mostly on unstructured codes.

Because the effects of quantization are taken into account already at the outset when deriving the performance criterion, heuristic modifications of the code design procedure in order to obtain good performance are not needed. This is in contrast to the methods in Chapter 5. Compared with that chapter, the development to follow thus constitutes a mathematically cleaner approach for incorporating quantized channel side information into the design of space-time codes.

Efficient unstructured codes are found by utilizing a gradient search technique for minimizing the performance criterion. Properties of optimal unstructured codes are derived for a number of different cases. Design procedures for linear dispersive codes and weighted OSTBC are

briefly described and implemented. Unstructured codes designed using the techniques developed in this work are shown to perform better than comparable OSTB codes, even without channel knowledge at the transmitter. Experimental investigations reveal that the design procedure for linear dispersive codes appears to automatically produce OSTB codes in the case of no channel knowledge.

Simulation results for a spatially uncorrelated flat Rayleigh fading scenario using two and four transmit antennas and one receive antenna demonstrate that the designed unstructured codes give significant gains relative conventional transmission schemes. In addition, the impact on performance due to the choice of code type is illustrated by means of one representative simulation example. The unstructured code is seen to perform significantly better than both weighted OSTBC and the linear dispersive code. It is also noted that the two latter have similar performance.

6.1 Introduction

In general, conventional space-time codes [GFBK99, TSC98, TJC99] have been designed for situations where the transmitter has no knowledge about the wireless MIMO/MISO channel. The present work, on the other hand, considers the design of space-time block codes that may utilize channel information at the transmitter, if available, for increasing the performance. Just as in Chapter 5, the design is targeted towards a non-power-controlled narrowband scenario in which the transmitter is provided with quantized channel information from the receiver via a dedicated feedback link. While the receiver has perfect channel knowledge, both quantization errors as well as feedback delay are assumed to plague the channel information at the transmitter. A model describing these detrimental effects is exploited in the code design. In this way, the constructed codes continue to work well even if the quality of the channel information is poor.

The channel knowledge is incorporated into the code construction process by utilizing a new design criterion that is developed specifically to deal with scenarios in which the channel side information is quantized. The design criterion is similar in structure to the one originally presented in Chapter 4, making it straightforward to modify the code design procedures described therein so that they utilize the new design criterion. Most of the work is focused on the design and analysis of unstructured

space-time block codes. However some effort is also spent on the design of linear dispersive codes and weighted OSTBC.

Related work includes the design of conventional space-time codes [GFBK99, TSC98, BBH00, ARU01, MBV02, HH02b], the design of modulation signal sets [FGW74, HSN91] and how to utilize non-perfect channel information at the transmitter [Wit95, NLTW98, HP98], respectively.

In [GFBK99, TSC98], code design criteria were introduced and used for manually constructing conventional space-time codes exhibiting a high degree of structure. The design criterion derived and used in this chapter constitutes a generalization of these criteria so as to handle also the case of (non-perfect) channel knowledge. A more automated design process was investigated in [BBH00], where an exhaustive search technique was employed in order to find appropriate space-time trellis codes. In [ARU01, MBV02] the code search was conducted over the set of unitary codeword matrices to produce codes suitable for a non-coherent detection scenario in which the channel is unknown at the receiver. Linear dispersive codes were designed in [HH02b] based on maximizing an information theoretic capacity measure. The design of efficient signal constellations and modulation signal sets for a SISO channel was treated in [FGW74] and [HSN91], respectively. Inspired by the numerical search approach in the two latter papers, the present work applies gradient search techniques for minimizing the design criterion and hence obtaining suitable channel side information dependent space-time block codes.

The present work contributes in several ways. A major contribution is the development of a new design criterion that specifically handles quantized channel side information. The design criterion lays the foundation of a general strategy for incorporating non-perfect quantized channel side information into the design of space-time codes. Another important contribution is the development of design procedures for realizing this strategy in the cases of unstructured and linear dispersive codes as well as for weighted OSTBC. The focus is however on unstructured codes for which a theoretical analysis is provided that illustrates the intuitively pleasing characteristics of the resulting transmission scheme and gives useful insights into the problem. Due to the unstructured nature of the codewords, these codes potentially have better performance than all other corresponding block codes with structure, albeit at the cost of a higher decoding complexity. Hence an additional field of application for our unstructured code design method is to produce benchmark codes that quantify the loss in performance that results from introducing structure to reduce decoding complexity. This is illustrated herein by simulation

examples and coding gain measurements showing that the unstructured codes perform better than comparable OSTB codes. Such gains are obtained even when no channel information is available. A particularly fascinating result is that the design procedure for linear dispersive codes seems to, in the case of no channel knowledge, automatically produce OSTB codes, if such codes exist for the setup under consideration. Finally, it should be noted that the design methods presented herein are easily extended to also cope with errors in the channel side information due to bit-errors introduced by the feedback channel.

The remaining part of the chapter is organized as follows. In Section 6.2, the generic system model is briefly reviewed and additional assumptions and preliminaries are described. The design criterion is derived in Section 6.3 and utilized for defining a design procedure for unstructured codes in terms of an optimization problem. The design of linear dispersive codes and weighed OSTBC is also addressed. Various properties of optimal unstructured codes are derived in Section 6.4. The numerical technique used for solving the optimization problem is described in Section 6.5. Section 6.6 gives explicit examples of designed codes of the unstructured type and discusses properties of several other codes, including a couple of constructed linear dispersive codes. Finally, the performance of some constructed codes is assessed by means of simulation of a spatially uncorrelated flat Rayleigh fading scenario and the results are presented in Section 6.7.

6.2 System Model

This work considers a MIMO wireless communication system making use of space-time block codes that utilize quantized channel side information at the transmitter for enhancing the performance. The channel side information is obtained from the receiver via a dedicated feedback link. The generic system model and the assumptions regarding code structures described in Chapter 3 apply, as indicated by the setup illustrated in Figure 6.1.

The channel side information $\boldsymbol{\zeta}$ now comes in the form of an integer i representing the quantized output of the feedback link. Hence, the channel code $\mathcal{C} = \mathcal{C}(i)$ currently in use is determined out of a set $\{\mathcal{C}(i)\}$ of space-time block codes. Codes of the types unstructured, linear dispersive or weighted OSTBC may be used. These code types are discussed in detail in Section 3.2. The focus in this work is however primarily on

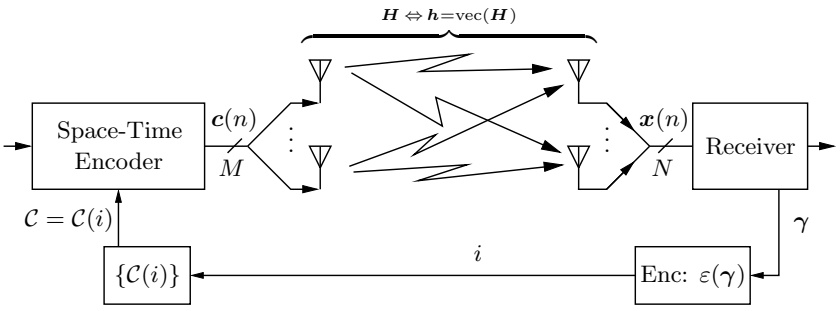

Figure 6.1: System model.

unstructured codes. Hence, the codes in $\{\mathcal{C}(i)\}$ are henceforth assumed to be unstructured, unless explicitly stated otherwise. Each code in $\{\mathcal{C}(i)\}$ will be designed so as to adapt the code to the specific channel knowledge, as described by the corresponding value of i. The resulting information carrying signals are transmitted over a wireless frequency-nonselective fading channel, picked up by the antennas at the receiver and thereafter decoded to produce an estimate of the transmitted message.

Recall from the generic system model that there are M transmit and N receive antennas and that the received signal vectors $\boldsymbol{x}(n) = \boldsymbol{H}^*\boldsymbol{c}(n) + \boldsymbol{e}(n)$ during the time interval $n = 0, 1, \cdots, L-1$ are grouped into a matrix $\boldsymbol{X}$ given by

$$\boldsymbol{X} = \boldsymbol{H}^*\boldsymbol{C} + \boldsymbol{E}\,, \tag{6.1}$$

where the zero-mean IID complex Gaussian elements of the noise matrix $\boldsymbol{E}$ have variance σ^2. Here, the MIMO channel is described by the $M \times N$ matrix $\boldsymbol{H}$ or its vectorized counterpart $\boldsymbol{h} \triangleq \mathrm{vec}(\boldsymbol{H})$ with elements $\{h_k\}$. The transmitted codeword $\boldsymbol{C}$ belongs to the current choice $\mathcal{C} \triangleq \{\boldsymbol{C}_k\}_{k=1}^{K}$ of channel code. The rate of the code is $\log_2(K)/L$ information bits per channel use. The receiver is assumed to know the channel perfectly and recovers the transmitted codeword by means of ML decoding according to

$$\hat{\boldsymbol{C}} = \arg \min_{\boldsymbol{C} \in \mathcal{C}(i)} \|\boldsymbol{X} - \boldsymbol{H}^*\boldsymbol{C}\|_{\mathrm{F}}^2\,, \tag{6.2}$$

where $\hat{\boldsymbol{C}}$ represents the estimated codeword.

The output power is limited in the same manner as previously. Hence, conditioned on the use of a certain code $\mathcal{C}(i)$, it is assumed that the transmission is such that all the codewords are equally probable and that

the average energy per information bit is equal to P. It was shown in (3.3) that this implies the power constraint

$$K\,\mathrm{E}[\|\boldsymbol{C}\|_{\mathrm{F}}^2|i] = \sum_{k=1}^{K} \|\boldsymbol{C}_k^{(i)}\|_{\mathrm{F}}^2 = PK\log_2(K)\,. \tag{6.3}$$

As is clear from (6.3), the focus in this work is on a scenario without power control, since the average output power is constant regardless of which code in the set $\{\mathcal{C}(i)\}$ that is chosen. However, power control may be performed on top of our transmission scheme using techniques similar to as in e.g. [SG00]. It is even possible to include power control directly into the code construction process by taking the average in (6.3) also over all the codes in $\{\mathcal{C}(i)\}$ and modifying our design method appropriately. These modifications are straightforward but complicate the resulting optimization problem.

6.2.1 The Feedback Link

As seen in Figure 6.1, the feedback link consists of an encoder, an ideal feedback channel, which does not introduce any bit-errors, and a decoder block $\{\mathcal{C}(i)\}$. At the encoder, the $MN \times 1$ vector $\boldsymbol{\gamma}$, with elements $\{\gamma_k\}$, represents the *initial channel information*. The initial channel information is quantized into a b bit integer $i = \varepsilon(\boldsymbol{\gamma})$, which is transported over the feedback channel to the transmitter. The encoder function $\varepsilon(\boldsymbol{\gamma})$ is such that it partitions the set of all possible $\boldsymbol{\gamma}$ into 2^b different encoder regions $\{\mathcal{S}_i\}_{i=0}^{2^b-1}$ based on the mapping $\boldsymbol{\gamma} \in \mathcal{S}_i \Rightarrow \varepsilon(\boldsymbol{\gamma}) = i$.

It will be apparent from the development in later sections that the design method and theoretical analysis presented herein are applicable to any mapping $\varepsilon(\boldsymbol{\gamma})$, unless explicitly stated otherwise. However, for the code construction and simulation examples the focus is on a specific quantization scheme based on the encoder function

$$\varepsilon(\boldsymbol{\gamma}) = \arg \min_{i\in\{0,1,\cdots,2^b-1\}} \|\boldsymbol{\gamma}/\gamma_1 - \hat{\boldsymbol{\gamma}}(i)\|^2\,, \tag{6.4}$$

with the codebook vectors $\{\hat{\boldsymbol{\gamma}}(i)\}_{i=0}^{2^b-1}$ given by

$$\hat{\boldsymbol{\gamma}}(i) \triangleq \begin{bmatrix} 1 & e^{j\phi_{i_0(i)}} & \cdots & e^{j\phi_{i_{MN-2}(i)}} \end{bmatrix}^{\mathrm{T}}\,, \tag{6.5}$$

where $j \triangleq \sqrt{-1}$ and $\phi_{i_k(i)} \triangleq 2\pi i_k(i)/2^{\bar{b}}$. Here, $\bar{b} \triangleq b/(MN-1)$ represents the number of feedback bits per complex-valued dimension

and $i_k(i) \in \{0, 1, \cdots, 2^{\bar{b}} - 1\}$ is implicitly defined through the relation $i = \sum_{k=0}^{MN-2} i_k(i) 2^{\bar{b}k}$.

The latter relation means that $i_0(i)$ represents the decimal number corresponding to the $\bar{b}$ least significant bits of a natural binary representation of i. Similarly, $i_1(i)$ corresponds to the next $\bar{b}$ bits and so on up to $i_{M-2}(i)$ which corresponds to the $\bar{b}$ most significant bits. Hence, it is seen that (6.4) and (6.5) together implement $\bar{b}$ bit uniform scalar quantization of the phases of γ_k/γ_1, $k = 2, \cdots, MN$. This is similar to the so-called partial phase combining scheme in [HP98] and is closely related to the feedback scheme in the closed-loop mode of the WCDMA system [3GP02b]. The same quantization scheme was also considered in connection with the information theoretic investigations in Section 2.6.3.

On the transmitter side, the decoder chooses the channel code as $\mathcal{C} = \mathcal{C}(i)$ from a set of 2^b different codes $\mathcal{C}(i) \triangleq \{\boldsymbol{C}_k^{(i)}\}_{k=1}^K$. The codes are designed to take the available channel knowledge into account using techniques described in later sections. These techniques can be extended to handle also joint design of codes $\{\mathcal{C}(i)\}$ and feedback encoder $\varepsilon(\boldsymbol{\gamma})$, as outlined in our work in [JSO02b].

6.2.2 Fading Statistics

The non-perfect nature of the channel information is not only due to quantization errors. The channel information may also suffer from feedback delay. This, in conjunction with the common situation of a time-varying channel, means that the channel information might be outdated by the time it reaches the transmitter. Such a phenomenon is here modeled by assuming that $\boldsymbol{\gamma}$ is a delayed copy of the current channel realization, correlated, to an arbitrary degree, with $\boldsymbol{h}$. More precisely, $\boldsymbol{h}$ and $\boldsymbol{\gamma}$ are jointly complex Gaussian with mean vectors

$$\boldsymbol{m}_{\boldsymbol{h}} \triangleq \mathrm{E}[\boldsymbol{h}], \quad \boldsymbol{m}_{\boldsymbol{\gamma}} \triangleq \mathrm{E}[\boldsymbol{\gamma}]$$

and cross- and covariance matrices

$$\begin{aligned}
\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}} &\triangleq \mathrm{E}[(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}})(\boldsymbol{\gamma} - \boldsymbol{m}_{\boldsymbol{\gamma}})^*] \\
\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}} &\triangleq \mathrm{E}[(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}})(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}})^*] \\
\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} &\triangleq \mathrm{E}[(\boldsymbol{\gamma} - \boldsymbol{m}_{\boldsymbol{\gamma}})(\boldsymbol{\gamma} - \boldsymbol{m}_{\boldsymbol{\gamma}})^*] \,.
\end{aligned}$$

All cross- and covariance matrices are assumed to be constant and invertible. It follows that the conditional mean and covariance of $\boldsymbol{h}$ are given

by [Kay93, p. 509]

$$\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}} \triangleq \mathrm{E}[\boldsymbol{h}|\boldsymbol{\gamma}] = \boldsymbol{m}_{\boldsymbol{h}} + \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}(\boldsymbol{\gamma} - \boldsymbol{m}_{\boldsymbol{\gamma}}) \tag{6.6}$$

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}} &\triangleq \mathrm{E}[(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}})(\boldsymbol{h} - \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}})^*|\boldsymbol{\gamma}] \\ &= \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}} - \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}^*\,. \end{aligned} \tag{6.7}$$

Note that $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}$ is the MMSE estimate of $\boldsymbol{h}$ based on $\boldsymbol{\gamma}$, with $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}}$ being the corresponding error covariance matrix.

As explained in Section 4.2.1, this way of modeling feedback delay in terms of correlation is reasonable in view of the well-known Jakes fading model [Jak94], which may be used to describe the time-variations of the channel as a temporally correlated stationary complex Gaussian process. The degree of correlation determines the quality of the initial channel information. Loosely speaking, the quality of the initial channel information improves as the correlation between $\boldsymbol{h}$ and $\boldsymbol{\gamma}$ grows larger, and conversely, deteriorates as the correlation becomes smaller. Likewise, the MMSE estimate error covariance $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}}$ is large (measured in some matrix norm) for low correlation and small for high. Guided by this and the fact that (6.7) can be rewritten as

$$\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}} = \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{1/2}\big(\boldsymbol{I}_{MN} - \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{-1/2}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}^*\big(\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{-1/2}\big)^*\big)\big(\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{1/2}\big)^*\,, \tag{6.8}$$

where $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{1/2}\big(\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{1/2}\big)^* = \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}$, we let

$$\begin{aligned}\rho &\triangleq \sqrt{\frac{\mathrm{tr}\left(\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{-1/2}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}^*\big(\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{-1/2}\big)^*\right)}{MN}} \\ &= \sqrt{\frac{\mathrm{tr}(\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}}^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\boldsymbol{R}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}^{-1}\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}^*)}{MN}} \end{aligned} \tag{6.9}$$

define a measure of the initial channel information quality. It is easily verified that $0 \leq \rho \leq 1$, with $\rho = 0$ if (and only if) there is no correlation and $\rho = 1$ corresponding to "full correlation". Thus, it makes sense to define the two extremes of initial channel knowledge as

- No initial channel knowledge: $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}} = \boldsymbol{0}$ ($\Leftrightarrow \rho = 0$)
- Perfect initial channel knowledge: $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}} \to \bar{\boldsymbol{R}}_{\boldsymbol{h}\boldsymbol{\gamma}}$ with $\bar{\boldsymbol{R}}_{\boldsymbol{h}\boldsymbol{\gamma}}$ such that $\rho \to 1$

Note that from (6.8) it follows that the perfect initial channel knowledge case also corresponds to $\|\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}}\| \to 0$.

A Simplified Fading Scenario

For illustrative purposes, the described fading model will occasionally be specialized into what is here called a simplified fading scenario, just as in previous chapters. In this scenario, the channel coefficients $\{h_k\}$ are assumed to be zero-mean IID with variance σ_h^2. The same goes for the initial channel information coefficients $\{\gamma_k\}$. Each γ_k is correlated with the corresponding h_k and uncorrelated with all others. It is easily verified that (6.9) then reduces to $\rho = |\tilde{\rho}|$, where $\tilde{\rho} \triangleq \mathrm{E}[h_k \gamma_k^*]/\sigma_h^2$ represents the normalized correlation coefficient. Thus the model is summarized by

$$\begin{aligned} &\boldsymbol{m_h} = \mathbf{0}, \quad \boldsymbol{m_\gamma} = \mathbf{0} \\ &\boldsymbol{R_{hh}} = \sigma_h^2 \boldsymbol{I}_{MN}, \quad \boldsymbol{R_{h\gamma}} = \sigma_h^2 \tilde{\rho} \boldsymbol{I}_{MN}, \quad \boldsymbol{R_{\gamma\gamma}} = \sigma_h^2 \boldsymbol{I}_{MN}\,. \end{aligned}$$

In cases such as in the simplified fading scenario when the channel mean $\boldsymbol{m_h}$ is zero and the covariance $\boldsymbol{R_{hh}}$ is a scaled identity matrix, we call the fading statistics *non-informative*. The name is motivated by the fact that in this case the channel fading does not provide any channel knowledge of its own, since the symmetric nature of the conditional channel distribution, in the absence of channel information from the feedback link, makes all channel "directions" seem equally good. Non-informative fading statistics combined with either $\rho = 0$ and/or $b = 0$ hence means that the transmitter has *no channel knowledge*.

6.3 Performance Bounds and Code Design

To systematically incorporate channel knowledge into the code design process, the codeword error probability $\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}]$ is used as a basis for developing a performance criterion. This strategy is similar to the one adopted in Chapter 4. A feasible design criterion will be derived by forming a lower bound on a certain upper bound on $\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}]$. The following subsections develop these bounds and show how the resulting design criterion may be used for constructing efficient channel side information dependent codes.

