SPACE-TIME CODING: FROM FUNDAMENTALS TO THE FUTURE

BY

Karen Su
(Trinity Hall)

First year report submitted for admission to candidacy
for the degree of Doctor of Philosophy,
Laboratory for Communication Engineering,
Department of Engineering,
University of Cambridge.

Space-time Coding:
From Fundamentals to the Future

First Year Report
Laboratory for Communication Engineering
Cambridge University Engineering Department
University of Cambridge

by Karen Su
September 2003

Abstract

In this report, we explore the fundamental concepts behind the emerging field of space-time coding for wireless communication systems. The first three chapters of background material discuss, respectively, signal fading and modelling in the wireless propagation environment, spatial diversity via Multi-Element Antenna (MEA) arrays, and the capacity of the Multiple Input Multiple Output (MIMO) wireless channel in Rayleigh fading. We find that at the heart of space-time coding lies the design of two-dimensional signal matrices to be transmitted over a period of time from a number of antennas. The structure of the signal enables us to exploit diversity in the spatial and temporal dimensions in order to obtain improved bit error performance and higher data rates without bandwidth expansion. Thus it is clear that transmit diversity plays an integral role in space-time code design. A brief survey of such existing communication techniques follows this discussion, leading naturally to a proposal for the work to be undertaken in our Ph.D. project.

Acknowledgements

Well it's not like this is a thesis or anything but it sure felt like it!

I would like to thank my PhD supervisor Dr Ian Wassell first for his extremely high speed yet meticulous read through several drafts of this unintentionally voluminous tome and for providing valuable feedback and suggestions to improve its content and presentation. Secondly for always having his office door open, pointing me in the right direction on a number of issues and giving me lots of books to read. I would also like to thank in advance my unsuspecting first year advisor Dr Arnaud Doucet for his time and patience in reading this report.

Next, I gratefully acknowledge the financial support of Universities UK, the Cambridge Commonwealth Trust, the Natural Sciences and Engineering Research Council of Canada, and Trinity Hall. Because of their generous assistance, I have been able to pursue my chosen program of study here at Cambridge University. I would also not have been able to realize this dream without the countless letters of reference written on my behalf. For these I am indebted to Dr Deepa Kundur, Dr Dimitrios Hatzinakos, Dr Frank Kschischang, and Dr Teng Joon Lim of the University of Toronto.

The last line is as always reserved for Colin, who now knows more about space-time coding than any sane person really wants to. My thanks for his most valiant efforts in proof-reading and more importantly "dummy-proofing" the material from the perspective of someone outside of the field. Also for providing me with a skeleton web library, sharing LaTeX and linear algebra tricks, teaching me about constrained optimization and digital filtering, putting up with me for the last few weeks, and the list goes on! Don't worry, one of these days I'll finally put a progress bar in...

Dimidium facti qui bene coepit habet - if only I should be so fortunate!

Karen Su
Cambridge, 2003.

Contents

List of Tables

List of Figures

List of Acronyms

AWGN	Additive White Gaussian Noise
BLAST	Bell Labs lAyered Space-Time
BPSK	Binary Phase Shift Keying
CCDF	Complementary Cumulative Distribution Function
CIR	Channel Impulse Response
CSI	Channel State Information
DFE	Decision-Feedback Equalizer
DFT	Discrete Fourier Transform
DMT	Discrete Multi-Tone
DMMT	Discrete Matrix Multi-Tone
EVD	EigenValue Decomposition
FDE	Frequency Domain Equalization
IBI	Inter-Block Interference
i.i.d.	Independent Identically Distributed
ISI	Inter-Symbol Interference
LTI	Linear Time-Invariant
LOS	Line Of Sight
MEA	Multi-Element Antenna
ML	Maximum Likelihood
MLSE	Maximum Likelihood Sequence Estimation
MMSE	Minimum Mean Squared Error
MRC	Maximal Ratio Combining
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
MUD	Multi-User Detection
OFDM	Orthogonal Frequency Division Multiplexing
PEP	Pair-wise Error Probability
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
SINR	Signal-to-Interference-and-Noise Ratio
SNR	Signal-to-Noise Ratio
SIMO	Single Input Multiple Output
SISO	Single Input Single Output
STC	Space-Time Code
STBC	Space-Time Block Code
STTC	Space-Time Trellis Code
SVD	Singular Value Decomposition
VA	Viterbi Algorithm

List of Symbols

τ_{coh}	Coherence time
B_{coh}	Coherence bandwidth
$\widehat{\sigma}_{\tau}$	RMS delay spread
d_{coh}	Coherence distance
T_s	Symbol period
N	Number of transmit antennas
M	Number of receive antennas
L	Number of symbol periods (per block)
Q	Number of distinct symbols (per block)
P	Total transmitted power (independent of N)
N_0	Noise power (of complex noise at each receive antenna)
$\mathcal{X}$	Symbol alphabet
x	Signal from alphabet $\mathcal{X}$ (scalar)
$\mathbf{x}$	Signals from alphabet $\mathcal{X}$ ($Q \times 1$ vector)
$s[l]$	Transmitted signal (scalar)
$\mathbf{s}[l]$	Transmitted signals ($N \times 1$ vector)
$\widetilde{\mathbf{S}}^{Z}$	Transmitted space-time signal matrix ($N \times Z$ matrix)
$r[l]$	Received signal (scalar)
$\mathbf{r}[l]$	Received signals ($M \times 1$ vector)
$\widetilde{\mathbf{R}}^{Z}$	Received space-time signal matrix ($M \times Z$ matrix)
$h[l]$	Channel fading coefficient (scalar)
$\mathbf{H}[l]$	Spatial channel fading coefficients ($M \times N$ matrix)
$\widetilde{\mathbf{H}}^{Z}$	Time-time channel fading coefficients ($Z \times Z$ matrix)
$n[l]$	Additive noise (scalar)
$\mathbf{n}[l]$	Additive noise ($M \times 1$ vector)
$\widetilde{\mathbf{N}}^{Z}$	Additive noise space-time matrix ($M \times Z$ matrix)
$\overline{s}[l]$	Detected signal (scalar, soft decision)
$\overline{\mathbf{s}}[l]$	Detected signal ($N \times 1$ vector, soft decision)
$\widehat{s}[l]$	Detected signal (scalar, hard decision)
$\widehat{\mathbf{s}}[l]$	Estimated signal ($N \times 1$ vector, hard decision)

Chapter 1

Introduction

Driven by the demand for increasingly sophisticated communication services available anytime, anywhere, wireless communications has emerged as one of the largest and most rapidly growing sectors of the global telecommunications industry. A quick glance at the status quo reveals that over 700 million people around the world subscribe to existing second and third generation cellular systems supporting data rates of 9.6 Kbps to 2 Mbps. More recently, IEEE 802.11 wireless LAN networks, enabling communication at rates of around 11 Mbps, have attracted more than $1.6 billion (USD) in equipment sales [12]. Over the next ten years, the capabilities of both of these technologies are expected to move toward the 100 Mbps - 1 Gbps range [32] and subscriber numbers to over 2 billion [44]. One of the most significant technological developments of the last decade, that promises to play a key role in realizing this tremendous growth, is wireless communication using MIMO antenna architectures.

The study of radio wave propagation was initiated by the works of Hertz and Marconi in the late 1800s. These experiments demonstrated that electrical signals could be transmitted via electromagnetic waves travelling at the speed of light. They can be described using the term Single Input Single Output (SISO), since they involve one circuit radiating energy into space, and another electrically disconnected circuit collecting this energy at some distance away. The SISO system model represents the space or wireless channel through which the electromagnetic wave travels to reach its destination. In Chapter 2, we will take a closer look at what researchers since then have discovered about the nature of propagation over the wireless channel, and how some typical communication environments are modelled in the literature.

In a MIMO system, Multi-Element Antenna (MEA) structures are deployed at both the transmitter and receiver.[1] From a communications engineering perspective, the challenge is to design the signals to be sent by the transmit array and the algorithms for processing those seen at the receive array, so that the quality of the transmission (i.e., bit error rate) and/or its data rate are improved. These gains can then be used to provide increased reliability, lower power requirements (per transmit antenna) or higher composite data rates (either higher rates per user or more users per link). What is especially exciting about the benefits offered by MIMO technology is that they can be attained *without the need for additional spectral resources*. In the last five years, the greatly enhanced performance that is possible over realistic fading channels has been shown both theoretically and demonstrated in experimental laboratory settings. Hence the recent explosion of interest from both academic and industrial researchers in the area of *space-time coding*.

[1] Since this project will not be concerned with antenna architectures and geometries, we will assume a MEA comprised of a linear array of uniformly spaced antenna elements and refer to it simply as an array.

Historically, work on transmit diversity techniques began as early as 1993. In [50], the authors consider transmitting delayed copies of the information-bearing signal on each antenna in order to obtain a diversity gain at the receiver. A more generalized approach presented in [63] proposes the use of a bank of linear time invariant precoding filters at the transmitter, combined with Maximum Likelihood (ML) detection at the receiver, to achieve the desired diversity gain. Up to this point, it had been well-known that a diversity gain proportional to the number of antennas at the receive array could be achieved using Maximal Ratio Combining (MRC) without any bandwidth expansion. These experiments were among the first to demonstrate that a diversity gain proportional to the number of antennas at the transmit array was also possible under certain channel conditions. They also highlight one of the main features of MIMO communication systems, which is the ability to benefit from the effects of multipath signal propagation. In Chapter 3, we will provide a more rigourous definition of diversity gain and study some of these approaches in more detail.

Naturally, the next issue that researchers tackled was determining the fundamental limits of multiple transmit antenna technologies. In 1995, a seminal work on the capacity of multiple antenna channels by Telatar [61] presented analytical equations for the mean and instantaneous capacities of wireless channels in flat, quasi-static, and spatially independent Rayleigh fading with perfect Channel State Information (CSI) at the receiver. These results were independently derived and extended with practical considerations by Foschini *et al.* [17]. The main finding of these information theoretic analyses was that in such a fading environment, the capacity of multiple antenna channels increases linearly with the smaller of the number of transmit and receive antennas. This and other major results relating to the capacity of MIMO channels in Rayleigh fading will be summarized Chapter 4.

The term Space-Time Code (STC) was originally coined in 1998 by Tarokh *et al.* to describe a new two-dimensional way of encoding and decoding signals transmitted over wireless fading channels using multiple transmit antennas [60]. In two key papers, the authors laid down the theories of the Space-Time Trellis Code (STTC) [60] and the Space-Time Block Code (STBC) [57] for flat independent Rayleigh fading channels. A number of other schemes employing multiple antenna arrays were also developed at about the same time, e.g., the simple and popular Alamouti STBC [3], a transmit diversity scheme using pilot symbol-assisted modulation [28] and the Bell Labs lAyered Space-Time (BLAST) multiplexing framework [15]. Since then, the term STC has been used more generally to refer to transmit diversity techniques in which the transmitted signals and corresponding receiver are designed to exploit spatial diversity. A more detailed overview of some fundamental techniques, along with a brief survey of core contributions to the field, can be found in Chapter 5.

In the closing chapter of this report, we will present a proposal for research to be undertaken during the next two years of our Ph.D. project. The main topic that we intend to address is *space-time block coding for frequency selective fading channels*. One key application area for this work is in low mobility and fixed broadband wireless access systems. Through this report we demonstrate that there is room for new developments in this area by approaching the problem from a capacity perspective. A table of completed and projected activities will also be provided.

Chapter 2

Signal fading and structures

When communicating over a wireless radio channel the received signal cannot be modelled simply as a copy of the transmitted signal corrupted by additive Gaussian noise. Instead, we observe signal *fading*, which can be defined as variations in the magnitude and/or phase of one or more frequency components, caused by the possibly time-varying characteristics of the propagation environment.

These variations can be divided into two categories: large and small scale signal fading. The first encapsulates long-term changes caused by environmental elements, such as shadowing by buildings and natural features or rain. Degradations of this type are heavily system-, application- and even terrain-dependent. Thus, solutions addressing large scale fading tend to be at the system protocol level (e.g., basestation placement, power control), whereas our focus is on enhancing performance through signal processing at the physical layer. Of more interest when designing widely-applicable space-time coding algorithms are the effects of small scale fading. These short-term fluctuations in the received envelope are caused by signal scattering off objects in the propagation environment.

This scattering leads to a phenomenon known as *multipath propagation*, where the received signal is comprised of a number of constructively and destructively interfering copies of the transmitted waveform. These copies are also referred to as multipaths, since they propagate along different paths to reach their destination. Because the properties of the transmission environment vary from path to path, each copy experiences different attenuations, phase shifts, angles of arrival, Doppler shifts, and time delays. In this report, we will be concerned primarily with the attenuations, arrival angles and excess delays, i.e., delays relative to the first arriving multipath, as these enable us to develop a simple yet relevant model of the wireless channel.

By considering the attenuations and excess delays of the arriving multipaths, we can characterize the fading channel as a linear system, where the length of its impulse response corresponds to the maximum excess delay. At any given time, the frequency response of the channel indicates how it affects transmitted signal components at different frequencies. Over time, the Channel Impulse Response (CIR) may change, leading to a time-variant linear system model. Finally, in the case of a MIMO transmission system, the impulse responses of the channels between each transmit-receive antenna pair may also be different. The spreads of departure angles from the transmit antenna and arrival angles at the receive antenna influence the relationships between these responses, as explained in Section 2.1.3.

The overall effect of this linear time-variant space-variant fading channel on a transmitted signal can be studied more easily by considering each of the domains in turn. Table 2.1 summarizes the kinds of small scale signal fading that may be encountered in the MIMO

wireless channel. The criteria used in defining each of the fading types is expressed in fairly standard notation, with the relevant system characteristics as follows:

- Symbol period T_s. Inverse of the symbol rate R_s.
- Signal bandwidth W. Nominal passband bandwidth needed to transmit the modulated signal, i.e., $W = R_s = \frac{1}{T_s}$.
- Inter-element distance d. MEA separation, assuming uniform linear array structure.
- Block length L. Length of transmitted block in symbols.

The propagation environment is defined by three parameters:

- Coherence time $\tau_{coh} \approx \frac{1}{\text{Doppler spread}}$.
- Coherence bandwidth $B_{coh} \approx \frac{1}{\text{rms delay spread}}$.
- Coherence distance d_{coh}.

Domain	**Type of signal fading**		
Time	Slow fading $\tau_{coh} \gg T_s$	Quasi-static fading $\tau_{coh} \approx LT_s$	Fast fading $\tau_{coh} < T_s$
Frequency	Flat (frequency non-selective) fading $B_{coh} \gg W$		Frequency selective fading $B_{coh} < W$
Space	Spatially correlated fading $d_{coh} \gg d$		Spatially independent fading $d_{coh} < d$

Table 2.1: Types of small scale signal fading and their defining criteria.

In the remainder of this chapter we review these parameters and how they relate to the definitions of the fading types outlined in Table 2.1. We also elaborate on their relevance to broadband fixed wireless systems. Then we define the mathematical notation, probability distributions and fading signal models that will be used throughout this report. In particular we pay special attention to the flat quasi-static, frequency selective, and spatially correlated fading scenarios. A more comprehensive study can be found in [6].

2.1 Channel parameters

In this section, we consider the three parameters described previously in terms of the domains in which they reside. Before we begin, it will be useful to briefly introduce the Channel Impulse Response (CIR) function, as well as the notions of coherence and selectivity.

In its most general form, the impulse response of a MIMO channel $h_{ij}(t, \tau)$ is a function of time, delay, and transmit and receive antenna positions, or as we show here, the indices of a particular pair of transmit and receive antennas on their respective uniform linear arrays. This complex-valued function describes the response of the channel seen by receive antenna i at time t to a unit impulse transmitted τ time units in the past from transmit antenna j. It is a generalization of the input delay-spread function that is most often used to represent the impulse response of a SISO channel.

The time domain is reflected by t, and as we shall see the frequency domain by the delay variable τ. The spatial domains at the transmit and receive arrays are captured by the indices j and i. In each of these domains, the fading channel is characterized by a *coherence* interval. Over this interval, the channel model is considered to be flat or invariant. Outside of this interval, the channel's response in that domain may change arbitrarily. Observe that coherence is a characteristic of the channel and does not depend on the properties of the signals being transmitted.

When we consider the transmission of a specific signal, it is clear that the properties of that signal play a role in determining whether the channel's effect on it are invariant in any given domain. This relationship between the coherence of the channel the properties of the signal, are captured by the notion of *selectivity*. If the channel is selective, then the region of support[1] of the transmitted signal is larger than the coherence interval. Thus the channel is not flat with respect to the signal in that domain. Conversely if the channel is not selective, then it is invariant with respect to the transmitted signal.

Observe that along each of the domains or rows of Table 2.1, selectivity increases through the columns from left to right. In particular, note that correlated channels have relatively low selectivity, while independent channels exhibit high selectivity in their respective domains. As we shall see in Chapters 3 and 4, selectivity is essential for obtaining increased diversity gains and capacities, and correlation is a property that reduces the achievable benefits.

2.1.1 Time selectivity

The coherence time τ_{coh} is the time difference at which the magnitude or envelope correlation coefficient between two signals at the same frequency falls below 0.5. In other words it can be assumed that the two signal components separated in time by τ_{coh} will undergo independent attenuations. Thus a signal experiences *slow* or *time non-selective* fading if its symbol period T_s is much smaller than the channel coherence time, and *fast* or *time selective* fading if $T_s > \tau_{coh}$. When a signal is slow fading, we can assume that the CIR is time invariant during a block transmission.

Another related type of time-oriented fading that is commonly used in signal analyses for space-time coding, especially when dealing with block-based algorithms, is *quasi-static* fading. In this case, the coherence time is on the order of LT_s. The channel attenuation is assumed to be constant over each block, but changes independently from block to block. In this project, we will be working almost exclusively with quasi-static fading channels. This assumption greatly simplifies analysis since it enables the channel to be modelled as an Linear Time-Invariant (LTI) system within each block.

Since the channel coherence time can be approximated by taking the inverse of the Doppler spread, it relates to the degree of mobility in the propagation environment. In particular, although there may be some motion of the scatterers, the fixed wireless channel has negligible Doppler spread, or equivalently a large coherence time, and is therefore a slow fading channel. In the context of mobile communications, since the symbol period is also the inverse of the signal bandwidth or symbol rate, fast fading arises in cases where the bandwidth is smaller than the Doppler spread of the channel, i.e., when the symbol rate is low and the mobile unit is moving rapidly. As the data rate increases and/or the mobile unit's mobility decreases, the channel becomes more slowly fading. Also note that the maximum Doppler shift is proportional to the carrier frequency. Thus time selectivity also becomes a more important consideration as transmission frequencies increase.

[1]The *region of support* of a function $f(x)$ is defined as the set $\mathscr{X} = \{x | f(x) \neq 0\}$. Also, in this discussion we define the size of such a region to be $\max_{x \in \mathscr{X}}(x) - \min_{x \in \mathscr{X}}(x)$.

2.1.2 Frequency selectivity

The coherence bandwidth B_{coh} captures the analogous notion for two signals of different frequencies transmitted at the same time. A signal experiences *flat* or *frequency non-selective* fading if its bandwidth W is much smaller than the channel coherence bandwidth, and *frequency selective* fading if $W > B_{coh}$.

Another way of thinking about frequency selectivity arises when considering the inverse Fourier transform of the channel frequency response. If the channel is frequency non-selective, then its frequency response is flat over the bandwidth of interest, or equivalently its impulse response is a scaled Dirac delta function. In this case its maximum excess delay τ_{max} is smaller than a symbol period. Although there are still many multipaths arriving at the receive antenna, we say that they are not *resolvable* or that only one significant multipath component can be resolved by the receiver. Effectively, the symbol period over which the receiver accumulates and then samples the signal energy, determines the resolution of the transmission system. Since all of the multipath images arrive at delays more closely spaced than T_s, they are blended into a single received sample.

In the frequency selective fading case, the response of the channel is not flat and its maximum excess delay is larger than T_s. A transmitted signal is sufficiently dispersed in time, so that the resolution of the receiver enables it to be seen during multiple symbol periods. Therefore we say that there is more than one resolvable or significant multipath component. Such channels are also known in the literature as *dispersive* or *multipath fading* channels. Note that in this report, we distinguish multipath fading from *multipath propagation*. We will use the latter term to refer to the scattering environment through which all wireless signals travel. When discussing a multipath fading channel, we will mean that more than one multipath may be resolved by the receiver.

So far we have considered how τ_{max}, which describes the maximum time dispersion of the transmitted signal, relates to frequency selectivity. Another quantity that is commonly used to capture dispersion is the *rms delay spread*, defined as the square root of the (normalized) second central moment of the *power delay profile* of the channel. The power delay profile $p_{h_{ij}}(t,\tau) = p_h(\tau) = |h_{ij}(t,\tau)|^2$ is generally assumed to be independent of time and space. It describes the power seen at the receiver as a function of delay, and is often used to specify standardized channel models in the literature. When simulating frequency selective channels, the delay profile is typically normalized so that $\int_0^{\tau_{max}} p_h(\tau)\, d\tau = 1$.

The rms delay spread is then given by

$$\begin{aligned} \widehat{\sigma}_\tau &= \sqrt{\frac{\int_0^{\tau_{max}} (\tau - \overline{\tau})^2 p_h(\tau)\, d\tau}{\int_0^{\tau_{max}} p_h(\tau)\, d\tau}}, \\ \overline{\tau} &= \frac{\int_0^{\tau_{max}} \tau p_h(\tau)\, d\tau}{\int_0^{\tau_{max}} p_h(\tau)\, d\tau}. \end{aligned} \tag{2.1}$$

The channel coherence bandwidth may be approximated by taking the inverse of the rms delay spread.[2] In absolute terms, the extent of the time dispersion induced by fading is an intrinsic property of the channel. The relative degree of dispersion becomes more severe as the transmitted symbol period decreases, i.e., as the signal bandwidth increases. For instance, observe that the same amount of absolute delay results in dispersion over a larger number of symbol periods as T_s decreases. Thus the broadband wireless channel tends to be a frequency selective fading channel.

[2]Note that some authors use the inverse of the maximum excess delay as an approximation of the channel coherence bandwidth. Since they are both approximations, the use of either definition is acceptable.

2.1.3 Spatial selectivity

When using MEA arrays, the coherence distance represents the minimum distance in space separating two antenna elements such that they experience independent fading. This distance clearly depends on the wavelength λ of the transmitted signal, as the phases of higher frequency signals are more sensitive to small distance changes than those of low frequency signals. Thus antenna separations and coherence distances are typically expressed in terms of wavelengths of the carrier signal. d_{coh} also depends on the presence of scatterers in the vicinity of the antenna array, and generally falls somewhere between $\frac{\lambda}{4}$ and 12λ.[3]

To understand why this is the case, we model the multipath propagation environment by representing the scatterers as sources, each reflecting the signal away from the transmit array and toward the receive array with some attenuation and delay. In a rich scattering environment, where many scatterers are approximately uniformly distributed over $[0, 2\pi)$ around the receive array, the envelope correlation between two signals seen at antennas separated by $d > \frac{\lambda}{4}$ is less than 0.5 [11]. Thus, given this angle spread and antenna separation, the channel exhibits *independent* or *spatially selective* fading.

As the spread of departure angles of the multipaths from the transmit array or arrival angles at the receive array decreases, it has been shown that the envelope correlation coefficient increases, in other words the coherence distance of the channel also increases [14]. When the correlation coefficient is greater than 0.5 for antennas separated by the desired distance,[4] the channel is said to exhibit *correlated* fading. This case is more typical for fixed broadband wireless access solutions, especially in rural deployments, and results in lower achievable capacity.

Observe that for the time and frequency domains, fading channels are modelled as either non-selective (invariant) or selective (changing independently) in the space-time coding literature. However for the spatial domain, we also consider channels that fall somewhere in the gray-region between full and partial selectivity. In the former case, the channel attenuations seen by two distinct receive antennas are statistically independent. In the latter case, there will be some non-zero correlation coefficient between these two signals, hence the notion of spatially correlated channels.

Table 2.2, adapted from [41], summarizes some typical wireless propagation environments and the parameters associated with the corresponding fading channels.

2.2 Mathematical notation

In this report, scalars will be denoted by lowercase letters: s_j^l and r_i^l for the transmitted and received signals respectively, h_j^l for the channel attenuation, and n_i^l the noise. Where appropriate, the spatial dimension, i.e., antenna number, will be indicated in the subscript, with the receive array index i followed by that of the transmit array j if both are required. The time index may be specified in the superscript or as a parameter $(t,\ l)$, where the discrete time index l corresponds to a sample at time $t = lT_s$.

Since the transmitted symbols do not necessarily co-incide with the transmit antennas or symbol periods in their number or order, we will reserve the variable s_j^l for use when

[3] In practice, an antenna separation of $\frac{\lambda}{2}$ is commonly chosen for subscriber units in rich local scattering environments with angle spreads of 360°. For urban roof-top and rural basestations with channels characterized by angle spreads on the order of 10°, d_{coh} is typically assumed to be around 10λ.

[4] For instance, the maximum possible antenna separation may be constrained by the physical size of the transceiver unit.

Environment	**Mobility**	**Doppler spread**	**Delay spread**	$R_{s,max}$ **for flat fading**[a]	**Angle spread**	**Spatial correlation**
Rural	High	190 Hz	0.5 μs	200 Kbaud	1°	High
Urban	Medium	120 Hz	5 μs	20 Kbaud	20°	Medium
Hilly	High	190 Hz	20 μs	5 Kbaud	30°	Medium
Mall	Low	10 Hz	0.3 μs	350 Kbaud	120°	Low
Office	Low	5 Hz	0.1 μs	1 Mbaud	360°	Low

[a]As a rule of thumb, a flat fading channel can be assumed if the symbol rate is less than $\frac{1}{10\widehat{\sigma}_\tau}$ [40].

Table 2.2: Typical wireless propagation environments and their associated parameters. The mobility and Doppler spread values apply to mobile systems, with approximate values given for a carrier frequency of 2.5 GHz.

referring to a symbol transmitted from antenna j during symbol period l. We will also make use of the following notation $x_q \in \mathcal{X},\ q = \{1, \ldots, Q\}$ to refer to the symbols independently of spatio-temporal position. They are chosen from alphabet $\mathcal{X} \subset \mathbb{C}$ of size $B = \log_2 |\mathcal{X}|$. At the receiver, soft decisions for the detected symbols (by antenna or generic symbol index) will be denoted by the corresponding overstruck transmitted symbol ($\overline{s}$, $\overline{x}$), and hard decisions covered with a hat ($\widehat{s}$, $\widehat{x}$).

There are a number of different types of vectors and matrices used in the space-time coding literature, which can be a source of confusion for readers. In this report we will distinguish between three kinds of vectors and four kinds of matrices (presented in order of appearance):

First, vectors whose elements span a non-temporal dimension, e.g., a set of signals transmitted at the same time from a MEA or an independent sequence of signals, will be denoted by boldface lowercase letters:

$$\mathbf{s} = \left[\begin{array}{cccc} s_1 & s_2 & \cdots & s_N \end{array} \right]^T,$$

$$\mathbf{x} = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_Q \end{array} \right]^T.$$

Spatial matrices representing a transformation between spatial dimensions, e.g., a mapping from the transmit array space to the receive array space, will be denoted by boldface uppercase letters ($\mathbf{H}$). In this case, the entries all share the same time index. When discussing receive diversity, we will be interested in considering the columns of the channel matrix $\mathbf{H}$, which will be denoted by $\mathbf{h}_j$, and to study transmit diversity, the rows of $\mathbf{H}$ will be denoted $\overline{\mathbf{h}}_i^T$. Observe that under this notation all vectors are column vectors by definition.

To represent a block of signals transmitted from each of the elements of an array over a number of symbol periods, space-time matrices are used. The rows of these matrices span a spatial dimension and their columns span time. They will be denoted by boldface uppercase letters covered with a tilde ($\widetilde{\mathbf{S}}$). The size of the matrix in the time dimension may be indicated in the superscript, or take the default value of the block length L. The corresponding time $\times$ time channel matrix ($\widetilde{\mathbf{H}}$) will be similarly annotated.

When working with frequency selective fading channels, signals will be expanded in the time dimension by the memory of the channel equivalent linear filter. To denote vectors whose elements span the time dimension, i.e., signals transmitted over a number of symbol periods from the same antenna, we will use boldface lowercase letters covered with a tilde.

As in the case of the space-time matrix, the length of the vector may be indicated in the superscript, or take the default value of L. In addition, the starting time index may be given as a parameter. For instance,

$$\widetilde{\mathbf{s}}_j \;=\; \widetilde{\mathbf{s}}_j^L \;=\; \left[\begin{array}{cccc} s_j^0 & s_j^1 & \ldots & s_j^{L-1} \end{array}\right]^T.$$

Transmission over MIMO frequency selective fading channels is expressed in two ways in the literature: as a linear combination of SISO channels [2] or using more compact block notation [26]. We will denote by a calligraphic letter $\mathcal{A} = \mathrm{Vec}(\widetilde{\mathbf{A}})$ the block vector obtained by stacking the columns of space-time matrix $\widetilde{\mathbf{A}}$. As before, the resulting space-time signal vector may be annotated by a superscript indicating the size of the source matrix in the time dimension and the starting time index may be given as a parameter.

To apply linear transformations to these space-time vectors, the relevant matrices may also have to be expanded. Wherever possible we will make use of the compact Kronecker product notation:

$$\mathbf{A} \otimes \mathbf{B} \;=\; \left[\begin{array}{ccc} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{array}\right],$$

where $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{A} \otimes \mathbf{B}$ are of sizes $m \times n$, $p \times q$ and $mp \times nq$, respectively. Singly time-expanded matrices, e.g., space $\times$ space-time, will be denoted by script letters ($\mathscr{H}$), annotated by a superscript as necessary. Doubly time-expanded matrices, e.g., space-time $\times$ space-time, will be denoted by script letters covered with a tilde ($\widetilde{\mathscr{H}}$).

In addition to studying blocks transmitted over a number of symbol periods, we will also be interested in considering transmissions spread over a number of frequency sub-channels, i.e., developing signal structures appropriate for multi-carrier modulation schemes. We will use L_f to denote the number of frequency sub-channels. The ensuing *space-frequency* signal matrices are straightforward duals of their space-time counterparts and will be indicated by a subscript $_f$, e.g., $\widetilde{\mathbf{S}}_f$.

The elements of these structures are scalar (discrete) Fourier transform coefficients representing the corresponding time domain signals in the frequency domain. They will be differentiated through the use of blackboard letters ($\mathbb{s}_j$, $\mathbb{h}_{ij}$), with the spatial dimension indicated in the subscript and the sub-channel index or frequency dimension specified as a parameter where appropriate (f).

Some other notations and operations that will be used in this report are summarized in Table 2.3

2.3 Statistical models for fading signals

At the beginning of this chapter, we briefly introduced the concepts of signal scattering and multipath. Because the combined effect of the scattered signals cannot easily be expressed in closed form, more tractable statistical descriptions for the resulting fading coefficients have been derived based on the nature of signal propagation in the wireless environment.

As there are a large number of scatterers in the wireless channel, the central limit theorem may be applied to obtain a limiting probability distribution for the composite received signal. Therefore, if there is no Line Of Sight (LOS) path from the transmitter to the receiver, we expect the real and imaginary parts of the complex baseband channel coefficients to be zero-mean Gaussian processes. Their magnitudes can then be modelled

x^*	Complex conjugate of x
x_{ij}	Element in the i^{th} row and j^{th} column of $\mathbf{X}$
$\mathbf{x}_{:j}$	j^{th} column of $\mathbf{X}$
$\mathbf{x}_{i:}$	i^{th} row of $\mathbf{X}$
$\mathbf{X}^\dagger$	Complex conjugate (Hermitian) transpose of $\mathbf{X}$
$\mathbf{X}^{-1}$	Inverse of $\mathbf{X}$
$\mathbf{X}^+$	Moore-Penrose pseudoinverse of $\mathbf{X} = (\mathbf{X}^\dagger\mathbf{X})^{-1}\mathbf{X}^\dagger$
$\lambda_i(\mathbf{X})$	Eigenvalues of $\mathbf{X}$ (parameter may be omitted where clear from context)
$\det \mathbf{X}$	Determinant of $\mathbf{X} = \prod_{i=1}^{n} \lambda_i$
$\operatorname{rank} \mathbf{X}$	Rank of $\mathbf{X}$ = Number of non-zero eigenvalues (or singular values)
$\operatorname{tr} \mathbf{X}$	Trace of $\mathbf{X} = \sum_{i=1}^{n} x_{ii}$
$\operatorname{diag} \mathbf{X}$	Diagonal elements of $\mathbf{X} = [x_{11} \ \ldots \ x_{nn}]$
$\operatorname{diag} \mathbf{x}$	Diagonal matrix constructed from elements of $\mathbf{x} = \mathbf{D}$ *s.t.* $d_{ii} = x_i$
$\mathbf{I}_n$	$n \times n$ identity matrix
$\mathbf{\Lambda}$	Diagonal matrix of eigenvalues
$\mathbf{\Sigma}$	Diagonal matrix of singular values
$\mathbf{U}$,$\mathbf{V}$	Unitary matrix
$\mathscr{F}_d\{\mathbf{x}\}$	Discrete Fourier Transform (DFT) of $\mathbf{x}$

Table 2.3: Summary of mathematical notation and operations.

according to a *Rayleigh* probability distribution and their phases are uniformly distributed over $[0, 2\pi)$.

The Rayleigh probability density function is written in terms of a parameter $\sigma > 0$, which corresponds to the standard deviation of the constituent real and imaginary Gaussian components:

$$f(\alpha) = \begin{cases} \frac{\alpha}{\sigma^2} e^{-\frac{\alpha^2}{2\sigma^2}}, & \alpha \geq 0 \\ 0, & \alpha < 0 \end{cases}.$$

Its mean value is $E(\alpha) = \sqrt{\frac{\pi}{2}}\sigma$ and its average power is $E(\alpha^2) = 2\sigma^2$.

In some cases, particularly in some existing commercial fixed wireless systems, there is a LOS path from the basestation to the subscriber units. The appropriate statistical model in this case is known as the *Rice* distribution. Its probability density function is given by

$$f(\alpha) = \begin{cases} \frac{\alpha}{\sigma^2} e^{-\frac{\alpha^2+A^2}{2\sigma^2}} I_0(\frac{A\alpha}{\sigma^2}), & \alpha \geq 0 \\ 0, & \alpha < 0 \end{cases},$$

where A is the peak amplitude of the dominant signal and $I_0(\cdot)$ is the zeroth modified Bessel function of the first kind. The Ricean channel is sometimes described using the *K-factor* $K = \frac{A^2}{2\sigma^2}$, which is the ratio of the power of the dominant signal, or *specular* component, to that of the scattered signals, or Rayleigh component. Observe that when $K = 0$ the Ricean distribution becomes the Rayleigh distribution.

Finally, the *Nakagami-m* distribution is an alternative two-parameter statistical model for the envelope of the channel response. Although it has been shown that this distribution provides the best fit for urban radio channels [55], we will likely use the Rayleigh distribution

in our work because of its simpler form and more common application in the literature. We include the Nagakami-m probability density function here for completeness:

$$f(\alpha) = \begin{cases} \frac{2}{\Gamma(m)} \left(\frac{m}{2\sigma^2}\right)^m \alpha^{2m-1} e^{-\frac{m\alpha^2}{2\sigma^2}}, & \alpha \geq 0 \\ 0, & \alpha < 0 \end{cases},$$

where σ represents the power of the real and imaginary Gaussian components as before and the *fading figure* is

$$m = \frac{(2\sigma^2)^2}{\mathsf{E}\left([\alpha^2 - 2\sigma^2]^2\right)}, \; m \geq \frac{1}{2}.$$

More detailed information and experimental results on the Nakagami-m distribution can be found in [55] and the references therein. We note that it reduces to the Rayleigh distribution when $m = 1$.

2.4 System models for fading channels

In this section we overview the key mathematical structures and features underlying three important wireless channel models. First we consider the quasi-static flat fading channel, which is the most popular in the literature because of its simplicity of analysis and relevance to narrowband communication schemes. Next, the quasi-static frequency selective fading channel is of interest, because its properties are more appropriate for broadband systems. Finally, since spatial correlation is found to some degree in all MIMO transmission systems, we will take a look at how such correlation is modelled mathematically.

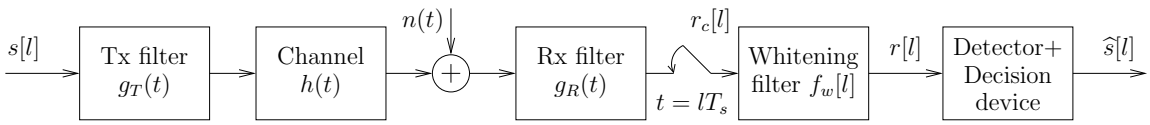
There are three standard assumptions that we make throughout this report about the channel impulse response function. First we assume that it is *wide-sense stationary*. This means that its autocorrelation is a function only of relative time differences and not absolute time references. Secondly we assume that it experiences *uncorrelated scattering*. This assumption is equivalent to wide-sense stationarity of the channel response in the frequency domain [54, Chapter 2]. This results in uncorrelatedness between the responses seen at different delays. Finally we make an analogous assumption in the spatial domain, that the spatial correlation between two transmit/receive antenna pairs is dependent only on their relative positions, i.e., indices on the uniform linear arrays, and not on their absolute positions.

In addition, we make the assumption that even if the channel itself is time-varying, its power delay profile remains time invariant as well as spatially invariant. These assumptions lead to a more tractable form for the autocorrelation function

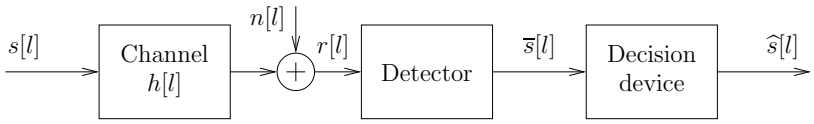
$$\begin{aligned} r_h(i_1, i_2, j_1, j_2, t_1, t_2, \tau_1, \tau_2) &= \mathsf{E}\left[h^*_{i_1 j_1}(t_1, \tau_1) h_{i_2 j_2}(t_2, \tau_2)\right] \\ &= \mathsf{E}\left[\sqrt{p_h(\tau)}\widehat{h}^*_{i_1 j_1}(t_1, \tau)\sqrt{p_h(\tau)}\widehat{h}_{i_2, j_2}(t_2, \tau)\right] \delta(\tau_2 - \tau_1) \\ &= p_h(\tau_1)(\mathbf{R}_M)_{i_2, i_1}(\mathbf{R}_N^\dagger)_{j_2, j_1} r_h(t_2 - t_1)\delta(\tau_2 - \tau_1), \end{aligned}$$

where $\widehat{h}$ are normalized random variables having unit variance, $\mathbf{R}_M$ and $\mathbf{R}_N$ are receive and transmit (spatial) correlation matrices that will be derived in Section 2.4.3, and $r_h(\theta)$ represents the temporal autocorrelation induced by fast fading. These reduce to $\delta(i_2 - i_1)$, $\delta(j_2 - j_1)$ and 1 respectively in the case of quasi-static spatially uncorrelated fading.

In this report, we will be working with the complex baseband representation of discrete time signals. For completeness, Figure 2.1 shows the applicable communication system diagrams and where the signals of interest can be found.



(a) Full system diagram, continuous time.



(b) Simplified system diagram, discrete time.

Figure 2.1: Complex baseband communication system diagrams.

