

# A Study of Wireless Modem Performance Using Multiple Element Antennas

by

Guillaume Lebrun

Ingénieur de l'Ecole Nationale Supérieure des Télécommunications (2000)

Submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

VICTORIA UNIVERSITY

March 2004

---

Thesis Supervisor Pr. Michael Faulkner

---

Centre for Telecommunications and Micro-Electronics

Victoria University

Footscray Park Campus

P.O. Box 14428 MCMC

Melbourne 8001

AUSTRALIA



# Contents

<b>Declaration of originality</b>	<b>ix</b>
<b>Acknowledgements</b>	<b>xi</b>
<b>List of acronyms</b>	<b>xiii</b>
<b>Notations</b>	<b>xv</b>
<b>Variables</b>	<b>xix</b>
<b>Abstract</b>	<b>xxiii</b>
<b>I Introduction</b>	<b>1</b>
I.1 History of telecommunications . . . . .	1
I.1.1 Origin of telecommunications . . . . .	1
I.1.2 Evolution and revolutions of telecommunications . . . . .	2
I.1.3 Telecommunication: a multi-faceted giant . . . . .	4
I.2 Recent trends in telecommunications . . . . .	5
I.2.1 Current telecommunications . . . . .	5
I.2.2 Future of telecommunications . . . . .	6
I.2.3 Towards understanding the Shannon capacity . . . . .	6
I.2.4 Multiple antennas wireless systems . . . . .	7
I.3 Statement of significance . . . . .	8
I.4 Contribution to knowledge . . . . .	8
I.5 List of publications . . . . .	10
I.6 Outline . . . . .	12

<b>II</b>	<b>Channel capacity</b>	<b>15</b>
II.1	Notion of capacity . . . . .	15
II.1.1	The sampling theorem . . . . .	16
II.1.2	Elementary information theory . . . . .	17
II.1.3	Capacity of a channel . . . . .	18
II.1.4	Coding theorem . . . . .	19
II.1.5	Channel model . . . . .	19
II.1.6	Gaussian complex random variable . . . . .	21
II.1.7	Capacity of the flat-fading channel . . . . .	22
II.1.8	Capacity of the non-ergodic channel . . . . .	23
II.1.9	Capacity of the ergodic channel . . . . .	23
II.2	MIMO Capacity . . . . .	24
II.2.1	MIMO channel model . . . . .	25
II.2.2	Circularly symmetric complex Gaussian distribution . . . . .	26
II.2.3	Capacity of the MIMO flat-fading channel . . . . .	26
II.3	Discussion of the assumptions . . . . .	40
II.3.1	Coherent detection . . . . .	41
II.3.2	Time and frequency properties of the channel . . . . .	42
II.3.3	Independence of the noise entries . . . . .	44
II.3.4	Single user channel . . . . .	44
II.4	Conclusion . . . . .	45
<b>III</b>	<b>Using the large capacity of MIMO channel</b>	<b>47</b>
III.1	Review of MIMO transmission techniques . . . . .	48
III.1.1	Uncoded systems . . . . .	48
III.1.2	Space-time coding . . . . .	54
III.1.3	BER simulation of MIMO transmission techniques . . . . .	56
III.2	Multiplexing vs diversity . . . . .	56
III.3	Capacity of correlated channels . . . . .	58
III.3.1	Ricean channel . . . . .	59
III.3.2	Capacity bounds for the Ricean channel . . . . .	61
III.3.3	Results and discussion . . . . .	63
III.3.4	Conclusion . . . . .	66

III.4 MIMO systems in highly specular channel . . . . .	67
III.5 Influence of CSI at the transmitter . . . . .	68
III.6 Conclusion . . . . .	71
<b>IV Optimum structure: the SVD</b>	<b>73</b>
IV.1 Introduction to the SVD structure . . . . .	74
IV.2 Optimality of the SVD . . . . .	75
IV.2.1 Optimal linear precoder and decoder . . . . .	76
IV.2.2 Optimal space-time structure . . . . .	77
IV.3 Practical SVD architecture . . . . .	80
IV.3.1 Performance loss of the practical architecture . . . . .	81
IV.3.2 Notion of system capacity . . . . .	81
IV.4 Channel estimation and SVD architecture . . . . .	82
IV.4.1 Channel estimation accuracy . . . . .	83
IV.4.2 Uniqueness of the SVD . . . . .	84
IV.5 Conclusion . . . . .	88
<b>V SVD architecture in TDD environment</b>	<b>91</b>
V.1 SVD architecture on reciprocal channels . . . . .	92
V.2 Hardware calibration procedure . . . . .	93
V.2.1 Effect of hardware errors on channel reciprocity . . . . .	93
V.2.2 Calibration procedure . . . . .	96
V.2.3 Effects of the calibration . . . . .	98
V.2.4 Choice of calibration matrices . . . . .	99
V.3 SVD-TDD system and singular values swapping . . . . .	99
V.3.1 Matrix perturbation theory, singular values . . . . .	100
V.3.2 Matrix perturbation theory, singular vectors . . . . .	101
V.3.3 Probability of robust transmission . . . . .	103
V.3.4 Singular value crossing . . . . .	105
V.3.5 Subspace swapping . . . . .	108
V.3.6 Correction of singular value crossing . . . . .	109
V.4 SVD architecture modified for TDD environment . . . . .	110
V.4.1 Correction of the CSI impairment . . . . .	111
V.4.2 Comparison with reference systems . . . . .	113

V.4.3	Simulation and results . . . . .	114
V.4.4	Conclusion . . . . .	120
V.5	CSI improvement through channel tracking . . . . .	121
V.5.1	Tracking of the CSI . . . . .	122
V.5.2	Tracking of the precoding and decoding matrices . . . . .	124
V.5.3	Correlation of pilots . . . . .	127
V.5.4	Simulation results, decoding matrix . . . . .	127
V.5.5	Simulation results, precoding matrix . . . . .	132
V.5.6	Conclusion . . . . .	134
V.6	Conclusion . . . . .	135
<b>VI</b>	<b>Conclusion</b>	<b>137</b>
VI.1	Summary . . . . .	137
VI.2	Future work . . . . .	140
VI.2.1	Pilot symbols . . . . .	140
VI.2.2	Coding . . . . .	141
VI.2.3	System level design . . . . .	141
VI.2.4	Study of the wireless channel . . . . .	142
VI.2.5	Implementation . . . . .	142
<b>A</b>	<b>Ricean channel capacity bounds</b>	<b>143</b>
I.1	Study of the Eigenvalues of $\mathbf{F}$ . . . . .	143
I.1.1	Ricean channel . . . . .	143
I.1.2	Singular Value Decompositions . . . . .	143
I.1.3	Eigenvalues of $\mathbf{F}$ . . . . .	144
I.1.4	Asymptotic eigenvalues of $\mathbf{F}$ . . . . .	145
I.2	Capacity Lower bound . . . . .	146
I.3	Capacity Upper bound . . . . .	148
<b>B</b>	<b>IEE Electronics Letters</b>	<b>151</b>
<b>C</b>	<b>IEEE Global Telecommunications Conference 2002</b>	<b>153</b>
<b>D</b>	<b>IEEE Vehicular Technology Conference 2003</b>	<b>155</b>

<b>E</b>	<b>IEEE International Conference on Communications 2004</b>	<b>157</b>
<b>F</b>	<b>IEEE International Conference on Communications 2004</b>	<b>159</b>
<b>G</b>	<b>IEEE Transactions on wireless communications</b>	<b>161</b>
<b>H</b>	<b>IEEE Transactions on wireless communications</b>	<b>163</b>



# Declaration of originality

I hereby declare that the research in this thesis itself was entirely composed and compiled by myself in the Centre for Telecommunications and Micro-Electronics at Victoria University.

---

Guillaume LEBRUN



# Acknowledgements

Though I am the author of this thesis, my supervisor Pr. Michael Faulkner is, in many ways, the driving force behind it. I was very fortunate to walk in his footsteps throughout this research project. Few PhD students receive support and guidance on both the methodology and the content of their project; I received them in abundance. Pr. Michael Faulkner not only showed me how to become a researcher but also taught me invaluable lessons on how to become a respectable human being.

I am deeply grateful to Pr. Mansoor Shaafi, Pr. Peter Smith and Pr. Jugdutt Singh for their time and their advice to improve the quality of my work as well as my understanding of telecommunications systems.

The present thesis represents three years of work that I shared with colleagues in the School of Electrical Engineering. I received much help from the group and would especially like to acknowledge Melvyn Pereira, Shane Spiteri, Dr. Tan Ying and Dr. Jason Gao for their support. I owe many thanks to Mrs Noreen Wheaton for accepting to patiently guide me through the subtleties of the English language.

This thesis is not only the result of the research conducted in the past three years but more generally the result of twenty seven years of education. I was lucky to meet inspiring teachers throughout high school and university. I would especially like to thank my teachers from the Ecole Nationale Supérieure des Télécommunications and from the Lycée Saint Louis.

Last but not least, I owe my happiness to my wonderful family and my faithful friends. No achievement would be possible without their love and support.



# List of acronyms

AD	Anno Domini
AWGN	Additive White Gaussian Noise
BC	Before Christ
bit	Binary Digit
BLAST	Bell laboratories Layered Space-Time
bps	bits per seconds
CCDF	Complementary Cumulative Density Function
CDMA	Code Division Multiple Access
CSI	Channel State Information
CSMA-CD	Carrier Sense Multiple Access - Collision Detection
dB	decibels
D-Blast	Diagonal Blast
FER	Frame Error Rate
FIR	Finite Impulse Response
GHz	Gigahertz
Hz	Hertz
IBM	International Business Machines corporation
i.i.d.	Independent and Identically Distributed
ISI	Inter-Symbol-Interference
ISM	Industrial, Scientific and Medical
km	kilometers
LAN	Local Area Network
MAC	Medium Access Control

MHz	Megahertz
MIMO	Multiple Input Multiple Output
MMSE	Minimum Mean Squared Error
ms	milliseconds
$\mu$ s	microseconds
OFDM	Orthogonal Frequency Division Multiplexing
PDF	Probability Density Function
PSAM	Pilot Symbol Assisted Modulation
QAM	Quadrature Amplitude Modulation
RF	Radio Frequency
s	Seconds
SISO	Single Input Single Output
SINR	Signal to Interference and Noise Ratio
SNR	Signal to Noise Ratio
ST	Space-Time
SVD	Singular Value Decomposition
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
USA	United States of America
USSR	Union of Soviet Socialist Republics
V-Blast	Vertical Blast
ZF	Zero-Forcing

# Notations

$\log_a(\cdot)$	base $a$ logarithm
$\exp(\cdot)$	exponential function
$e$	exponential of 1, $\exp(1)$
$\Gamma(\cdot)$	gamma function
$J_0(\cdot)$	zeroth-order Bessel function of the first kind
$\delta(\cdot, \cdot)$	Kronecker function
$F(w)$	fourier transform of $f(t)$
$\mathbb{N}$	semi-ring of the positive integer numbers
$\mathbb{Z}$	ring of integer numbers
$\mathbb{Q}$	field of rational numbers
$\mathbb{R}$	field of real numbers
$\mathbb{C}$	field of the complex numbers
$\Re(\cdot)$	real part function
$\Im(\cdot)$	imaginary part function
$\mathbb{C}^i$	vector space of dimension $i$ over $\mathbb{C}$
$x$	scalar
$\mathbf{x}$	vector
$\mathbf{x}_i$	$i^{\text{th}}$ entry of $\mathbf{x}$
$\mathbf{X}$	matrix
$\mathbf{X}_{i,j}$	entry of the $i^{\text{th}}$ line, $j^{\text{th}}$ column of $\mathbf{X}$
$\mathbf{X}_{:,i}$	$i^{\text{th}}$ column of $\mathbf{X}$

$\mathbf{x}^\dagger$	transpose of $\mathbf{x}$
$\mathbf{x}^*$	transpose-conjugate of $\mathbf{x}$
$\tilde{\mathbf{x}}$	decomposition of $\mathbf{x}$ into its real and imaginary parts
$(\cdot)$	slicing (hard decision) operation
$\text{Vect}(\mathbf{x}_i)$	vector of zeros (length of $\mathbf{x}$ ) except the $i^{\text{th}}$ entry $x_i$
$\hat{\mathbf{x}}$	estimate of $\mathbf{x}$
$\bar{\mathbf{x}}$	average of all the possible $\mathbf{x}$
$\vec{\mathbf{X}}_{i,j}(lT)$	vector of time samples 0 to $lT$ of $\mathbf{X}_{i,j}$
$(\cdot)_+$	only non-negative values are acceptable
$ \cdot $	absolute value
$\ \cdot\ _2$	norm 2
$\ \cdot\ _F$	Frobenius norm
$\ \cdot\ _u$	unitary invariant norm
$SVD(\cdot)$	SVD function
$f^u(\cdot)$	output subspace function (component of the SVD)
$f^v(\cdot)$	input subspace function (component of the SVD)
$\mathcal{R}(\mathbf{H})$	rank of a matrix $\mathbf{H}$
$\mathbf{X}^+$	pseudo-inverse of $\mathbf{X}$
$\mathbf{I}_k$	Identity matrix of dimension $k$
$\mathbf{0}$	matrix of zeros
$\mathbf{1}$	matrix of ones
$\cdot^\Pi$	permutation of a matrix
$\mathcal{P}(\cdot)$	probability function
$\underline{x}, \underline{\mathbf{x}}, \underline{\mathbf{X}}$	random variable (scalar, vector and matrix)
$\mu_{\underline{x}}$	mean of the random variable $\underline{x}$
$\sigma_{\underline{x}}^2$	variance of the random variable $\underline{x}$
$\text{pdf}_{\underline{x}}$	probability density function of $\underline{x}$
$\gamma$	Gaussian probability density function
$\chi_k^2$	chi-square variate with $k$ degrees of freedom

$E_{\text{pdf}_x}[\cdot]$  denotes the expectancy with probability density function  $\text{pdf}_x$

$\mathcal{H}_a(\cdot) = -\log_a(\cdot)$  entropy function

$\mathcal{H}(\cdot) = \mathcal{H}_2(\cdot)$  entropy function (in bits)

$\mathcal{I}(\cdot)$  mutual information function



# Variables

$x, \mathbf{x}, \mathbf{X}$	transmitted signal
$\tilde{\mathbf{x}}$	signal transmitted on the transmission eigenmodes
$y, \mathbf{y}, \mathbf{Y}$	received signal
$\tilde{\mathbf{y}}$	signal received on the transmission eigenmodes
$n, \mathbf{n}, \mathbf{N}$	noise
$\tilde{\mathbf{n}}$	noise on the transmission eigenmodes
$\mathbf{N}_{est}$	estimation noise
$\mathbf{N}_{tvc}$	time-varying channel perturbation (pseudo-noise)
$\mathbf{P}$	pilot symbols
$h, \mathbf{h}, \mathbf{H}$	wireless channel
$\mathbf{H}^{A \rightarrow B}$	wireless channel from transceiver A to transceiver B
$\mathbf{H}^{B \rightarrow A}$	wireless channel from transceiver B to transceiver A
$\mathbf{H}^{sp}$	specular component of a Ricean channel
$\mathbf{H}^{sc}$	scattering component of a Ricean channel
$K$	Ricean $K$ -factor expressed in dB
$M_T$	number of transmitting antennas
$M_T^{act}$	number of active transmitting antennas
$M_R$	number of receiver antennas
$M = \min(M_R, M_T)$	minimum number of antennas at either the transmitter or the receiver
$\alpha = M_T/M_R$	ratio between the number of transmitting antennas and the number of receiving antennas
$\beta = M_R/M_T$	ratio between the number of receiving antennas and the number of transmitting antennas

$M_A$	number of antennas at transceiver A
$M_B$	number of antennas at transceiver B
$p$	probability of an event
$p_e$	probability of error
$p_e^{\min}$	minimum probability of error
$p_{out}$	outage probability
$C$	capacity
$C_{out}$	outage capacity
$C_{ergodic}$	ergodic capacity
$P$	total transmitted power
$P_{wf}$	level chosen to satisfy the waterfilling requirements
$\mathbf{Q}$	correlation of the transmitted signal
$\tilde{\mathbf{Q}}$	correlation of the signal transmitted on the transmission eigenmodes
$\text{SNR}_{\text{dB}}$	SNR expressed in dB
$\text{SNR}_{\text{est}}$	SNR in the channel estimation
$\mathbf{A}, \mathbf{B}$	decoding matrices for MIMO receivers
$\mathbf{a}$	decoding vectors for MIMO receivers
$\mathbf{w}_{opt}$	optimum linear FIR filter (Wiener filter)
$W$	bandwidth of the signal
$w$	angular frequency
$T_s$	sampling period
$\tau$	maximum delay of multipath
$F_d$	Doppler frequency
$F_{svc}$	frequency of singular value crossings
$\mathcal{X}$	MIMO constellation
$\mathcal{S}$	SISO constellation

$s$	symbol of a SISO constellation
$R$	rate of a code
$L_t$	length of a space-time code
$d$	diversity of a MIMO system
$t$	time index
$\delta t$	time delay
$t_c$	time index of a singular value crossing
$i, j, k, l$	indexes
$r$	real number
$c$	complex number
$\epsilon$	perturbation
$\kappa$	chi-square variate
$\Pi$	permutation matrix
$\mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{Z}$	unitary matrices
$\mathbf{U}_T, \mathbf{V}_T$	precoding and decoding matrices derived at the transmitter (when specific)
$\mathbf{U}_R, \mathbf{V}_R$	precoding and decoding matrices derived at the receiver (when specific)
$\mathbf{D}, \Sigma$	diagonal matrix with non-negative entries
$\Phi$	diagonal matrix with complex entries of norm 1
$\mathbf{E}r^{T,A}, \mathbf{E}r^{R,A}$	matrices of errors in respectively the RF transmitter and receiver chains, transceiver A
$\mathbf{E}r^{T,B}, \mathbf{E}r^{R,B}$	matrices of errors in respectively the RF transmitter and receiver chains, transceiver B
$\mathbf{R}$	upper triangular matrix with real elements on the diagonal
$\mathcal{W}$	matrix from the Wishart ensemble



# Abstract

Next generation wireless modems are likely to be based on multiple antenna hardware platforms, following the demand for high data rate. Multiple antenna at both end of the link can potentially increase the spectrum efficiency of wireless systems since the capacity of Multiple Input Multiple Output (MIMO) systems is generally higher than the capacity of their Single Input Single Output (SISO) counterparts. Most proposed MIMO architectures assume that the Channel State Information (CSI) is available at the receiver only, or not available at all. This assumption is justified by the fact that CSI is difficult to obtain. Furthermore the gain in capacity due to available CSI at the transmitter is not large on heavy multipath channels.

However, in correlated channels, CSI at the transmitter provides a significant capacity increase over conventional MIMO systems. For example, the capacity of MIMO Ricean channels reduces to the capacity of their Rayleigh component in the asymptotic limit of a large number of antennas when the CSI is not available at the transmitter. This result remains largely valid, even for a small number of antennas. In other words, MIMO systems with no CSI at the transmitter do not use the antenna gain at the transmitter, which is detrimental in correlated propagation environment. Besides, under any propagation environment, the optimal capacity is obtained with CSI at the transmitter. Finally the CSI can be easily obtained at the transmitter in Time Division Duplex (TDD) channels. These three reasons justify a thorough study of MIMO architecture with CSI at transmitter. In such a case, the Singular Value Decomposition (SVD) of the channel matrix gives the optimal precoder and decoder.

This thesis studies the performance of the SVD architecture under varying propagation environments, as well as its robustness to various impairments, e.g. incorrect channel estimation, hardware errors. The aim is to provide a comprehensive understanding of the advantages and weaknesses of the SVD-based transmission architecture.

The precoding and decoding matrix of the SVD architecture are in theory perfectly matched with the CSI matrix. The channel estimation requirements are shown to be mild when the system is synchronized, i.e. the same CSI is used at both the transmitter and the receiver. This is due to the fact that unitary matrices are well conditioned, allowing the error in the estimation to be considered as additional noise. However, in practice, most systems obtain the CSI at transmitter and receiver separately. In such a case, the non-linearity of the SVD imposes stringent constraints on the channel estimation. Furthermore, the SVD architecture relies on the possibility for transmitter and receiver to agree on a unique SVD of the channel matrix. The unicity of the SVD is studied theoretically and an upper bound on the probability of the SVD not being unique is derived, demonstrating that a better than 20 dB of SNR channel estimation is required for SVD system to operate correctly.

In Time Division Duplex (TDD) systems, the channel is reciprocal, allowing the transmitter to obtain the CSI through standard channel estimation on the reverse link. Though the wireless channel is reciprocal, the transmitter and receiver electronics are usually not matched, destroying the reciprocity of the overall transmission channel. SVD systems cannot operate in the presence of such hardware errors. A calibration procedure is proposed to insure that the overall channel is reciprocal and that hardware errors do not affect the system. Delayed CSI at the transmitter is another source of performance loss. MIMO systems without CSI at transmitter outperform their SVD-based counterpart at  $F_d\delta t = 0.038$ , where  $F_d$  is the Doppler frequency,  $\delta t$  the time delay on the CSI at the transmitter, assuming Jake's fading on an i.i.d. channel, SNR=20dB, and perfect channel estimation.  $F_d\delta t = 0.038$  corresponds to walking speed (2m/s) at 5.725 GHz (802.11a frequency band) and CSI delayed by 1ms at the transmitter. The loss of performance of SVD systems can be reduced by filtering the CSI at the transmitter, i.e. prediction of the CSI to match it with the channel. Another identified source of performance loss for SVD systems is the occurrence of rare channel events named singular value crossings, which result in bursts of errors. Singular value crossing can be detected and corrected at the transmitter if the channel is tracked over time. Finally, a new scheme is proposed, combining the advantages of systems with and without CSI at the transmitter: by including the precoding matrix in the channel, the receiver can both recover the CSI for the next transmission in the reverse link and decode the stream without harmful effect of incorrect CSI at the transmitter. This proposed transmission scheme does shift seamlessly from an

SVD system at low mobility to a standard MIMO system for higher mobility.



# Chapter I

## Introduction

Most of us consider telecommunications as nothing more than a useful tool. Calling a person on the other side of the planet hardly surprises us. In an age where technical feats are mere consumer products, what are the goals and challenges of telecommunications?

### I.1 History of telecommunications

Telecommunication has become a part of our everyday life in the past two centuries, so much so that the very meaning of the word itself has been blurred and is mistaken as being synonymous with mobile phones. The word *telecommunications* conveys information about its own etymology. The prefix *tele-* comes from the Greek *τηλη*, and means "remote" or "afar". *Communication* comes from the Latin *communio*, which means "to impart" or "to share", literally "to make common". *Telecommunication* means, strictly speaking, to share something over a distance.

However, the word communication has evolved to the present day meaning of the exchange of information. Hence, telecommunications are adequately defined as the exchange of information over a distance, or over time. As such, telecommunications are the foundation of any society which requires people to send and receive information outside of the "here and now" realm.

#### I.1.1 Origin of telecommunications

Obviously, telecommunications were born with the very first human civilizations. The best example of simple telecommunications is writing, which dates back to 3500 BC. A written

message can be sent to a recipient in a remote place, or can be transmitted to a reader at a later time. Messengers are the second example of telecommunications, e.g. the famous marathon runner Philippides who, in 490 BC, prevented Athen's destruction by delivering a warning message that the Persian army was arriving. Together with these two inventions, writing and messengers, came the first two requirements of telecommunications, namely speed and reliability. Thereafter all telecommunications would be measured in terms of speed of transfer and accuracy in the transmission. Even today's telecommunications industry still focuses on satisfying these two fundamental requirements.

In ancient times, systems were designed to transmit information faster than a human could move, using either sound (drum telegraph) or visible signals (fire and/or smoke). By the year 150 AD, the Roman Empire was covered with a telecommunications network which had a total length of approximately 4500 kilometres. The network consisted of closely spaced towers which enabled them to exchange visual information through predetermined smoke signals. During the French Revolution Claude Chappe rediscovered the idea of an optical telegraph and built a line 240km long which could transfer 196 different signals.

Ultimately, telecommunications were to be linked with electricity and really developed in the middle of the XIXth century. Samuel Morse invented the electromagnetic telegraph in 1837. Elisha Gray and Alexander Graham Bell took out patents for telephones in 1876 and by 1895 Guglielmo Marconi was demonstrating wireless transmission of signals. The engineers involved with early telecommunications innovations were mostly inventors interested in practical progress and consumer products. This explains the rapid impact of telecommunications on society: by 1902 worldwide communication was available on ocean ships; by 1903 more than 3 million phones were installed in the USA alone, and broadcasting stations were commercialised as early as 1922 in Russia, France, England and the USA. Thereafter, progress in telecommunications equipment would mirror the advances in electronics, starting with the invention of the vacuum tube in 1910.