6.3.1 Bounds Related to the Codeword Error Probability

In order to get a tractable closed-form expression, we start by exploiting the well-known union bound technique for upperbounding the conditional

codeword error probability $\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}|i,\boldsymbol{h}]$ so as to obtain

$$\begin{aligned}\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}|i,\boldsymbol{h}] &\leq \frac{1}{K}\sum_{k\neq l} P(\boldsymbol{C}_k^{(i)} \rightarrow \boldsymbol{C}_l^{(i)}|i,\boldsymbol{h}) \\ &= \frac{1}{K}\sum_{k\neq l} Q\left(\frac{\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(i)} - \boldsymbol{C}_l^{(i)})\|_{\mathrm{F}}}{2\sqrt{\sigma^2/2}}\right) \\ &\leq \frac{1}{2K}\sum_{k\neq l} e^{-\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(i)}-\boldsymbol{C}_l^{(i)})\|_{\mathrm{F}}^2/(4\sigma^2)} \\ &= \frac{1}{K}\sum_{k<l} e^{-\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(i)}-\boldsymbol{C}_l^{(i)})\|_{\mathrm{F}}^2/(4\sigma^2)}\,, \end{aligned} \tag{6.10}$$

where $P(\boldsymbol{C}_k^{(i)} \rightarrow \boldsymbol{C}_l^{(i)}|i,\boldsymbol{h})$, with $k \neq l$, denotes the pairwise error probability

$$\Pr\left[\|\boldsymbol{X} - \boldsymbol{H}^*\boldsymbol{C}_k^{(i)}\|_{\mathrm{F}}^2 > \|\boldsymbol{X} - \boldsymbol{H}^*\boldsymbol{C}_l^{(i)}\|_{\mathrm{F}}^2 \,|\, \boldsymbol{C} = \boldsymbol{C}_k^{(i)}, i, \boldsymbol{h}\right] \tag{6.11}$$

and where the Gaussian tail function $Q(x)$ is upperbounded [WJ90, p. 84] by $0.5\exp(-x^2/2)$. Note that (6.11) represents the probability that the ML decoding metric for $\boldsymbol{C}_l^{(i)}$ is smaller than the metric for $\boldsymbol{C}_k^{(i)}$, conditioned on $\boldsymbol{h}$, i and on the event that the transmitted codeword is $\boldsymbol{C} = \boldsymbol{C}_k^{(i)}$.

An upper bound $P_{\mathrm{UB}}(\{\mathcal{C}(i)\})$ on the codeword error probability is obtained by averaging over i and $\boldsymbol{h}$ in (6.10) to arrive at

$$\begin{aligned}\Pr[\hat{\boldsymbol{C}} \neq \boldsymbol{C}] &\leq \frac{1}{K}\sum_{k<l} \mathrm{E}_{i,\boldsymbol{h}}\left[e^{-\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(i)}-\boldsymbol{C}_l^{(i)})\|_{\mathrm{F}}^2/(4\sigma^2)}\right] \\ &= \frac{1}{K}\sum_{k<l} \mathrm{E}_{i,\boldsymbol{\gamma}}\left[\mathrm{E}_{\boldsymbol{h}}\left[e^{-\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(i)}-\boldsymbol{C}_l^{(i)})\|_{\mathrm{F}}^2/(4\sigma^2)}|i,\boldsymbol{\gamma}\right]\right] \\ &= \frac{1}{K}\sum_{k<l} \mathrm{E}_{i,\boldsymbol{\gamma}}\left[\mathrm{E}_{\boldsymbol{h}}\left[e^{-\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(i)}-\boldsymbol{C}_l^{(i)})\|_{\mathrm{F}}^2/(4\sigma^2)}|\boldsymbol{\gamma}\right]\right] \\ &\triangleq P_{\mathrm{UB}}(\{\mathcal{C}(i)\})\,, \end{aligned} \tag{6.12}$$

where the second equality is due to the fact that $\boldsymbol{h}$, $\boldsymbol{\gamma}$ and i form a Markov chain $\boldsymbol{h} - \boldsymbol{\gamma} - i$. Comparing $\mathrm{E}_{\boldsymbol{h}}\left[e^{-\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(i)}-\boldsymbol{C}_l^{(i)})\|_{\mathrm{F}}^2/(4\sigma^2)}|\boldsymbol{\gamma}\right]$ with the definition of the previous codeword pair criterion in (4.7), while keeping in mind that $\boldsymbol{\gamma}$ corresponds to $\boldsymbol{\zeta}$ in (4.7), shows that (4.11) can be used

to obtain

$$V(\boldsymbol{C}_k^{(i)}, \boldsymbol{C}_l^{(i)}|\boldsymbol{\gamma}) \triangleq \frac{1}{2}\,\mathrm{E}_{\boldsymbol{h}}\left[e^{-\|\boldsymbol{H}^*(\boldsymbol{C}_k^{(i)}-\boldsymbol{C}_l^{(i)})\|_{\mathrm{F}}^2/(4\sigma^2)}|\boldsymbol{\gamma}\right] = \frac{e^{\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^* \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1}\left(\boldsymbol{\Psi}(\boldsymbol{C}_k^{(i)}-\boldsymbol{C}_l^{(i)})^{-1}-\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\right)\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}}}{2\det(\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}})\det\left(\boldsymbol{\Psi}(\boldsymbol{C}_k^{(i)}-\boldsymbol{C}_l^{(i)})\right)}, \quad (6.13)$$

where now

$$\boldsymbol{\Psi}(\boldsymbol{C}) \triangleq \boldsymbol{I}_N \otimes \boldsymbol{C}\boldsymbol{C}^*\eta + \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1}, \quad \eta \triangleq 1/(4\sigma^2)\,.$$

Inserting (6.13) into (6.12) then gives

$$\begin{aligned} P_{\mathrm{UB}}(\{\mathcal{C}(i)\}) &= \frac{2}{K}\sum_{k<l}\mathrm{E}_{i,\boldsymbol{\gamma}}\left[V(\boldsymbol{C}_k^{(i)}, \boldsymbol{C}_l^{(i)}|\boldsymbol{\gamma})\right] \\ &= \frac{2}{K}\sum_{k<l}\sum_{i=0}^{2^b-1} p_i\,\mathrm{E}_{\boldsymbol{\gamma}}\left[V(\boldsymbol{C}_k^{(i)}, \boldsymbol{C}_l^{(i)}|\boldsymbol{\gamma})|i\right] \\ &= \frac{2}{K}\sum_{i=0}^{2^b-1} p_i \sum_{k<l}\mathrm{E}_{\boldsymbol{\gamma}}\left[V(\boldsymbol{C}_k^{(i)}, \boldsymbol{C}_l^{(i)}|\boldsymbol{\gamma})|i\right], \end{aligned} \quad (6.14)$$

where p_i denotes the probability that the feedback encoder outputs the integer value i.

With the aim of obtaining a closed-form approximation of $P_{\mathrm{UB}}(\{\mathcal{C}(i)\})$, let

$$\begin{aligned} \ell(\boldsymbol{C}_k^{(i)}, \boldsymbol{C}_l^{(i)}|\boldsymbol{\gamma}) \triangleq\ & \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^* \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1}\boldsymbol{\Psi}(\boldsymbol{C}_k^{(i)}-\boldsymbol{C}_l^{(i)})^{-1}\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}} \\ & - \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^* \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}} - \log\det\left(\boldsymbol{\Psi}(\boldsymbol{C}_k^{(i)}-\boldsymbol{C}_l^{(i)})\right), \end{aligned} \quad (6.15)$$

which we note corresponds to the criterion function in (4.15) but with an additional codeword independent term $-\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^* \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}$. Inserting (6.13) into (6.14) and utilizing the convexity of e^x and Jensen's inequality [CT91, p. 25] gives a lower bound $P_{\mathrm{LBUB}}(\{\mathcal{C}(i)\})$ on the upper bound $P_{\mathrm{UB}}(\{\mathcal{C}(i)\})$ as follows

$$P_{\mathrm{UB}}(\{\mathcal{C}(i)\}) = \frac{1}{K\det(\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}})}\sum_{i=0}^{2^b-1} p_i \sum_{k<l}\mathrm{E}_{\boldsymbol{\gamma}}\left[e^{\ell(\boldsymbol{C}_k^{(i)}, \boldsymbol{C}_l^{(i)}|\boldsymbol{\gamma})}|i\right]$$

$$\geq \frac{1}{K \det(\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}})} \sum_{i=0}^{2^b-1} p_i \sum_{k<l} e^{\mathrm{E}_{\boldsymbol{\gamma}}[\ell(\boldsymbol{C}_k^{(i)}, \boldsymbol{C}_l^{(i)}|\boldsymbol{\gamma})|i]} \tag{6.16}$$

$$\triangleq P_{\mathrm{LBUB}}(\{\mathcal{C}(i)\}),$$

which holds with equality if $\ell(\boldsymbol{C}_k^{(i)}, \boldsymbol{C}_l^{(i)}|\boldsymbol{\gamma})$, for a fixed i, is a constant.

In general, the lower bound $P_{\mathrm{LBUB}}(\{\mathcal{C}(i)\})$ approximates $P_{\mathrm{UB}}(\{\mathcal{C}(i)\})$ well when the variance of $\ell(\boldsymbol{C}_k^{(i)}, \boldsymbol{C}_l^{(i)}|\boldsymbol{\gamma})$ conditioned on i is small. From (6.15) it is seen that this is the case if the quantization is dense (so that $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}$ is concentrated around $\mathrm{E}_{\boldsymbol{\gamma}}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}|i]$), σ^2 is low and/or the correlation between $\boldsymbol{\gamma}$ and $\boldsymbol{h}$ is small. The final form of $P_{\mathrm{LBUB}}(\{\mathcal{C}(i)\})$ is obtained by utilizing the relation $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$ and the fact that $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}$ is constant with respect to $\boldsymbol{\gamma}$, as evident from (6.7). With a slight abuse of notation, the exponent in (6.16) can then be rewritten as

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{\gamma}}\left[\ell(\boldsymbol{C}_k^{(i)}, \boldsymbol{C}_l^{(i)}|\boldsymbol{\gamma})|i\right] &= \mathrm{E}_{\boldsymbol{\gamma}}\left[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^* \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1} \boldsymbol{\Psi}^{-1} \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}|i\right] \\
&\quad - \mathrm{E}_{\boldsymbol{\gamma}}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^* \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}|i] - \mathrm{E}_{\boldsymbol{\gamma}}\left[\log\det(\boldsymbol{\Psi})|i\right] \\
&= \mathrm{tr}(\boldsymbol{\Psi}^{-1} \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1} \mathrm{E}_{\boldsymbol{\gamma}}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*|i] \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1}) \\
&\quad - \mathrm{tr}(\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1} \mathrm{E}_{\boldsymbol{\gamma}}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*|i]) - \log\det(\boldsymbol{\Psi}).
\end{aligned} \tag{6.17}$$

Using (6.6) it is easily verified that $\mathrm{E}_{\boldsymbol{\gamma}}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*|i]$ can be written in terms of the first and second order moments $\mathrm{E}[\boldsymbol{\gamma}|i]$, $\mathrm{E}[\boldsymbol{\gamma}\boldsymbol{\gamma}^*|i]$. In the code design process, both these moments are replaced by the corresponding sample estimates taken from a Monte-Carlo simulation of the encoder in the feedback link.

Note that the development in this subsection is general in the sense that it does not rely on the assumption of a specific space-time code. Hence, the derived bounds can be used for assessing the performance of any space-time code.

6.3.2 The Code Design Problem

In principle, as suggested in our work in [JSO02b], it is possible to construct codes by minimizing $P_{\mathrm{UB}}(\{\mathcal{C}(i)\})$ with respect to the codewords, while satisfying the output power constraint. However, the computational complexity of such an approach is challenging, to say the least, since the conditional expectation in (6.14) is not available in closed-form and since it depends on the codewords to be optimized. Motivated by

the fact that $P_{\mathrm{LBUB}}(\{\mathcal{C}(i)\})$ approximates $P_{\mathrm{UB}}(\{\mathcal{C}(i)\})$ well in certain cases, we therefore take on an alternative approach in which the goal is to minimize $P_{\mathrm{LBUB}}(\{\mathcal{C}(i)\})$. This criterion function is more suitable from a computational point of view since the conditional expectations do not depend on the codewords and can hence be evaluated prior to the code search.

Unstructured codes, linear dispersive codes and weighted OSTBC, as described in Section 3.2, are considered. Design procedures for these code types are developed below.

Unstructured Codes

Note from (6.16) and the power constraint (6.3) that the design problem decouples and hence each unstructured code $\mathcal{C}(i)$ can be optimized separately from all others without loss of optimality. Consequently, based on (6.16) and (6.17), the *design criterion* is defined as

$$W_{\mathrm{q}}(\boldsymbol{\mathcal{C}}|i) \triangleq \sum_{k<l} V_{\mathrm{q}}(\boldsymbol{C}_k - \boldsymbol{C}_l|i)\,, \tag{6.18}$$

where $\boldsymbol{\mathcal{C}} \triangleq \begin{bmatrix}\boldsymbol{C}_1 & \boldsymbol{C}_2 & \cdots & \boldsymbol{C}_K\end{bmatrix}$ contains K arbitrary codewords and

$$V_{\mathrm{q}}(\boldsymbol{C}|i) \triangleq \frac{e^{\mathrm{tr}(\boldsymbol{\Psi}(\boldsymbol{C})^{-1}\boldsymbol{R}^{-1}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\,\mathrm{E}_{\boldsymbol{\gamma}}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}^{*}_{\boldsymbol{h}|\boldsymbol{\gamma}}|i]\boldsymbol{R}^{-1}_{\boldsymbol{hh}|\boldsymbol{\gamma}})}}{\det(\boldsymbol{\Psi}(\boldsymbol{C}))}\,. \tag{6.19}$$

The index "q" has here been introduced to emphasize that *quantized* feedback is considered. For the present case of unstructured codes, we minimize $P_{\mathrm{LBUB}}(\{\mathcal{C}(i)\})$, subject to the power constraint in (6.3), by designing the codes as

$$\boldsymbol{\mathcal{C}}^{(i)} = \arg \min_{\substack{\boldsymbol{\mathcal{C}} \\ \|\boldsymbol{\mathcal{C}}\|_{\mathrm{F}}^2 = PK\log_2(K)}} W_{\mathrm{q}}(\boldsymbol{\mathcal{C}}|i), \qquad i = 0, \cdots, 2^b - 1\,, \tag{6.20}$$

where the codewords in $\boldsymbol{\mathcal{C}}^{(i)}$ correspond to the ith code $\mathcal{C}(i)$. A strategy for numerically solving (6.20) will be presented in Section 6.5.

Observe that the solution to (6.20) is not unique. Clearly, all codeword matrices can be multiplied from the right by a common $L \times L$ unitary matrix without affecting neither the cost function nor the power constraint. However, if desired, this ambiguity can be removed by imposing a lower triangular structure onto one of the codewords.

Note also that the above criterion function is closely related to the one that was derived in Chapter 4 based on the assumption that $\boldsymbol{h}|\boldsymbol{\zeta}$ is complex Gaussian. In fact, if $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}$ and $\mathrm{E}_{\boldsymbol{\gamma}}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^{*}|i]$ in (6.19) are replaced with $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ and $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}^{*}$, respectively, then $V_{\mathrm{q}}(\boldsymbol{C}_k - \boldsymbol{C}_l|i)$ is equal to the corresponding term $V(\boldsymbol{C}_k - \boldsymbol{C}_l|\boldsymbol{\zeta})$ in (4.14). Hence, the criterion functions in (6.20) and in the corresponding design problem for unstructured codes in (4.18) differ only in the choice of certain parameters, but not in structure.

Design procedures based on the present criterion function are also easily developed for both linear dispersive codes as well as weighted OSTBC. This since the design procedures may be directly inferred from (4.20) and (4.26), respectively, by simply modifying the constituent criterion functions. For completeness, the specifics for each code type are given below.

Linear Dispersive Codes

From the performance criterion in (6.18) and from the previously described design procedure given by (4.20) it is clear that the weights $\{\boldsymbol{B}_m^{(i)}\}_{m=1}^{L_{\mathrm{d}}}$ in the ith linear dispersive code may be designed as

$$\boldsymbol{\mathcal{B}}^{(i)} = \arg \min_{\substack{\boldsymbol{\mathcal{B}} \\ \|\boldsymbol{\mathcal{B}}\|_{\mathrm{F}}^2 = P\log_2(K)}} W_{\mathrm{q}}(\boldsymbol{\mathcal{C}}(\boldsymbol{\mathcal{B}})|i)\,, \tag{6.21}$$

where

$$\begin{aligned}
\boldsymbol{\mathcal{B}} &\triangleq \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 & \cdots & \boldsymbol{B}_{L_{\mathrm{d}}} \end{bmatrix} \\
\boldsymbol{\mathcal{B}}^{(i)} &\triangleq \begin{bmatrix} \boldsymbol{B}_1^{(i)} & \boldsymbol{B}_2^{(i)} & \cdots & \boldsymbol{B}_{L_{\mathrm{d}}}^{(i)} \end{bmatrix} \\
\boldsymbol{\mathcal{C}}(\boldsymbol{\mathcal{B}}) &\triangleq \begin{bmatrix} \sum_{m=1}^{L_{\mathrm{d}}} \boldsymbol{B}_m s_m^{(1)} & \sum_{m=1}^{L_{\mathrm{d}}} \boldsymbol{B}_m s_m^{(2)} & \cdots & \sum_{m=1}^{L_{\mathrm{d}}} \boldsymbol{B}_m s_m^{(K)} \end{bmatrix},
\end{aligned}$$

with $s_m^{(k)}$, $m = 1, 2, \cdots, L_{\mathrm{d}}$ representing the symbol sequence corresponding to the kth codeword.

Weighted OSTBC

Similarly to as in Section 4.5, the transmit weighting can be designed by only considering the worst codeword pairs in the fixed OSTB code

$\{\bar{\boldsymbol{C}}_k\}_{k=1}^K$. In other words, after substituting $\boldsymbol{C}_k = \boldsymbol{W}\bar{\boldsymbol{C}}_k$ into (6.18), only the terms corresponding to codeword pairs such that

$$\boldsymbol{A}(\bar{\boldsymbol{C}}_k, \bar{\boldsymbol{C}}_l) \triangleq (\bar{\boldsymbol{C}}_k - \bar{\boldsymbol{C}}_l)(\bar{\boldsymbol{C}}_k - \bar{\boldsymbol{C}}_l)^* = \mu_{\min}\boldsymbol{I}_{M'}, \quad k \neq l$$

are retained. Here, $\mu_{\min}$ denotes the minimum value over all $\mu_{kl} \triangleq \sum_{m=1}^{L_\mathrm{o}} \left(s_m^{(k)} - s_m^{(l)}\right)^2$. Let

$$\ell_\mathrm{q}(\boldsymbol{W}\boldsymbol{W}^*, \mu_{kl}|i) \triangleq \log\left(V_\mathrm{q}(\boldsymbol{W}(\bar{\boldsymbol{C}}_k - \bar{\boldsymbol{C}}_l)|i)\right).$$

Since the remaining terms are all equal, they are minimized by designing the codebook of transmit weighting matrices $\{\boldsymbol{W}^{(i)}\}_{i=0}^{2^b-1}$ as

$$\boldsymbol{W}^{(i)} = \arg \min_{\substack{\boldsymbol{W} \\ \|\boldsymbol{W}\|_\mathrm{F}^2 = P_\mathrm{o}}} \ell_\mathrm{q}(\boldsymbol{W}\boldsymbol{W}^*, \mu_{\min}|i), \quad i = 0, 1, \cdots, 2^b.$$

Just as in Section 4.5, the above optimization problem may be solved in two steps. The first step is to re-parameterize using $\boldsymbol{W}\boldsymbol{W}^* = \boldsymbol{Z}$ and then solve the equivalent problem

$$\boldsymbol{Z}^{(i)} = \arg \min_{\substack{\boldsymbol{Z} \\ \boldsymbol{Z} = \boldsymbol{Z}^* \succeq \boldsymbol{0} \\ \mathrm{rank}(\boldsymbol{Z}) \leq \min\{M', M\} \\ \mathrm{tr}(\boldsymbol{Z}) = P_\mathrm{o}}} \ell_\mathrm{q}(\boldsymbol{Z}|i), \quad i = 0, 1, \cdots, 2^b, \tag{6.22}$$

where

$$\begin{aligned}\ell_\mathrm{q}(\boldsymbol{Z}|i) &\triangleq \ell_\mathrm{q}(\boldsymbol{Z}, \mu_{\min}|i) \\ &= \mathrm{tr}\left(\left((\boldsymbol{I}_N \otimes \boldsymbol{Z})\tilde{\eta} + \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}}^{-1}\right)^{-1} \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}}^{-1} \mathrm{E}_{\boldsymbol{\gamma}}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*|i]\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}}^{-1}\right) \\ &\qquad - \log\det\left((\boldsymbol{I}_N \otimes \boldsymbol{Z})\tilde{\eta} + \boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}}^{-1}\right),\end{aligned} \tag{6.23}$$

with $\tilde{\eta} \triangleq \mu_{\min}/(4\sigma^2)$. In the second and final step, the transmit weighting is obtained as $\boldsymbol{W}^{(i)} = \left(\boldsymbol{Z}^{(i)}\right)^{1/2}$, where $(\cdot)^{1/2}$ denotes a matrix square-root.

Because of the previously mentioned similarity between $V_\mathrm{q}(\boldsymbol{C}_k - \boldsymbol{C}_l|i)$ in (6.19) and $V(\boldsymbol{C}_k - \boldsymbol{C}_l|\boldsymbol{\zeta})$ in (4.14), it is not difficult to show that basically all the theoretical results based on the latter criterion function carry over, sometimes in somewhat modified form, to the present case. In particular, it should be noted that also (6.22) is a convex optimization problem if the transmit weighting is square, i.e., $M' = M$. Moreover, it is straightforward to modify the algorithm for a simplified scenario described in Section 4.7 to deal with the new criterion function. Thus, the transmit weight design problem given by (6.22) can be efficiently solved using simple modifications of techniques discussed earlier.