The sampled received signals (before and after whitening) can be expressed as

$$\begin{aligned} r_c[l] &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(\tau) \sum_{k=-\infty}^{\infty} s[k] g_T(\theta - \tau - kT_s)\, d\tau + n(\theta) \right] g_R(lT_s - \theta)\, d\theta \\ &= \sum_{k=-\infty}^{\infty} h_c[l-k]s[k] + n[l]_c \\ r[l] &= \sum_{p=0}^{P_w-1} f_w[p] z^{-p} \left(\sum_{k=-\infty}^{\infty} h_c[l-k]s[k] + n_c[l] \right) \\ &= \sum_{k=-\infty}^{\infty} s[k] \sum_{p=0}^{P_w-1} f_w[p] h_c[l-p-k] + \sum_{p=0}^{P-1} f_w[p] n_c[l-p], \end{aligned}$$

where the discrete time whitening filter is of length P_w and the discrete time, and for quasi-static fading time invariant, channel impulse response is defined by the discrete filtered convolution ($\star$)

$$h[l] = \sum_{p=0}^{P-1} f_w[p] \iint_{-\infty}^{\infty} h(\tau) g_T(\theta) g_R([l-p]T_s - \tau - \theta)\, d\theta\, d\tau.$$

With these definitions in hand, the simplified equivalent system diagram shown in Figure 2.1(b) will be used for the remainder of this report. See Appendix D for a more detailed discussion.

2.4.1 Flat quasi-static fading channel

The simplest channel model is that of flat quasi-static fading. In this case the path attenuations do not vary over the duration of the block transmission, thus the channel becomes an

LTI system within the block. The attenuations also do not vary over the spectrum of the transmitted signal, and as discussed in Section 2.1.2, its impulse response is a scaled Dirac delta function. Thus the process can be modelled mathematically as a single-tap filter with complex coefficient $h[l]$, or equivalently by real path attenuation $\alpha[l]$ and phase shift $\theta[l]$:

$$\begin{aligned} r[l] &= \alpha[l]e^{j\theta[l]}s[l] + n[l] \\ &= h[l]s[l] + n[l]. \end{aligned} \tag{2.2}$$

Equation 2.2 actually illustrates the possibly time selective fading case with one transmit antenna and one receive antenna. Because the channel under consideration is slowly fading, the coefficients will be independent of time, i.e., $h[l] = h$. Also note that we have dropped the delay and spatial parameters to simplify the notation.

In the MIMO case, we may have M receive antennas and N transmit antennas, leading to the following matrix equation:

$$\mathbf{r} = \mathbf{Hs} + \mathbf{n}, \tag{2.3}$$

where $\mathbf{r}$ is an $M \times 1$ vector of signals received at each antenna, $\mathbf{s}$ is an $N \times 1$ vector of signals transmitted by each antenna, and $\mathbf{n}$ is an $M \times 1$ vector of complex Additive White Gaussian Noise (AWGN) signals seen at each receive antenna. The channel matrix $\mathbf{H}$ is an $M \times N$ matrix, whose elements h_{ij} represent the complex fading coefficients experienced by a signal transmitted from transmit antenna j to receive antenna i.

In the case of block transmission over a flat quasi-static fading channel, the full transmission may be represented using space-time transmit, receive and noise signal matrices $\widetilde{\mathbf{R}}$, $\widetilde{\mathbf{S}}$ and $\widetilde{\mathbf{N}}$ of sizes $M \times L$, $N \times L$ and $M \times L$, respectively.

$$\widetilde{\mathbf{R}} = \mathbf{H}\widetilde{\mathbf{S}} + \widetilde{\mathbf{N}}. \tag{2.4}$$

2.4.2 Frequency selective quasi-static fading channel

In the case of frequency-selective fading, the coherence bandwidth of the channel is smaller than the bandwidth of the signal. Equivalently, the delay spread of the channel is larger than the symbol period T_s. Thus each transmitted symbol is dispersed over K symbol periods, where $\tau_{max} \leq KT_s$. These K images of the signal are referred to as *resolvable paths* as discussed in Section 2.1.2.

The linear system representation of a dispersive fading channel is then a finite impulse response filter of length K or of *memory* $K-1$. The first multipath is assumed to arrive at time 0. In this case, we must include the delay parameter, but can simplify the presentation by recalling that within each block over a quasi-static channel $h(t,\tau) = h(\tau)$ is time invariant. We also define the discrete delay variable k as corresponding to a delay of $\tau = kT_s$. Thus in the derivations that follow, we will write the CIR as $h(t, kT_s) = h^k$.

Equation 2.2 representing the SISO case then becomes

$$r[l] = \sum_{k=0}^{K-1} h^k s[l-k] + n[l]. \tag{2.5}$$

In a block transmission scheme, e.g., Orthogonal Frequency Division Multiplexing (OFDM), directly applying (2.5) exposes the first symbol in each new block to interference from the

last $K-1$ symbols of the previous block. A common strategy used to protect the current block from Inter-Block Interference (IBI) is to flush out these leftover symbols by inserting a prefix of $K-1$ guard symbols. Thus a small rate loss is incurred because $\widetilde{L} = L + K - 1$ symbols must be transmitted to convey L information-bearing symbols.

The following expressions describe the signals involved in such a block transmission, where we assume that the channel is quasi-static, i.e., its impulse response is LTI:

$$\begin{aligned} \widetilde{\mathbf{r}}^{\widetilde{L}} &= \widetilde{\mathbf{H}}^{\widetilde{L}}\widetilde{\mathbf{s}}^{\widetilde{L}} + \widetilde{\mathbf{n}}^{\widetilde{L}} \qquad (2.6) \\ \widetilde{\mathbf{r}}^{\widetilde{L}} &= \left[\begin{array}{ccc|ccc} r[-K+1] & \cdots & r[-1] & r[0] & \cdots & r[L-1] \end{array}\right]^T \end{aligned}$$

where the channel matrix and transmitted signal and noise vectors are given by

$$\widetilde{\mathbf{H}}^{\widetilde{L}} = \left[\begin{array}{ccccccc} h^0 & 0 & \cdots & 0 & h^{K-1} & \cdots & h^1 \\ h^1 & h^0 & 0 & \cdots & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \cdots & 0 & h^{K-1} \\ \hline h^{K-1} & \cdots & h^1 & h^0 & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \cdots & \ddots & h^0 & 0 \\ 0 & \cdots & 0 & h^{K-1} & \cdots & h^1 & h^0 \end{array}\right] \qquad (2.7)$$

$$\begin{aligned} \widetilde{\mathbf{s}}^{\widetilde{L}} &= \left[\begin{array}{ccc|ccc} s[-K+1] & \cdots & s[-1] & s[0] & \cdots & s[L-1] \end{array}\right]^T \\ \widetilde{\mathbf{n}}^{\widetilde{L}} &= \left[\begin{array}{ccc|ccc} n[-K+1] & \cdots & n[-1] & n[0] & \cdots & n[L-1] \end{array}\right]^T \end{aligned}$$

Observe that the $\widetilde{L} \times \widetilde{L}$ CIR matrix has a special structure. Each of its columns differs from the previous column by a circular shift; such a matrix is called *circulant*. Some important properties of circulant matrices will be discussed later on. The other vectors in (2.6) are of length $\widetilde{L}$. The symbols to the left and channel coefficients above the annotated lines are associated with the inserted prefix. Thus the information-bearing block can be recovered simply by discarding the first $K-1$ received symbols. It is straightforward to verify that the L data symbols $r[0], \ldots, r[L-1]$ correspond to those given by (2.5).

We also note that the first symbol depends on the current symbol of interest as well as the previous prefix symbols. If these are chosen to be the cyclic extension of the block of data symbols, i.e., $s[-k] = s[L-k],\ k = \{1, \ldots, K-1\}$, then we can rewrite (2.6) in the following equivalent form:

$$\widetilde{\mathbf{r}} = \widetilde{\mathbf{H}}\widetilde{\mathbf{s}} + \widetilde{\mathbf{n}}, \qquad (2.8)$$

where the superscripts have been dropped since the vectors are all of length L. $\widetilde{\mathbf{H}}$ is again a circulant matrix, now of size $L \times L$, generated by the same channel response as before, and the transmitted, received and noise vectors correspond precisely to the L elements appearing to the right of the annotated lines in the previous expressions.

Because it is circulant, the SISO channel matrix (2.7) inherits some useful properties:

1. It is specified completely by its first column.

2. All other columns can be generated by circularly shifting the first, and all circular shifts of the first column necessarily appear as columns.

3. Circulant matrices are always square and *Toeplitz*, i.e., the elements along every diagonal are the same.

4. The set of $n \times n$ complex-valued circulant matrices with binary operations matrix addition and multiplication form a field. In particular, matrix multiplication is *commutative* within this field.

5. The circulant matrix is diagonalized by the inverse DFT, i.e., its EigenValue Decomposition (EVD) is given by $\widetilde{\mathbf{H}} = \mathbf{U}_{DFT,L}^{\dagger} \mathbf{\Lambda} \mathbf{U}_{DFT,L}$, where $\mathbf{U}_{DFT,L}$ is the $L \times L$ unitary DFT matrix,[5] $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\widetilde{\mathbf{H}}$, and $\operatorname{diag} \mathbf{\Lambda} = [\lambda^0 \cdots \lambda^{L-1}] = [\mathbb{h}^0 \cdots \mathbb{h}^{L-1}] = \mathscr{F}_d \left\{ [h^0 \cdots h^{K-1} \ 0 \cdots 0] \right\}$, the DFT of the zero-extended CIR.

In particular, Property 5 is of interest because it emphasizes the dual nature of the multipath fading channel as a frequency selective fading channel. The frequency selectivity is represented by the different eigenvalues found on the diagonal of $\mathbf{\Lambda}$, i.e., the non-flat frequency response of the channel. We can also see that the EVD decouples the frequency selective block transmission channel into $L_f = L$ parallel flat fading channels as follows:

$$\begin{aligned} \widetilde{\mathbf{r}} &= \mathbf{U}_{DFT,L}^{\dagger} \mathbf{\Lambda} \mathbf{U}_{DFT,L} \widetilde{\mathbf{s}} + \widetilde{\mathbf{n}} \\ \mathbf{U}_{DFT,L} \widetilde{\mathbf{r}} &= \mathbf{\Lambda} \mathbf{U}_{DFT,L} \widetilde{\mathbf{s}} + \mathbf{U}_{DFT,L} \widetilde{\mathbf{n}} \\ \mathscr{F}_d \{\widetilde{\mathbf{r}}\} &= \mathbf{\Lambda} \mathscr{F}_d \{\widetilde{\mathbf{s}}\} + \mathscr{F}_d \{\widetilde{\mathbf{n}}\} . \end{aligned}$$

By *decoupling* we mean that the coefficients received in each of the frequency divided sub-channels depend only on those transmitted in that sub-channel, and noise. Because $\mathbf{\Lambda}$ is diagonal, the composite signal mixtures that would otherwise be seen by the receiver have been reduced to simple single variable equations. Furthermore, the symbols transmitted on any of the L_f sub-channels or *tones* experience only flat fading, rather than the more difficult to combat frequency selective type. This result is well known to be advantageous, among other reasons because it removes the need for complex equalizers at the receiver. Because of its frequency domain structure, this block transmission strategy is known as Discrete Multi-Tone (DMT) or OFDM.

So far in this section, we have only been considering transmissions from a single antenna to a single antenna. Next we will take a look at some structures required for analysis of the frequency selective MIMO fading channel. A single symbol transmission from N transmit antennas to M receive antennas may be expressed as follows:

$$\mathbf{r} = \mathscr{H}^K \mathcal{S}^K[-K+1] + \mathbf{n}$$

where $\mathscr{H}^K$ is a $1 \times K$ space $\times$ space-time block matrix of $M \times N$ blocks of spatial channel coefficients,

$$\mathscr{H}^K = \underbrace{\begin{bmatrix} h_{11}^{K-1} & \cdots & h_{N1}^{K-1} \\ \vdots & \ddots & \vdots \\ h_{1M}^{K-1} & \cdots & h_{NM}^{K-1} \end{bmatrix}}_{\text{Delay } K\text{-1}} \cdots \underbrace{\begin{bmatrix} h_{11}^{0} & \cdots & h_{N1}^{0} \\ \vdots & \ddots & \vdots \\ h_{1M}^{0} & \cdots & h_{NM}^{0} \end{bmatrix}}_{\text{Delay 0}} = \begin{bmatrix} \mathbf{H}[K-1] & \cdots & \mathbf{H}[0] \end{bmatrix},$$

[5] $\mathbf{U}_{DFT,L} = \frac{1}{\sqrt{L}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & e^{-j\frac{2\pi}{L}} & e^{-j\frac{2\pi(2)}{L}} & \cdots & e^{-j\frac{2\pi(L-1)}{L}} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & e^{-j\frac{2\pi(L-1)}{L}} & e^{-j\frac{2\pi(L-1)(2)}{L}} & \cdots & e^{-j\frac{2\pi(L-1)(L-1)}{L}} \end{bmatrix}$, $\mathbf{U}_{IDFT,L} = \mathbf{U}_{DFT,L}^{\dagger}$.

and $\mathcal{S}^K[-K+1] = \text{Vec}\left(\widetilde{\mathbf{S}}^K[-K+1]\right)$ is a $NK \times 1$ space-time block vector of transmitted signals

$$\widetilde{\mathbf{S}}^K[-K+1] = \begin{bmatrix} s_1[-K+1] & \cdots & s_1[0] \\ \vdots & \ddots & \vdots \\ s_N[-K+1] & \cdots & s_N[0] \end{bmatrix}, \; \mathcal{S}^K[-K+1] = \begin{bmatrix} \mathbf{s}[-K+1] \\ \vdots \\ \mathbf{s}[0] \end{bmatrix}, \tag{2.9}$$

and $\mathbf{r}$ and $\mathbf{n}$ are $M \times 1$ received signals and complex AWGN vectors as before.

We will also be interested in the case of block transmission over the MIMO frequency selective quasi-static fading channel, which is described mathematically in an analogous manner to (2.8) by

$$\mathcal{R} \;\; = \;\; \widetilde{\mathscr{H}}\mathcal{S} + \mathcal{N},$$

where $\mathcal{R} = \text{Vec}(\widetilde{\mathbf{R}})$ and $\mathcal{N} = \text{Vec}(\widetilde{\mathbf{N}})$ are $ML \times 1$ space-time block vectors of received and noise signals

$$\widetilde{\mathbf{R}} = \begin{bmatrix} r_1[0] & \cdots & r_1[L-1] \\ \vdots & \ddots & \vdots \\ r_M[0] & \cdots & r_M[L-1] \end{bmatrix}, \; \mathcal{R} = \begin{bmatrix} \mathbf{r}[0] \\ \vdots \\ \mathbf{r}[L-1] \end{bmatrix},$$

$$\widetilde{\mathbf{N}} = \begin{bmatrix} n_1[0] & \cdots & n_1[L-1] \\ \vdots & \ddots & \vdots \\ n_M[0] & \cdots & n_M[L-1] \end{bmatrix}, \; \mathcal{N} = \begin{bmatrix} \mathbf{n}[0] \\ \vdots \\ \mathbf{n}[L-1] \end{bmatrix},$$

$\widetilde{\mathscr{H}}$ is a space-time $\times$ space-time block circulant matrix of $L \times L$ blocks, each containing the $M \times N$ MIMO spatial channel matrix corresponding to a particular multipath delay,

$$\widetilde{\mathscr{H}} \;\; = \;\; \begin{bmatrix} \mathbf{H}[0] & 0 & \cdots & 0 & \mathbf{H}[K-1] & \cdots & \mathbf{H}[1] \\ \mathbf{H}(1) & \mathbf{H}[0] & 0 & \cdots & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \cdots & 0 & \mathbf{H}[K-1] \\ \mathbf{H}[K-1] & \cdots & \mathbf{H}[1] & \mathbf{H}[0] & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \cdots & \ddots & \mathbf{H}[0] & 0 \\ 0 & \cdots & 0 & \mathbf{H}(K-1) & \cdots & \mathbf{H}[1] & \mathbf{H}[0] \end{bmatrix}, \tag{2.10}$$

and $\mathcal{S}$ is a $NL \times 1$ space-time block vector of transmitted signals as described in (2.9), except that in this case the sequence of signals sent from each antenna is of length L and starts from time index t.

Comparing (2.7) and (2.10), we can see that $\widetilde{\mathscr{H}}$ could equivalently be constructed by appropriately interleaving the NM SISO channel impulse response matrices $\widetilde{\mathbf{H}}_{ij}$, which correspond to each pair of transmit and receive antennas. Finally we note that the MIMO frequency selective channel matrix is block diagonalized by the unitary block inverse DFT matrices $\mathbf{U}_{DFT,NL} = \mathbf{U}_{DFT} \otimes \mathbf{I}_N$ and $\mathbf{U}_{DFT,ML} = \mathbf{U}_{DFT} \otimes \mathbf{I}_M$, where $\mathbf{U}_{DFT}$ is the $L \times L$

unitary DFT matrix as before [53]:

$$\begin{aligned}\mathbf{U}_{DFT,ML}\widetilde{\mathscr{H}}\mathbf{U}^{\dagger}_{DFT,NL} &= \mathbf{\Lambda}_{\widetilde{\mathscr{H}}} \\ \text{diag}\left(\mathbf{\Lambda}_{\widetilde{\mathscr{H}}}\right)_l &= \mathbb{H}[l] \quad (l^{th} \text{ diagonal block of } \mathbf{\Lambda}_{\widetilde{\mathscr{H}}}) \\ &= \sum_{k=0}^{K-1} \mathbf{H}[k]e^{-j\frac{2\pi}{L}kl}.\end{aligned}$$

The L diagonal blocks comprising space-frequency $\times$ space-frequency channel matrix $\mathbf{\Lambda}_{\widetilde{\mathscr{H}}}$ are $M \times N$ matrices representing the frequency response of the MIMO channel in each of the $L_f = L$ discrete frequency bins. Because of its similarity to the DMT solution for SISO frequency selective fading channels, the term Discrete Matrix Multi-Tone (DMMT) has been coined to describe this block transmission structure [43].

2.4.3 Flat quasi-static spatially correlated fading channel

To model the spatial selectivity of the channel, we begin by considering the flat quasi-static multi-antenna model given in (2.4):

$$\widetilde{\mathbf{R}} = \mathbf{H}\widetilde{\mathbf{S}} + \widetilde{\mathbf{N}}.$$

Each column of $\mathbf{H}$ represents the path gains and phase shifts seen by a signal propagating from one of the N transmit antennas to all M antennas in the receive array. To consider the true statistical dependence between these random variables, we would have to look at the $NM \times NM$ covariance matrix $\mathsf{E}\left[\text{Vec}(\mathbf{H})\text{Vec}(\mathbf{H})^{\dagger}\right]$. However it has been shown via simulation that a more compact separable product form provides a good approximation of the correlation (both in terms of the distribution of eigenvalues of $\mathbf{H}\mathbf{H}^{\dagger}$ [51] and the magnitudes of the resulting coefficients [11]).

This product form of the channel correlation is derived by making a few simplifying assumptions [11]:

- The channel coefficients are random variables with unit variance.
- The correlation between the fading coefficients corresponding to the paths to two different receive antennas does not depend on the transmit antenna, i.e., $\mathsf{E}\left(h^*_{ij}h_{i'j}\right) = \mathsf{E}\left(h^*_{ij'}h_{i'j'}\right)$ for all j'. Hence this term is referred to as the *receive correlation*.
- Analogously, the correlation corresponding to the paths from two different transmit antennas does not depend on the receive antenna, and is referred to as the *transmit correlation*.
- Finally, the correlation corresponding to two distinct paths having no common transmit or receive antennas is computed by taking the product of the appropriate receive and transmit correlation coefficients, i.e., $\mathsf{E}\left(h^*_{ij}h_{i'j'}\right) = \mathsf{E}\left(h^*_{ij}h_{i'j}\right)\mathsf{E}\left(h^*_{i'j}h_{i'j'}\right)$.

For simplicity, we use a zero mean distribution for the channel coefficients h in the following. First we form the receive correlation matrix $\mathbf{R}_{M,j}$ for the signal sent by transmit antenna j:

$$\mathbf{R}_{M,j} = \mathsf{E}\left(\mathbf{h}_j\mathbf{h}_j^{\dagger}\right).$$

Note that $\mathbf{h}_j$ is a column vector containing the channel coefficients seen by all M elements of the receive array corresponding to paths from transmit antenna j. $\mathbf{R}_{M,j}$ then encapsulates the spatial correlation between the signals received at each of the M antennas, i.e., $\mathbf{h}_j = \mathbf{R}_{M,j}^{\frac{1}{2}} \widehat{\mathbf{h}}_j$, where $\widehat{h}_{1j}, \ldots, \widehat{h}_{Mj}$ are uncorrelated random variables with zero mean and unit variance for all j.[6]

Applying the first assumption, we get that $\mathbf{R}_{M,j} = \mathbf{R}_M$ is independent of j, allowing us to express $\mathbf{H}$ in the following partially-decorrelated form:

$$\mathbf{H} \;=\; \mathbf{R}_M^{\frac{1}{2}} \widehat{\mathbf{H}},$$

where the rows of $\widehat{\mathbf{H}}$ are uncorrelated random vectors. By this we mean that $\mathsf{E}\left(\widehat{\overline{\mathbf{h}}}_i^{\dagger} \widehat{\overline{\mathbf{h}}}_{i'}\right) = \sum_{j=1}^{N} \mathsf{E}\left(\widehat{h}_{ij}^{*} \widehat{h}_{i'j}\right) = 0 \;\forall\; i' \neq i$. Also note that $(\mathbf{R}_M)_{i'i}$ corresponds to the receive correlation coefficient $\mathsf{E}\left(h_{ij}^{*} h_{i'j}\right)$ for any j.

We can also form an analogous transmit correlation matrix encapsulating the spatial correlation between signals transmitted from each of the N antennas:

$$\mathbf{R}_{N,i} \;=\; \mathsf{E}\left(\widehat{\overline{\mathbf{h}}}_i^{*} \widehat{\overline{\mathbf{h}}}_i^{T}\right),$$

where $\overline{\mathbf{h}}_i^T$ is a row vector containing the channel coefficients seen by receive antenna i along paths from all N elements of the transmit array, after receive correlation has been removed. As in the previous argument, we see that $\widehat{\overline{\mathbf{h}}}_i = \mathbf{R}_{N,i}^{\frac{*}{2}} \overline{\mathbf{h}}_{w,i}$, where $h_{w,i1}, \ldots, h_{w,iN}$ are uncorrelated random variables with zero mean and unit variance for all i.[7] Applying the assumption that the fading statistics are the same for all receive antennas, i.e., $\mathbf{R}_{N,i} = \mathbf{R}_N$ is independent of i, we can fully decorrelate $\mathbf{H}$ as follows:

$$\widehat{\mathbf{H}}^T \;=\; \mathbf{R}_N^{\frac{*}{2}} \mathbf{H}_w^T$$

$$\mathbf{H} \;=\; \mathbf{R}_M^{\frac{1}{2}} \mathbf{H}_w \mathbf{R}_N^{\frac{\dagger}{2}},$$

where all entries of $\mathbf{H}_w$ are uncorrelated (i.e., white) random variables with zero mean and unit variance. Similarly to the previous decomposition, $(\mathbf{R}_N^{\dagger})_{j'j}$ corresponds to the transmit correlation coefficient $\mathsf{E}\left(h_{ij}^{*} h_{ij'}\right)$ for any i.

Thus spatial selectivity is modelled by the matrices $\mathbf{R}_M$ and $\mathbf{R}_N$, which represent the spatial correlation between signals received by and transmitted from the receive and transmit antenna arrays, respectively. We note that they are both by definition Hermitian positive semi-definite.

2.5 Discussion

Much of the existing work in space-time coding concentrates on slow or quasi-static, flat, and spatially independent fading channels. These assumptions are relevant to narrowband

[6] This can be seen by observing that $\mathsf{E}\left(\widehat{\mathbf{h}}_j \widehat{\mathbf{h}}_j^{\dagger}\right) = \mathbf{R}_{M,j}^{-\frac{1}{2}} \mathsf{E}\left(\mathbf{h}_j \mathbf{h}_j^{\dagger}\right) \mathbf{R}_{M,j}^{-\frac{\dagger}{2}} = \mathbf{I}_M$.

[7] Similarly, this can be seen by observing that $\mathsf{E}\left(\overline{\mathbf{h}}_{w,i}^{*} \overline{\mathbf{h}}_{w,i}^{T}\right) = \mathbf{R}_{N,i}^{-\frac{1}{2}} \mathsf{E}\left(\widehat{\overline{\mathbf{h}}}_i^{*} \widehat{\overline{\mathbf{h}}}_i^{T}\right) \mathbf{R}_{N,i}^{-\frac{\dagger}{2}} = \mathbf{I}_N$.

communications with low mobility in rich scattering environments. However, they are not representative of the channels applicable to broadband fixed wireless access, a technology that we believe will become increasingly important in years to come.

The broadband fixed wireless access channel is a slow, frequency selective fading channel, which may experience medium to high spatial correlation, depending on the applications being considered. Although spatial correlation reduces the achievable capacity, its frequency selectivity provides additional diversity which can be exploited to improve the performance of a system communicating over this channel.

Chapter 3

Diversity and spatial diversity

The wireless environment presents a challenging communications problem because of the possibly time-, frequency- and spatially-varying degradations caused by signal fading. As we shall see, these impairments are not necessarily harmful. Under certain conditions it is possible to take advantage of the variations in the channel's responses to improve the received Signal-to-Noise Ratio (SNR). For instance, suppose that the channel is such that two identical signals, transmitted in parallel over two distinct frequency sub-channels, experience independent fading effects. The receiver can then obtain two copies of the desired signal, and the probability that both are severely degraded is lower than in the case where only one observation is available. Thus a better overall estimate may be recovered by combining these together in some manner.

The idea of obtaining a number of different copies of the same signal is called *diversity*. Such techniques provide a powerful toolset for achieving reliable transmission over fading channels. Although there are a number of means by which signal diversity can be obtained, the desired end remains the same: Enable the receiver to recover a more robust replica of the transmitted signal by combining a number of independently faded copies. Thus, diversity techniques can only be applied in cases and domains where the channel is sufficiently selective.

In this report we will be concerned primarily with spatial diversity, i.e., that derived from using MEA arrays. However, there are also four other kinds of diversity that are of current interest in the literature:

Delay diversity arises in a multipath channel, where the receiver can resolve multiple delayed copies of the same signal. Equalization may be used to provide diversity gain in this case, or for instance, a RAKE receiver [32]. This type of diversity may be referred to as *frequency diversity*, as such a multipath channel is equivalently modelled as a frequency selective fading channel.

Time diversity is available when the channel is fast fading. Thus signals sent from one symbol period to the next experience independent fading and may be combined to achieve a diversity gain.

Polarization diversity takes advantage of single antenna structures supporting orthogonal polarizations to provide independently fading channels [36]. These may present a promising cost- and space-effective alternative to MEAs, however recent results reveal that transmit diversity schemes generally suffer a performance loss when combined with polarization diversity [39].

Modal diversity or *pattern diversity* is a benefit available when using multimode antennas. It has been shown theoretically that the correlations between signals received by the different modes are sufficiently low that a significant diversity gain can be realized [56]. Such approaches are also envisioned to provide an alternative solution to MEAs, especially in applications with stringent physical size requirements.

We begin by introducing the ideas underlying the use of spatial diversity, briefly sketching its historical development and defining the term diversity gain. Then we derive the performance gains available to systems using receive only, transmit only, and combined transmit and receive diversity. Throughout this report we will make use of the following standard assumptions, which are commonly applied in the literature for analysis of wireless systems in fading environments:

- **Limited transmitter power.** The total transmitted power, regardless of the number of transmit antennas, is constrained to be less than P.
- **Rayleigh fading channel.** The coefficients $h[l]$ are drawn from a circularly symmetric complex Gaussian distribution with zero mean and unit variance (or independent real and imaginary parts each having variance $\frac{1}{2}$).
- **Spatially independent fading.** At any given time, these coefficients $h_{ij}[l]$ are independent, i.e., there is enough physical separation between the antenna elements such that the signals fade independently across the arrays.
- **Perfect CSI at the receiver.** The receiver has perfect knowledge of the channel coefficients. This is a reasonable assumption when the fading is slow enough to allow estimation of the channel state with negligible error, as in the case of fixed wireless systems.
- **Circularly symmetric complex Gaussian noise.** In addition to fading, the signals are also corrupted by additive noise components $n_i[l]$ that are modelled as independent circularly symmetric complex Gaussian random variables with zero mean and variance N_0 (or $\frac{N_0}{2}$ per dimension).

3.1 Spatial diversity

MEA arrays are used in wireless communications to improve system performance at the expense of processing complexity at either the transmitter, receiver, or both. The purpose of this chapter is to illustrate how the signals transmitted using MEAs can be designed and processed to provide a diversity advantage, i.e., improved SNR and hence bit error performance at the receiver. In the case of the MIMO channel, where MEAs are used at both the transmitter and receiver, increased capacity or a multiplexing gain may also be realized; this topic will be touched upon in Section 3.4 and discussed in more detail in Chapter 4.

Research in spatial diversity focused initially on receiver techniques, motivated by the goal of mitigating degradations in the signal caused by multipath propagation. Under the assumption that the paths taken by each of the copies result in statistically independent fading effects, we can conclude that they are unlikely to all be in a deep fade, i.e., strongly distorted, simultaneously. Thus an improved signal may be obtained by forming a weighted combination of the received copies. Three common receive diversity strategies used when communicating over Rayleigh fading channels are selection diversity, equal gain combining,

and the well-known optimal SNR approach, MRC which will be discussed in more detail in Section 3.2.

Because of the physical size of the relevant antennas,[1] as well as restrictions on the processing power available at subscriber terminals, receive diversity was appropriate for improving signal quality only at the basestation, i.e., in the uplink. Interest in transmit diversity techniques arose in an attempt to realize similar performance benefits in the downlink, while displacing the additional processing complexity and the physical burden of the MEA from subscriber units to the basestation. The general structure of a transmit diversity system will be presented in Section 3.3, and a survey of some proposed techniques in Section 5.2.

Performance improvements attributed to the receiver having obtained multiple copies of the same information signal are referred to in the literature as *diversity advantage*, *order*, *level*, or *gain*. Such improvements are most obviously manifest in the form of steeper bit error curves. Perhaps a more intuitive way of thinking of diversity advantage is as the ratio of the SNR obtained using diversity to that obtained with a single transmission path (given the same total transmitted power and signal bandwidth). We will show how this SNR relates to the bit error performance and define diversity advantage more formally as the asymptotic slope of the bit error curve on a log-log scale [1]. Thus it reflects the exponential decrease of the bit error rate against SNR.

We will say that a technique achieves *full spatial diversity* if its diversity advantage is equal to the number of paths from the transmit to the receive arrays. We shall see that the notion of diversity gain is not as clear cut for instance as that of coding gain. In particular, two systems can have the same diversity gain but achieve different bit error performance curves (e.g., see Fig. 3.5). These ideas will be made more clear as the chapter progresses.

3.2 Receive only diversity

A simple illustration of diversity advantage can be seen by considering the well-known MRC receiver. We begin with a system having no diversity and show how receive diversity can be used to improve its performance without increasing transmitted power or bandwidth. In Figure 3.1 the $N = 1$, $M = 1$ SISO flat fading channel and a optimal SNR detector are shown. The optimality of this detector can be seen by applying Schwarz' Inequality and is also discussed in Appendix A in more detail. The received signal is

$$r[l] \;=\; h[l]s[l] + n[l],$$

and the detected signal is then

$$\begin{aligned} \overline{s}[l] &= h^*[l]\,[h[l]s[l] + n[l]] \\ &= \alpha^2[l]s[l] + \alpha[l]\widetilde{n}[l], \end{aligned}$$

where $\alpha[l] = |h[l]|$ is the magnitude of the channel coefficient or path gain, and the noise term $\widetilde{n}[l]$ has the same distribution as $n[l]$ since the circularly symmetric Gaussian distribution is invariant to multiplication by unit phase components. We note that this detector is sometimes referred to in the space-time coding literature as a *discrete matched filter*, since it maximizes the post-detection SNR by matching to the complex channel coefficient $h[l]$.

[1] A nominal quarter wavelength antenna tuned to the 800-900 MHz frequency bands used in second generation cellular systems would be around 8-9cm in size.

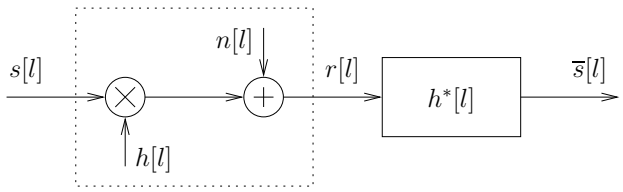


Figure 3.1: System diagram for $N = 1$ transmit antenna, $M = 1$ receive antenna SISO channel with optimal SNR detector.

The conditional SNR of the SISO system given perfect knowledge of the channel coefficients $h[l]$ (and hence their magnitudes $\alpha[l]$) is

$$\begin{aligned} \rho_{SISO|\alpha[l]} &= \frac{\left|\alpha^2[l]s[l]\right|^2}{\mathsf{E}\left(|\alpha[l]\widetilde{n}[l]|^2\right)} \\ &= \rho\alpha^2[l], \end{aligned} \tag{3.1}$$

where $\rho = \frac{P}{N_0}$ is the SNR of the channel without fading. This expression may also be referred to as the instantaneous SNR since it corresponds to the SNR of a particular channel realization. Analogously, the instantaneous bit error rate of a coherent Binary Phase Shift Keying (BPSK) system communicating over this channel is given by

$$p_{\varepsilon|\alpha[l]}(\rho) = \mathrm{Q}\left(\sqrt{2\rho}\alpha[l]\right),$$

The average bit error rate can be determined by simulation or analytically by integrating over the Rayleigh probability density function of the channel coefficient:

$$\begin{aligned} p_{\varepsilon,M=1}(\rho) &= \int_0^\infty p_{\varepsilon|\alpha}(\rho)\cdot f(\alpha)\,d\alpha \\ &= \int_0^\infty \mathrm{Q}\left(\sqrt{2\rho}\alpha\right)\cdot 2\alpha e^{-\alpha^2}\,d\alpha \\ &= -\mathrm{Q}\left(\sqrt{2\rho}\alpha\right)\cdot e^{-\alpha^2}\Big|_0^\infty - \int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-\rho\alpha^2}\cdot\sqrt{2\rho}e^{-\alpha^2}\,d\alpha \\ &= \frac{1}{2} - \sqrt{\frac{\rho}{\pi}}\int_0^\infty e^{-(1+\rho)\alpha^2}\,d\alpha \\ &\overset{u=\alpha\sqrt{1+\rho}}{=} \frac{1}{2} - \sqrt{\frac{\rho}{\pi(1+\rho)}}\int_0^\infty e^{-u^2}\,du \\ &= \frac{1}{2}\left(1-\sqrt{\frac{\rho}{1+\rho}}\right). \end{aligned} \tag{3.2}$$

At high SNRs, (3.2) can be approximated as $\frac{1}{4\rho}$ [42]. Observe that the exponent of the SNR ρ is -1, hence we say that the SISO system has a diversity advantage of 1. In the plot shown in Figure 3.3 at the end of the section, we also note that the asymptotic slope of the no diversity curve is -1, or equivalently that the bit error rate falls by an order of magnitude for every additional 10 dB of SNR.

Next, if we consider the $N = 1$, M Single Input Multiple Output (SIMO) channel shown in Figure 3.2, we can write the signal received along the i^{th} branch as

$$r_i[l] \quad = \quad h_i[l]s[l] + n_i[l].$$

The SNR per branch along each transmission path can be expressed similarly as in (3.1):

$$\rho_{i|\alpha_i[l]} \quad = \quad \rho\alpha_i^2[l].$$

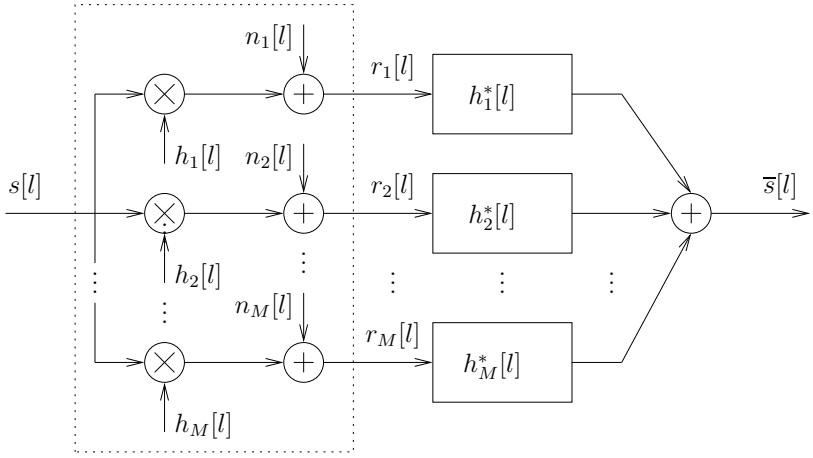


Figure 3.2: System diagram for $N = 1$ transmit antenna, M receive antennas SIMO channel with optimal MRC receiver.

In the MRC diversity combining scheme, the estimated signal is obtained by forming a linear combination of the signals received at each antenna. It is well-known to be an optimal SNR linear processing receive diversity algorithm. In fact, it can also be shown that MRC combining is optimal in the ML sense, and thus achieves the best performance over the set of all linear and non-linear receiver structures. See Appendix A for more details.

$$\begin{aligned} \overline{s}[l] \quad &= \quad \sum_{i=1}^{M} c_i[l] r_i[l] \\ &= \quad \sum_{i=1}^{M} c_i[l] \left[h_i[l]s[l] + n_i[l] \right]. \end{aligned}$$

The conditional SNR of the combined signal given the channel coefficients $h_1[l], \ldots, h_M[l]$ is then

$$\begin{aligned}
\rho_{SIMO,MRC|\alpha_i[l],i=1,\ldots,M} &= \frac{\left|\sum_{i=1}^{M} c_i[l]h_i[l]s[l]\right|^2}{\mathsf{E}\left(\left|\sum_{i=1}^{M} c_i[l]n_i[l]\right|^2\right)} \\
&\leq \rho \cdot \frac{\sum_{i=1}^{M}|c_i[l]|^2 \sum_{i=1}^{M}\alpha_i^2[l]}{\sum_{i=1}^{M}|c_i[l]|^2} \\
&= \rho\sum_{i=1}^{M}\alpha_i^2[l] \qquad (3.3) \\
&= M\overline{\rho}_{i|\alpha_i[l]}\,,
\end{aligned}$$

where the average branch SNR $\overline{\rho}_i$ is equal to the SNR of the system with no diversity given in (3.1), thus giving a full diversity advantage of M. The MRC combining scheme is designed to achieve this optimal SNR by setting the combining weights $c_i[l] = h_i^*[l]$.[2] Observe that the magnitudes of the combining weights are precisely those of the path gains, thus "good" paths with high SNRs will figure more prominently in the resulting combination, whereas paths with low SNRs will play a lesser role.

The signal estimate is therefore

$$\begin{aligned}
\overline{s}[l] &= \sum_{i=1}^{M} h_i^*[l]\left[h_i[l]s[l] + n_i[l]\right] \\
&= s[l]\sum_{i=1}^{M}\alpha_i^2[l] + \sum_{i=1}^{M}\alpha[l]\widetilde{n}_i[l],
\end{aligned}$$

and decoding can proceed from this form. The instantaneous SNR of this signal is given by (3.3) and thus the instantaneous bit error rate of a coherent BPSK system communicating over this channel is

$$\begin{aligned}
p_{\varepsilon|\alpha_i[l],j=1,\ldots,M}(\rho) &= \mathrm{Q}\left(\sqrt{2\rho\sum_{i=1}^{M}\alpha_i^2[l]}\right) \\
&= \mathrm{Q}\left(\sqrt{2\rho\chi_{2M}^2}\right),
\end{aligned}$$

where χ_{2M}^2 is a chi-squared random variable with $2M$ degrees of freedom. Note that when plotting this curve, we would like to see the error rate in terms of the average (received) SNR per bit. Again from (3.3), we can see that this is

$$\begin{aligned}
\overline{\rho} &= \mathsf{E}\left(\rho_{SIMO,MRC|\alpha_i[l],j=1,\ldots,M}\right) \\
&= M\rho.
\end{aligned}$$

[2]This follows from Schwarz' Inequality, which states that $|\langle x,y\rangle|^2 \leq \langle x,x\rangle\langle y,y\rangle$, where $\langle x,y\rangle$ denotes the inner product of x and y, i.e., $x^\dagger y$ for vectors, $\int x^*(t)y(t)\,dt$ for continuous time signals, $\sum_l x^*[l]y[l]$ for discrete time signals, or $\mathsf{E}(x^*y)$ for random variables. Equality is achieved if and only if $x = Cy$ for some constant C.