### **I.1.2 Evolution and revolutions of telecommunications**

While telecommunications engineers were focusing on incremental advances until the second world war, a major paradigm shift was soon under way, leading to modern telecommunications theory: the shift towards digital communication. Telecommunication theory was founded by Nyquist [1, 2] in his pioneering work on the maximum signalling rate

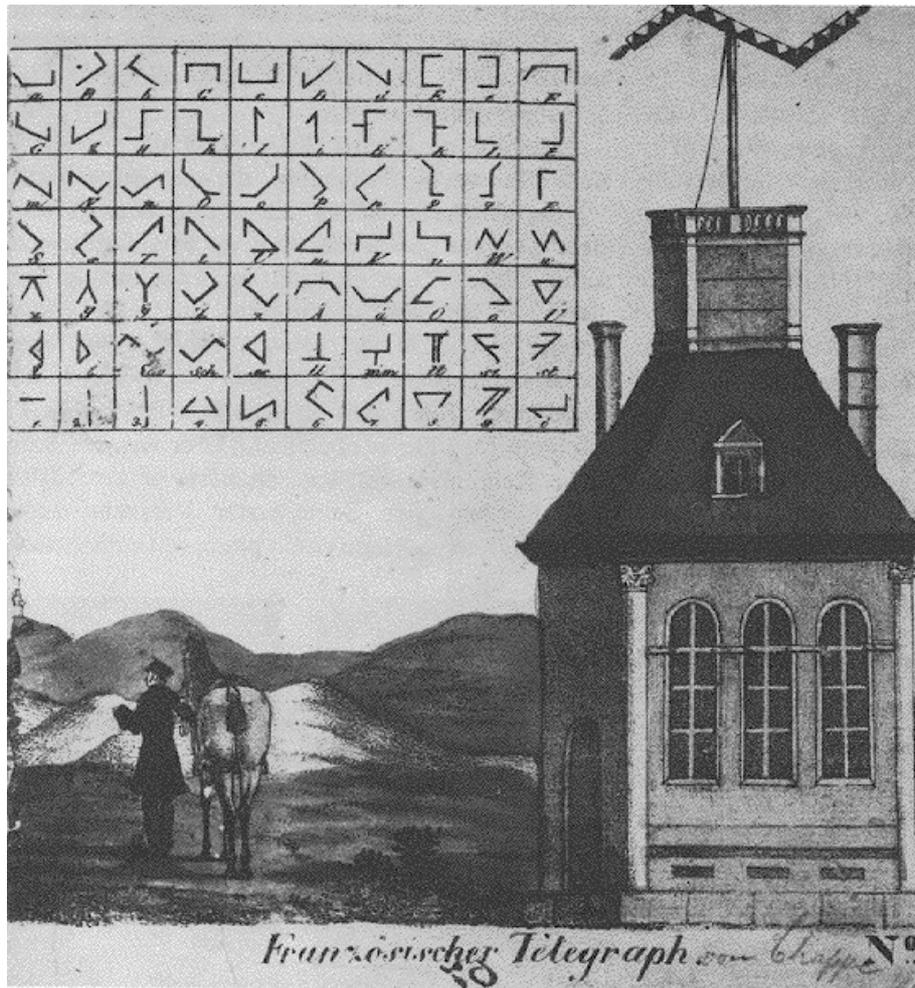


Figure I.1: Chapppe’s telegraph is an early example of digital optical communication using source coding. A message could be transmitted over 240 km in less than half an hour... that is, when the weather was not foggy.

achievable on a telegraph line of a given bandwidth. He was soon followed by Hartley [3] who studied the amount of data that can be transmitted when multiple amplitude levels are used. Both proposed the use of a logarithmic measure of information. Their work was completed by Shannon in two papers published in 1948 [4, 5] where for the first time, a mathematical theory of telecommunications was proposed. This work remains the precursor of most telecommunications advances and defines in particular the capacity of a channel (the maximum amount of information that can be transmitted per second on a channel). This work also proposes a probabilistic framework for telecommunications theory. Kolmogorov and Wiener contributed to the development of the new branch of telecommunications theory known as information theory. Their work consisted of mathematically modelling and deriving an optimal filter for the reception of telecommunications

signals [6, 7]. Information theory refers to the combination of mathematical modelling and telecommunications theory.

Telecommunication was to be revolutionized once again with the development of a new type of user: computers communicating with computers. In 1961 IBM computers started communicating over the telephone line, using what was soon to be known as modulator/demodulators or modems. But the rapid rise of computers and the need to communicate between them led to the development of packet switched networks, more adapted to computer traffic. Coincidentally, vast projects of interconnected computers appeared, more specifically, Arpanet, which initially began as a network connecting universities and the military and armament industry, and later evolved into what is known today as the Internet.

The Cold War and the space race between USSR and the USA paved the way for the rapid development of satellites. As early as 1960 Echo 1, the first communication satellite, was in orbit. It was essentially a mirror reflecting radio waves on earth, but successful transmissions between the USA and France demonstrated the benefits obtainable through satellite telecommunications.

Quite surprisingly, 1960 was also to see the rebirth of optical telecommunications, one of the oldest forms of telecommunications. The invention of the laser ensured the availability of high quality light sources. It triggered research into optical fibres, which were impractical at the time due to high loss. However, by 1980, worldwide optical networks were deployed.

### **I.1.3 Telecommunication: a multi-faceted giant**

In the 50 years, from 1920 to 1970, telecommunications evolved from specialized electrical engineering into its own engineering discipline with its own theory, a wide range of applications and several specialised domains. Telecommunication brings together:

- mathematicians for information theory, coding theory, digital signal processing and network theory
- electronic engineers for the development of all enabling technologies, especially chipsets, which are at the heart of telecommunications equipment
- computer scientists since most telecommunications equipment is implemented partly

in software, allowing, among other advantages, reconfigurability.

- electromagnetic researchers, for all aspects of radio wave emission, reception and propagation
- optical physicists, involved in all aspects of the development of optical communications
- mechanical engineers, involved in satellite communication, but also in the development of micro-electro-mechanical devices.

This diversity illustrates the challenges inherent in the design of a telecommunications system as well as the necessity for the telecommunications engineer to develop an understanding of traditionally unrelated engineering fields.

## **I.2 Recent trends in telecommunications**

### **I.2.1 Current telecommunications**

The development of information and coding theory combined with the development of enabling technologies (especially electronic components) allowed telecommunications to become an everyday commodity readily available in developed countries. Wireless communication has become a part of our everyday life as recently as the past 20 years (whereas telephones were already widespread more than 30 years ago). The development of second generation cellular phone networks redefined mobile phones from a luxury item into a consumer product.

The push for wireless telecommunications was not limited to cellular phones. Relatively inexpensive satellite phones became a reality, allowing consumers to be reached anywhere in the world. Cordless phones, and other low bit rate wireless devices (for example Bluetooth), contributed to proving wireless communication was a strong candidate for future consumer products. The advantages of wireless products over their wired counterparts include straightforward installation, freedom of movement and cutting of all costs associated with cables.

The other major evolution of the telecommunications industry was due to the prevailing use of computers. Voice was no longer the main content of telecommunications networks. Telecommunication networks now had to accommodate for the massive amount

of electronic data being transferred. Consumers are demanding ever growing data rates to support applications such as real time video.

In this context, third generation cellular networks have been designed to accommodate both voice and data, and provide a unique, high data rate, wireless access to the telecommunications network. The deployment of third generation equipment has been delayed due to economic reasons and corresponds to a downturn in the telecommunications industry.

## **I.2.2 Future of telecommunications**

From a technical point of view, it has become increasingly clear that it is extremely difficult to answer the diverse requirements of users with a single standard. Therefore, the industry is considering the possibility of developing wireless products that reconfigure themselves depending on their environment and data transfer requirements. For example, a single device would use a low bit rate cellular standard for a phone call from a car and switch over to a high-speed standard when the user requires internet connectivity from an office. This philosophy of "best connected" is sometimes referred to as the elusive fourth generation.

As mentioned earlier, progress in telecommunications has mostly been measured in terms of achievable data rate. Focusing on the speed of the link has been questioned in the context of cellular networks, arguing that most users do not use high data rate applications. However, this focus has never been questioned concerning Local Area Network (LAN) applications because they are data rate hungry. Computer networks are becoming widespread, not only in business but also for private consumers. However, the costs associated with the installation and maintenance of networking cables are high and make wireless LANs an extremely viable alternative to their wired counterparts, provided they can offer comparable data rates.

## **I.2.3 Towards understanding the Shannon capacity**

The wireless spectrum is a scarce resource which everybody has to share. If two users transmit at the same frequency, they interfere with each other. The radio frequency spectrum is regulated internationally to control user interference. In particular, fixed frequency bands are allocated to telecommunications companies. These companies, in turn, want to provide high data rates, to a maximum number of users, over a fixed bandwidth.

However, Shannon proved that there exists a maximum data rate for a wireless channel and that this maximum data rate (or channel capacity) is a direct function of the bandwidth of the channel, the power of the signal and the amount of noise in the channel. Therefore, until recently, it was considered that the only solution to increasing the data rate of a wireless system was either to increase its bandwidth or to increase the power it radiates. Neither of these two solutions is satisfactory. Since the total bandwidth of the telecommunications system is fixed, increasing the bandwidth of one user directly limits the number of users in the system. It is no more reasonable to increase the power of the wireless system. The first consideration is the simple fact that radiating a large power is impractical for battery-operated devices. The second reason is concern about the effect on human health of intense electromagnetic fields.

Therefore, telecommunications engineers were trying to transmit as closely as possible to the Shannon capacity, which severely limited further increases in the data rate.

This conclusion was shown to be incorrect for a simple reason: the Shannon capacity of a channel also depends on the number of transmitting and receiving antennas. This simple fact had always been overlooked since, intuitively (although incorrectly), transmitting with two antennas at the same frequency appears to be equivalent to creating ones own interference.

## **I.2.4 Multiple antennas wireless systems**

The fact that the capacity of a channel depends on the number of antennas at both the transmitter and the receiver has enormous implications. Higher data rates can be achieved by using multiple antennas without increasing the radiated power or bandwidth. Channels with multiple antennas at both ends are usually referred to as Multiple Input Multiple Output (MIMO) channels.

Several practical systems have been proposed to benefit from the large capacity offered by MIMO channels. The proposed architectures usually aim at either increasing the data rate, or reducing the power requirement of the user. Most of these systems are designed for specific communication scenarios and most of them rely on pessimistic assumptions, e.g. the propagation environment is unknown at the transmitter. Because of these assumptions, most proposed architectures are transmitting at only a fraction of the capacity of the channel.

Theoretical results on the capacity of MIMO channels are now well understood. In particular, the variation in achievable data rate when the transmitter does not know the channel can be well understood by using a mathematical operation: the Singular Value Decomposition (SVD) of the channel. The SVD also suggests a communication architecture that allows transmission at the channel capacity. This well-known architecture has not been studied extensively in the literature, being commonly considered too complex to be implemented.

### I.3 Statement of significance

The typical requirements of users of a wireless LAN are reliable and high data rates. For the wireless propagation channel in such an environment an obvious solution is a MIMO system. Given the wireless LAN environment and user requirements, it is likely that the SVD architecture will be used for the implementation of such a system.

The goal of this thesis is to study the advantages and drawbacks of the SVD architecture. It aims at providing exact results on whether it is implementable, under which assumptions, in which propagation environment and with what expected benefits.

### I.4 Contribution to knowledge

The main contributions of this thesis include:

**-Ergodic capacity series expansions.** Series expansions of the Gaussian approximation of the ergodic MIMO channel capacity are presented in Section II.2.3.3 and highlight the importance of the minimum number of antennas at either the transmitter or the receiver. The series expansions are derived in the asymptotic case of the number of antennas at the receiver (transmitter) being much larger than the number of antennas at the transmitter (receiver). Simulation results support the theoretical results of the series expansion.

**-Ricean channel study.** Section III.5 demonstrates that the ergodic normalized capacity of the Ricean MIMO channel approaches the corresponding normalized capacity of the underlying scattering channel when the antenna numbers are large and no Channel State Information (CSI) is available at the transmitter. Section III.5 also

demonstrates that the capacity variance of the Ricean channel approaches the corresponding variance of the underlying scattering channel when the number of antennas is large. These results highlight the importance of CSI at the transmitter. Upper and lower bounds for the ergodic capacity of the Ricean channel are derived. The accuracy of the bounds is confirmed via simulation.

**-Channel estimation requirements for SVD systems.** An analysis of the effect of incorrect channel estimation on the performance of SVD transmission systems is proposed in Section IV.4. The analysis demonstrates that these effects are negligible when the SNR on the estimation of the channel is much larger than the SNR on the received data. On the contrary, under channel estimation errors, the performance of SVD systems does not increase with increasing SNR: the capacity of SVD systems plateaus at high SNR. Simulation results support this analysis

**-Uniqueness of the SVD of a complex matrix.** The SVD of a complex matrix is not unique, as shown in Section IV.4.2. However, it is possible to select a unique SVD following some criteria when the complex matrix is square and the singular values of multiplicity one. Further technical considerations allow to extend this result to all matrices with singular values of multiplicity one and to handle the very rare occurrence of channel matrices with singular values of multiplicity higher than one. Therefore the SVD of the channel matrix can be derived separately at the transmitter and the receiver.

**-Calibration procedure for SVD systems over reciprocal channels.** MIMO SVD systems are well suited for reciprocal channels since the CSI can be obtained at the transmitter without overhead. However, this technique relies on symmetric reception and transmission chains at each transceiver. This assumption is unrealistic. A calibration procedure is proposed in Section V.2.2 which forces the channel to be reciprocal. This calibration procedure is a form of handshaking at the start of the transmission and relies on the assumption that the imperfections of the RF chains vary slowly in time, allowing calibration to remain valid for large periods of time.

**-Singular value crossing and singular value swapping.** The application of matrix perturbation theory shows in Section V.3 that SVD systems are robust to imperfect channel estimation only when the singular values of the channel matrix have

a multiplicity of one. Imperfections of the system allow this theoretical case study to gain practical applicability. E.g. when the transmitter and the receiver deduce their precoding and decoding matrices from the CSI at two different time-slots, the transmission eigenmodes might be completely different on both time-slots. This event is referred to as a singular value crossing. In practice, singular value crossing usually involves singular subspace swapping: the data sent on subchannel 1 is received on subchannel 2 and the data sent on subchannel 2 is received on subchannel 1. Therefore, simple mechanisms can mitigate the harmful effect of singular value crossings.

**-Modified SVD architecture.** Some of the practical issues occurring on SVD systems can be mitigated by modifying the transmission architecture. In Section V.4, the pilot symbols are sent through the precoding matrix, providing the receiver with information on the channel matrix and the precision of the channel estimation at the transmitter at the same time. This proposed architecture promises high performance when the CSI is accurate at both end of the channel and enjoys a graceful loss of performance, down to the performance of MIMO systems without CSI at the transmitter, when the CSI is inaccurate at the transmitter. Furthermore, the newly proposed architecture has no additional complexity.

**-MIMO PSAM and SVD structure.** The requirement of accurate channel estimation can be achieved through added pilot symbols (larger overhead) or added complexity through the natural extension of PSAM to MIMO channels. An SVD system can filter the CSI in time or directly filter the precoding and decoding matrices. In Section V.5, it is shown that filtering the CSI achieves better performance.

These results are, to the best knowledge of the author, either new and unpublished or previously published by the author.

## I.5 List of publications

The results presented in this thesis have been partly published or accepted for publication in the following articles:

- G. Lebrun, M. Faulkner, P. J. Smith and M. Shaafi, "MIMO Ricean channel capacity," in *Proc. IEEE International Conference on Communications (ICC 2004)*,

20-24 Jun. 2004, pp. 2939-2943,

- G. Lebrun, M. Faulkner, P. J. Smith and M. Shaafi, "MIMO Ricean channel capacity: an asymptotic analysis," *IEEE Transactions on wireless communications*, Vol. 5, No. 5, May 2006,
- G. Lebrun, S. Spiteri and M. Faulkner, "Channel estimation for an SVD-MIMO System," in *Proc. IEEE International Conference on Communications (ICC 2004)*, 20-24 Jun. 2004, pp. 3025-3029,
- S. Spiteri, G. Lebrun and M. Faulkner, "Prediction for time varying SVD Systems," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2004)*,
- S. Spiteri, G. Lebrun and M. Faulkner, "Capacity performance of hardware erred MIMO systems in slowly fading Rayleigh channels," in *Proc. Australian Communication Theory Workshop 2004 (AusCTW)*,
- G. Lebrun, S. Spiteri and M. Faulkner, "Antenna selection and time-varying channel," in *Proc. Australian Telecommunication Cooperative Research Center (ATCRC) conference 2003*,
- S. Spiteri, G. Lebrun and M. Faulkner, "Capacity performance of hardware erred MIMO systems in slowly fading Rayleigh channels," in *Proc. Australian Telecommunication Cooperative Research Center (ATCRC) conference 2003*,
- G. Lebrun, S. Spiteri and M. Faulkner, "MIMO complexity reduction through antenna selection," *Proc. Australian Telecommunications, Networks and Applications Conference (ATNAC) 2003*, Dec. 8-10, 2003, Melbourne, Australia,
- S. Spiteri, G. Lebrun and M. Faulkner, "Effects of hardware induced errors on TDD-based SVD systems," *Proc. Australian Telecommunications, Networks and Applications Conference (ATNAC) 2003*, Dec. 8-10, 2003, Melbourne, Australia,
- G. Lebrun, J. Gao and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *IEEE Transactions on wireless communications*, Vol. 4, No. 2, March 2005, pp. 757-764,

- T. Ying, G. Lebrun and M. Faulkner, "An adaptive channel SVD tracking strategy in time-varying TDD system," *Proc. IEEE Vehicular technology conference 2003, (VTC 2003-Spring)*, Apr. 22-25, 2003 Vol. 1, pp. 769-773,
- G. Lebrun and M. Faulkner, "Analysis of singular value decomposition applied to wireless communications," *Proc. Australian Telecommunication Cooperative Research Center (ATCRC) conference 2002*,
- G. Lebrun, T. Ying and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *Proc. IEEE Global Telecommunications Conference, (Globecom '02)*, Nov. 17-21, 2002 pp. 414-418,
- G. Lebrun, T. Ying, M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *IEE Electronics Letters*, vol. 37, no. 32, pp. 1363-1364, Oct. 2001.

## I.6 Outline

Chapter II introduces the notion of capacity, as the maximum data rate that can be transferred over a channel, under given assumptions. The intimate connection between coding and capacity is highlighted. The impact on the capacity of the number of antennas at both ends of a wireless link is presented in details, both for the ergodic and non-ergodic channels.

Chapter III presents some well-known MIMO transmission architectures and highlights their specificities. The fundamental concepts of diversity and spatial multiplexing are introduced. These concepts allow to answer the challenge of exploiting the high capacity of MIMO channels. Finally, the study of the capacity of Ricean channels stresses the benefits associated with CSI at the transmitter.

The SVD transmission architecture is presented in Section IV. The optimality of the SVD is demonstrated. Two issues linked with the implementation of the SVD are studied in detail: the channel estimation requirements of the system and the uniqueness of the SVD of a complex matrix.

Chapter V analyses the SVD system in a Time Division Duplex (TDD) environment. The impact of delayed and incorrect CSI at the transmitter is examined. The degradation of performance due to singular value crossings is assessed and a practical solution to mitigate this loss of performance is proposed, through the discovery of singular subspace

swapping. A novel transmission architecture is proposed to prevent catastrophic loss of performance when the CSI is inaccurate. The new SVD architecture combines the performance of SVD systems when the CSI is accurate with the performance of MIMO systems without CSI at the transmitter when the CSI is inaccurate. Finally, an extension of PSAM to MIMO channel is proposed to improve the accuracy of channel estimation without increasing the overhead.



# Chapter II

## Channel capacity

### II.1 Notion of capacity

The theoretical study of telecommunications systems relies on the choice of a theoretical model to represent the systems. Telecommunication (wired or wireless) consists usually of the emission of a physical signal that varies in time, denoted in the following  $x(t)$ , where  $t$  represents the time index. In most telecommunications systems, this signal is an electromagnetic wave. This signal is received as another signal  $y(t)$ . The relationship between  $x(t)$  and  $y(t)$  depends only on the channel, i.e. the medium over which the signal is transmitted. Telecommunication theory is concerned with:

- determining for a given channel how much information can be transmitted over the channel, i.e. the capacity of the channel,
- designing  $x(t)$  to transmit information in a fast and reliable way,
- reliably recovering the information transmitted  $x(t)$  from the received signal  $y(t)$ .

Though very general, this model is difficult to analyse. Further assumptions enable simplification of this model, specifically the sampling theorem allows to restrict the analysis to discrete time systems when the signals are bandlimited (Section II.1.1). Information theory provides the theoretical framework to define the notion of information (Section II.1.2) as well as the capacity of the channel (Section II.1.3).

Chapter II provides a literature review of the main results concerning the capacity of both SISO and MIMO single user channels in a variety of situations: CSI at the receiver/transmitter, ergodic or block fading channel. MIMO channel capacity is shown to

increase with the number of antennas at both ends of the wireless link in Section II.2. A novel series expansion of the capacity of the ergodic channel with no CSI at the transmitter is presented in Section II.2.3.3. The series expansion is new and unpublished to the knowledge of the author. The series expansion shows that symmetric antenna allocation (same number of antennas at the transmitter and the receiver) maximizes the capacity of a MIMO channel with a given total number of antennas. Finally, Section II.3 shows that the assumptions are compatible with wireless LANs standards 802.11a and Hyperlan 2.

### II.1.1 The sampling theorem

The sampling theorem is central in telecommunications since it demonstrates that analog and digital communications are equivalent, provided the analog transmission is band limited. Though the sampling theorem has been applied to telecommunications by Nyquist [1], it had been demonstrated previously in other forms by mathematicians [8]. The form given here corresponds to the theorem stated by Shannon [9].

**Theorem 1.** *Sampling Theorem If a function  $f(t)$  contains no frequencies higher than  $W$  Hz, it is completely determined by giving its ordinates at a series of points spaced  $1/2W$  seconds apart.*

The intuitive justification is that, if  $f(t)$  contains no frequencies higher than  $W$ , it cannot change to a substantially new value in a time less than one-half cycle of the highest frequency, that is,  $1/2W$ . The exact mathematical proof follows.

Let  $F(w)$  be the spectrum of  $f(t)$ . Then

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(w) \exp(\sqrt{-1}wt)dw = \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} F(w) \exp(\sqrt{-1}wt)dw, \quad (\text{II.1})$$

since  $F(w)$  is assumed zero outside the band  $W$ . This relationship is verified at the sampling points  $t = \frac{k}{2W}$  where  $k$  is any integer ( $k \in \mathbb{Z}$ ):

$$f\left(\frac{k}{2W}\right) = \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} F(w) \exp(\sqrt{-1}w \frac{k}{2W})dw. \quad (\text{II.2})$$

The integral on the right is the  $k^{\text{th}}$  coefficient in the Fourier-series expansion of the function  $F(w)$ , taking the interval  $-W$  to  $+W$  as a fundamental period. So, the samples  $f(\frac{k}{2W})$  determine the Fourier coefficients in the series expansion of  $F(w)$ . Furthermore, since  $F(w)$  is zero outside  $[-W, W]$ ,  $F(w)$  is uniquely defined by the Fourier coefficients of its

series expansion and, in turn, the samples  $f(\frac{k}{2W})$ . Therefore, the samples determine the function  $f(t)$  completely, since  $F(w)$  determines  $f(t)$ .

Additionally,  $f(t)$  can be reconstructed from its samples:

$$f(t) = \sum_{k=-\infty}^{\infty} f\left(\frac{k}{2W}\right) \frac{\sin(2\pi W(t - \frac{k}{2W}))}{2\pi W(t - \frac{k}{2W})}. \quad (\text{II.3})$$

Finally, it is possible to prove a similar result if the signal is not limited to the frequency  $[0, W]$  but bandlimited to the band  $[W_c - W/2, W_c + W/2]$  [9].

Most wireless communication systems are bandlimited, due to bandwidth allocation, as explained in Section I.2.3. Therefore, applying the sampling theorem, the study of telecommunications systems can be limited to the study of sampled systems without loss of generality. The transmitted signal and received signals can be restricted to  $x(kT_s)$  and  $y(kT_s)$  respectively, with  $k \in \mathbb{Z}$ .

## II.1.2 Elementary information theory

It was shown in Section I that the goal of a telecommunications system is to transmit information. However, no precise definition of information was given. A practical example can help to clarify the concept of information.

Consider a system transmitting, every five minutes, the answer to the question: "Is there a fire in John's house?" Most of the time, the system indicates that John's house is fine, which John naturally discards as a lack of information. However, if the system raises the alarm about a fire, it is natural to consider that very important information has been transmitted. Intrinsically, this event carries more information because it happens less often. Thus, information is somehow related to the inverse of the probability of occurrence. For our communication system, consider the event  $x = s_1$ , where  $s_1$  is any symbol, i.e. any value of the physical signal used to carry the message. The information carried by the event  $x = s_1$  depends on the probability of this event  $p_1 = \mathcal{P}(x = s_1)$ . It is, therefore, useful to define a function  $\mathcal{H}(p)$ , which measures the amount of information in the occurrence of an event of probability  $p$ , and has the following properties:

- $\mathcal{H}(p) \geq 0$ , the measure of information is real, non-negative,
- $\mathcal{H}(p_1 p_2) = \mathcal{H}(p_1) + \mathcal{H}(p_2)$  for independent events, i.e. the amount of information carried by two unrelated events is the sum of the amount of information carried by each event,

- $\mathcal{H}(p)$  is a continuous function of  $p$ .

The only functions following these requirements are  $\mathcal{H}_a(p) = -\log_a(p)$ , where  $\log_a(\cdot)$  denotes the logarithm in base  $a$  [10]. The base 2 logarithm is usually used, and the resulting unit of information is called a bit (binary digit). In the following, it is considered that  $a = 2$ , and the corresponding index is dropped.

Assume that the transmitted signal  $x(kT_s)$  can take the values  $s_1, s_2, \dots, s_q$  with respective probabilities  $p_1, p_2, \dots, p_q$ .  $x(kT_s)$  is a random variable denoted  $\underline{x}(kT_s)$ . The average information transmitted  $\mathcal{H}(\underline{x})$  is called the entropy of the source, with

$$\mathcal{H}(\underline{x}) = \sum_{i=1}^q p_i \mathcal{H}(p_i) = \sum_{i=1}^q p_i \log_2\left(\frac{1}{p_i}\right). \quad (\text{II.4})$$

### II.1.3 Capacity of a channel

The transmission of information can now be defined precisely. Consider  $\mathbf{x} = [x(0), x(T_s), \dots, x(kT_s)]$  the overall transmitted signal and  $\mathbf{y} = [y(0), y(T_s), \dots, y(kT_s)]$  the overall received signal. The actual transmission can be modelled as the set of conditional probabilities  $\mathcal{P}(\mathbf{y}|\mathbf{x})$ , the probability of receiving  $\mathbf{y}$ , having transmitted  $\mathbf{x}$ . This set of probabilities determines a channel over which transmission occurs.

Prior to reception, the probability of  $\mathbf{x}$  is  $\mathcal{P}(\mathbf{x})$ . After reception of  $\mathbf{y}$ , the probability that the input symbol was  $\mathbf{x}$  becomes  $\mathcal{P}(\mathbf{x}|\mathbf{y})$ . The change in probability measures how much the receiver learned from the reception of  $\mathbf{y}$ , and is called the mutual information, defined as

$$\mathcal{I}(\mathbf{x}; \mathbf{y}) = \log_2\left(\frac{1}{\mathcal{P}(\mathbf{x})}\right) - \log_2\left(\frac{1}{\mathcal{P}(\mathbf{x}|\mathbf{y})}\right) = \log_2\left(\frac{\mathcal{P}(\mathbf{x}|\mathbf{y})}{\mathcal{P}(\mathbf{x})}\right) \quad (\text{II.5})$$

It is interesting to determine the average mutual information (also called system mutual information):

$$\begin{aligned} \mathcal{I}(\underline{\mathbf{x}}; \underline{\mathbf{y}}) &= \sum_{\mathbf{x}} \mathcal{P}(\underline{\mathbf{x}} = \mathbf{x}) \mathcal{I}(\underline{\mathbf{x}} = \mathbf{x}; \underline{\mathbf{y}}) \\ &= \sum_{\mathbf{x}} \mathcal{P}(\underline{\mathbf{x}} = \mathbf{x}) \sum_{\mathbf{y}} \mathcal{P}(\underline{\mathbf{y}} = \mathbf{y} | \underline{\mathbf{x}} = \mathbf{x}) \mathcal{I}(\underline{\mathbf{x}} = \mathbf{x}; \underline{\mathbf{y}} = \mathbf{y}) \\ &= \sum_{\mathbf{x}} \sum_{\mathbf{y}} \mathcal{P}(\underline{\mathbf{x}} = \mathbf{x}, \underline{\mathbf{y}} = \mathbf{y}) \mathcal{I}(\underline{\mathbf{x}} = \mathbf{x}; \underline{\mathbf{y}} = \mathbf{y}), \end{aligned} \quad (\text{II.6})$$

where  $\underline{\mathbf{x}}$  and  $\underline{\mathbf{y}}$  are random variables. In a slight abuse of notation,  $\mathcal{P}(\mathbf{x})$  denotes  $\mathcal{P}(\underline{\mathbf{x}} = \mathbf{x})$  in the following.