6.4 Unstructured Codes – Analysis and Interpretations

It is in general a difficult task to explicitly characterize the optimal set of codewords. However, in a few special cases it is possible to derive some properties that the optimal codewords must posses. The analysis presented in this section focuses entirely on unstructured codes. It is however straightforward to perform similar analyses for the two other code types. The results to be derived herein are important for understanding the behavior of the transmitter and also help to explain the simulation results in Section 6.7.

Several of the following cases under consideration are asymptotic in the sense that optimal unstructured codes obtained from (6.20) are to be investigated as some parameter tends to a limit. In such asymptotic cases, we rely, similarly to as in Appendix 4.A, on a theorem [SS89, p. 221] stating that it is possible to interchange the order of the limit and minimization operator in an optimization problem provided that 1) the criterion function converges uniformly over the feasibility set to a limit function, 2) the feasibility set is compact, 3) the limit function is continuous and has a unique global minimum within the feasibility set. If the limit function possesses several (i.e., non-unique) global minima then the theorem reduces to a weaker form claiming that an asymptotically optimal solution belongs to the set of global minima of the limit function.

Applying the weaker form of the theorem to the design procedure means that (6.20) may be studied with $W_{\mathrm{q}}(\boldsymbol{\mathcal{C}}|i)$ replaced by the corresponding limit function. Using techniques similar to the ones used in Appendix 4.A, it can be shown that such an interchange of order is allowed. Therefore, to simplify the presentation we will in the following omit the proofs on uniform convergence and instead tacitly assume that the limit function can be used for deriving the asymptotic properties of the optimal code.

6.4.1 Perfect Channel Knowledge

Consider first the case of perfect *initial* channel information, i.e., $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}} \to \bar{\boldsymbol{R}}_{\boldsymbol{h}\boldsymbol{\gamma}}$, where $\bar{\boldsymbol{R}}_{\boldsymbol{h}\boldsymbol{\gamma}}$ is such that $\rho \to 1$. Introduce the $M \times N$ channel estimate matrix $\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}} \triangleq \mathrm{E}[\boldsymbol{H}|\boldsymbol{\gamma}]$, also obtained by rearranging $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}$ so as to satisfy the relation $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}} = \mathrm{vec}(\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}})$. A Taylor expansion of $\log(V_{\mathrm{q}}(\boldsymbol{C}_k - \boldsymbol{C}_l|i))$ then shows that a criterion function equivalent to

$W_{\mathrm{q}}(\boldsymbol{\mathcal{C}}|i)$ in (6.20) may be written as

$$\begin{aligned}\tilde{W}_1(\boldsymbol{\mathcal{C}}|i) &\triangleq W_{\mathrm{q}}(\boldsymbol{\mathcal{C}}|i)/e^{\operatorname{tr}(\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1}\,\mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^{*}|i])+\log\det(\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}})}\\ &= \sum_{k<l} e^{-\eta\operatorname{tr}(\boldsymbol{I}_N\otimes\boldsymbol{A}_{kl}(\mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^{*}|i]+\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}))+\mathrm{O}(\eta^2\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\|)}\\ &= \sum_{k<l} e^{-\eta\operatorname{tr}(\boldsymbol{A}_{kl}(\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}^{*}|i]+\tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}}))+\mathrm{O}(\eta^2\|\tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\|)}\,, \end{aligned} \tag{6.24}$$

where $\mathrm{O}(\cdot)$ is the big ordo operator and where

$$\begin{aligned}\boldsymbol{A}_{kl} &\triangleq (\boldsymbol{C}_k-\boldsymbol{C}_l)(\boldsymbol{C}_k-\boldsymbol{C}_l)^*\\ \tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}} &\triangleq \mathrm{E}[(\boldsymbol{H}-\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}})(\boldsymbol{H}-\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}})^*|\boldsymbol{\gamma}]\,.\end{aligned}$$

Note that $\tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}}$ can also be expressed as the sum of the N blocks of size $M\times M$ on the diagonal of $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}$, explaining why $\mathrm{O}(\eta^2\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\|)$ may be replaced by $\mathrm{O}(\eta^2\|\tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\|)$. Observing that $\boldsymbol{A}_{kl}$ is bounded due to the power constraint, dropping η from the ordo and making use of the Taylor expansion $e^x = 1+\mathrm{O}(x)$ leads to

$$\begin{aligned}\tilde{W}_1(\boldsymbol{\mathcal{C}}|i) &= \sum_{k<l} e^{-\eta\operatorname{tr}(\boldsymbol{A}_{kl}\,\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}^{*}|i])}e^{\mathrm{O}(\|\tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\|)}\\ &= \sum_{k<l} e^{-\eta\operatorname{tr}(\boldsymbol{A}_{kl}\,\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}^{*}|i])}+\mathrm{O}(\|\tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\|)\,.\end{aligned}$$

As previously noted, $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\to\bar{\boldsymbol{R}}_{\boldsymbol{h}\boldsymbol{\gamma}}$ implies $\|\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\|\to 0$ and hence also $\|\tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\|\to 0$. Thus, $\tilde{W}_1(\boldsymbol{\mathcal{C}}|i)$ converges to the limit function

$$\begin{aligned}\bar{W}_1(\boldsymbol{\mathcal{C}}|i) &\triangleq \lim_{\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\to\bar{\boldsymbol{R}}_{\boldsymbol{h}\boldsymbol{\gamma}}}\tilde{W}_1(\boldsymbol{\mathcal{C}}|i)\\ &= \sum_{k<l} e^{-\eta\operatorname{tr}((\boldsymbol{C}_k-\boldsymbol{C}_l)(\boldsymbol{C}_k-\boldsymbol{C}_l)^*\boldsymbol{R}_i)}\,,\end{aligned} \tag{6.25}$$

where

$$\boldsymbol{R}_i \triangleq \mathrm{E}[\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\gamma}}\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\gamma}}^{*}|i]\,,$$

with $\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\gamma}} \triangleq \lim_{\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\to\bar{\boldsymbol{R}}_{\boldsymbol{h}\boldsymbol{\gamma}}}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}$.

Before discussing the above result, we consider the related case of perfect channel knowledge *at the transmitter*. Hence, in addition to perfect initial channel information, assume the feedback is such that

the quantization is infinitely dense in the $\boldsymbol{\gamma}$ domain[1]. In such a case, $\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}^{*}|i]$ can be expected to be essentially equal to $\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}^{*}$ for a $\boldsymbol{\gamma}$ corresponding to some infinitely small encoder region $\mathcal{S}_i$. We formalize this as follows. Assume a sequence of quantization schemes is chosen so that the corresponding sequence of encoder regions $\mathcal{S}_i^{(b)}$, $b = 0, 1, \cdots$, for a particular i, all enclose a certain fixed point $\hat{\boldsymbol{\gamma}}$ in the vector space $\mathbb{C}^{MN\times 1}$. Let

$$r_b \triangleq \max_{\boldsymbol{\gamma}\in\mathcal{S}_i^{(b)}} \|\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}\|$$

be the radius of a sphere centered at $\hat{\boldsymbol{\gamma}}$ that encloses $\mathcal{S}_i^{(b)}$ and assume that $\mathcal{S}_i^{(b)}$ becomes smaller as b is increased in the sense that $\lim_{b\to\infty} r_b = 0$. From the mean-value theorem it can then be shown that

$$\begin{aligned}\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}^{*}|i] &= \int_{\mathcal{S}_i^{(b)}} \boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}^{*} p_{\boldsymbol{\gamma}|i}(\boldsymbol{\gamma}|i)d\boldsymbol{\gamma} \\ &= \boldsymbol{M}_{\boldsymbol{H}|\hat{\boldsymbol{\gamma}}}\boldsymbol{M}_{\boldsymbol{H}|\hat{\boldsymbol{\gamma}}}^{*} + \mathrm{O}(r_b)\,, \end{aligned} \tag{6.26}$$

where $\boldsymbol{M}_{\boldsymbol{H}|\hat{\boldsymbol{\gamma}}} \triangleq \mathrm{E}[\boldsymbol{H}|\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}]$. With (6.24) as a starting point, inserting (6.26) into (6.24), using the Taylor expansion $e^x = 1 + \mathrm{O}(x)$ and taking the limit $(\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}, r_b) \to (\bar{\boldsymbol{R}}_{\boldsymbol{h}\boldsymbol{\gamma}}, 0)$ gives a limit function exactly like the one in (6.25) but now with

$$\boldsymbol{R}_i \triangleq \bar{\boldsymbol{M}}_{\boldsymbol{H}|\hat{\boldsymbol{\gamma}}}\bar{\boldsymbol{M}}_{\boldsymbol{H}|\hat{\boldsymbol{\gamma}}}^{*}\,,$$

where $\bar{\boldsymbol{M}}_{\boldsymbol{H}|\hat{\boldsymbol{\gamma}}} \triangleq \lim_{\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}}\to\bar{\boldsymbol{R}}_{\boldsymbol{h}\boldsymbol{\gamma}}} \boldsymbol{M}_{\boldsymbol{H}|\hat{\boldsymbol{\gamma}}}$.

Based on the limit functions, properties of the optimal code are derived as follows. Let $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{*}$ be the EVD of $\boldsymbol{R}_i$, where $\boldsymbol{U}$ is a unitary $M \times M$ matrix with columns containing the eigenvectors of $\boldsymbol{R}_i$ and $\boldsymbol{\Lambda}$ is an $M \times M$ diagonal matrix with ordered diagonal elements $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_M = 0$ representing the eigenvalues. Here, r is the rank of $\boldsymbol{R}_i$. Due to the unitary nature of $\boldsymbol{U}$, the optimization problem with (6.25) as criterion function can be re-parameterized based on $\bar{\boldsymbol{C}}_k \triangleq \boldsymbol{U}^{*}\boldsymbol{C}_k$ and $\bar{\boldsymbol{\mathcal{C}}} \triangleq \boldsymbol{U}^{*}\boldsymbol{\mathcal{C}}$, without affecting the power con-

[1] Obviously, the phase-only quantization scheme used as an example in this work does not belong to this category.

straint since $\|\bar{\boldsymbol{\mathcal{C}}}\|_{\mathrm{F}}^2 = \|\boldsymbol{\mathcal{C}}\|_{\mathrm{F}}^2$, to arrive at

$$\begin{aligned}\bar{W}_1(\boldsymbol{U}\bar{\boldsymbol{\mathcal{C}}}|i) &= \sum_{k<l} e^{-\eta\,\mathrm{tr}((\bar{\boldsymbol{C}}_k-\bar{\boldsymbol{C}}_l)(\bar{\boldsymbol{C}}_k-\bar{\boldsymbol{C}}_l)^*\boldsymbol{\Lambda})} \\ &= \sum_{k<l} e^{-\eta\sum_{m=1}^{r}\lambda_m\|\bar{\boldsymbol{c}}_m^{(k)}-\bar{\boldsymbol{c}}_m^{(l)}\|^2}\,, \end{aligned} \tag{6.27}$$

where $(\bar{\boldsymbol{c}}_m^{(k)})^*$ is the mth row of $\bar{\boldsymbol{C}}_k$. It is seen that the value of the asymptotic criterion function in (6.27) does not depend on $\{\bar{\boldsymbol{c}}_m^{(k)}\}_{k=1}^K$, $m = r+1, \cdots, M$, i.e., the $M-r$ last rows of the re-parameterized codewords. Also, observe that letting these rows be non-zero decreases the available output power that can be allocated to the remaining rows and is hence wasteful. Since the optimal cost decreases as a function of the output power it can be concluded that the rows $\{\bar{\boldsymbol{c}}_m^{(k)}\}_{m=r+1}^M$ in all codewords must be zero. On the other hand, for the r first (potentially non-zero) rows, each eigenvalue λ_m weighs the corresponding row in $\bar{\boldsymbol{\mathcal{C}}}$. Because of the weighting, rows with a large eigenvalue are favored which in turn means that the energy allocated to a particular row tends to be larger for a row with a large eigenvalue than for one with a small eigenvalue. After solving the re-parameterized problem in (6.27), the code is formed as $\boldsymbol{\mathcal{C}}^{(i)} = \bar{\boldsymbol{U}}\bar{\boldsymbol{\mathcal{C}}}^{(i)}$, where $\bar{\boldsymbol{\mathcal{C}}}^{(i)}$ denotes the optimal $\bar{\boldsymbol{\mathcal{C}}}$. Consequently, for the asymptotically optimal code, the transmitted power is distributed only along the first r eigenvectors of $\boldsymbol{R}_i$, with typically large power in "eigen-directions" with large eigenvalues.

The above result is worth elaborating upon. Assume perfect channel knowledge at the transmitter so that $\boldsymbol{R}_i = \bar{\boldsymbol{M}}_{\boldsymbol{H}|\hat{\boldsymbol{\gamma}}}\bar{\boldsymbol{M}}_{\boldsymbol{H}|\hat{\boldsymbol{\gamma}}}^*$. From the SVD theorem [HJ96, p. 414], recall that $\boldsymbol{U}$ and $\{\sqrt{\lambda_m}\}_{m=1}^r$ also represent the left singular eigenvectors and singular values, respectively, of $\bar{\boldsymbol{M}}_{\boldsymbol{H}|\hat{\boldsymbol{\gamma}}}$. Due to the assumption of perfect channel knowledge, $\bar{\boldsymbol{M}}_{\boldsymbol{H}|\hat{\boldsymbol{\gamma}}}$ is now essentially equal to the channel matrix $\boldsymbol{H}$ and we may base the following discussion on the commonly used SVD of the channel matrix. Hence, since the last $M-r$ rows of $\bar{\boldsymbol{\mathcal{C}}}^{(i)}$ are zero, the transmission is confined to the space spanned by the left singular eigenvectors corresponding to the non-zero singular values of the channel matrix. Moreover, the power tends to be primarily allocated to strong eigen-directions, i.e., where the corresponding singular values are large. Such a behavior makes sense since there is no point wasting energy in directions which contribute little or not at all to the received signal. In particular, for the case of one receive antenna we have $r = 1$ and the optimal codewords are therefore of rank

one, resulting in a beamforming type of transmission.

In comparison, an information theoretic result derived in e.g. [Tel95] states that capacity in a MIMO system can be achieved by transmitting along the left singular eigenvectors of the channel matrix with a power allocation obeying a water-filling distribution. The water-filling procedure ensures that the power allocated along strong eigen-directions is large and that no power is wasted on eigen-directions having singular values equal to zero. This clearly agrees well with the perfect channel knowledge property derived above.

Note also that although the perfect channel knowledge case in the present work and in Section 4.6 may seem to be similar, it is risky to draw conclusions for the problem at hand based on the earlier development in that section, where the totally different weighted OSTBC structure was considered. In particular, if the channel knowledge is perfect, the asymptotically optimal codewords mentioned in Section 4.6 are always of rank one, regardless of the number of receive antennas, while the same is in general not true for the code structure herein, unless only one receive antenna is used.

6.4.2 Low SNR

The low SNR case is defined as $P/\sigma^2 = 4P\eta \to 0$. For technical reasons, we start with re-parameterizing the optimization problem in (6.20) based on $\tilde{\boldsymbol{C}}_k \triangleq \boldsymbol{C}_k/\sqrt{P}$ and $\tilde{\boldsymbol{\mathcal{C}}} \triangleq \boldsymbol{\mathcal{C}}/\sqrt{P}$. With η replaced by $P\eta$, (6.24) can be used as a starting point for deriving the first order (in terms of $P\eta$) Taylor expansion

$$\begin{aligned}
\tilde{W}_2(\tilde{\boldsymbol{\mathcal{C}}}|i) &\triangleq W_{\mathrm{q}}(\sqrt{P}\tilde{\boldsymbol{\mathcal{C}}}|i) \\
&= \sum_{k<l} e^{-P\eta\,\mathrm{tr}(\tilde{\boldsymbol{A}}_{kl}(\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}^*_{\boldsymbol{H}|\boldsymbol{\gamma}}|i]+\tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}}))+\mathrm{O}(P^2\eta^2)} \\
&= \sum_{k<l} 1 - P\eta\,\mathrm{tr}(\tilde{\boldsymbol{A}}_{kl}(\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}^*_{\boldsymbol{H}|\boldsymbol{\gamma}}|i] + \tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}})) + \mathrm{O}(P^2\eta^2) \\
&= -P\eta\,\mathrm{tr}\Big(\Big(\sum_{k<l}\tilde{\boldsymbol{A}}_{kl}\Big)(\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}^*_{\boldsymbol{H}|\boldsymbol{\gamma}}|i] + \tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}})\Big) \\
&\qquad + \frac{K(K-1)}{2} + \mathrm{O}(P^2\eta^2)\,, \quad (6.28)
\end{aligned}$$

where $\tilde{\boldsymbol{A}}_{kl} \triangleq (\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l)(\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l)^*$ and $K(K-1)/2$ is the number of different codeword pairs. From (6.28) it is clear that the asymptotic prop-

erties of the optimal code can be analyzed by considering maximization of the limit function

$$\begin{aligned}\bar{W}_2(\tilde{\boldsymbol{\mathcal{C}}}|i) &\triangleq \lim_{P\eta\to 0}\frac{-1}{P\eta}\left(\tilde{W}_2(\tilde{\boldsymbol{\mathcal{C}}}|i)-\frac{K(K-1)}{2}\right)\\ &= \mathrm{tr}\left(\Big(\sum_{k<l}\tilde{\boldsymbol{A}}_{kl}\Big)(\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}^*_{\boldsymbol{H}|\boldsymbol{\gamma}}|i]+\tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}})\right).\end{aligned}$$

Assume that the rank of $\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}^*_{\boldsymbol{H}|\boldsymbol{\gamma}}|i]+\tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}}$ is r and let $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$ represent the corresponding EVD. With a similar approach as was used to obtain (6.27), the limit function can be re-parameterized using $\bar{\boldsymbol{C}}_k \triangleq \boldsymbol{U}^*\tilde{\boldsymbol{C}}_k$, $\bar{\boldsymbol{\mathcal{C}}} \triangleq \boldsymbol{U}^*\tilde{\boldsymbol{\mathcal{C}}}$ and written as

$$\bar{W}_2(\boldsymbol{U}\bar{\boldsymbol{\mathcal{C}}}|i)=\sum_{m=1}^{r}\lambda_m\sum_{k<l}\|\bar{\boldsymbol{c}}_m^{(k)}-\bar{\boldsymbol{c}}_m^{(l)}\|^2\,. \tag{6.29}$$

Note that also in this case must the $M-r$ last rows $\{\bar{\boldsymbol{c}}_m^{(k)}\}_{k=1}^K$, $m=r+1,\cdots,M$ be equal to zero for optimality. To proceed, define

$$f(\tilde{P}) \triangleq \max_{\substack{\{\bar{\boldsymbol{c}}_m^{(k)}\}_{k=1}^K\\ \sum_{k=1}^K\|\bar{\boldsymbol{c}}_m^{(k)}\|^2=\tilde{P}}}\sum_{k<l}\|\bar{\boldsymbol{c}}_m^{(k)}-\bar{\boldsymbol{c}}_m^{(l)}\|^2$$

and again re-parameterize based on $\bar{\boldsymbol{c}}^{(k)}=\bar{\boldsymbol{c}}_m^{(k)}/\sqrt{\tilde{P}}$ to obtain

$$\begin{aligned}f(\tilde{P}) &= \tilde{P}\max_{\substack{\{\bar{\boldsymbol{c}}^{(k)}\}_{k=1}^K\\ \sum_{k=1}^K\|\bar{\boldsymbol{c}}^{(k)}\|^2=1}}\sum_{k<l}\|\bar{\boldsymbol{c}}^{(k)}-\bar{\boldsymbol{c}}^{(l)}\|^2\\ &= \alpha\tilde{P}\,,\end{aligned} \tag{6.30}$$

where α is a constant which can be interpreted as a measure of how well the constellation of rows are separated. Optimization of the rows in (6.29) can be carried out by first optimizing each row separately under a fixed power constraint $\sum_{k=1}^K\|\bar{\boldsymbol{c}}_m^{(k)}\|^2=\tilde{P}_m$, producing the concentrated criterion function $f(\tilde{P}_m)$, and then using $f(\tilde{P}_m)$ for optimizing the power values $\{\tilde{P}_m\}$ subject to $\sum_{k=1}^r\tilde{P}_m=K\log_2(K)$. By utilizing (6.30), it is seen that the optimal power values may hence be obtained as

$$\{P_m\}_{m=1}^r=\arg\max_{\substack{\{\tilde{P}_m\}_{m=1}^r\\ \sum_{m=1}^r\tilde{P}_m=K\log_2(K)}}\sum_{m=1}^{r}\lambda_m f(\tilde{P}_m)$$

$$= \arg \max_{\substack{\{\tilde{P}_m\}_{m=1}^r \\ \sum_{m=1}^r \tilde{P}_m = K\log_2(K)}} \sum_{m=1}^{r} \lambda_m \tilde{P}_m \,. \tag{6.31}$$

For simplicity, assume that λ_1 is strictly larger than all other eigenvalues. It is then easily verified that in the optimal solution of (6.31), all power is allocated to the first eigen-direction, i.e., $P_1 = K\log_2(K)$, $P_m = 0, m = 2, \cdots, M$. Thus, only the first row of an optimal $\tilde{\boldsymbol{\mathcal{C}}}$ is non-zero and, because of $\tilde{\boldsymbol{\mathcal{C}}} = \boldsymbol{U}\bar{\boldsymbol{\mathcal{C}}}$, the asymptotically optimal code distributes all power along the strongest eigenvector of $\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}^*|i] + \tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}}$.