Thus we are actually interested in

$$p_{\varepsilon|\alpha_i[l],j=1,\ldots,M}\left(\overline{\rho}\right) = \mathrm{Q}\left(\sqrt{2\frac{\overline{\rho}}{M}\chi^2_{2M}}\right),$$

and when simulating the system, we set the transmitted power to $P = \frac{E_b}{M}$. Thus P represents the transmitted power per antenna that achieves an average SNR per bit of $\frac{E_b}{N_0}$. Exact expressions for the bit error probabilities using MRC with $M = 2, 4$ have been derived as follows and are plotted in Figure 3.3.

$$p_{\varepsilon,M=2}(\rho) = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\rho}{1+\rho}} - \frac{1}{4}\sqrt{\frac{\rho}{(1+\rho)^3}} \approx \frac{1}{8\rho^2}$$

$$p_{\varepsilon,M=4}(\rho) = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\rho}{1+\rho}} - \frac{1}{4}\sqrt{\frac{\rho}{(1+\rho)^3}} - \frac{3}{16}\sqrt{\frac{\rho}{(1+\rho)^5}} - \frac{15}{96}\sqrt{\frac{\rho}{(1+\rho)^7}} \approx \frac{1}{32\rho^4}$$

It can be shown that in the limit as $M \to \infty$, the performance of MRC approaches that obtained over an equivalent AWGN channel, i.e., with the same N_0, but with no fading. See Appendix B for more details.

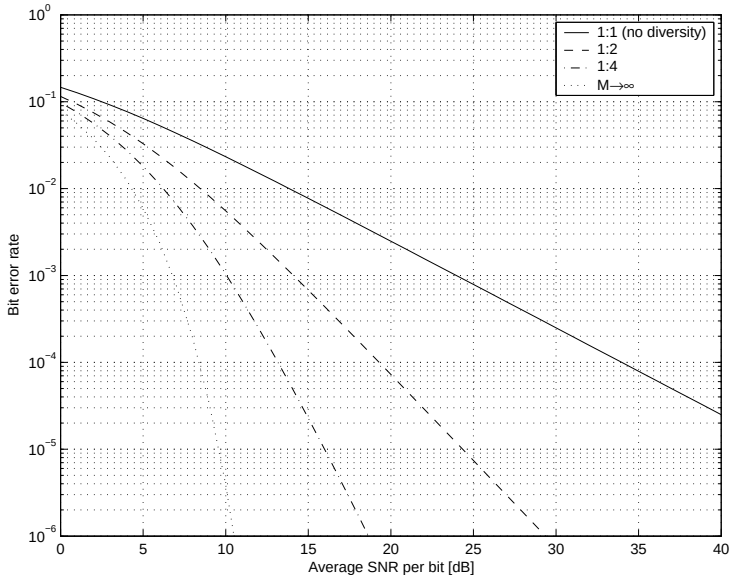


Figure 3.3: Bit error rate vs. average SNR per bit for 1:M systems in independent flat Rayleigh fading using uncoded BPSK with a coherent optimal MRC receiver.

Also in Appendix B, we show that in the high SNR regime, the bit error probability is approximately proportional to ρ^{-M}. Hence the definition of the diversity advantage given at the beginning of Section 3.1, as the asymptotic slope of the bit error performance curve on a log-log scale.

3.3 Transmit only diversity

The success of receive diversity in improving the performance of wireless communication systems led to the wide deployment of multiple element antenna arrays, particularly at

basestations where hardware size and cost are less important considerations. There they could be used to enhance the uplink channel from the subscriber unit to the basestation. To achieve similar benefits in the downlink channel, while not requiring multiple element arrays at the subscriber unit, researchers turned their attention to transmit diversity.

Transmit diversity is inherently a more difficult problem than receive diversity. In the case of receive diversity we obtain copies of the signal that are assumed to have undergone independent fading. The task at hand is to combine them optimally to recover the original transmitted signal. One fundamental difference between such systems and those employing transmit diversity is that in the latter case the signals are already combined when they reach the receiver. Even assuming that the receiver has perfect CSI, separating this mixture of signals in an optimal manner is a great challenge.

We note that there are two classes of problems that are very closely related to transmit diversity detection: The more general is blind source separation, where a number of multiplicative or convolutive mixtures are observed and the goal is to recover pure copies of the constituent sources. There is a large body of signal processing literature on this subject, for instance a recent survey can be found in [10]. Source separation is analogous to extracting the streams corresponding to each antenna in the transmit array. However, the transmit diversity detection problem is simpler, since the sources are discrete rather than continuous random variables.

Another superset of transmit diversity detection is the class of problems known as Multi-User Detection (MUD). In this case, the receive antenna observes N signals coming not only from different antennas, but also from different users. Detecting the signals of the individual users is also analogous to extracting the streams transmitted by each antenna in a transmit diversity system. However, transmit diversity detection is simplified by the knowledge that the antennas are co-located. Therefore the near-far problem does not have to be addressed with complicated power control algorithms.

To investigate transmit diversity, we first consider the N transmit, $M = 1$ receive antenna Multiple Input Single Output (MISO) channel shown in Figure 3.4.

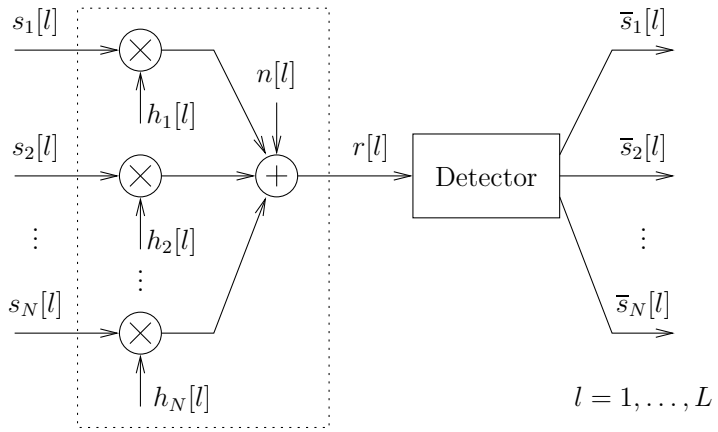


Figure 3.4: System diagram for N transmit antenna, $M = 1$ receive antenna MISO channel.

The received signal, assuming slow or quasi-static and flat fading, can be written as

$$\begin{aligned} r[l] &= \sum_{j=1}^{N} h_j[l]s_j[l] + n[l], \;\; l = 1, \ldots, L \\ \widetilde{\mathbf{r}}^T &= \overline{\mathbf{h}}^T \widetilde{\mathbf{S}} + \widetilde{\mathbf{n}}^T \\ \widetilde{\mathbf{r}} &= \widetilde{\mathbf{S}}^T \overline{\mathbf{h}} + \widetilde{\mathbf{n}}, \end{aligned} \tag{3.4}$$

where $\overline{\mathbf{h}}[l]$ and $\mathbf{s}[l]$ are vectors of channel coefficients and transmitted signals respectively and $\widetilde{\mathbf{S}}$ is a $N \times L$ space-time signal matrix.

From matrix (3.4) we see immediately that in order to resolve this system of equations, the block length L must be at least as large as the number of transmit antennas N. Otherwise we are left with an underdetermined system of equations. We use the term resolve here rather than solve because the noise vector is an unknown random variable. Thus using linear techniques, the best we can hope to do is to provide an estimate of the solution. Zero-forcing type estimates are obtained by inverting the matrix (ignoring the noise), leading to receivers of the form $\mathbf{H}^{-1}$ or $\mathbf{H}^{+} = (\mathbf{H}^{\dagger}\mathbf{H})^{-1}\mathbf{H}^{\dagger}$. Whereas Minimum Mean Squared Error (MMSE) estimates take the noise into consideration by minimizing the residual error term. In this case the linear receiver and has the form $\mathbf{H}^{\dagger}(\mathbf{H}\mathbf{H}^{\dagger} + \frac{1}{\rho}\mathbf{I}_M)^{-1}$.

Also observe that (3.4) as it stands does not enable us to resolve anything. Given ideal CSI at the receiver, $\overline{\mathbf{h}}$ is already known and it is an estimate for $\widetilde{\mathbf{S}}$ that we are trying to obtain. Thus in order to make use of this framework, we have to impose some additional structure on $\widetilde{\mathbf{S}}$. The following approach for two transmit antennas is known as the Alamouti STBC and has been shown to be an optimal SNR linear processing transmit diversity algorithm [3]:

$$\begin{aligned} \widetilde{\mathbf{S}} &= \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix} \\ \begin{bmatrix} r[1] \\ r[2] \end{bmatrix} &= \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} n[1] \\ n[2] \end{bmatrix} \\ &= \begin{bmatrix} x_1 h_1 + x_2 h_2 \\ -x_2^* h_1 + x_1^* h_2 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} n[1] \\ n[2] \end{bmatrix} \\ \begin{bmatrix} r[1] \\ r^*[2] \end{bmatrix} &= \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n[1] \\ n^*[2] \end{bmatrix}, \end{aligned} \tag{3.5}$$

The ML solution can then be found by inverting the channel matrix derived in (3.5). Note that it is not usually the case that the optimal ML estimate can be obtained using linear techniques. Since the constructed channel matrix is a scaled unitary matrix, i.e., $\det(\mathbf{H})\,\mathbf{H}^{-1} = \mathbf{H}^{\dagger}$, the ML, zero-forcing and MMSE solutions are all equivalent. See Appendix A for a more detailed discussion:

$$\begin{aligned} \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \end{bmatrix} &= \begin{bmatrix} h_1^* & h_2 \\ h_2^* & -h_1 \end{bmatrix} \begin{bmatrix} r[1] \\ r^*[2] \end{bmatrix} \\ &= \left(\alpha_1^2 + \alpha_2^2\right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} h_1^* & h_2 \\ h_2^* & -h_1 \end{bmatrix} \begin{bmatrix} n[1] \\ n^*[2] \end{bmatrix} \\ &= \left(\alpha_1^2 + \alpha_2^2\right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \widetilde{n}_1[1] & \widetilde{n}_2[2] \\ \widetilde{n}_2[1] & \widetilde{n}_1[2] \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \end{aligned}$$

Note that for this code $Q = 2$ signals are being transmitted using $N = 2$ antennas, $s_1[1] = x_1$, $s_1[2] = -x_2^*$, $s_2[1] = x_2$ and $s_2[2] = x_1^*$, therefore $|x_1|^2 + |x_2|^2 = |s_1|^2 + |s_2|^2$. The conditional SNR achieved by the Alamouti STBC given perfect knowledge of the channel coefficients h_1, h_2 (and hence their magnitudes α_1, α_2) is

$$\begin{aligned} \rho_{MISO,Alamouti-2:1|\alpha_1,\alpha_2} &= \frac{\left|(\alpha_1^2+\alpha_2^2)x\right|^2}{\mathsf{E}\left(\left|\alpha_1\widetilde{n}+\alpha_2\widetilde{n}'\right|^2\right)} \\ &= \rho(\alpha_1^2+\alpha_2^2) \end{aligned} \tag{3.6}$$

where $\widetilde{n}$ and $\widetilde{n}'$ are independent random noise terms having the same probability distribution as $n[l]$. Comparing this SNR to that of the SISO system (i.e., with no diversity) in (3.1) we get a full diversity advantage of 2. (This follows from the post-detection SNR in (3.6), which indicates that the bit error performance curve is based on a χ^2 random variable with four degrees of freedom.) The scheme is also optimal in terms of SNR, since it results in the same benefit as an MRC receiver with the same number of diversity paths.

The instantaneous bit error rate of a coherent BPSK system communicating over this channel using the Alamouti code is

$$\begin{aligned} p_{\varepsilon|\alpha_1,\alpha_2}(\rho) &= \mathrm{Q}\left(\sqrt{2\rho(\alpha_1^2+\alpha_2^2)}\right) \\ &= \mathrm{Q}\left(\sqrt{2\rho\chi_4^2}\right), \end{aligned}$$

where χ_4^2 is a chi-squared random variable with four degrees of freedom. However when simulating this system, we will be interested in its performance compared to that of other schemes, given the same total transmitted power. Thus the transmitted power per antenna (or signal) will be reduced by a factor of 2 and (3.6) becomes

$$\rho_{MISO,Alamouti-2:1\left|\alpha_1,\alpha_2,|s_1|^2+|s_2|^2\le P\right.} = \frac{\rho}{2}(\alpha_1^2+\alpha_2^2) \tag{3.7}$$

which is the same SNR as in the SISO case in expected value. This does not mean that there is no diversity advantage. From the plots in Figure 3.5 we can see that the performance curves for the Alamouti STBC and no diversity cases differ. Because the diversity advantage of the code is 2, the asymptotic slope (as the SNR increases) of its bit error curve is steeper than that of the SISO curve by a factor of 2.

We can also see that the decreased transmit power results in a the 3 dB shift to the right with respect to the curve for the MRC receiver. However, the shape and slope of the two curves are the same, because they have the same diversity advantage. We make one final comment about the fact that these curves are plotted in terms of average SNR per bit per receive branch. This ensures a fair comparison, for in the case of transmit diversity we are only looking at the average SNR per bit for one of the transmitted signals, i.e., one transmit branch.[3]

[3]Plots in terms of average SNR per bit per receive branch are also presented in [3]. One way to explain why this comparison is more fair is to consider the elements of the antenna arrays as separate users communicating with a basestation. In this analogy, the receive diversity scenario corresponds to the downlink from the single basestation antenna to the multiple users, and the transmit diversity scenario to the uplink from all users to the basestation. To compare the uplink and downlink performance fairly, we should plot the bit error rate against the average SNR per bit per user in both directions, rather than against the average SNR per bit (i.e., for all users) in the downlink and only per a single user in the uplink.

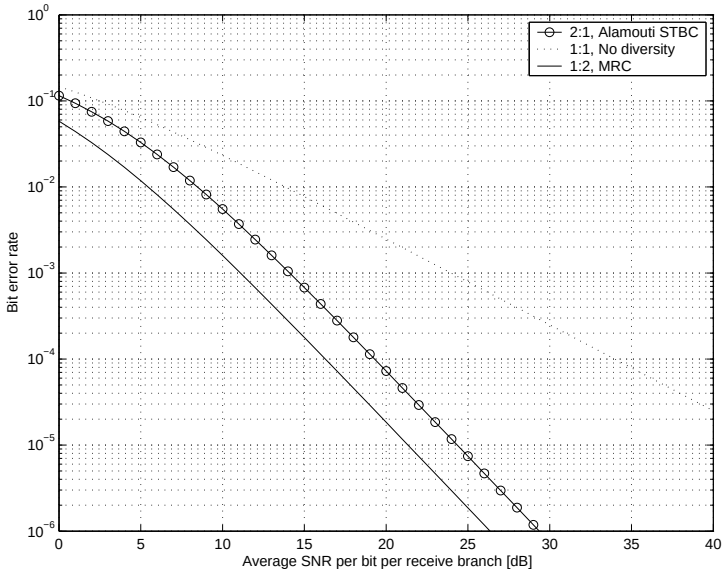


Figure 3.5: Bit error rate vs. average SNR per bit per receive branch for 2:1 systems in independent flat Rayleigh fading using uncoded coherent BPSK with the Alamouti STBC. Analogous curves for 1:1 (no diversity) and 1:2 MRC receiver system having the same diversity advantage are also shown for comparison.

The Alamouti code illustrates that for $N = 2$, the full diversity advantage can be achieved using a STBC signal structure with a block length of $L = 2$. Since only linear operations are employed at the receiver, this performance improvement can also be realized with little increase in computational complexity. We note that although this code is generally accepted to be one of the best solutions in the case where there are two transmit and one receive antenna, transmit diversity is by no means a solved problem. In particular, observe that the code cannot be extended to accommodate more than two transmit antennas. More general transmit diversity strategies have been proposed in the literature; details on these can be found in Chapter 5. The Alamouti code was chosen for presentation here because we feel that it most clearly illustrates the main ideas behind transmit diversity.

Another important point that should be made about the Alamouti STBC is the following: If we look more closely at (3.5), we can see that the code is in fact using the spatial diversity available at the transmitter to create time selectivity in the quasi-static fading channel. Because the transmitted signals are combined by the channel, transmit diversity by itself does not provide any SNR benefit. If we consider the signals seen at the receiver over $l = 1, \ldots, L$ separately, then those sent from the other antenna elements are actually sources of co-channel interference, with respect to detecting a particular transmitted signal of interest. Thus the Signal-to-Interference-and-Noise Ratio (SINR) is further reduced in comparison to the baseline SISO system SNR. A diversity advantage is only achieved through extending the transmission in the time domain and exploiting the temporal diversity induced by the mathematical properties of the space-time signal matrix $\widetilde{\mathbf{S}}$.

We note that the underlying structure of the Alamouti code could just as easily have been used to create frequency selectivity in a flat fading channel from the spatial diversity at the transmitter. In this case the transmission resources are divided into $L_f = 2$ parallel sub-channels in the frequency domain, each of bandwidth $\frac{W}{L_f}$. Because of the reduced

transmission bandwidth, the symbol period must be extended by the same factor to $L_f T_s$. We can then define the following space-frequency code:

$$\widetilde{\mathbf{S}}_f = \begin{bmatrix} \mathbb{x}_1 & -\mathbb{x}_2^* \\ \mathbb{x}_2 & \mathbb{x}_1^* \end{bmatrix}$$

$$\begin{bmatrix} \mathbb{r}[1] \\ \mathbb{r}^*[2] \end{bmatrix} = \begin{bmatrix} \mathbb{h}_1 & \mathbb{h}_2 \\ \mathbb{h}_2^* & -\mathbb{h}_1^* \end{bmatrix} \begin{bmatrix} \mathbb{x}_1 \\ \mathbb{x}_2 \end{bmatrix} + \begin{bmatrix} \mathbb{n}[1] \\ \mathbb{n}^*[2] \end{bmatrix}$$

Although the symbol period has been doubled, the rate of this code is the same as that of the space-time version. Symbols take twice as long to send, however two can be sent simultaneously over the disjoint frequency sub-channels. Thus a block length in the time domain of $L = 1$ is sufficient. We can see immediately from its structure, which is identical to that given by (3.5), that the same diversity advantage is achieved.

As observed in [27], the maximum diversity arises when the channel is time, frequency, and spatially selective. The degree of selectivity may be represented by the number of independent sub-channels per transmission block in each of these domains: L, L_f and M, respectively. Spatial diversity at the transmitter cannot increase the diversity advantage beyond the maximum level of LL_fM, *it can only be used to help achieve this diversity level if the fading on the sub-channels in the other domains is correlated or non-selective.*

3.4 Combined transmit and receive diversity

So far we have considered deploying MEAs at the transmitter or at the receiver, as a means of obtaining a spatial diversity advantage. MEAs can also be used at both the transmit and receive arrays, in conjunction with STBCs of block length $L = N$, to provide a diversity advantage of up to NM over flat quasi-static fading channels. In a manner similar to the Alamouti code discussed previously, spatial diversity at the transmitter may be converted into selectivity in the time or frequency domains.

The generic MIMO system diagram is shown in Figure 3.6. As illustrated, if the receiver is linear, we will denote it by the matrix $\mathbf{W}$.

Observe that unlike the MISO channel, where the block length must be greater than one in order for the transmitted signals to be estimated, in this case we do not require any conditions on the block length. From a purely mathematical perspective, the received signal at any sampling time can be expressed as

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{3.8}$$

where $\mathbf{H}$ is an $M \times N$ matrix of channel coefficients. Assuming that $\mathbf{H}$ is full rank, the solvability of this matrix equation is determined by the relationship between M and N. As long as $M \geq N$, the transmitted signal vector $\mathbf{s}$ may be estimated. If $M < N$, we are left with an underdetermined system, which may still be partially solved, usually given some additional structure imposed on the set of transmitted signals.

It is important to emphasize that this way of thinking about the MIMO channel is fundamentally different from those that we have considered so far in this chapter. There is still spatial diversity available at both ends of the communication link. However in this case we are not interested in diversity combining or in obtaining an SNR advantage. Instead we are looking at the problem of transmitting up to $\min(M, N)$ different signals simultaneously, and recovering them at the receiver. As noted in Section 3.3, this is essentially the problem of blind source separation.

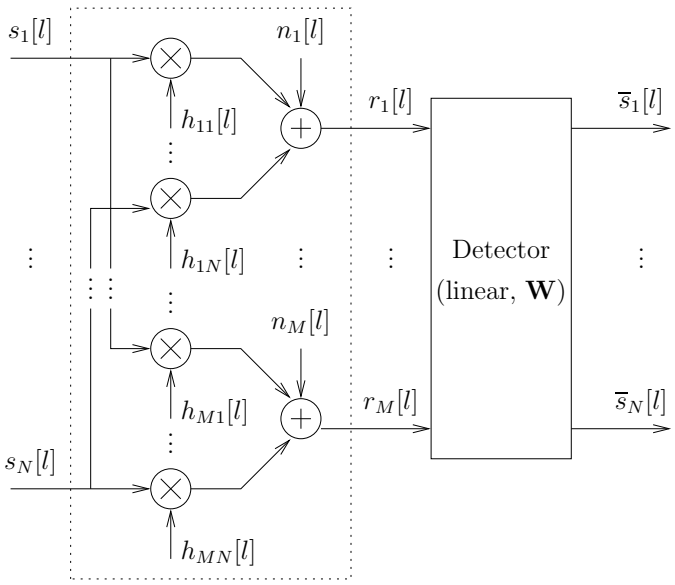


Figure 3.6: System diagram for N transmit, M receive antenna MIMO channel.

In this section we take a closer look at these two classes of communication strategies for MIMO channels. They may be referred to in the literature using the terms *diversity* and *spatial multiplexing* techniques.

3.4.1 Improving the received SNR

The first simple yet illustrative code is obtained by constructing an extended version of the Alamouti STBC for two transmit and two receive antennas:

$$\widetilde{\mathbf{S}} = \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix}$$

$$\begin{bmatrix} r_1[1] & r_2[1] \\ r_1[2] & r_2[2] \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix} \begin{bmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{bmatrix} + \begin{bmatrix} n[1] \\ n[2] \end{bmatrix}$$

$$\begin{bmatrix} r_1[1] \\ r_1^*[2] \\ r_2[1] \\ r_2^*[2] \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{12}^* & -h_{11}^* \\ h_{21} & h_{22} \\ h_{22}^* & -h_{21}^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n[1] \\ n^*[2] \\ n[1] \\ n^*[2] \end{bmatrix}.$$

ML detection can be done via a linear operation by applying the Moore-Penrose pseudoinverse of the constructed channel matrix at the receiver:

$$\begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \end{bmatrix} = \begin{bmatrix} h_{11}^* & h_{12} & h_{21}^* & h_{22} \\ h_{12}^* & -h_{11} & h_{22}^* & -h_{21} \end{bmatrix} \begin{bmatrix} r_1[1] \\ r_1^*[2] \\ r_2[1] \\ r_2^*[2] \end{bmatrix}$$

$$= \left(\alpha_{11}^2 + \alpha_{12}^2 + \alpha_{21}^2 + \alpha_{22}^2\right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \widetilde{n}_{11}[1] & \widetilde{n}_{12}[2] & \widetilde{n}_{21}[1] & \widetilde{n}_{22}[2] \\ \widetilde{n}_{12}[1] & \widetilde{n}_{11}[2] & \widetilde{n}_{22}[1] & \widetilde{n}_{21}[2] \end{bmatrix} \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{21} \\ \alpha_{22} \end{bmatrix}$$

We can see immediately from the structure of the estimated signal that the full diversity advantage of $LM = 4$ is achieved. (As before, this follows since the post-detection SNR is $\rho(\alpha_{11}^2+\alpha_{12}^2+\alpha_{21}^2+\alpha_{22}^2)$ and therefore the bit error performance curve is based on a χ^2 random variable with eight degrees of freedom.) Thus as shown in Figure 3.8, its performance is the same as that of an MRC receiver with 4 diversity branches, except that the transmitted power per antenna is reduced by a factor of 2 to maintain the same total transmitted power. Note that the post-detection SNR of both signals is the same.

Finally, we also observe that there is a repetition-like structure to the Alamouti STBC. The same signals are sent in each symbol period, but they are conjugated and their order is permuted to induce temporal selectivity. Therefore, although the full diversity advantage is achieved, we are only able to transmit $Q = 2$ signals in $L = 2$ symbols periods, giving an overall communication rate of 1 sym/s/Hz.

3.4.2 Increasing the data capacity

As noted previously, when using MEAs at both the transmit and receive arrays, it is not necessary to add redundancy to the transmitted signal matrix in order to detect the symbols. From (3.8), we can see immediately that $\min(M, N)$ different signals can be recovered at the receiver using standard processing techniques based on linear algebra. This increase in the communication rate is known as a *spatial multiplexing gain*.

The most popular of spatial multiplexing strategies is known as the Bell Labs lAyered Space-Time (BLAST) transmission scheme. It is a simpler version of it, known as (Vertical) V-BLAST that we will discuss here [15, 25]. The challenge of spatial multiplexing lies in the design of the receiver processing algorithms. In V-BLAST, four main approaches are proposed for resolving the transmitted signals.

As before, we can obtain a zero-forcing type estimate of the transmitted signal vector by inverting the channel matrix. The linear detector matrix is given by $\mathbf{W} = (\mathbf{H}^\dagger\mathbf{H})^{-1}\mathbf{H}^\dagger$. The performance of this basic receiver can be improved by detecting the transmitted signals one at a time, and then feeding back the symbol decisions as follows:

$$\mathbf{W} = (\mathbf{H}^\dagger\mathbf{H})^{-1}\mathbf{H}^\dagger \tag{3.9}$$

For j= 1 to N:

$$\overline{s}_j = \mathbf{e}_j^T\mathbf{W}\mathbf{r} \quad \text{(nulling)} \tag{3.10}$$

$$= s_j + \mathbf{e}_j^T\mathbf{W}\mathbf{n} \tag{3.11}$$

$$\widehat{s}_j = \text{Quantize}(\overline{s}_j) \tag{3.12}$$

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n} - \widehat{s}_j\mathbf{h}_j \quad \text{(cancellation)}$$

$$= \mathbf{H}_{\setminus\mathbf{h}_j}\mathbf{s}_{\setminus s_j} + \mathbf{n}, \tag{3.13}$$

where $\mathbf{e}_j$ is an elementary vector of appropriate length (N) in direction j, i.e., $e_{j'} = 1$ if $j' = j$, 0 otherwise. Since independent symbols are transmitted from each of the antennas, $s_j = x_j$ and so we work with $\overline{s}_j$ and $\widehat{s}_j$ here.

Left multiplication by $\mathbf{e}_j^T$ in (3.10) has the effect of selecting the j^{th} row of $\mathbf{W}$, hence only the j^{th} signal is estimated. The quantization operation in (3.12) represents the non-linear decision device. The symbol decision is fed back and subtracted from subsequent detection loops, with (3.13) being based on the assumption that the decision is correct, i.e., that $\widehat{s}_j = s_j$. $\mathbf{H}_{\setminus\mathbf{h}_j}$ and $\mathbf{s}_{\setminus s_j}$ are the channel matrix and transmitted signal vector with the j^{th} column and j^{th} entry, respectively, removed.

The strategy employed by this non-linear receiver is known in the MUD literature as *successive interference cancellation* and in V-BLAST it is referred to as zero-forcing nulling and cancellation. It has also been shown to be equivalent to the *generalized Decision-Feedback Equalizer (DFE)* [25]. To consider the bit error performance of V-BLAST with zero-forcing nulling and cancellation, we derive the SNR of the j^{th} signal from (3.11):

$$\rho_{MIMO,ZF-V-BLAST|\mathbf{h}_j} = \frac{\rho}{\sum_{i=1}^{M} |w_{ji}|^2}. \tag{3.14}$$

It can be shown that

$$\frac{1}{\sum_{i=1}^{M} |h_{ij}|^2} \leq \sum_{i=1}^{M} |w_{ji}|^2 \leq \frac{1}{\min(\sigma_i^2)},$$

where σ_i are the singular values of $\mathbf{H}$. The SNR can then be bounded as follows:

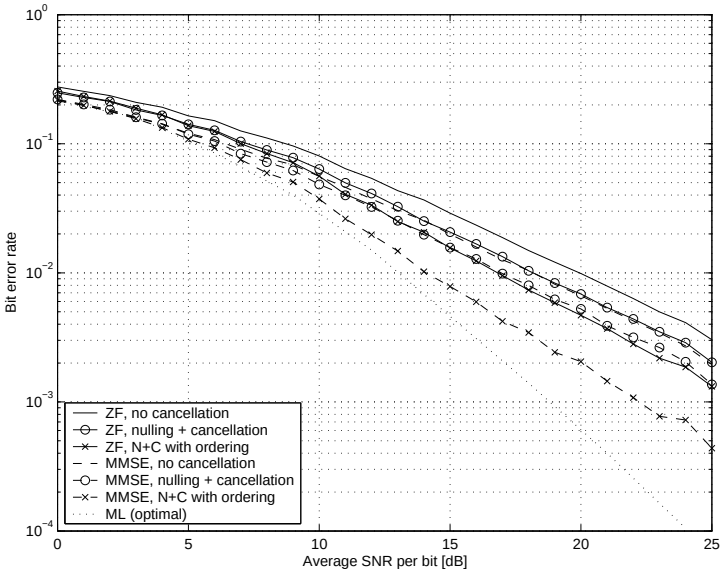
$$\rho \min(\sigma_i^2) \leq \rho_{MIMO,ZF-V-BLAST|\mathbf{h}_j} \leq \rho \sum_{i=1}^{M} |h_{ij}|^2 .$$

Thus the potential diversity advantage offered by V-BLAST is the same as that supported by MRC strategies. Intuitively, we can see that this conclusion is reasonable by considering that the signals transmitted from each of the N antennas are seen by all M receive antennas. Ignoring the difficulty of separating the sources from the observed mixtures, the combined SNR seen at the receiver for each of the transmitted signals is M times that obtained without spatial diversity. This level of diversity corresponds to the slopes of the optimal ML decoder performance curves shown in Figure 3.7.[4] However, we can also see from the plots that the sub-optimal linear receiver structures achieve almost no diversity gain.

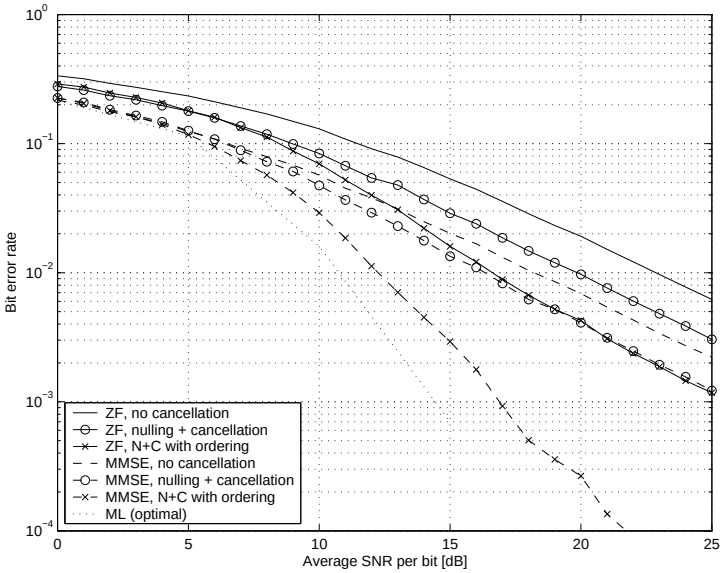
Another observation that can be made from (3.14) is that since the SNR varies inversely with the squared magnitude of the receiver vector, it is not the same for all of the transmitted signals. Given this disparity in the signal SNRs, we expect that better performance can be obtained by choosing the order of detection of the transmitted signals such that those with higher SNRs, i.e., that are the least likely to be in error, are detected first. This strategy is known as zero-forcing nulling and cancellation with ordering, and as can be seen in Figure 3.7, its performance is indeed superior to that obtained when the transmitted signals are detected in an arbitrary order. A proof that the globally optimal solution is attained by applying this ordering is given in [64].

Successive interference cancellation describes the general idea of detecting a set of transmitted symbols sequentially, feeding back symbol decisions after each loop and removing the interference caused by that symbol from the received vector. In the case of zero-forcing nulling, the detector matrix corresponds to the inverse of the channel matrix. Such approaches lead to the problem of noise amplification and generally exhibit poorer performance compared to solutions that take the noise vector into consideration. The well-known best estimate in AWGN is given by the MMSE detector matrix $\mathbf{W} = \mathbf{H}^{\dagger}(\mathbf{H}^{\dagger}\mathbf{H} + \frac{1}{\rho}\mathbf{I}_M)^{-1}$. In Figure 3.7, the performance of the MMSE nulling and cancellation V-BLAST receivers with and without ordering are shown for two and four antenna systems using Quadrature Phase Shift Keying (QPSK) modulation.

[4]The performance data for the ML curves were extracted from [24] and [22], respectively. We may undertake to implement an ML decoder (e.g., the sphere decoder [62]) at some point in the future. The plots for the other curves were generated via simulation.



(a) 2:2 V-BLAST system.



(b) 4:4 V-BLAST system.

Figure 3.7: Bit error rate vs. average SNR per bit for V-BLAST systems in independent flat Rayleigh fading using uncoded coherent QPSK. The curves corresponding to the no diversity case are also shown for comparison.

The zero-forcing and MMSE nulling strategies presented previously are both sub-optimal detectors. Although their performance falls short of that offered by an optimal ML detector, they remain more widely used in practice because of their much lower computational complexities. Comparing the sub-optimal methods, we see that the diversity advantage offered by the zero-forcing receivers is slightly less than 1, and that of the best sub-optimal receiver, MMSE with cancellation and ordering, is nearly as good as that of the ML approach. Studying the two plots together, we observe that the bit error curves become steeper as the number of transmit and receive antennas is increased. However the maximum potential diversity advantage of M is not realized.

Finally, we take a look at the spatial multiplexing gain offered by V-BLAST. In the cases simulated here, we have two transmit, two receive antenna (Fig. 3.7(a)) and four transmit, four receive antenna (Fig. 3.7(b)) systems using QPSK modulation. Both send $Q = N$ symbols per symbol period and thus achieve the desired spatial multiplexing gain of $\min(M, N)$. (If $M \geq N$, all Q symbols can be recovered, otherwise at most M symbols can be reliably transmitted.) This multiplexing gain is said to be optimal, since it is not mathematically possible to separate more than $\min(M, N)$ independent sources from M observed mixtures.

The increased data capacity offered by the V-BLAST scheme does not come without a cost in performance. In Figure 3.8, we compare the bit error curves of V-BLAST to those of an Alamouti STBC operating over the same 2:2 MIMO channel. To make a fair comparison, we would like to transmit using both systems at the same rate. The rate of the Alamouti code can be increased by using higher order modulation. Thus we show simulated performance results for the STBC using QPSK, and BPSK for V-BLAST, giving both schemes maximum rates of 2 b/s/Hz. Also, to maintain reasonable levels of receiver complexity, the V-BLAST MMSE with cancellation and ordering approach is used.

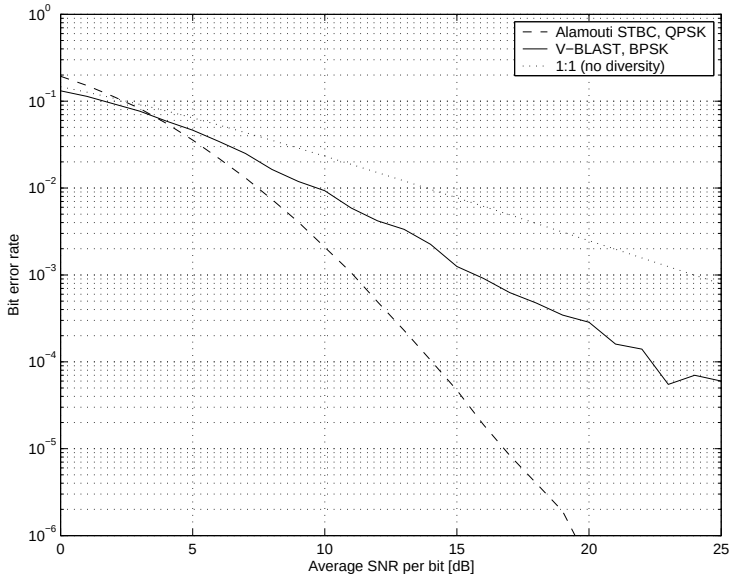


Figure 3.8: Bit error rate vs. average SNR per bit for 2:2 systems in independent flat Rayleigh fading using uncoded coherent modulation: QPSK with the Alamouti STBC and BPSK with V-BLAST. The curve corresponding to the no diversity case is also shown for comparison.

Although it appears that the block code outperforms BLAST at most SNRs, we note that the target rates for state-of-the-art BLAST systems, which are presently under test at Bell Labs, range from 20-40 b/s/Hz [18]. These sorts of rates can easily be achieved using the BLAST architecture, for instance 10 antennas each transmitting symbols from an 8-Quadrature Amplitude Modulation (QAM) alphabet achieve an overall rate of 3 b/s/Hz per spatial dimension, or 30 b/s/Hz. In order to achieve comparable rates with a STBC, an unrealistic modulation scheme supporting 30 bits per symbol would be required.

3.5 Summary

In this chapter we have taken a look at spatial diversity advantage from the perspective of SNR gain. We have seen that the diversity advantage reflects the steepness of the bit error performance curve. To be more precise, in the high SNR regime, it is approximately proportional to the slope of the bit error performance curve on a log-log scale. A number of well-known systems have also been simulated and plots of their performance in the flat quasi-static fading channel are shown throughout.

Table 3.1 summarizes the maximum diversity advantage and rates that can be achieved in the various spatial diversity scenarios.

Spatial diversity scenario	**Maximum diversity advantage**	**Maximum rate**
No diversity	1	1 sym/s/Hz
Receive diversity	M	1 sym/s/Hz
Transmit diversity	N	1 sym/s/Hz
Transmit and receive diversity	NM	$\min(M, N)$ sym/s/Hz

Table 3.1: Summary of achievable performance for different spatial diversity scenarios in flat quasi-static Rayleigh fading.

We stress that of these spatial diversity scenarios, only the MIMO channel can support higher communication rates. This unique benefit is commonly referred to as a spatial multiplexing gain. In the next chapter we will study the information theoretic capacity of MIMO channels and prove that the maximum gain is $\min(M, N)$.

Chapter 4

Capacity of MIMO Rayleigh fading channels

Previously in Chapter 3, we considered the potential improvement in received SNR offered by employing spatial diversity at the transmit or receive antenna arrays or both. Research in these areas was primarily aimed at displacing processing complexity from subscriber units to base stations. It was a key result rediscovered independently by Telatar [61] and Foschini *et al.* [17] that revitalized interest in MIMO channels. These analyses demonstrate that not only can multiple antenna systems provide enhanced transmission reliability, but they are also the key to achieving greatly increased data rates over the wireless channel without any bandwidth expansion.