The natural question that arises is, given the conditional probabilities  $\mathcal{P}(\mathbf{y}|\mathbf{x})$ , what is the maximum amount of information that can be transmitted over the channel? The only

parameter that can be chosen is  $\mathcal{P}(\mathbf{x})$  the set of probabilities of the possible transmitted symbol. The capacity of the channel is defined as

$$C = \max_{\mathcal{P}(\mathbf{x})} \mathcal{I}(\mathbf{x}; \mathbf{y}), \quad (\text{II.7})$$

the maximum (over every set of possible transmitted signals) average mutual information, and is expressed either in bits per second per Hertz (bps/Hz) or alternatively in bits per channel use.

### II.1.4 Coding theorem

The importance of the capacity of channels in telecommunications is due to the Shannon theorem, which links this theoretical tool with transmission devices.

**Theorem 2.** *Shannon Theorem* For any  $R < C$  and any probability of error  $p_e$ , given a channel of capacity  $C$ , there exists a code with rate  $R$  that can transmit with a probability of error smaller than  $p_e$ .

On the contrary, for any  $R > C$ , there exists a minimum probability of error  $p_e^{\min}$ , i.e. no code with  $R > C$  has a probability of error smaller than  $p_e^{\min}$ .

The demonstration of the Shannon theorem is given in [9]. However it is quite technical and is not reproduced here. The main insights from the Shannon theorem are:

- optimal codes are of infinite length,
- random codes are good candidates to be optimal, i.e. random codes can be as close to optimal as desired by increasing the code's length.

Because of the first property, it is impractical to use a random code to transmit close to the capacity. A finite delay in the decoding is a requirement of telecommunications systems. Therefore, the capacity of the channel is considered as an upper bound on achievable rate. It is useful to understand how effective a code is and how much improvement can be achieved on a given system by improving the coding.

### II.1.5 Channel model

Following the results from the previous section, it is necessary to obtain the set of probabilities  $\mathcal{P}(\mathbf{x}, \mathbf{y})$  to study the telecommunications link. This set of probabilities is directly

linked to the physical properties of the propagation medium, e.g. the optic fiber, the wireless spectrum, the copper cable. It is also linked to the equipment in the telecommunications system. Both the propagation medium and the telecommunications equipment distort the signal. The channel is defined as the system formed by the combination of the equipment and the medium.

Distortion occurs due to several factors:

- non-linear distortion due to non-linearities in transmission equipment,
- frequency offset, which results from the use of a carrier,
- phase jitter, a low frequency modulation due to coupling of the power line,
- impulse noise, due to equipment switching,
- thermal noise, due to thermal agitation of electrons in the receiving device
- interference noise, due to other transmissions
- fading, due to the attenuation of the signal, depending on the distance, the frequency and the propagation environment.

However, all impairments due to equipment imperfection can be reduced through better engineering design, with an associated cost. The trade-off between performance and cost of the equipment is a complex topic which would require a thorough study and therefore is kept out of the scope of this thesis. The only impairment that cannot be reduced through better engineering design is thermal noise, since it is inherent to electronic equipment. If the model is limited to the intrinsic distortion of the transmission medium and equipment, the channel can be modeled as a linear filter that introduces amplitude and delay distortion and adds thermal noise. Specifically the channel is composed of a time-variant impulse response  $\mathbf{h}(kT_s) = [h_0(kT_s), h_1(kT_s), \dots, h_\tau(kT_s)]$  and additive noise  $n(kT_s)$ . Therefore, the transmission equation becomes

$$y(kT_s) = \left[ \sum_{j=0}^{\tau} h_j((k-j)T_s)x((k-j)T_s) \right] + n(kT_s). \quad (\text{II.8})$$

Obviously, the set of probabilities  $\mathcal{P}(\mathbf{x}, \mathbf{y})$  is entirely defined by the set of  $\mathbf{h}(kT_s)$  and  $n(kT_s)$  for all  $k$ .

Additional assumptions allow us to further simplify the model.

- Thermal noise can be modelled as Additive White Gaussian Noise (AWGN) with mean zero, and a chosen variance, depending on the Signal to Noise Ratio (SNR).
- The channel frequency response can be considered flat over the bandwidth of interest, i.e.  $h_l = 0, \forall l \neq 0$ . The assumption of flat-fading is further justified in Section II.3.
- The channel response  $h(kT_s)$  can be modelled as a random variable. The random variable is generally assumed to be zero mean complex Gaussian (Rayleigh fading channel) or non-zero mean complex Gaussian (Ricean fading channel).

The transmission equation becomes

$$\underline{y} = \underline{h}x + \underline{n}, \quad (\text{II.9})$$

where  $\underline{h}$  and  $\underline{n}$  are complex Gaussian random variables.

Finally, the detection is assumed coherent, i.e. the receiver knows  $h$  the realization of the channel  $\underline{h}$ , also referred to as the CSI. The practicality of this assumption is discussed in Section II.3.

## II.1.6 Gaussian complex random variable

The Gaussian complex distribution has a central role in telecommunications and is defined by the Probability Density Function (PDF) of the random variable  $\underline{y}$  with mean  $\mu_{\underline{y}}$  and variance  $\sigma_{\underline{y}}^2$

$$\gamma(y) = \frac{1}{\pi\sigma_{\underline{y}}^2} \exp\left(-\frac{(y - \mu_{\underline{y}})^*(y - \mu_{\underline{y}})}{\sigma_{\underline{y}}^2}\right), \quad (\text{II.10})$$

where  $y^*$  denotes the transpose conjugate of  $y$ . Therefore,

$$\begin{aligned} \mathcal{H}(\underline{y}) &= E_{\gamma}[-\log_2(\gamma(y))] \\ &= \log(\pi\sigma_{\underline{y}}^2) + \log_2(e)E_{\gamma}\left[\frac{(y - \mu_{\underline{y}})^*(y - \mu_{\underline{y}})}{\sigma_{\underline{y}}^2}\right] \\ &= \log_2(\pi\sigma_{\underline{y}}^2e), \end{aligned} \quad (\text{II.11})$$

where  $E_{\gamma}[\cdot]$  denotes the expectancy with  $\gamma$ . The importance of the complex Gaussian distribution comes from the following theorem.

**Theorem 3.** *The complex Gaussian distribution is the distribution of which entropy is maximal.*

Proof: let  $\text{pdf}_{\underline{y}}$  be any density function satisfying  $\int_{\mathbb{C}} \text{pdf}_{\underline{y}}(y)yy^*dy = \sigma_{\underline{y}}^2$ . Clearly  $\log_2(\gamma(y))$  is a linear function of  $yy^*$ . Furthermore,

$$\begin{aligned} \int_{\mathbb{C}} \text{pdf}_{\underline{y}}(y)yy^*dy &= \sigma_{\underline{y}}^2 = \int_{\mathbb{C}} \gamma(y)yy^*dy \\ \Rightarrow \int_{\mathbb{C}} \text{pdf}_{\underline{y}}(y)(c_1yy^* + c_2)dy &= \int_{\mathbb{C}} \gamma(y)(c_1yy^* + c_2)dy \quad \forall c_1, c_2 \in \mathbb{C}, \end{aligned} \quad (\text{II.12})$$

implying that

$$E_{\text{pdf}_{\underline{y}}}[\log_2(\gamma(y))] = E_{\gamma}[\log_2(\gamma(y))]. \quad (\text{II.13})$$

Then

$$\begin{aligned} \mathcal{H}(\text{pdf}_{\underline{y}}) - \mathcal{H}(\underline{y}) &= -\int_{\mathbb{C}} \text{pdf}_{\underline{y}}(y) \log_2(\text{pdf}_{\underline{y}}(y))dy + \int_{\mathbb{C}} \gamma(y) \log_2(\gamma(y))dy \\ &= -\int_{\mathbb{C}} \text{pdf}_{\underline{y}}(y) \log_2(\text{pdf}_{\underline{y}}(y))dy + \int_{\mathbb{C}} \text{pdf}_{\underline{y}}(y) \log_2(\gamma(y))dy \\ &= \int_{\mathbb{C}} \text{pdf}_{\underline{y}}(y) \log_2\left(\frac{\gamma(y)}{\text{pdf}_{\underline{y}}(y)}\right)dy, \end{aligned} \quad (\text{II.14})$$

and  $\log_2$  is a concave function, so

$$\begin{aligned} \log_2(\text{pdf}_{\underline{y}}(y)) - \log_2(\gamma(y)) &\geq \frac{1}{\text{pdf}_{\underline{y}}(y)} \times (\text{pdf}_{\underline{y}}(y) - \gamma(y)) \quad \forall y \\ \Rightarrow \int_{\mathbb{C}} \text{pdf}_{\underline{y}}(y) \log_2\left(\frac{\text{pdf}_{\underline{y}}(y)}{\gamma(y)}\right)dy &\geq \int_{\mathbb{C}} \text{pdf}_{\underline{y}}(y)dy - \int_{\mathbb{C}} \gamma(y)dy = 0 \end{aligned} \quad (\text{II.15})$$

Combining (II.14) and (II.15) concludes the demonstration.

## II.1.7 Capacity of the flat-fading channel

With the previous definitions, it is possible to derive the capacity of the channel. Considering  $h$  fixed and known at the receiver, the channel is effectively similar to the well-known AWGN channel. The definition of the capacity combined with (II.5) leads to

$$C(h) = \max_{\text{pdf}_{\underline{x}}} \mathcal{I}(\underline{x}; \underline{y}) = \mathcal{H}(\underline{y}) - \mathcal{H}(\underline{y}|\underline{x}) = \mathcal{H}(\underline{y}) - \mathcal{H}(\underline{n}), \quad (\text{II.16})$$

since the receiver knows  $h$ . Thus, maximising  $\mathcal{I}(\underline{x}; \underline{y})$  is equivalent to maximizing  $\mathcal{H}(\underline{y})$ . That implies that  $\underline{y}$  has to be a complex Gaussian random variable. Therefore, observing (II.9), the transmitted signal has to be Gaussian to achieve the capacity of the channel.

To achieve the capacity,  $\underline{x}$  is a complex Gaussian random variable with variance  $\sigma_{\underline{x}}^2$ ,  $\underline{n}$  is a complex Gaussian random variable with variance  $\sigma_{\underline{n}}^2$  and  $\underline{y}$  is a complex Gaussian random variable with variance  $\sigma_{\underline{n}}^2 + hh^*\sigma_{\underline{x}}^2$ , justifying

$$\begin{aligned} C(h) &= \mathcal{H}(\underline{y}) - \mathcal{H}(\underline{n}) \\ &= \log_2(\pi e(\sigma_{\underline{n}}^2 + hh^*\sigma_{\underline{x}}^2)) - \log_2(\pi e\sigma_{\underline{n}}^2) \\ &= \log_2\left(1 + \frac{\sigma_{\underline{x}}^2}{\sigma_{\underline{n}}^2}hh^*\right). \end{aligned} \quad (\text{II.17})$$

## II.1.8 Capacity of the non-ergodic channel

Suppose  $h$  is chosen randomly at the beginning of the transmission and fixed thereafter. This assumption corresponds to a system transmitting frames, with coding/decoding performed on a frame by frame basis and the length of the frame smaller than the coherence time of the channel, i.e. the channel is fixed for the length of a frame.

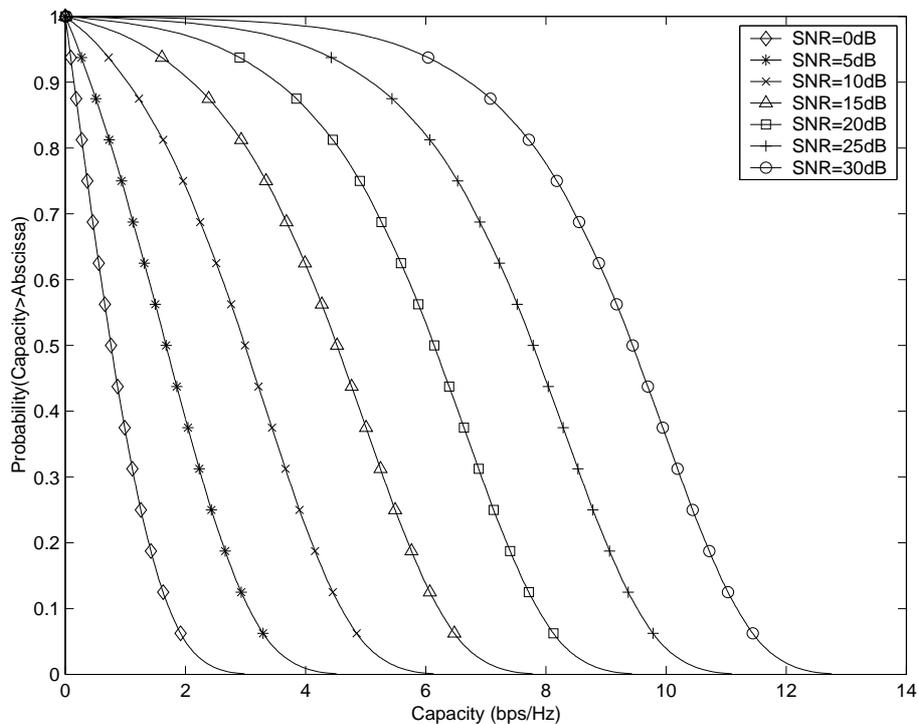


Figure II.1: CCDF of SISO Rayleigh channel

The capacity depends on  $h$ , the realization of the channel and  $\underline{h}$  is a random variable. Therefore, the capacity itself is a random variable. The Complementary Cumulative Density Function (CCDF) of the channel is plotted in Fig. II.1 for different SNRs.

It is possible to define  $C_{out}(p)$ , the outage capacity, as the maximum capacity achieved  $100 \times p\%$  of the time. Using a code designed for a channel capacity of  $C_{out}(p)$  directly leads to a frame error rate of  $1 - p$ .

## II.1.9 Capacity of the ergodic channel

When either coding occurs on a large number of frames, or the coherence time of the channel is much smaller than the length of the frame, the channel is ergodic. The capacity of the channel can then be expressed as:

$$C_{ergodic} = E_{\underline{h}}[\log_2(1 + \frac{\sigma_x^2}{\sigma_n^2} \underline{h} \underline{h}^*)]. \quad (\text{II.18})$$

The ergodic capacity of the channel is plotted as a function of the SNR in Fig. II.2.

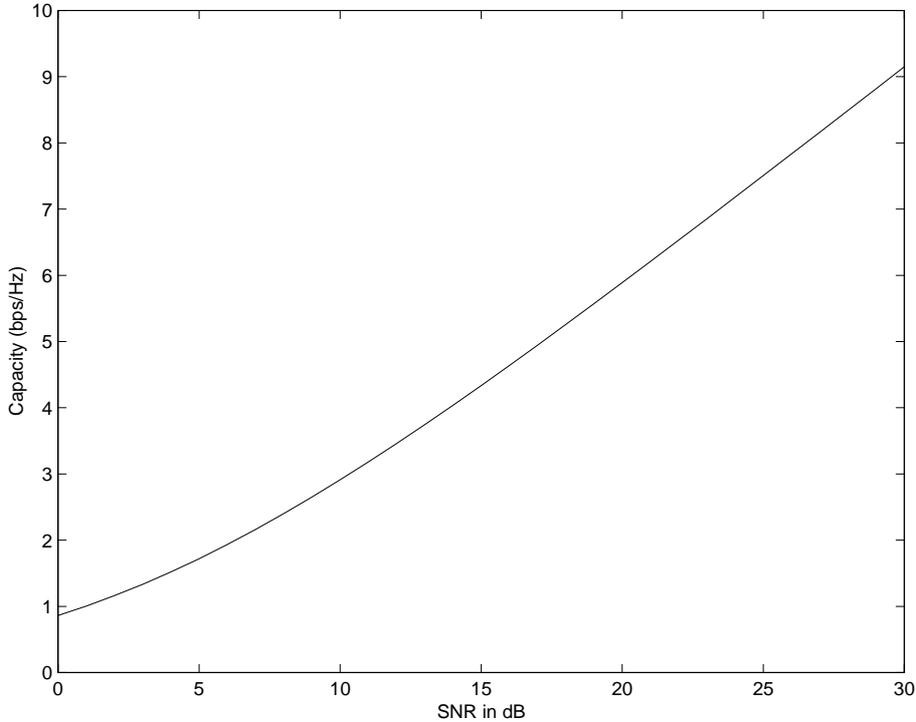


Figure II.2: Capacity of the ergodic SISO Rayleigh channel vs SNR in dB.

Interestingly, at high SNRs, the capacity increases linearly with the SNR expressed in decibels (dB). Therefore, the power at the transmitter has to be doubled to gain an extra bit of capacity.

## II.2 MIMO Capacity

Most of the research effort in the field of telecommunications has been spent on improving coding techniques to achieve data rates close to the capacity at acceptable probability of error (depending on the application). The discovery of turbo-codes by Berrou et al. [11] was a considerable step in this regard. Therefore, it became increasingly clear that further modification of the coding would not result in a significant improvement of the transmission link. It is clear from the previous section that for a given  $\underline{h}$  and  $\underline{n}$ , the only other solution for increasing the capacity of the channel was to increase the power of the signal  $\sigma_x^2$ . This

is undesirable since it also increases the interference to other co-channel users. Higher data rates could only be achieved by discovering channels with higher capacity.

Section II.1 implicitly assumes that the transmitter and the receiver use only one antenna. Several practical systems implemented in cellular networks proved that using several antennas could improve the quality of the wireless link. This directly triggered interest in MIMO channels.

## II.2.1 MIMO channel model

Consider a system with  $M_T$  antennas at the transmitter and  $M_R$  antennas at the receiver. Transmission is over a flat fading channel. Therefore the time index is dropped and the analysis is conducted on the transmission of a single sample.

$\mathbf{x}$  is a vector of the input symbols ( $\mathbf{x} \in \mathbb{C}^{M_T}$ ), i.e.  $\mathbf{x} = [x_1, x_2, \dots, x_{M_T}]$  with  $x_1$  the symbol transmitted on the first antenna.

$\mathbf{H}$  the channel matrix ( $\mathbf{H} \in \mathbb{C}^{M_R \times M_T}$ ) is the realization of a random variable matrix assumed to have complex Gaussian random variable entries with zero mean (Rayleigh channel) or non-zero mean (Ricean channel). The entries of  $\mathbf{H}$  can be correlated random variables (correlated channel) or Independent and Identically Distributed (i.i.d. channel).

$\mathbf{n}$  is a vector of AWGN on the receiving antennas ( $\mathbf{n} \in \mathbb{C}^{M_R}$ ).  $\mathbf{n}$  is complex Gaussian, with zero mean and equal variance in the independent real and imaginary components. It is also assumed that the noise at each receiving antenna is independent and the transmitted power is normalized by the noise power at a single receiving antenna. This can be written as:

$$E[\mathbf{n}\mathbf{n}^*] = \mathbf{I}_{M_R}, \quad (\text{II.19})$$

where  $\mathbf{I}_{M_R}$  is the identity matrix of dimension  $M_R$ .

Assumptions are similar for Single Input Single Output (SISO) or MIMO channels. The only significant difference between the two cases is the possibility in the MIMO case for the entries of the vector or matrices to be correlated. The influence of this parameter is discussed in Section II.3.

The vector of received symbols can be expressed as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}. \quad (\text{II.20})$$

## II.2.2 Circularly symmetric complex Gaussian distribution

The vector equivalent of the Gaussian distribution is a circularly symmetric Gaussian distribution, i.e. a vector  $\underline{\mathbf{x}}$  with entries being complex Gaussian random variable and,

$$\text{defining } \check{\underline{\mathbf{x}}} = \begin{bmatrix} \Re(\underline{\mathbf{x}}) \\ \Im(\underline{\mathbf{x}}) \end{bmatrix},$$

$$E[(\check{\underline{\mathbf{x}}} - E[\check{\underline{\mathbf{x}}}])(\check{\underline{\mathbf{x}}} - E[\check{\underline{\mathbf{x}}}]^*) = \frac{1}{2} \begin{bmatrix} \Re(\mathbf{Q}) & -\Im(\mathbf{Q}) \\ \Im(\mathbf{Q}) & \Re(\mathbf{Q}) \end{bmatrix}, \quad (\text{II.21})$$

where  $\Re(c)$  is the real part of  $c$ ,  $\Im(c)$  the complex part of  $c$  and  $\mathbf{Q}$  an hermitian positive definite matrix. Therefore a circularly complex Gaussian random vector is specified by prescribing  $E[\underline{\mathbf{x}}]$  and  $E[(\underline{\mathbf{x}} - E[\underline{\mathbf{x}}])(\underline{\mathbf{x}} - E[\underline{\mathbf{x}}])^*] = \mathbf{Q}$ .

The probability density of a circularly symmetric complex Gaussian with mean  $\underline{\mu}_{\underline{\mathbf{x}}}$  and covariance  $\mathbf{Q}$  is given by

$$\gamma(\underline{\mathbf{x}}) = \frac{1}{\det(\pi\mathbf{Q})} \exp(-(\underline{\mathbf{x}} - \underline{\mu}_{\underline{\mathbf{x}}})^* \mathbf{Q}^{-1} (\underline{\mathbf{x}} - \underline{\mu}_{\underline{\mathbf{x}}}). \quad (\text{II.22})$$

The differential entropy of a circularly symmetric complex Gaussian with mean zero and covariance  $\mathbf{Q}$  is given by

$$\begin{aligned} \mathcal{H}(\gamma) &= E_{\gamma}[-\log_2 \gamma(\underline{\mathbf{x}})] \\ &= \log_2 \det(\pi e \mathbf{Q}). \end{aligned} \quad (\text{II.23})$$

As in the SISO case, the importance of the Gaussian distribution comes from the following theorem.

**Theorem 4.** *Suppose the complex random vector  $\underline{\mathbf{x}} \in \mathbb{C}^{M_R}$  is zero-mean and satisfies  $E[\underline{\mathbf{x}}\underline{\mathbf{x}}^*] = \mathbf{Q}$ , then the entropy of  $\underline{\mathbf{x}}$  satisfies  $\mathcal{H}(\underline{\mathbf{x}}) \leq \log_2 \det(\pi e \mathbf{Q})$  with equality if and only if  $\underline{\mathbf{x}}$  is a circularly symmetric complex Gaussian.*

The demonstration is similar to its equivalent in the SISO case [12].

Finally, consider  $\mathbf{A}$  a complex matrix,  $\mathbf{A} \in \mathbb{C}^{i \times j}$ , if  $\underline{\mathbf{x}}$  is circularly symmetric complex Gaussian, so is  $\mathbf{A}\underline{\mathbf{x}}$  [12]. If  $\underline{\mathbf{x}}$  and  $\underline{\mathbf{n}}$  are circularly symmetric complex Gaussian, so is  $\underline{\mathbf{x}} + \underline{\mathbf{n}}$  [12].

## II.2.3 Capacity of the MIMO flat-fading channel

Assume  $\mathbf{H}$  fixed and known at the receiver. The capacity of the channel is given by

$$C(h) = \max_{\text{pdf}_{\underline{\mathbf{x}}}} \mathcal{I}(\underline{\mathbf{x}}; \underline{\mathbf{y}}) = \mathcal{H}(\underline{\mathbf{y}}) - \mathcal{H}(\underline{\mathbf{y}}|\underline{\mathbf{x}}) = \mathcal{H}(\underline{\mathbf{y}}) - \mathcal{H}(\underline{\mathbf{n}}), \quad (\text{II.24})$$

since the receiver knows  $\mathbf{h}$ . Thus, maximizing  $\mathcal{I}(\underline{\mathbf{x}}; \underline{\mathbf{y}})$  is equivalent to maximizing  $\mathcal{H}(\underline{\mathbf{y}})$ .

Theorem 4 implies that the capacity is reached if and only if  $\underline{\mathbf{y}}$  is circularly symmetric complex Gaussian, which implies in turn, from (II.20), that  $\underline{\mathbf{x}}$  is also circularly symmetric complex Gaussian.

The transmitted signal is usually constrained by a power limitation, i.e.  $E[\underline{\mathbf{x}}^* \underline{\mathbf{x}}] \leq P$ . This, in turn, helps specifying the nature of  $\underline{\mathbf{x}}$ : if  $\underline{\mathbf{x}}$  satisfies the power limitation, so does  $\underline{\mathbf{x}} - E[\underline{\mathbf{x}}]$ . Thus, we can restrict our attention to zero-mean  $\underline{\mathbf{x}}$  since a non-zero mean would only result in a non-zero mean  $\underline{\mathbf{y}}$  which has the same entropy as the corresponding zero-mean  $\underline{\mathbf{y}}$ .

Therefore, if  $\underline{\mathbf{x}}$  is zero-mean with covariance  $E[\underline{\mathbf{x}} \underline{\mathbf{x}}^*] = \mathbf{Q}$  then  $\underline{\mathbf{y}}$  is zero-mean with covariance  $E[\underline{\mathbf{y}} \underline{\mathbf{y}}^*] = \mathbf{H} \mathbf{Q} \mathbf{H}^* + \mathbf{I}_{M_R}$ . The mutual information of the channel is given by

$$\mathcal{I}(\underline{\mathbf{x}}; \underline{\mathbf{y}}) = \log_2 \det(\mathbf{I}_{M_R} + \mathbf{H} \mathbf{Q} \mathbf{H}^*). \quad (\text{II.25})$$

The exact formula for the capacity of the channel depends on further assumptions, which in turn determine the matrix  $\mathbf{Q}$  providing the capacity achieving distribution.

### II.2.3.1 Ergodic Rayleigh i.i.d. channel, no CSI at the transmitter

Assume the entries of  $\underline{\mathbf{H}}$  are independent and zero-mean Gaussian with independent real and imaginary parts, each with variance 1/2. Therefore each entry of  $\underline{\mathbf{H}}$  has uniformly distributed phase and Rayleigh distributed magnitude, with expected magnitude square equal to unity.