A comparison with the perfect channel knowledge case reveals several similarities. Studying (6.27) and (6.29) we see that the perfect channel knowledge and low SNR cases are closely related. In both cases the transmission power is distributed along some eigen-directions and the amount of energy in each eigen-direction depends on weights in the form of certain eigenvalues. However, they primarily differ in that in the low SNR case only the *strongest* eigen-direction is used while for the perfect channel knowledge case the energy may be distributed among *all* eigen-directions with non-zero eigenvalues. In addition, for the low SNR case, the EVD is based not only on a feedback dependent term $\mathrm{E}[\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}\boldsymbol{M}_{\boldsymbol{H}|\boldsymbol{\gamma}}^*|i]$ but also on an additional term $\tilde{\boldsymbol{R}}_{\boldsymbol{hh}|\boldsymbol{\gamma}}$.

6.4.3 No Feedback / No Channel Knowledge

In this section, the cases of no feedback and no channel knowledge are analyzed. Consider first a scenario with no feedback, or more precisely, assume that $b = 0$ and/or $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{\gamma}} = \boldsymbol{0}$. In other words, the channel information at the transmitter is statistically independent of $\boldsymbol{\gamma}$ and hence also of $\boldsymbol{h}$. Consequently, $\mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*|i] = \mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*] = \boldsymbol{m}_{\boldsymbol{h}}\boldsymbol{m}_{\boldsymbol{h}}^*$ and $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} = \boldsymbol{R}_{\boldsymbol{hh}}$, which means that the design criterion no longer depends on the initial channel information $\boldsymbol{\gamma}$. In order to also remove the exponential in (6.19), the scenario is further specialized by assuming that $\boldsymbol{m}_{\boldsymbol{h}} = \boldsymbol{0}$. These assumptions model a non-line-of-sight scenario with a feedback delay that is very long compared to the coherence time of the channel. After scaling with $1/\det(\boldsymbol{R}_{\boldsymbol{hh}})$ the design criterion in (6.18) then takes on the equivalent form

$$\bar{W}_3(\boldsymbol{\mathcal{C}}|i) \triangleq \sum_{k<l} \frac{1}{\det(\boldsymbol{I}_{MN} + \boldsymbol{R}_{\boldsymbol{hh}}(\boldsymbol{I}_N \otimes \boldsymbol{A}_{kl})\eta)} \,, \tag{6.32}$$

which can be shown to be a valid upper bound on the codeword error probability also when $\boldsymbol{R_{hh}}$ is singular. Now, assume fading such that the columns $\{\boldsymbol{h}_l\}$ of $\boldsymbol{H}$ are uncorrelated but with the same *transmitter covariance* $\boldsymbol{R}_{\mathrm{T}} \triangleq \mathrm{E}[\boldsymbol{h}_l \boldsymbol{h}_l^*]$ or, equivalently, that the channel covariance matrix has the Kronecker structure $\boldsymbol{R_{hh}} = \boldsymbol{I}_N \otimes \boldsymbol{R}_{\mathrm{T}}$. This so-called "one-ring" channel model may be appropriate in scenarios where the transmitter is placed above the rooftops far away from scatterers while at the receiver there is plenty of local scattering [SFGK00]. The Kronecker structure of $\boldsymbol{R_{hh}}$ means that (6.32) simplifies to

$$\bar{W}_3(\boldsymbol{\mathcal{C}}|i) = \sum_{k<l} \frac{1}{\det^N(\boldsymbol{I}_M + \eta \boldsymbol{R}_{\mathrm{T}} \boldsymbol{A}_{kl})} \,. \tag{6.33}$$

To see how the criterion function in (6.33) affects the code design, let $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$ represent the EVD of $\boldsymbol{R}_{\mathrm{T}}$, with $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_M)$, $\boldsymbol{U}\boldsymbol{U}^* = \boldsymbol{I}_M$. Furthermore, use the relation $\det(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{I} + \boldsymbol{B}\boldsymbol{A})$, found in Appendix C, to express the determinant as

$$\det(\boldsymbol{I}_M + \eta \boldsymbol{R}_{\mathrm{T}} \boldsymbol{A}_{kl}) = \det\left(\boldsymbol{I}_L + \eta \sum_{m=1}^{M} \lambda_m (\bar{\boldsymbol{c}}_m^{(k)} - \bar{\boldsymbol{c}}_m^{(l)})(\bar{\boldsymbol{c}}_m^{(k)} - \bar{\boldsymbol{c}}_m^{(l)})^*\right),$$

where $(\bar{\boldsymbol{c}}_m^{(k)})^*$ is the mth row of $\bar{\boldsymbol{C}}_k \triangleq \boldsymbol{U}^* \boldsymbol{C}_k$. Hence, we again see that the transmission is along some eigen-directions, given by the columns of $\boldsymbol{U}$, and that the corresponding eigenvalues serve to weigh the eigen-directions.

Unfortunately, further development is complicated by the large number of terms in (6.33). Therefore, in order to obtain tractable expressions, a special case with only two codewords, $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$, is studied next. It is reasonable to assume that this provides valuable insight also into the general case of an arbitrary number of codewords. In the derivation, a necessary condition on the codewords of an optimal code will be utilized. The following lemma summarizes the result, which is valid for any number of codewords and for the general assumptions described in the system model.

Lemma 1 (A Necessary Condition for Optimality) Let $\{\boldsymbol{C}_k\}_{k=1}^K$ denote the codewords in an optimal code that solves the design problem in (6.20). These codewords then satisfy the necessary condition

$$\sum_{k=1}^{K} \boldsymbol{C}_k = \boldsymbol{0}_{M \times L} \,, \tag{6.34}$$

regardless of the quality of the channel information.

Proof:

The proof relies on the KKT conditions [BS93, p. 162] that are necessary for optimality. In order to obtain compact expressions, it is convenient to first define some additional notation for derivatives. Let $\frac{\partial f}{\partial \boldsymbol{X}}$ denote a matrix where element (k,l), is equal to $\frac{\partial f}{\partial [\boldsymbol{X}]_{kl}}$, i.e., it is equal to the derivative of $f \triangleq f(\boldsymbol{X})$ with respect to element $[\boldsymbol{X}]_{kl}$ of $\boldsymbol{X}$. Since the elements in the matrices of interest are typically complex-valued, the notion of a *complex derivative* [Kay93, p. 517] is utilized and defined as

$$\frac{\partial f}{\partial x} \triangleq \frac{1}{2}\left(\frac{\partial f}{\partial \operatorname{re}(x)} - j\frac{\partial f}{\partial \operatorname{im}(x)}\right), \tag{6.35}$$

where j is the imaginary unit, f is a function of x, x is complex-valued and $\operatorname{re}(x)$ and $\operatorname{im}(x)$ denote the real and imaginary part of x, respectively.

Let $\frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{C}_k}(\boldsymbol{\mathcal{C}})$ denote the derivative of $W_{\mathrm{q}}(\boldsymbol{\mathcal{C}}|i)$ with respect to the complex-valued elements of $\boldsymbol{C}_k$, evaluated at $\boldsymbol{\mathcal{C}}$. With δ_n representing the Kronecker delta function, the derivative can be written as

$$\frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{C}_k}(\boldsymbol{\mathcal{C}}) = \sum_{\substack{k'<l' \\ k'=k \text{ or } l'=k}} (1-2\delta_{l'-k})\boldsymbol{V}'_{\mathrm{q}}(\boldsymbol{C}_{k'} - \boldsymbol{C}_{l'}), \tag{6.36}$$

where, as shown in Appendix 6.A, the derivative $\boldsymbol{V}'_{\mathrm{q}}(\boldsymbol{C})$ of $V_{\mathrm{q}}(\boldsymbol{C}|i)$, evaluated at $\boldsymbol{C}$, is equal to

$$\boldsymbol{V}'_{\mathrm{q}}(\boldsymbol{C}) \triangleq \frac{\partial V_{\mathrm{q}}}{\partial \boldsymbol{C}}(\boldsymbol{C}) = -\eta V_{\mathrm{q}}(\boldsymbol{C}|i)(\boldsymbol{\Omega}\boldsymbol{C})^{\mathrm{c}}. \tag{6.37}$$

Here, $(\cdot)^{\mathrm{c}}$ denotes the complex conjugate and $\boldsymbol{\Omega}$ represents the sum of the N blocks of size $M \times M$ on the diagonal of

$$\boldsymbol{\Psi}(\boldsymbol{C})^{-1} + \boldsymbol{\Psi}(\boldsymbol{C})^{-1}\boldsymbol{R}^{-1}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\,\mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}^{*}_{\boldsymbol{h}|\boldsymbol{\gamma}}|i]\boldsymbol{R}^{-1}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\boldsymbol{\Psi}(\boldsymbol{C})^{-1}.$$

It is now straightforward to verify that the KKT conditions applied to $W_{\mathrm{q}}(\boldsymbol{\mathcal{C}}|i)$ and the power constraint $\|\boldsymbol{\mathcal{C}}\|^2_{\mathrm{F}} = PK\log_2(K)$ lead to the following system of equations

$$\frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{C}_k}(\boldsymbol{\mathcal{C}}) + \mu(\boldsymbol{C}_k)^{\mathrm{c}} = \boldsymbol{0}_{M\times L}, \quad k = 1, 2, \cdots, K \tag{6.38}$$

$$\sum_{k=1}^{K} \|\boldsymbol{C}_k\|^2_{\mathrm{F}} = PK\log_2(K), \tag{6.39}$$

where μ is a Lagrange multiplier. Adding all K equations in (6.38) and noting that $\sum_k \frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{C}_k}(\boldsymbol{\mathcal{C}}) = \boldsymbol{0}_{M\times L}$, due to the coefficients $(1-2\delta_{l'-k})$ in (6.36), gives

$$\mu \sum_{k=1}^{K} \boldsymbol{C}_k = \boldsymbol{0}_{M\times L}\,. \tag{6.40}$$

It is realized that μ must be non-zero, otherwise (6.38) and (6.39) imply that all K gradients $\frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{C}_k}(\boldsymbol{\mathcal{C}})$ are identically zero for a code that satisfies the power constraint. The latter is only possible if all codewords are all equal since it can easily be checked that for any other code $W_{\mathrm{q}}(\alpha\boldsymbol{\mathcal{C}}|i)$ is a strictly decreasing function of α, meaning that the gradient is non-zero. Clearly, the codewords cannot all be equal in an optimal code. Hence, μ is non-zero and (6.40) reduces to the necessary condition in (6.34), which all optimal codes must obey. □

Returning to the $K=2$ codeword case, the lemma is utilized for simplifying $\bar{W}_3(\boldsymbol{\mathcal{C}}|i)$ in (6.33). We solve for $\boldsymbol{C}_2$ in (6.34) and hence substitute $\boldsymbol{C}_2 = -\boldsymbol{C}_1$ into $\bar{W}_3(\boldsymbol{\mathcal{C}}|i)$ to arrive at the equivalent problem

$$\boldsymbol{A}_{12} = \arg \max_{\substack{\tilde{\boldsymbol{A}}_{12} \\ \mathrm{rank}(\tilde{\boldsymbol{A}}_{12})\le r \\ \mathrm{tr}\,\tilde{\boldsymbol{A}}_{12}=4P}} \det(\boldsymbol{I}_M + \boldsymbol{R}_{\mathrm{T}}\tilde{\boldsymbol{A}}_{12}/(4\sigma^2))\,, \tag{6.41}$$

where $r = \min\{M, L\}$ denotes the maximum possible rank of $\boldsymbol{A}_{12} = (\boldsymbol{C}_1 - \boldsymbol{C}_2)(\boldsymbol{C}_1 - \boldsymbol{C}_2)^* = 4\boldsymbol{C}_1\boldsymbol{C}_1^*$. Ignoring for a moment the rank constraint, this is a classical optimization problem that also occurs, for example, when computing the capacity of a set of parallel channels with colored Gaussian noise. It turns out that the rank constraint is easily handled within the scope of previous classical derivations (compare with how a thin transmit weighting matrix is handled in Section 4.7). Hence, with $4\sigma^2(\boldsymbol{R}_{\mathrm{T}})^{-1}$ here playing the role of the covariance of the colored noise, it follows from [CT91, p. 253-255] that the optimal solution is $\boldsymbol{A}_{12} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^*$, where $\boldsymbol{V}$ is an $M \times r$ matrix containing the r first columns of $\boldsymbol{U}$ and $\boldsymbol{D}$ is a diagonal $r \times r$ matrix with elements given by the water-filling solution

$$d_m = \max\{\nu - 4\sigma^2\lambda_m^{-1}, 0\}, \quad m = 1, 2, \cdots, r\,. \tag{6.42}$$

Here, ν is a Lagrange multiplier that can be determined from the power constraint. Thus, since $\boldsymbol{A}_{12} = 4\boldsymbol{C}_1\boldsymbol{C}_1^*$, the optimal codewords can be taken as

$$\boldsymbol{C}_1 = -\boldsymbol{C}_2 = 0.5\begin{bmatrix}\boldsymbol{V}\boldsymbol{D}^{1/2} & \boldsymbol{0}_{M\times(L-r)}\end{bmatrix}$$

and we conclude that the transmission power is distributed along the r first eigen-vectors of $\boldsymbol{R}_{\mathrm{T}}$.

A related water-filling result concerning the properties of linear dispersive codes in correlated fading scenarios is found in [SP01]. That result, as well as the solution of (6.41), can also be used to directly determine a semi closed-form expression for the optimal transmit weighting matrix in the weighted OSTBC design procedure given by (4.26). This since (4.26) reduces to an optimization problem of the same form as (6.41), when $\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\zeta}}$ is zero and $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\zeta}}$ has the same structure as the covariance matrix $\boldsymbol{R}_{\boldsymbol{hh}}$ considered herein.

Specializing the one-ring model by assuming non-informative fading statistics $\boldsymbol{R}_{\boldsymbol{hh}} = \boldsymbol{I}_N \otimes \sigma_h^2 \boldsymbol{I}_M = \sigma_h^2 \boldsymbol{I}_{MN}$ leads to the no channel knowledge case. Since the eigenvalues $\lambda_m = \sigma_h^2$ of $\boldsymbol{R}_{\mathrm{T}}$ then are all equal, it follows from (6.42) that $\boldsymbol{D}$ is also a scaled identity matrix and hence the power is distributed equally in all eigen-directions. Thus, for $M \leq L$ ($M \geq L$) the rows (columns) of $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ must be orthogonal and of equal magnitude. Another way of formulating this is to say that the singular values of each codeword are equal. In particular, note that $M = L$ means that the codewords are unitary matrices. How the statement generalizes for codes with more than two codewords is still an open issue. However, the experimental code construction results later presented in Section 6.6 indicate that the derived orthogonality property hold in many cases. The appealing properties of unitary codewords have also been noted in previous work such as [BV01], where it is shown that unitary codewords minimize the union bound of the codeword error probability under the assumption of an unknown channel at *both* the transmitter and receiver. This is in line with previous information theoretic based arguments in [MH99] for a similar non-coherent scenario.

6.4.4 High SNR

The high SNR case is defined as $P/\sigma^2 = 4P\eta \to \infty$. It is assumed in the following derivation that the codeword matrices are wide, i.e., $M \leq L$. In practice, this is not a particularly restrictive assumption since space-time codewords with shorter duration are unable to satisfy the *rank criterion* [GFBK99, TSC98] and hence cannot reach the maximum diversity order of the system.

Similar to as was done in the low SNR derivation, the optimization problem in (6.20) is re-parameterized using $\tilde{\boldsymbol{C}}_k \triangleq \boldsymbol{C}_k/\sqrt{P}$ and $\tilde{\boldsymbol{\mathcal{C}}} \triangleq \boldsymbol{\mathcal{C}}/\sqrt{P}$. Furthermore, the resulting power constraint $\|\tilde{\boldsymbol{\mathcal{C}}}\|_{\mathrm{F}}^2 = K \log_2(K)$

is augmented with an additional constraint

$$\tilde{\boldsymbol{A}}_{kl} \triangleq (\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l)(\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l)^* \succeq \varepsilon \boldsymbol{I}_M$$

to ensure that $\tilde{\boldsymbol{A}}_{kl}$ is invertible. Here, ε is a positive constant. It can be argued that, for sufficiently small ε, introducing the additional constraint does not change the asymptotically optimal solution. Hence, the properties of the asymptotically optimal code that will be derived next are valid also for the original optimization problem, see Appendix 4.A.1 concerning a similar approach. Based on the Taylor expansions

$$\begin{aligned}\operatorname{tr}\left((\boldsymbol{I}_N \otimes \tilde{\boldsymbol{A}}_{kl} P\eta + \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\boldsymbol{\gamma}})^{-1} \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} \operatorname{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}} \boldsymbol{m}^*_{\boldsymbol{h}|\boldsymbol{\gamma}}|i]^{-1} \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}\right) \\ = \mathrm{O}(P^{-1}\eta^{-1}\varepsilon^{-1}) \\ \log\det(\boldsymbol{I}_N \otimes \tilde{\boldsymbol{A}}_{kl} P\eta + \boldsymbol{R}^{-1}_{\boldsymbol{hh}|\boldsymbol{\gamma}}) = MN\log(P\eta) \\ + N\log\det(\tilde{\boldsymbol{A}}_{kl}) + \mathrm{O}(P^{-1}\eta^{-1}\varepsilon^{-1})\,,\end{aligned}$$

an equivalent criterion function can be expressed as

$$\begin{aligned}\tilde{W}_4(\tilde{\boldsymbol{\mathcal{C}}}|i) &\triangleq W_{\mathrm{q}}(\sqrt{P}\tilde{\boldsymbol{\mathcal{C}}}|i) e^{MN\log(P\eta)} \\ &= \sum_{k<l} e^{-N\log\det(\tilde{\boldsymbol{A}}_{kl}) + \mathrm{O}(P^{-1}\eta^{-1}\varepsilon^{-1})} \\ &= \sum_{k<l} e^{-N\log\det(\tilde{\boldsymbol{A}}_{kl})} + \mathrm{O}(P^{-1}\eta^{-1}\varepsilon^{-1})\,.\end{aligned}$$

Thus, letting $P/\sigma^2 = 4P\eta \to \infty$, while keeping ε constant and sufficiently small, we conclude that the code properties can be inferred from the limit function

$$\begin{aligned}\bar{W}_4(\tilde{\boldsymbol{\mathcal{C}}}|i) &\triangleq \lim_{P\eta\to\infty} \tilde{W}_4(\tilde{\boldsymbol{\mathcal{C}}}|i) \\ &= \sum_{k<l} \frac{1}{\det^N((\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l)(\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l)^*)}\,. \end{aligned} \tag{6.43}$$

It is seen that the differences $\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l$ must all be full-rank in the asymptotically optimal code. Otherwise, the determinant of $(\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l)(\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l)^*$ is zero for at least one codeword pair and the corresponding term in (6.43) would be infinitely large, implying that this cannot be the optimal solution. Consequently, performing the code design at a sufficiently high SNR ensures full-diversity space-time codes.

Note also that the criterion in (6.43) does not depend on neither the feedback information i nor the channel statistics. Thus, as the SNR increases the optimal code makes less and less use of channel knowledge and in the limit does not use it at all. A similar observation was made in connect with information theoretic results depicted in Figure 2.5. The similarities with the previous no channel knowledge case suggest that the optimal code tends to a no channel knowledge like code with codewords whose rows are often more or less orthogonal. This can be proved for the case of two codeword codes ($K = 2$) by arguing similarly as was done to solve (6.41).

6.4.5 Parameter Insensitivity of Some Codes

In general, the codes designed in this work depend on a number of parameters in addition to the explicit channel information contained in i. The power budget P, noise variance σ^2, covariance and cross-covariance matrices and the number of receive antennas N are all examples of such parameters. This dependence may be a drawback in situations were the parameters are not known, or are not known with great accuracy, since the performance typically degrades when the assumptions in the design process are different from those in the scenario in which the codes are used. It would therefore be beneficial to have codes that remain optimal even if the mentioned parameters are varied. Although the codes are in general not parameter independent, it is in this section made plausible that under certain circumstances the proposed code design process will generate codes that work well regardless of the values of the parameters. These theoretical findings are in line with code design and simulation results presented in later sections.

Below, we will show that, for the simplified fading scenario with no channel knowledge, a code $\mathcal{C} \triangleq \{\sqrt{P/\tilde{P}}\tilde{\boldsymbol{C}}_k\}$ satisfies the KKT conditions (6.38) - (6.39), regardless of the values of σ^2, σ_h^2, P and N, if the codewords $\{\tilde{\boldsymbol{C}}_k\}$ are such that they satisfy the KKT conditions for some parameters $\sigma^2 = \tilde{\sigma}^2$, $\sigma_h^2 = \tilde{\sigma}_h^2$, $P = \tilde{P}$, $N = \tilde{N}$ and in addition obey the orthogonality property

$$\tilde{\boldsymbol{A}}_{kl} \triangleq (\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l)(\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l)^* = \tilde{\lambda}_{kl}\boldsymbol{I}_M, \quad k \neq l\,, \tag{6.44}$$

with $\tilde{\lambda}_{kl} = \tilde{\lambda}$ constant over all codeword pairs. As shown in Section 6.6, the orthogonality property in (6.44) is a characteristic of one of the con-

structed codes. Note that the well-known OSTB codes satisfy the orthogonality property in (6.44), albeit not with a constant $\tilde{\lambda}_{kl}$.