In this chapter we overview core results published in the literature on the capacity of MIMO channels. We begin by sketching the proof that in a flat, quasi-static, spatially independent fading environment, the capacity of the MIMO channel increases linearly with the smaller of the number of transmit and receive antennas. For completeness, capacity expressions based on the optimal input signal distributions are derived for both the cases where CSI is and is not available at the transmitter. Then we consider the cases of frequency selective and spatially correlated fading scenarios. Another important finding that will be of interest to us is a derivation of the effective maximum capacity attained by some STBCs.

4.1 Flat quasi-static fading channel

We begin by considering the flat quasi-static multi-antenna model given in (3.8), dropping the time-dependence for simplicity in the subsequent derivations:

$$\mathbf{r} = \mathbf{H}\mathbf{s} + \mathbf{n}.$$

Applying the Singular Value Decomposition (SVD) to $\mathbf{H}$ allows us to write the following:

$$\begin{aligned} \mathbf{r} &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\dagger}\mathbf{s} + \mathbf{n} \\ \mathbf{U}^{\dagger}\mathbf{r} &= \boldsymbol{\Sigma}\mathbf{V}^{\dagger}\mathbf{s} + \mathbf{U}^{\dagger}\mathbf{n} \\ \widetilde{\mathbf{r}} &= \boldsymbol{\Sigma}\widetilde{\mathbf{s}} + \widetilde{\mathbf{n}}, \end{aligned} \tag{4.1}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary transformations, thus the distributions of $\widetilde{\mathbf{r}}$, $\widetilde{\mathbf{s}}$ and $\widetilde{\mathbf{n}}$ are equivalent to those of $\mathbf{r}$, $\mathbf{s}$ and $\mathbf{n}$, respectively. $\boldsymbol{\Sigma}$ is a diagonal matrix containing the singular

values of $\mathbf{H}$, i.e., the positive square roots of the eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$. The number of non-zero singular values is $\operatorname{rank}\mathbf{H} \leq \min(M,N)$, effectively decoupling the MIMO channel into at most $\min(M,N)$ parallel complex AWGN fading sub-channels. Note that if the elements of $\mathbf{H}$ are i.i.d., $\operatorname{rank}\mathbf{H} = \min(M,N)$ with probability 1.

Given a total transmitted power constraint of P per symbol period, i.e., $\sum_{j=1}^{N} |\widetilde{s}_j|^2 = \operatorname{tr}\mathsf{E}\left[\widetilde{\mathbf{s}}\widetilde{\mathbf{s}}^\dagger\right] \leq P$, the *instantaneous capacity* of the MIMO channel can be determined by maximizing the mutual information

$$\begin{aligned}\mathcal{I}(\widetilde{\mathbf{s}};\widetilde{\mathbf{r}}) &= \mathcal{H}(\widetilde{\mathbf{r}}) - \mathcal{H}(\widetilde{\mathbf{n}}) \\ &= \log_2 \det\left(\frac{\mathbf{K}_{\widetilde{\mathbf{r}}}}{N_0}\right),\end{aligned}$$

where $\mathcal{H}(x)$ denotes the entropy of random variable x and $\mathbf{K}_{\widetilde{\mathbf{r}}} = \mathbf{K}_{\mathbf{r}}$ is the covariance matrix of the received signal vector:

$$\begin{aligned}\mathbf{K}_{\mathbf{r}} &= \mathsf{E}\left(\mathbf{r}\mathbf{r}^\dagger\right) \\ &= N_0\mathbf{I}_M + \mathbf{H}\mathbf{K}_{\mathbf{s}}\mathbf{H}^\dagger \\ \mathbf{K}_{\widetilde{\mathbf{r}}} &= N_0\mathbf{I}_M + \mathbf{\Sigma}\mathbf{K}_{\widetilde{\mathbf{s}}}\mathbf{\Sigma}.\end{aligned}$$

Thus the capacity is given by

$$\begin{aligned}\mathcal{C}_{|\mathbf{H}} &= \max_{\operatorname{tr}\mathbf{K}\leq\rho} \log_2 \det\left(\mathbf{I}_M + \mathbf{H}\mathbf{K}\mathbf{H}^\dagger\right) \\ &= \max_{\operatorname{tr}\mathbf{K}\leq\rho} \log_2 \det\left(\mathbf{I}_M + \mathbf{\Sigma}\mathbf{K}\mathbf{\Sigma}\right),\end{aligned}$$

This capacity is referred to as an instantaneous capacity because it is conditioned on a particular realization of the fading coefficients $\mathbf{H}$. When CSI is not available at the transmitter, the capacity achieving distribution is the circularly symmetric complex Gaussian with zero mean and covariance matrix $\mathbf{K} = \frac{\rho}{N}\mathbf{I}_N$. In this case, the expression for the capacity is then

$$\mathcal{C}_{|\mathbf{H}} = \log_2 \det\left(\mathbf{I}_M + \frac{\rho}{N}\mathbf{H}\mathbf{H}^\dagger\right). \tag{4.2}$$

If the transmitter has access to CSI, usually obtained through feedback from the receiver, the capacity achieving distribution has a diagonal covariance matrix whose entries are optimized by *water filling* over the inverse squared singular value spectrum of the spatial channel matrix [13, 61]: $k_{ii} = (\mu - \frac{1}{\sigma_i^2})^+$, where μ is chosen such that $\operatorname{tr}\mathbf{K} = \sum_{i=1}^{\min(M,N)} k_{ii} = \rho$. The instantaneous capacity in this case is given by

$$\mathcal{C}_{|\mathbf{H}} = \sum_{i=1}^{\min(M,N)} \left[\log_2(\mu\sigma_i^2)\right]^+.$$

As we do not presume that CSI is available at the transmitter, we will use the simpler form in 4.2 as the instantaneous capacity for the remainder of this section.

The Shannon capacity of the MIMO channel is given by the maximum rate at which error-free communication is possible over all realizations of $\mathbf{H}$. Observe that for all rates, no matter how small, there is a non-zero probability that the channel matrix cannot support

reliable communication at that rate. Thus the Shannon capacity of the Rayleigh fading channel is zero. Instead, we may be interested in its *mean* or *ergodic capacity*, determined by evaluating the expectation

$$\overline{\mathcal{C}} = \mathsf{E}_{\mathbf{H}}\left[\log_2 \det\left(\mathbf{I}_M + \frac{\rho}{N}\mathbf{H}\mathbf{H}^\dagger\right)\right] \tag{4.3}$$

over all possible channel realizations $\mathbf{H}$. We note that this expectation is often more easily taken over $\mathbf{H}\mathbf{H}^\dagger$, which is known as a Wishart matrix when the entries of $\mathbf{H}$ are independent Rayleigh random variables. See [61, 48] and the references therein for more details.

Because the capacity of the Rayleigh fading channel is itself a random variable, it can be characterized by its probability density and distribution functions $f_{\mathcal{C}}(R;\rho) = P(\mathcal{C}_{|\mathbf{H}} = R)$ and $F_{\mathcal{C}}(R;\rho) = P(\mathcal{C}_{|\mathbf{H}} < R)$. In addition to these standard representations, the following more descriptive statistics are also commonly used:

- **Outage probability**. Given target rate R and transmitted power constraint in terms of SNR ρ, the probability that the capacity of the channel is not sufficient to support reliable communication

$$\begin{aligned} P_{OUT}(R;\rho) &= P(\mathcal{C}_{|\mathbf{H}} < R) \\ &= F_{\mathcal{C}}(R;\rho). \end{aligned}$$

- **Complementary Cumulative Distribution Function (CCDF)**. Given transmitted SNR constraint ρ, the distribution over target rates R of the probability that the capacity of the channel is sufficient to support reliable communication

$$\begin{aligned} F_{CCDF}(R;\rho) &= P(\mathcal{C}_{|\mathbf{H}} \geq R) \\ &= 1 - F_{\mathcal{C}}(R;\rho). \end{aligned}$$

- **Supportable rate**. Given a target maximum outage probability and transmitted SNR constraint ρ, the maximum rate at which reliable communication is theoretically possible

$$\begin{aligned} R_{MAX}(\varepsilon;\rho) &= \max R \ s.t. \ P(\mathcal{C}_{|\mathbf{H}} < R) \leq \varepsilon \\ &= F_{\mathcal{C}}^{-1}(\varepsilon;\rho). \end{aligned}$$

  This quantity may also be referred to as *outage capacity* in the literature.

The capacity distribution and related quantities are difficult to evaluate analytically for the MIMO channel. However, in the sections to follow we will briefly consider the results derived for some special cases.

4.1.1 Receive only diversity

When $N = 1$, it is convenient to recall that $\det\left(\mathbf{I}_M + \frac{\rho}{N}\mathbf{H}\mathbf{H}^\dagger\right) = \det\left(\mathbf{I}_N + \frac{\rho}{N}\mathbf{H}^\dagger\mathbf{H}\right)$. Then assuming that the entries of $\mathbf{h}$ are independent complex Gaussian random variables, with independent real and imaginary parts having mean zero and variance $\frac{1}{2}$, we have that $\mathbf{h}^\dagger\mathbf{h}$

is a χ^2_{2M} random variable. A closed form analytical expression can be derived for the probability distribution function of the capacity:

$$\begin{aligned} F_{\mathcal{C},1:M}(R;\rho) &= P(\mathcal{C}_{|\mathbf{h}} < R) \\ &= P\left(\log_2 \det\left[\mathbf{I}_N + \rho \mathbf{h}^\dagger \mathbf{h}\right] < R\right) \\ &= P\left(\log_2\left[1 + \rho\chi^2_{2M}\right] < R\right) \\ &= P\left(\chi^2_{2M} < \frac{2^R - 1}{\rho}\right) \\ &= \frac{\gamma\left(M, \frac{2^R-1}{\rho}\right)}{\Gamma(M)}, \end{aligned} \tag{4.4}$$

where $\gamma(\alpha, x)$ is the incomplete gamma function.[1] This expression for the capacity distribution function allows us to compute and study a number of descriptive statistics of interest.

Figure 4.1 shows the outage probability plotted against target rate for systems with one transmit antenna and a number of receive antennas varying from one to 10. The SNR is set to 15 dB for these simulations. As expected, we see that for a given rate, as the number of receive antennas increases, the outage probability decreases. Also, for a desired outage probability, the maximum achievable rate increases with the number of antennas. However, the rate of this increase is slower than linearly, as will be made more clear in Figure 4.3.

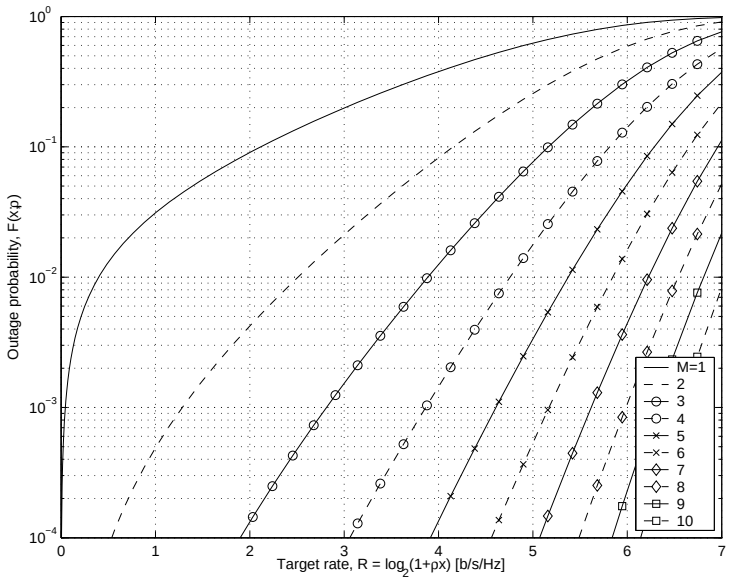


Figure 4.1: Outage probability vs. target rate for 1:M channels in flat quasi-static Rayleigh fading, $\rho = 15$ dB.

Because of the close relationship between the CCDF and the outage probability, the curve depicted in Figure 4.2 is simply the reflection of the previous plot, shown on a linear rather than logarithmic scale. Both are frequently used in the literature to present capacity results.

[1] The incomplete gamma function $\gamma(\alpha, x) = \int_0^x t^{\alpha-1} e^{-t}\, dt$.

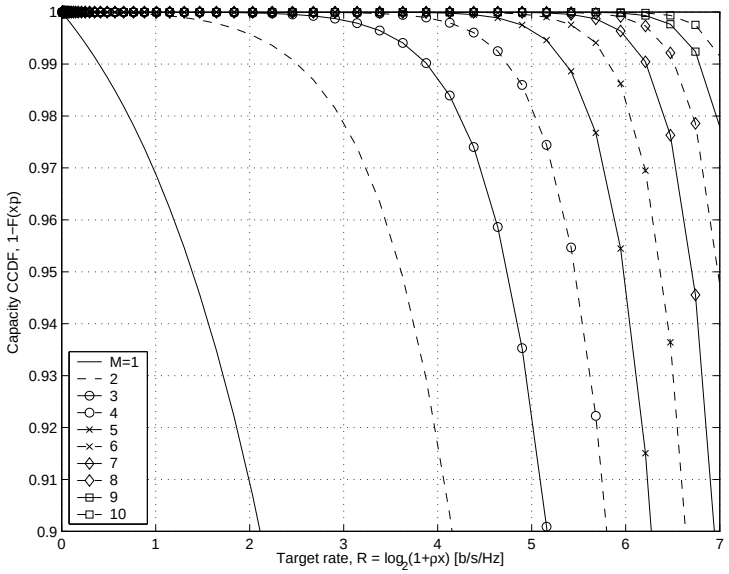


Figure 4.2: Capacity CCDF as a function of target rate for 1:M channels in flat quasi-static Rayleigh fading, $\rho = 15$ dB.

In Figure 4.3, the supportable rate is plotted against the number of receive antennas for target maximum outage probabilities varying from 0.01% to 50%. The SNR is set to 15 dB again. Although an outage probability of 50% does not represent a very high quality link, this curve is shown because as demonstrated by the simulations it provides a good upper bound on the ergodic capacity. Alternatively, the ergodic capacity may be determined by evaluating the integral

$$\overline{\mathcal{C}}_{1:M} = \frac{1}{\Gamma(M)} \int_0^\infty \log_2(1+\rho x) x^{M-1} e^{-x}\, dx. \tag{4.5}$$

From the plot, we can see that for all outage levels, the growth in rate is approximately logarithmic in the number of receive antennas. It has been shown that the ergodic capacity converges to $\log_2(1+\rho M)$ as M becomes large [61]. Also, note that the ergodic capacity does not become a very good estimate of the more applicable outage capacities until the number of antennas is quite large.

4.1.2 Transmit only diversity

When $M = 1$, the capacity of the MISO flat fading channel can be expressed as

$$\begin{aligned} \mathcal{C}_{N:1|\overline{\mathbf{h}}} &= \log_2\left(1 + \frac{\rho}{N}\overline{\mathbf{h}}^T\overline{\mathbf{h}}^*\right) \\ &= \log_2\left(1 + \frac{\rho}{N}\chi^2_{2N}\right), \end{aligned} \tag{4.6}$$

where again we make the assumption that the entries of $\overline{\mathbf{h}}$ are independent complex Gaussian random variables, with independent real and imaginary parts having mean zero and variance $\frac{1}{2}$. Thus $\overline{\mathbf{h}}^T\overline{\mathbf{h}}^*$ is a χ^2_{2N} distributed random variable. A closed form expression for the

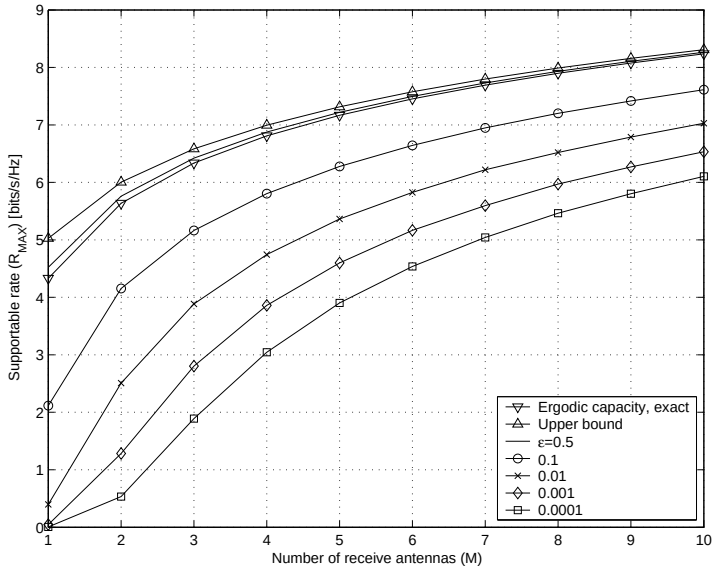


Figure 4.3: Supportable rate vs. number of receive antennas for 1:M channels in flat quasi-static Rayleigh fading, $\rho = 15$ dB.

capacity distribution can then be derived in a manner similar to that used in the receive diversity case:

$$\begin{aligned} F_{\mathcal{C},N:1}(R;\rho) &= P(\mathcal{C}_{N:1|\overline{\mathbf{h}}} < R) \\ &= P\left(\log_2\left[1+\frac{\rho}{N}\chi^2_{2N}\right] < R\right) \\ &= P\left(\chi^2_{2N} < \frac{N[2^R-1]}{\rho}\right) \\ &= \frac{\gamma\left(N, \frac{N[2^R-1]}{\rho}\right)}{\Gamma(N)}. \end{aligned} \tag{4.7}$$

For $N = M > 1$, the distribution of the MISO channel (4.7) is strictly greater than that of the SIMO channel (4.4). Thus its outage probability is higher for any target rate, its CCDF drops off more slowly, and its supportable rate is lower for any desired outage level. It has been shown that the ergodic capacity in this case converges to a constant $\log_2(1+\rho)$ as N becomes large [61].

4.1.3 Combined transmit and receive diversity

In the case where both transmit and receive diversity are available, it is helpful to define $m = \min(M, N)$ and $n = \max(M, N)$. We also note that the at most m non-zero eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$ are the same as those of $\mathbf{H}^\dagger\mathbf{H}$. The ergodic capacity can then be obtained from (4.3) by applying some results on the probability density of eigenvalues of Wishart matrices

$\underline{\lambda}_i = \sigma_i^2$ [61]:

$$\begin{aligned}
\overline{\mathcal{C}}_{N:M} &= \mathsf{E}_{\mathbf{H}}\left[\log_2 \det\left(\mathbf{I}_M + \frac{\rho}{N}\mathbf{H}\mathbf{H}^\dagger\right)\right] \\
&= \mathsf{E}_{\underline{\lambda}_i}\left[\sum_{i=1}^{m} \log_2\left(1 + \rho\frac{\underline{\lambda}_i}{N}\right)\right] \\
&= m\,\mathsf{E}_{\underline{\lambda}}\left[\log_2\left(1 + \rho\frac{\underline{\lambda}}{N}\right)\right] \\
&= m\int_0^\infty \log_2\left(1 + \rho\frac{\underline{\lambda}}{N}\right) p_{\underline{\lambda}}(\underline{\lambda})\, d\underline{\lambda} \\
&\approx m\int_0^4 \log_2(1+\rho x)\frac{1}{\pi}\sqrt{\frac{1}{x} - \frac{1}{4}}\, dx. \qquad (4.8)
\end{aligned}$$

Note that the capacity growth in this case is linear in m, the smaller of the number of receive or transmit antennas. The plot in Figure 4.4 shows how the ergodic capacity grows with the number of receive or transmit antennas for the spatial diversity scenarios considered in this section. The expectations were evaluated numerically using MATLAB® given a received SNR of 15 dB.

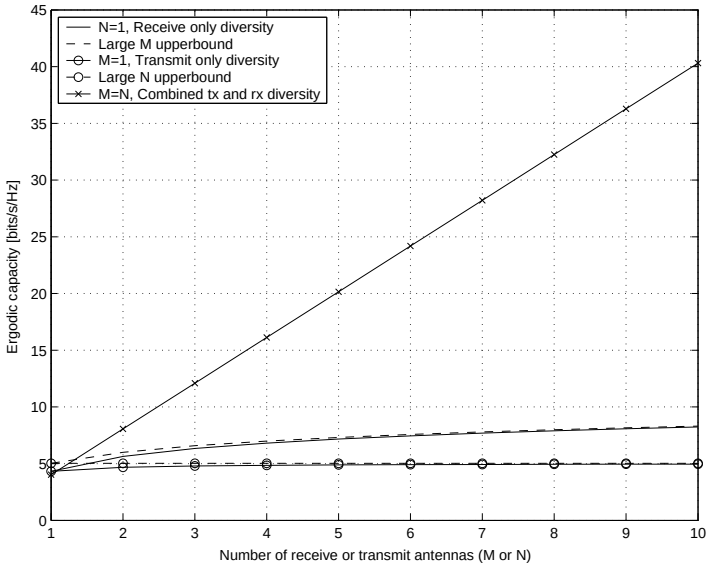


Figure 4.4: Ergodic capacity vs. number of receive/transmit antennas for receive only, transmit only, and combined transmit and receive diversity channels in flat quasi-static Rayleigh fading, $\rho = 15$ dB.

Table 4.1 summarizes the instantaneous and ergodic capacities achievable using the different spatial diversity scenarios.

The ergodic capacity expressions for the no diversity case were derived by evaluating the integral in (4.5) with $M = 1$ and applying the approximations that at low SNR, i.e., $\rho \approx 1$, $\log_2(1+\rho x) \approx \log_2(1+\rho)\log_2(1+x)$, and at high SNR, $\log_2(1+\rho x) \approx \log_2(1+\rho) + \log_2(x)$.

Spatial diversity scenario	Instantaneous capacity	Ergodic capacity
No diversity	$\log_2(1+\rho\chi^2_2)$	$\approx \log_2(1+\rho)\frac{0.2194e}{\ln 2}$ (at low SNR)[a] $\gtrsim \log_2(1+\rho)+\frac{\gamma}{\ln 2}$ ($\approx$ at high SNR)[b]
Receive diversity	$\log_2(1+\rho\chi^2_{2M})$	$\lesssim \log_2(1+\rho M)$ ($\approx$ for large M)
Transmit diversity	$\log_2\left(1+\frac{\rho}{N}\chi^2_{2N}\right)$	$\lesssim \log_2(1+\rho)$ ($\approx$ for large N)
Transmit and receive diversity	$\log_2\det\left(\mathbf{I}_M+\frac{\rho}{N}\mathbf{H}\mathbf{H}^\dagger\right)$	$\approx m\left[\log_2(1+\rho)\frac{0.2194e}{\ln 2}\right]$ (at low SNR) $\gtrsim m\left[\log_2(1+\rho)-\frac{1}{\ln 2}\right]$ ($\approx$ at high SNR)

[a]Integrated numerically, $\int_1^\infty \ln(x)e^{-x}\,dx = 0.2194\ldots$, and $e = 2.7183\ldots$.
[b]The Euler-Mascheroni constant $\gamma = \int_0^\infty \ln(x)e^{-x}\,dx = -0.5772\ldots$.

Table 4.1: Summary of achievable capacity in b/s/Hz for different spatial diversity scenarios in flat quasi-static Rayleigh fading with no transmitter CSI.

These were then verified graphically and extended in a straightforward manner to the equal transmit and receive diversity scenario, by observing that the integral in (4.8) is less than that in (4.5) at all relevant SNRs.

4.2 Frequency selective quasi-static fading channel

If the channel is frequency selective, we consider the case of block transmissions using the DMT or DMMT structures. Its capacity can be derived by first applying the EVD to $\mathbf{H}$. This transformation decouples the channel into $L_f = L$ parallel flat fading sub-channels. The overall capacity can then be determined by averaging over those of the constituent sub-channels. Since the fading in each of the sub-channels is flat, these capacities can be computed using the techniques introduced in the previous section.

4.2.1 No spatial diversity

We begin by taking a look at the SISO frequency selective fading channel. Recall from Section 2.4.2 that the channel matrix $\widetilde{\mathbf{H}}$ is circulant, and its first column is given by the zero-extended CIR $\widetilde{\mathbf{h}} = [\, h^0 \;\cdots\; h^{K-1} \;\; 0 \;\cdots\; 0\,]^T$ of length $L_f = L$. $\widetilde{\mathbf{H}}$ is diagonalized by the unitary inverse DFT matrix, and its eigenvalues $\operatorname{diag}\mathbf{\Lambda}$ correspond to the DFT coefficients of $\widetilde{\mathbf{h}}$ in order.[2] We can then write the following expression for the instantaneous block capacity:

$$\begin{aligned}\mathcal{C}_{1:1|\widetilde{\mathbf{H}}} &= \max_{\operatorname{tr}\mathbf{K}\le L\rho} \log_2\det\left(\mathbf{I}_L+\widetilde{\mathbf{H}}\mathbf{K}\widetilde{\mathbf{H}}^\dagger\right)\\ &= \max_{\operatorname{tr}\mathbf{K}\le L\rho} \log_2\det\left(\mathbf{I}_L+\mathbf{\Lambda}\mathbf{K}\mathbf{\Lambda}^\dagger\right).\end{aligned}$$

Without any CSI available at the transmitter, the capacity achieving distribution is the circularly symmetric complex Gaussian with zero mean and covariance matrix $\mathbf{K} = \rho\mathbf{I}_L$. Observe that a *per block* transmitted energy constraint of $L\rho$ is being applied to the composite channel use of length L symbol periods. This constraint is weaker than the *per symbol* constraint applied to study the flat fading scenario. However, it can be shown that

[2]We note that the unitary DFT matrix is also used by some authors to diagonalize $\widetilde{\mathbf{H}}$. By definition of the EVD, the eigenvalues must remain the same. In this case, the order of the eigenvalues reflects the DFT of the first row of $\widetilde{\mathbf{H}}$, rather than that of the first column.

the capacities derived under both of these constraints are equal [30]. The block capacity is then averaged over the block length for comparison to that of the symbol-based flat fading channel. With this power constraint, the capacity becomes

$$\begin{aligned} \mathcal{C}_{1:1|\widetilde{\mathbf{H}}} &= \frac{1}{L}\log_2 \det\left(\mathbf{I}_L + \rho\mathbf{\Lambda}\mathbf{\Lambda}^\dagger\right) \\ &= \frac{1}{L}\sum_{l=0}^{L-1}\log_2\left(1+\rho\left|\lambda_l\right|^2\right). \end{aligned} \tag{4.9}$$

If the transmitter has access to CSI, the capacity achieving distribution has a diagonal covariance matrix whose entries are optimized by water filling over the inverse power spectrum of the channel: $k_{ll} = (\mu - \frac{1}{|\lambda_l|^2})^+$, where μ is chosen such that $\operatorname{tr}\mathbf{K} = \sum_{l=0}^{L-1} k_{ll} = L\rho$. The capacity in this case is given by

$$\mathcal{C}_{1:1|\widetilde{\mathbf{H}}} = \frac{1}{L}\sum_{l=0}^{L-1}\left[\log_2\left(\mu\left|\lambda_l\right|^2\right)\right]^+.$$

Note that the Toeplitz channel matrix is always square and full rank. Thus block transmission over the SISO frequency selective channel is equivalent to transmission over $L_f = L$ parallel flat fading sub-channels. The gains experienced along each are determined by $|\lambda_l|^2$, the squared magnitudes of the eigenvalues of the channel matrix $\widetilde{\mathbf{H}}$, or the power spectrum of the CIR. Hence the sub-channels may be referred to as *spectral eigenmodes*. However, the magnitudes of some of the eigenvalues may be quite small. These weak eigenmodes correspond to frequency bins where nulls appear in the power spectrum.

Unlike the flat fading channel, the relationships between elements of $\widetilde{\mathbf{H}}$ are not modelled statistically. Instead it is typical for a multipath channel to be specified in terms of its power delay profile $p_{\widetilde{\mathbf{h}}}[k] = \mathsf{E}\left(\left|h^k\right|^2\right)$ for $k = 0, \ldots, K$. Then we can express h^k as $\sqrt{p_{\widetilde{\mathbf{h}}}[k]}\widehat{h}^k$, where $\widehat{h}^k$ are independent complex Gaussian random variables with zero mean and unit variance. Next, we recall that by Parseval's Theorem the sum of the eigenvalues is constrained as follows:

$$\sum_{l=0}^{L-1}|\lambda_l|^2 = L\sum_{k=0}^{K-1}p_{\widetilde{\mathbf{h}}}[k]|\widehat{h}^k|^2.$$

Therefore, the sum in (4.9) is maximized when all of the eigenvalues are equal.[3] Finally, since $\widehat{h}^k$ are i.i.d., the total power is maximized when the delay profile is uniform, i.e., $\sum_{l=0}^{L-1}|\lambda_l|^2 = \frac{L}{K}\sum_{k=0}^{K-1}|\widehat{h}^k|^2$. In this case, the eigenvalues have unit variance and their squared magnitudes are χ^2_{2K} distributed. We then arrive at the following bound for the capacity of the SISO frequency selective channel:

$$\begin{aligned} \mathcal{C}_{1:1|\widetilde{\mathbf{H}}} &\leq \log_2\left(1+\rho\left|\lambda\right|^2\right) \\ &= \log_2\left(1+\frac{\rho}{K}\chi^2_{2K}\right). \end{aligned} \tag{4.10}$$

Note the similarity between (4.10) and the capacity of the MISO flat fading channel (4.6). Thus, the distribution of the lower bound of the SISO frequency selective channel

[3] This conclusion can also be reached by observing that $\log_2(x)$ is a concave function and applying Jensen's Inequality to (4.9).

capacity can be derived in a similar manner as previously:

$$\begin{aligned} F_{\mathcal{C},1:1,FS}(R;\rho) &= P(\mathcal{C}_{1:1|\tilde{\mathbf{H}}} < R) \\ &\geq P\left[\log_2\left(1+\frac{\rho}{K}\chi^2_{2K}\right) < R\right] \\ &= \frac{\gamma\left(K, \frac{K[2^R-1]}{\rho}\right)}{\Gamma(K)}. \end{aligned} \tag{4.11}$$

Equation 4.11 reveals that the outage probability of the SISO frequency selective channel is lower bounded by that of a MISO flat fading channel with $N = K$. We have also confirmed via simulation that in the case of equal power multipath components, this bound is a good average case approximation. In particular, the capacity of the frequency selective fading channel is generally greater than that of the underlying SISO flat fading channel. Equivalently, its outage probability is lower, except at very high target rates where the likelihood of an outage is very high in both cases (e.g., >20%). The following plots illustrate our findings:

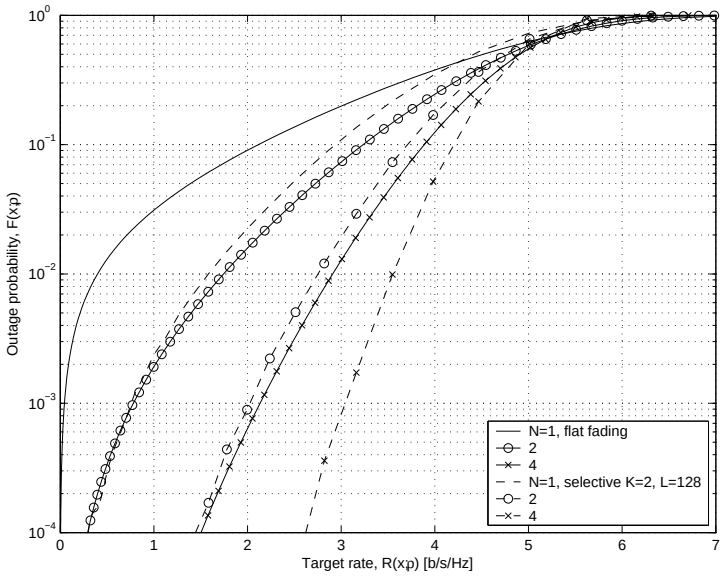
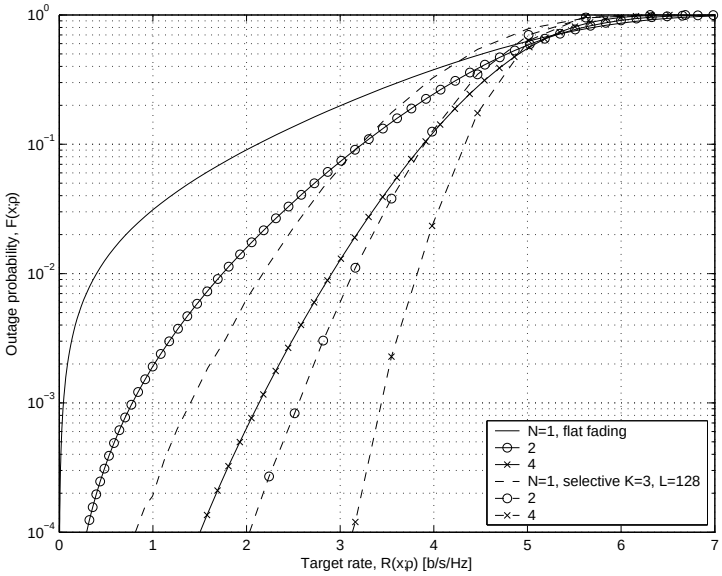


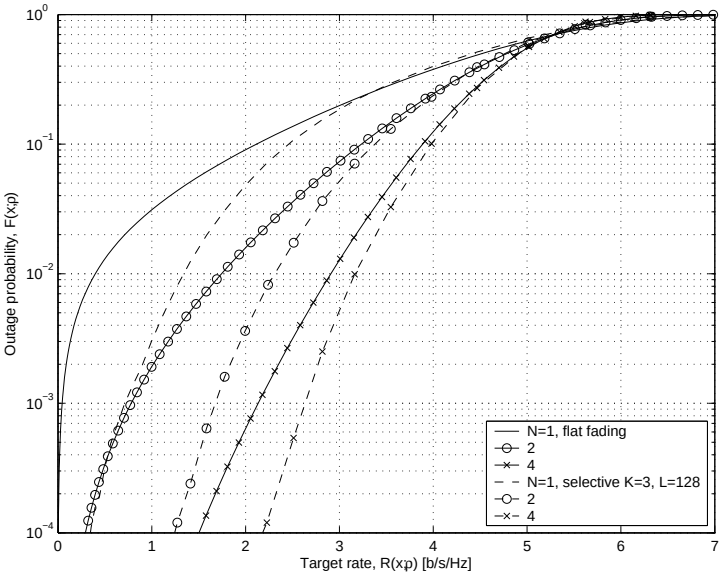
Figure 4.5: Outage probability vs. target rate for N:1 channels in frequency selective quasi-static Rayleigh fading, ρ=15 dB, K=2 equal power multipath components. The corresponding flat fading curves are also shown for comparison.

Figure 4.5 clearly illustrates the relationship between frequency selectivity and MISO spatial diversity by plotting the outage probabilities of two-tap equal power multipath channels alongside those of flat fading channels for various values of N. An SNR of 15 dB is used in the simulations with a block length of L=128. Similar performance curves are observed for different block lengths. Although we have omitted the derivations for the MISO frequency selective fading channel, it is straightforward to show that its capacity is governed by a χ^2_{2KN} random variable. Therefore we observe that the outage curves for the 4:1 flat fading channel and 2:1 frequency selective case with $K = 2$ nearly coincide.

Figure 4.6 shows the outage probabilities of two sample three-tap multipath channels for the same values of N. The flat fading results are also included for comparison. The



(a) Equal power multipath components.



(b) Multipath components with SUI-2 power delay profile (0 dB, -12 dB and 15 dB).

Figure 4.6: Outage probability vs. target rate for N:1 channels in frequency selective quasi-static Rayleigh fading, ρ=15 dB, K=3 multipath components. The corresponding flat fading curves are also shown for comparison.

differences between the two plots arise from the distribution of power over the delay profile of the channel. In Figure 4.6(a) an equal power distribution is applied, and in Figure 4.6(b) the SUI-2 profile of 0 dB, -12 dB and -15 dB is used. Observe that for any target rate, the outage probability of the uniform power profile channel is lower than that of the more realistic scenario. This experiment also has implications for LOS or near LOS channels. Since they have a single dominant tap in their delay profiles, the resulting capacity curves more closely resemble those of Figure 4.6(b). Thus they are less able to take advantage of the enhanced capacity potential offered by frequency selective fading channels, especially at higher target rates.

4.2.2 Combined transmit and receive diversity

Next, we consider the MIMO frequency selective fading channel. As we saw in Section 2.4.2, the block circulant channel matrix is block diagonalized by appropriately sized unitary block inverse DFT matrices. The resulting block diagonal matrix $\mathbf{\Lambda}_{\widetilde{\mathscr{H}}}$ contains the L_f spatial channel matrices corresponding to each of the frequency sub-channels. The elements of these blocks are the flat fading channel coefficients seen along the paths between each transmit-receive antenna pair.

Following [48], we can express the block capacity as follows:

$$\begin{aligned}\mathcal{C}_{N:M\left|\widetilde{\mathscr{H}}\right.} &= \max_{\mathrm{tr}\,\mathcal{K}\leq L\rho} \log_2 \det\left(\mathbf{I}_{ML} + \widetilde{\mathscr{H}}\mathcal{K}\widetilde{\mathscr{H}}^{\dagger}\right) \\ &= \max_{\mathrm{tr}\,\mathcal{K}\leq L\rho} \log_2 \det\left(\mathbf{I}_{ML} + \mathbf{\Lambda}_{\widetilde{\mathscr{H}}}\mathcal{K}\mathbf{\Lambda}^{\dagger}_{\widetilde{\mathscr{H}}}\right).\end{aligned}$$

With the assumption of no CSI at the transmitter, the capacity achieving distribution is given by the circularly symmetric complex Gaussian with zero mean and covariance matrix $\mathcal{K} = \frac{\rho}{N}\mathbf{I}_{NL}$. In this case, the block capacity can be expressed as

$$\begin{aligned}\mathcal{C}_{N:M\left|\widetilde{\mathscr{H}}\right.} &= \log_2 \det\left(\mathbf{I}_{ML} + \frac{\rho}{N}\mathbf{\Lambda}_{\widetilde{\mathscr{H}}}\mathbf{\Lambda}^{\dagger}_{\widetilde{\mathscr{H}}}\right) \\ &= \sum_{l=0}^{L-1} \log_2 \det\left(\mathbf{I}_M + \frac{\rho}{N}\mathbb{H}[l]\mathbb{H}[l]^{\dagger}\right).\end{aligned}$$

To continue with the derivation we need to take a closer look at $\mathbb{H}[l]$. From Section 2.4.2, we have that $\mathbb{H}[l] = \sum_{k=0}^{K-1} \mathbf{H}[k]e^{-j\frac{2\pi}{L}kl}$. Observe that this sum corresponds to an element-wise DFT, i.e., denoting the elements of $M \times N$ matrix $\mathbb{H}[l]$ by $\mathbb{h}_{ij}^{l}$, we have that

$$\begin{aligned}\mathbb{h}_{ij}^{l} &= \sum_{k=0}^{K-1} h_{ij}^{k} e^{-j\frac{2\pi}{L}kl} \\ &= \sum_{k=0}^{K-1} \sqrt{p_{\tilde{\mathbf{h}}}[k]}\,\widehat{h}_{ij}^{k} e^{-j\frac{2\pi}{L}kl},\end{aligned}$$

$$\begin{aligned}\mathsf{E}\left[\mathbb{h}_{ij}^{l}\left(\mathbb{h}_{ij}^{l}\right)^{*}\right] &= \sum_{k=0}^{K-1} p_{\tilde{\mathbf{h}}}[k] \\ &= 1,\end{aligned}$$

by normalization of the power delay profile. Each DFT consists of the sum over a distinct set of K independent circularly symmetric complex Gaussian random variables with mean zero and variance determined by the delay profile.