The mutual information is given by

$$\begin{aligned} \mathcal{I}(\underline{\mathbf{x}}; (\underline{\mathbf{y}}, \underline{\mathbf{H}})) &= \mathcal{I}(\underline{\mathbf{x}}; \underline{\mathbf{H}}) + \mathcal{I}(\underline{\mathbf{x}}; \underline{\mathbf{y}} | \underline{\mathbf{H}}) \\ &= \mathcal{I}(\underline{\mathbf{x}}; \underline{\mathbf{y}} | \underline{\mathbf{H}}) \\ &= E_{\underline{\mathbf{H}}}[\mathcal{I}(\underline{\mathbf{x}}; \underline{\mathbf{y}} | \underline{\mathbf{H}} = \mathbf{H})]. \end{aligned} \quad (\text{II.26})$$

Since  $\mathbf{Q}$  is positive definite,  $\exists \mathbf{U}$  unitary such that  $\mathbf{Q} = \mathbf{U} \mathbf{D} \mathbf{U}^*$  where  $\mathbf{D}$  is non-negative and diagonal. Therefore the mutual information can be expressed as

$$\begin{aligned} \mathcal{I}(\underline{\mathbf{x}}; (\underline{\mathbf{y}}, \underline{\mathbf{H}})) &= E_{\underline{\mathbf{H}}}[\log_2 \det(\mathbf{I}_{M_R} + \underline{\mathbf{H}} \mathbf{Q} \underline{\mathbf{H}}^*)] \\ &= E_{\underline{\mathbf{H}}}[\log_2 \det(\mathbf{I}_{M_R} + (\underline{\mathbf{H}} \mathbf{U}) \mathbf{D} (\underline{\mathbf{H}} \mathbf{U})^*)]. \end{aligned} \quad (\text{II.27})$$

**Theorem 5.** *Suppose  $\underline{\mathbf{H}}$  is a complex Gaussian i.i.d. matrix, each entry with zero-mean and equal variance. The distribution of  $\underline{\mathbf{H}}$  is the same as the distribution of  $\mathbf{U} \underline{\mathbf{H}} \mathbf{V}^*$  for any unitary matrices  $\mathbf{U}$  and  $\mathbf{V}$ .*

The theorem is proved in [12]. Combining the theorem with (II.27) leads to

$$\mathcal{I}(\underline{\mathbf{x}}; (\underline{\mathbf{y}}, \underline{\mathbf{H}})) = E_{\underline{\mathbf{H}}}[\log_2 \det(\mathbf{I}_{M_R} + (\underline{\mathbf{H}})\mathbf{D}(\underline{\mathbf{H}})^*)]. \quad (\text{II.28})$$

Therefore the optimal  $\mathbf{Q}$  is non-negative diagonal. Given any permutation matrix  $\mathbf{\Pi}$ , consider  $\mathbf{Q}^{\mathbf{\Pi}} = \mathbf{\Pi}\mathbf{Q}\mathbf{\Pi}^*$ . Since  $\underline{\mathbf{H}}\mathbf{\Pi}$  has the same distribution as  $\underline{\mathbf{H}}$ ,

$$E_{\underline{\mathbf{H}}}[\log_2 \det(\mathbf{I}_{M_R} + (\underline{\mathbf{H}})\mathbf{Q}(\underline{\mathbf{H}})^*)] = E_{\underline{\mathbf{H}}}[\log_2 \det(\mathbf{I}_{M_R} + \underline{\mathbf{H}}\mathbf{Q}^{\mathbf{\Pi}}\underline{\mathbf{H}}^*)]. \quad (\text{II.29})$$

For any  $\underline{\mathbf{H}}$ , if  $\mathbf{Q}$  is positive definite, then  $\mathbf{I}_{M_R} + \underline{\mathbf{H}}\mathbf{Q}\underline{\mathbf{H}}^*$  is positive definite, and  $\log_2 \det(\cdot)$  is concave on the set of positive definite matrices. Thus, defining

$$\bar{\mathbf{Q}} = \frac{1}{M_T!} \sum_{\mathbf{\Pi}} \mathbf{Q}^{\mathbf{\Pi}} \quad (\text{II.30})$$

gives us

$$E_{\underline{\mathbf{H}}}[\log_2 \det(\mathbf{I}_{M_R} + \underline{\mathbf{H}}\bar{\mathbf{Q}}\underline{\mathbf{H}}^*)] \geq E_{\underline{\mathbf{H}}}[\log_2 \det(\mathbf{I}_{M_R} + \underline{\mathbf{H}}\mathbf{Q}\underline{\mathbf{H}}^*)], \quad (\text{II.31})$$

and  $\bar{\mathbf{Q}}$  is a multiple of the identity matrix. Therefore, the optimal  $\mathbf{Q}$  must be of the form  $\alpha\mathbf{I}$ . Clearly, the maximum is achieved when  $\alpha$  is the largest possible, i.e.  $\alpha = P/M_T$  since  $\mathbf{Q}$  is constrained to  $\text{trace}(\mathbf{Q}) \leq P$ . It is interesting to note that  $P$  is the average SNR at each receiving antenna.

The capacity of the channel is given by

$$C = E_{\underline{\mathbf{H}}}[\log_2 \det(\mathbf{I}_{M_R} + \frac{P}{M_T}\underline{\mathbf{H}}\underline{\mathbf{H}}^*)]. \quad (\text{II.32})$$

An analytical expression of the capacity can be derived by identifying  $\underline{\mathbf{W}} = \underline{\mathbf{H}}\underline{\mathbf{H}}^*$  when  $M_R \leq M_T$  (or  $\underline{\mathbf{W}} = \underline{\mathbf{H}}^*\underline{\mathbf{H}}$  when  $M_R > M_T$ ) as a random matrix following a Wishart distribution [12].

The capacity of the channel is plotted against the SNR in Fig. II.3. Obviously, the capacity of the channel increases with the number of antennas at both ends of the link. The capacity of SISO and MIMO channels is linear, for high SNRs, with the SNR expressed in dB, but the slope of the curve depends on the number of antennas.

The capacity of the channel at 20 dB of SNR is plotted against the number of antennas in Fig. II.4. The capacity of the channel increases linearly with the number of antennas at both ends of the link, i.e. with the minimum of either the number of antennas at the transmitter or the number of antennas at the receiver. This point is made clear by the example of a system where  $M_T$  increases and  $M_R = \lceil M_T/2 \rceil$ . When the system gains an additional antenna at the transmitter, the capacity hardly changes. However, when both

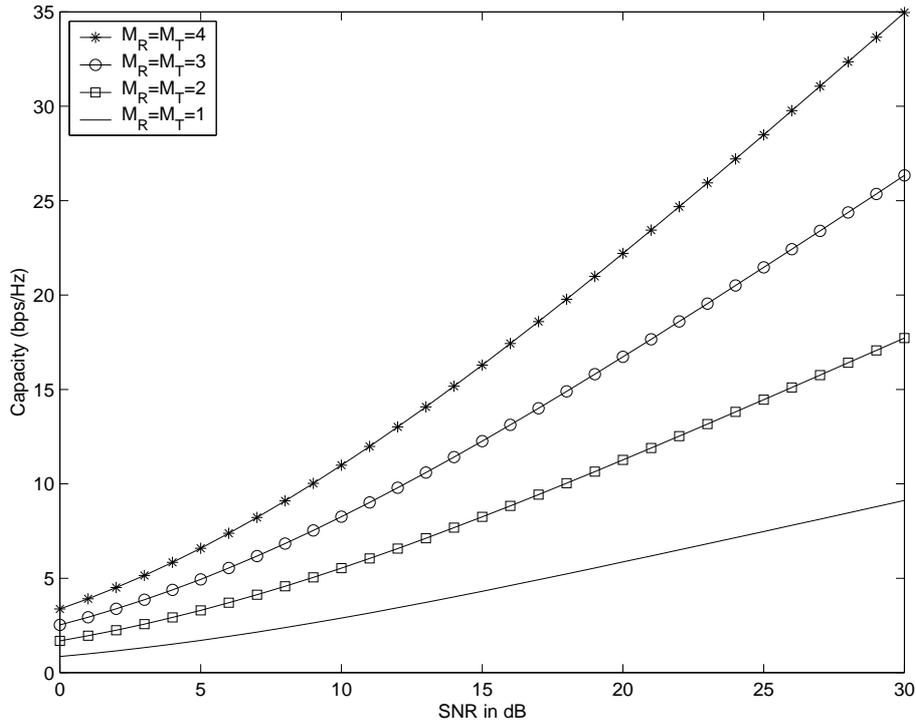


Figure II.3: Capacity of the Rayleigh channel vs SNR, ergodic Rayleigh i.i.d. channel, no CSI at the transmitter

the transmitter and the receiver obtain an additional antenna, the capacity of the system increases significantly.

Note that the channel is not symmetric. The capacity of the channel with  $M_R > M_T$  is always higher than the capacity with  $M_R < M_T$ . This is simply due to the fact that the CSI is available at the receiver but not the transmitter. This point is further discussed in Section III.5.

### II.2.3.2 Non-ergodic Rayleigh i.i.d. channel, no CSI at the transmitter

The mutual information of the channel is given by

$$\mathcal{I}(\underline{x}; \underline{y}) = \log_2 \det(I_{M_R} + \underline{H} \underline{Q} \underline{H}^*), \quad (\text{II.33})$$

and is a random variable.

Therefore, however small the rate we attempt to communicate at, there is a non-zero probability that the realized  $\underline{H}$  is incapable of supporting it, no matter the length of the code applied. A good example is the case where the entries of  $\underline{H}$  are zeros (however this case in itself has a probability of zero).

It is possible to examine the trade-off between the outage probability and the supported

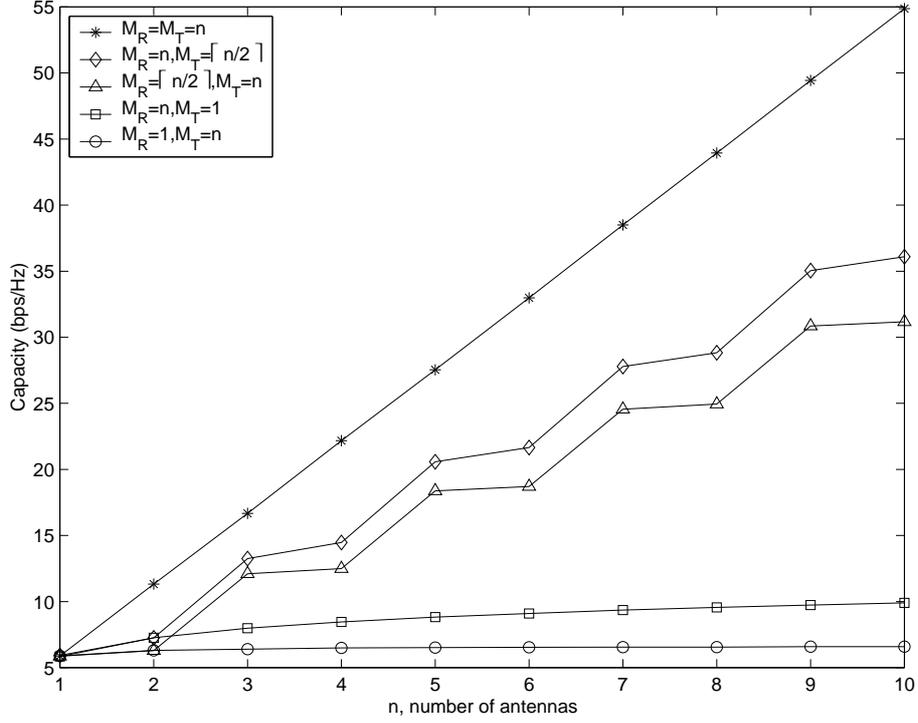


Figure II.4: Capacity of the Rayleigh channel vs number of antennas, SNR=20dB, ergodic Rayleigh i.i.d. channel, no CSI at the transmitter

rate. Namely, given a rate  $R$  and a power  $P$ , one can find  $p_{out}(R, P)$  such that  $R$  is less than the capacity of  $\mathbf{H}$  with total transmitted power  $P$ , except on a set of  $\mathbf{H}$  with probability less than  $p_{out}$ .

From theorem 5, for any unitary matrix  $\mathbf{U}$ ,

$$\log_2 \det(\mathbf{I}_{M_R} + \mathbf{H}\mathbf{Q}\mathbf{H}^*) \quad (\text{II.34})$$

and

$$\log_2 \det(\mathbf{I}_{M_R} + \mathbf{H}\mathbf{U}\mathbf{Q}\mathbf{U}^*\mathbf{H}^*) \quad (\text{II.35})$$

have the same distribution. Hence, it is only necessary to examine diagonal  $\mathbf{Q}$ . The choice of a diagonal  $\mathbf{Q}$  is not only a theoretical simplification, but also corresponds to the simplest hardware implementation possible (the symbols are independent from one antenna to the other).

The symmetry of the problem suggests the following conjecture.

**Conjecture.** *The optimal  $\mathbf{Q}$  is, up to a permutation of the indexes of the antennas, of the form*

$$\frac{P}{M_T^{act}} \text{diag}(\underbrace{1, \dots, 1}_{M_T^{act} \text{ ones}}, \underbrace{0, \dots, 0}_{M_T - M_T^{act} \text{ zeros}}) \quad (\text{II.36})$$

where  $M_T^{act}$  is the number of antennas actually used for transmission (active antennas). The value of  $M_T^{act}$  depends on the rate: the higher the rate (i.e. the higher the outage probability), the smaller the  $M_T^{act}$ .

The ordering of the entries of  $\mathbf{Q}$  has been applied here without loss of generality: it is not an antenna reordering but simply corresponds to a reordering of the columns of  $\mathbf{U}$ .

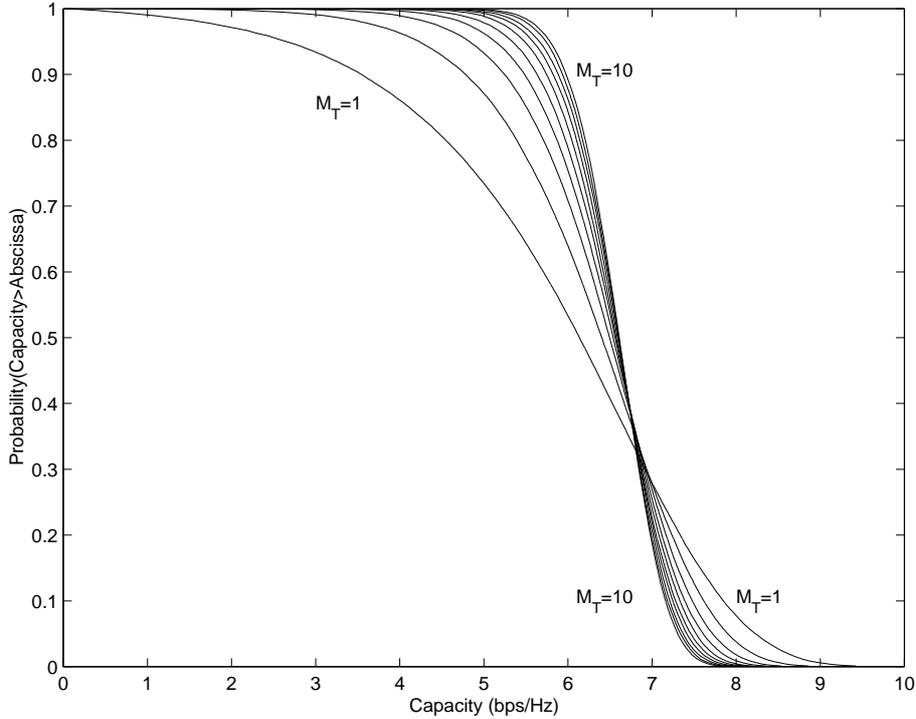


Figure II.5: CCDF of the capacity of the channel for varying  $M_T$ , with  $M_R = 1$  and  $\text{SNR} = 20\text{dB}$

The CCDF of channels with varying number of transmitting antennas and equal power allocation among the transmitter is presented in Fig. II.5. At low outage probability (low data rate), the best channel is the Multiple Input Single Output (MISO) channel. However, at high outage probability (high data rate), the SISO channel has a higher outage capacity than the MISO channel.

These results can be understood intuitively as follows: when the transmitter does not have access to the CSI and the transmission system has only one receiver, transmitting on several antennas is equivalent to transmitting on an average channel, i.e. the average of the individual paths from each transmitting antenna to the receiver. Hence, the behaviour of the system is likely to be more stable and the capacity of the channel does not vary much. In the asymptotic case of an infinite number of antennas, the capacity of the channel is

fixed. This is in contrast with a single transmitting antenna system facing in turn excellent and poor transmission environments.

A theoretical justification follows. The capacity of the channel can be expressed as

$$C(\mathbf{h}) = \log_2\left(1 + \frac{1}{M_T} \left(\sum_{i=1}^{M_T} \mathbf{h}_i \mathbf{h}_i^*\right)\right) \quad (\text{II.37})$$

where  $\mathbf{h} = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{M_T})$ . Obviously  $\underline{\kappa} = \sum_{i=1}^{M_T} \mathbf{h}_i \mathbf{h}_i^*$  is a chi-square variate with  $2 \times M_T$  degrees of freedom (denoted  $\chi_{2M_T}^2$ ) normalized so that  $E[\underline{\kappa}] = 1$ . The PDF of  $\frac{1}{M_T} \underline{\kappa}$  is given by [13]

$$\text{pdf}_{\frac{1}{M_T} \underline{\kappa}}(\kappa) = \frac{M_T}{(\sqrt{0.5})^{2M_T} 2^{M_T} \Gamma(M_T)} (M_T \times \kappa)^{M_T-1} \exp(-M_T \times \kappa), \quad (\text{II.38})$$

where  $\Gamma(\cdot)$  is the gamma function defined as

$$\begin{aligned} \Gamma(r) &= \int_0^\infty t^{r-1} \exp(-t) dt, \quad r > 0 \\ \Gamma(r) &= (r-1)!, \quad r \text{ an integer} > 0 \\ \Gamma(1/2) &= \sqrt{\pi}, \\ \Gamma(3/2) &= \frac{1}{2} \sqrt{\pi}. \end{aligned} \quad (\text{II.39})$$

The PDF of  $\frac{1}{M_T} \underline{\kappa}$  is plotted in Fig. II.6 for varying  $M_T$ .

The analysis is more complicated when the receiver side has more than one antenna. Fig. II.7 presents CCDF of the outage capacity for varying  $M_T$  and  $M_R = 2, M_R = 3, M_R = 4$  and  $M_R = 5$ . For  $M_T < M_R$ , increasing the number of transmitting antennas leads to a significant gain in capacity. However, for  $M_T \geq M_R$ , increasing the number of transmitting antennas only leads to a limited gain in outage capacity for low outage probability. Intuitively,  $M_R$  receiving antennas can separate  $M_R$  spatially independent signals, e.g. independent signals coming from  $M_R$  transmitting antennas. Therefore, for  $M_T < M_R$ , increasing the number of transmitting antennas creates a new spatial orthogonal mode of excitation [14], whereas for  $M_T \geq M_R$ , any new transmitting antenna can only be used to increase the reliability of the existing  $M_R$  channel transmission eigenmodes. These results are discussed in Section III.2 and further justified in the following section.

### II.2.3.3 Non-ergodic Rayleigh i.i.d. channel, fixed power allocation at the transmitter

Most papers in the literature (e.g. [15]) assume that the power allocation at the transmitter is of the form

$$E[\underline{\mathbf{x}} \underline{\mathbf{x}}^*] = \mathbf{Q} = \frac{P}{M_T} \mathbf{I}_{M_T}. \quad (\text{II.40})$$

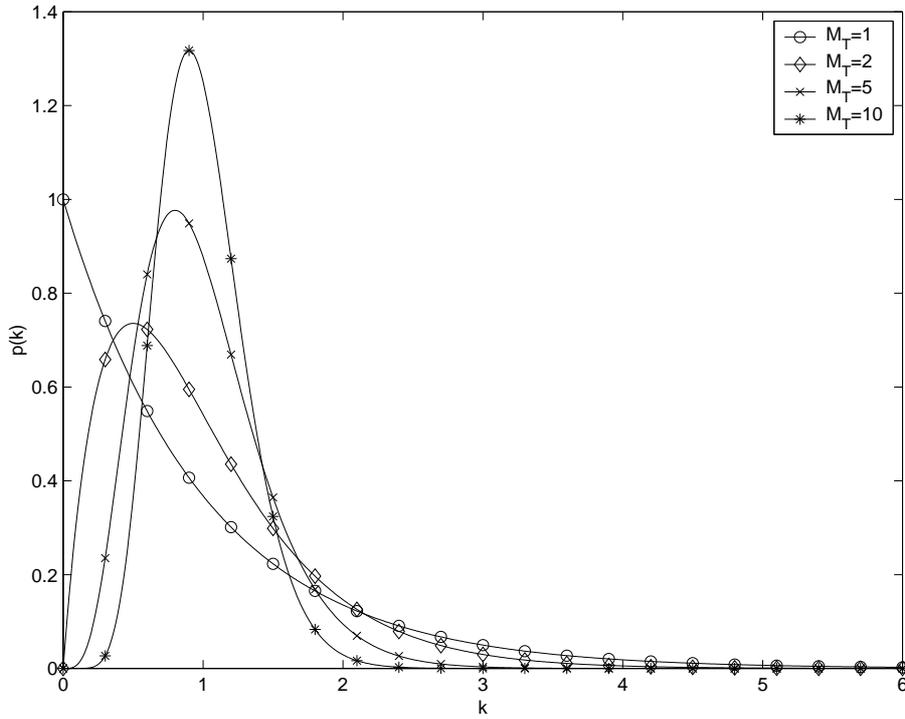


Figure II.6: PDF of  $\frac{1}{M_T} \kappa$  with  $\kappa$  a chi-square random variable with  $2 \times M_T$  degrees of freedom, normalized so that  $E[\chi_2^2] = 1$ .

This assumption is rarely discussed and is assumed "reasonable" or "intuitive".

However, this particular choice of  $\mathbf{Q}$  is not always optimal: in Fig II.5 the capacity of the channel is higher for  $M_T = 1$  than for  $M_T > 1$  for a high outage probability.  $\mathbf{Q}$  can be chosen from the ensemble of diagonal matrices as explained in Section II.2.3.2. However this choice of  $\mathbf{Q}$  can be justified by examining the results of Fig. II.7. For  $M_R > 1$ , the capacity of the MIMO channels increases with  $M_T$ . Allocating more power to a specific antenna, rather than allocating power uniformly on all antenna is an intermediate case where the diversity of the channel is not fully exploited.

Finally, this specific choice of  $\mathbf{Q}$  leads to the simplest MIMO hardware implementation at the transmitter.

In such a case, the capacity is given by

$$C(\mathbf{H}) = \log_2 \det(\mathbf{I}_{M_R} + \frac{P}{M_T} \mathbf{H} \mathbf{H}^*). \quad (\text{II.41})$$

When  $\mathbf{H}$  is Rayleigh and the number of antennas is large, the normalized capacity ( $\frac{C(\mathbf{H})}{\min(M_T, M_R)}$ ) can be approximated by a Gaussian random variable [16]. The importance of this result comes from the fact that this approximation remains accurate, even for a small number of antennas.

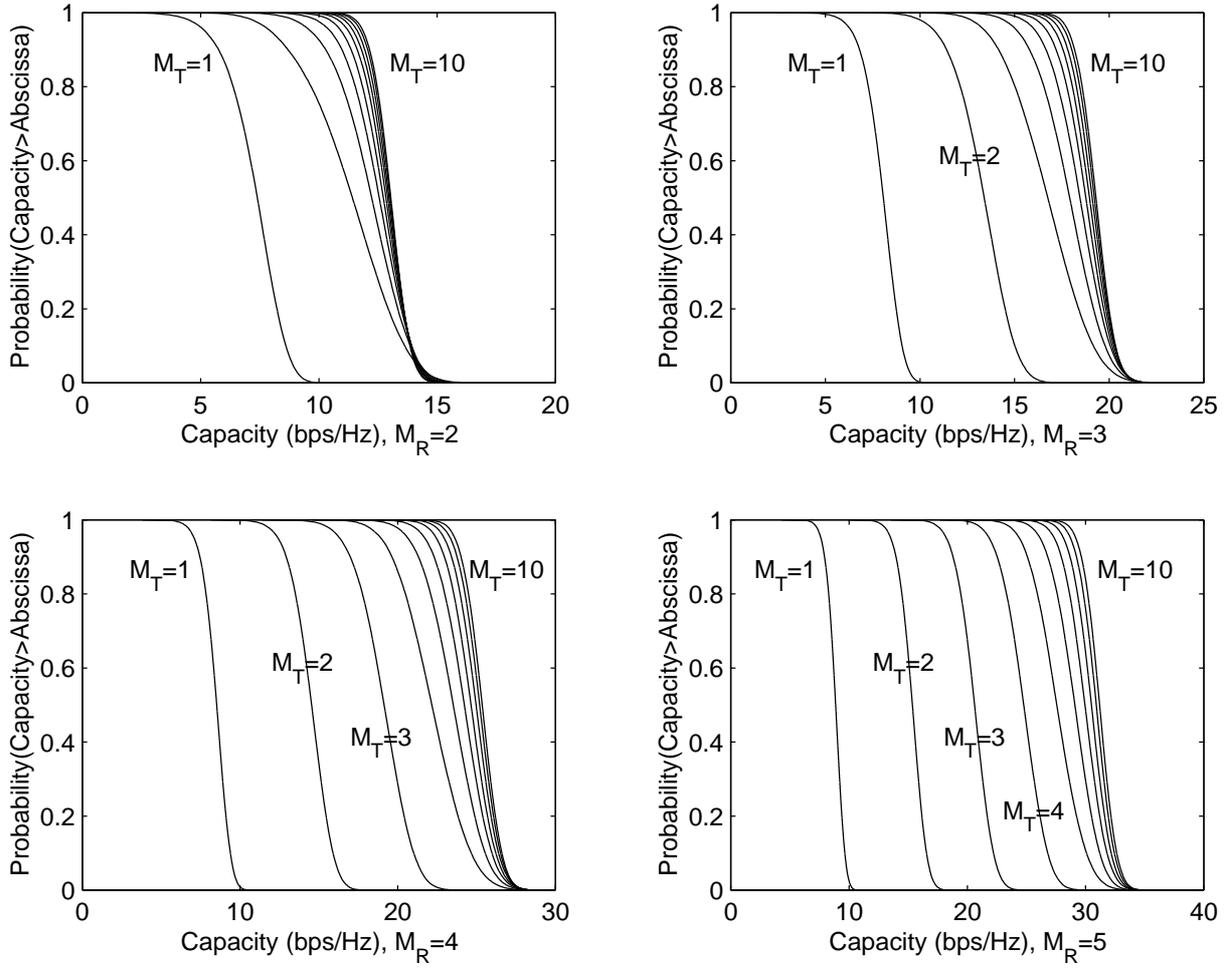


Figure II.7: CCDF of the capacity of the channel for varying  $M_T$ , with  $M_R = \{2, 3, 4, 5\}$  and  $SNR = 20dB$

Suppose  $M_R \rightarrow \infty$ ,  $M_T \rightarrow \infty$  with  $M_T/M_R = \alpha$ , then the mean is given by [17]

$$E[C/\min(M_T, M_R)] = (\log_2(r_+P) + (1 - \alpha) \log_2(1 - r_-) - \frac{r_- \alpha}{\ln 2}) \max(1, 1/\alpha). \quad (\text{II.42})$$

where

$$r_{\pm} \triangleq (r \pm \sqrt{r^2 - 4/\alpha})/2 \quad (\text{II.43})$$

and

$$r \triangleq 1 + \frac{1}{\alpha} + \frac{1}{P}. \quad (\text{II.44})$$

The variance of C is also given in [17] as,

$$\begin{aligned} \sigma_C^2 = & -\log_2 e \log_2 |1 - \alpha \times (\frac{P}{4})^2 \times (1/\alpha - 1 - 1/P + \sqrt{(1/\alpha - 1 - P)^2 + \dots} \\ & 4/(\alpha P))^2 \times (1 - 1/\alpha - 1/P + \sqrt{(1 - 1/\alpha - P)^2 + 4/P^2})|, \end{aligned} \quad (\text{II.45})$$

where  $|\cdot|$  denotes the absolute value.