Since the KKT conditions are only necessary and not sufficient for optimality, this parameter insensitivity result merely provides an indication that a scaled version of an optimal code $\{\tilde{\boldsymbol{C}}_k\}$ remains optimal even if the parameters are changed. However, numerous experimental investigations of designed codes that satisfy the above conditions seem to support the claim.

The proof of the parameter insensitivity result proceeds as follows. Since $\{\tilde{\boldsymbol{C}}_k\}$ satisfies the KKT conditions and (6.44) for some $\tilde{\lambda} = \lambda_{kl}$, consider (6.38) - (6.39) and the expression for the gradient in (6.37) and tailor them for the problem at hand to arrive at

$$\begin{aligned} \boldsymbol{G}_k(\tilde{\boldsymbol{\mathcal{C}}}, \tilde{\sigma}^2, \tilde{\sigma}_h^2, \tilde{P}, \tilde{N}) + \tilde{\mu}(\tilde{\boldsymbol{C}}_k)^{\mathrm{c}} = \boldsymbol{0}_{M\times L},\ \forall k \\ \sum_k \|\tilde{\boldsymbol{C}}_k\|_{\mathrm{F}}^2 = \tilde{P}K\log_2(K)\,, \end{aligned} \tag{6.45}$$

where $\tilde{\mu}$ is a Lagrange multiplier and, since

$$\begin{aligned} \boldsymbol{V}'_{\mathrm{q}}(\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l) &= \frac{-\tilde{\eta}(\boldsymbol{\Omega}(\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l))^{\mathrm{c}}}{\det(\boldsymbol{I}_N \otimes \tilde{\boldsymbol{A}}_{kl}\tilde{\eta} + \tilde{\sigma}_h^{-2}\boldsymbol{I}_{MN})} \\ &= \frac{-\tilde{N}(\tilde{\boldsymbol{C}}_k - \tilde{\boldsymbol{C}}_l)^{\mathrm{c}}}{4\tilde{\sigma}^2(\tilde{\lambda}/(4\tilde{\sigma}^2) + \tilde{\sigma}_h^{-2})^{MN+1}}\,, \end{aligned}$$

the gradient is given by

$$\begin{aligned} &\boldsymbol{G}_k(\tilde{\boldsymbol{\mathcal{C}}}, \tilde{\sigma}^2, \tilde{\sigma}_h^2, \tilde{P}, \tilde{N}) \triangleq \frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{C}_k}(\tilde{\boldsymbol{\mathcal{C}}}) \\ &= -\frac{\tilde{N}}{4\tilde{\sigma}^2(\tilde{\lambda}/(4\tilde{\sigma}^2) + \tilde{\sigma}_h^{-2})^{M\tilde{N}+1}} \sum_{\substack{k'<l' \\ k'=k \text{ or } l'=k}} (1 - 2\delta_{l'-k})(\tilde{\boldsymbol{C}}_{k'} - \tilde{\boldsymbol{C}}_{l'})^{\mathrm{c}}\,. \end{aligned}$$

After noting that changing the parameters only re-scales the gradient $\boldsymbol{G}_k(\tilde{\boldsymbol{\mathcal{C}}}, \tilde{\sigma}^2, \tilde{\sigma}_h^2, \tilde{P}, \tilde{N})$, it is clear that, because

$$\boldsymbol{G}_k\Big(\frac{\sqrt{P}}{\sqrt{\tilde{P}}}\tilde{\boldsymbol{\mathcal{C}}}, \sigma^2, \sigma_h^2, P, N\Big) = \alpha \boldsymbol{G}_k(\tilde{\boldsymbol{\mathcal{C}}}, \tilde{\sigma}^2, \tilde{\sigma}_h^2, \tilde{P}, \tilde{N})\,,$$

where

$$\alpha \triangleq \frac{\tilde{\sigma}^2 N\sqrt{P/\tilde{P}}(\tilde{\lambda}/(4\tilde{\sigma}^2) + \tilde{\sigma}_h^{-2})^{MN+1}}{\sigma^2\tilde{N}(\sqrt{P/\tilde{P}}\tilde{\lambda}/(4\sigma^2) + \sigma_h^{-2})^{MN+1}}\,,$$

the code $\{\sqrt{P/\tilde{P}}\tilde{\boldsymbol{C}}_k\}$ and the Lagrange multiplier $\mu = \alpha\tilde{\mu}$ simply scale the left hand side of (6.45) and hence the KKT conditions are satisfied. This concludes the proof.

6.4.6 A Symmetric Feedback Scenario

It often makes sense to design the feedback link so that the quantization is, in some sense, symmetric. For example, in the commonly assumed Rayleigh fading scenario, the phase of each channel coefficient is uniformly distributed and hence it is reasonable to use an encoder that quantizes the phase uniformly over the phase interval $[0, 2\pi)$. This motivates the partial phase combining [HP98] type of method of this work. The feedback method in the closed-loop mode of the WCDMA system [3GP02b] is another example of uniform phase quantization. The encoder regions of these quantization schemes exhibit symmetries in the sense that they are rotated copies of each other. This section investigates how similar rotational symmetries affect the code design and how they can be exploited for reducing the computational complexity of the design method. A similar concept of symmetric feedback was introduced in Section 2.6.3 for simplifying numerical evaluation of a channel capacity expression.

Recall that a unitary matrix may correspond to a rotation. With this in mind, we say that the feedback scenario is symmetric if $\mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*|i]$ and $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}$ can be written on the form

$$\mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*|i] = \tilde{\boldsymbol{D}} \otimes \boldsymbol{Q}_i\boldsymbol{R}\boldsymbol{Q}_i^* \tag{6.46}$$

$$\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}} = \boldsymbol{D} \otimes \boldsymbol{I}_M\,, \tag{6.47}$$

where $\boldsymbol{Q}_i$ is a unitary $M \times M$ matrix that depends on i, $\boldsymbol{R}$ is a constant $M \times M$ matrix, $\tilde{\boldsymbol{D}} \triangleq \mathrm{diag}(\tilde{d}_1, \tilde{d}_2, \cdots, \tilde{d}_N)$ and $\boldsymbol{D} \triangleq \mathrm{diag}(d_1, d_2, \cdots, d_N)$ for some arbitrary $\{\tilde{d}_n\}_{n=1}^N$ and $\{d_n\}_{n=1}^N$, respectively.

The above conditions affect the code design as follows. By using the Kronecker relations (C.3) and (C.4) in Appendix C, repeated below for convenience

$$(\boldsymbol{X}_1 \otimes \boldsymbol{X}_2)(\boldsymbol{X}_3 \otimes \boldsymbol{X}_4) = \boldsymbol{X}_1\boldsymbol{X}_3 \otimes \boldsymbol{X}_2\boldsymbol{X}_4$$
$$(\boldsymbol{X}_1 \otimes \boldsymbol{X}_2)^{-1} = \boldsymbol{X}_1^{-1} \otimes \boldsymbol{X}_2^{-1}\,,$$

and assuming (6.46), (6.47) hold, it is possible to write

$$\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1}\,\mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*|i]\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1} = \tilde{\boldsymbol{D}}\boldsymbol{D}^{-2} \otimes \boldsymbol{Q}_i\boldsymbol{R}\boldsymbol{Q}_i^*\,,$$

which inserted into (6.18) leads to the re-parameterized criterion function

$$W_{\mathrm{q}}(\boldsymbol{Q}_i\bar{\boldsymbol{\mathcal{C}}}|i) = \sum_{k<l} \frac{e^{\mathrm{tr}((\boldsymbol{I}_N\otimes\bar{\boldsymbol{A}}_{kl}\eta+\boldsymbol{D}^{-1}\otimes\boldsymbol{I}_M)^{-1}(\tilde{\boldsymbol{D}}\boldsymbol{D}^{-2}\otimes\boldsymbol{R}))}}{\det(\boldsymbol{I}_N\otimes\bar{\boldsymbol{A}}_{kl}\eta+\boldsymbol{D}^{-1}\otimes\boldsymbol{I}_M)} \tag{6.48}$$

and power constraint $\|\bar{\boldsymbol{\mathcal{C}}}\|_{\mathrm{F}}^2 = PK\log_2(K)$, where $\bar{\boldsymbol{\mathcal{C}}} \triangleq \boldsymbol{Q}_i^*\boldsymbol{\mathcal{C}}$, $\bar{\boldsymbol{C}}_k \triangleq \boldsymbol{Q}_i^*\boldsymbol{C}_k$, $\bar{\boldsymbol{A}}_{kl} \triangleq (\bar{\boldsymbol{C}}_k-\bar{\boldsymbol{C}}_l)(\bar{\boldsymbol{C}}_k-\bar{\boldsymbol{C}}_l)^*$.

Clearly, (6.48) is independent of i and the original code design problem may therefore be solved by first minimizing (6.48) with respect to $\{\bar{\boldsymbol{C}}_k\}$, subject to the power constraint, and then extracting the ith code through the linear mapping $\boldsymbol{\mathcal{C}}^{(i)} = \boldsymbol{Q}_i\bar{\boldsymbol{\mathcal{C}}}$, where $\bar{\boldsymbol{\mathcal{C}}}$ now represents the optimal solution in the re-parameterized problem. Hence, the inherent symmetry of the feedback carries over to a corresponding symmetry of the codes. Because all codes are easily derived by means of a linear transformation from the single code in $\bar{\boldsymbol{\mathcal{C}}}$, this result can be utilized to substantially expedite the design process and reduce the memory required to store the code.

Symmetric Feedback in the Simplified Fading Scenario

As an illustrative example of a symmetric feedback situation, consider the feedback scheme in this work operating in the simplified fading scenario with $N = 1$ receive antenna. Since the relative phases are uniformly quantized, the codebook vectors in (6.5) that the encoder uses can be expressed in terms of the first codebook vector as $\hat{\boldsymbol{\gamma}}(i) = \boldsymbol{Q}_i\hat{\boldsymbol{\gamma}}(0)$, $i = 0,\cdots,2^b-1$, where

$$\boldsymbol{Q}_i \triangleq \mathrm{diag}(1, e^{j\phi_{i_0(i)}}, e^{j\phi_{i_1(i)}}, \cdots, e^{j\phi_{i_{MN-2}(i)}})$$

rotates $\hat{\boldsymbol{\gamma}}(0)$ to the desired position. To see that also the encoder regions are rotations of each other, let $\boldsymbol{\gamma}$ belong to encoder region $\mathcal{S}_0$ and note that

$$\begin{aligned}\varepsilon(\boldsymbol{Q}_k\boldsymbol{\gamma}) &= \arg\min_i \|\boldsymbol{Q}_k\boldsymbol{\gamma}/\gamma_1 - \boldsymbol{Q}_i\hat{\boldsymbol{\gamma}}(0)\|^2 \\ &= \arg\min_i \|\boldsymbol{\gamma}/\gamma_1 - \boldsymbol{Q}_k^*\boldsymbol{Q}_i\hat{\boldsymbol{\gamma}}(0)\|^2 = k\,,\end{aligned}$$

since $\boldsymbol{Q}_k^*\boldsymbol{Q}_i \in \{\boldsymbol{Q}_i\}$ and $\boldsymbol{Q}_k^*\boldsymbol{Q}_k = \boldsymbol{I}_M = \boldsymbol{Q}_0$. This means that for each point $\boldsymbol{\gamma}$ in $\mathcal{S}_0$ there is a corresponding rotated point $\boldsymbol{Q}_k\boldsymbol{\gamma}$ in $\mathcal{S}_k$ (the converse can also be established in a similar manner, compare with Appendix 2.D.3) and we conclude that $\{\mathcal{S}_i\}$ are rotated versions

of $\mathcal{S}_0$. Together with the zero-mean IID complex Gaussian distribution's invariance to a unitary linear mapping, i.e., the PDF of $\boldsymbol{\gamma}$ has the property $p_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}) = p_{\boldsymbol{\gamma}}(\boldsymbol{Q}_i\boldsymbol{\gamma})$, it follows from similar derivations as in Appendix 2.D.1 that the conditional PDF of $\boldsymbol{\gamma}|i$ satisfies the following relation $p_{\boldsymbol{\gamma}|i}(\boldsymbol{\gamma}|0) = p_{\boldsymbol{\gamma}|i}(\boldsymbol{Q}_i\boldsymbol{\gamma}|i)$. Using this relation it is readily shown that

$$\begin{aligned}\mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^*|i] &= |\tilde{\rho}|^2\,\mathrm{E}[\boldsymbol{\gamma}\boldsymbol{\gamma}^*|i] = \rho^2\boldsymbol{Q}_i\,\mathrm{E}[\boldsymbol{\gamma}\boldsymbol{\gamma}^*|0]\boldsymbol{Q}_i^* \\ &= \boldsymbol{Q}_i\boldsymbol{R}\boldsymbol{Q}_i^*\,,\end{aligned}$$

with $\boldsymbol{R} = \rho^2\,\mathrm{E}[\boldsymbol{\gamma}\boldsymbol{\gamma}^*|0]$ and, consequently, the first symmetric feedback condition in (6.46) is satisfied. The second symmetric feedback condition (6.47) is also satisfied since for the simplified fading scenario it follows from (6.7) that $\boldsymbol{R}_{\boldsymbol{h}\boldsymbol{h}|\boldsymbol{\gamma}} = \sigma_h^2(1-|\tilde{\rho}|^2)\boldsymbol{I}_{MN} = \sigma_h^2(1-\rho^2)\boldsymbol{I}_N \otimes \boldsymbol{I}_M$. Thus, we have shown that the partial phase combining type of feedback scheme used in the simplified fading scenario indeed constitutes a symmetric feedback situation and hence all the 2^b different codes can be derived from a single code $\{\bar{\boldsymbol{C}}_k\}$ by means of known linear transformations. This will be utilized later, both in Section 6.6 to describe constructed codes in a compact manner as well as for simulation purposes in Section 6.7.

6.5 Numerical Optimization

The design procedures for unstructured and linear dispersive codes described in Section 6.3.2 represent non-convex optimization problems, which are in general difficult to solve. The present section describes how numerical gradient search [BS93] type of techniques can be utilized for "solving" these optimization problems. With such an approach, the solution can only be expected to be locally optimal. To remedy this, the chance of finding the global optimum is improved by repeating the gradient search several times, each time initiated differently, and picking the best of the resulting solutions. Although the technique is tailored to the application under consideration, it is similar in structure to the optimization procedure utilized in [FGW74] for the design of efficient signal constellations.

6.5.1 Unstructured Codes

The numerical optimization procedure for unstructured codes implements a gradient search which updates the code $\boldsymbol{\mathcal{C}}$ by taking a step along the

direction of the negative gradient of $W_{\mathrm{q}}(\boldsymbol{\mathcal{C}}|i)$. The result is thereafter normalized so as to satisfy the output power constraint of the codeword constellation. Starting with an initial guess $\boldsymbol{\mathcal{C}}_0$ of the solution, the codeword constellation is hence updated, for $n = 0, 1, \cdots$, according to

$$\boldsymbol{\mathcal{C}}_{n+1} = \frac{\sqrt{PK\log_2(K)}\left(\boldsymbol{\mathcal{C}}_n - \mu_n\left(\frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{\mathcal{C}}}(\boldsymbol{\mathcal{C}}_n)\right)^{\mathrm{c}}\right)}{\|\boldsymbol{\mathcal{C}}_n - \mu_n\left(\frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{\mathcal{C}}}(\boldsymbol{\mathcal{C}}_n)\right)^{\mathrm{c}}\|_{\mathrm{F}}}, \tag{6.49}$$

where μ_n is the so-called step size and, based on (6.36), the derivatives with respect to all the codewords have been collected in a matrix representing the gradient as

$$\frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{\mathcal{C}}}(\boldsymbol{\mathcal{C}}) = \left[\frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{C}_1}(\boldsymbol{\mathcal{C}}) \quad \frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{C}_2}(\boldsymbol{\mathcal{C}}) \quad \cdots \quad \frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{C}_K}(\boldsymbol{\mathcal{C}})\right].$$

The complex conjugate in (6.49) is due to the negative sign of the imaginary part in the definition of the complex derivative in (6.35).

An appropriate step size is obtained by using successively smaller values of μ_n until the design criterion is found to have decreased. Several alternative methods for choosing the step size exist in the literature [BS93]. A candidate solution is found by iterating (6.49) until convergence.

The gradient search is repeated several times using different $\boldsymbol{\mathcal{C}}_0$ set at random. Finally, the best of the resulting candidate solutions is chosen and designated as the constructed code $\boldsymbol{\mathcal{C}}^{(i)}$. The random initialization strategy can be complemented in various ways. For example, a $\boldsymbol{\mathcal{C}}_0$ corresponding to an OSTB code or beamforming could also be used. In this way, the constructed code is guaranteed to perform equal or better (as measured by the criterion function) than either of these two well-known transmission methods.

6.5.2 Linear Dispersive Codes

It should be clear that a similar numerical optimization technique can be used to "solve" the design problem for linear dispersive codes in (6.21). The major difference compared with the previous gradient search in (6.49) is that the gradient is now instead evaluated with respect to the transmit weights in $\boldsymbol{\mathcal{B}}$.

To explicitly describe the gradient search in the present case, let $\frac{\partial W_{\mathrm{q}}}{\partial \boldsymbol{B}_m}(\boldsymbol{\mathcal{B}})$ denote the derivative of $W_{\mathrm{q}}(\boldsymbol{\mathcal{C}}(\boldsymbol{\mathcal{B}})|i)$ with respect to the mth

weighting matrix $\boldsymbol{B}_m$, evaluated at $\boldsymbol{\mathcal{B}}$. Since $\boldsymbol{C}_k = \sum_{m=1}^{L_\mathrm{d}} \boldsymbol{B}_m s_m^{(k)}$, the derivative is easily inferred from (6.37). Hence, it is realized that

$$\frac{\partial W_\mathrm{q}}{\partial \boldsymbol{B}_m}(\boldsymbol{\mathcal{B}}) = \sum_{k<l} (s_m^{(k)} - s_m^{(l)}) \boldsymbol{V}_\mathrm{q}' \left(\sum_{m'=1}^{L_\mathrm{d}} \boldsymbol{B}_{m'} (s_{m'}^{(k)} - s_{m'}^{(l)}) \right) .$$

Collect now all the L_d derivatives and form the gradient as

$$\frac{\partial W_\mathrm{q}}{\partial \boldsymbol{\mathcal{B}}}(\boldsymbol{\mathcal{B}}) = \left[\frac{\partial W_\mathrm{q}}{\partial \boldsymbol{B}_1}(\boldsymbol{\mathcal{B}}) \quad \frac{\partial W_\mathrm{q}}{\partial \boldsymbol{B}_2}(\boldsymbol{\mathcal{B}}) \quad \cdots \quad \frac{\partial W_\mathrm{q}}{\partial \boldsymbol{B}_{L_\mathrm{d}}}(\boldsymbol{\mathcal{B}}) \right] .$$

Similarly to as in the case of unstructured codes, the gradient search starts with an initial guess $\boldsymbol{\mathcal{B}}_0$ of the solution and updates the weighting matrices according to

$$\boldsymbol{\mathcal{B}}_{n+1} = \frac{\sqrt{P \log_2(K)} \left(\boldsymbol{\mathcal{B}}_n - \mu_n \left(\frac{\partial W_\mathrm{q}}{\partial \boldsymbol{\mathcal{B}}}(\boldsymbol{\mathcal{B}}_n) \right)^\mathrm{c} \right)}{\| \boldsymbol{\mathcal{B}}_n - \mu_n \left(\frac{\partial W_\mathrm{q}}{\partial \boldsymbol{\mathcal{B}}}(\boldsymbol{\mathcal{B}}_n) \right)^\mathrm{c} \|_\mathrm{F}} , \tag{6.50}$$

until convergence.

6.5.3 Computational Complexity Issues

It is clear that the numerical optimization may be performed off-line. Thus, although the computing power needed for the codeword search is substantial, the resulting transmission scheme can be efficiently implemented via the use of lookup tables.

Consider first the case of unstructured codes. The time it takes to complete the code design increases with the size of the problem. For modest values of M, L and K, a solution is obtained within a reasonable short time period. The main bottleneck is typically the computation of the design criterion and gradient when the number of codewords K is large. Since the number of terms in the design criterion and gradient grows quadratically with K (c.f. (6.28)), the time it takes to perform the update in (6.49) quickly becomes prohibitive, resulting in an essentially unsolvable code design problem. Nevertheless, we have managed to construct $M = 8$, $L = 8$ unstructured codes with up to $K = 4096$ codewords using the techniques presented in this chapter. Of course, the design of codes with codewords of smaller dimension and at least this many codewords is also feasible.