As before, we can obtain an upper bound on the capacity by assuming a uniform delay profile and observing that since the complex exponential is a unitary operation,

$$\begin{aligned} \widetilde{\mathbf{H}}^l[k] &= \mathbf{H}[k]e^{-j\frac{2\pi}{L}kl} \\ &\sim \mathbf{H}_w. \end{aligned}$$

Therefore the elements of $\mathbb{H}[l]$ are independent circularly symmetric complex Gaussian random variables with mean zero, unit variance, and whose squared magnitudes have $2K$ degrees of freedom. The rotated coefficient matrices $\widetilde{\mathbf{H}}^l[k]$ are also independent over k. The (symbol-based or block averaged) instantaneous capacity of the MIMO frequency selective channel can then be bounded as follows:

$$\begin{aligned} \mathcal{C}_{N:M\left|\widetilde{\mathscr{H}}\right.} &= \frac{1}{L}\sum_{l=0}^{L-1}\log_2\det\left(\mathbf{I}_M + \frac{\rho}{KN}\sum_{k=0}^{K-1}\widetilde{\mathbf{H}}^l[k]\widetilde{\mathbf{H}}^l[k]^\dagger\right) \\ &\leq \frac{1}{L}\sum_{l=0}^{L-1}\sum_{k=0}^{K-1}\log_2\det\left(\mathbf{I}_M + \frac{\rho}{KN}\widetilde{\mathbf{H}}^l[k]\widetilde{\mathbf{H}}^l[k]^\dagger\right) \end{aligned} \tag{4.12}$$

Further simplification is complicated by the fact that $\mathbb{H}[l]$ and $\mathbb{H}(l')$ (hence also $\widetilde{\mathbf{H}}^l[k]$ and $\widetilde{\mathbf{H}}^{l'}[k]$) are not independent for $l \neq l'$. However, we can make the observation that the sum in (4.12) is maximized when all of its arguments are the same, in other words when they are independent of l and k. The resulting upper bound on the ergodic capacity of the frequency selective MIMO channel is given by

$$\overline{\mathcal{C}}_{N:M\left|\widetilde{\mathscr{H}}\right.} \leq K\mathsf{E}_{\mathbf{H}_w}\left[\log_2\det\left(\mathbf{I}_M + \frac{\rho}{KN}\mathbf{H}_w\mathbf{H}_w^\dagger\right)\right],$$

which can be evaluated numerically or approximated using the techniques demonstrated in [61] for flat fading channels. The following preliminary simulations reveal some characteristics of this channel.

Figure 4.7 shows the outage probability of MIMO frequency selective channels where $N = M$. The SNR is 15 dB and there are two equal power multipath components with a block length of 128. Similar performance curves are observed for different numbers of equal and unequal power multipath components and block lengths. The difference between the outage probabilities of the frequency selective and flat fading scenarios can be attributed to the additional degrees of freedom provided by the K multipath components. As in Figure 4.5, Figure 4.7 shows improved outage characteristics exhibited at target rates lower than the approximate ergodic capacity (outage capacity at which $P_{OUT} = \frac{1}{2}$).

Figure 4.8 illustrates the supportable rate of MIMO frequency selective channels where $N = M$. Again an SNR of 15 dB is used in these simulations and there are two equal power multipath components with a block length of 128. We find that at outage levels ranging from 0.1% to 10% the supportable rates available in the frequency selective fading channel are higher than those offered by its flat fading counterpart.

Table 4.2 summarizes the instantaneous capacities and upper bounds for various frequency selective fading scenarios.

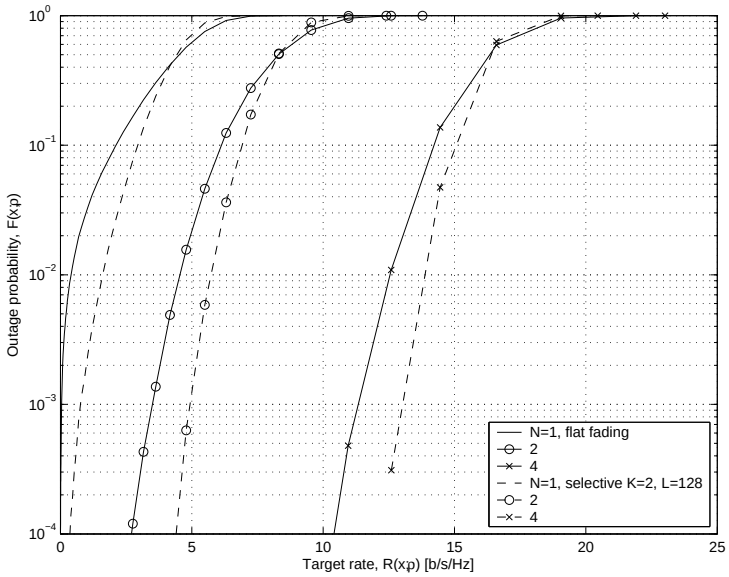


Figure 4.7: Outage probability vs. target rate for $N{=}M$ channels in frequency selective quasi-static Rayleigh fading, ρ=15 dB, K=2 equal power multipath components. The corresponding flat fading curves are also shown for comparison.

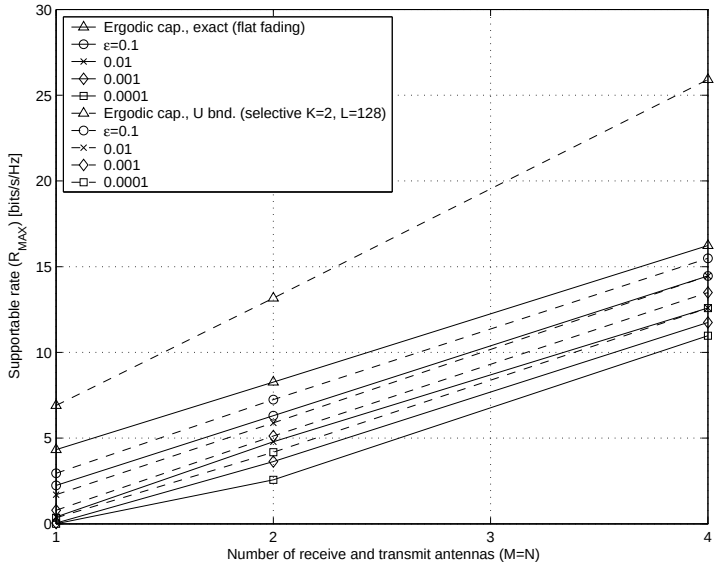


Figure 4.8: Supportable rate vs. number of receive antennas for $N{=}M$ channels in frequency selective quasi-static Rayleigh fading, ρ=15 dB, K=2 equal power multipath components.

Spatial diversity scenario	**Instantaneous capacity**	**Upper bound**
No diversity	$\frac{1}{L}\sum_{l=0}^{L-1}\log_2(1+\rho\left\|\mathbb{h}^l\right\|^2)$	$\log_2\left(1+\frac{\rho}{K}\chi^2_{2K}\right)$
Receive diversity	$\frac{1}{L}\sum_{l=0}^{L-1}\log_2(1+\rho\sum_{i=1}^{M}\left\|\mathbb{h}_i^l\right\|^2)$	$\log_2\left(1+\frac{\rho}{K}\chi^2_{2KM}\right)$
Transmit diversity	$\frac{1}{L}\sum_{l=1}^{L-1}\log_2\left(1+\frac{\rho}{N}\sum_{j=1}^{N}\left\|\mathbb{h}_j^l\right\|^2\right)$	$\log_2\left(1+\frac{\rho}{KN}\chi^2_{2KN}\right)$
Transmit and receive diversity	$\frac{1}{L}\sum_{l=0}^{L-1}\log_2\det\left(\mathbf{I}_M+\frac{\rho}{KN}\sum_{k=0}^{K-1}\widetilde{\mathbf{H}}^l[k]\widetilde{\mathbf{H}}^l[k]^\dagger\right)$	$K\log_2\det\left(\mathbf{I}_M+\frac{\rho}{KN}\mathbf{H}\mathbf{H}^\dagger\right)$

Table 4.2: Summary of achievable capacity and upper bounds in b/s/Hz for different spatial diversity scenarios in frequency selective quasi-static Rayleigh fading with no transmitter CSI.

Finally, we note that there is a small fractional rate loss of $\frac{L}{L+K-1}$ associated with block transmission over frequency selective channels, due to the cyclic prefix which is inserted to combat IBI. This prefix is also necessary for our study of the frequency selective channel from a block-based perspective and the application of the DFT in the analysis. It has been shown that in the limit as $L \to \infty$, the capacity of the DMMT channel considered here converges to that of the more general frequency selective fading scenario [43].

4.3 Flat quasi-static spatially correlated fading channel

From (4.1), we can see that transmission over the MIMO channel is equivalent to transmission over $m = \min(M, N)$ parallel SISO sub-channels. Note that these sub-channels do not represent particular physical paths between the transmit and receive arrays. The gains experienced along each are determined by the squared singular values of the spatial channel matrix $\mathbf{H}$, or the eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$. Hence the sub-channels are referred to as *spatial eigenmodes* [51, 9].

Just as the squared magnitudes of the eigenvalues of the channel impulse response matrix represent gains associated with each of the spectral eigenmodes, the squared magnitudes of the singular values of the spatial channel matrix indicate gains applied to each of the spatial eigenmodes. Spatial nulls occur when the magnitudes of the singular values are small, and also result in degraded overall capacity. In [51], the notion of *effective degrees of freedom* is introduced, roughly characterizing the number of spatial sub-channels over which reliable communication can occur.

4.3.1 Receive only correlation

To study what happens to the capacity when the fading channel is spatially correlated, we begin by taking a look at a typical uplink scenario. In this case, it is assumed that there is only significant spatial correlation at the receiver (basestation). Recall that the receive correlation matrix is Hermitian and positive semi-definite by definition. Following [11] we also make the assumption that the spatial correlation between fading coefficients depends only on the relative positions of the antennas, which are arranged in a uniform linear array. Therefore $(\mathbf{R}_M)_{kl} = (\mathbf{R}_M)_{k-l+1,1}$ for all valid row and column indices, and $\mathbf{R}_M$ is Toeplitz. It has been shown that this is a reasonable structure for the correlation matrix when the angle spread is small [4].

Since the receive correlation matrix is Hermitian and positive semi-definite, its SVD is given by $\mathbf{R}_M = \mathbf{U}_M\mathbf{\Lambda}_M\mathbf{U}_M^\dagger$, i.e., its singular values correspond to its (non-negative real)

eigenvalues and it is diagonalized by unitary matrix $\mathbf{U}_M$. The following expression can then be derived about the channel matrix:

$$\begin{aligned}\mathbf{H}\mathbf{H}^\dagger &= \mathbf{R}_M^{\frac{1}{2}}\mathbf{H}_w\mathbf{H}_w^\dagger\mathbf{R}_M^{\frac{\dagger}{2}} \\ &= \mathbf{\Lambda}_M^{\frac{1}{2}}\mathbf{U}_M\mathbf{H}_w\mathbf{H}_w^\dagger\mathbf{U}_M^\dagger\mathbf{\Lambda}_M^{\frac{1}{2}} \\ &\sim \mathbf{\Lambda}_M\mathbf{H}_w\mathbf{H}_w^\dagger,\end{aligned}$$

where the $\sim$ relation indicates equality in terms of probability distribution. The capacity of this channel, under the assumption of no CSI at the transmitter, is then

$$\mathcal{C}_{|\mathbf{H}} = \sum_{i=1}^{\min[m,\operatorname{rank}\mathbf{R}_M]} \log_2\left(1+\rho\frac{\lambda_{M,i}\sigma_i^2}{N}\right).$$

The received power on each of the parallel spatial sub-channels is now scaled by the non-flat spatial eigenmode spectrum. Further, the number of sub-channels making significant contributions to the overall communication capacity is constrained not only by the smaller of the number of transmit or receive antennas, but also by the rank of the receive correlation matrix. Unlike the previous analysis presented for the frequency selective channel, where $\widetilde{\mathbf{H}}$ was circulant and therefore full rank, the rank of $\mathbf{R}_M$ can become close to unity as the angle spread at the receiver approaches zero [7].

4.3.2 Combined transmit and receive correlation

When both transmit and receive correlation are present, the analysis is complicated by the "sandwiched" transmit correlation matrix. As a result, the channel capacity can only be bounded above and below as shown in [51]:

$$\begin{aligned}\mathbf{H}\mathbf{H}^\dagger &= \mathbf{R}_M^{\frac{1}{2}}\mathbf{H}_w\mathbf{R}_N\mathbf{H}_w^\dagger\mathbf{R}_M^{\frac{\dagger}{2}} \\ &\sim \mathbf{\Lambda}_M\mathbf{H}_w\mathbf{\Lambda}_N\mathbf{H}_w^\dagger \\ \mathcal{C}_{|\mathbf{H}} &\geq \sum_{i=1}^{\min[m,\operatorname{rank}\mathbf{R}_M]} \log_2\left[1+\rho\frac{\lambda_{M,i}}{N}\left(\lambda_{N,i}\left|\mathbf{R}_{ii}\right|^2\right)\right] \qquad (4.13) \\ &\leq \sum_{i=1}^{\min[m,\operatorname{rank}\mathbf{R}_M]} \log_2\left[1+\rho\frac{\lambda_{M,i}}{N}\left(\lambda_{N,i}\left|\mathbf{R}_{ii}\right|^2+\sum_{i'=i+1}^{N}\lambda_{N,i'}\left|\mathbf{R}_{ii'}\right|^2\right)\right], \qquad (4.14)\end{aligned}$$

where $\mathbf{R}$ is the upper triangular matrix arising from the QR decomposition of $\mathbf{H}_w$. Similarly to the previous case, we can see that the spatial sub-channels are scaled by the transmit and receive spatial eigenmode spectra, both of which may be non-flat.

For the case where there is no receive correlation, the authors note that the mean values of the lower (4.13) and upper bounds (4.14) differ by less than 1 b/s/Hz per spatial dimension. The lower bound had previously been derived [15] and shown to converge to the same value as the exact capacity in the large m (minimum number of transmit and receive antennas) limit. In more recent work, Chuah *et al.* prove that even with spatial correlation, the capacity still scales linearly with m [11]. It is also demonstrated via simulation that the reduction in theoretical capacity due to correlated fading is about 10-20%.

4.4 Effective capacity of some STBCs

We saw in Chapter 3 that STBCs such as that proposed by Alamouti enjoy optimal diversity gains and simple linear detection algorithms. However their performance falls short of that offered by spatial multiplexing schemes in terms of realized capacity. In this section, we consider the question of whether a greater capacity is achievable by STBCs.

Typically the term capacity is associated with a particular communication channel rather than a transmission scheme. However, it is also possible to consider the maximum achievable rate of some STBCs, by treating the channel as a standard AWGN channel and re-assigning the received SNR according to the properties of the code. We will refer to this quantity $\mathcal{C}_E = \log_2(1+\rho_{code})$ as the *effective capacity*, where ρ_{code} is the post-detection SNR achieved the code. The derivations presented here are based on results reported in [29] and [46].

We begin by considering the 2:1 Alamouti STBC described in Section 3.3. Recall that the channel matrix is $\overline{\mathbf{h}}^T = [h_1 \;\; h_2]$ and the SNR seen at the receiver is given by $\frac{\rho}{2}(|h_1|^2+|h_2|^2)$. The instantaneous capacity supported by the channel is then

$$\begin{aligned}
\mathcal{C}_{2:1|\overline{\mathbf{h}}} &= \log_2 \det\left(\mathbf{I}_1 + \frac{\rho}{2}\overline{\mathbf{h}}^T\overline{\mathbf{h}}^*\right) \\
&= \log_2\left[1+\frac{\rho}{2}\left(|h_1|^2+|h_2|^2\right)\right] \\
&= \mathcal{C}_{E,Alamouti-2:1}.
\end{aligned}$$

Therefore we can conclude that the Alamouti code is able to achieve the maximum capacity offered by the 2:1 MISO channel. Note that this statement is not meant to imply that the Alamouti STBC does in fact achieve full channel capacity, only that it is not restricted by its structure to some fraction of the available capacity.

Next we take another look at the 2:2 Alamouti code described in Section 3.4. In this case, the channel matrix and received SNR are given by

$$\begin{aligned}
\mathbf{H} &= \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \\
\rho_{Alamouti-2:2} &= \frac{\rho}{2}\left(|h_{11}|^2+|h_{12}|^2+|h_{21}|^2+|h_{22}|^2\right) \\
&= \frac{\rho}{2}\|\mathbf{H}\|_F^2,
\end{aligned}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. However, the instantaneous capacity of the channel is

$$\begin{aligned}
\mathcal{C}_{2:2|\mathbf{H}} &= \log_2 \det\left(\mathbf{I}_2 + \frac{\rho}{2}\mathbf{H}\mathbf{H}^\dagger\right) \\
&= \log_2\left[1+\frac{\rho}{2}\|\mathbf{H}\|_F^2 + \left(\frac{\rho}{2}\right)^2 \det\left(\mathbf{H}\mathbf{H}^\dagger\right)\right] \\
&> \mathcal{C}_{E,Alamouti-2:2},
\end{aligned}$$

where strict inequality follows by defining $\det\left(\mathbf{H}\mathbf{H}^\dagger\right) \triangleq \prod_{i=1}^{\operatorname{rank}\mathbf{H}} \sigma_i^2$. Thus we can see that although the structured use of channel resources made by the Alamouti block codes leads to efficient detection algorithms, there is an implicit sacrifice of effective capacity. This class of codes is only optimal in terms of diversity advantage and rate for the 2:1 MISO channel.

We note that this result was extended in [29] to STBCs based on orthogonal designs ([57], to be overviewed in Chapter 5). One of the first analytical proofs that we will be looking at in our Ph.D. project is the effective capacity of the time-reversal STBCs, which were recently proposed for frequency selective fading channels and will also be discussed in Chapter 5. More details on our ideas for this study are given in Chapter 6.

Chapter 5

Space-time coding

Now that we have studied the underlying phenomena making it possible for MIMO systems employing space-time coding strategies to support greatly enhanced performance, we turn to a more detailed exploration of what researchers in the field have done with this potential. First, to help us understand the relationships between different types of codes, as well as the codes within each class, we give a system-level overview of space-time transmission through generic transmitter and receiver models. This brief section is followed by a survey of some key algorithms and results from the space-time coding literature.

5.1 Transmitter and receiver system models

To formalize the main ideas behind space-time coding and express them in a clear mathematical framework, we begin by presenting generic system models for the space-time transmitter (Fig. 5.1) and receiver (Fig. 5.2). These diagrams encapsulate many of the structures and ideas that are mentioned in the space-time coding literature. More importantly, they provide us with a starting point for studying the similarities and differences between existing approaches, and also lay the groundwork for further discussion and analysis.

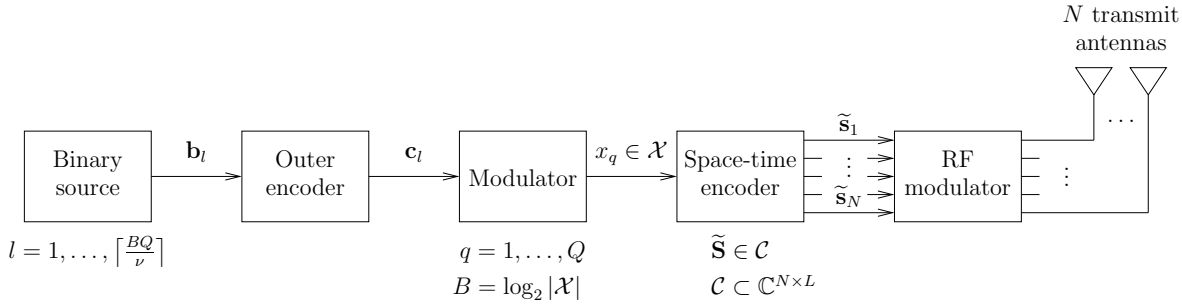


Figure 5.1: System model of generic space-time transmitter.

A space-time transmission starts from the source, which generates κ-bit data vectors $\mathbf{b}_l \in \{0,1\}^\kappa$. For reasons that will become clear, we consider $\left\lceil \frac{BQ}{\nu} \right\rceil$ such vectors at a time. The outer encoder represents a traditional error correcting code of rate $\frac{\kappa}{\nu}$. Therefore it produces from this input $\left\lceil \frac{BQ}{\nu} \right\rceil$ ν-bit codewords $\mathbf{c}_l \in \{0,1\}^\nu$. The modulator is a bit-to-symbol mapper that outputs Q symbols from finite and generally complex alphabet $\mathcal{X} \in \mathbb{C}$, where the modulation order is $B = \log_2 |\mathcal{X}|$. Note that it is possible for the outer encoding

to be done by a coded modulation scheme. In this case we might think of the modulator as a sub-block of the outer encoder.

The next step in our generic system is the space-time encoder. It transforms the Q symbols x_q into N vectors of complex signals to be transmitted from the N antennas. Each is of length L, which is the number of symbol periods that it takes to complete the transmission. These vectors form the rows of space-time signal matrix $\widetilde{\mathbf{S}}$, which is an element of codebook $\mathcal{C} \subset \mathbb{C}^{N \times L}$ of size 2^{BQ}. The overall rate of the code is therefore $\frac{\kappa}{\nu} B \frac{Q}{L}$, where the first term arises from the outer code, the second from the modulation order, and the third from the inner space-time code. We note that whereas the rate of a traditional code is always less than one, that of a space-time code is limited by its spatial multiplexing gain, which as shown previously can exceed one.

We will see that some of the currently proposed space-time techniques can best be described as space-time coded modulation schemes, e.g., a Space-Time Trellis Code (STTC). Effectively these approaches treat the modulator as a sub-block of the space-time encoder. In the discussion to follow, we concentrate on the modulator and space-time encoder blocks. However the others are illustrated here to show how space-time codes fit into the overall design of transmission systems.

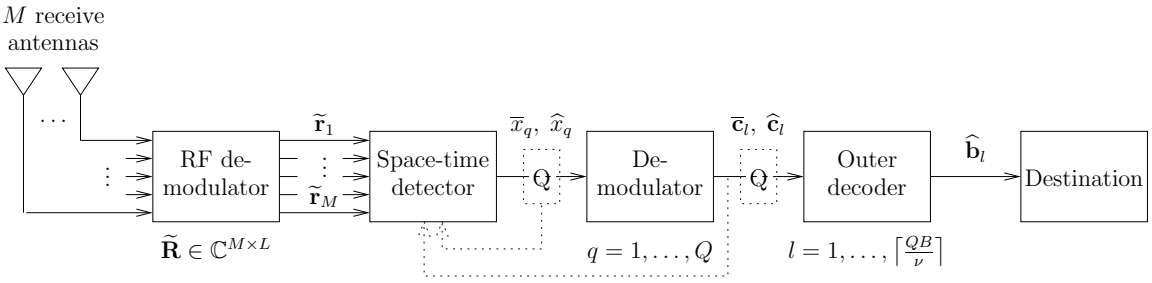


Figure 5.2: System model of generic space-time receiver.

Figure 5.2 depicts the system model of a space-time receiver. As is typical in transmission systems, the receiver blocks perform the inverse operations of their transmitter-side counterparts. We note that the quantizer or (symbol) decision device is shown in two places as a dotted block. It would typically be placed before the demodulator, except in the case of a detection approach making use of soft-decision information, e.g., the soft-decision Viterbi algorithm applied to a STTC. The space-time detector may also make use of feedback from the output of the quantizer, as we saw in the V-BLAST receiver.

5.2 Overview of existing space-time techniques

In this section, we aim to highlight important developments that have been reported in the body of literature on space-time coding. Because the field is growing and changing so rapidly, it is not feasible to assemble a comprehensive up-to-date survey. However, we have selected schemes that are representative of the types of solutions being proposed to address communication over various fading channels. To assist in presenting a logically structured discussion, we have divided the existing works into four major directions of current research interest. These are summarized by the classification tree shown in Figure 5.3. Because the proposed project focuses on STBCs for frequency selective fading channels, our survey will be concentrated on these strategies.

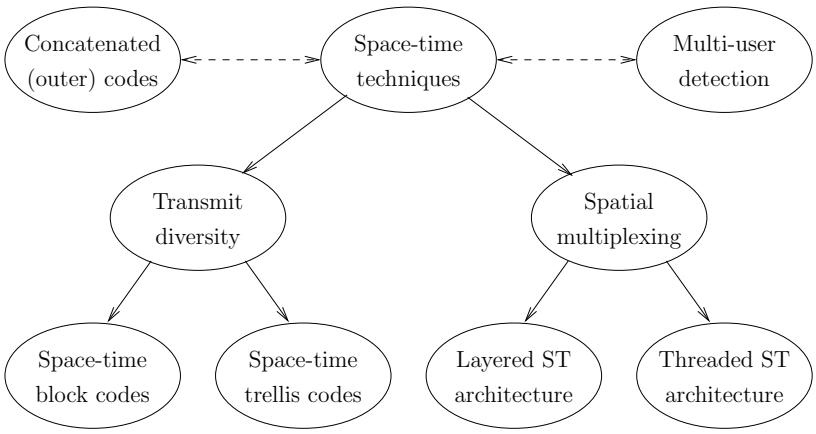


Figure 5.3: Classification of space-time coding techniques and related areas of research.

The four leaf nodes represent areas that have produced interesting recent publications. Two other related fields of study are also shown in the diagram: Concatenated codes, which involve wrapping a generally one-dimensional outer code around an inner space-time technique to improve its performance, and Multi-User Detection (MUD). To keep this report focused on space-time coding, these topics will not be discussed here; we merely mention them for completeness since they are often combined with space-time ideas in the literature.

There are a number of criteria that have been applied in the design and evaluation of space-time codes. Among the most popular are maximizing diversity gain or achievable rate (spatial multiplexing gain). Without going too far into the underlying analysis, we will highlight the performance of the codes that we survey with respect to these parameters, and summarize their properties in a concise tabular format for convenient reference (Table 5.1). We will also briefly consider the decoding complexities of the various schemes, as well as limitations, such as constraints on the number of transmit antennas, and other benefits, such as coding gain.

5.2.1 Space-time block codes

Our discussion of space-time block coding techniques is divided into three sections. The first concentrates on techniques proposed for use over flat quasi-static spatially independent fading channels. These schemes are appropriate for communication systems operating under sufficiently narrowband conditions, so that the flat fading assumption is applicable. The next codes were developed for wideband systems and therefore address issues raised by the frequency selectivity of the channel. Finally, recent approaches to the difficult challenge of coding for spatial correlated channels are briefly overviewed.

Flat quasi-static fading channel

There are three key STBCs that we will discuss in this section. Although new block codes appear in the literature every few months, these three are still widely used as performance benchmarks and form the foundation of insightful new analytical results. The first is the Alamouti code, which we will cover only lightly here since it has already been studied in the previous chapters. Next, we have an extended version of Alamouti's work, which accommodates larger numbers of transmit antennas, proposed by Tarokh *et al.* under the name of *orthogonal designs*. Finally we take a look at the *linear dispersion* codes of Hassibi *et al.*,

which address the capacity limitations of both of these codes and also support arbitrary numbers of transmit antennas.

The Alamouti block code [3] The Alamouti code is the first and probably most well-known STBC. It is designed from the perspective of diversity gain, with the goal of enabling a multiple antenna transmission scheme to achieve the same performance benefits as the optimal SNR multiple antenna MRC receiver. The Alamouti block code succeeds in realizing this desired diversity gain, in the case where there are two transmit antennas, by arranging the symbols and their complex conjugates in a special 2×2 matrix

$$\widetilde{\mathbf{S}} = \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix}.$$

We have already seen that the Alamouti code achieves an optimal diversity gain, while still admitting a linear ML detection algorithm. To understand why this is the case, we highlight two aspects of the design of the space-time encoder.

First, each column of $\widetilde{\mathbf{S}}$ contains the symbols transmitted from the pair of antennas during a particular symbol period. Observe that the second column is a permutation and a reflection of the complex conjugate of the first. Therefore when $\widetilde{\mathbf{S}}$ is transmitted over a flat fading channel, we can write the following:

$$\begin{aligned} \overline{\mathbf{h}}^T\widetilde{\mathbf{S}} &= \begin{bmatrix} \overline{\mathbf{h}}^T\mathbf{x} & \overline{\mathbf{h}}^T\mathbf{P}\mathbf{x}^* \end{bmatrix} \\ \begin{bmatrix} \left(\overline{\mathbf{h}}^T\widetilde{\mathbf{S}}\right)_1 & \left(\overline{\mathbf{h}}^T\widetilde{\mathbf{S}}\right)_2^* \end{bmatrix} &= \begin{bmatrix} \overline{\mathbf{h}}^T & \left(\overline{\mathbf{h}}^T\mathbf{P}\right)^* \end{bmatrix}\mathbf{x}, \end{aligned}$$

where $\mathbf{P}$ is the appropriate permutation and reflection matrix. The manipulation described by these expressions allows the algebraic structure of $\widetilde{\mathbf{S}}$ to be transferred to the effective channel matrix seen by the receiver. One of the key questions that space-time block code design must ask is what kind of structure can be transferred in this manner to the effective channel? For instance, are we limited to permutations and reflections? Another important consideration is to understand what structures are advantageous.

The second ingredient that makes the Alamouti code so effective is its simplicity. Because the algebraic characteristics passed on to the channel give it a unitary structure, ML detection is possible via linear operations. However, as we saw in Chapter 4, it is also this scaled unitary channel matrix that prevents the block code from achieving the full capacity available in the MIMO channel. Thus it seems to be the case that there may not be a single structure that is optimal for all applications.

STBCs based on orthogonal designs [57, 58] Aside from the capacity limitations of the Alamouti STBC, there is also the issue that its design only applies when there are two transmit antennas. The extension of the ideas behind the code to systems with larger numbers of transmit antennas, and also to those employing complex symbol alphabets, is considered by Tarokh *et al.*.

The authors define a *linear processing orthogonal design* in a set of indeterminate variables $x_1, \ldots, x_n$ as an $n \times n$ matrix $\mathbf{X}$ such that the entries of $\mathbf{X}$ are linear combinations of x_i, x_i^*, and $\mathbf{X}^\dagger\mathbf{X} = c\mathbf{I}_n$ for some constant c.[1] We can see immediately that orthogonal designs capture exactly those properties of a channel matrix that make linear ML detection

[1] We have merged the definitions of the real and complex orthogonal designs for brevity.

possible. It can also be shown that this structure results in optimal diversity gain, and in fact the complex orthogonal design (or unitary design) for $N = 2$ is precisely the Alamouti STBC.

The construction and existence of such matrices is explained in detail in [57]. We note here that real orthogonal designs exist only for $N = 2$, 4, and 8. For systems using real symbol alphabets and having other numbers of transmit antennas, *generalized real orthogonal designs* are derived. STBCs based on real designs are rate 1 by definition; a number codes based on generalized real designs are constructed explicitly for $N \leq 8$.

To make these ideas more clear, let us first take a closer look at a real orthogonal design. The code for a $N = 4$ transmit antenna system is given by

$$\widetilde{\mathbf{S}} = \begin{bmatrix} x_1 & -x_2 & -x_3 & -x_4 \\ x_2 & x_1 & x_4 & -x_3 \\ x_3 & -x_4 & x_2 & x_2 \\ x_4 & x_3 & -x_2 & x_1 \end{bmatrix}.$$

Again we see that each column of $\widetilde{\mathbf{S}}$ differs from the first by a permutation and a reflection. Thus when transmitting this code over a slow fading channel, its structure will be transferred to the effective channel matrix, as we saw in the case of the Alamouti STBC. There are $Q = 4$ symbols being sent over $L = 4$ symbol periods, thus the rate of the space-time encoder is 1.

Next we consider a generalized real orthogonal design, for instance the one that is constructed for the $N = 3$ transmit antenna case:

$$\widetilde{\mathbf{S}} = \begin{bmatrix} x_1 & -x_2 & -x_3 & -x_4 \\ x_2 & x_1 & x_4 & -x_3 \\ x_3 & -x_4 & x_1 & x_2 \end{bmatrix}.$$

Although it may seem counter-intuitive at first, this design is also of rate 1. There are $Q = 4$ symbols being sent over $L = 4$ symbol periods. In addition, the receiver still sees $N = 3$ independently faded copies of each symbol. Therefore the optimal diversity gain of NM for any number of receive antennas M is also realized. However we note that the optimal decoding delay, given the constraint $L \geq N$ to ensure that the system of equations at the detector is not underdetermined (see Section 3.3), is $L = N$. Thus the codes based on generalized real designs are not delay optimal.

These ideas can also be extended to systems using complex symbol alphabets. It is shown that *complex orthogonal designs* only exist for $N = 2$, namely the Alamouti STBC. Therefore the analogous notion of *generalized complex orthogonal designs* is derived, and various codes are constructed. However, since codes based on generalized complex designs may nominally be constructed by concatenating the corresponding real design and its complex conjugate, they generally achieve a rate of $\frac{1}{2}$. The generalized design for $N = 4$ is given by

$$\widetilde{\mathbf{S}} = \begin{bmatrix} x_1 & -x_2 & -x_3 & -x_4 & x_1^* & -x_2^* & -x_3^* & -x_4^* \\ x_2 & x_1 & x_4 & -x_3 & x_2^* & x_1^* & x_4^* & -x_3^* \\ x_3 & -x_4 & x_2 & x_2 & x_3^* & -x_4^* & x_2^* & x_2^* \\ x_4 & x_3 & -x_2 & x_1 & x_4^* & x_3^* & -x_2^* & x_1^* \end{bmatrix}.$$

$L = 8$ symbol periods are required to transmit $Q = 4$ symbols, resulting in a significantly reduced rate, especially taking into consideration the increased capacity offered by competitive MIMO schemes such as BLAST. Two closely related works include STBC based on *amicable designs*, which achieve higher rates than those based on orthogonal designs for some

numbers of transmit and receive antennas [23], and quasi-orthogonal STBC, which sacrifice diversity to achieve rate 1 for some scenarios with more than two transmit antennas [34]. However, we note that because these STBCs share the scaled orthogonal/unitary effective channel structure as induced by the Alamouti block code, the same capacity limitations are also inherent in their design.

Linear dispersion codes [29] To realize rates higher than 1 sym/s/Hz using space-time block coded transmission, Hassibi *et al.* study the effective capacity of codes based on orthogonal designs. The authors show that it is not possible for these codes to achieve the maximum capacity supported by the channel. They then develop a new class of block codes designed to maximize the mutual information between the transmitted and received signals. The resulting designs are called *linear dispersion codes.*

The proposed codes can be defined in terms of space-time modulation using a set of $2Q$ *dispersion matrices* $\mathbf{A}_q, \mathbf{B}_q \in \mathbb{C}^{N\times L}$:

$$\widetilde{\mathbf{S}} \;=\; \sum_{q=1}^{Q} \left(x_{Rq}\mathbf{A}_q + jx_{Iq}\mathbf{B}_q\right), \tag{5.1}$$

where the subscript $_R$ is used to denote the real part of a complex-valued structure, and the subscript $_I$ its imaginary component. They subsume two important classes of STBCs, those based on orthogonal designs and V-BLAST, by defining the dispersion matrices appropriately. For instance, if $Q = 2$ and

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

then the linear combination of (5.1) gives

$$\begin{aligned} \widetilde{\mathbf{S}} \;&=\; \begin{bmatrix} x_{R1} + jx_{I1} & -x_{R2} + jx_{I2} \\ x_{R2} + jx_{I2} & -x_{R1} - jx_{I1} \end{bmatrix} \\ &=\; \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix}, \end{aligned}$$

which is precisely the design of the Alamouti STBC. In V-BLAST, $Q = LN$ and $\mathbf{A}_q = \mathbf{B}_q = \mathbf{e}_j\mathbf{e}_l^T$, where j and l are specified by $Q = N(l-1)+j,\ j = 1,\ldots,N,\ l = 1,\ldots,L$. Thus the transmitted signal matrix becomes

$$\widetilde{\mathbf{S}} \;=\; \begin{bmatrix} x_1 & \cdots & x_{N(L-1)+1} \\ \vdots & \ddots & \vdots \\ x_N & \cdots & x_{NL} \end{bmatrix}.$$

What differentiates the linear dispersion code from others that we have considered previously in this report is the approach taken to decoding. It is based on detecting $2Q$ dimensional, rather than N dimensional, vectors of transmitted symbols. A key step in the development of this algorithm is the transformation of the received signal matrix

$\widetilde{\mathbf{R}} = \mathbf{H}\widetilde{\mathbf{S}} + \widetilde{\mathbf{N}}$ into the following form:

$$\widetilde{\mathbf{R}} = \mathbf{H}\sum_{q=1}^{Q}(x_{Rq}\mathbf{A}_q + jx_{Iq}\mathbf{B}_q) + \widetilde{\mathbf{N}}$$

$$\underbrace{\begin{bmatrix}\widetilde{\mathbf{R}}_R^T \\ \widetilde{\mathbf{R}}_I^T\end{bmatrix}}_{\triangleq\underline{\mathcal{R}}} = \sum_{q=1}^{Q}\left(x_{Rq}\underbrace{\begin{bmatrix}\mathbf{A}_{Rq}^T & -\mathbf{A}_{Iq}^T \\ \mathbf{A}_{Iq}^T & \mathbf{A}_{Rq}^T\end{bmatrix}}_{\triangleq\underline{\mathcal{A}}_q}\underbrace{\begin{bmatrix}\mathbf{H}_R^T \\ \mathbf{H}_I^T\end{bmatrix}}_{\triangleq\underline{\mathcal{H}}} + x_{Iq}\underbrace{\begin{bmatrix}-\mathbf{B}_{Iq}^T & -\mathbf{B}_{Rq}^T \\ \mathbf{B}_{Rq}^T & -\mathbf{B}_{Iq}^T\end{bmatrix}}_{\triangleq\underline{\mathcal{B}}_q}\begin{bmatrix}\mathbf{H}_R^T \\ \mathbf{H}_I^T\end{bmatrix}\right) + \underbrace{\begin{bmatrix}\widetilde{\mathbf{N}}_R^T \\ \widetilde{\mathbf{N}}_I^T\end{bmatrix}}_{\triangleq\underline{\mathcal{N}}}$$

$$\mathrm{Vec}(\underline{\mathcal{R}}) = \underbrace{\begin{bmatrix}\underline{\mathcal{A}}_1\underline{\mathcal{H}}_{:1} & \underline{\mathcal{B}}_1\underline{\mathcal{H}}_{:1} & \cdots & \underline{\mathcal{A}}_Q\underline{\mathcal{H}}_{:1} & \underline{\mathcal{B}}_Q\underline{\mathcal{H}}_{:1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \underline{\mathcal{A}}_1\underline{\mathcal{H}}_{:M} & \underline{\mathcal{B}}_1\underline{\mathcal{H}}_{:M} & \cdots & \underline{\mathcal{A}}_Q\underline{\mathcal{H}}_{:M} & \underline{\mathcal{B}}_Q\underline{\mathcal{H}}_{:M}\end{bmatrix}}_{\triangleq\underline{\mathscr{H}}}\mathrm{Vec}\left(\underbrace{\begin{bmatrix}\mathbf{x}_R^T \\ \mathbf{x}_I^T\end{bmatrix}}_{\triangleq\underline{\mathcal{X}}}\right) + \mathrm{Vec}(\underline{\mathcal{N}}). \quad (5.2)$$

$\mathrm{Vec}(\underline{\mathcal{R}})$, $\mathrm{Vec}(\underline{\mathcal{X}})$ and $\mathrm{Vec}(\underline{\mathcal{N}})$ are vectors of length $2LM$, $2Q$ and $2LM$ respectively. The constructed channel matrix $\underline{\mathscr{H}}$ is of size $2LM \times 2Q$ and therefore the overall system of equations can be resolved, i.e., it is not underdetermined, as long as the number of transmitted symbols $Q \leq LM$. This limit agrees with the maximum rate that we expect to be achievable by space-time codes over MIMO channels.

From the structure of (5.2), it is clear that the methods described previously for detecting V-BLAST transmissions are identically applicable to decoding linear dispersion codes. However, generally the complexity of the detector is increased because the dimension of the transmitted symbol vector is higher.