Fig. II.8, presents the CCDF of the capacity for  $M_T = M_R$ , as well as the corresponding Gaussian approximation. The capacity is well approximated by a Gaussian random

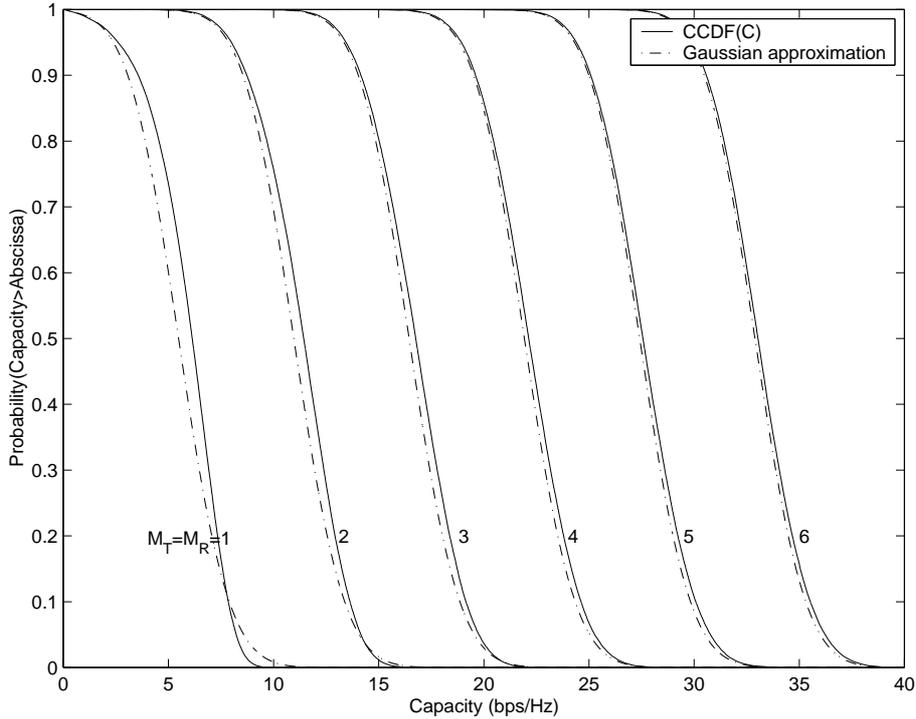


Figure II.8: CCDF of the capacity of the channel for varying  $M_T = M_R$  and SNR= 20dB

variable. The two curves are only easy to distinguish for the special cases  $M_T = M_R = 1$  and  $M_T = M_R = 2$ . The accuracy of the Gaussian approximation increases with the number of antennas.

Fig II.9 and II.10 present both the CCDF of the capacity and its Gaussian approximation to illustrate the accuracy of the model for asymmetrical channels. In [17], it is proven that in the asymptotical case of a large number of antennas, the CCDF becomes Gaussian.

The Gaussian model for the capacity provides us with a good understanding of the behaviour of the capacity for varying parameters  $M_R$ ,  $M_T$  and  $P$ . The effect of an increase in the number of antennas can be understood by examining (II.42) in asymptotic situations.

Telatar proved in [12] that the capacity of the i.i.d. MIMO channel increases linearly with  $\min(M_T, M_R)$  and logarithmically with  $P$ . Novel series expansions of the Gaussian approximation of the mean capacity are presented in the following to highlight the importance of the minimum number of antennas at either the transmitter or the receiver.

Suppose  $M_T \gg M_R$ ,  $\beta = 1/\alpha$  is close to zero and for reasonable variations of  $M_T$  and  $M_R$ ,  $\beta$  remains close to zero. The series expansion of (II.42) with respect to  $\beta$  around the point  $\beta = 0$  is given by:

$$\begin{aligned} E[C/M_R] &= \log_2(P+1) - \frac{1}{2} \times \frac{P^2}{(P+1)^2 \ln(2)} \beta + O(\beta^2) \\ &= \log_2(P+1) + O(\beta). \end{aligned} \quad (\text{II.46})$$

Increasing  $M_R$  leads to a nearly linear gain in capacity. Increasing  $M_T$  only varies the magnitude of the terms factor of  $\beta$  and those terms are negligible. Therefore increasing  $M_T$  leads to a negligible increase in capacity. CCDF of the capacity in this asymptotic situation is given in Fig. II.11.

Suppose  $M_R \gg M_T$ ,  $\alpha$  is close to zero and for reasonable variations of  $M_T$  and  $M_R$ ,  $\alpha$  remains close to zero. The series expansion of (II.42) with respect to  $\alpha$  around the point  $\alpha = 0$  is given by:

$$\begin{aligned} E[C/M_T] &= \log_2(P) - \log_2(\alpha) - \frac{1}{2} \times \frac{P-2}{P \ln(2)} \alpha + O(\alpha^2) \\ &= \log_2(P) - \log_2(M_T) + \log_2(M_R) + O(\alpha). \end{aligned} \quad (\text{II.47})$$

Increasing  $M_R$  only leads to a logarithmic increase in capacity. Furthermore, since  $M_R \gg M_T$ ,  $M_R \gg 1$ . Therefore increasing  $M_R$  leads to a negligible increase in capacity. In-

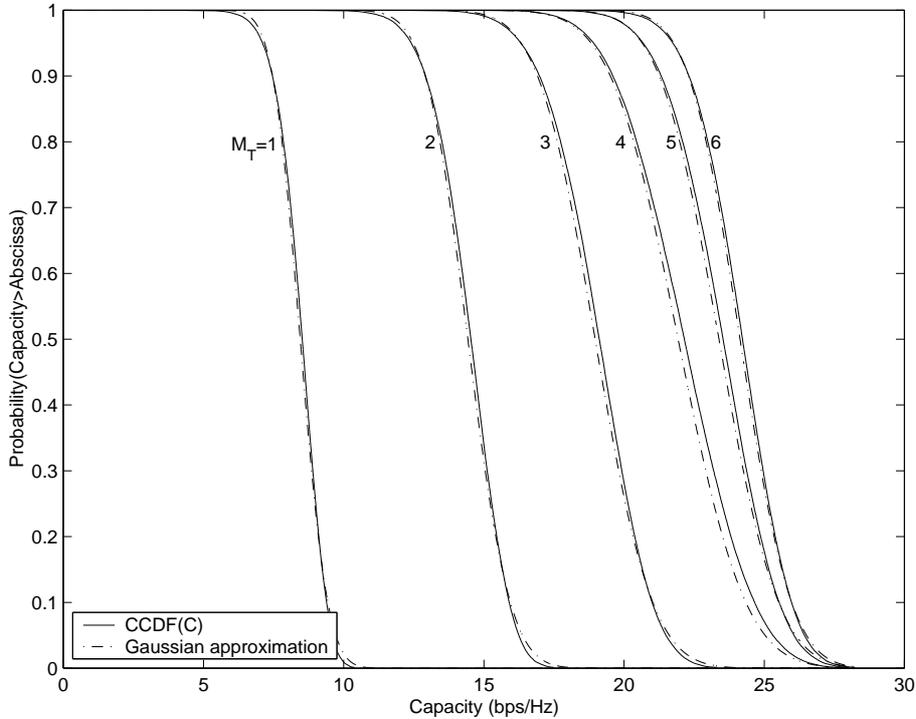


Figure II.9: CCDF of the capacity of the channel for varying  $M_T$ ,  $M_R = 4$  and  $\text{SNR} = 20\text{dB}$

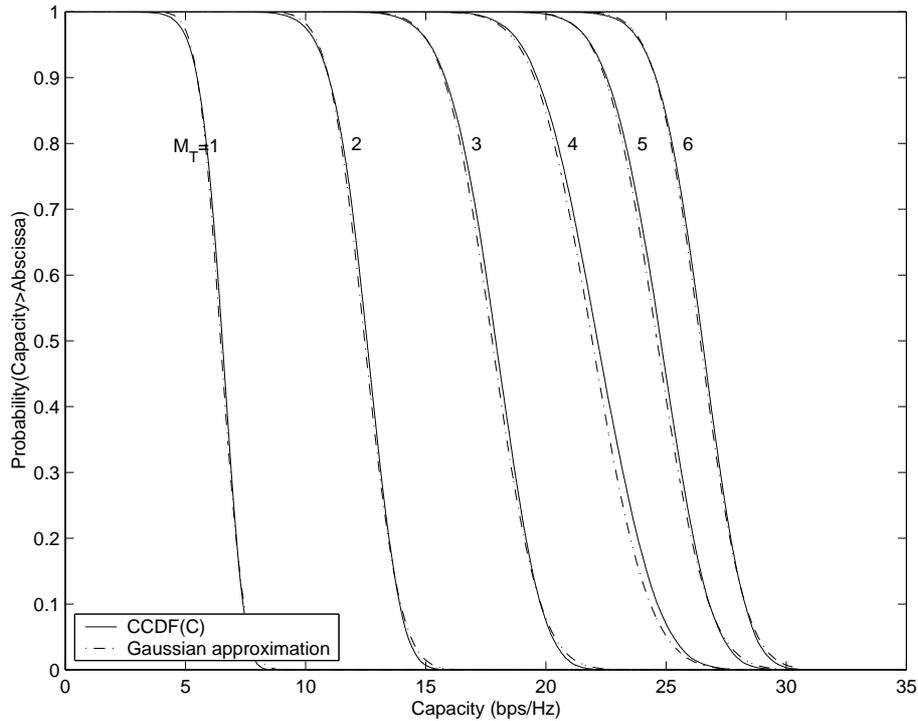


Figure II.10: CCDF of the capacity of the channel for varying  $M_R$ ,  $M_T = 4$  and  $\text{SNR} = 20\text{dB}$

ing  $M_T$  does lead to an increase in capacity (since  $\log_2(M_R) > \log_2(M_T)$ ). The CCDF of the capacity in this asymptotic situation is given in Fig. II.12.

The outage probability might be imposed by requirements of the system, e.g. the Medium Access Control (MAC) layer might require a Frame Error Rate (FER) better than a given constant. If it is not the case, designing the system to allow for maximum throughput is a sensible design criterion. The throughput is the rate of data actually received at the transmitter. Considering that every outage leads to a lost frame, an upper bound of the throughput consists of the product of the attempted data rate multiplied by the outage probability. Fig. II.13 presents the throughput of MIMO systems versus the assumed capacity of the channel.

Fig. II.13 further justifies the choice of a small outage probability for practical telecommunications systems. The maximum throughput of most systems corresponds to a small outage probability, e.g.  $p_{out} = 0.245$  for  $M_R = M_T = 1$  and  $\text{SNR} = 20\text{dB}$ ,  $p_{out} = 0.125$  for  $M_R = M_T = 2$  and  $\text{SNR} = 20\text{dB}$ , and  $p_{out} = 0.028$  for  $M_R = M_T = 5$  and  $\text{SNR} = 20\text{dB}$ .

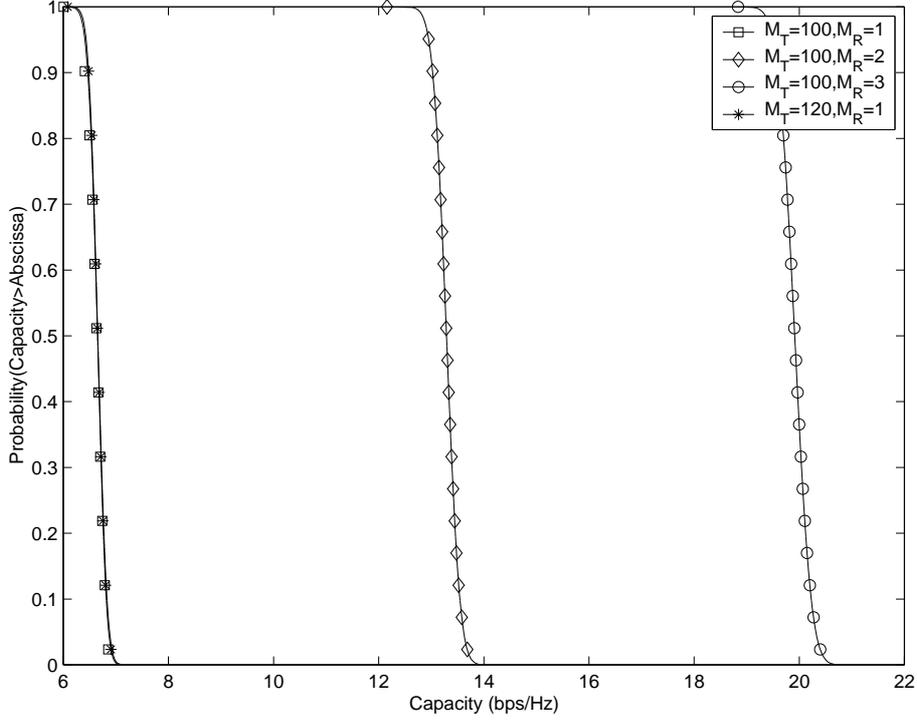


Figure II.11: CCDF of the capacity for varying  $M_R, M_T$ ,  $M_T \gg M_R$  and  $\text{SNR}= 20\text{dB}$

#### II.2.3.4 Non-ergodic channel, CSI at the transmitter

Consider a channel matrix  $\mathbf{H}$  fixed and known at both sides of the link. The mutual information is given by

$$\mathcal{I}(\underline{\mathbf{x}}; \underline{\mathbf{y}}) = \log_2 \det(\mathbf{I}_{M_R} + \mathbf{H}\mathbf{Q}\mathbf{H}^*). \quad (\text{II.48})$$

The SVD of  $\mathbf{H}$  is defined by

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \quad (\text{II.49})$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are unitary and  $\mathbf{\Sigma}$  is diagonal non-negative. Therefore,

$$\mathcal{I}(\underline{\mathbf{x}}; \underline{\mathbf{y}}) = \log_2 \det(\mathbf{I}_{M_R} + \mathbf{\Sigma}\mathbf{V}^*\mathbf{Q}\mathbf{V}\mathbf{\Sigma}^*). \quad (\text{II.50})$$

Observe that  $\tilde{\mathbf{Q}} = \mathbf{V}^*\mathbf{Q}\mathbf{V}$  is non-negative definite if and only if  $\mathbf{Q}$  is non-negative definite, and  $\text{trace}(\mathbf{Q}) = \text{trace}(\tilde{\mathbf{Q}})$ . Hence, the maximization can be carried over  $\tilde{\mathbf{Q}}$ .

Note that for any non-negative definite  $\mathbf{A}$ ,  $\det \mathbf{A} \leq \prod_i \mathbf{A}_{i,i}$ . Hence,

$$\det(\mathbf{I}_{M_R} + \mathbf{\Sigma}\tilde{\mathbf{Q}}\mathbf{\Sigma}^*) \leq \prod_i (1 + \tilde{\mathbf{Q}}_{i,i}\mathbf{\Sigma}_{i,i}^2) \quad (\text{II.51})$$

with equality when  $\tilde{\mathbf{Q}}$  is diagonal. Thus the maximising  $\tilde{\mathbf{Q}}$  is diagonal with entries found by waterfilling power allocation [13]:

$$\tilde{\mathbf{Q}}_{i,i} = (P_{wf} - \mathbf{\Sigma}_{i,i}^{-2})_+, \forall i \in \{1, \dots, M_T\}, \quad (\text{II.52})$$

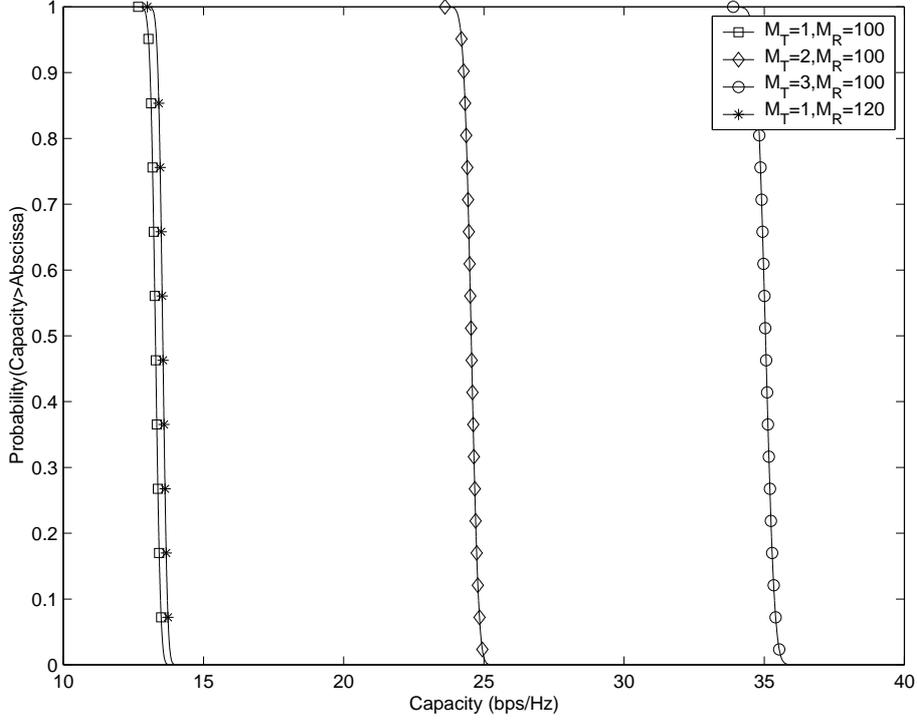


Figure II.12: CCDF of the capacity for varying  $M_T, M_R$ ,  $M_R \gg M_T$  and SNR= 20dB

where  $P_{wf}$  is chosen to satisfy  $\sum_i \tilde{Q}_{i,i} = P$  and  $(\cdot)_+$  indicates that only non-negative values are acceptable. The capacity of the channel is given by

$$C(\mathbf{H}) = \sum_i (\log_2(P_{wf} \Sigma_{i,i}^2))_+. \quad (\text{II.53})$$

The CCDF of the capacity of the channel is given for varying  $M_R = M_T$  with and without CSI at the transmitter in Fig. II.14. The capacity of the channel is higher when the CSI is available at the transmitter. This is easily justified since the transmitter allocates more power to the best spatial transmission eigenmodes, resulting in better transmission. However, the capacity gained by providing CSI to the transmitter is negligible when the channel is i.i.d.. This fact is further discussed in Section III.5.

The CCDF of the capacity of the channel is given for varying  $M_R$  and  $M_T$  in Fig. II.15. The other major difference when CSI is available at the transmitter is that a  $(M_T = i, M_R = j)$  channel has the same capacity as a  $(M_R = i, M_T = j)$  channel,  $\forall i, j \in \mathbb{N}$ .

### II.2.3.5 Ergodic channel, CSI at the transmitter

The capacity of the ergodic channel is simply given by

$$C = E_{\mathbf{H}} \left[ \sum_i (\log_2(P_{wf}(\mathbf{H}) \Sigma_{i,i}^2)) \right], \quad (\text{II.54})$$

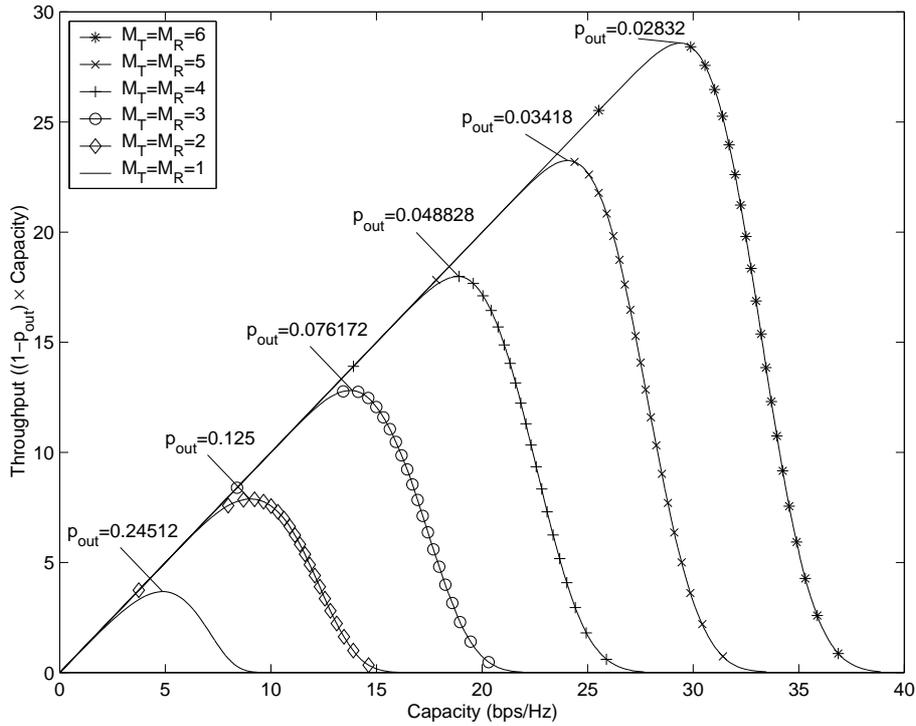


Figure II.13: Maximum throughput of the system for varying  $M_R = M_T$  and SNR= 20dB

where  $P_{wf}(\mathbf{H})$  is the level determined by waterfilling of the singular values of  $\mathbf{H}$  and  $\mathbf{\Sigma}$  is a diagonal matrix with entries the singular values of  $\mathbf{H}$ .

Therefore the capacity of the ergodic channel is simply the mean corresponding to the CCDF of Fig. II.14 and II.15. The capacity of the ergodic channel for varying  $M_R, M_T$  is given in Fig. II.16. Obviously, a  $(M_T = i, M_R = j)$  ergodic channel has the same capacity as a  $(M_R = i, M_T = j)$  ergodic channel,  $\forall i, j \in \mathbb{N}$ .

Finally, at high SNR, the capacity of the ergodic channel is increasing linearly with the SNR expressed in dB, and the slope of the curve depends on  $M_R, M_T$ , as shown in Fig. II.17.

## II.3 Discussion of the assumptions

The assumptions taken in Sections II.1 and II.2 are discussed here, to gain insight into their respective meaning.

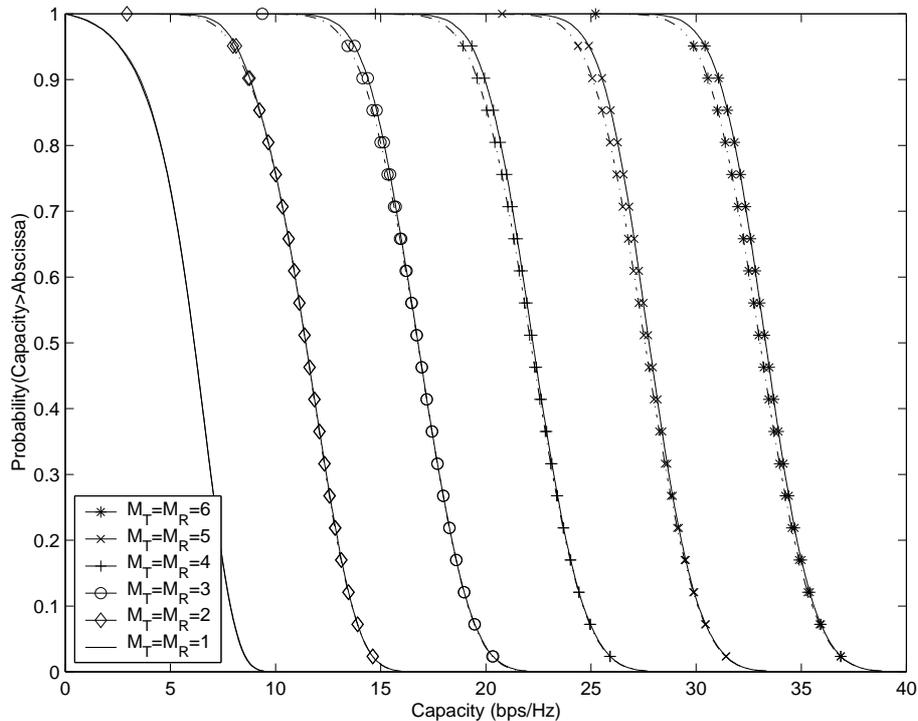


Figure II.14: CCDF of the capacity of the channel with CSI at the transmitter (solid lines) and without CSI at the transmitter (dotted lines) for varying  $M_R = M_T$  and SNR= 20dB

### II.3.1 Coherent detection

The detection is assumed coherent, i.e. the receiver knows the CSI. Measuring the channel at the receiver is actually a complicated task and the accuracy of the CSI estimation is critical in the performance of the system as is further discussed in Section IV.4. Most systems use pilot symbols (known symbols sent by the transmitter) to estimate the channel. Pilot symbols can be inserted at the beginning of the frame (802.11a), in the middle of the frame (GSM) or as a separate signal (IS-95). Blind channel estimation can be applied to avoid the overhead due to pilot symbols.

Coherent detection offers a gain of 3dB in performance over differential detection when the channel estimation is perfect. Non-coherent detection is usually applied in the form of differential coding schemes and unitary space- time codes [18].

Overall, coherent detection is a fitting assumption for most wireless LAN standards (including 802.11a and Hyperlan 2). However, imperfect CSI estimation invariably reduces the performance of the system and so special emphasis must be put on the design of the channel estimation techniques.