Both high and low rate codes are possible, depending on the length L of the codewords. For an $M = 2$, $L = 2$ code and assuming a maximum of 4096 codewords, the rate may be as high as $\log_2(4096)/2 = 6$ bits per channel use while for the $M = 8$, $L = 8$ case a more moderate rate of $\log_2(4096)/8 = 1.5$ bits per channel use is possible. To obtain much higher rates it may be necessary to use other design methods that do not rely on pairwise error probabilities and possibly also reduce the number of parameters in the problem by introducing additional constraints/structure. The use of linear dispersive codes is one example of how structure may be introduced to drastically reduce the number of parameters that need to be optimized. However, the benefits of structure typically come at the price of lower performance since structure generally limits the degrees of freedom in the design.

The problems associated with the rapid growth of the computational complexity with respect to K may also be circumvented by employing linear dispersive codes. A viable strategy is then to design the weighting matrices so as to maximize the information theoretic capacity of a suitable channel. Similarly as was done in [HH02b] for the no channel knowledge case, the well-known MIMO capacity formula [Tel95], or the related expression in (2.7), can be used with channel information incorporated into the design through the statistics of the channel and feedback (c.f. mean and covariance feedback [VM01]).

It should be noted that such criteria are only appropriate for long block lengths with the transmitted codeword experiencing many different channel realizations, the latter because the relevant capacity formulas rely on the assumption of ergodic fading. In other words, code design methods based on these criteria are geared toward applications which can tolerate a rather high processing delay. Our design procedures, on the other hand, are intended for significantly shorter block lengths, thus targeting delay sensitive applications, and do not assume coding over different channel realizations. Furthermore, a fundamental assumption behind capacity maximizing code construction methods as in [HH02b] is that the designed code is concatenated with an outer code to obtain low error probability. In contrast, the code design methods proposed herein strive for minimizing the error probability directly.

Another important issue is decoding complexity. As evident from (6.2), the unstructured codes designed in this work require an exhaustive search over all codewords. The resulting computational complexity may be tolerable for codes with a moderate number of codewords, but for high rate applications the high decoding complexity essentially excludes

the use of unstructured codes. Linear dispersive codes may then be a more appealing choice since they permit, as pointed out in Section 3.2.2 the use of decoding algorithms [VH02, JMO03] with significantly lower complexity than an exhaustive search.

Obviously, weighted OSTBC is an even better choice if low computational complexity is the only issue under consideration. This applies both to the transmit weight design procedure as well as to the ML detection algorithm.

6.6 Code Design Results

The design method described by (6.20) and implemented as (6.49) has been used for constructing numerous channel side information dependent unstructured codes. This section presents a few illustrative examples of such constructed codes and their properties. In addition, properties of a couple of linear dispersive codes, designed using (6.21), are also presented. All codes presented in this section were designed under the assumption of the simplified fading scenario with $N = 1$ receive antenna, a power budget $P = 1$ and fading variance $\sigma_h^2 = 1$.

A Channel Side Information Dependent Unstructured Code

In Table 6.1, three unstructured $M = 2$, $L = 2$, $K = 4$ codes are presented. The codes have been designed for a noise variance $\sigma^2 = 0.1$, corresponding to an SNR of 10 dB in the graph depicted in Figure 6.2, and using an initial channel information quality ρ set to 0, 0.95 and 0.9999, respectively. In the two latter cases, $b = 2$ bits were used to quantize the initial channel information using the phase quantization scheme described in Section 6.2.1. Only the codewords of the first code, $\mathcal{C}(0)$, are explicitly displayed. This constitutes all needed codes in the no channel knowledge $\rho = 0$ case. For the $\rho = 0.95$ and $\rho = 0.9999$ cases, the feedback is symmetric, as explained in Section 6.4.6, and the remaining codes $\mathcal{C}(i)$, $i = 1, 2, 3$ are hence obtained through the linear transformations $\boldsymbol{C}_k^{(i)} = \boldsymbol{Q}_i \boldsymbol{C}_k^{(0)}$, where $\boldsymbol{Q}_i \triangleq \mathrm{diag}(1, e^{j\pi i/2})$. All the codes in the table together with their linearly transformed counterparts can be said to form one channel side information dependent code.

The condition number $\kappa(\boldsymbol{C}_k^{(i)})$ (with respect to the spectral norm) of each codeword $\boldsymbol{C}_k^{(i)}$ has been computed and the results are summarized and displayed above the corresponding code. Since the condition number

	$\rho = 0 \quad \Rightarrow \quad \kappa(\boldsymbol{C}_k^{(0)}) \approx 1$	
$\boldsymbol{C}_1^{(0)}$	$-0.71126 + j0.36087$ $-0.58663 + j0.14285$	$+0.40735 - j0.44435$ $-0.65934 + j0.44859$
$\boldsymbol{C}_2^{(0)}$	$+0.01832 + j0.21581$ $+0.81098 + j0.54269$	$+0.08399 + j0.97295$ $-0.18061 - j0.12080$
$\boldsymbol{C}_3^{(0)}$	$+0.15579 + j0.09515$ $+0.20541 - j0.96129$	$-0.94038 - j0.28723$ $-0.00773 - j0.18309$
$\boldsymbol{C}_4^{(0)}$	$+0.53715 - j0.67183$ $-0.42976 + j0.27575$	$+0.44904 - j0.24136$ $+0.84767 - j0.14470$
	$\rho = 0.95 \quad \Rightarrow \quad \kappa(\boldsymbol{C}_k^{(i)}) \approx 6$	
$\boldsymbol{C}_1^{(0)}$	$-0.92111 + j0.37685$ $-0.88850 + j0.32513$	$-0.04889 - j0.10390$ $-0.30045 + j0.10687$
$\boldsymbol{C}_2^{(0)}$	$+0.48440 + j0.49155$ $+0.67026 + j0.56673$	$-0.03616 + j0.72504$ $-0.09747 + j0.46532$
$\boldsymbol{C}_3^{(0)}$	$+0.24692 - j0.47884$ $+0.25685 - j0.72680$	$-0.77306 - j0.34031$ $-0.55128 - j0.31430$
$\boldsymbol{C}_4^{(0)}$	$+0.18980 - j0.38956$ $-0.03860 - j0.16506$	$+0.85812 - j0.28083$ $+0.94920 - j0.25788$
	$\rho = 0.9999 \quad \Rightarrow \quad \kappa(\boldsymbol{C}_k^{(i)}) > 10^4$	
$\boldsymbol{C}_1^{(0)}$	$-0.41928 + j0.80405$ $-0.41680 + j0.80166$	$+0.24074 - j0.35123$ $+0.23961 - j0.35014$
$\boldsymbol{C}_2^{(0)}$	$+0.40759 - j0.35120$ $+0.40569 - j0.35052$	$-0.50008 - j0.68115$ $-0.49916 - j0.67815$
$\boldsymbol{C}_3^{(0)}$	$-0.33466 - j0.68957$ $-0.33429 - j0.68668$	$+0.61490 + j0.19504$ $+0.61289 + j0.19353$
$\boldsymbol{C}_4^{(0)}$	$+0.34635 + j0.23672$ $+0.34539 + j0.23554$	$-0.35556 + j0.83734$ $-0.35334 + j0.83476$

Table 6.1: Constructed $\mathcal{C}(0)$ codes for different ρ. Design parameters: $M = 2$, $L = 2$, $N = 1$, $K = 4$, $b = 2$, $P = 1$, $\sigma^2 = 0.1$.

is the ratio between the maximum and minimum singular value [HJ96, p. 442], it serves as a measure of how close the codewords are to unitary and rank one matrices (the latter because there are only two singular values), respectively.

The condition number is seen to increase with ρ. In particular, since $\kappa(\boldsymbol{C}_k^{(0)}) \approx 1$ the codewords in the $\rho = 0$ code are all essentially scaled unitary matrices, in line with the $K = 2$ no channel knowledge result derived in Section 6.4.3. Correspondingly, the almost perfect initial channel knowledge $\rho = 0.9999$ code has codewords with very high condition numbers, meaning that the codewords are close to rank one. This is explained by studying the eigenvalues of $\boldsymbol{R}_i = \mathrm{E}[\bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\gamma}} \bar{\boldsymbol{M}}_{\boldsymbol{H}|\boldsymbol{\gamma}} | i]$ in (6.25). It can be verified that the difference between the two eigenvalues is large and hence almost all power is allocated along the strongest eigen-direction, in good agreement with the discussion related to perfect initial channel information in Section 6.4.1.

Properties in the No Channel Knowledge Case

No channel knowledge constitutes an important special case that warrants further investigation. Table 6.2 presents some interesting properties of a few constructed codes and, as benchmarks, corresponding rate one OSTB codes as found in [TJC99]. This time, both unstructured and linear dispersive codes were designed. Because of the no channel knowledge assumption, only the determinant term of $\ell(\boldsymbol{C}_k^{(i)}, \boldsymbol{C}_l^{(i)}|\boldsymbol{\gamma})$ in (6.15) remains, implying that $P_{\mathrm{LBUB}}(\{\mathcal{C}(i)\}) = P_{\mathrm{UB}}(\{\mathcal{C}(i)\})$ in (6.16). Hence, both design procedures now strive for minimizing an upper bound on the union bound of the codeword error probability.

A BPSK signal constellation was used for the constituent information symbols in each OSTB and linear dispersive code. Every row represents a code with the first column displaying dimensions and the type of code, "U", "L" and "O" stand for unstructured, linear dispersive and OSTB code, respectively. In the second column, the spread of energy over the codewords is measured using the ratio

$$\varepsilon_C^{\max}/\varepsilon_C^{\min} \triangleq \max_k \|\boldsymbol{C}_k\|_{\mathrm{F}}^2 / \min_k \|\boldsymbol{C}_k\|_{\mathrm{F}}^2 \,.$$

To measure how close the codewords are to unitary matrices, the third column shows the maximum codeword condition number

$$\kappa_C \triangleq \max_k \kappa(\boldsymbol{C}_k)$$

of each code. Similarly, in the fourth and fifth column,

$$\varepsilon_A^{\max}/\varepsilon_A^{\min} \triangleq \max_{k\neq l} \|\boldsymbol{A}_{kl}\|_{\mathrm{F}}^2 / \min_{k\neq l} \|\boldsymbol{A}_{kl}\|_{\mathrm{F}}^2, \quad \kappa_A \triangleq \max_{k\neq l} \kappa(\boldsymbol{A}_{kl}) \,,$$

where we recall that $\boldsymbol{A}_{kl} = (\boldsymbol{C}_k - \boldsymbol{C}_l)(\boldsymbol{C}_k - \boldsymbol{C}_l)^*$. Together these two measures show whether the code is parameter insensitive as defined by the orthogonality property in (6.44). Finally, to facilitate a comparison of the performance at high SNR values, the last column displays the coding gain

$$\mathrm{CG} \triangleq \left(\sum_{k<l} \frac{1}{\det^N((\boldsymbol{C}_k - \boldsymbol{C}_l)(\boldsymbol{C}_k - \boldsymbol{C}_l)^*)} \right)^{-1/(MN)}$$

for each code. The definition of the coding gain stems from the fact that at high SNR values the upper bound on the codeword error probability in (6.32) can, for the no channel knowledge case and assuming full diversity codes that satisfy the rank criterion [TSC98], be approximated

(M, L, K)	$\varepsilon_C^{\max}/\varepsilon_C^{\min}$	κ_C	$\varepsilon_A^{\max}/\varepsilon_A^{\min}$	κ_A	CG[dB]
U: $(2, 2, 4)$	1.000	1.001	1.000	1.004	0.369
U: $(4, 4, 16)$	1.084	1.002	3.297	4.036	-2.47
U: $(8, 8, 256)$	1.130	1.136	11.41	502.2	-4.58
L: $(2, 2, 4)$	1.000	1.000	4.000	1.004	-0.256
L: $(4, 4, 16)$	1.000	1.006	16.00	1.013	-3.87
L: $(8, 8, 256)$	1.000	1.006	64.00	1.018	-6.78
O: $(2, 2, 4)$	1.000	1.000	4.000	1.000	-0.256
O: $(4, 4, 16)$	1.000	1.000	16.00	1.000	-3.87
O: $(8, 8, 256)$	1.000	1.000	64.00	1.000	-6.78

Table 6.2: Code properties for the no channel knowledge case.

by $\alpha/(\mathrm{CG} \cdot \mathrm{SNR})^{MN}$ for some constant α (c.f. [TSC98]). To roughly optimize the coding gain, all unstructured codes were designed with a high SNR scenario in mind using a small noise variance $\sigma^2 = 0.001$.

Observe first that the linear dispersive and OSTB codes are essentially equal in terms of the measured properties. In particular they have the same coding gain. This is no coincidence. A closer study of the weighting matrices in the constructed linear dispersive codes shows that the matrices satisfy (with great accuracy) the conditions in (3.11) and (3.12) that define the notion of an OSTB code. Experimental investigations indicate that the conditions are satisfied also for other codeword sizes and code rates than the ones presented in the table, as long as there exists an OSTB code for the given rate. In other words, our design procedure seems to automatically produce OSTB codes. This is quite a remarkable result since such codes have so far been handcrafted. The fact that OSTB codes are obtained indicates that they are optimal in the sense of minimizing, within the class of linear dispersive codes, the previously mentioned upper bound on the codeword error probability. Theoretical analyses regarding the optimality of OSTB codes within more restrictive classes of linear dispersive codes than considered herein point in the same direction [SHP00, SP01].

The unstructured and OSTB codes also have a lot in common. For example, from the ratio $\varepsilon_C^{\max}/\varepsilon_C^{\min}$ we see that for both code types the energy is distributed more or less equally among the codewords and, since κ_C is close to one, the codewords are essentially scaled unitary matrices. Hence, the no channel knowledge orthogonality property that was derived for a $K = 2$ codeword code seems to hold for other K as

well. For the unstructured $(2, 2, 4)$ code, $\kappa_A \approx 1$, which means that there are constructed codes that even share the pairwise orthogonality property (6.44) with the OSTB codes. Moreover, since $\varepsilon_A^{\max}/\varepsilon_A^{\min} \approx 1$, the scaling factor $\tilde{\lambda}_{kl}$ is almost constant. The code therefore satisfies all the requirements for the parameter insensitivity result to apply, which provides an indication that the code will perform well even if σ^2, σ_h^2, P and N are changed. Tests conducted for various parameter values show that this is indeed the case since the unstructured $(2, 2, 4)$ code gives approximately the same values of the criterion function as the solutions obtained by performing the numerical optimization procedure. From the parameter insensitivity it follows that the codewords can be considered to be the same as for the $\rho = 0$ code in Table 6.1.

Performance wise, however, the unstructured and OSTB codes differ significantly, even in the present case of no channel knowledge. The former are seen to have consistently higher coding gains than the latter. For example, the unstructured $(2, 2, 4)$ code is 0.6 dB better than the corresponding OSTB code while the difference is 1.4 dB for the $(4, 4, 16)$ codes. These numbers are also approximately confirmed by simulation results in Section 6.7. Of course, the performance advantage exhibited by the designed unstructured codes is to be expected considering the design freedom offered by the absence of structural constraints.

6.7 Numerical Examples

To illustrate the performance of some constructed codes, simulations of the communication system in Section 6.2 were conducted for the simplified fading scenario using an $M = 2$ and $M = 4$ transmit antenna system, respectively. To focus on a common situation in practice, $N = 1$ receive antenna was assumed throughout the simulations. By considering three different cases, $\rho = 0$ (no channel knowledge), $\rho = 0.95$ and $\rho = 0.9999$ with $b = 2$ or $b = 6$ feedback bits, depending on the number of transmit antennas, the impact of the channel information quality on the performance was studied. Note that the $\rho = 0$ case also corresponds to an open-loop system for which conventional space-time codes are appropriate. Therefore, an OSTBC system was used as a benchmark. A comparison with beamforming was also performed. Using beamforming, the transmitted signal can be written on the form $\boldsymbol{c}(n) = \boldsymbol{v}s(n)$, where $s(n)$ represents the nth information bearing data symbol and $\boldsymbol{v}$ is a transmit weighting vector. For *ideal* beamforming the true channel $\boldsymbol{h}$ is used

as a weighting vector while for *conventional* beamforming $\boldsymbol{v} = \hat{\boldsymbol{\gamma}}(i)$, with the feedback encoder[2] now defined as $i = \varepsilon(\boldsymbol{\gamma}) \triangleq \arg\max_k \|\boldsymbol{\gamma}^* \hat{\boldsymbol{\gamma}}(k)\|_{\mathrm{F}}^2$. This choice of encoder serves to maximize the received signal power in the case of perfect initial channel knowledge [HP98].

The remaining assumptions in the simulations were as follows. For all the examined cases, equally probable bits were mapped into codewords/blocks of length L samples and conveyed over the spatially uncorrelated flat Rayleigh fading channel. An energy of $P = 1$ per information bit was used. The variance of the channel coefficients was set at $\sigma_h^2 = 1$. In the case of the systems employing OSTBC or beamforming, a BPSK signal constellation was employed. To make fair comparisons, the effective data rates of the constructed codes were fixed at 1 information bit per channel use. The SNR was measured for the conventional OSTBC system and defined as

$$\mathrm{SNR} \triangleq \frac{\mathrm{E}[\|\boldsymbol{H}^* \boldsymbol{C}\|_{\mathrm{F}}^2]}{LN\sigma^2},$$

where $\boldsymbol{C} = \sqrt{P_{\mathrm{o}}/M}\bar{\boldsymbol{C}}$ with $\bar{\boldsymbol{C}}$ denoting the codeword output from the OSTB encoder. The expression for the SNR is equal to the total received average signal energy, divided by the total average noise energy.

We start with investigating the performance of unstructured codes. Unstructured codes will then be compared with linear dispersive codes and weighted OSTBC by means of one representative simulation example.

Unstructured Codes versus Conventional Schemes

In Figure 6.2, the codeword/block error probability (BLER) as a function of the SNR is depicted for some unstructured codes, OSTBC and beamforming. An $M = 2$, $L = 2$ system is considered. The channel information was coarsely quantized using $\bar{b} = b = 2$ bits per complex-valued dimension. The $\rho = 0$ code was taken from Table 6.1 while all other unstructured codes were designed based on the same parameter values as used in the simulations.

Clearly, the performance of the constructed codes, regardless of channel information quality ρ, is significantly better than the performance of conventional OSTBC and beamforming. In particular, it should be noted that the OSTB code is outperformed by the unstructured codes even if there is no channel knowledge. As predicted by Table 6.2, a gain on the

[2] For an $M = 2$, $N = 1$ system, this encoder is equivalent to the one in (6.4).

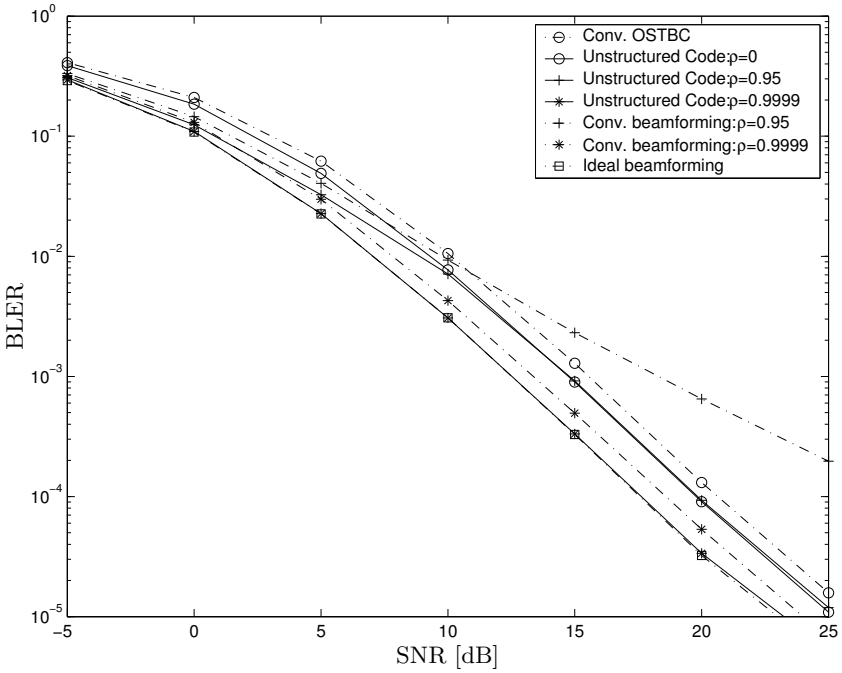

Figure 6.2: Performance comparisons for an $M = 2$, $L = 2$, $N = 1$, $b = 2$ system.

order of 0.6 - 0.8 dB compared with OSTBC is observed for high SNR values using the $\rho = 0$ code. A similar gain for the constructed channel side information dependent $\rho = 0.9999$ code compared with the corresponding conventional beamforming method is also seen. With higher SNR, the curve for the $\rho = 0.95$ code (eventually also the $\rho = 0.9999$ code) approaches the one for the no channel knowledge $\rho = 0$ code. This agrees well with the asymptotic high SNR result in Section 6.4.4 stating that channel knowledge is used less and less as the SNR is increased. In fact, the two channel side information dependent codes tend to the $\rho = 0$ code, since the latter can be viewed as parameter insensitive, and hence designed at an arbitrary SNR.