The main drawback of linear dispersion codes is that good designs are not known to follow systematic or algebraic rules. Choosing the dispersion matrices involves choosing a target block rate Q given M, N and L, and then optimizing the effective ergodic capacity

$$\mathcal{C}_{E,LD}(\rho, M, N, L) = \max_{\substack{\mathbf{A}_q,\mathbf{B}_q, \\ q=1,\ldots,Q}} \frac{1}{2L}\mathsf{E}\left[\log_2\det\left(\mathbf{I}_{2LM} + \frac{\rho}{N}\underline{\mathscr{H}}\,\underline{\mathscr{H}}^\dagger\right)\right],$$

subject to a power constraint on $\mathbf{A}_q, \mathbf{B}_q$. The authors note that the problem can be solved using gradient-based optimization methods and have presented a number of codes for various parameter values. They observe that linear dispersion codes are highly non-unique, in particular any $2Q \times 2Q$ orthogonal matrix transformation preserves the mutual information of the code. They show how this property can be used to optimize the performance of the code for other criteria, e.g., diversity gain, without sacrificing effective capacity.

In general, it is shown via simulation that linear dispersion codes optimized first for spatial multiplexing gain and then for diversity gain outperform STBCs based on orthogonal designs and V-BLAST over a wide range of SNRs and target rates. Other authors have also applied the linear dispersion framework to search for good codes, for instance Heath *et al.* construct codes with optimal ergodic capacity that are further optimized for bit error performance [45]. However, the design procedure as it stands cannot be directly extended to multipath fading channels where the symbol period exceeds the coherence time of the channel.

Frequency selective fading channel

There are two main strategies that researchers have applied to generalizing space-time block codes for transmission over frequency selective or multipath fading channels. In the first class

are those techniques for single-carrier modulation systems that focus on reducing equalization complexity. The key proposal in this area is a *time-reversal* approach by Lindskog *et al.* that takes advantage of the space-time code structure to decrease the dimensionality of the equalization step. The second class of techniques are built around block processing operations that effectively convert the frequency selective channel into a set of flat fading sub-channels. These may employ OFDM with multi-carrier modulation or Frequency Domain Equalization (FDE) with single-carrier modulation. A modified STBC is then designed to communicate over the resulting set of frequency sub-channels.

Time reversal (TR) STBC [37, 35] Time reversal block codes are essentially an extension of the class of STBC for flat fading channels based on orthogonal designs. They are designed for use with single-carrier modulation with the goal of simplifying the equalization procedure by decoupling the problem from LN dimensions to N L-dimensional tasks which may be executed in parallel. To gain some insight into how this decoupling is achieved, let us consider the simplest TR-STBC based on the Alamouti code with one receive antenna.

Recall that the original code transmits over $L = 2$ symbol periods using $N = 2$ antennas and an arbitrary number of receive antennas. Each column of the transmitted signal matrix is comprised of the symbols sent from the transmit array during a given symbol period. In a frequency selective fading channel, these symbol columns will suffer Inter-Symbol Interference (ISI) because the delay spread of the channel exceeds the symbol period. Thus the Alamouti code cannot be directly applied in a frequency selective setting without a computationally expensive LN-dimensional Maximum Likelihood Sequence Estimation (MLSE) equalizer.

The TR-STBC strategy involves protecting the data symbol columns by enclosing each of them between guard columns of known symbols. We will refer to these guard blocks as the prefix and suffix, both must be of length at least $K - 1$, and denote by $\widehat{L}$ the net length of the protected data block. It is clear that there is some rate loss associated with the guard blocks, which can be reduced by increasing the size of the data block. However, the maximum size of the data blocks is also limited by the coherence time of the channel.[2] In addition, data columns where complex conjugation is applied in the underlying code are transmitted in time-reversed order, hence the name given to the code. The accompanying guard blocks are also conjugated and time-reversed.

To assist in the following derivations, we define the matrix $\mathbf{I}_n^A$ to be the $n \times n$ permutation matrix with ones along the anti-diagonal. Thus left multiplication by $\mathbf{I}_n^A$ corresponds to order reversal in the row dimension and right multiplication to reversal in the column dimension. Clearly $\mathbf{I}_n^A$ is symmetric. We also recall that matrix multiplication of circulant matrices is commutative and note that $\mathbf{I}_n^A \mathbf{T} \mathbf{I}_n^A = \mathbf{T}^T$ for all Toeplitz matrices $\mathbf{T}$ (of which circulant matrices are a subset).

[2]The authors suggest making use of the guard blocks for channel estimation and also note that because of the requirement for accurate channel estimation, there is some rate loss associated with all coherent coding schemes.

The transmitted signal matrix has the following general structure:

$$\widetilde{\mathbf{S}} = \Bigg[\underbrace{\begin{bmatrix} d_1^{-K+1} & \cdots & d_1^{-1} \\ \vdots & \ddots & \vdots \\ d_N^{-K+1} & \cdots & d_N^{-1} \end{bmatrix}}_{\text{Guard prefix 1}} \underbrace{\widetilde{\mathbf{S}}^{\widehat{L}}(1)}_{\text{Data block 1}} \underbrace{\begin{bmatrix} d_1^{\widehat{L}} & \cdots & d_1^{\widehat{L}+K-2} \\ \vdots & \ddots & \vdots \\ d_N^{\widehat{L}} & \cdots & d_N^{\widehat{L}+K-2} \end{bmatrix}}_{\text{Guard suffix 1}} \cdots$$

$$\cdots \underbrace{\begin{bmatrix} (d_1^{\widehat{L}+K-2})^* & \cdots & (d_1^{\widehat{L}})^* \\ \vdots & \ddots & \vdots \\ (d_N^{\widehat{L}+K-2})^* & \cdots & (d_N^{\widehat{L}})^* \end{bmatrix}}_{\text{Guard prefix } L_0} \underbrace{\widetilde{\mathbf{S}}^{\widehat{L}}(L_0)}_{\text{Data block } L_0} \underbrace{\begin{bmatrix} (d_1^{-1})^* & \cdots & (d_1^{-K+1})^* \\ \vdots & \ddots & \vdots \\ (d_N^{-1})^* & \cdots & (d_N^{-K+1})^* \end{bmatrix}}_{\text{Guard suffix } L_0} \Bigg]$$

Observe that the channel must be sufficiently slowly fading so that it is appropriate to assume that the fading coefficients are constant over $\widetilde{L} = L_0[\widehat{L} + 2(K-1)]$ symbol periods, where as before $\widetilde{L}$ denotes the gross block length including guard symbols, and L_0 is the block length of the underlying STBC design for flat fading.

In the case of the time reversal Alamouti block code, $L_0 = 2$ and the data blocks take the following form:

$$\widetilde{\mathbf{S}}^{\widehat{L}}(1) = \begin{bmatrix} x_1^0 & \cdots & x_1^{\widehat{L}-1} \\ x_2^0 & \cdots & x_2^{\widehat{L}-1} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \end{bmatrix}$$

$$\widetilde{\mathbf{S}}^{\widehat{L}}(2) = \begin{bmatrix} -(x_2^{\widehat{L}-1})^* & \cdots & -(x_2^0)^* \\ (x_1^{\widehat{L}-1})^* & \cdots & (x_1^0)^* \end{bmatrix} = \begin{bmatrix} -\mathbf{x}_2^\dagger \mathbf{I}_{\widehat{L}}^A \\ \mathbf{x}_1^\dagger \mathbf{I}_{\widehat{L}}^A \end{bmatrix},$$

Since $Q = L_0\widehat{L}$ data symbols are transmitted during $\widetilde{L}$ symbol periods, the code has an overall rate of $\frac{\widehat{L}}{\widehat{L}+2(K-1)}$.[3]

In [35] it is argued that the so-called border effects[4] may be ignored asymptotically and are therefore removed from the analysis by making a simplifying approximation. We have found that by defining the prefix and suffix blocks as cyclic extensions of the data block, the border effects are eliminated without the need for any approximations and we may then perform more elegant block analysis using the structures developed in Section 2.4.2. Thus after dropping the prefix and suffix blocks, the received signal can be decoded as follows:

$$\widetilde{\mathbf{r}}(b) = \begin{bmatrix} \widetilde{\mathbf{H}}_1 & \widetilde{\mathbf{H}}_2 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{s}}_1(b) \\ \widetilde{\mathbf{s}}_2(b) \end{bmatrix} + \widetilde{\mathbf{n}}(b), \quad b = 1, 2 \tag{5.3}$$

$$\begin{bmatrix} \widetilde{\mathbf{r}}(1) \\ \mathbf{I}_{\widehat{L}}^A \widetilde{\mathbf{r}}^*(2) \end{bmatrix} = \begin{bmatrix} \widetilde{\mathbf{H}}_1 & \widetilde{\mathbf{H}}_2 \\ \mathbf{I}_{\widehat{L}}^A \widetilde{\mathbf{H}}_2^* \mathbf{I}_{\widehat{L}}^A & -\mathbf{I}_{\widehat{L}}^A \widetilde{\mathbf{H}}_1^* \mathbf{I}_{\widehat{L}}^A \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbf{n}}(1) \\ \mathbf{I}_{\widehat{L}}^A \widetilde{\mathbf{n}}^*(2) \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \widetilde{\mathbf{H}}_1 & \widetilde{\mathbf{H}}_2 \\ \widetilde{\mathbf{H}}_2^\dagger & -\widetilde{\mathbf{H}}_1^\dagger \end{bmatrix}}_{\mathscr{H}} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbf{n}}(1) \\ \mathbf{I}_{\widehat{L}}^A \widetilde{\mathbf{n}}^*(2) \end{bmatrix} \tag{5.4}$$

$$\mathscr{H}^\dagger \begin{bmatrix} \widetilde{\mathbf{r}}(1) \\ \mathbf{I}_{\widehat{L}}^A \widetilde{\mathbf{r}}^*(2) \end{bmatrix} = \left[\mathbf{I}_2 \otimes \left(\widetilde{\mathbf{H}}_1 \widetilde{\mathbf{H}}_1^\dagger + \widetilde{\mathbf{H}}_2 \widetilde{\mathbf{H}}_2^\dagger \right) \right] \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \mathscr{H}^\dagger \begin{bmatrix} \widetilde{\mathbf{n}}(1) \\ \mathbf{I}_{\widehat{L}}^A \widetilde{\mathbf{n}}^*(2) \end{bmatrix} \tag{5.5}$$

[3] For STBCs based on orthogonal designs with more than two transmit antennas, the number of data symbols transmitted during $\widetilde{L}$ symbol periods is $Q \leq L_0\widehat{L}$.

[4] Small error terms arising in the first and last $K-1$ symbols due to undefined inputs at the beginning and end of the block transmission.

The scaled block unitary nature of $\mathscr{H}$ is revealed in (5.5). This structure enables decoupling of the generally LN-dimensional vector equalization of (5.4) into two L-dimensional scalar equalization steps. Since the time-reversed codes are detected in an analogous manner to those based on orthogonal designs, it can be shown that they achieve the optimal diversity gain of NMK, while still sharing the same modest decoding complexity. The main drawback of the TR-STBC approach is its limited rate compared to the potential multiplexing gain available in the MIMO channel. In addition, it still incurs the complexity of L-dimensional MLSE equalization.

STBC with frequency domain processing [38, 2, 67] A number of researchers have also considered extensions of the Alamouti scheme to systems using frequency domain processing. One of the first proposals for combining STBC with OFDM and multi-carrier modulation was put forward by Mudulodu *et al.*. Subsequently, two works based on single-carrier transmission systems with frequency domain processing at the receiver were presented by Al-Dhahir and Zhou *et al.*. All three approaches share substantially similar signal matrix structures and thus we will follow [2] here.

In this work STBC over frequency selective fading channels is proposed in combination with FDE. As we shall see, it exhibits a structure that bears some resemblance to time-reversal, and thus shares many properties of the TR-STBC. The transmitted signal matrix is of the form

$$\begin{aligned}
\widetilde{\mathbf{S}} &= \Big[\underbrace{\begin{bmatrix} s_1^{-K+1} & \cdots & s_1^{-1} \\ s_2^{-K+1} & \cdots & s_2^{-1} \end{bmatrix}}_{\text{Cyclic prefix 1}} \underbrace{\widetilde{\mathbf{S}}^{\widehat{L}}(1)}_{\text{Data block 1}} \underbrace{\begin{bmatrix} s_1^{\widehat{L}} & \cdots & s_1^{\widehat{L}+K-1} \\ s_2^{\widehat{L}} & \cdots & s_2^{\widehat{L}+K-1} \end{bmatrix}}_{\text{Cyclic prefix 2}} \underbrace{\widetilde{\mathbf{S}}^{\widehat{L}}(2)}_{\text{Data block 2}} \Big] \\
\widetilde{\mathbf{S}}^{\widehat{L}}(1) &= \begin{bmatrix} x_1^0 & \cdots & x_1^{\widehat{L}-1} \\ x_2^0 & \cdots & x_2^{\widehat{L}-1} \end{bmatrix} = \begin{bmatrix} \widetilde{\mathbf{x}}_1^T[l] \\ \widetilde{\mathbf{x}}_2^T[l] \end{bmatrix} \\
\widetilde{\mathbf{S}}^{\widehat{L}}(2) &= \begin{bmatrix} -(x_2^0)^* & -(x_2^{\widehat{L}-1})^* & \cdots & -(x_2^1)^* \\ (x_1^0)^* & (x_1^{\widehat{L}-1})^* & \cdots & (x_1^1)^* \end{bmatrix} = \begin{bmatrix} -\widetilde{\mathbf{x}}_2^\dagger[-l \bmod \widehat{L}] \\ \widetilde{\mathbf{x}}_1^\dagger[-l \bmod \widehat{L}] \end{bmatrix},
\end{aligned}$$

Like TR-STBC, this approach involves transmitting a number of protected data blocks. Recall that the SISO frequency selective fading channel matrix is circulant and thus diagonalized by the unitary inverse DFT matrix. Following (5.3) for the case where there is one receive antenna, we can drop the cyclic prefix, diagonalize the channel matrix, and write the following:

$$\widetilde{\mathbb{r}}(b) = \begin{bmatrix} \mathbf{\Lambda}_{\widetilde{\mathbf{H}}_1} & \mathbf{\Lambda}_{\widetilde{\mathbf{H}}_2} \end{bmatrix} \begin{bmatrix} \widetilde{\mathbb{s}}_1(b) \\ \widetilde{\mathbb{s}}_2(b) \end{bmatrix} + \widetilde{\mathbb{n}}(b), \;\; b = 1, 2$$

$$\begin{bmatrix} \widetilde{\mathbb{r}}(1) \\ \widetilde{\mathbb{r}}^*(2) \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{\Lambda}_{\widetilde{\mathbf{H}}_1} & \mathbf{\Lambda}_{\widetilde{\mathbf{H}}_2} \\ \mathbf{\Lambda}^*_{\widetilde{\mathbf{H}}_2} & -\mathbf{\Lambda}^*_{\widetilde{\mathbf{H}}_1} \end{bmatrix}}_{\mathbf{\Lambda}_{\mathscr{H}}} \begin{bmatrix} \widetilde{\mathbb{x}}_1 \\ \widetilde{\mathbb{x}}_2 \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbb{n}}(1) \\ \widetilde{\mathbb{n}}^*(2) \end{bmatrix} \tag{5.6}$$

$$\mathbf{\Lambda}^\dagger_{\mathscr{H}} \begin{bmatrix} \widetilde{\mathbb{r}}(1) \\ \widetilde{\mathbb{r}}^*(2) \end{bmatrix} = \Big[\mathbf{I}_2 \otimes \Big(|\mathbf{\Lambda}_{\widetilde{\mathbf{H}}_1}|^2 + |\mathbf{\Lambda}_{\widetilde{\mathbf{H}}_2}|^2 \Big) \Big] \begin{bmatrix} \widetilde{\mathbb{x}}_1 \\ \widetilde{\mathbb{x}}_2 \end{bmatrix} + \mathbf{\Lambda}^\dagger_{\mathscr{H}} \begin{bmatrix} \widetilde{\mathbb{n}}(1) \\ \widetilde{\mathbb{n}}^*(2) \end{bmatrix} \tag{5.7}$$

where (5.6) follows from properties of the DFT.

The elements of diagonal matrices $\mathbf{\Lambda}_{\widetilde{\mathbf{H}}_j}$ are the DFT coefficients of the channel impulse response from transmit antenna j to the receiver. As we saw in Section 4.2 these random variables have unit variance and under certain assumptions their squared magnitudes are

χ^2_{2K} distributed. Since diversity combining at the receiver enables us to see two copies of the transmitted signals, each scaled by an independent CIR DFT coefficient, the STBC with frequency domain processing provides an SNR of $\rho\left(|\lambda^l_{\widetilde{\mathbf{H}}_1}|^2 + |\lambda^l_{\widetilde{\mathbf{H}}_2}|^2\right)$. Therefore the scheme can achieve a diversity gain of $2K$. Another maximum diversity argument based on the Pair-wise Error Probability (PEP) is given in [67].

We note that the rate achieved by this transmission scheme is fractionally higher than that of the TR-STBC because it does not require a guard suffix block. Commenting on their decoding complexities, we see from (5.7) that not only is the equalization step decoupled in space, but since $\mathbf{\Lambda}_{\widetilde{\mathbf{H}}_j}$ are diagonal it is also decoupled in time. As expected by virtue of the frequency domain processing, no equalization is actually needed. The LN-dimensional problem has been reduced to LN 1-dimensional detection steps. There is however a small computational requirement imposed by the DFT operations.

5.2.2 Space-time trellis codes [60, 59]

One of the primary goals of STBC design is to maximize the diversity gain attained by transmit diversity systems. As such, it is perhaps more appropriate to call them space-time modulation or transmission techniques, since they do not address the error-correction function that is traditionally associated with channel coding in modern communications. For instance, although there is a repetition component in the design of orthogonal STBCs, its purpose is not to serve as a repetition code but rather to improve the received SNR via diversity combining. One way to introduce a coding component into the STBC framework is to wrap an outer error-correction code around the inner space-time modulation. Another is provided by space-time trellis codes.

STTCs were originally proposed by Tarokh *et al.* as an an extension of trellis coding to space-time signal structures. Just as trellis codes impose structure within each codeword (over the code space) and also between codewords transmitted in sequence (over time), STTCs impose similar constraints in physical space and time by distributing the symbols comprising each codeword over the elements of the transmit antenna array. Thus they are able to provide a diversity gain as well as a coding gain.

With reference to Figure 5.1, during each symbol period, the space-time encoder takes as its input one symbol x and generates output symbols $s_1 \cdots s_N$. Note that the encoder is always initialized to its zero state before beginning a codeword transmission. Therefore the block length $L = Q$, and the rate of the STTC is 1 sym/s/Hz. The overall data rate is then specified by the base-2 logarithm B of the symbol alphabet.

The diversity gain of STTCs is determined via a PEP argument. The PEP expresses the probability of transmitting $\widetilde{\mathbf{S}}_c$ and deciding in favour of $\widetilde{\mathbf{S}}_\varepsilon$ at the decoder. Thus it provides a lower bound on the average probability of error, as it only considers errors to a single other codeword. An upper bound on the conditional PEP, given channel coefficients $\mathbf{H}$, can be obtained by applying the Chernoff bound (e.g., see [27]). Observing that the resulting PEP is the product of Independent Identically Distributed (i.i.d.) random expressions, the average PEP can then be written as the product of their expectation. Defining the codeword difference matrix $\mathbf{B} = \widetilde{\mathbf{S}}_c - \widetilde{\mathbf{S}}_\varepsilon$ with SVD $\mathbf{B} = \mathbf{U\Sigma V}^\dagger$ and $r = \operatorname{rank}\mathbf{B}$ a sketch of the

derivation is given as follows:

$$\begin{aligned} p(\widetilde{\mathbf{S}}_c \to \widetilde{\mathbf{S}}_\varepsilon \,|\mathbf{H}) &\leq e^{-d^2(\widetilde{\mathbf{S}}_c \to \widetilde{\mathbf{S}}_\varepsilon|\mathbf{H})\frac{\rho}{4}} \quad \text{(Chernoff bound)} \\ d^2(\widetilde{\mathbf{S}}_c \to \widetilde{\mathbf{S}}_\varepsilon \,|\mathbf{H}) &= \sum_{i=1}^{M} \overline{\mathbf{h}}_i^T \mathbf{B}\mathbf{B}^\dagger \overline{\mathbf{h}}_i^* \\ &= \sum_{i=1}^{M}\sum_{j=1}^{r} |\widetilde{h}_{ij}|^2 \sigma_j^2 \quad (\widetilde{h}_{ij} \sim h_{ij} \sim \chi_2^2) \\ \text{and so} \quad p(\widetilde{\mathbf{S}}_c \to \widetilde{\mathbf{S}}_\varepsilon \,|\mathbf{H}) &\leq \prod_{i=1}^{M}\prod_{j=1}^{r} e^{-|\widetilde{h}_{ij}|^2 \sigma_j^2 \frac{\rho}{4}} \\ p(\widetilde{\mathbf{S}}_c \to \widetilde{\mathbf{S}}_\varepsilon) &< \prod_{i=1}^{M}\prod_{j=1}^{r} \left(\sigma_j^2 \frac{\rho}{4}\right)^{-1} \\ &= \left(\det\left[\mathbf{B}\mathbf{B}^\dagger\right]\right)^{-M} \left(\frac{\rho}{4}\right)^{-Mr}, \end{aligned} \tag{5.8}$$

where (5.8) follows by defining $\det\left(\mathbf{B}\mathbf{B}^\dagger\right) = \prod_{j=1}^{r} \sigma_j^2$.

As we saw in Section 3.2, the exponent of the SNR term in the average probability of error expression (5.8) gives the diversity gain of the code. Also, recall that the *coding gain* is defined as the additional power (in dB) that an uncoded system must transmit in order to achieve the same error performance as the coded system. Therefore from (5.8) we can see that a coding gain of approximately $\gamma = [\det\left(\mathbf{B}\mathbf{B}^\dagger\right)]^{\frac{1}{r}}$ is achieved. The challenges in STTC design lie in maximizing these benefits. To this end Tarokh *et al.*, and Guey *et al.* in an independent work, derived the following fundamental criteria for space-time trellis coding over flat quasi-static fading channels:

- **Rank criterion** To achieve the optimal diversity gain of MN, the block length $L \geq N$ and the codeword difference matrix $\mathbf{B}$ must be full rank over all pairs of codewords. Otherwise the diversity gain is Mr, where

$$r = \min_{\substack{\mathbf{B}=\widetilde{\mathbf{S}}_c - \widetilde{\mathbf{S}}_{c'}, \\ \widetilde{\mathbf{S}}_c, \widetilde{\mathbf{S}}_{c'} \in \mathcal{C}}} \operatorname{rank} \mathbf{B}.$$

  It can be shown that this criterion is equivalent to the *effective Hamming distance* metric applied in the error analysis of trellis coded modulation systems [27].

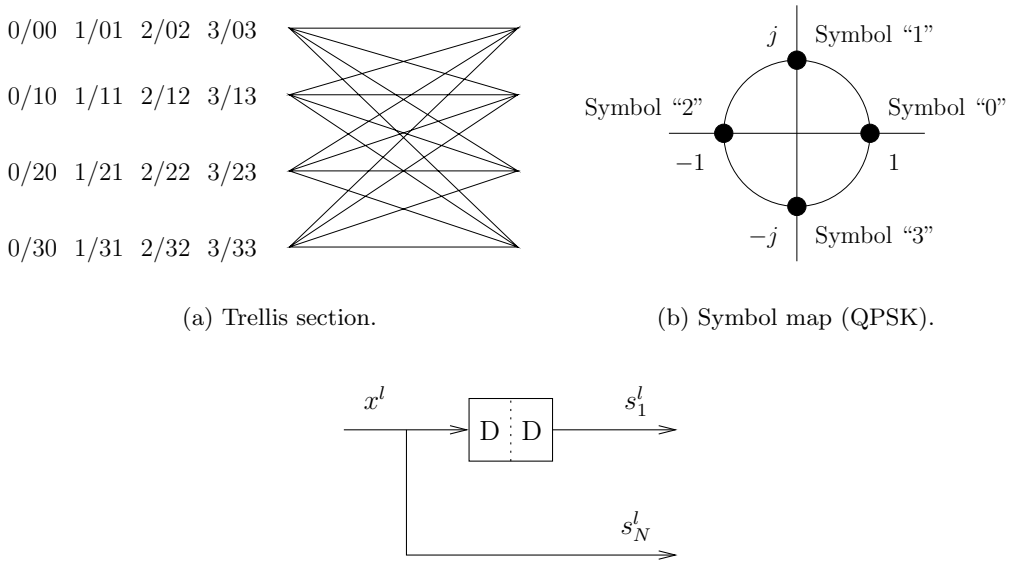
- **Determinant criterion** The coding gain can be optimized by designing the codebook such that the minimum determinant

$$\gamma_{min} = \min_{\substack{\mathbf{B}=\widetilde{\mathbf{S}}_c - \widetilde{\mathbf{S}}_{c'}, \\ \widetilde{\mathbf{S}}_c, \widetilde{\mathbf{S}}_{c'} \in \mathcal{C}}} \det\left(\mathbf{B}\mathbf{B}^\dagger\right),$$

  is maximized. It can be shown that this criterion is equivalent to the *effective product distance* applied in the error analysis of trellis coded modulation systems [27].

Having derived the design criteria for space-time trellis coding, we now turn to a simple illustrative example. Figure 5.4 shows the trellis section, symbol map and convolutional

encoder of a four state STTC designed for a two transmit antenna system using QPSK modulation. It was constructed using two design rules proposed by Tarokh *et al.* that result in optimal diversity designs in some cases (see [60, 27] for more details). The labels to the left of the trellis correspond to the branches leaving each mode, from top to bottom. They show the data symbol and signals transmitted by each of the two transmit antennas as $x^l/s_1^l s_2^l$.



(a) Trellis section.

(b) Symbol map (QPSK).

(c) Convolutional encoder.

Figure 5.4: A four state STTC over the QPSK symbol alphabet for $N = 2$ transmit antennas.

As in standard trellis coding notation, each node corresponds to a particular encoder state. In this example, there are four states, or equivalently the convolutional encoder has two one-bit memory elements.[5] It can be shown that in order to support a diversity gain of Mr, there must be at least $2^{B(r-1)}$ states in the trellis of the STTC. Since they are decoded using the Viterbi Algorithm (VA), which scales exponentially with the number of trellis states, the decoder complexity grows exponentially with the spectral efficiency and diversity of the scheme. This very high complexity is one of the main disadvantages of space-time trellis coding.

The original codes designed by Tarokh *et al.* achieve the optimal diversity gain and some portion of the available coding gain. Since their original proposal, a number of authors have tackled the problem of searching for STTCs with higher coding gains. Some notable results include improved codes found using computer assisted search techniques by Grimm *et al.* [27] and the current state-of-the-art by Yan *et al.* [65], as well as those constructed using more systematic approaches by Baro *et al.* [5] and El Gamal *et al.* [1, 21]. The coding gain offered by these codes is around 1-2 dB higher than that of the first STTCs.

5.2.3 Layered space-time architecture

As introduced in Section 3.4.2, layered space-time systems are designed to optimize spatial multiplexing rather than diversity gain. They generally have $M > N$ receive antennas and

[5]Since a QPSK alphabet is being used, two memory elements are needed to store each delayed symbol.

therefore achieve rates of N sym/s/Hz by transmitting independent sub-streams from each antenna. Although this is not necessarily the case, we will assume that $M > N$ for simplicity in our comments to follow. Some diversity gain is still available via diversity combining at the receiver. As we shall see, some coding gain may also be obtained by applying 1 dimensional outer codes to each of the sub-streams. There are three main approaches that we will discuss here: uncoded V-BLAST, coded (Horizontal) H-BLAST, and coded (Diagonal) D-BLAST.

Vertical BLAST [18, 64] In the vertically layered space-time architecture proposed by Foschini *et al.* at Bell Labs, $Q = NL$ symbols are transmitted over L symbol periods, resulting in a rate of N sym/s/Hz as desired. Since the signals transmitted during each symbol period are independent, i.e., there is no temporal code structure, we can consider detection in each time step separately. Detection is done by a strategy known as *successive interference cancellation*, whereby each of the N symbols is detected in sequence and the hard-decision produced at the end of each loop is used to cancel out interference caused by the detected symbol from the residual observation vector.

The V-BLAST transmitted signal matrix is given by

$$\widetilde{\mathbf{S}} = \begin{bmatrix} s_1^0 & \cdots & s_1^{L-1} \\ \vdots & \ddots & \vdots \\ s_N^0 & \cdots & s_N^{L-1} \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_{N(L-1)+1} \\ \vdots & \ddots & \vdots \\ x_N & \cdots & x_{NL} \end{bmatrix}$$

There are three key tasks performed at the receiver: ordering, nulling and cancellation. The ordering operation involves selecting the order of detection of the symbols at each time step. It is clear that depending on the channel characteristics, the attenuation experienced along each path may vary. As we saw in Section 3.4.2, in order to improve receiver performance, it is advantageous to detect those symbols with higher post-detection SNRs first. These symbols are the least likely to be in error, which is an important property since errors in detected symbols are propagated via the cancellation step to subsequent detection loops.

The nulling step is analogous to the feedforward filter of a generalized Decision-Feedback Equalizer (DFE). Its goal is to produce the best estimate of a particular transmitted symbol, given the presence of interference and noise. This step can be implemented using sub-optimal (i.e., not ML) zero-forcing or MMSE linear detectors, or the sphere decoder. The latter approach is non-linear, ML, and incurs a higher complexity cost than the linear techniques.

The purpose of the cancellation step is to improve the performance of subsequent nulling loops by removing interference caused by the most recently decoded symbol. It is analogous to the feedback filter of a generalized DFE. This step is non-linear since it is based on the hard-decision or quantized symbol. A block diagram of the V-BLAST detector demonstrating this correspondence is shown in Figure 5.5.

We also note that the diversity gain achievable by V-BLAST is potentially M, since M independently faded copies of each transmitted symbol are seen by the receiver. However, based on simulation results presented in Section 3.4.2, the zero-forcing nulling approach yields diversity gains on the order of 1 and only the ML sphere decoder comes close to achieving the maximum achievable gain of M.

Horizontal BLAST [16] The performance of V-BLAST is such that with a moderate number of antennas and reasonably low decoding complexity (e.g., a 4:4 system using an MMSE detector with cancellation and ordering), an average SNR per bit of 15-20 dB is required to achieve an error probability of 10^{-3}. This level is generally accepted as the minimum tolerable bit error rate for voice transmission. To improve upon the performance

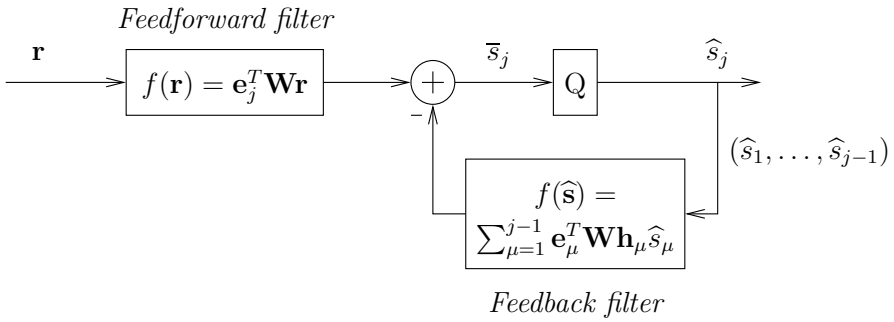


Figure 5.5: The V-BLAST detector as a generalized DFE.

of V-BLAST, one approach is to add more antennas at the receiver, thus making more diversity gain available through observation of an increased number of redundant signal copies.

Another approach is to introduce traditional error control coding in the time dimension, thus improving the overall performance by some amount of coding gain. This strategy is applied in H-BLAST, where $\nu = L$ and $Q = NL$. The $NB\kappa$ information-bearing bits carried in each space-time matrix are first partitioned into N words of length $B\kappa$ and these words are then encoded by rate $\frac{\kappa}{L}$ outer encoders. The resulting codewords $\mathbf{c}_j$ of length BL bits are modulated to produce N codewords ξ_j, each consisting of L symbols. These are transmitted as shown in the following matrix structure:

$$\widetilde{\mathbf{S}} = \begin{bmatrix} \boxed{\begin{matrix} \xi_1^1 & \cdots & \xi_1^L \end{matrix}} \\ \begin{matrix} \vdots & \ddots & \vdots \end{matrix} \\ \begin{matrix} \xi_N^1 & \cdots & \xi_N^L \end{matrix} \end{bmatrix},$$

where the symbols corresponding to the first transmitted codeword are shaded to highlight the horizontal layering that gives the scheme its name.

The H-BLAST receiver tasks are nearly the same as those for V-BLAST, with a few differences: In H-BLAST the ordering step must take into consideration the post-detection SNRs of each complete codeword in making its selection. This change reflects the fact that the subsequent nulling and cancellation operations are based on codeword rather than symbol decisions. In addition, there is a small wrinkle in nulling step, which may be implemented using a symbol-by-symbol detector (as before in V-BLAST). Alternatively, improved performance may be attained by using a sequence estimator or a soft-decision symbol-based approach. Overall the receiver complexity is higher than that of V-BLAST, first because of the additional complexity imposed by the outer code and secondly because of the increased dimensionality incurred by working with codewords rather than symbols.

Clearly the rate of the overall system is decreased by a factor of $\frac{\kappa}{L}$ and the amount of coding gain depends on the particular outer code applied. The diversity gain available in H-BLAST should be the same as that for V-BLAST, although we have not seen any proofs or comments to this effect in the literature.

Diagonal BLAST [15] Finally, we consider the structure and properties of the best and most sophisticated of the BLAST techniques. D-BLAST extends the outer coding introduced in H-BLAST to span both the space and time dimensions. In this case $\nu = N\delta$, $Q = NL$ and $\frac{L\kappa}{\delta}$ information-bearing bits are transmitted per space-time matrix. They are

first partitioned into $\frac{L}{\delta} - (N-1)$ words of length κ and these words are then encoded by rate $\frac{\kappa}{NB\delta}$ outer encoders. The resulting codewords $\mathbf{c}_j$ of length $NB\delta$ bits are modulated to produce codewords ξ_l $(l = 1, \ldots, \frac{L}{\delta} - (N-1))$ each consisting of $N\delta$ symbols. These are arranged for transmission in an $N \times L$ symbol matrix as follows:

$$\widetilde{\mathbf{S}} = \begin{bmatrix} \xi_1^1 & \cdot & \xi_1^\delta & \xi_2^1 & \cdot & \xi_2^\delta & \cdots & \cdots & \cdots & 0 \\ 0 & & & \xi_1^{\delta+1} & \cdot & \xi_1^{2\delta} & \cdots & \cdots & \cdots & \vdots \\ \vdots & & & & \ddots & & \cdots & \cdots & \cdots & 0 \\ 0 & & & \cdots & & & 0 & \xi_1^{(N-1)\delta+1} & \cdot \; \xi_1^{N\delta} & \cdots \; \xi_{\frac{L}{\delta}-(N-1)}^{N\delta} \end{bmatrix},$$

where the symbols corresponding to the first transmitted codeword are shaded to highlight the diagonal layering that gives the scheme its name. The *dwelling* time δ captures the duration in symbol periods of each codeword-transmit antenna association.

In terms of decoding, the D-BLAST detector performs essentially the same operations as those of the other BLAST variants. However, it is termed *balanced* since the symbols constituting each codeword are spread in both space and time. In particular, the order of their detection does not affect performance. As for H-BLAST, nulling and cancellation are based on codeword rather than symbol decisions. The decoding complexity of D-BLAST is comparable to that of H-BLAST.

The rate of the transmission using the D-BLAST architecture is decreased by an outer coding factor of $\frac{\kappa}{NB\delta}$ as well as another factor of $\frac{L-(N-1)\delta}{L\delta}$, owing to the diagonal layering strategy, which limits the number of codewords that can be arranged in the transmission matrix. This second factor is sometimes referred to in the literature as *border wastage*.[6] Simulations reported in the literature show that the D-BLAST architecture achieves higher outage capacities than H-BLAST [16]. It is also argued in [66] that the diversity gain of D-BLAST is $\frac{N(N+1)}{2}$. However, we believe that the gain captured by this expression includes coding gain and the component arising from diversity remains the same as that for V-BLAST.

5.2.4 Threaded space-time architecture [20, 22]

The last space-time architecture that we consider in this section is Threaded Space-Time (TST) proposed by El Gamal *et al.* It was developed to enable the construction of full rate and full diversity MIMO transmission systems by combining layering ideas with constituent space-time codes. It is based on partitioning the space-time signal matrix into non-overlapping threads, which as we shall see have dispersion properties that allow them to be effectively treated as distinct antenna elements.

One obvious advantage is that no thread is associated with a particular set of spatial propagation paths for very long. Thus performance is improved by spreading the symbols of a codeword over space and enabling the decoder to average over the channel attenuations seen by each symbol, just as D-BLAST improved over H-BLAST by balancing out the layers across the spatial dimension.

[6] Although we have not seen this discussed in the literature, it seems that rate loss due to border wastage could be avoided entirely by packing $N-1$ more codewords into the empty slots at the beginning and end of each transmission block. These codewords would experience a delay of L rather than N, since they could not be decoded until the entire block had been received.

In the TST scheme $\nu = L$, $Q = NL$ and $NB\kappa$ information-bearing bits are transmitted by each space-time matrix. The bits are first partitioned into D words of length $B\kappa$, where generally $D \leq N$ but for simplicity we set $D = N$ in this discussion. These words are then encoded by rate $\frac{\kappa}{L}$ outer encoders. These channel codes are referred to in TST as *constituent codes* and generally may be of different rates. The resulting codewords $\mathbf{c}_j$ of length BL bits are modulated to produce D codewords ξ_j, each consisting of L symbols.

Next, each of these D modulated codewords is associated with a *thread*. The thread is the special structure that gives TST its name. Formally a thread is defined as an indexing set $d \subset \{1, \ldots, N\} \times \{0, \ldots, L-1\}$ such that the following conditions are met:

- $(j, l), (j', l') \in d \Rightarrow l \neq l'$ or $j = j'$.
- The *spatial span* defined as $\max\{j \,|\, (j, l) \in d \text{ for some } l\} - \min\{j \,|\, (j, l) \in d \text{ for some } l)\}$ must be full, i.e., N.[7]
- The *temporal span* defined as $\max\{l \,|\, (j, l) \in d \text{ for some } j\} - \min\{l \,|\, (j, l) \in d \text{ for some } j)\}$ must be full, i.e., L.[8]

Since each of the D modulated codewords must be assigned to a thread, we require a set of D threads in the TST architecture. Good standard thread sets have the properties that each thread is active during all L symbol periods, and that over the block each thread is associated with each transmit antenna equally often. Further, a set of N threads should form a partition of $\{1, \ldots, N\} \times \{0, \ldots, L-1\}$, i.e., they should not overlap.

Observe that a thread may be specified by an $N \times L$ matrix $\mathbf{D}$ with the space-time resources that it uses marked by 1 entries. In this context, the thread defining conditions can be summarized as follows: None of the column sums of $\mathbf{D}$ may exceed 1, all of its column sums must be at least 1 and all of its row sums must also be at least 1. The thread set defining conditions imply again that the column sums should be precisely 1, and that the row sums should be equal, i.e., $\approx \frac{L}{N}$. This observation reveals that there are many structural similarities between the TST and linear dispersion frameworks, similarities that we believe remain unexplored in the literature.

As an example, for $N = L = 4$ a standard set of $D = N$ thread matrices is given by

$$\mathbf{D}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{D}_2 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{D}_3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{D}_4 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

[7] I do not believe that any ordering on the spatial antenna elements was intended by the authors, and therefore a better definition of spatial span is simply $\sigma_S = \{j \,|\, (j, l) \in d \text{ for some } l\}$, with full spatial span then being achieved when $|\sigma_S| = N$.