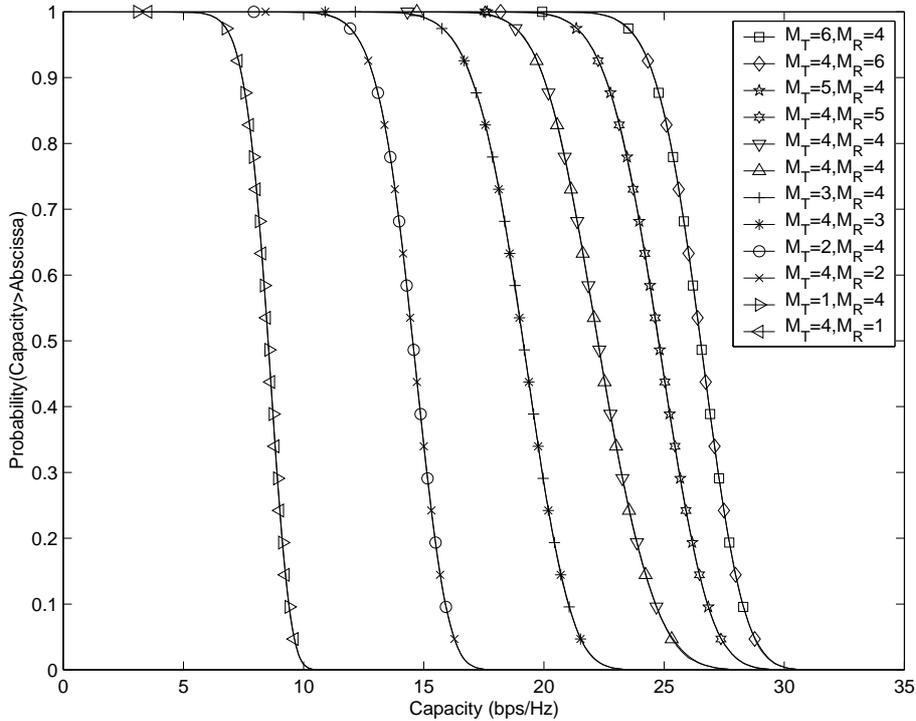


Figure II.15: CCDF of the capacity of the channel with CSI at the transmitter for varying  $M_R$ ,  $M_T$  and SNR= 20dB

### II.3.2 Time and frequency properties of the channel

The three main assumptions about the channel are:

- flat frequency fading,
- slow fading,
- block fading.

The channel is assumed flat over the frequency bandwidth of interest, i.e. the distortion of the signal due to the channel is simply a multiplication by a complex number. It is obvious that the wideband channels used in modern wireless communication are not flat over the whole frequency bandwidth (20 MHz for 802.11a). However, several telecommunications standards, including 802.11a and Hyperlan 2, use Orthogonal Frequency Division Multiplexing (OFDM), a modulation technique which divides the available bandwidth into subchannels of reduced bandwidth. Hence, the assumption of flat-fading is justified in these cases.

The channel is assumed slow fading, i.e. the CSI is constant over the length of a symbol. OFDM modulation increases the length of symbols, placing more stress on the assumption

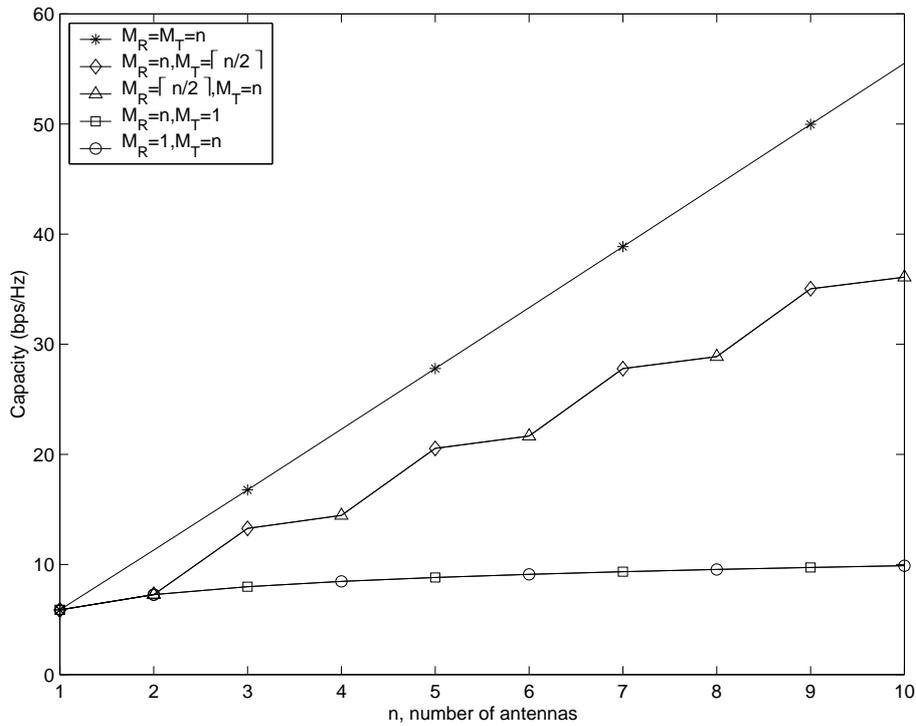


Figure II.16: Capacity of the channel with CSI at the transmitter vs number of antennas, at SNR= 20dB

of slow fading. The combination of the slow fading and coherent demodulation implies that the channel must be really slow fading in time. Specifically, the time coherence of the channel has to be several times longer than a symbol period, since the CSI estimated by the pilot symbols is applied to decode a number of surrounding data symbols, e.g. for the 802.11a standard, the time coherence of the channel should be at least an order of magnitude higher than the  $4\mu\text{s}$  duration of an OFDM symbol. Finally, the transmission model chosen is relevant and applicable to OFDM since OFDM eliminates Inter-Symbol-Interference (ISI).

The channel is assumed block fading, i.e. the channel fading is independent from one block to the next. Though a poor approximation in practice since the channel is continuously fading, most communication devices recover this property of uncorrelated fading on adjacent blocks through the use of interleavers. An interleaver simply permutes the transmitted symbols, while at the receiver the corresponding de-interleaver regenerates the original order. Interleavers allow fading to be independent on consecutive symbols.

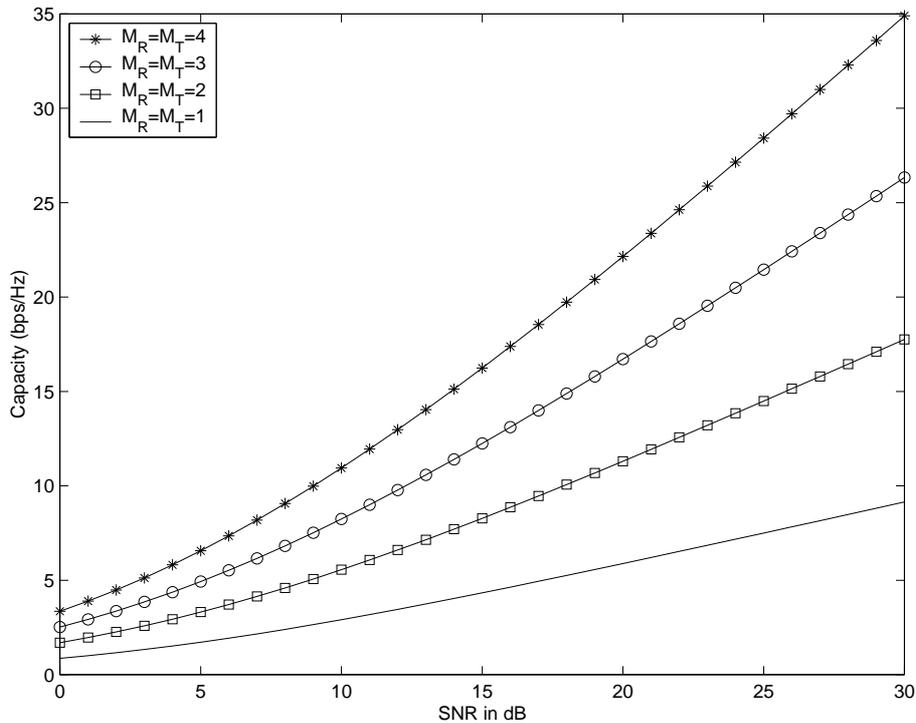


Figure II.17: Capacity of the channel with CSI at the transmitter vs SNR

### II.3.3 Independence of the noise entries

The noise is assumed uncorrelated at the receiver antennas. This is equivalent to discarding all other sources of noise than thermal noise in the receiver front ends. In particular, the assumption cannot account for interference due to other users on adjacent frequency bands, or even on the same band in the case of unlicensed frequency bands such as the Industrial, Scientific and Medical (ISM) band used in the 802.11x and Hyperlan 2 standards. The theoretical capacity of MIMO systems in the interference regime is still an open question.

The distribution of the noise is assumed identical on each receiver antenna. This assumption suggests that the receiver antennas' front ends are supposed identical.

### II.3.4 Single user channel

Throughout Section II.1 single user channels are considered. The capacity of multiuser channels is higher than the capacity of single user channels due to the multiuser diversity, i.e. the ability to allocate the channel to the user facing the best propagation conditions. Insights on the capacity of MIMO multiuser channels are presented in [19]. However, several practical systems, including 802.11a, 802.11g and Hiperlan 2, use time division multiple access, where a single user transmits during a time slot. Therefore, on each time

slot, the relevant capacity corresponds to the capacity of single user MIMO channels.

## II.4 Conclusion

The capacity of a channel is a fundamental concept in telecommunications since it determines how much information can be transmitted. It is possible to increase the capacity of a channel by increasing the number of antennas at the receiver and transmitter. The non-ergodic channel can be studied in term of outage capacity which is well approximated by a Gaussian variable. The ergodic capacity of a Rayleigh i.i.d. channel (with CSI at the receiver only) increases linearly with the minimum number of antennas at the transmitter or at the receiver. These results are demonstrated for flat-fading channels under the assumptions of slow fading, block fading, perfect channel estimation and spatial and temporal independence of the noise at the receivers.



# Chapter III

## Using the large capacity of MIMO channel

Telecommunications theory is concerned with determining the capacity of the channel, designing the transmitted signal  $x$  and reliably recovering the information transmitted in  $x$  from the received signal  $y$ . Chapter II has demonstrated that MIMO channels have a large capacity, but has not indicated how to exploit this large capacity.

Chapter III aims at providing an overview of some well-known techniques to transmit over the MIMO channel and detect the transmitted symbols. Systems with no spatial coding at the transmitter are introduced in Section III.1.1 with the corresponding decoding algorithms. Space-time codes are introduced in Section III.1.2. The performance of uncoded and coded systems are compared in terms of BER in Section III.1.3. The results vary greatly with the simulation parameters and the number of parameters is large. A simple comparison of MIMO schemes is therefore impractical. The concepts of diversity and multiplexing are introduced in Section III.2 to gain insight into the performance of MIMO transmission schemes.

Most theoretical results are only valid when the channel is i.i.d.. Section III.3 presents some new results concerning the capacity of the Ricean channel. These results have been partly accepted for publication [20] and submitted for publication [21]. Section III.3 demonstrates that the ergodic normalized capacity of the Ricean MIMO channel approaches the corresponding normalized capacity of the underlying scattering channel when the antenna numbers are large and there is no CSI available at the transmitter. Theoretical bounds of the normalized capacity of the Ricean channel are proven. These

bounds allow the capacity of the Ricean channel to be estimated in the asymptotic limit of a large number of antennas without recourse to simulation.

BER simulations in Section III.4 demonstrate that uncoded MIMO transmission techniques are badly designed for highly specular MIMO channels. On the contrary, space-time codes can be robust to channel correlation. However, most space-time codes do not benefit from spatial multiplexing and therefore need to use high order constellations to achieve a non-negligible portion of the channel capacity when the number of antennas grows large.

The shortcomings of both uncoded and coded MIMO transmission schemes are partially due to the fact that the CSI is not available at the transmitter. CSI at the transmitter increases the capacity of the MIMO channel and is especially beneficial when the channel is highly specular as demonstrated in Section III.5. The capacity gain when providing the CSI at the transmitter grows linearly with the number of antennas.

## III.1 Review of MIMO transmission techniques

The large capacity offered by MIMO channels triggered numerous attempts to design transmission techniques using several antennas. Some of the most widespread techniques are presented in this section.

### III.1.1 Uncoded systems

The simplest method to transmit over the MIMO channel is to send uncorrelated data streams to each transmitter with equal power. The resulting transmission equation is (II.20) with  $E[\mathbf{x}\mathbf{x}^*] = P/M_T \times \mathbf{I}_{M_T}$ . The transmitter does not require the CSI and the transmission architecture can easily be implemented: the MIMO encoding simply consists of a serial to parallel converter, as presented in Fig. III.1. The decoder uses a MIMO detection block either to separate the transmitted symbols (which are then detected by SISO detectors) or to fully detect the transmitted symbols. The MIMO detection block is a critical component of the transmission system and is likely to be difficult to implement due to its high complexity. MIMO uncoded systems suffer from the disadvantage that  $M_R$  has to be greater than or equal to  $M_T$  for linear MIMO detection schemes to operate. This point is further discussed in Section III.2. Though other detection schemes, e.g. the maximum likelihood scheme presented in Section III.1.1.1, are applicable even for  $M_T$

greater than  $M_R$ , their complexity is usually higher than the complexity of linear detection schemes.

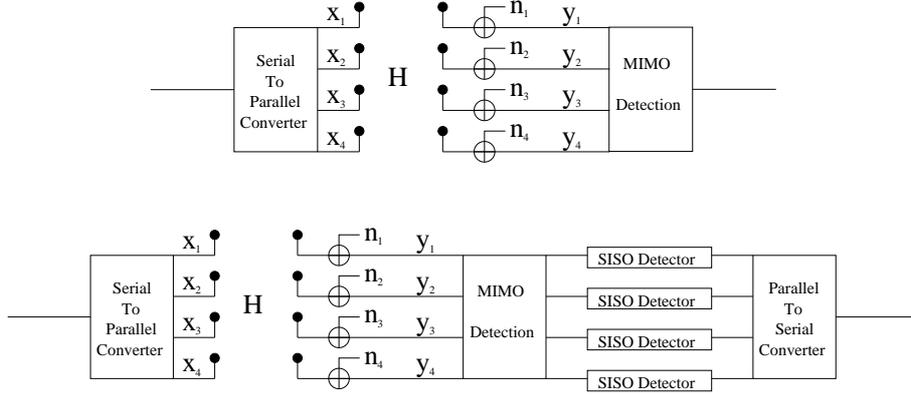


Figure III.1: Block diagrams of the two possible uncoded MIMO transmission architectures.

Several classical MIMO detection schemes for uncoded transmission are presented in the following. All of them assume perfect CSI at the receiver.

### III.1.1.1 Maximum-likelihood detection

The MIMO detection algorithms' role is to detect the transmitted symbols on each of the transmitting antennas. To this effect, the decoder not only knows the channel, but also the constellation used at each transmitting antenna  $\mathcal{S} = \{s_1, s_2, \dots, s_k\}$  for a constellation transmitting  $\log_2(k)$  bits per symbol. Therefore, the transmitter knows the set of possible transmitted symbols, i.e. the MIMO constellation

$$\mathcal{X} = \left\{ \mathbf{x}_1 = \begin{pmatrix} s_1 \\ s_1 \\ \vdots \\ s_1 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} s_2 \\ s_1 \\ \vdots \\ s_1 \end{pmatrix}, \dots, \mathbf{x}_{k^{M_T}} = \begin{pmatrix} s_k \\ s_k \\ \vdots \\ s_k \end{pmatrix} \right\}. \quad (\text{III.1})$$

Note that the cardinality of the MIMO constellation  $\mathcal{X}$  is  $k^{M_T}$  if the constituent constellation  $\mathcal{S}$  is of size  $k$ . The receiver can determine the set of noiseless received symbols  $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{k^{M_T}}\} = \{\mathbf{H}\mathbf{x}_1, \mathbf{H}\mathbf{x}_2, \dots, \mathbf{H}\mathbf{x}_{k^{M_T}}\}$  and deduce the probability

$$\begin{aligned} \forall i \quad \mathcal{P}(\mathbf{x}_i | \mathbf{H}, \mathbf{y}) &= \mathcal{P}(\underline{\mathbf{n}} = \mathbf{y} - \mathbf{y}_i) \\ &= \frac{1}{\pi^{M_R}} \exp(-(\mathbf{y} - \mathbf{y}_i)^*(\mathbf{y} - \mathbf{y}_i)) \end{aligned} \quad (\text{III.2})$$

since  $\underline{\mathbf{n}}$  is AWGN with variance one and mean zero on each receiving antenna.

Therefore it is straightforward to determine the likelihood of each possible transmitted symbol  $\mathbf{x}_i$  and to select the most likely (hard decision) or to send forward the probability of each symbol (soft decision). The Maximum Likelihood (ML) detector is the best detector, under the assumption of perfect channel estimation. However, the MIMO constellation rapidly grows in size with the number of antennas and the size of the component constellation. Therefore the complexity of the ML detector becomes prohibitive. For each received symbol, the ML detector has to obtain  $k^{M_T}$  distances and select the smallest. Additionally, for each frame, the ML detector has to determine the set of noiseless received symbols  $\{\mathbf{y}_i\}$ , i.e. compute  $k^{M_T} (M_R, M_T) \times (M_T, 1)$  complex matrices multiplications.

Alternative algorithms have been proposed to perform ML detection with reduced complexity, e.g. the sphere decoder [22], [23], [24].

### III.1.1.2 Linear detection

A linear detector obtains an estimate of the transmitted symbols by a linear combination of the received symbols. Specifically, the detector obtains

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{y} + \mathbf{B}, \quad (\text{III.3})$$

where  $\hat{\mathbf{x}}$  is an estimate of  $\mathbf{x}$  and  $\mathbf{A}$  and  $\mathbf{B}$  are two matrices assumed fixed with  $\mathbf{H}$ . The main advantage of linear detectors is their reasonable complexity. The complexity can be separated into the computation of  $\mathbf{A}$  and  $\mathbf{B}$  on a frame by frame basis (i.e. for each realisation of  $\mathbf{H}$ ) and the MIMO detection on a symbol per symbol basis. The complexity of the computation of  $\mathbf{A}$  and  $\mathbf{B}$  varies depending on the specific linear receiver considered. The complexity of the MIMO detection is low, i.e. a matrix multiplication and a matrix addition. Another advantage of the linear receiver is that the symbols are detected simultaneously, on the contrary of some non-linear techniques. The main disadvantage of the linear receiver is that higher performance can be obtained with non-linear detection.

The choice of  $\mathbf{A}$  and  $\mathbf{B}$  is discussed in the following. The detector aims at estimating the symbols as well as possible, i.e. minimising the expectancy of the power of the error

$$\begin{aligned} E[\|\hat{\mathbf{x}} - \mathbf{x}\|_F^2] &= E[(\mathbf{A}\mathbf{y} + \mathbf{B} - \mathbf{x})^*(\mathbf{A}\mathbf{y} + \mathbf{B} - \mathbf{x})] \\ &= E[(\mathbf{A}\mathbf{y} - \mathbf{x})^*(\mathbf{A}\mathbf{y} - \mathbf{x})] + \dots \\ &\quad E[\mathbf{B}^*(\mathbf{A}\mathbf{y} - \mathbf{x}) + (\mathbf{A}\mathbf{y} - \mathbf{x})^*\mathbf{B}] + E[\mathbf{B}^*\mathbf{B}], \end{aligned} \quad (\text{III.4})$$

where  $\|\cdot\|_F$  denotes the Frobenius norm. The best choice of the additive component is  $\mathbf{B} = \mathbf{0}$ , a matrix of zeros, since  $E[\mathbf{A}\mathbf{y} - \mathbf{x}] = \mathbf{0}$ .

The Zero-Forcing (ZF) linear detector is defined by  $\mathbf{A} = \mathbf{H}^+$ , where  $\cdot^+$  denotes the pseudo-inverse of a matrix. The estimated symbols are

$$\hat{\mathbf{x}} = \mathbf{x} + \mathbf{H}^+ \mathbf{n}. \quad (\text{III.5})$$

The ZF receivers suffer from noise amplification when the channel matrix is ill-conditioned.

The Minimum Mean Squared Error (MMSE) detector aims at minimising the expectancy of the power of the error (mean-squared error). This criteria is equivalent to insuring that the error is orthogonal to each received symbol

$$\begin{aligned} & E[(\mathbf{x} - \hat{\mathbf{x}})\mathbf{y}^*] = 0 \\ \Leftrightarrow & E[\mathbf{x}\mathbf{y}^*] - E[\hat{\mathbf{x}}\mathbf{y}^*] = 0 \\ \Leftrightarrow & E[\mathbf{x}(\mathbf{H}\mathbf{x} + \mathbf{n})^*] - \mathbf{A} \times E[\mathbf{y}\mathbf{y}^*] = 0 \\ \Leftrightarrow & E[\mathbf{x}\mathbf{x}^*]\mathbf{H}^* - \mathbf{A} \times E[(\mathbf{H}\mathbf{x} + \mathbf{n})(\mathbf{H}\mathbf{x} + \mathbf{n})^*] = 0 \\ \Leftrightarrow & \frac{P}{M_T}\mathbf{H}^* - \mathbf{A} \times \mathbf{H}E[\mathbf{x}\mathbf{x}^*]\mathbf{H}^* - \mathbf{A} \times E[\mathbf{n}\mathbf{n}^*] = 0 \\ \Leftrightarrow & \frac{P}{M_T}\mathbf{H}^* - \frac{P}{M_T}\mathbf{A} \times \mathbf{H}\mathbf{H}^* - \mathbf{A} = 0 \\ \Leftrightarrow & \mathbf{A} \times \left(\frac{P}{M_T}\mathbf{H}\mathbf{H}^* + \mathbf{I}_{M_R}\right) = \frac{P}{M_T}\mathbf{H}^* \\ \Leftrightarrow & \mathbf{A} = \frac{P}{M_T}\mathbf{H}^*(\mathbf{I}_{M_R} + \frac{P}{M_T}\mathbf{H}\mathbf{H}^*)^{-1}, \end{aligned} \quad (\text{III.6})$$

where the inverse is justified by the fact that  $\mathbf{I}_{M_R} + \frac{P}{M_T}\mathbf{H}\mathbf{H}^*$  is hermitian positive definite.

### III.1.1.3 Vertical Bell laboratories layered space-time

The Bell laboratories Layered Space-Time (BLAST) architecture encompasses a series of systems proposed by the Bell Laboratories, some including coding at the transmitter such as Diagonal BLAST (D-BLAST) and some using uncorrelated data streams on the transmission antennas (uncoded MIMO system) such as Vertical BLAST (V-BLAST) [25].

The V-BLAST detection algorithm follows the logical steps:

1. deciding to estimate  $\mathbf{x}_i$ , choosing index  $i$  according to some criterion,
2. finding the vector  $\mathbf{a}(i)$  of minimum norm such that  $(\mathbf{a}(i))^*\mathbf{H}_{:,i} = 1$  and  $(\mathbf{a}(i))^*\mathbf{H}_{:,j} = 0 \forall j \neq i$  with  $\mathbf{H}_{:,i}$  the  $i^{\text{th}}$  column of  $\mathbf{H}$ ,
3. estimate  $\hat{\mathbf{x}}_i = ((\mathbf{a}(i))^*\mathbf{y})$ , where  $(\cdot)$  is a slicing (hard decision) operation,
4. subtract the estimated effect of  $\mathbf{x}_i$  from the received symbol, i.e.  $\mathbf{y} = \mathbf{y} - \mathbf{H}\text{Vect}(\hat{\mathbf{x}}_i)$ , where  $\text{Vect}(\hat{\mathbf{x}}_i)$  is the  $M_T \times 1$  vector with  $\hat{\mathbf{x}}_i$  at the  $i^{\text{th}}$  line,

5. reduce the transmission equation to the non-detected symbols, i.e. replace the  $i^{th}$  column of  $\mathbf{H}$  with zeros,
6. reiterate until all symbols are decoded.

The first step determines in which order the symbols are detected. The last symbol to be detected is facing no interference from other symbols if the other symbols have been correctly detected. This is due to the fact that the effect of previously detected symbols is removed from the received signal. Therefore the last symbol is likely to be detected correctly. On the other hand, the first symbol has  $M_T - 1$  other symbols interfering while being detected, making it the most difficult symbol to detect. Furthermore, the correct estimation of the first symbol is critical in the correct operation of the system: if the symbol is not correctly detected, the fourth step of the detection algorithm corresponds to an increase of the noise for the other symbols. The result is an event named error propagation, which usually creates errors in the detection of subsequent symbols. Therefore it is reasonable to select the received symbol with the highest SNR as the first symbol to be detected. This criterion is optimum in maximising the post-detection SNR [26].

The advantage of V-BLAST is that it out-performs linear detectors. The disadvantage of V-BLAST is its complexity. For each frame, the receiver has to compute  $M_T$  pseudo-inverses of matrices of respective sizes  $M_R \times M_T, M_R \times M_T - 1, \dots, M_R \times 1$ . For each symbol, the linear detection of the symbol is followed by a recoding of the symbol and its subtraction from the received vector (step 4). The complexity is large compared with the linear receiver.

#### III.1.1.4 QR detection

The QR detector is based on the QR-decomposition of the channel matrix  $\mathbf{H}$ . Given any complex matrix  $\mathbf{H}$  of dimension  $M_R \times M_T$ , there exist two matrices  $\mathbf{U}$  and  $\mathbf{R}$  of dimensions respectively  $M_R \times M_R$  and  $M_R \times M_T$ , such that  $\mathbf{H} = \mathbf{U}\mathbf{R}$ ,  $\mathbf{U}$  is unitary and  $\mathbf{R}$  is upper triangular with real elements on the diagonal, i.e.  $\mathbf{R}_{i,j} \neq 0 \Rightarrow i \leq j$  and  $\mathbf{R}_{i,i} \in \mathbb{R}$ .

Now consider the partially detected symbols

$$\begin{aligned}
 \mathbf{y}^{pd} &= \mathbf{U}^* \mathbf{y} \\
 &= \mathbf{U}^* (\mathbf{H}\mathbf{x} + \mathbf{n}) \\
 &= \mathbf{R}\mathbf{x} + \mathbf{U}^* \mathbf{n}.
 \end{aligned} \tag{III.7}$$

Since  $\mathbf{U}$  is unitary, so is  $\mathbf{U}^*$  and the distribution of  $\tilde{\mathbf{n}} = \mathbf{U}^* \mathbf{n}$  is identical to the distribution of  $\mathbf{n}$ . The symbol  $\mathbf{x}_{M_T}$  can be detected since  $\mathbf{y}_{M_T}^{pd} = \mathbf{R}_{M_T, M_T} \mathbf{x}_{M_T} + \tilde{\mathbf{n}}_{M_T}$ . Following the same procedure as in the V-Blast detection algorithm, the symbol  $M_T$  is detected and its contribution is removed from the received symbol. Then, the algorithm detects iteratively the symbols  $M_T - 1$  to 1.