Figure 6.3 shows how the performance is affected if the number of transmit antennas and the code length are increased to $M = 4$ and $L = 4$, respectively. By using $\bar{b} = b/3 = 2$ bits, the quantization is roughly as

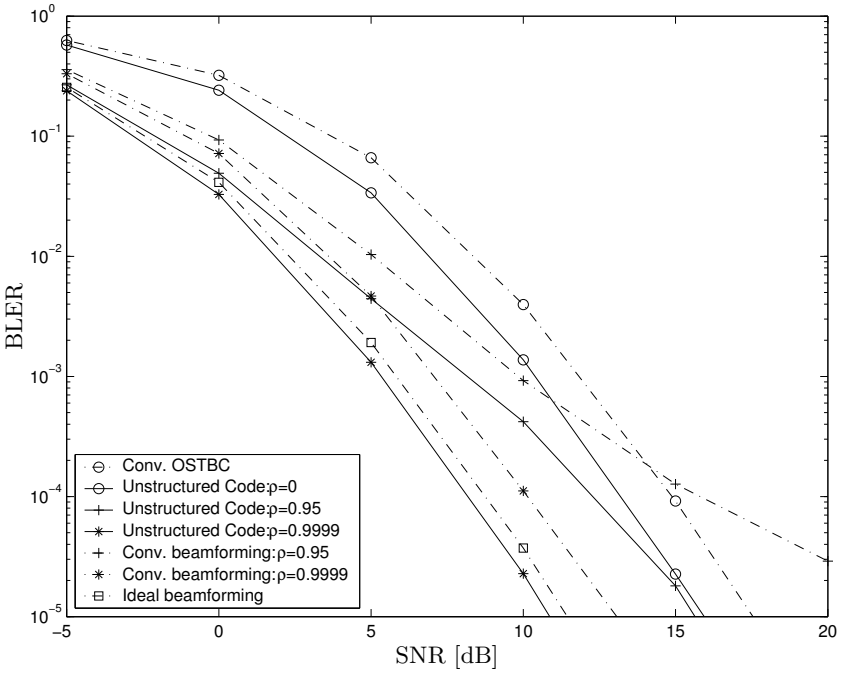

Figure 6.3: Performance comparisons for an $M = 4$, $L = 4$, $N = 1$, $b = 6$ system.

coarse as in the previous figure. As expected, the gains due to channel knowledge increase substantially with more transmit antennas. Again, the 1.4 dB coding gain in Table 6.2 agrees well with simulated results. It is also interesting to note that even though the unstructured $\rho = 0.9999$ code is based on coarsely quantized channel side information, it is significantly better than ideal beamforming, which has access to perfect (i.e., non-quantized) channel state information. Thus, the proposed code design procedure succeeds in exploiting the unstructured nature of the code for compensating for the imperfections of the channel feedback.

Comparing the Three Code Structures

Design procedures for both linear dispersive codes as well as weighted OSTBC have also been implemented. Linear dispersive codes were designed using the gradient search based optimization algorithm in (6.50). For weighted OSTBC, the transmit weight design procedure in (6.22) was implemented by means of an interior point algorithm (c.f. (2.54)) that finds the global optimal solution.

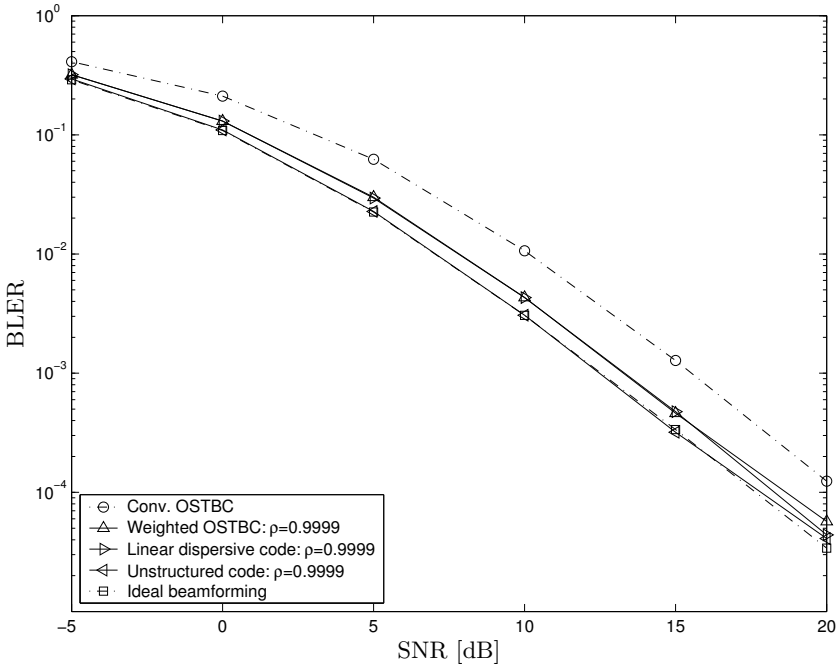

Figure 6.4: Illustrating how the choice of code structure affects the performance for an $M = 2$, $L = 2$, $N = 1$, $b = 2$ system.

To give an idea of how the choice of code structure influences the performance, constructed codes of the three code types are compared in Figure 6.4 for a system using $M = 2$ transmit antennas, $N = 1$ receive antennas, $L = 2$ and $b = 2$ feedback bits, under similar conditions as for the previous graphs. The quality of the initial channel information is here arbitrarily set to $\rho = 0.9999$, i.e., it is essentially perfect.

As expected, the unstructured code is better than the two other code

types. What may be more surprising is that the linear dispersive code and weighted OSTBC basically have the same performance. Numerical results not presented here show that this is also the case also for other values of ρ. Indeed, for the extreme case of no channel knowledge, it is known from Section 6.6 that the linear dispersive code is equivalent with OSTBC.

Preliminary experimental investigations indicate that linear dispersive codes and weighted OSTBC exhibit similar performance in many different scenarios. Thus, it may seem like linear dispersive codes offer little, if any, performance advantage over weighted OSTBC. However, keep in mind that the former code type exists for all possible combinations of antenna array sizes and data rates whereas weighted OSTBC is to a large degree limited by the fact that an OSTB code with the maximum data rate of two real-valued information symbols per channel use only exists for the case of two transmit antennas, as previously discussed in Section 3.2.3. A thin transmit weighting matrix, i.e., $M' < M$, may to some extent offset this limitation. Nevertheless, it is reasonable to expect that linear dispersive codes have clear performance advantages over weighted OSTBC for systems with more than two transmit antennas when the required symbol rate is higher than the low rate of one generally considered in the present section.

6.8 Conclusions

This work considered the design and analysis of space-time block codes utilizing non-perfect quantized channel side information. A new design criterion was derived that directly takes the quantized nature of the channel side information into account, thus avoiding heuristic modifications as in the previous chapter. While the focus was on unstructured block codes, design procedures for linear dispersive codes and weighted OSTBC were also proposed and implemented.

The constructed unstructured codes showed better performance than both conventional OSTBC as well as beamforming, even if there was no channel knowledge at the transmitter. In the case of no channel knowledge, the design procedure for linear dispersive codes was seen to automatically produce OSTB codes, which is interesting since OSTB codes previously have been handcrafted. Further investigations, experimental as well as theoretical, are however needed to more fully understand the mechanisms behind this result.

Appendix 6.A The Gradient of $V_{\mathrm{q}}(\boldsymbol{C}|i)$

The gradient of $V_{\mathrm{q}}(\boldsymbol{C}|i)$ with respect to the parameters in $\boldsymbol{C}$ will now be derived. In order to simplify the notation, let $\tilde{\boldsymbol{R}} \triangleq \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1} \mathrm{E}[\boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}} \boldsymbol{m}_{\boldsymbol{h}|\boldsymbol{\gamma}}^{*}|i] \boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}^{-1}$, $\boldsymbol{\Psi} \triangleq \boldsymbol{\Psi}(\boldsymbol{C})$ and drop the subscript of $\boldsymbol{R}_{\boldsymbol{hh}|\boldsymbol{\gamma}}$. For later reference, note the easily verifiable complex derivative relations

$$\frac{\partial x}{\partial x} = 1, \quad \frac{\partial x^{\mathrm{c}}}{\partial x} = 0\,, \tag{6.51}$$

where x is a complex-valued scalar. Observe that the second relation of (6.51) means that any function of x^{c} can be treated as a constant.

By evaluating the complex derivative with respect to element $[\boldsymbol{C}]_{kl}$ of $\boldsymbol{C}$ and making use of the identities [Kay93, p. 73]

$$\frac{\partial \boldsymbol{X}(\theta)^{-1}}{\partial \theta} = -\boldsymbol{X}(\theta)^{-1} \frac{\partial \boldsymbol{X}(\theta)}{\partial \theta} \boldsymbol{X}(\theta)^{-1} \tag{6.52}$$

$$\frac{\partial \log \det(\boldsymbol{X}(\theta))}{\partial \theta} = \mathrm{tr}\left(\boldsymbol{X}(\theta)^{-1} \frac{\partial \boldsymbol{X}(\theta)}{\partial \theta} \right)\,, \tag{6.53}$$

as well as (6.51) and the fact that $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$ we have

$$\begin{aligned}
\frac{\partial \log\left(V_{\mathrm{q}}(\boldsymbol{C}|i)\right)}{\partial [\boldsymbol{C}]_{kl}} &= -\eta\, \mathrm{tr}\left(\boldsymbol{\Psi}^{-1}(\boldsymbol{I}_N \otimes \boldsymbol{e}_k^{(M)}(\boldsymbol{e}_l^{(L)})^{*}\boldsymbol{C}^{*})\right) \\
&\quad -\eta\, \mathrm{tr}\left(\boldsymbol{\Psi}^{-1}\left(\boldsymbol{I}_N \otimes \boldsymbol{e}_k^{(M)}(\boldsymbol{e}_l^{(L)})^{*}\boldsymbol{C}^{*}\right)\boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{R}}\right) \\
&= -\eta\, \mathrm{tr}\left(\boldsymbol{\Psi}^{-1}(\boldsymbol{I}_N \otimes \boldsymbol{e}_k^{(M)}(\boldsymbol{e}_l^{(L)})^{*}\boldsymbol{C}^{*})\right) \\
&\quad -\eta\, \mathrm{tr}\left(\boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{R}}\boldsymbol{\Psi}^{-1}(\boldsymbol{I}_N \otimes \boldsymbol{e}_k^{(M)}(\boldsymbol{e}_l^{(L)})^{*}\boldsymbol{C}^{*})\right) \\
&= -\eta \sum_{n=1}^{N} \mathrm{tr}\left(\boldsymbol{\Omega}_n(\boldsymbol{e}_k^{(M)}(\boldsymbol{e}_l^{(L)})^{*}\boldsymbol{C}^{*})\right) \\
&= -\eta(\boldsymbol{e}_l^{(L)})^{*}\boldsymbol{C}^{*}\left(\sum_{n=1}^{N}\boldsymbol{\Omega}_n\right)\boldsymbol{e}_k^{(M)}\,,
\end{aligned} \tag{6.54}$$

where $\boldsymbol{\Omega}_n$ represents the nth block of size $M \times M$ on the diagonal of $\boldsymbol{\Psi}^{-1} + \boldsymbol{\Psi}^{-1}\tilde{\boldsymbol{R}}\boldsymbol{\Psi}^{-1}$. The first equality is due to the fact that differentiating each element of $\boldsymbol{C}$ with respect to $[\boldsymbol{C}]_{kl}$ yields

$$\frac{\partial \boldsymbol{C}}{\partial [\boldsymbol{C}]_{kl}} = \boldsymbol{e}_k^{(M)}(\boldsymbol{e}_l^{(L)})^{*}\,,$$

where $\boldsymbol{e}_k^{(K)}$ denotes the kth column of $\boldsymbol{I}_K$. After collecting the derivatives (6.54) in a matrix it is straightforward to show that

$$\frac{\partial \log\big(V_{\mathrm{q}}(\boldsymbol{C})\big)}{\partial \boldsymbol{C}} = -\eta(\boldsymbol{\Omega}\boldsymbol{C})^{\mathrm{c}}\,,$$

where $\boldsymbol{\Omega} = \sum_{n=1}^{N} \boldsymbol{\Omega}_n$. The gradient of $V_{\mathrm{q}}(\boldsymbol{C}|i)$ is finally obtained as

$$\frac{\partial V_{\mathrm{q}}(\boldsymbol{C}|i)}{\partial \boldsymbol{C}} = -\eta V_{\mathrm{q}}(\boldsymbol{C}|i)(\boldsymbol{\Omega}\boldsymbol{C})^{\mathrm{c}}\,.$$

Appendix A

Acronyms

BER	Bit Error Rate
BLER	BLock Error Rate
BPSK	Binary Phase Shift Keying
COVQ	Channel Optimized Vector Quantization
DOA	Direction Of Arrival
EVD	EigenValue Decomposition
FDD	Frequency Division Duplex
GSM	Global System for Mobile communication or Groupe Spécial Mobile
IID	Independent and Identically Distributed
KKT	Karush-Kuhn-Tucker
ML	Maximum Likelihood
MIMO	Multi-Input-Multi-Output
MISO	Multi-Input-Single-Output
MMSE	Minimum Mean-Square Error
OSTB	Orthogonal Space-Time Block
OSTBC	Orthogonal Space-Time Block Coding
OFDM	Orthogonal Frequency Division Multiplexing
PAM	Pulse Amplitude Modulation
PDF	Probability Density Function
PMF	Probability Mass Function

QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
SIMO	Single-Input-Multi-Output
SISO	Single-Input-Single-Output
SINR	Signal-to-Interference-plus-Noise-Ratio
SNR	Signal-to-Noise-Ratio
SVD	Singular Value Decomposition
TDD	Time Division Duplex
VQ	Vector Quantization
WCDMA	Wideband Code Division Multiple-Access

Appendix B

Notation

In this thesis, vectors and matrices are usually denoted by boldface lowercase and uppercase letters, respectively. For example, $\boldsymbol{x}$ denotes a vector while $\boldsymbol{X}$ represents a matrix. Vectors are generally represented by one-column matrices. A non-boldface calligraphic typeface is used to denote sets, e.g. $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and so on. The list below and on the following pages defines additional notation.

$M \times N$	The dimension of a matrix with M rows and N columns.
$\boldsymbol{X}_{M\times N}$	An $M \times N$ matrix $\boldsymbol{X}$.
$\boldsymbol{I}_M$	The identity matrix of dimension $M \times M$.
$[\boldsymbol{X}]_{kl}$	The element on the kth row and lth column, also referred to as element (k, l), of the matrix $\boldsymbol{X}$.
$\boldsymbol{X}^{\mathrm{T}}$	The transpose of the matrix $\boldsymbol{X}$.
$\boldsymbol{X}^*$	The complex conjugate transpose of the matrix $\boldsymbol{X}$.
$\boldsymbol{X}^{\mathrm{c}}$	The matrix obtained by taking the complex conjugate on each element of the matrix $\boldsymbol{X}$, i.e., $\boldsymbol{X}^{\mathrm{c}} = (\boldsymbol{X}^*)^{\mathrm{T}}$.
$\|\boldsymbol{x}\|$	The Euclidean norm of a vector $\boldsymbol{x}$, i.e., $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^*\boldsymbol{x}}$.

$\|\boldsymbol{x}\|_{\boldsymbol{K}}$	The norm of the vector $\boldsymbol{x}$ defined through $\|\boldsymbol{x}\|_{\boldsymbol{K}}^2 = \boldsymbol{x}^*\boldsymbol{K}\boldsymbol{x}$, where $\boldsymbol{K}$ is a positive definite weight matrix.
$\|\boldsymbol{X}\|$	The spectral norm of the matrix $\boldsymbol{X}$, i.e., $\|\boldsymbol{X}\| = \sigma_{\max}$, where $\sigma_{\max}$ is the maximum singular value of $\boldsymbol{X}$. Reduces to the usual Euclidian norm if $\boldsymbol{X}$ is a vector.
$\|\boldsymbol{X}\|_{\mathrm{F}}$	The Frobenius norm of the matrix $\boldsymbol{X}$, i.e., $\|\boldsymbol{X}\|_{\mathrm{F}} = \sqrt{\sum_{k,l} \lvert[\boldsymbol{X}]_{kl}\rvert^2}$.
$\mathrm{rank}(\boldsymbol{X})$	The rank of the matrix $\boldsymbol{X}$.
$\mathrm{tr}(\boldsymbol{X})$	The trace of the matrix $\boldsymbol{X}$.
$\det(\boldsymbol{X})$	The determinant of the matrix $\boldsymbol{X}$.
$\mathrm{vec}(\boldsymbol{X})$	The vectorized counterpart of the matrix $\boldsymbol{X}$, i.e., $\mathrm{vec}(\boldsymbol{X})$ is a vector containing the columns of $\boldsymbol{X}$ stacked on top of each other. The columns are ordered from the leftmost at the top to the rightmost at the bottom.
$\mathrm{diag}(x_1, x_2, \cdots, x_M)$	A diagonal matrix with elements $x_1, x_2, \cdots, x_M$ on the main diagonal.
$\boldsymbol{A} \succeq \boldsymbol{B}$	$\boldsymbol{A} - \boldsymbol{B}$ is a positive semi definite matrix.
$\boldsymbol{A} \otimes \boldsymbol{B}$	The Kronecker product of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.
δ_n	The Kronecker delta function, i.e., $\delta_0 = 1$ and zero otherwise. The argument n is here assumed to be an integer.
$\mathrm{E}[\cdot]$	The expectation operator. Alternatively, $\mathrm{E}_{x_1,x_2,\ldots}[\cdot]$ may be used to emphasize that the expectation is with respect to the random variables $x_1, x_2, \cdots$.
$x\|y$	Should be read as "x conditioned on y" except when it appears within a function name such as for example $f(x\|y)$. In the latter case, $x\|y$ is used for emphasizing that f is a function of x parameterized by y.
$\mathrm{E}[x\|y]$	Expected value of x conditioned on y, i.e., the expectation is with respect to the PDF/PMF of $x\|y$.

$\Pr[\mathcal{E}]$	Probability of the event $\mathcal{E}$. For example, $\Pr[i = k]$ represents the probability that the random variable i is equal to the deterministic value k.
$x \triangleq a$	Denotes that x is equal to a per definition (and vice versa).
$\mathrm{O}(\cdot)$	The big ordo operator.
$\mathrm{re}(x)$ and $\mathrm{im}(x)$	Real and imaginary part of x, respectively.
$\arg(x)$	The phase in radians, in the range $-\pi$ to π, of the complex-valued scalar x.
$\arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$	The value of $\boldsymbol{x}$ that minimizes the criterion function $f(\boldsymbol{x})$.
$\arg\max_{\boldsymbol{x}} f(\boldsymbol{x})$	The value of $\boldsymbol{x}$ that maximizes the criterion function $f(\boldsymbol{x})$.
$\sup_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x})$	The supremum of $f(\boldsymbol{x})$ over the set of $\boldsymbol{x}$ such that $\boldsymbol{x} \in \mathcal{X}$.
$\exp(x)$ or e^x	The exponential function.
$\log(x)$	The natural logarithm of x.
$\log_2(x)$	The base 2 logarithm of x.
$\mathbb{C}$	The complex number field.
$\mathbb{R}$	The real number field.
$\{A_k\}$	Denotes the set of all A_k.
$\{A_k\}_{k=1}^{K}$	The set $\{A_1, A_2, \cdots, A_K\}$.
$\lvert\mathcal{A}\rvert$	The cardinality of the set $\mathcal{A}$, i.e., the number of elements in the set.

Appendix C

Matrix Relations

Some useful formulas for manipulating expressions with matrices are given below. Proofs of all matrix relations are found in e.g. [SS89, Appendix A] combined with [Gra81]. Let $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ denote matrices with dimensions so that the operations make sense. It then holds that

$$\operatorname{tr}(\boldsymbol{AB}) = \big(\operatorname{vec}(\boldsymbol{A}^*)\big)^* \operatorname{vec}(\boldsymbol{B}) \tag{C.1}$$

$$\operatorname{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{B}) \tag{C.2}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC} \otimes \boldsymbol{BD}) \tag{C.3}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1} \tag{C.4}$$

$$\det(\boldsymbol{I} + \boldsymbol{AB}) = \det(\boldsymbol{I} + \boldsymbol{BA})\,. \tag{C.5}$$

Bibliography

[3GP02a] 3rd Generation Partnership Project (3GPP). Physical channels and mapping of transport channels onto physical channels. Technical Report TS 25.211 V3.11.0, 2002.

[3GP02b] 3rd Generation Partnership Project (3GPP). Physical layer procedures. Technical Report 3GPP TS 25.214 V3.10.0, 2002.

[AHD+99] S. Anderson, B. Hagerman, H. Dam, U. Forssen, J. Karlsson, F. Kronestedt, S. Mazur, and K. J. Molnar. Adaptive antennas for GSM and TDMA systems. *IEEE Personal Communications*, 6(3):74–86, June 1999.

[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.

[AMVW91] S. Andersson, M. Millnert, M. Viberg, and B. Wahlberg. An adaptive array for mobile communication systems. *IEEE Transactions on Vehicular Technology*, 40(1):230–236, February 1991.

[And63] T. W. Anderson. Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, 34:122–148, February 1963.

[AO98] D. Astély and B. Ottersten. MLSE and spatio-temporal interference rejection combining with antenna arrays. In *Proceedings European Signal Processing Conference*, pages 1341–1344, September 1998.