[8] Likewise, I would suggest that the intended meaning of temporal span is $\sigma_T = \{l \,|\, (j, l) \in d \text{ for some } j\}$, and full temporal span is achieved when $|\sigma_T| = L$.

The resulting space-time signal matrix can then be written as

$$\widetilde{\mathbf{S}} = \begin{bmatrix} \xi_1^1 & \xi_4^2 & \xi_3^3 & \xi_2^4 \\ \xi_2^1 & \xi_1^2 & \xi_4^3 & \xi_3^4 \\ \xi_3^1 & \xi_2^2 & \xi_1^3 & \xi_4^4 \\ \xi_4^1 & \xi_3^2 & \xi_2^3 & \xi_1^4 \end{bmatrix},$$

where the symbols corresponding to the first thread are shaded to highlight the threading strategy that gives the scheme its name.

This example enables us to visualize some of the more subtle benefits offered by the threaded architecture. Observe that if the channel experiences spatially independent fading, threading induces temporally independent fading through spatial hopping. The effect of this strategy is somewhat similar to that of using an interleaver to reduce temporal correlation in a slow fading channel. In addition, TST may enable multiple users to share the transmission resource in a more balanced and efficient manner.

Many existing space-time codes, most obviously the family of BLAST solutions, can be captured by the TST framework. The key difference between the TST approach and that of BLAST lies in their decoding strategies. For instance in D-BLAST, although the layering of the transmitted signal matrices may appear to be the same as presented above, detection is effected via a MUD-like receiver based on successive interference cancellation. In TST, space-time codes may be applied over a subset of threads, as if each thread were a distinct antenna element. Thus enabling the threaded architecture to offer the potentially reduced complexity and diversity gains as those realized by its constituent codes.

Analyses are presented in [20, 22] predicting the expected gains supported by the TST framework, and simulation results demonstrating the excellent performance achieved by space-time block codes combined with TST are given in [22]. We are particularly interested in the observation that the thread matrices exhibit very similar structure to the dispersion matrices of linear dispersion space-time codes. It may even be argued that through the definition of its thread criteria, TST provides a way of constraining the design of the dispersion matrices, with added channel coding, to achieve certain performance advantages.[9] It may also be the case that the study of space-time block coding over frequency selective channels will be facilitated by one or the other framework.

5.2.5 Discussion

Our goal in this chapter was to present an overview of some fundamental approaches to space-time coding. A concise summary comparing the surveyed techniques with respect to diversity gain, achievable rate and decoding complexity is provided in Table 5.1. Some other major categories of related work include:

- **Other statistical models for fading.** Although we will use the Rayleigh distribution to model fading coefficients in our work, primarily because of its analytical simplicity, many authors develop and generalize their codes for the Ricean fading channel.

[9] It should be clear that the linear dispersion framework is more general than TST, for instance in that it allows overlapping among the "marked" entries of the dispersion matrices.

- **Space-time codes for correlated fading channels.** It has been shown that when the fading channel exhibits spatial correlation, the achievable diversity gain is limited by its degrees of freedom. More precisely, the optimal diversity gain is reduced from MN to $\text{rank}\,(\mathbf{R}_M)\,\text{rank}\,(\mathbf{R}_N)$ [8]. Proposals for improving performance using concatenated outer codes [52] and maximized diversity designs for block fading channels [19] have recently been proposed. Space-time coding solutions for correlated fading channels are generally referred to in the literature by the term *robust*.

- **Systems with transmitter CSI.** *Feedback* or *closed loop* schemes make use of CSI at the transmitter to enhance code design. There are two main classes of approaches that fall into this category:

  - Matching the power or statistics of the transmitted signals to the channel.
  - Selecting only a subset of the antennas for transmission based on current channel conditions.

- **Channel state estimation.** Most existing space-time codes are designed under the assumption of perfect receiver CSI. Studies have been undertaken to analyze performance degradations in the presence of channel estimation errors and also to take advantage of the structure of some space-time codes to facilitate estimation.

- **Differential space-time coding.** For cases where CSI is not available at the receiver, e.g., in very fast fading environments, differential space-time block coding schemes based on codebooks of unitary space-time signal matrices [31] and unitary group codes [33] have been proposed. The latter codes have the desirable property of being able to achieve the same performance as regular block codes when CSI is available.

Code	**Diversity gain**	**Achievable rate [sym/s/Hz]**	**Decoding complexity**	**Additional comments**
Alamouti STBC	$2M$	1	•	Only for $N = 2$
Orthogonal designs	NM	1	•	$\mathcal{X} \subset \mathbb{R}$ or $N = 2$
	NM	$\left[\frac{1}{2}, 1\right)$	•	$\mathcal{X} \not\subset \mathbb{R}$ and $N > 2$
Linear dispersion	$< NM$	$\lesssim \min(M, N)$	••	
TR-STBC	NML	< 1	•••	Same constraints as orthogonal designs
STBC+frequency domain processing	NML	< 1	••	
Tarokh STTC	NM	1	••••	Provides coding gain
V-BLAST	$\leq M$	$\min(M, N)$	••	
H-BLAST	$\leq M$	$\lesssim \min(M, N)$	•••	Provides coding gain
D-BLAST	$\leq M$	$\lesssim \min(M, N)$	•••	Provides more coding gain than H-BLAST

Table 5.1: Comparative summary of the performance and properties some representative space-time codes.

Our research is concerned with computationally efficient ways of exploiting the MIMO frequency selective fading channel to achieve higher communication rates and improved bit

error performance. Potential application areas include low mobility and fixed broadband wireless access systems. To realize these benefits, some computational cost will almost certainly have to be incurred. Although it is acceptable to perform complicated offline optimizations when designing a code, it is desirable to ensure that decoding complexity is kept at a reasonable level. Therefore we have chosen to focus our research on space-time block coding strategies.

Space-time block coding has been an active area of research for about five years. Existing proposals for the frequency selective fading channel, which we are interested in because of its relevance to wideband transmissions, are primarily extensions of the Alamouti STBC or of those based on orthogonal designs. As such they enjoy optimal diversity gains but suffer from limited data rates of less than 1 sym/s/Hz.

A natural question to ask is how a space-time block code might be designed for frequency selective channels so as to achieve a higher spectral efficiency, while still maintaining a relatively low decoding complexity. This is an issue that has not been well addressed in the literature and that we intend to study in our Ph.D. project. In the next chapter, we provide more details on how we will approach this topic.

Chapter 6

Research Proposal and Activities

The purpose of this chapter is twofold: First, I will summarize the tasks that have been undertaken during the first year of this Ph.D. project to prepare for its successful and timely completion. I believe that the comprehensive coverage of this report itself provides evidence of having developed a solid background in the field of space-time coding. Thus I will strive to be relatively brief in this section. Secondly, I will present a research plan describing some specific tasks and milestones that I anticipate completing during the course of the project.

6.1 Completed activities

The first year began with an in-depth study of diversity and capacity, two fundamental concepts that are undoubtedly key ingredients in space-time coding design. This process also involved reading about important related communication principles such as wireless channel modelling and simulation, ML detection, and coding theory, as well as reviewing core mathematical results in linear algebra and matrix theory. General background material on the wireless channel and applicable signal models is presented in Chapter 2, and further discussion of diversity and capacity, with reference to existing space-time code structures, in Chapters 3 and 4 of this report, as well as in Appendix A.

The following list highlights key activities completed during the first year in preparation for this project:

- **Preliminary analysis of space-time coding structures.** Through the study and presentation of fundamental results on diversity and capacity in Chapters 3 and 4, I have gained greater facility working with the matrix manipulations, inequalities and techniques used to analyze space-time codes. I have also consolidated mathematical models for the systems of interest and assigned consistent naming conventions for the structures involved.

- **Simulation of fundamental space-time block codes.** To demonstrate understanding of the concepts behind and methodologies applicable to space-time code design, as well as to compare the performance of existing techniques, I have conducted numerous preliminary simulations during the first year. These include investigating diversity transmission systems, as well as the Alamouti STBC and V-BLAST space-time algorithms.

- **Preliminary simulations of MIMO channel capacity.** The study of MIMO channel capacities is complicated by the fact that often results cannot be expressed

in closed form or in a form that clarifies the relationships between the capacity and parameters of interest. As illustrated in Chapter 4, preliminary simulations have been conducted to investigate outage probabilities, capacity distributions and ergodic capacities for some of the more challenging MIMO Rayleigh fading scenarios.

- **Capacity study of the frequency selective MIMO channel.** In particular, because of the intended focus of this project on frequency selective fading, I have concentrated my efforts on studying the potential capacity offered by this channel. The original derivations and simulations presented in Section 4.2 may lead to some interesting extensions to the capacity results currently found in the literature.

- **Survey of space-time coding techniques.** Finally, to complement the detailed survey in Chapter 5, which is focused on block coding schemes, I have compiled an extensive bibliography of existing works on general space-time coding, spatial multiplexing and related topics. This more comprehensive listing is available online.

The simulations have been primarily conducted in MATLAB®, and the resulting plots can be found throughout Chapters 3 and 4. With the exception of recent efforts analyzing the capacity of frequency selective fading channels, most of the plots confirm results already published in the literature. However, much insight into the structures and procedures involved in space-time coding has been gained through these valuable exercises, in addition to practical implementation experience that will facilitate development and testing of new space-time designs. Since I chose to focus on two very fundamental schemes, I also expect that the simulation tools created during the first year will continue to serve as useful benchmarks.

6.2 Proposed research plan

During the first year, I acquired a grasp of the theoretical and simulation tools applied in space-time coding research, and prepared a literature survey of key existing works. These studies have led to the conclusion that investigating *space-time block codes for frequency selective fading channels from a capacity perspective* is an important and interesting direction for study that has not yet been well addressed in the literature. It is also an area that I feel well equipped to undertake research in after my readings and preliminary work this year.

The overall goals of my project proposal are first, to develop a novel space-time code for the frequency selective MIMO fading channel that, unlike current proposals, strives to attain a larger portion of the available capacity. Secondly, acknowledging the importance of low complexity for practical applications, the design will be centered around the tradeoff between capacity, diversity and decoding complexity. In the remainder of this section, some specific challenges that I propose to tackle in this Ph.D. project are outlined.

6.2.1 Frequency selective MIMO fading channel study

I will begin by continuing the study of the frequency selective MIMO fading channel. In Chapter 4 we discussed the capacity of this channel, which is reasonably well-understood for the SISO and MISO cases. We were able to obtain a simple expression for the capacity of these channels based on χ^2_{2K} distributed random variables. However, we have not seen any results in the literature on simplified or asymptotic approximations for the capacity of the full frequency selective MIMO channel. In the absence of such expressions, I have conducted simulations and found that the outage probability offered by SISO and MISO

frequency selective channels is lower bounded by that of a MISO flat fading channel with KN transmit antennas. We have also seen that the communication rates supported by the MIMO frequency selective channel are greater than those offered by the underlying flat fading channel.

In addition to capacity benefits, the frequency selective channel provides frequency diversity. This diversity gain can be obtained by using MLSE equalization in the time domain, e.g., the VA, or Frequency Domain Equalization (FDE)(see [47] and the references therein). The FDE is based on the key observation that while the equalization operation in the time domain is based on an L-dimensional vector optimization:

$$\widehat{\mathbf{s}} \;=\; \underset{\widetilde{\mathbf{s}}\in\mathcal{X}^L}{\operatorname{argmin}} \left|\widetilde{\mathbf{r}} - \widetilde{\mathbf{H}}\widetilde{\mathbf{s}}\right|^2,$$

where $\widetilde{\mathbf{H}}$ is a circulant matrix, the corresponding equalization in the frequency domain is based on L scalar optimizations:

$$\mathbf{U}_{DFT,L}\widehat{\mathbf{s}} \;=\; \underset{\widetilde{\mathbb{s}}\in\mathcal{X}_{DFT}^L}{\operatorname{argmin}} \left|\widetilde{\mathbb{r}} - \mathbf{\Lambda}\widetilde{\mathbb{s}}\right|^2$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the DFT coefficients of the channel impulse response and the dual alphabet $\mathcal{X}_{DFT}^L = \left\{\widetilde{\mathbb{s}} \,|\, \widetilde{\mathbb{s}} = \mathbf{U}_{DFT,L}\widetilde{\mathbf{s}},\, \widetilde{\mathbf{s}} \in \mathcal{X}^L\right\}$. This alphabet is difficult to visualize, however since the DFT is a unitary transform, we can apply the inverse DFT and perform the optimizations in the time domain:

$$\widehat{s}^l \;=\; \underset{s^l\in\mathcal{X}}{\operatorname{argmin}} \left|\left(\mathbf{U}_{DFT,L}^{\dagger}\mathbb{W}\mathbf{U}_{DFT,L}\widetilde{\mathbf{r}}\right)^l - s^l\right|^2, \quad l = 0,\ldots,L,$$

where the frequency domain equalizer matrix $\mathbb{W} = (|\mathbf{\Lambda}|^2)^{-1}\mathbf{\Lambda}^{\dagger}$ for a zero-forcing type or $\mathbf{\Lambda}^{\dagger}(|\mathbf{\Lambda}|^2 + \frac{1}{\rho}\mathbf{I}_L)^{-1}$ for an MMSE type operation. Because of its much lower complexity, FDE is a useful tool that I expect to play an important role in this project.

There is also an intriguing analogy between the time domain equalization performed by MISO or MIMO receivers and the decoding of rate $\frac{1}{n}$ or rate $\frac{k}{n}$ convolutional codes. At first glance it would seem that it might be possible to apply the FDE approach directly to the convolutional decoding problem. Thus providing a much lower complexity alternative to the VA in a frequency-like dual code domain. However, I have encountered some difficulty here with respect to the nature of this frequency-like dual domain and the appropriate definition of the DFT, i.e., whether it should be over GF(2) or $\mathbb{C}$ or another field entirely.

A brief sketch of the problem is provided in Appendix C. This study is not the primary focus of the project and is only mentioned here because of its close relationship to equalization for transmit diversity systems in frequency selective fading channels. Of more interest in my work will be the application of FDE and related frequency domain processing concepts to decoding space-time transmissions over frequency selective channels.

6.2.2 Trading off capacity, diversity and complexity

As discussed in Chapter 5, it is clear that the rates of many existing STBC proposals for frequency selective fading channels, i.e., those based on orthogonal designs, are limited to rates of $\lesssim 1$ sym/s/Hz. The first analytical contribution of this project will include a formal proof that this is not merely an incidental property of these codes, but rather a fundamental limitation because of the way in which they manipulate the channel. I expect that as in the analogous analysis for STBCs over flat fading channels, the proof will reveal certain

properties of the induced channel matrix that simultaneously enable full diversity gain and inhibit capacity.

An important question to address is then how a code might be designed to achieve a good diversity gain while also taking advantage of the potential capacity growth offered by the channel. In [29], the authors consider first maximizing mutual information and then applying unitary transformations to increase the diversity gain offered by the resulting codes. This approach seems to have worked well in the flat fading channel and it is one that I will try to apply to the frequency selective case.

In addition, I note an important observation in [24]: that low computational complexities have facilitated uptake of STBCs, particularly the Alamouti code, by telecommunication standards bodies. Therefore I will be interested not only in studying the tradeoff between diversity and capacity, but also that between these properties and decoding complexity. In general, establishing simple expressions governing the tradeoffs between performance and complexity is a very difficult problem. However as we have seen throughout this report, for many space-time codes these parameters are heavily dependent on the characteristics of the channel matrix $\mathbf{H}$ and $\mathbf{H}^\dagger\mathbf{H}$. Thus I hope to be able to make some progress in understanding how they are related within this context.

6.2.3 Design methodology

A number of different criteria have been used to design and evaluate space-time coding schemes in the literature. At first it was believed that optimal diversity gain should be the primary goal. However based on recent publications [24, 29], the focus is shifting toward capacity and computational efficiency.

With respect to the design of a novel space-time code, one approach that I will consider is the extension of techniques such as linear dispersion codes, which have been recently developed to support higher data rates over flat fading channels, to frequency selective channels. Since the frequency selective channel actually provides another dimension of diversity (i.e., frequency diversity), I expect it to be the case that the most effective approaches to code design for systems transmitting over this environment may not be simple extensions of codes for flat fading channels. However, the linear dispersion code framework is very general and I hope that this starting point will give more insight into the limitations and opportunities offered by the frequency selective channel.

6.2.4 Verification of designs via simulation

Ultimately new space-time designs will have to be tested via simulation. I expect to continue using MATLAB® in this work and conduct preliminary simulations over a basic frequency selective quasi-static fading channel. The analysis and simulations presented in this report were done using this channel. It is implemented by generating $M \times N$ spatial channel matrices for each multipath, scaling them by the appropriate power delay profile, and then constructing the block circulant $\mathscr{H}$ as described in Section 2.4.2.

For the capacity simulations in Section 4.2.2, I used a hypothetical uniform power delay profile and one of the three tap power delay profiles specified in the Stanford University Interim (SUI) channel models. In the next stage of the project I may consider using channels specified by the Multiple Element Transmit and Receive Antennas (METRA) projects, as they are based on the four and six tap power delay profiles that are expected to form the core of the 3rd Generation Partnership Project (3GPP) Spatial Channel Model (SCM) standard proposal [49].

The complete SUI and METRA models include a Ricean fading component as well as Doppler spreading. Since it is not the immediate purpose to study the design of space-time codes for these channels, these parameters are likely to remain omitted in future simulations. The use of a simplified channel model is typical of results presented in the literature. Finally, in addition to simulations supporting the development of new space-time code designs, it will likely be necessary to simulate a few existing space-time codes to provide some basis for performance comparison. I anticipate that the TR-STBC and STBC with frequency domain processing techniques will be two good benchmarks for the proposed work.

6.2.5 Additional topics

Should promising results arise from our work designing block codes for frequency selective fading channels, there are a few avenues of additional research that may be of interest. First it will be informative to consider extensions to other fading scenarios such as combined frequency selective and correlated fading. This particular case has been studied in the context of OFDM-like space-time codes and preliminary results showed that in some cases the diversity benefit provided by the frequency selectivity of the channel could be used to compensate for the loss of degrees of freedom caused by spatial correlation. Initially this study would involve simulating a spatially correlated channel and investigating the performance of our codes.

Another interesting consideration is the applicability of any techniques that are developed to applications for multi-access channels. This theme is already beginning to emerge in the literature, for instance in architectures such as threaded space-time, where the notion of a constituent code is easily extended to the code of a particular user in a multi-user system.

Finally, related to the study of space-time block coding schemes that can achieve spectral efficiencies of more than 1 sym/s/Hz is that of space-time trellis coding schemes capable of realizing the same goal. As we saw in Section 5.2.2, current STTC designs are focused on maximizing diversity and coding gains. At the heart of each lies a simple rate $\frac{1}{N}$ convolutional encoder. There do not seem to be any obvious reasons why higher rate encoders could not be used. Since the VA scales with the number of states in the trellis, the decoder complexity should remain comparable to that of rate 1 STTCs, as long as design criteria are devised to ensure that the number of memory elements in the encoder remains as low as possible. It has been shown that in order to achieve higher rates, some diversity gain must be sacrificed [60].

The proposed research plan is summarized visually with realistic scheduling details, including time allocated for writing tasks, in Figure 6.1.

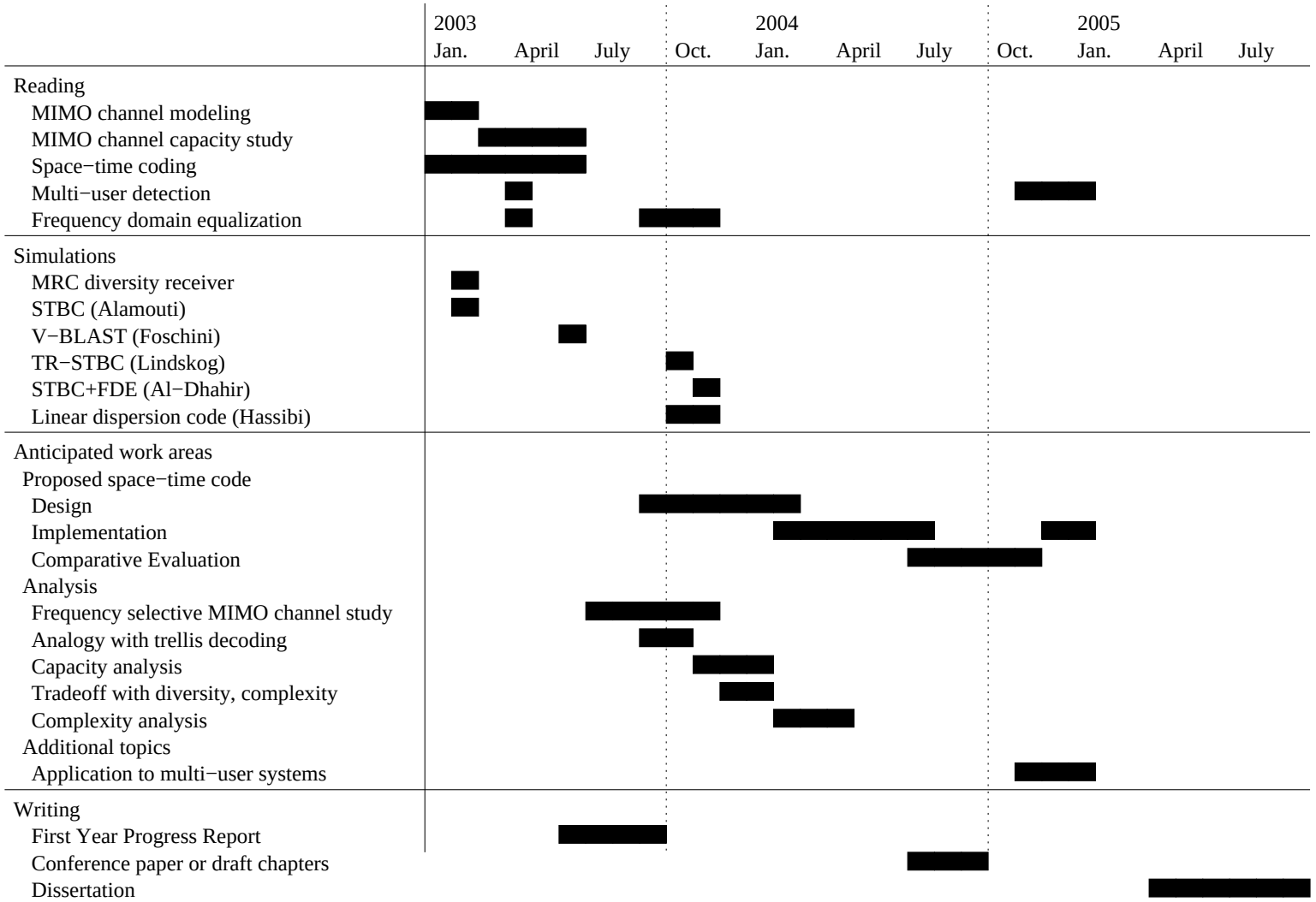


Figure 6.1: Projected milestones and completed activities, January 2003 to October 2005.

Bibliography

[1] Jr. A. Roger Hammons and Hesham El Gamal. On the theory of space-time codes for PSK modulation. *IEEE Transactions on Information Theory*, 46(2):524–542, March 2000.

[2] Naofal Al-Dhahir. Single-carrier frequency-domain equalization for space-time block-coded transmissions over frequency-selective fading channels. *IEEE Communications Letters*, 5(7):304–306, July 2001.

[3] Siavash M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16(8):1451–1458, October 1998.

[4] David Asztély. On antenna arrays in mobile communication systems: Fast fading and GSM base station receiver algorithms. Technical report, March 1996.

[5] Stephan Bäro, Gerhard Bauch, and Axel Hansmann. Improved codes for space-time trellis-coded modulation. *IEEE Communications Letters*, 4(1):20–22, January 2000.

[6] Ezio Biglieri, John G. Proakis, and Shlomo Shamai. Fading channels: Information-theoretic and communications aspects. *IEEE Transactions on Information Theory*, 44(6):2619–2692, October 1998.

[7] Helmut Bölcskei, David Gesbert, and Arogyaswami J. Paulraj. On the capacity of OFDM-based spatial multiplexing systems. *IEEE Transactions on Communications*, 50(2):225–234, February 2002.

[8] Helmut Bölcskei and Arogyaswami J. Paulraj. Performance of space-time codes in the presence of spatial fading correlation. In *Asilomar Conference on Signals, Systems, and Computers*, volume 1, pages 687–693, October 2000.

[9] Alister G. Burr. Matrix theory in MIMO systems: The MIMO channel as an eigensystem. In *Second IMA Conference on Mathematics in Communications*, December 2002.

[10] Jean-François Cardoso. Blind signal separation: Statistical principles. *Proceedings of the IEEE*, 86(10):2009–2025, October 1998.

[11] Chen-Nee Chuah, David N. C. Tse, Joseph M. Kahn, and Reinaldo A. Valenzuela. Capacity scaling in MIMO wireless systems under correlated fading. *IEEE Transactions on Information Theory*, 48(3):637–650, March 2002.

[12] CommWeb. Wireless industry statistics, 2001.

[13] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley series in telecommunications. John Wiley & Sons, 1991.

[14] Richard B. Ertel, Paulo Cardieri, Kevin W. Sowerby, Theodore S. Rappaport, and Jeffrey H. Reed. Overview of spatial channel models for antenna array communication systems. *IEEE Personal Communications Magazine*, 5(1):10–22, February 1998.

[15] Gerard J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multiple antennas. *Bell Labs Technical Journal*, 1(2):41–59, September 1996.

[16] Gerard J. Foschini, Dmitry Chizhik, Micahel J. Gans, Constantinos B. Papadias, and Reinaldo A. Valenzuela. Analysis and performance of some basic space-time architectures. *IEEE Journal on Selected Areas in Communications*, 21(3):303–320, April 2003.

[17] Gerard J. Foschini and Micahel J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6(3):311–335, March 1998.

[18] Gerard J. Foschini, Glen D. Golden, Reinaldo A. Valenzuela, and Peter W. Wolniansky. Simplified processing for high spectral efficiency wireless communication employing multi-element arrays. *IEEE Journal on Selected Areas in Communications*, 17(11):1841–1852, November 1999.

[19] Hesham El Gamal. On the robustness of space-time coding. *IEEE Transactions on Signal Processing*, 50(10):2417–2428, October 2002.

[20] Hesham El Gamal and Jr. A. Roger Hammons. A new approach to layered space-time coding and signal processing. *IEEE Transactions on Information Theory*, 47(6):2321–2334, September 2001.

[21] Hesham El Gamal and Jr. A. Roger Hammons. On the design and performance of algebraic space-time codes for BPSK and QPSK modulation. *IEEE Transactions on Communications*, 50(6):907–913, June 2002.

[22] Hesham El Gamal and Mohamed Oussama Damen. Universal space-time coding. *IEEE Transactions on Information Theory*, 49(5):1097–1119, May 2003.

[23] Girish Ganesan and Petre Stoica. Space-time diversity using orthogonal and amicable orthogonal designs. *Wireless Personal Communications*, 18(2):165–178, August 2001.

[24] David Gesbert, Mansoor Shafi, Da-Shan Shiu, Peter J. Smith, and Ayman Naguib. From theory to practice: An overview of MIMO space-time coded wireless systems. *IEEE Journal on Selected Areas in Communications*, 21(3):281–302, April 2003.

[25] George Ginis and John M. Cioffi. On the relation between V-BLAST and the GFDE. *IEEE Communications Letters*, 5(9):364–366, September 2001.

[26] Dhananjay Gore, Sumeet Sandhu, and Arogyaswami J. Paulraj. Delay diversity codes for frequency selective channels. In *IEEE International Conference on Communications*, volume 3, pages 1949–1953, April 2002.

[27] Jimm Grimm. *Transmitter diversity code design for achieving full diversity on Rayleigh fading channels*. PhD thesis, Purdue University, 1998.

[28] Jiann-Ching Guey, Michael P. Fitz, Mark R. Bell, and Wen-Yi Kuo. Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels. In *IEEE Vehicular Technology Conference*, volume 1, pages 136–140, April 1996.

[29] Babak Hassibi and Bertrand Hochwald. High-rate codes that are linear in space and time. *IEEE Transactions on Information Theory*, 48(7):1804–1824, July 2002.

[30] Walter Hirt and James L. Massey. Capacity of the discrete-time Gaussian channel with intersymbol interference. *IEEE Transactions on Information Theory*, 34(3):380–388, May 1988.

[31] Bertrand Hochwald, Thomas L. Marzetta, Thomas J. Richardson, Wim Sweldens, and Rüdiger Urbanke. Systematic design of unitary space-time constellations. *IEEE Transactions on Information Theory*, 46(6):1962–1973, September 2000.

[32] Ari Hottinen, Olav Tirkkonen, and Risto Wichman. *Multi-antenna transceiver techniques for 3G and beyond*. John Wiley & Sons, 2003.

[33] Brian L. Hughes. Differential space-time modulation. *IEEE Transactions on Information Theory*, 46(7):2567–2578, November 2000.

[34] Hamid Jafarkhani. A quasi-orthogonal space-time block code. *IEEE Communications Letters*, 49(1):1–4, January 2001.

[35] Erik G. Larsson, Petre Stoica, Erik Lindskog, and Jian Li. Space-time block coding for frequency-selective channels. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 2405–2408, May 2002.

[36] Björn Lindmark and Martin Nilsson. On the available diversity gain from different dual-polarized antennas. *IEEE Journal on Selected Areas in Communications*, 19(2):287–294, February 2001.

[37] Erik Lindskog and Arogyaswami J. Paulraj. A transmit diversity scheme for channels with intersymbol interference. In *IEEE International Conference on Communications*, volume 1, pages 307–311, June 2000.

[38] Sriram Mudulodu and Arogyaswami J. Paulraj. A transmit diversity scheme for frequency selective fading channels. In *IEEE Global Telecommunications Conference*, volume 2, pages 1089–1093, November 2000.

[39] Rohit U. Nabar, Helmut Bölcskei, Vinko Erceg, David Gesbert, and Arogyaswami J. Paulraj. Performance of multiantenna signaling techniques in the presence of polarization diversity. *IEEE Transactions on Signal Processing*, 50(10):2553–2562, October 2002.

[40] Kaveh Pahlavan and Allen H. Levesque. *Wireless information networks*. Wiley Series in telecommunications and signal processing. John Wiley & Sons, 1995.

[41] Arogyaswami J. Paulraj and Constantinos B. Papadias. Space-time processing for wireless communications. *IEEE Transactions on Communications*, 14(5):49–83, November 1997.

[42] John G. Proakis. *Digital communications*. McGraw-Hill series in communications and signal processing. McGraw-Hill, 1995.

[43] Gregory G. Raleigh and John M. Cioffi. Spatio-temporal coding for wireless communication. *IEEE Transactions on Communications*, 46(3):357–366, March 1998.

[44] Theodore S. Rappaport, A. Annamalai, R. M. Buehrer, and William H. Tranter. Wireless communications: Past events and a future perspective. *IEEE Communications Magazine*, 40(5):148–161, May 2002.

[45] Jr. Robert W. Heath and Arogyaswami J. Paulraj. Capacity maximizing linear space-time codes. *IEICE Transactions on Electronics*, E85-C(3):428–435, March 2002.

[46] Sumeet Sandhu and Arogyaswami J. Paulraj. Space-time block codes: A capacity perspective. *IEEE Communications Letters*, 4(12):384–386, December 2000.

[47] Hikmet Sari, Georges Karam, and Isabelle Jeanclaude. Frequency-domain equalization of mobile radio and terrestrial broadcast channels. In *IEEE Global Telecommunications Conference*, volume 1, pages 1–5, November 1994.

[48] Anna Scaglione. Statistical analysis of the capacity of MIMO frequency selective Rayleigh fading channels with arbitrary number of inputs and outputs. In *International Symposium on Information Theory*, page 278, June 2002.

[49] Laurent Schumacher, Jean-Philippe Kermoal, Frank Frederiksen, Klaus I. Pedersen, Albert Algans, and Preben E. Mogensen. MIMO channel characterisation. Project ist-1999-11729 metra deliverable d2, Information Society Technologies, February 2001.

[50] Nambi Seshadri and Jack H. Winters. Two signalling schemes for improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitter antenna diversity. In *IEEE Vehicular Technology Conference*, pages 508–511, May 1993.

[51] Da-Shan Shiu, Gerard J. Foschini, Micahel J. Gans, and Joseph M. Kahn. Fading correlation and its effect on the capacity of multielement antenna systems. *IEEE Transactions on Communications*, 48(3):502–513, March 2000.

[52] Siwaruk Siwamogsatham and Michael P. Fitz. Robust space-time codes for correlated Rayleigh fading channels. *IEEE Transactions on Signal Processing*, 50(10):2408–2416, October 2002.

[53] Eva Sjöström. *Singular value computations for Toeplitz matrices*. PhD thesis, Linköping University, 1996.

[54] Raymond Steele and Lajos Hanzo, editors. *Mobile radio communications*. John Wiley & Sons, 1999.

[55] Hirofumi Suzuki. A statistical model for urban radio propagation. *IEEE Transactions on Communications*, COM-25(7):673–680, July 1977.

[56] Thomas Svantesson. Correlation and channel capacity of MIMO systems employing multimode antennas. *IEEE Transactions on Vehicular Technology*, 51(6):1304–1312, November 2002.

[57] Vahid Tarokh, Hamid Jafarkhani, and A. Robert Calderbank. Space-time block codes from orthogonal designs. *IEEE Transactions on Information Theory*, 45(5):1456–1467, July 1999.

[58] Vahid Tarokh, Hamid Jafarkhani, and A. Robert Calderbank. Space-time block coding for wireless communications: Performance results. *IEEE Journal on Selected Areas in Communications*, 17(3):451–460, March 1999.

[59] Vahid Tarokh, Ayman Naguib, Nambi Seshadri, and A. Robert Calderbank. Space-time codes for high data rate wireless communication: Performance criteria in the presence of channel estimation errors, mobility, and multiple paths. *IEEE Transactions on Information Theory*, 47(2):199–207, February 1999.

[60] Vahid Tarokh, Nambi Seshadri, and A. Robert 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.

[61] I. Emre Telatar. Capacity of multi-antenna Gaussian channels. Technical report, October 1995.

[62] Emanuele Viterbo and Joseph Boutros. A universal lattice code decoder for fading channels. *IEEE Transactions on Information Theory*, 45(5):1639–1642, July 1999.

[63] Armin Wittneben. A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation. In *IEEE International Conference on Communications*, volume 3, pages 1630–1634, May 1993.

[64] Peter W. Wolniansky, Gerard J. Foschini, Glen D. Golden, and Reinaldo A. Valenzuela. V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel. In *International Symposium on Signals, Systems, and Electronics*, pages 295–300, September 1998.

[65] Qing Yan and Rick S. Blum. Optimum space-time convolutional codes. In *IEEE Wireless Communications and Networking Conference*, volume 3, pages 1351–1355, September 2000.

[66] Lizhong Zheng and David N. C. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels. *IEEE Transactions on Information Theory*, 49(5):1073–1096, May 2003.

[67] Shengli Zhou and Georgios B. Giannakis. Space-time coding with maximum diversity gains over frequency-selective fading channels. *IEEE SIgnal Processing Letters*, 8(10):269–272, October 2001.

Appendix A

ML decoding in fading channels with perfect receiver CSI

In this appendix we review some derivations and concepts relating to ML decoding in fading channels with complex AWGN, where the receiver has perfect channel knowledge. For simplicity, we consider single symbol period transmissions over *memoryless* channels, e.g., flat quasi-static fading.

Let the complex baseband transmitted symbol s be drawn from alphabet $\mathcal{X}$ of cardinality B. We assume that the channel coefficients are known at the receiver, and that the complex AWGN is of zero mean and variance N_0. In the derivations that follow, D-dimensional complex-valued random variables are represented as $2D$-dimensional real-valued random variables annotated by an underline, i.e., given complex x or $\mathbf{x}$,

$$\underline{\mathbf{x}} \;=\; \begin{bmatrix} x_R \\ x_I \end{bmatrix} \text{ or } \begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{bmatrix},$$

where the subscripts $_R$ and $_I$ are used to denote the real and imaginary parts of complex-valued variables. Note that the channel matrix must be handled in a slightly different manner. For a memoryless channel, each received signal can be considered in isolation from those seen previously. In this case the general MIMO received signal can be expressed mathematically as

$$\begin{aligned} \mathbf{r} &= \mathbf{H}\mathbf{s} + \mathbf{n} \\ \text{or equivalently} \quad \underline{\mathbf{r}} &= \underbrace{\begin{bmatrix} \mathbf{H}_R & -\mathbf{H}_I \\ \mathbf{H}_I & \mathbf{H}_R \end{bmatrix}}_{\triangleq \underline{\mathbf{H}}} \underline{\mathbf{s}} + \underline{\mathbf{n}}. \end{aligned}$$

Figure A.1 provides a geometric view of the fading process for a single transmitted QPSK signal. Observe that in a fading channel, the effective signal constellation is a rotated and scaled version of the one from which the transmitted symbol was actually selected. The complex noise term adds a random circularly symmetric component to intermediate point hs after the fading transformation. Thus as we saw in Chapter 3, the SNR without fading is $\rho = \frac{P}{N_0}$, compared to $\frac{P}{N_0}\alpha = \rho\alpha$ when the fading is taken into consideration.

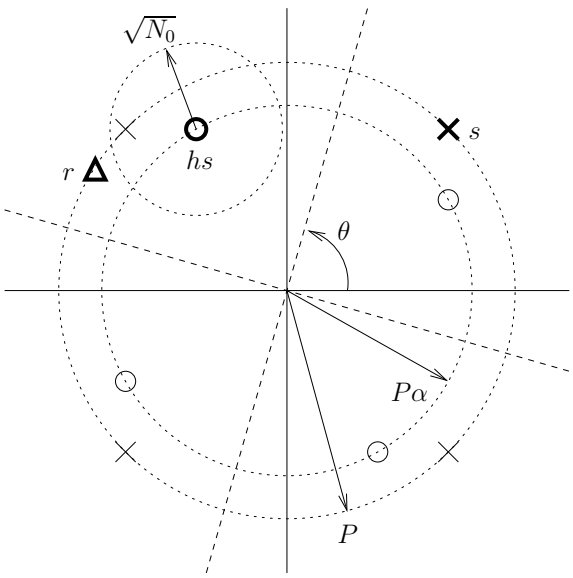


Figure A.1: Illustration of received (bold triangle), intermediate (bold circle) signal points, and rotated and scaled fading constellation (circles) when a signal (bold cross) drawn from a QPSK constellation (crosses) is transmitted over a complex AWGN fading channel.