The advantage of QR detection is its moderate complexity. The QR decomposition of the channel matrix, computed on a frame by frame basis, is far less complex than the inversion of the channel matrix (inversion of matrices are commonly implemented as iterative applications of the QR algorithm). The complexity on a symbol by symbol basis is modest. The disadvantage of the QR algorithm lies in its modest performance. The QR suffers, as V-BLAST does, from error propagation and the optimal ordering of the decoding (linked to the indexing of the columns of  $\mathbf{H}$ ) is difficult to obtain.

### III.1.1.5 BER of decoding algorithms

Bit Error Rate (BER) simulation results of the detection schemes previously presented are plotted in Fig. III.3.

As expected,

- the ML detector has lowest BER at any SNR,
- the performance of each of the detectors increases with the number of antennas at the receiver,
- the MMSE detector performs better than the ZF detector, but the performance gap between MMSE and ZF receivers reduces when the SNR increases,
- the V-Blast detector outperforms linear receivers at high SNR but not at low SNR (due to error propagation),
- the QR detector is outperformed by the V-blast detector at any SNR.

Unexpected results include:

- at high SNR,  $\log_{10}(BER)$  decreases linearly with the SNR expressed in dB. Furthermore, the asymptotic slope of the ML detector depends only on the number of receiving antennas. The asymptotic slopes of the other detectors depends on the

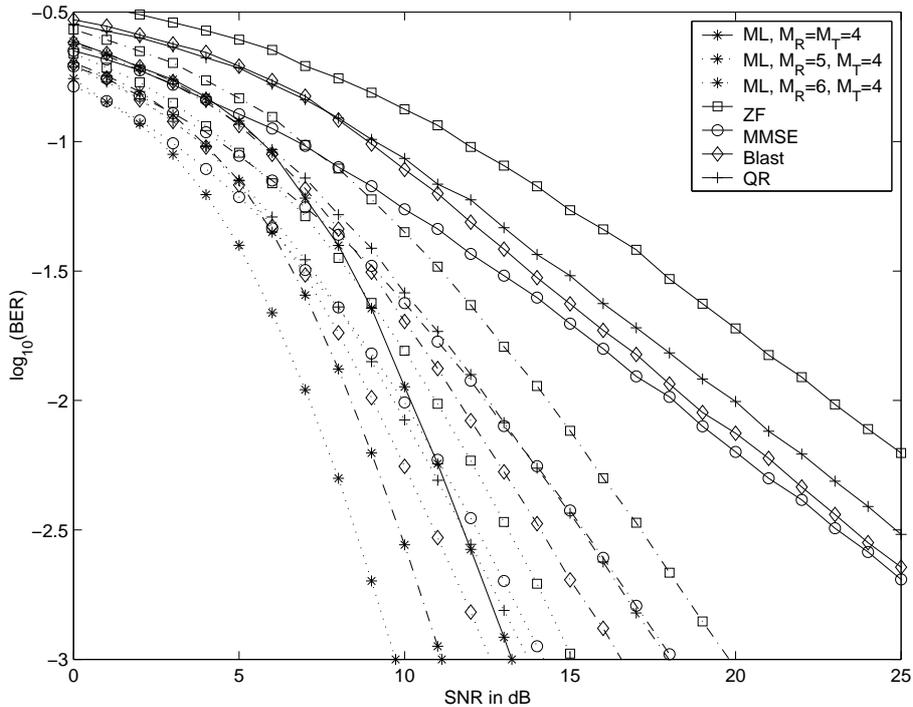


Figure III.2: BER of several decoding algorithms for uncoded MIMO transmission over i.i.d. channel, rate of 8 bits/s/Hz (4-QAM on each transmitting antenna). Legend: plain, dot-dashed and dotted lines denote respectively  $M_R = 4$ ,  $M_R = 5$  and  $M_R = 6$  cases, whereas stars, squares, circles and crosses denote respectively ML, ZF, MMSE, Blast and QR detection algorithms.

difference between the numbers of receiving and transmitting antennas,  $M_R - M_T$ . This point is further discussed in Section III.2.

- The QR detector outperforms linear receivers at high SNR.

### III.1.2 Space-time coding

SISO systems use channel coding to lower the probability of error of received symbols. Transmission errors occur when the channel is in deep fade or when the AWGN is large. The correct transmitted symbol can be recovered if the symbol is transmitted again at a later time. Retransmission, in essence, is channel coding. Channel coding can detect errors or correct them, depending on the code design. In SISO systems, symbols are coded in time. MIMO channels offer another degree of freedom since coding can occur across the antennas. A code operating on MIMO systems combines bits in space and time, and therefore is referred to as a Space-Time (ST) code.

SISO error correction codes are usually binary codes followed by a mapping operation. ST codes are usually not binary: words of ST codes are matrices of symbols (with the rows representing transmit antennas and columns the transmission time).

Similarly to their SISO counterparts, ST block codes are finite in time, whereas ST trellis codes are semi-infinite: transmission has a start but no end. In practice, ST trellis codes are terminated at the end of the transmission. Therefore, ST trellis and block codes mainly differ regarding implementation issues and theoretical analysis, but not in essence.

The transmission of a finite time ST code (length  $L_t$ ) is represented by the transmission equation

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}, \quad (\text{III.8})$$

where  $\mathbf{X}$  and  $\mathbf{N}$  are matrices of size  $M_T \times L_t$  and  $\mathbf{Y}$  is a matrix of size  $M_R \times L_t$ . A simple example of ST block code is the Alamouti scheme, a ST code with  $M_T = 2$  and  $L_t = 2$  [27]

$$\mathbf{X} = \begin{pmatrix} s_0 & s_1 \\ -s_1^* & s_0^* \end{pmatrix}, \quad (\text{III.9})$$

where  $s_0$  and  $s_1$  are the two symbols transmitted by the system. The Alamouti scheme achieves a rate of 1 (1 symbol transmitted per channel usage), and offers a diversity of 2, since every symbol is transmitted over both channels available. The concept of diversity is further discussed in Section III.2.

In coherent detection, the metric considered for the decoding of ST codes is similar to (III.2), i.e.  $\mathcal{P}(\mathbf{X}|\mathbf{H}, \mathbf{Y})$ . However, direct application of a ML decoder is unlikely, due to its extreme complexity. Most space-time codes are designed to obtain a simple decoding algorithm, e.g. the main attraction of the Alamouti scheme remains its very simple decoding process. Both symbols are estimated as

$$\begin{aligned} \hat{s}_0 &= h_0^*y_0 + h_1y_1 = (h_0^*h_0 + h_1h_1^*)s_0 + h_0^*n_0 + h_1n_1^* \\ \hat{s}_1 &= h_1^*y_0 - h_0y_1^* = (h_0^*h_0 + h_1h_1^*)s_1 - h_0n_1^* + h_1^*n_0. \end{aligned} \quad (\text{III.10})$$

Bit Error Rate (BER) simulation results of the Alamouti code are presented in Fig. III.3. The mapping is either 4-Quadrature Amplitude Modulation (QAM, 2 bits/s/Hz) or 16-QAM (4 bits/s/Hz).

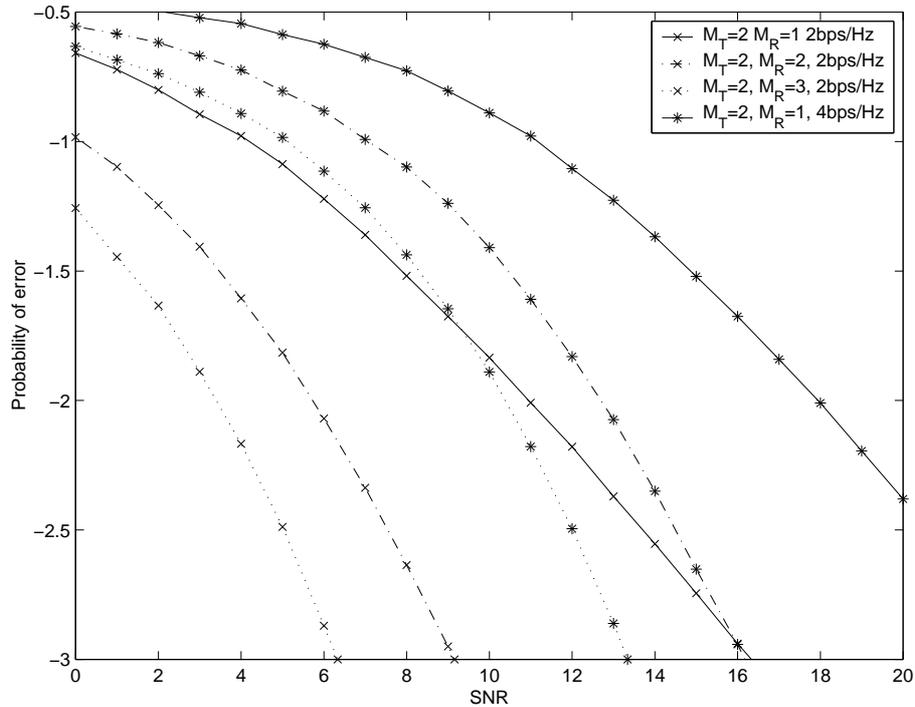


Figure III.3: BER of Alamouti code over i.i.d. channel

### III.1.3 BER simulation of MIMO transmission techniques

Simulation results of MIMO transmission techniques are presented in Fig. III.4. The BER of transmission techniques is plotted for varying SNRs over the i.i.d. Rayleigh fading channel. Parameters include data rate and number of antennas. Perfect channel estimation is assumed.

Fig. III.4 illustrates the fact that it is difficult to compare MIMO transmission techniques. The difficulty is due to the large number of parameters involved. The relative performance of the transmission techniques depends heavily on the choice of those parameters and general results cannot be obtained. A better understanding can be gained through the introduction of the theoretical concepts of multiplexing and diversity. These concepts allow us to predict the effect of a parameter on the BER performance of MIMO transmission systems.

## III.2 Multiplexing vs diversity

As presented in the previous section, popular MIMO transmission systems can be classified as coded or uncoded. This classification reveals a major difference in the way the

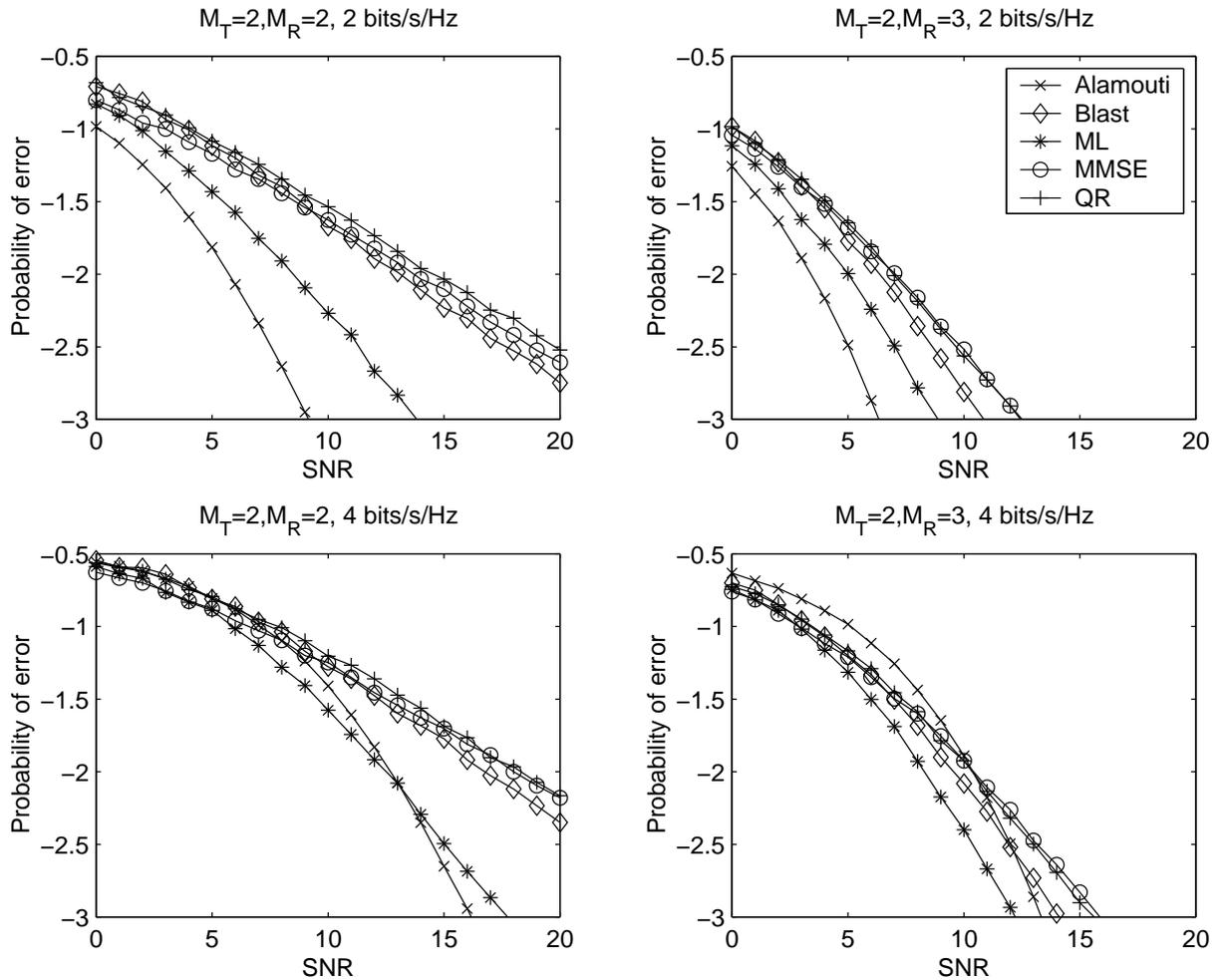


Figure III.4: BER of MIMO transmission techniques over i.i.d. channel

large capacity of MIMO channels is exploited, i.e. using multiple antennas to increase the reliability of the transmission (decrease the probability of error) or to increase the data rate (throughput).

Coded multiple antennas systems increase diversity to combat channel fading. Diversity consists in the transmission of a signal through different means to increase the reliability of the transmission: if one copy of the signal is poorly received, it can be recovered by observing the other copies of the signal. It is well-known that if the fading is independent across antenna pairs (i.i.d. channel assumption), a maximum diversity order of  $M_R \times M_T$  can be achieved, i.e. every symbol can be transmitted through  $M_R \times M_T$  independent SISO channels for every realisation of the MIMO channel [28]. The average error probability of a system achieving a diversity order of  $d$  can be made to decay like  $1/\text{SNR}^d$  at high SNR, which corresponds to a linear variation of the BER with  $\text{SNR}_{\text{dB}}$ , the SNR expressed in dB. The slope of the linear variation is  $-d \times \text{SNR}_{\text{dB}}$  [28]. Most MIMO

systems use reception diversity: each receive antenna obtains a different signal when the receive antennas are uncorrelated. This result is clearly illustrated in Fig. III.2 and Fig. III.4: increasing the number of receiving antennas leads to a direct increase of the slope of the BER curves. As illustrated by Fig. III.2, not all detectors exploit the maximal diversity available: while the ML detector enjoys the maximum diversity available, other detectors lose diversity order. ST-codes further increase the diversity by sending each symbol on several transmit antennas. Therefore coded MIMO systems achieve a higher diversity gain than uncoded systems, justifying the steeper slope of the Alamouti scheme's BER curves (when compared to those of uncoded systems).

Uncoded MIMO systems consider fading as beneficial: fading increases the number of degrees of freedom available for communication since i.i.d. fading increases the probability of a high rank channel matrix (well-conditioned matrix). Uncoded MIMO systems operate at a higher rate than ST-codes since they transmit  $M_T$  symbols every time-slot. Therefore ST-codes need to operate with higher order constellations to transmit at the same data rate. This point is clearly demonstrated by Fig. III.4: to achieve high spectral efficiency (4 bits/s/Hz), the Alamouti scheme is using 16-QAM, whereas uncoded systems achieve the same efficiency with 4-QAM. As a result, the linear portion of the performance curve starts approximately at 6 dB of SNR for uncoded systems but at 9 dB of SNR for the Alamouti scheme.

The Alamouti scheme and uncoded systems are two extremes: the Alamouti scheme achieves maximum diversity with no multiplexing, whereas uncoded systems obtain maximum multiplexing at the expense of diversity. The choice of a diversity order and a multiplexing level is a trade-off and an optimal trade-off curve can be derived, providing the optimal diversity order achievable for each multiplexing level [28].

### III.3 Capacity of correlated channels

Decoding algorithms of uncoded MIMO systems all assume that the channel matrix is well-conditioned. If the matrix is not full rank, the correct reception of the  $M_T$  transmitted symbols is not possible using a linear receiver. This is simply due to the fact that the channel projects a vector space of dimension  $M_T$  into a vector space of dimension  $\mathcal{R}(\mathbf{H})$ , where  $\mathcal{R}(\mathbf{H})$  is the rank of the channel matrix. Any information contained in the null space of the channel matrix cannot be retrieved at the receiver by a linear receiver.

The BER performance degradation (see Section III.4) is not linked to the decoding algorithms. It corresponds to the fact that correlated channels have a lower capacity than their full rank counterparts. A large body of knowledge has been compiled in recent years on the capacity of MIMO correlated channels.

- Several measurement campaigns have been conducted to determine the correlation properties of real-life MIMO channels. The impact and extent of correlation on indoor MIMO channels is discussed in [29, 30, 31, 32]. In [33] it is shown that the MIMO capacity increases as the signal correlation decreases but the SNR has a greater impact than correlation on the MIMO capacity, as shown in [34].
- Several theoretical correlation models have been proposed to simulate correlated channels. Though the most widespread correlation model is arguably the one ring model [35], other models include the exponential correlation matrix [36, 37] and the virtual representation channel matrix [38, 39, 40].
- The effects of fading correlation have been studied for the Rayleigh non-i.i.d. channel through simulation [41] and analysis [35, 42, 43, 44]. Further results investigated the impact of correlation on the variations of the capacity [45]. An upper bound on mean capacity is given in [46, 47, 48].

The effects of correlation on MIMO channels and systems are difficult to analyse due to the complexity of correlation models. Meaningful insight can be gained through the study of a simplified model, such as the Ricean channel model. It is shown in the following that a fully correlated channel has a negligible capacity compared with the capacity of an i.i.d. channel when the number of antennas grows large.

### III.3.1 Ricean channel

The effect of correlation can be easily understood by studying the Ricean channel. In the Ricean case, the flat-fading channel is composed of a Line Of Sight (LOS) component and a Rayleigh component. The choice of the Ricean  $K_{dB}$ -factor (expressed in dB) varies the Ricean channel from a Rayleigh channel ( $K_{dB} \rightarrow -\infty$  dB) to a pure LOS channel ( $K_{dB} \rightarrow +\infty$  dB).

### III.3.1.1 Literature review

Simulation results of the capacity of the Ricean channel are presented in [49]. A geometric approach to interpret the capacity of Ricean channels is described [50]. Simulation indicated that the capacity of the Ricean channel can be approximated by a Gaussian random variable [16]. Finally, analytical results on the capacity of Ricean channel are now emerging for finite numbers of antennas [51] or in the special case of the low power regime [52].

### III.3.1.2 Ricean channel model

In Ricean fading the elements of  $\mathbf{H}$  are non-zero mean complex Gaussians. Hence we can express  $\mathbf{H}$  in matrix notation as [53]

$$\mathbf{H} = \sqrt{\frac{10^{K_{dB}/10}}{1 + 10^{K_{dB}/10}}} \mathbf{H}^{sp} + \sqrt{\frac{1}{1 + 10^{K_{dB}/10}}} \mathbf{H}^{sc} \quad (\text{III.11})$$

where the specular and scattered components of  $\mathbf{H}$  are denoted by superscripts and  $K_{dB}$  is the Ricean  $K$ -factor expressed in dB. The entries of  $\mathbf{H}^{sc} = (h_{i,j})$  are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and unit magnitude variance. It is assumed that  $\mathbf{H}^{sp} = \mathbf{1}$ , where  $\mathbf{1}$  is an  $M_R \times M_T$  matrix of unit entries.

The correlation of the MIMO Ricean channel differs slightly from the widespread idea of correlation because the entries of the channel matrix are not zero-mean Gaussian random variable. The correlation between the entries of the channel matrix is solely due to the non-zero mean of the subchannels.

### III.3.1.3 Pure LOS channel: $K_{dB} \rightarrow +\infty$

In general, a MIMO LOS channel has a capacity of

$$C(K_{dB} = +\infty, M_T, M_R, P) = \log_2(1 + PM_R). \quad (\text{III.12})$$

Since the channel is not random, the capacity is fixed and the ergodic capacity and the capacity are equal. It should be noted that the capacity does not depend on the number of transmit antennas, and only increases logarithmically with the number of receive antennas. In the special case  $M_T = M_R = 1$ , the channel reduces to a Single Input Single Output (SISO) Additive White Gaussian Noise (AWGN) channel.

### III.3.1.4 Pure Rayleigh channel: $K_{dB} \rightarrow -\infty$ , $M_R = 1$ or $M_T = 1$

For the Rayleigh channel,  $|\mathbf{H}_{i,j}|^2$  is a  $\chi^2_2$  variate (chi-squared variate with two degrees of freedom) but normalized so that  $E[|\mathbf{H}_{i,j}|^2] = 1$ . For one transmit antenna, the channel capacity is [15]

$$C(K_{dB} \rightarrow -\infty, M_T = 1, M_R, P) = \log_2(1 + P\chi_{2M_R}^2), \quad (\text{III.13})$$

and using one receive antenna the channel capacity is [15]

$$C(K_{dB} \rightarrow -\infty, M_T, M_R = 1, P) = \log_2(1 + (P/M_T)\chi_{2M_T}^2). \quad (\text{III.14})$$

Notice that

$$\begin{aligned} E[1 + (P/M_T)\chi_{2M_T}^2] &= (1 + P) \\ E[1 + P\chi_{2M_R}^2] &= (1 + PM_R), \end{aligned} \quad (\text{III.15})$$

and  $\log_2(\cdot)$  is a convex function, that is  $\forall r > 0$ ,  $E[\log_2(r)] \leq \log_2(E[r])$ . Therefore

$$E[C(K_{dB} \rightarrow -\infty, M_T, M_R = 1, P)] \leq E[C(K_{dB} \rightarrow +\infty, M_T, M_R = 1, P)] \quad (\text{III.16})$$

and

$$E[C(K_{dB} \rightarrow -\infty, M_T = 1, M_R, P)] \leq E[C(K_{dB} \rightarrow +\infty, M_T = 1, M_R, P)], \quad (\text{III.17})$$

Hence, for a Single Input Multiple Output (SIMO) or Multiple Input Single Output (MISO) channel, the ergodic capacity is greater in a LOS case than in a Rayleigh case (see Fig. III.5).

## III.3.2 Capacity bounds for the Ricean channel

For large numbers of antennas, it has been suggested in the literature that the capacity of the Ricean channel tends to the capacity of its Rayleigh component [54] and that the capacity can be upper bounded by the sum of the capacities of the Rayleigh and LOS component matrices [55].

This section studies the limiting case where  $M_R \rightarrow \infty$ ,  $M_T \rightarrow \infty$  and  $M_T/M_R = \alpha$ . Lower and upper bounds are derived for the Ricean channel capacity. The normalized Ricean ergodic capacity is defined as the ergodic capacity of the Ricean channel divided by  $\min(M_T, M_R)$ . Since the Rayleigh capacity grows linearly with  $\min(M_T, M_R)$  and the LOS capacity only grows logarithmically, it is easily deduced that the normalized Ricean

ergodic capacity approaches that of the underlying Rayleigh channel when the number of antennas  $(M_T, M_R)$  grows large. Also, the Ricean ergodic capacity should be greater than that of the underlying Rayleigh channel. Both results are demonstrated in this section. In order to understand the asymptotic capacity behaviour for the Ricean channel it is instructive to study the eigenvalues of  $\mathbf{H}$ .

### III.3.2.1 Towards the capacity of the Ricean channel

While the capacities of LOS and Rayleigh channels are well understood, the capacity of the Ricean channel is not straightforward to study since the capacity is not a linear operator.

To begin, note that

$$\log_2 \left| \mathbf{I}_{M_R} + \frac{P}{M_T} \mathbf{H} \mathbf{H}^* \right| = \log_2 \left| \mathbf{I}_{M_R} + \frac{1}{1 + 10^{K_{dB}/10}} \frac{P}{M_T} \mathbf{H}^{sc} (\mathbf{H}^{sc})^* + \frac{P}{M_T} \mathbf{F} \right|, \quad (\text{III.18})$$

where  $\mathbf{F}$  is the  $M_R \times M_R$  hermitian matrix,

$$\mathbf{F} = \frac{\sqrt{10^{K_{dB}/10}}}{1 + 10^{K_{dB}/10}} (\mathbf{H}^{sc} (\mathbf{H}^{sp})^* + \mathbf{H}^{sp} (\mathbf{H}^{sc})^*) + \frac{10^{K_{dB}/10}}{1 + 10^{K_{dB}/10}} \mathbf{H}^{sp} (\mathbf{H}^{sp})^*. \quad (\text{III.19})$$

From Appendix I.1,  $\mathbf{F}$  is a matrix of maximum rank two, with one negative eigenvalue and one positive eigenvalue. The positive eigenvalue tends to

$$\lambda_1(\mathbf{F}) \rightarrow \frac{10^{K_{dB}/10}}{1 + 10^{K_{dB}/10}} M_R M_T. \quad (\text{III.20})$$

Though the positive eigenvalue of  $\mathbf{F}$  grows quadratically with the number of antennas,  $\mathbf{F}$  has a fixed number of eigenvalues (2) when the number of antennas increases. Therefore,  $\mathbf{F}$  has a negligible effect on the normalized capacity of the Ricean channel when the number of antennas tends to  $\infty$ , since it contributes approximately as  $\frac{\log_2(M_T^2)}{M_T}$ . This insight is demonstrated in the following section.