[ARU01] D. Agrawal, T. J. Richardson, and R. L. Urbanke. Multiple-antenna signal constellations for fading channels. *IEEE Transactions on Information Theory*, 47(6):2618–2626, September 2001.

[Asz95] D. Asztély. On antenna arrays in mobile communication systems, fast fading and GSM base station receiver algorithms. Master's thesis, Royal Institute of Technology, February 1995.

[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*, pages 2232–2236, May 1998.

[BBH00] S. Bäro, G. Bauch, and A. Hansmann. Improved codes for space-time trellis-coded modulation. *IEEE Communications Letters*, 4(1):20–22, January 2000.

[BCT01] E. Biglieri, G. Caire, and G. Taricco. Limiting performance of block-fading channels with multiple antennas. *IEEE Transactions on Information Theory*, 47(4):1273–1289, May 2001.

[BD91] W. R. Braun and U. Dersch. A physical mobile radio channel model. *IEEE Transactions on Vehicular Technology*, 40(2):472–482, May 1991.

[BFAOHY02] D. W. Bliss, K. W. Forsythe, III A. O. Hero, and A. F. Yegulalp. Environmental issues for MIMO capacity. *IEEE Transactions on Signal Processing*, 50(9):2128–2142, September 2002.

[Bin90] J. A. C. Bingham. Multicarrier modulation for data transmission: An idea whose time has come. *IEEE Communications Magazine*, 28(5):5–14, May 1990.

[BJ95] G. E. Bottomley and K. Jamal. Adaptive arrays and MLSE equalization. In *Proc. IEEE Vehicular Technology Conference*, pages 50–54, July 1995.

[Blu02] R. S. Blum. Some analytical tools for the design of space-time convolutional codes. *IEEE Transactions on Communications*, 50(10):1593–1599, October 2002.

[BMC99] G. E. Bottomley, K. J. Molnar, and S. Chennakeshu. Interference cancellation with an array processing MLSE receiver. *IEEE Transactions on Vehicular Technology*, 48(5):1321–1331, September 1999.

[BO02] M. Bengtsson and B. Ottersten. *Handbook of Antennas in Wireless Communications*, chapter 18. The Electrical Engineering and Applied Signal Processing Series. CRC Press, 2002.

[BS91] P. Balaban and J. Salz. Dual diversity combining and equalization in digital cellular mobile radio. *IEEE Transactions on Vehicular Technology*, 40(2):342–354, May 1991.

[BS93] M. S. Bazaraa and H. D. Sherali. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, 2nd edition, 1993.

[Bur96] A. G. Burr. The multipath problem: An overview. In *IEE Colloquium on Multipath Countermeasures*, pages 1/1–1/7, May 1996.

[BV99] S. Boyd and L. Vandenberghe. Introduction to convex optimization with engineering applications. Course Notes, http://www.stanford.edu/class/ee364, 1999.

[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.

[CGS94] M. V. Clark, L. J. Greenstein, and W. K. Shafi. Optimum linear diversity receivers for mobile communications. *IEEE Transactions on Vehicular Technology*, 43(1):47–56, February 1994.

[Cim85] L. J. Cimini Jr. Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. *IEEE Transactions on Communications*, COM-33(7):665–675, July 1985.

[CS99] G. Caire and S. Shamai. On the capacity of some channels with channel state information. *IEEE Transactions on Information Theory*, 45(6):2007–2019, September 1999.

[CT91] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[DGI+02] R. T. Derryberry, S. D. Gray, D. M. Ionescu, G. Mandyam, and B. Raghothaman. Transmit diversity in 3G CDMA systems. *IEEE Communications Magazine*, 40(4):68–75, April 2002.

[DHSH00] A. Duel-Hallen, H. Shengquan, and H. Hallen. Long-range prediction of fading signals. *IEEE Signal Processing Magazine*, 17(3):62–75, May 2000.

[DN02] A. Das and P. Narayan. Capacities of time-varying multiple-access channels with side information. *IEEE Transactions on Information Theory*, 48(1):4–25, January 2002.

[ECS+98] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed. Overview of spatial channel models for antenna array communication systems. *IEEE Personal Communications*, 5(1):10–22, February 1998.

[Far90] N. Farvardin. A study of vector quantization for noisy channels. *IEEE Transactions on Information Theory*, 36(4):799–809, July 1990.

[FG98] G. J. Foschini and M. J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6(3):311–335, March 1998.

[FGW74] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein. Optimization of two-dimensional signal constellations in the presence of Gaussian noise. *IEEE Transactions on Communications*, 22(1):28–37, January 1974.

[Fos96] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Technical Journal*, pages 41–59, 1996.

[FU98] G. D. Forney Jr. and G. Ungerboeck. Modulation and coding for linear Gaussian channels. *IEEE Transactions on Information Theory*, 44(6):2384–2415, October 1998.

[FV91] N. Farvardin and V. Vaishampayan. On the performance and complexity of channel-optimized vector quantizers. *IEEE Transactions on Information Theory*, 37(1):155–160, January 1991.

[Gal68] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[GBGP02] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj. Outdoor MIMO wireless channels: Models and performance prediction. *IEEE Transactions on Communications*, 50(12):1926–1934, December 2002.

[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.

[GFK96] J.-C. Guey, M. P. Fitz, and W.-Y. Kuo. Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels. In *Proc. IEEE Vehicular Technology Conference*, pages 136–140, 1996.

[GG92] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

[GP94a] D. Gerlach and A. Paulraj. Adaptive transmitting antenna arrays with feedback. *IEEE Signal Processing Letters*, 1(10):150–152, October 1994.

[GP94b] D. Gerlach and A. Paulraj. Adaptive transmitting antenna methods for multipath environments. In *Proc. IEEE Global Telecommunications Conference*, pages 425–429, November 1994.

[Gra81] A. Graham. *Kronecker Products and Matrix Calculus with Applications*. Ellis Horwood Ltd, 1981.

[GS00] G. Ganesan and Petre Stoica. *Signal Processing Advances in Wireless & Mobile Communications: Trends in Single-User and Multi-User Systems*, chapter 2. Prentice Hall, 2000.

[GS01] G. Ganesan and P. Stoica. Space-time block codes: A maximum SNR approach. *IEEE Transactions on Information Theory*, 47(4):1650–1656, May 2001.

[GS02] G. Ganesan and P. Stoica. Differential modulation using space-time block codes. *IEEE Signal Processing Letters*, 9(2):57–60, February 2002.

[GS03] J. Giese and M. Skoglund. Space-time constellations for unknown frequency-selective channels. In *Proc. IEEE International Conference on Communications*, May 2003.

[GV97] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43(6):1986–1992, November 1997.

[HAN92] A. Hiroike, F. Adachi, and N. Nakajima. Combined effects of phase sweeping transmitter diversity and channel coding. In *Proc. IEEE Vehicular Technology Conference*, May 1992.

[HH78] T. Hattori and K. Hirade. Multitransmitter digital signal transmission by using offset frequency strategy in land-mobile radio telephone system. *IEEE Transactions on Vehicular Technology*, 27(4):231–238, November 1978.

[HH02a] B. Hassibi and B. M. Hochwald. Cayley differential unitary space-time codes. *IEEE Transactions on Information Theory*, 48(6):1485–1503, June 2002.

[HH02b] 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.

[HJ96] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1996.

[HP98] R. W. Heath Jr. and A. Paulraj. A simple scheme for transmit diversity using partial channel feedback. In *Proc. 32th Asilomar Conference on Signals, Systems and Computers*, pages 1073–1078, November 1998.

[HS00] B. M. Hochwald and W. Sweldens. Differential unitary space-time modulation. *IEEE Transactions on Communications*, 48(12):2041–2052, December 2000.

[HSN91] M. L. Honig, K. S., and S. A. Norman. Optimization of signal sets for partial-response channels–part I: Numerical techniques. *IEEE Transactions on Information Theory*, 37(5):1327–1341, September 1991.

[HT02] H. Holma and A. Toskala. *WCDMA for UMTS*. John Wiley & Sons, 2002.

[Hug00] B. L. Hughes. Differential space-time modulation. *IEEE Transactions on Information Theory*, 46(7):2567–2578, November 2000.

[HW01] J. Hämäläinen and R. Wichman. The effect of feedback delay to the closed-loop transmit diversity in FDD WCDMA. In *12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, volume 1, pages D–27–D–31, September 2001.

[Jak94] W. C. Jakes. *Microwave Mobile Communications*. IEEE press, 1994.

[JAO00] G. Jöngren, D. Astély, and B. Ottersten. Structured spatial interference rejection combining. In *Proceedings European Signal Processing Conference*, September 2000.

[JMO03] J. Jaldén, C. Martin, and B. Ottersten. Semidefinite programming for detection in linear systems – optimality conditions and space-time decoding. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume IV, pages 9–12, April 2003.

[JO99] G. Jöngren and B. Ottersten. Combining transmit antenna weights and orthogonal space-time block codes by utilizing side information. In *Proc. 33th Asilomar Conference on Signals, Systems and Computers*, October 1999.

[JS00] G. Jöngren and M. Skoglund. Utilizing quantized feedback information in orthogonal space-time block coding. In *Proceedings IEEE Global Telecommunications Conference*, November 2000.

[JS01] G. Jöngren and M. Skoglund. Improving orthogonal space-time block codes by utilizing quantized feedback information. In *Proceedings International Symposium on Information Theory*, June 2001.

[JSO00] G. Jöngren, M. Skoglund, and B. Ottersten. Combining transmit beamforming and orthogonal space-time block codes by utilizing side information. In *Proc. First IEEE Sensor Array and Multichannel Signal Processing Workshop*, March 2000.

[JSO02a] G. Jöngren, M. Skoglund, and B. Ottersten. Combining beamforming and orthogonal space-time block coding. *IEEE Transactions on Information Theory*, 48(3):611–627, March 2002.

[JSO02b] 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*, October 2002.

[Kay93] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[Kay98] S. M. Kay. *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall, 1998.

[KF97] W.-Y. Kuo and M. P. Fitz. Design and analysis of transmitter diversity using intentional frequency offset for wireless communications. *IEEE Transactions on Vehicular Technology*, 46(4):871–881, November 1997.

[Koh98] R. Kohno. Spatial and temporal communication theory using adaptive antenna array. *IEEE Personal Communications*, 5(1):28–35, February 1998.

[KSP+02] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen. A stochastic MIMO radio

channel model with experimental validation. *IEEE Journal on Selected Areas in Communications*, 20(6):1211–1226, August 2002.

[KTT+99] A. Kuchar, M. Taferner, M. Tangemann, C. Hoek, W. Rauscher, M. Strasser, G. Pospischil, and E. Bonek. Real-time smart antenna processing for GSM1800 base station. In *Proc. IEEE Vehicular Technology Conference*, volume 1, pages 664–669, July 1999.

[KV96] H. Krim and M. Viberg. Two decades of array signal processing research: The parametric approach. *IEEE Signal Processing Magazine*, 13(4):67–94, July 1996.

[LGSW02] E. G. Larsson, G. Ganesan, P. Stoica, and W.-H. Wong. On the performance of orthogonal space-time block coding with quantized feedback. *IEEE Communications Letters*, 6(11):487–489, November 2002.

[LP00] E. Lindskog and A. Paulraj. A transmit diversity scheme for channels with intersymbol interference. In *Proc. IEEE International Conference on Communications*, pages 307–311, 2000.

[May97] S. Mayrargue. Practical implementation of a multisensor array receiver structure for wireless communications. In *Proc. SPAWC*, pages 201–204, April 1997.

[MBV02] M. L. McCloud, M. Brehler, and M. K. 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.

[ME86] J. W. Modestino and V. M. Eyuboglu. Integrated multielement receiver structures for spatially distributed interference channels. *IEEE Transactions on Information Theory*, IT-32(2):195–219, March 1986.

[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.

[MM80] R. A. Monzingo and T. W. Miller. *Introduction to Adaptive Arrays*. John Wiley & Sons, 1980.

[Mog93] P. E. Mogensen. GSM base-station antenna diversity using soft decision combining on up-link and delayed-signal transmission on down-link. In *Proc. IEEE Vehicular Technology Conference*, pages 611–616, May 1993.

[MSA01] K. K. Mukkavilli, A. Sabharwal, and B. Aazhang. Design of multiple antenna coding schemes with channel feedback. In *Proc. of 35th Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1009–1013, November 2001.

[NLTW98] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell. Efficient use of side information in multiple-antenna data transmission over fading channels. *IEEE Journal on Selected Areas in Communications*, 16(8):1423–1436, October 1998.

[NN94] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM Publications, 1994.

[NTC99] R. Negi, A. M. Tehrani, and J. M. Cioffi. Adaptive antennas for space-time coding over block-time invariant multipath fading channels. In *Proc. IEEE Vehicular Technology Conference*, pages 70–74, May 1999.

[NTW99] A. Narula, M. D. Trott, and G. W. Wornell. Performance limits of coded diversity methods for transmitter antenna arrays. *IEEE Transactions on Information Theory*, 45(7):2418–2433, November 1999.

[OGDH01] E. N. Onggosanusi, A. Gatherer, A. G. Dabak, and S. Hosur. Performance analysis of closed-loop transmit diversity in the presence of feedback delay. *IEEE Transactions on Communications*, 49(9):1618–1630, September 2001.

[PP97] A. J. Paulraj and C. B. Papadias. Space-time processing for wireless communication. *IEEE Signal Processing Magazine*, 14(6):49–83, November 1997.

[Pro95] J. G. Proakis. *Digital Communications*. McGraw-Hill, 3rd edition, 1995.

[RC98] G. G. Raleigh and J.M. Cioffi. Spatio-temporal coding for wireless communication. *IEEE Transactions on Communications*, 46(3):357–366, March 1998.

[RDJP95] G. G. Raleigh, S. N. Diggavi, V. K. Jones, and A. Paulraj. A blind adaptive transmit antenna algorithm for wireless communication. In *Proc. IEEE International Conference on Communications*, volume 3, pages 1494–1499, June 1995.

[RFLT98] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas. Transmit beamforming and power control for cellular wireless systems. *IEEE Journal on Selected Areas in Communications*, 16(8):1437–1450, October 1998.

[RJ99] G. G. Raleigh and V. K. Jones. Multivariate modulation and coding for wireless communication. *IEEE Journal on Selected Areas in Communications*, 17(5):851–866, May 1999.

[RWO95] S. Redl, M. K. Weber, and M. Oliphant. *An Introduction to GSM*. Artech House, 1995.

[Sal92] M. Salehi. Capacity and coding for memories with real-time noisy defect information at encoder and decoder. *IEE Proceedings-I*, 139(2):113–117, April 1992.

[SBEM90] S. C. Swales, M. A. Beach, D. J. Edwards, and J. P. McGeehan. The performance enhancement of multibeam adaptive base-station antennas for cellular land mobile radio systems. *IEEE Transactions on Vehicular Technology*, 39(1):56–67, February 1990.

[SD01] A. Stefanov and T. M. Duman. Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: System model, decoding approaches, and practical considerations. *IEEE Journal on Selected Areas in Communications*, 19(5):958–968, May 2001.

[SFGK00] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn. Fading correlation and its effect on the capacity of multi-element antenna systems. *IEEE Transactions on Communications*, 48(3):502–513, March 2000.

[SG00] N. Sharma and E. Geraniotos. Analyzing the performance of the space-time block codes with partial channel state feedback. In *Proc. IEEE Wireless Communications and Networking Conference*, volume 3, pages 1362–1366, September 2000.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3 and 4):379–423 and 623–656, July and Oct. 1948.

[Sha58] C. E. Shannon. Channels with side information at the transmitter. *IBM J. Res. Devel.*, 2:289–293, 1958.

[SHP00] S. Sandhu, R. W. Heath Jr., and A. Paulraj. Union bound for linear space-time codes. In *Proc. 38th Annual Allerton Conference on Communication, Control and Computing*, September 2000.

[SJ03] M. Skoglund and G. Jöngren. On the capacity of a multiple-antenna communication link with channel side information. *IEEE Journal on Selected Areas in Communications*, 21(3):395–405, April 2003.

[SP01] S. Sandhu and A. Paulraj. Union bound for linear space-time block codes under different channel conditions. In *Proc. 39th Annual Allerton Conference on Communication, Control and Computing*, October 2001.

[SS89] T. Söderström and P. Stoica. *System Identification*. Prentice Hall, 1989.

[SW93] N. Seshadri and J. H. Winters. Two signaling schemes for improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitter antenna diversity. In *Proc. IEEE Vehicular Technology Conference*, pages 508–511, May 1993.

[SX03] W. Su and X.-G. Xia. Two generalized complex orthogonal space-time block codes of rates 7/11 and 3/5 for 5 and 6 transmit antennas. *IEEE Transactions on Information Theory*, 49(1):313–316, January 2003.

[TAP98] V. Tarokh, S. M. Alamouti, and P. Poon. New detection schemes for transmit diversity with no channel estimation. In *Proc. IEEE International Conference on Personal Communications*, pages 917–920, October 1998.

[Tel95] E. Telatar. Capacity of multi-antenna Gaussian channels. *Bell Laboratories Technical Memorandum*, October 1995.

[TJ00] V. Tarokh and H. Jafarkhani. A differential detection scheme for transmit diversity. *IEEE Journal on Selected Areas in Communications*, 18(7):1169–1174, July 2000.

[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.

[VB96] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996.

[VH02] H. Vikalo and B. Hassibi. Maximum-likelihood sequence detection of multiple antenna systems over dispersive channels via sphere decoding. *EURASIP Journal on Applied Signal Processing*, 2002(5):525–531, May 2002.

[Vis99] H. Viswanathan. Capacity of markov channels with receiver CSI and delayed feedback. *IEEE Transactions on Information Theory*, 45(2):761–771, March 1999.

[VM00] E. Visotsky and U. Madhow. Space-time precoding with imperfect feedback. In *Proc. IEEE International Symposium on Information Theory*, page 312, June 2000.

[VM01] E. Visotsky and U. Madhow. Space-time transmit precoding with imperfect feedback. *IEEE Transactions on Information Theory*, 47(6):2632–2638, September 2001.

[WE71] S. B. Weinstein and P. M. Ebert. Data transmission by frequency-division multiplexing using the discrete fourier transform. *IEEE Transactions on Communication Technology*, COM-19(5):628–634, October 1971.

[Wee93] V. Weerackody. Characteristics of simulated fast fading indoor radio channel. In *Proc. IEEE Vehicular Technology Conference*, pages 231–235, 1993.

[Win94] J. H. Winters. The diversity gain of transmit diversity in wireless systems with Rayleigh fading. In *Proc. IEEE International Conference on Communications*, pages 1121–1125, May 1994.

[Win98] J. H. Winters. The diversity gain of transmit diversity in wireless systems with Rayleigh fading. *IEEE Transactions on Vehicular Technology*, 47(1):119–123, February 1998.

[Wit91] A. Wittneben. Basestation modulation diversity for digital SIMULCAST. In *Proc. IEEE Vehicular Technology Conference*, pages 848–853, May 1991.

[Wit93] A. Wittneben. A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation. In *Proc. IEEE International Conference on Communications*, pages 1630–1634, May 1993.

[Wit95] A. Wittneben. Optimal predictive TX combining diversity in correlated fading for microcellular mobile radio applications. In *Proc. IEEE Global Telecommunications Conference*, pages 48–54, 1995.

[WJ90] J. M. Wozencraft and I. M. Jacobs. *Principles of Communication Engineering*. Waveland Press, Inc, 1990.

[WJ02] J. W. Wallace and M. A. Jensen. Modeling the indoor MIMO wireless channel. *IEEE Transactions on Antennas and Propagation*, 50(5):591–599, May 2002.

[WT97] G. W. Wornell and M. D. Trott. Efficient signal processing techniques for exploiting transmit antenna diversity on fading channels. *IEEE Transactions on Signal Processing*, 45(1):191–205, January 1997.

[XL95] G. Xu and H. Liu. An effective transmission beamforming scheme for frequency-division-duplex digital wireless communication systems. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 1729–1732, May 1995.

[Zet97] P. Zetterberg. *Mobile Cellular Communications with Base Station Antenna Arrays.* PhD thesis, Royal Institute of Technology, Stockholm, Sweden, 1997.

[Zet99] P. Zetterberg. A comparison of two systems for downlink communication with base station antenna arrays. *IEEE Transactions on Vehicular Technology*, 48(5):1356–1370, September 1999.

[ZG02a] S. Zhou and G. B. Giannakis. Optimal transmitter eigenbeamforming and space-time block coding based on channel mean feedback. *IEEE Transactions on Signal Processing*, 50(10):2599–2613, October 2002.

[ZG02b] S. Zhou and G. B. Giannakis. Optimal transmitter eigenbeamforming and space-time block coding based on channel correlations. In *Proc. IEEE International Conference on Communications*, volume 1, pages 553–557, April 2002.

[ZM92] W. Zhang and M. J. Miller. Baseband equivalents in digital communication system simulation. *IEEE Transactions on Education*, 35(4):376–382, November 1992.

[ZO95] P. Zetterberg and B. Ottersten. The spectrum efficiency of a base station antenna array system for spatially selective transmission. *IEEE Transactions on Vehicular Technology*, 44(3):651–660, August 1995.