The likelihood function of the received signal given perfect CSI and that $\mathbf{s}$ was sent is given by the conditional probability density

$$\begin{aligned} f(\underline{\mathbf{r}}\,|\mathbf{s} \text{ was sent}, \mathbf{H}) &= \frac{1}{\sqrt{2\pi}^M \sqrt{\det \boldsymbol{\Sigma}}}\, e^{-\frac{1}{2}(\underline{\mathbf{r}}-\underline{\mathbf{H}}\underline{s})^\dagger \boldsymbol{\Sigma}^{-1}(\underline{\mathbf{r}}-\underline{\mathbf{H}}\underline{s})}, \\ \text{where} \quad \boldsymbol{\Sigma} &= \frac{N_0}{2}\mathbf{I}_{2M}, \\ \text{and so} \quad f(\underline{\mathbf{r}}\,|\mathbf{s} \text{ was sent}, \mathbf{H}) &= \frac{1}{(\pi N_0)^M}\, e^{-\frac{1}{N_0}|\underline{\mathbf{r}}-\underline{\mathbf{H}}\underline{\mathbf{s}}|^2}. \end{aligned} \tag{A.1}$$

Since $\log(x)$ is concave, the ML estimate of the transmitted signal can be found by choosing the one that maximizes the log-likelihood function. Constants terms and factors are typically removed to simplify the presentation. Because the exponent in (A.1) is negative, the likelihood and log-likelihood are maximized when its magnitude is minimized:

$$\widehat{\mathbf{s}}_{ML} = \underset{\mathbf{s}\in\mathcal{X}^N}{\operatorname{argmin}} |\mathbf{r} - \mathbf{H}\mathbf{s}|^2 . \tag{A.2}$$

For the SISO and SIMO cases, where there is a single transmitted signal, the ML solution corresponds to the discrete matched filter and MRC receivers, respectively. Since the channel matrix is either square (a scalar) or tall (a vector), with statistically independent entries, $\mathbf{h}^\dagger\mathbf{h}$ is non-singular with probability 1 and the Moore-Penrose pseudoinverse is well-defined. It is then straightforward to show that the linear receiver $\mathbf{w} = \mathbf{h}^\dagger$ is optimal:

$$\begin{aligned} |\mathbf{r} - \mathbf{h}s|^2 &= \frac{\left|\mathbf{h}^\dagger\mathbf{r} - |\mathbf{h}|^2 s\right|^2}{|\mathbf{h}|^2}, \\ \text{and so} \quad \widehat{s}_{ML} &= \underset{s\in\mathcal{X}}{\operatorname{argmin}} \left|\mathbf{h}^\dagger\mathbf{r} - |\mathbf{h}|^2 s\right|^2 . \end{aligned} \tag{A.3}$$

Observe that $|\mathbf{h}|^2$ is a scalar quantity. Thus this coherent receiver has the effect of inverting the rotation operation caused by the fading channel. However, when transmit diversity is employed, this sort of channel inversion is usually no longer possible. For instance in the case of a system using transmit diversity only, we have the following minimization problem over the N-dimensional space $\mathcal{X}^N$:

$$\widehat{\mathbf{s}}_{ML} = \underset{\mathbf{s}\in\mathcal{X}^N}{\operatorname{argmin}} \left| r - \overline{\mathbf{h}}^T \mathbf{s} \right|^2 .$$

Because the rows and columns of $\overline{\mathbf{h}}^* \overline{\mathbf{h}}^T$ are multiples of each other, it is a singular matrix and the Moore-Penrose pseudoinverse of $\overline{\mathbf{h}}^T$ is not well-defined. The structure of the optimal receiver may not be linear, and solutions may not be unique. MIMO transmission systems where $M < N$, i.e., the channel matrix is wide, suffer from the same difficulty.[1]

For a MIMO channel where $M \geq N$, solving (A.2) can be shown to be equivalent to minimizing

$$\begin{aligned}
\widehat{\mathbf{s}}_{ML} &= \underset{\mathbf{s}\in\mathcal{X}^N}{\operatorname{argmin}} (\mathbf{r} - \mathbf{H}\mathbf{s})^\dagger (\mathbf{r} - \mathbf{H}\mathbf{s}) \\
&= \underset{\mathbf{s}\in\mathcal{X}^N}{\operatorname{argmin}} \operatorname{Re}\left[\mathbf{s}^\dagger (\mathbf{H}^\dagger \mathbf{H}\mathbf{s} - 2\mathbf{H}^\dagger \mathbf{r}) \right] \qquad \text{(A.4)} \\
&= \underset{\mathbf{s}\in\mathcal{X}^N}{\operatorname{argmin}} (\mathbf{H}^\dagger \mathbf{r} - \mathbf{H}^\dagger \mathbf{H}\mathbf{s})^\dagger (\mathbf{H}^\dagger \mathbf{H})^{-1} (\mathbf{H}^\dagger \mathbf{r} - \mathbf{H}^\dagger \mathbf{H}\mathbf{s}) \qquad \text{(A.5)} \\
&\triangleq \underset{\mathbf{s}\in\mathcal{X}^N}{\operatorname{argmin}} \left| \mathbf{H}^\dagger \mathbf{r} - \mathbf{H}^\dagger \mathbf{H}\mathbf{s} \right|^2_{(\mathbf{H}^\dagger \mathbf{H})^{-1}} .
\end{aligned}$$

It is well known that the complexity of the ML problem is generally exponential in the dimension of the search space ($2N$ in this case) [32]. However the *sphere decoder* is a recent advance in lattice coding theory that enables a solution to be found in polynomial time ($\mathcal{O}(n^3)$ to $\mathcal{O}(n^6)$, where n is the dimension of the search space) [62, 32]. Further complexity improvements can only be realized by exploiting special properties of the channel matrix, or by inducing a channel matrix with desirable characteristics.

For instance in the case of the Alamouti STBC, where $\mathbf{H}^\dagger = \det(\mathbf{H})\, \mathbf{H}^{-1}$, (A.5) can be reduced as follows:

$$\begin{aligned}
\left| \mathbf{H}^\dagger \mathbf{r} - \mathbf{H}^\dagger \mathbf{H}\mathbf{s} \right|^2_{(\mathbf{H}^\dagger \mathbf{H})^{-1}} &= \frac{\left| \mathbf{H}^\dagger \mathbf{r} - \det(\mathbf{H})\, \mathbf{s} \right|^2}{\det \mathbf{H}}, \\
\text{and so} \quad \widehat{\mathbf{s}}_{ML} &= \underset{\mathbf{s}\in\mathcal{X}^N}{\operatorname{argmin}} \left| \mathbf{H}^\dagger \mathbf{r} - \det(\mathbf{H})\, \mathbf{s} \right|^2 . \qquad \text{(A.6)}
\end{aligned}$$

Note the similarity between (A.3) and (A.6): $\det \mathbf{H}$ is also a scalar quantity, hence the ML solution for the Alamouti code can be found using N scalar minimizations, rather than the more computationally costly vector optimization in (A.5).

[1] In this case, although the pseudoinverse may exist, it is still not possible to recover more than M signals, assuming that the transmitted symbols are independent. To see why this is the case, consider that geometrically a wide matrix corresponds to a projection onto a lower dimensional subspace. Because some information about the higher dimensions is lost, such a transformation can only be partially inverted at best.

Appendix B

Average bit error rates for MRC receive diversity with M antennas

In this appendix we derive exact expressions for the average bit error rates for receive diversity when using MRC with two and four receive antennas. We then generalize these expressions in the limit as $M \to \infty$. Recall from Chapter 3 that the instantaneous bit error rate conditioned on the channel coefficients $h_1(t), \ldots, h_M(t)$ is

$$p_{\varepsilon|h_i[l],j=1,\ldots,M}(\rho) = \mathrm{Q}\left(\sqrt{2\rho\chi^2_{2M}}\right),$$

where χ^2_{2M} is a chi-squared random variable with $2M$ degrees of freedom. The general expressions for the probability density and distribution functions of a chi-squared random variable with r degrees of freedom, derived from Gaussian random variables with zero mean and variance $\frac{1}{2}$, are given by

$$f_r(x) = \frac{x^{\frac{r}{2}-1}e^{-x}}{\Gamma(\frac{r}{2})}$$

$$F_r(x) = \frac{\gamma(\frac{r}{2},x)}{\Gamma(\frac{r}{2})},$$

where the gamma function is defined over the integers as $\Gamma(x) = (x-1)!$ and the incomplete gamma function as $\gamma(\alpha,x) = \int_0^x t^{\alpha-1}e^{-t}\,dt$. In particular we will be working with the chi-squared distributions with four and eight degrees of freedom:

$$\begin{aligned} f_4(x) &= xe^{-x} \\ F_4(x) &= 1 - e^{-x} - xe^{-x} \\ f_8(x) &= \frac{x^3e^{-x}}{6} \\ F_8(x) &= 1 - e^{-x} - xe^{-x} - \frac{x^2}{2}e^{-x} - \frac{x^3}{6}e^{-x} \end{aligned}$$

Some other helpful identities that we will make use of are

$$\begin{aligned} \int_0^\infty u^{2n}e^{-u^2}\,du &= \frac{\sqrt{\pi}}{2}\frac{1}{2}\frac{3}{2}\cdots\frac{2n-1}{2} \\ &= \sqrt{\pi}\frac{(2n-1)!!}{2^{n+1}}, \end{aligned}$$

where the double factorial is defined over the odd integers as $n!! = n \cdot (n-2) \cdot (n-4) \cdots 5 \cdot 3 \cdot 1$, and the geometric series sum:

$$\sum_{k=0}^{N} r^k = \frac{1 - r^{N+1}}{1 - r},\ |r| < 1.$$

The average bit error probability for the case with M=2 receive antennas is then

$$\begin{aligned}
p_{\varepsilon,M=2}(\rho) &= \int_0^\infty p_{\varepsilon|h_i[l],j=1,M}(\rho) \cdot f_4(x)\, dx \\
&= \int_0^\infty \mathrm{Q}\left(\sqrt{2\rho x}\right) \cdot f_4(x)\, dx \\
&= \mathrm{Q}\left(\sqrt{2\rho x}\right) \cdot F_4(x)\Big|_0^\infty + \int_0^\infty \sqrt{\frac{\rho}{4\pi x}} e^{-\rho x} \cdot F_4(x)\, dx \\
&\overset{x=z^2}{=} \sqrt{\frac{\rho}{\pi}} \int_0^\infty e^{-\rho z^2} - e^{-z^2(1+\rho)} - z^2 e^{-z^2(1+\rho)}\, dz \\
&= \sqrt{\frac{\rho}{\pi}} \left(\frac{1}{\sqrt{\rho}} \frac{\sqrt{\pi}}{2} - \frac{1}{\sqrt{(1+\rho)}} \frac{\sqrt{\pi}}{2} - \frac{1}{\sqrt{(1+\rho)^3}} \frac{\sqrt{\pi}}{4} \right) \\
&= \frac{1}{2} - \frac{1}{2}\sqrt{\frac{\rho}{1+\rho}} - \frac{1}{4}\sqrt{\frac{\rho}{(1+\rho)^3}} \\
&= p_{\varepsilon,M=1}(\rho) - \frac{1}{4}\sqrt{\frac{\rho}{(1+\rho)^3}}.
\end{aligned}$$

Using a similar approach, we can also derive an expression for the bit error probability for the case with $M = 4$ receive antennas:

$$\begin{aligned}
p_{\varepsilon,M=4}(\rho) &= \int_0^\infty p_{\varepsilon|h_i[l],j=1,\ldots,M}(\rho) \cdot f_8(x)\, dx \\
&= \int_0^\infty \sqrt{\frac{\rho}{4\pi x}} e^{-\rho x} \cdot F_8(x)\, dx \\
&\overset{x=z^2}{=} \sqrt{\frac{\rho}{\pi}} \int_0^\infty e^{-\rho z^2} - e^{-z^2(1+\rho)} - z^2 e^{-z^2(1+\rho)} - \frac{z^4}{2} e^{-z^2(1+\rho)} - \frac{z^6}{6} e^{-z^2(1+\rho)}\, dz \\
&= p_{\varepsilon,M=2}(\rho) - \sqrt{\frac{\rho}{\pi}} \left(\frac{1}{2\sqrt{(1+\rho)^5}} \frac{3\sqrt{\pi}}{8} - \frac{1}{6\sqrt{(1+\rho)^7}} \frac{15\sqrt{\pi}}{16} \right) \\
&= p_{\varepsilon,M=2}(\rho) - \frac{3}{16}\sqrt{\frac{\rho}{(1+\rho)^5}} - \frac{15}{96}\sqrt{\frac{\rho}{(1+\rho)^7}}.
\end{aligned}$$

Based on these derivations, a general expression for the bit error probability is given by

$$\begin{aligned} p_{\varepsilon,M}(\rho) &= \frac{1}{2}\left[1-\sum_{i=1}^{M}\sqrt{\frac{\rho}{(1+\rho)^{2i-1}}}\frac{(2i-3)!!}{(2i-2)!!}\right] \\ &= \frac{1}{2}\left[1-\mu\sum_{i=1}^{M}(1-\mu^2)^{i-1}\frac{(2i-3)!!}{(2i-2)!!}\right], \end{aligned}$$

where $\mu = \sqrt{\frac{\rho}{(1+\rho)}}$. Since $\frac{(2i-3)!!}{(2i-2)!!} \approx \frac{1}{2^{i-1}}$, and in the high SNR regime we can make the approximations $\frac{1}{2}(1+\mu) \approx 1$ and $\frac{1}{2}(1-\mu) \approx \frac{1}{4\rho}$ [42], the following approximate average error probability expression can also be derived:

$$\begin{aligned} p_{\varepsilon,M}(\rho \gg 1) &\approx \frac{1}{2}\left[1-\mu\sum_{i=1}^{M}\frac{1}{\rho^{i-1}}\frac{(2i-3)!!}{(2i-2)!!}\right] \\ &\leq \frac{1}{2}\left[1-\mu\sum_{i=1}^{M}\frac{1}{(2\rho)^{i-1}}\right] \\ &= \frac{1}{2}\left[1-\mu\frac{1-\frac{1}{(2\rho)^M}}{1-\frac{1}{2\rho}}\right] \\ &\approx \frac{1}{2}\left[\frac{\mu-\mu-\frac{\mu}{(2\rho)^M}}{\mu}\right] \\ &= \frac{1}{2(2\rho)^M} \end{aligned}$$

Naturally, next we would like to consider what happens to the average bit error probability in the limit as $M \to \infty$. We begin by taking a look at what happens to $\Upsilon = \frac{\chi^2_{2M}}{M}$ as M becomes large:

$$\begin{aligned} \mathsf{E}\Upsilon &= \frac{1}{M}\cdot\mathsf{E}\chi^2_{2M} \\ &= \frac{\int_0^\infty x\cdot x^{M-1}e^{-x}\,dx}{M\cdot\Gamma(M)} \\ &= \frac{\Gamma(M+1)}{M!} \\ &= 1 \\ \mathsf{E}\Upsilon^2 &= \frac{1}{M^2}\cdot\mathsf{E}\left([\chi^2_{2M}]^2\right) \\ &= \frac{\int_0^\infty x^2\cdot x^{M-1}e^{-x}\,dx}{M^2\cdot\Gamma(M)} \\ &= \frac{\Gamma(M+2)}{M\cdot M!} \\ &= \frac{M+1}{M} \end{aligned}$$

Thus for all M, the expected value of Υ is 1, and as $M \to \infty$ its variance tends to 0. Thus as M becomes large, the density function of Υ approaches a Dirac delta function centered around 1:

$$\lim_{M\to\infty} f_\Upsilon(y) \quad = \quad \delta(y-1).$$

The average bit error probability of the SIMO system as $M \to \infty$ can then be determined:

$$\begin{aligned}
p_{\varepsilon,M\to\infty}(\rho) \quad &= \quad \lim_{M\to\infty} \int_0^\infty p_{\varepsilon|h_i[l],j=1,\ldots,M}(\rho) \cdot f_{2M}(x)\, dx \\
&= \quad \lim_{M\to\infty} \int_0^\infty \mathrm{Q}\left(\sqrt{2\rho x}\right) \cdot f_{2M}(x)\, dx \\
&\overset{x=My}{=} \quad \lim_{M\to\infty} \int_0^\infty \mathrm{Q}\left(\sqrt{2\overline{\rho} y}\right) \cdot f_\Upsilon(y)\, dy \\
&= \quad \int_0^\infty \mathrm{Q}\left(\sqrt{2\overline{\rho} y}\right) \cdot \delta(y-1)\, dy \\
&= \quad \mathrm{Q}\left(\sqrt{2\overline{\rho}}\right),
\end{aligned}$$

which is the same error performance as that obtained over an AWGN channel with the same noise power and no fading. (Recall that $\overline{\rho} = M\rho$ is the average SNR per bit.) ■

Appendix C

On decoding rate $\frac{1}{n}$ convolutional codes

The well-known VA is widely applied to solve the problem of decoding convolutional codes. It is essentially a ML sequential processing detection technique and thus can also be used for MLSE equalization. With respect to the latter problem, we have encountered cases in the space-time coding literature where the structure of the channel matrix may be exploited to decouple the equalization procedure, both in space and in time [2]. In this appendix we explore how these strategies may also be applied to simplify the decoding of rate $\frac{1}{n}$ convolutional codes.

We believe that this approach inspired by space-time coding may be more computationally efficient than the VA under certain conditions. Like the VA, it exhibits a small rate loss due to border effects. For the case of the VA, this is caused by appending a suffix of zeros to reset the encoder state between codewords. In the algorithm that we are considering, a cyclic prefix is prepended to enable block processing in a frequency-like dual code domain.

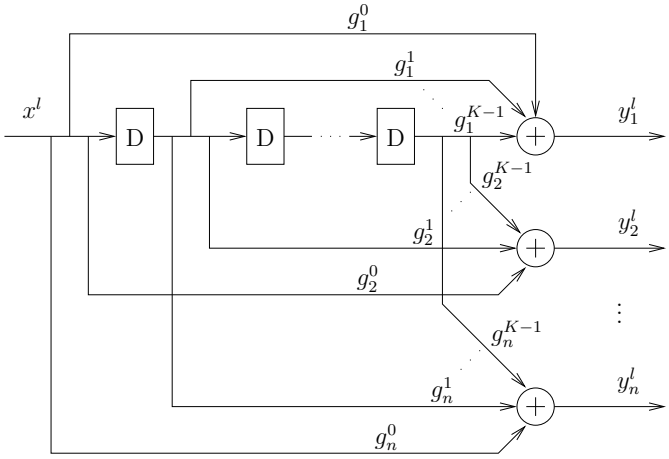


Figure C.1: A rate $\frac{1}{n}$ convolutional encoder.

We begin by considering the structure of a $\frac{1}{n}$ convolutional code as shown in Figure C.1. Let us define the constraint length of the code as K, where $K-1$ is the minimum number of zero input symbols required to reset the state of the encoder, i.e., the number of delays in the longest delay line or equivalently, the memory of the encoder. At each time step l,

the encoder accepts an input symbol x_l from alphabet $\mathcal{A}$ and produces n-symbol codewords $(y_1 y_2 \cdots y_n)^l$. We are interested in decoding blocks of codewords of length L.

The codeword-wise encoding operation can be expressed as a linear filter:

$$y_j^l \;=\; \sum_{k=0}^{K-1} g_j^k x^{l-k}, \quad l = 1, \ldots, L.$$

The output of this filter at time step l depends on the current input symbol and the $K-1$ previous inputs. To ensure that all required inputs are well-defined, we let $x^0, \ldots, x^{L-1}$ contain the data symbols and set $x^{-k} = x^{L-k}$ for $K = 1, \ldots, K-1$. This strategy is known as prepending a cyclic prefix. The decoder ignores the first $n(K-1)$ received symbols and the block codeword encoding operation can then be written in the form of an nL-dimensional linear matrix equation:

$$\underbrace{\begin{bmatrix} \mathbf{y}^0 \\ \vdots \\ \mathbf{y}^{L-1} \end{bmatrix}}_{\triangleq \mathcal{Y}} = \underbrace{\begin{bmatrix} \mathbf{g}^0 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{g}^{K-1} & \cdots & \mathbf{g}^1 \\ \mathbf{g}^1 & \mathbf{g}^0 & \mathbf{0} & \cdots & \mathbf{0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{g}^{K-1} \\ \mathbf{g}^{K-1} & \cdots & \mathbf{g}^1 & \mathbf{g}^0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \cdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \cdots & \ddots & \mathbf{g}^0 & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{g}^{K-1} & \cdots & \mathbf{g}^1 & \mathbf{g}^0 \end{bmatrix}}_{\triangleq \mathcal{G}} \underbrace{\begin{bmatrix} x^0 \\ \vdots \\ x^{L-1} \end{bmatrix}}_{\triangleq \mathcal{X}}, \tag{C.1}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are $L \times 1$ and $nL \times 1$ vectors formed by stacking the input and encoded bits, respectively, first by codeword and then by time index, i.e., ,

$$\mathbf{Y} \;=\; \begin{bmatrix} \mathbf{y}^0 & \cdots & \mathbf{y}^{L-1} \end{bmatrix} \;=\; \begin{bmatrix} y_1^0 & \cdots & y_1^{L-1} \\ \vdots & \ddots & \vdots \\ y_n^0 & \cdots & y_n^{L-1} \end{bmatrix}, \;\; \mathcal{Y} \;=\; \mathrm{Vec}\,(\mathbf{Y})$$

$$\mathbf{X} \;=\; \begin{bmatrix} x_1^0 & \cdots & x_1^{L-1} \end{bmatrix}, \;\; \mathcal{X} \;=\; \mathrm{Vec}\,(\mathbf{X}),$$

and $\mathcal{G}$ is an $nL \times L$ block circulant matrix with

$$\begin{bmatrix} \mathbf{g}^0 & \cdots & \mathbf{g}^{K-1} \end{bmatrix} \;=\; \begin{bmatrix} g_1^0 & \cdots & g_1^{K-1} \\ \vdots & \ddots & \vdots \\ g_n^0 & \cdots & g_n^{K-1} \end{bmatrix}.$$

Recall that $\mathcal{G}$ is then block diagonalized by the unitary inverse DFT: $\mathcal{G} = \mathbf{U}_{DFT,nL}^{\dagger} \mathbb{G} \mathbf{U}_{DFT,L}$, where element ij of $\mathbf{U}_{DFT,L} = \frac{1}{\sqrt{L}} e^{-j\frac{2\pi}{L}(i-1)(j-1)}$ and $\mathbf{U}_{DFT,nL}$ is specified by the Kronecker product $\mathbf{U}_{DFT,L} \otimes \mathbf{I}_n$.

If $\mathcal{Y}$ is transmitted over an AWGN channel, the signal seen at the receiver is given by $\mathcal{Z} = \mathcal{G}\mathcal{X} + \mathcal{N} \in \mathbb{C}^{nL}$, where the elements of $\mathcal{N}$ are zero mean noise samples with variance N_0. The ML detection problem can then be expressed as

$$\begin{aligned} \widehat{\mathcal{X}} \;&=\; \underset{\mathcal{X} \in \mathcal{A}^L}{\mathrm{argmin}} \left| \mathcal{Z} - \mathcal{G}\mathcal{X} \right|^2 \\ &=\; \underset{\mathcal{X} \in \mathcal{A}^L}{\mathrm{argmin}} \left| \mathcal{Z} - \mathbf{U}_{DFT,nL}^{\dagger} \mathbb{G} \mathbf{U}_{DFT,L} \mathcal{X} \right|^2 \end{aligned}$$

Next let us define the dual code alphabet

$$\mathcal{A}_{\mathbf{U}}^L = \left\{ \mathbb{X} \mid \mathbb{X} = \mathbf{U}\mathcal{X}, \mathcal{X} \in \mathcal{A}^L \right\},$$

where $\mathbf{U}$ is an $L \times L$ unitary matrix. Because $\mathbf{U}$ is unitary, $\left|\mathcal{A}_{\mathbf{U}}^L\right| = \left|\mathcal{A}^L\right|$. Also note that metrics are preserved by unitary transformations, i.e., given $c, d \in \mathbb{C}^N$, $\mu(c,d) = \mu(\mathbf{U}_N c, \mathbf{U}_N d)$ for all metrics μ. In particular, Euclidean distance is a metric and therefore defining $\mathbb{Z} = \mathbf{U}_{DFT,nL}\mathcal{Z}$ gives the following equivalent minimization problem:

$$\begin{aligned} \widehat{\mathcal{X}} &= \mathbf{U}_{DFT,L}^{\dagger}\widehat{\mathbb{X}} \\ \widehat{\mathbb{X}} &= \underset{\mathbb{X}\in\mathcal{A}_{\mathbf{U}}^L}{\operatorname{argmin}} \left|\mathbb{Z} - \mathbb{G}\mathbb{X}\right|^2. \end{aligned} \tag{C.2}$$

To address this problem in the frequency-like dual code domain, we exploit the structures of $nL \times L$ block diagonal matrix $\mathbb{G}$ and $L \times L$ diagonal matrix $\mathbb{G}^{\dagger}\mathbb{G}$:

$$\mathbb{G} = \begin{bmatrix} \mathrm{g}^0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathrm{g}^1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathrm{g}^{L-1} \end{bmatrix}, \quad \mathbb{G}^{\dagger}\mathbb{G} = \begin{bmatrix} |\mathrm{g}^0|^2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & |\mathrm{g}^1|^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & |\mathrm{g}^{L-1}|^2 \end{bmatrix},$$

where $\mathrm{g}^l = [\mathrm{g}_1^l \ \cdots \ \mathrm{g}_n^l]^T$ and $\mathrm{g}_i^l = \sum_{k=0}^{K-1} g_i^k e^{-j\frac{2\pi}{L}kl}$. The nL-dimensional minimization in (C.2) can then be decoupled into L n-dimensional problems:

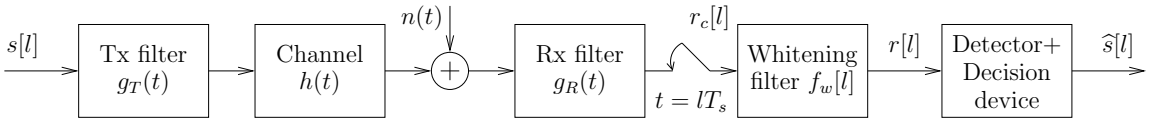
$$\begin{aligned} \widehat{\mathbb{X}} &= \underset{\mathbb{X}\in\mathcal{A}_{\mathbf{U}}^L}{\operatorname{argmin}} \left(\mathbb{Z}^{\dagger} - \mathbb{X}^{\dagger}\mathbb{G}^{\dagger}\right)\left(\mathbb{Z} - \mathbb{G}\mathbb{X}\right) \\ &= \underset{\mathbb{X}\in\mathcal{A}_{\mathbf{U}}^L}{\operatorname{argmin}} \left(\mathbb{Z}^{\dagger}\mathbb{Z} - \mathbb{Z}^{\dagger}\mathbb{G}\mathbb{X} - \mathbb{X}^{\dagger}\mathbb{G}^{\dagger}\mathbb{Z} + \mathbb{X}^{\dagger}\mathbb{G}^{\dagger}\mathbb{G}\mathbb{X}\right) \\ &= \underset{\mathbb{X}\in\mathcal{A}_{\mathbf{U}}^L}{\operatorname{argmin}} \sum_{l=0}^{L-1} \left[(\mathbb{z}^l)^{\dagger}\mathbb{z}^l - (\mathbb{z}^l)^{\dagger}\mathrm{g}^l\mathbb{x}^l - (\mathbb{x}^l)^*(\mathrm{g}^l)^{\dagger}\mathbb{z}^l + (\mathbb{x}^l)^*(\mathrm{g}^l)^{\dagger}\mathrm{g}^l\mathbb{x}^l\right] \\ &= \underset{\mathbb{X}\in\mathcal{A}_{\mathbf{U}}^L}{\operatorname{argmin}} \sum_{l=0}^{L-1} \left|\mathbb{z}^l - \mathrm{g}^l\mathbb{x}^l\right|^2 \end{aligned}$$

(Aside: At this point in the development, we pause since there is a hopefully small but important flaw in the reasoning. Equation C.1 is not quite correct unless we specify that $x^l, y_i^l, g_i^l \in \mathrm{GF}(2)$ and redefine addition and multiplication appropriately within the finite field. However this may invalidate the block diagonalization. We are looking at ways to address this, more specifically decompositions within the finite field (e.g., using a modified Hadamard transform instead of the DFT) and also, since it is clear that $\mathcal{Z} \in \mathbb{C}^{nL}$, understanding the transition from $\mathrm{GF}(2^{nL})$ to $\mathbb{C}^{nL}$. We are still exploring what implications these issues have for decoding convolutional codes in a frequency-like dual domain, but believe that there is potential here for complexity savings.)

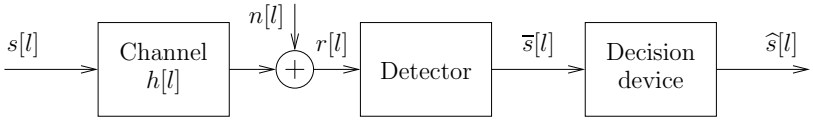
Appendix D

On the design of transmit and receive filters

In this appendix we consider the effect of the transmit and receive filters on the continuous time and discrete time systems presented in Chapter 2. We begin by repeating the complex baseband system diagrams for convenient reference.



(a) Full system diagram, continuous time.



(b) Simplified system diagram, discrete time.

Complex baseband communication system diagrams.

To simplify the analysis, we will present our derivations in the frequency domain as much as possible. Let the continuous time Fourier transforms of the transmitted and received signals (before sampling) be denoted by $\mathbb{s}(f)$ and $\mathbb{r}_c(f)$ respectively, that of the transmit and receive filters by $\mathbb{g}_T(f)$ and $\mathbb{g}_R(f)$ respectively, that of the channel impulse response by $\mathbb{h}(f)$ and that of the noise by $\mathbb{n}(f)$. Then the frequency domain representation of the received signal can be written as

$$\mathbb{r}_c(f) \;=\; \mathbb{s}(f) \star \underbrace{\underbrace{\mathbb{g}_T(f)\mathbb{h}(f)}_{\triangleq\, \mathbb{c}(f)}\, \mathbb{g}_R(f)}_{\triangleq\, \mathbb{p}(f)} + \mathbb{n}(f)\mathbb{g}_R(f),$$

where as we shall see, the subscript $_c$ is used to indicate that the samples $r_c[l]$ may be statistically correlated.

Our first goal is to design the transmit and receive filters to minimize ISI. The well-known *Nyquist criterion* states that in order to ensure ISI-free transmission, the overall pulse shape $p(t)$ must be such that its impulse and frequency responses satisfy

$$\begin{aligned} p(kT_s) \;=\; p[k] \;&=\; \delta[k] \quad \text{(time domain)} \\ \sum_{n=-\infty}^{\infty} \mathbb{p}\left(f - \frac{n}{T_s}\right) \;&=\; T_s \quad \text{(frequency domain)}, \end{aligned}$$

where $\delta[k]$ is the Kronecker delta function. This goal can be achieved by choosing a pulse shape that satisfies the Nyquist criterion, e.g., the raised cosine pulse, and then defining the transmit and receive filters such that the product of their frequency responses is equal to that of the product of the pulse and the inverse of the channel:

$$\mathbb{g}_T(f)\mathbb{g}_R(f) \;=\; \frac{\mathbb{h}^*(f)\mathbb{p}(f)}{|\mathbb{h}(f)|} \tag{D.1}$$

We note that there are an infinite number of filters that satisfy (D.1). However, the Nyquist criterion only applies for transmission in the absence of noise. Our second goal is to attain good bit error performance, in other words $\mathbb{g}_T(f)$ and $\mathbb{g}_R(f)$ must also be designed to reduce the effects of channel noise in some optimal manner.

Optimal detection in an AWGN channel is typically accomplished using a matched filter. In this case the time domain representation of the output of the sampler is given by

$$r_c[l] \;=\; \sum_{k=-\infty}^{\infty} s[k] \underbrace{\int_{-\infty}^{\infty} c(\theta - kT_s) g_R(lT_s - \theta)\, d\theta}_{\triangleq\, h_c[l-k]} + \underbrace{\int_{-\infty}^{\infty} n(\theta) g_R(lT_s - \theta)\, d\theta}_{\triangleq\, n_c[l]}$$

and its SNR by

$$\rho_{r_c[l]} \;=\; \frac{P\left|\int_{-\infty}^{\infty} c(\theta - kT_s) g_R(lT_s - \theta)\, d\theta\right|^2}{N_0 \int_{-\infty}^{\infty} |g_R(lT_s - \theta)|^2\, d\theta}, \tag{D.2}$$

where as before P is the power of the transmitted signal and N_0 that of the noise. By applying Schwarz' Inequality to the numerator of (D.2) and matching the receive filter to the l^{th} transmitted symbol, we find that the received SNR is maximized by choosing

$$\begin{aligned} g_R(t) \;&=\; Kc^*(-t) \\ \text{or equivalently} \quad \mathbb{g}_R(f) \;&=\; K\mathbb{h}^*(f)\mathbb{g}_T^*(f), \end{aligned} \tag{D.3}$$

for some real scaling factor K.

D.1 Flat fading channel

If the channel is flat fading, $\mathbb{h}(f) = \mathbb{h}$ is constant over the frequency band of interest. As long as $\mathbb{h} \neq 0$ (D.1) is well defined and ISI-free transmission is possible. Combining the filter design criteria of (D.1) and (D.3) leads to the following transmit and receive filter pair:

$$\mathbb{g}_T(f) \;=\; \mathbb{q}(f),\;\; \mathbb{g}_R(f) \;=\; \frac{\mathbb{h}^*}{|\mathbb{h}|}\mathbb{q}^*(f),\;\; \text{such that } \mathbb{q}(f)\mathbb{q}^*(f) = \mathbb{p}(f). \tag{D.4}$$

Next we are interested in the implications of the transmit and receive filters for the discrete time CIR and noise processes. The continuous time noise process $n(t)$ is assumed to be circularly symmetric complex Gaussian with zero mean and variance N_0. The autocorrelation of the discrete time noise process $r_{n_c}[k]$ can be derived as follows:

$$\begin{aligned}
\mathsf{E}\left(n_c^*[l]n_c[l+k]\right) &= \mathsf{E}\left(\int_{-\infty}^{\infty} n^*(\theta)g_R^*(lT_s-\theta)\,d\theta \int_{-\infty}^{\infty} n(\theta')g_R(lT_s+kT_s-\theta')\,d\theta'\right) \\
&= \iint_{-\infty}^{\infty} \mathsf{E}\left[n^*(\theta)n(\theta')\right] g_R^*(lT_s-\theta)g_R(lT_s+kT_s-\theta')\,d\theta\,d\theta' \\
&\overset{\phi=lT_s-\theta}{=} N_0 \int_{-\infty}^{\infty} g_R^*(\phi)g_R(kT_s+\phi)\,d\phi \\
&= N_0 r_{g_R}[k], \qquad \text{(D.5)}
\end{aligned}$$

where $r_{g_R}[k]$ is the discrete time autocorrelation function of the receive filter. In order for the filtered and sampled noise process to be spectrally white, its autocorrelation must be equal to the Kronecker delta function. In other words, the power spectrum of the receive filter must correspond to the frequency response of a Nyquist pulse.

Applying the filter design in (D.4), we find that the power spectrum of the receive filter is $|\mathbb{g}_R(f)|^2 = \mathbb{p}(f)$. Assuming that the transmitted pulse satisfies the Nyquist criterion, its autocorrelation function is then $r_{g_R}[k] = p[k] = \delta[k]$. Thus the sampled noise process is spectrally white with variance N_0. In this case, the whitening filter is not needed since $n[l] = n_c[l]$ is already white. Also recall that the noise convolution integral is the limit of an infinite sum of independent Gaussian random variables. Therefore the output sample is also Gaussian and we have that the filtered and sampled noise process is complex AWGN with the same statistics as $n(t)$.

To consider the properties of the filtered and sampled channel impulse response, we assume that the fading is Rayleigh. Then we can write the complex-valued response in terms of its real and imaginary components, which are i.i.d. Gaussian random processes with zero mean and unit variance. The frequency domain representation of the filtered real (or imaginary) channel response can be expressed as

$$\mathbb{h}_R\mathbb{g}_T(f)\mathbb{g}_R(f) \;=\; \mathbb{h}_R\mathbb{p}(f)$$

Similarly to the previous argument, if the transmitted pulse satisfies the Nyquist criterion, the real and imaginary components of the sampled channel impulse response are then spectrally white, i.e., $h_R[l] = h_c^R[l]$. Further, $h_R[l]$ and $h_I[l]$ have the same unit variance as the continuous time processes $h_R(t)$ and $h_I(t)$ respectively. The convolution integral again yields the limit of an infinite sum of independent Gaussian random variables, which is also Gaussian. Therefore the filtered and sampled channel impulse response is Rayleigh, with the same statistics as its continuous time counterpart.

D.2 Frequency selective fading channel

If the channel is frequency selective, by definition its rms delay spread exceeds T_s. In the frequency domain this implies that the coherence bandwidth of the channel is smaller than the nominal bandwidth of the transmitted signal. Therefore, there are at least one and

at most $\left\lfloor \frac{\widehat{\sigma}_\tau}{T_s} \right\rfloor$ nulls in its frequency response $\mathbb{h}(f)$. Because of these spectral nulls, the denominator term in (D.1) is effectively zero at some frequencies and the filter condition for ISI-free transmission is not well defined.

In this case, the transmit and receive filters are designed such that they do not introduce any additional ISI into the system. The ISI caused by the channel is then handled using equalization techniques as discussed in previous sections of this report that address frequency selective fading channels. The validity of the matched filtering operation is not affected by the frequency selectivity of the channel. Therefore the criterion in (D.3) still applies. For convenience, we set $K = 1$ and the resulting transmit and receive filter pair is

$$\mathbb{g}_T(f) \;=\; \mathbb{q}(f),\;\; \mathbb{g}_R(f) \;=\; \mathbb{h}^*(f)\mathbb{q}^*(f),\;\; \text{such that } \mathbb{q}(f)\mathbb{q}^*(f) = \mathbb{p}(f).$$

As before, we can now study the effect of these transmit and receive filters on the statistics of the discrete time channel coefficients and noise process. The power spectral density (before sampling) and sampled autocorrelation function of the filtered discrete time noise process are

$$\begin{aligned} s_{n_c}(f) &= N_0 \left|\mathbb{h}(f)\right|^2 \mathbb{p}(f) \\ r_{n_c}[k] &= N_0 r_h[k], \end{aligned} \tag{D.6}$$

where we recall that the autocorrelation function of the channel is specified by its power delay profile, which takes non-zero values for $k = 0, \ldots, K$. For the following analysis, we define the two-sided autocorrelation of length $2K - 1$ as $r_h[k] = p_h[|k|]$.

Observe that the noise process $n_c[l]$ is not spectrally white. Since it is mathematically simpler to work with a white noise process in subsequent analysis of the discrete time system, a discrete *whitening* filter is inserted after the sampler [42]. This filter is designed by factoring the z-transform of the sampled autocorrelation function

$$\begin{aligned} \mathbb{H}(z) &= \sum_{k=-K}^{K} r_h[k] z^{-k} \\ &= \mathbb{F}(z)\mathbb{F}^*(z^{-1}) \end{aligned}$$

The desired discrete whitening filter is then an IIR filter with z-transform

$$\frac{1}{\mathbb{F}^*(z^{-1})} \;\approx\; z^{P_w} \sum_{p=0}^{P_w - 1} f_w[p] z^{-p} \quad \text{(truncated and causal)}$$

where P_w is the length of the truncated filter. After whitening the discrete power spectral density of the resulting noise is $N_0\mathbb{p}(f)$. As before, the filtered noise retains its Gaussian characteristic and with the whitening filter we have that the discrete time noise process $n[l]$ is complex AWGN with the same statistics as $n(t)$.

The properties of the filtered and sampled channel impulse response, can be studied by writing the complex-valued response in terms of its real and imaginary components. The frequency domain representation of the filtered real (or imaginary) channel response can be expressed as

$$\mathbb{h}_R(f)\mathbb{g}_T(f)\mathbb{g}_R(f) \;=\; \left|\mathbb{h}_R(f)\right|^2 \mathbb{p}(f) \tag{D.7}$$

Similarly to the previous argument, we assume that the transmitted pulse satisfies the Nyquist criterion. The real and imaginary components of the sampled channel impulse response are then not spectrally white. However, it can be seen from the mathematical structures of (D.6) and (D.7) that applying the discrete whitening filter to the sampled convolution of the signal and correlated channel coefficients $h_c[l]$ is equivalent to sampling the convolution of the signal and statistically white channel $h[l]$.

The unit variance and Gaussian characteristic of the continuous time processes $h_R(t)$ and $h_I(t)$ is also preserved. Therefore the filtered and sampled channel impulse response is Rayleigh, with the same statistics as its continuous time counterpart.

Author Index