### III.3.2.2 Asymptotic capacity of a Ricean channel

This section states the main contribution of this chapter: for a Ricean channel that is not pure LOS ( $K_{dB} \neq +\infty$ ), when the number of antennas grows large, the normalized capacity of the Ricean channel tends to the normalized capacity of its scattering component, i.e. when  $M_R, M_T \rightarrow \infty, M_T/M_R = \alpha$ ,

$$E \left[ \frac{C(K_{dB}, M_T, M_R, P)}{\min(M_T, M_R)} \right] \rightarrow E \left[ \frac{C(K_{dB} = -\infty, M_T, M_R, \frac{1}{1+10^{K_{dB}/10}} P)}{\min(M_T, M_R)} \right]. \quad (\text{III.21})$$

**III.3.2.2.1 Lower bound** Appendix I.2 proves that  $\forall M_R, M_T, P, K_{dB}$ ,

$$E \left[ \frac{C(K_{dB}, M_T, M_R, P)}{\min(M_T, M_R)} \right] \geq E \left[ \frac{C(K_{dB} = -\infty, M_T, M_R, \frac{1}{1+10^{K_{dB}/10}} P)}{\min(M_T, M_R)} \right]. \quad (\text{III.22})$$

**III.3.2.2.2 Higher bound** Appendix I.3 proves that as  $M_T, M_R \rightarrow \infty$ ,

$$E \left[ \frac{C(K_{dB}, M_T, M_R, P)}{\min(M_T, M_R)} \right] \leq E \left[ \frac{C(K_{dB} = -\infty, M_T, M_R, \frac{1}{1+10^{K_{dB}/10}} P)}{\min(M_T, M_R)} \right] + \Delta, \quad (\text{III.23})$$

with  $\Delta \rightarrow 0$  as  $M_T, M_R \rightarrow \infty$ .

### III.3.3 Results and discussion

#### III.3.3.1 Mean and variance of capacity for the Ricean channel

Fig. III.5 plots the average normalized capacity with  $\alpha = 1$  for an increasing number of antennas and different  $K$ -factors. As indicated in our analysis, for  $M_T = 1$ , the mean capacity of Ricean channels is higher than the mean capacity of Rayleigh channels. This trend is inverted for  $M_T, M_R > 1$ . This result is in contrast with the outage capacity of the Ricean channel which remains higher than the corresponding capacity of the Rayleigh channel for small number of antenna (i.e. up to four antennas) when the targeted outage probability is very small (i.e. 0.01) [49]. Such a result is not surprising since the capacity of the Ricean channel is a random variable with a smaller variance than the capacity of the Rayleigh channel [16].

For  $K_{dB} = -1000$ , the capacity converges rapidly to a limit as  $M_T \rightarrow \infty$ , as indicated in [17]. For other values of  $K_{dB}$ , the capacity decreases with the number of antennas over this range. For increasing  $M_T$  (as soon as  $M_T \geq 2$ ), the capacity of the Ricean channel is a decreasing function of  $K_{dB}$ . The results shown in Fig. III.5 are for  $M_T = M_R$ .

Fig. III.6 plots the variance of the total capacity for different  $K_{dB}$  values versus the numbers of transmit and receive antennas when SNR= 20 dB and  $\alpha = 1$ . For the purposes of comparison, the corresponding analytical results in [17] are also shown for the Rayleigh channel. From these results it is clear that the Ricean channel conditions do not impact the capacity variance beyond the corresponding Rayleigh values for large antenna conditions.

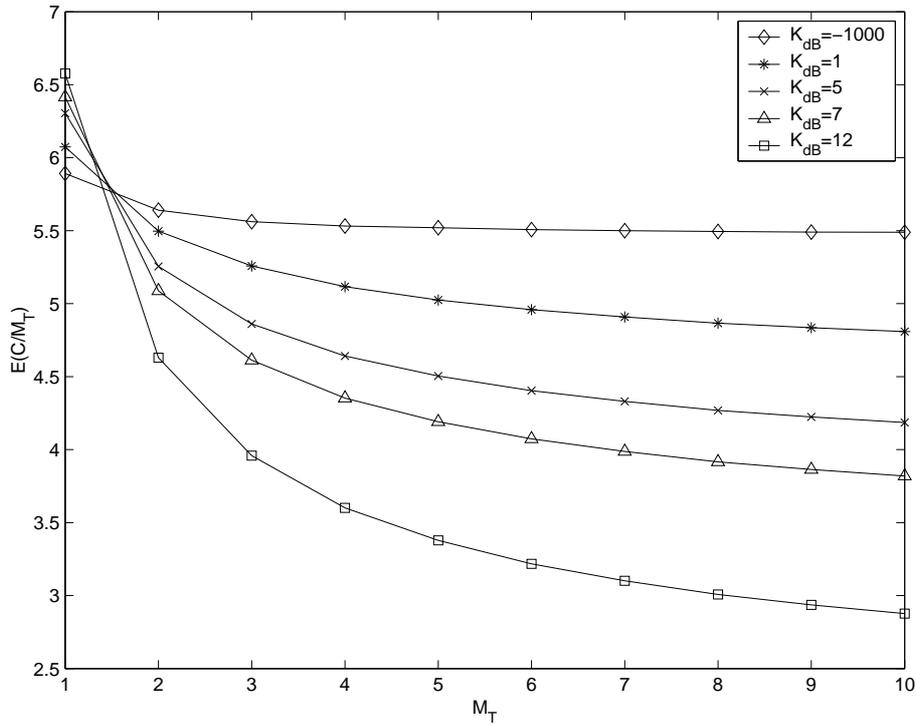


Figure III.5: Normalized mean capacity per antenna with  $M_T/M_R = 1$ , Ricean fading and SNR= 20dB

### III.3.3.2 Asymptotic mean and variance of the capacity for the Ricean channel

Fig. III.7 shows the behaviour of the normalized capacity of a Ricean channel as the number of antennas grows large. For all values of  $K_{dB}$ , the normalized capacity of the Ricean channel tends to the normalized capacity of its scattering component (the lower bound on the capacity). This lower bound is tighter when  $K_{dB}$  is smaller, and for  $K_{dB} = -1000$  it is impossible to discern the simulation from the lower bound.

The upper bound converges slowly to the lower bound and is tight for large values of  $K_{dB}$ . An explanation for the slowness of convergence can be found in (A.31) where it is shown that  $\Delta$  (see Section III.3.2.2.2) tends to zero like  $\log(M_R)/M_T$ , which itself converges very slowly. Although the upper bound is only strictly valid for a large number of antennas (see the assumptions in (A.32)), the simulations show excellent correlation for values of  $M_T$  as low as 20, and for  $K_{dB} \leq 12$ .

The capacity bounds for values of  $\alpha$  other than 1 are shown in Fig. III.8. The tightness of both bounds appears to remain unaffected by the value of  $\alpha$ .

The results shown in Fig. III.6 are further reinforced by Fig. III.9. This shows the

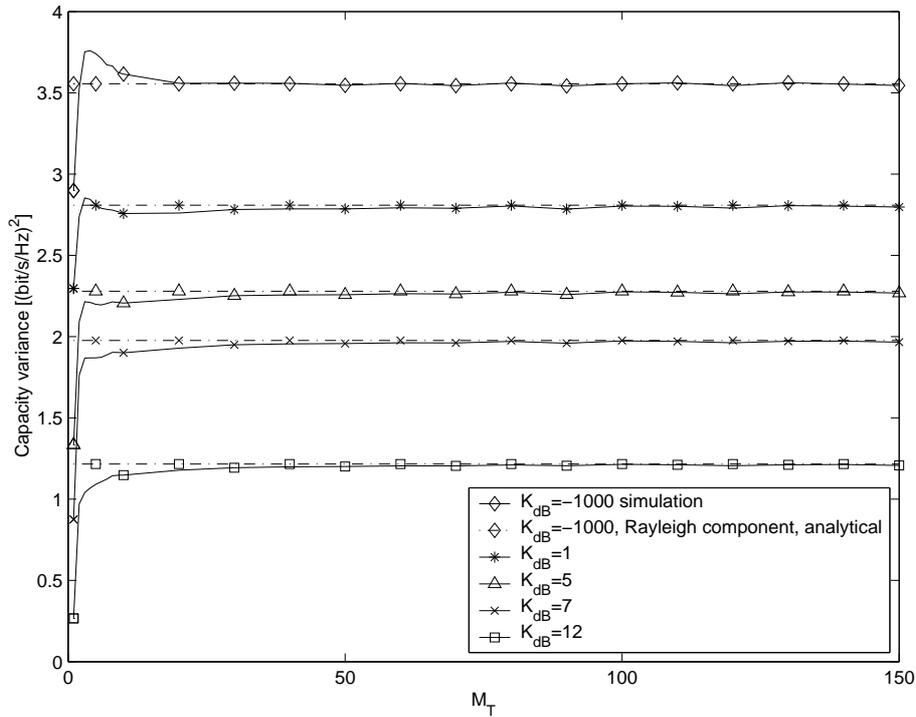


Figure III.6: Capacity variance with  $M_T/M_R = 1$ , Ricean fading and SNR= 20dB

behaviour of the variance of the capacity of a Ricean channel as the number of antennas grows large. For all values of  $K_{dB}$ , the variance clearly tends to the corresponding variance of its scattering component, the variance is maximum for  $\alpha = 1$  and the maximum decreases as the channel approaches LOS condition. We note from results not reported here that this asymptotic behaviour of the capacity variance is only achieved for very large number of antennas  $M_T, M_R(M_T = M_R) = 100$  or more.

### III.3.3.3 Capacity bounds

Fig. III.8 shows the behaviour of the normalized capacity, for varying  $\alpha$ , in the asymptotic case of a large number of antennas ( $\min(M_T, M_R) = 100$ ). As in Fig. III.7, the lower bound is tight for small  $K_{dB}$ , whereas the upper bound is tight for large  $K_{dB}$ . Note that the tightness of the bounds appears uniform across all  $\alpha$  values, indicating that the tightness depends on the ratio  $M_T/M_R$ , and not on their actual values.

The results demonstrate that the upper and lower bounds provide a fast and reliable way to bound the mean capacity of a Ricean channel for large  $\min(M_T, M_R)$ , without extensive simulations. Furthermore, depending on the  $K$ -factor, it is straightforward to deduce which of the bounds is the tightest.

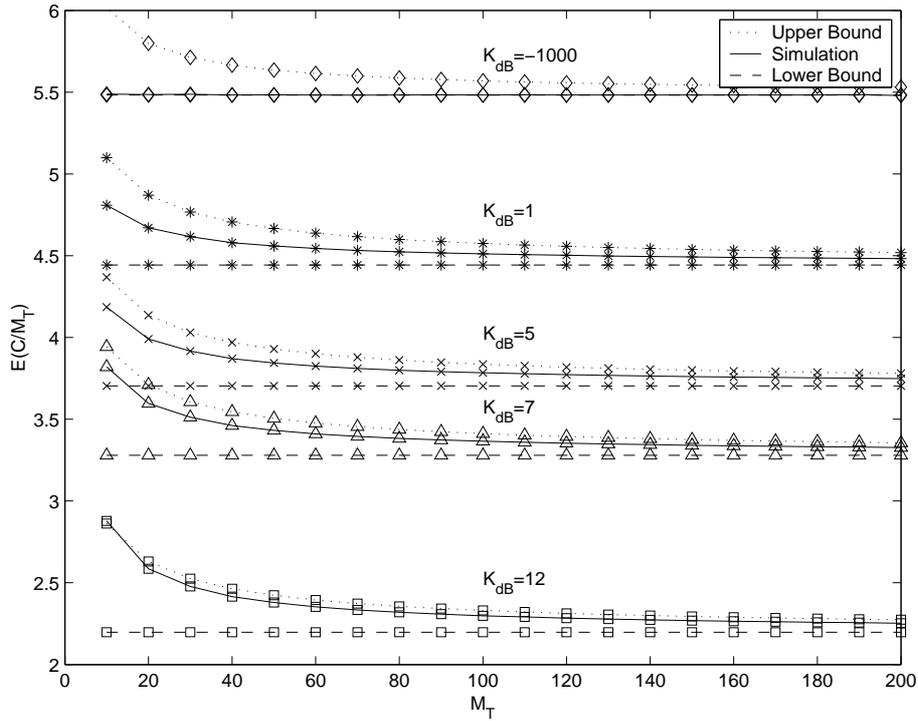


Figure III.7: Ergodic capacity per antenna with  $M_T, M_R \rightarrow \infty$ , Ricean fading and SNR=20dB

### III.3.4 Conclusion

The capacity of the Rayleigh and LOS channels have been studied extensively and are well-known, both for a small number of antennas and in the asymptotic case of a large number of antennas. For a Ricean channel, the capacity is more difficult to derive.

For a large number of antennas, the normalized mean capacity of a Ricean channel tends to the normalized mean capacity of its Rayleigh component. Precisely, the capacity of the Ricean channel is lower bounded by the capacity of its Rayleigh component and upper bounded by a quantity that tends to the capacity of its Rayleigh component as the number of antennas grows large.

The lower bound is valid for any number of antennas, and depending on the choice of a constant  $\min(M_T, M_R)$ , the upper bound can be valid for any number of antennas, or only for a large number of antennas (in which case the upper bound becomes tighter as the number of antennas grows large). The lower bound is tighter when the  $K$ -factor is smaller, whereas the upper bound is tighter with increasing  $K$ . The two bounds allow us to estimate the capacity of a Ricean channel without extensive simulations.

A very similar behaviour is observed for the capacity variance that also tends to the cor-

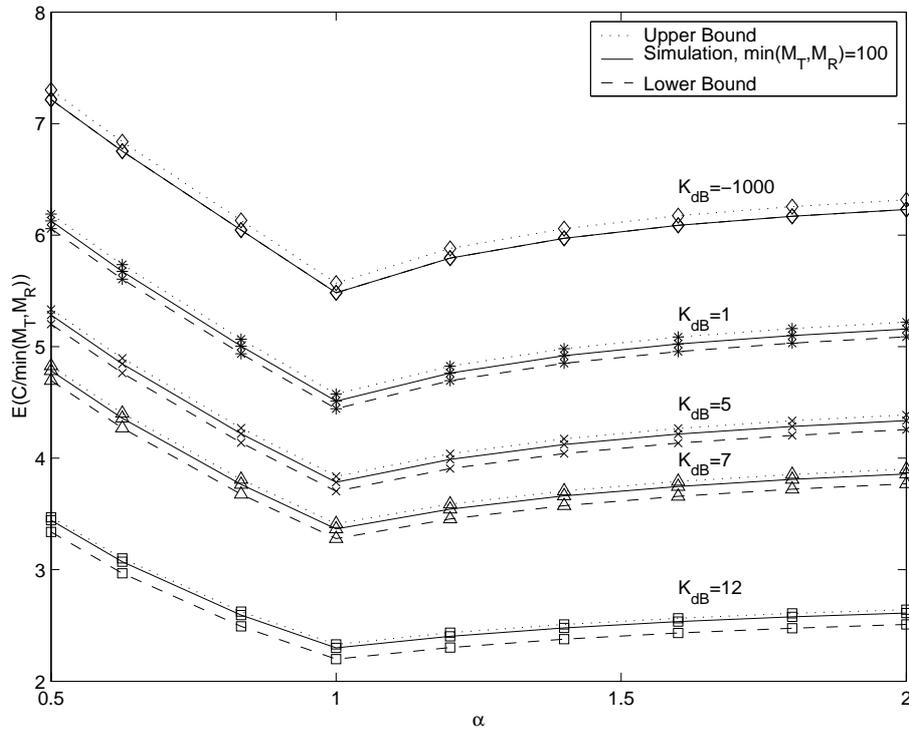


Figure III.8: Ergodic capacity per antenna versus  $\alpha$ , Ricean fading and SNR=20dB

responding variance of the scattering component especially when the number of antennas grows very large.

### III.4 MIMO systems in highly specular channel

As mentioned earlier, decoding algorithms of uncoded MIMO systems typically assume that the channel matrix is well-conditioned. Any system with a MIMO multiplexing gain cannot operate properly over a channel of rank one, since the vector space of the received symbol is of dimension one. Furthermore, the capacity of highly specular MIMO channels is usually lower than the capacity of i.i.d. MIMO channels. The influence of channel correlation on the performance of MIMO systems is further evaluated in this Section through BER simulation results.

Linear detection schemes presented in Section III.1.1 all assume that the channel matrix is full rank. The QR detection algorithm produces a division by zero when the channel matrix is not full rank, whereas other detection schemes can be applied but cannot decode correctly the transmitted signals, regardless of the SNR. It is reasonable to assume that limited correlation leads to a loss of performance for all detection schemes. Fig. III.10

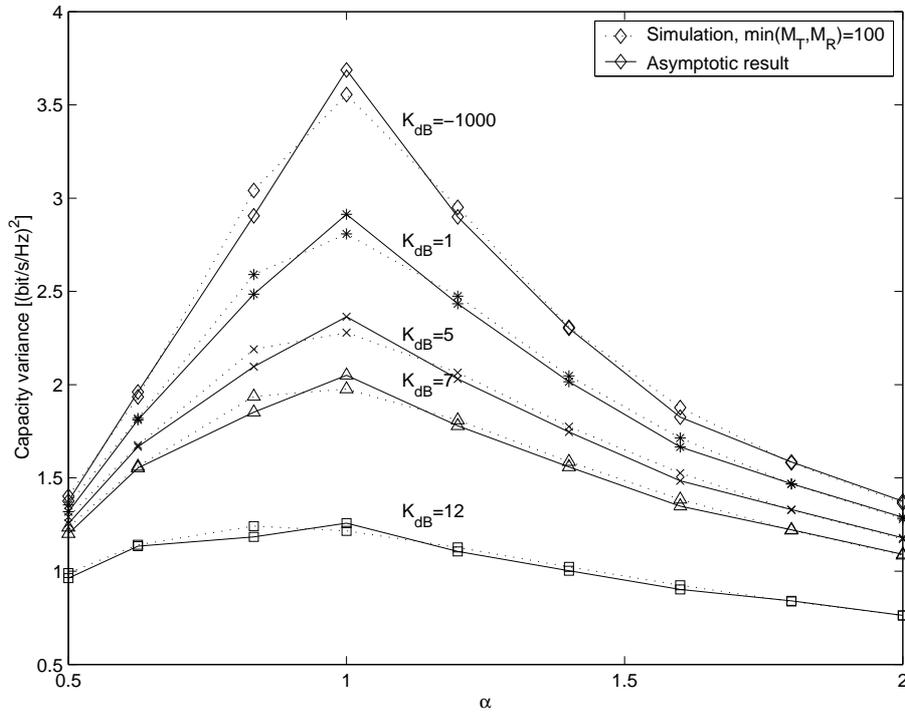


Figure III.9: Asymptotic capacity variance, Ricean fading and SNR= 20dB

presents BER simulation results for several uncoded MIMO systems over the Ricean channel. For all systems, an increase of the  $K$ -factor directly leads to a loss of performance which can be related to the loss of power of the i.i.d. component of the channel matrix.

Fig. III.11 presents BER simulation results of the Alamouti scheme over the Ricean channel. The performance of the Alamouti scheme does not degrade with channel correlation since no spatial multiplexing is used. The two transmitted signals are coded to be orthogonal in space and time, regardless of the rank of the channel. Therefore, increasing the  $K$ -factor of the Ricean channel leads to a constant transmission environment. This explains the lower BER at high SNR: at high SNR, errors only occur when the channel is in deep fade, which is less likely for a highly specular Ricean channel than for an i.i.d. channel.

### III.5 Influence of CSI at the transmitter

As discussed previously, correlated channels have generally a lower capacity than i.i.d. channels. As a result, uncoded schemes suffer from a loss of performance when the channel is highly specular. ST-codes do not suffer from the same loss of performance since they do not take advantage of the multiplexing gain available in i.i.d. MIMO channels. However,

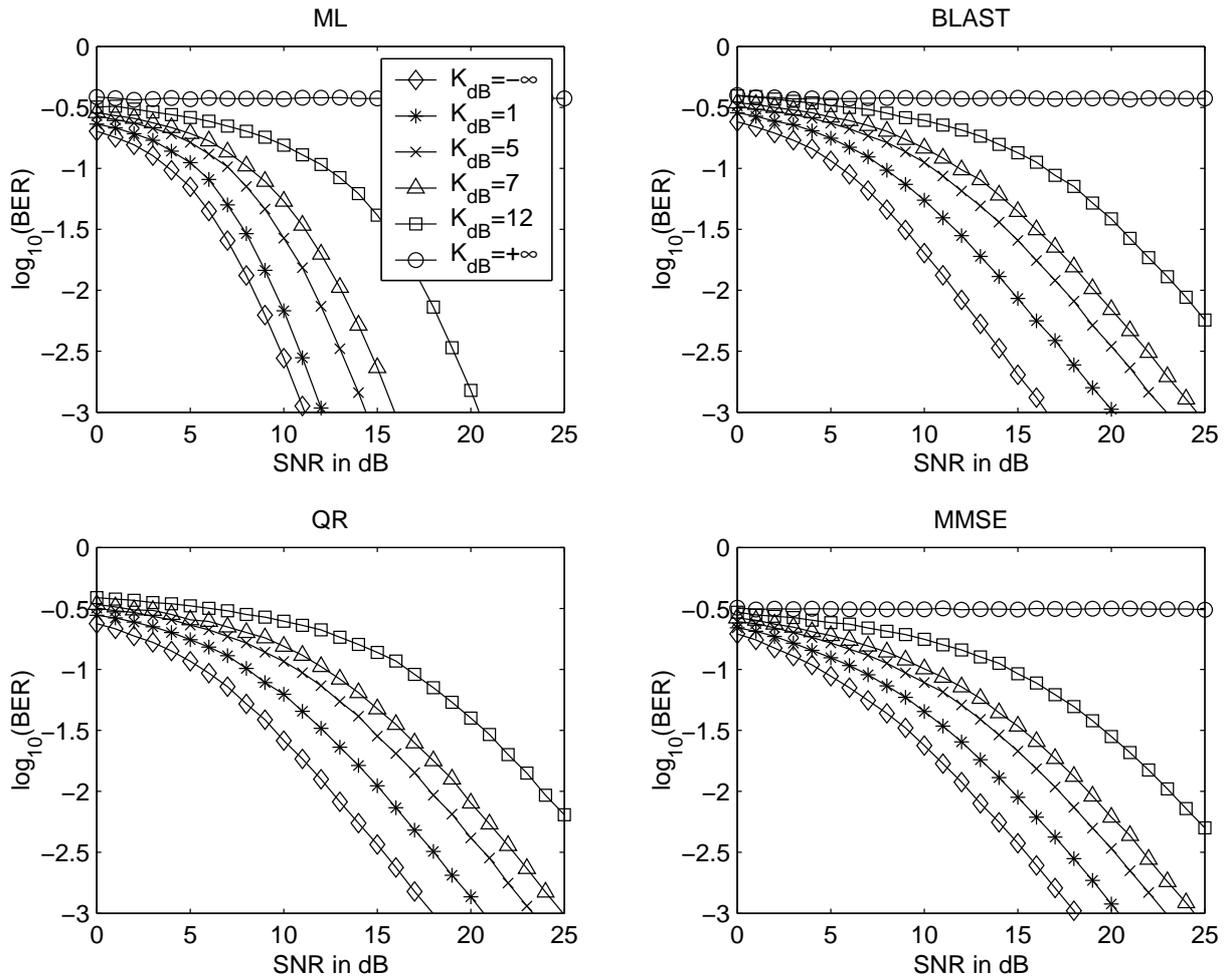


Figure III.10: BER of several decoding algorithms for uncoded MIMO transmission over Ricean channel,  $M_T = 4$ ,  $M_R = 5$ , rate of 8 bits/s/Hz (4-QAM on each transmitting antenna).

ST-codes do not use the multiplexing gain. As a consequence, high-order constellations are required to transmit at a rate close to the channel capacity, especially if the number of antennas grows large. For example, the Alamouti scheme using a 16-QAM should be compared with an uncoded scheme using a 4-QAM. However, if the number of transmitting antennas is increased to three, a ST-code (without multiplexing gain) has to use a 64-QAM to be compared with an uncoded scheme using a 4-QAM. For  $M_T = 4$ , an uncoded system using a 4-QAM should be compared with a ST-code (without multiplexing gain) using a 256-QAM. This comparison highlights the difficulty in designing space-time codes to achieve a non-negligible fraction of the capacity when the number of antennas grows large.

Systems with CSI at the transmitter are potentially able to

- achieve a non-negligible fraction of the capacity even when the number of antennas

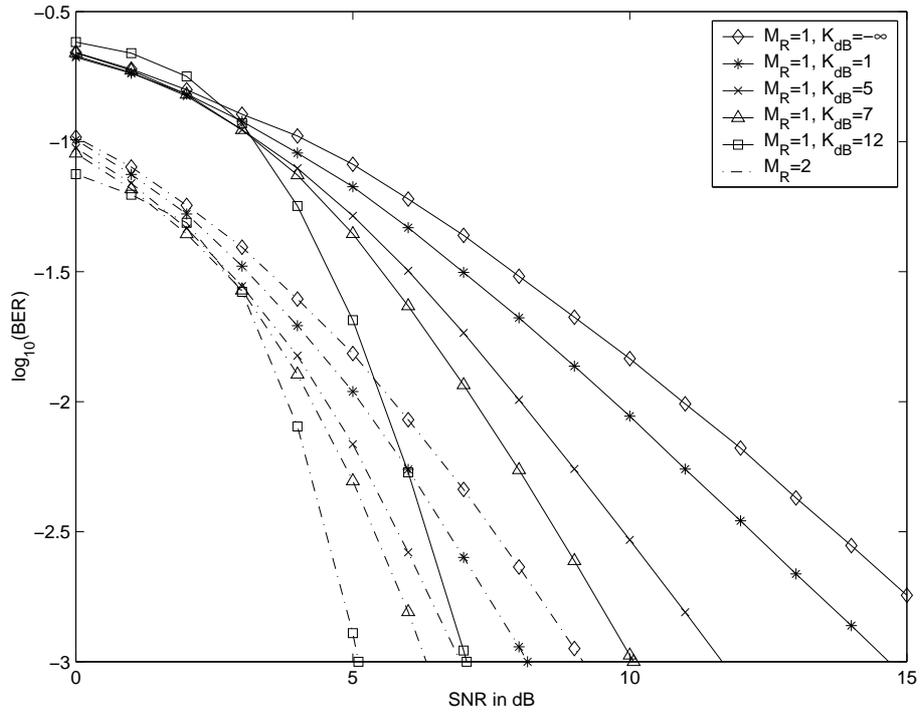


Figure III.11: BER of Alamouti scheme over Ricean channel.

is large,

- operate robustly over correlated channels.

However, these are not the only benefits linked with providing the CSI at the transmitter. Systems with CSI at the transmitter can benefit from the array gain at the transmitter. Hence, they transmit over a channel with higher capacity. Fig. III.12 presents simulation results of the capacity of Ricean channels with and without CSI at the transmitter.

The normalized capacity gain due to CSI at the transmitter is constant when the number of antennas grows large, which means that the capacity gain due to CSI at the transmitter increases linearly with the number of antennas. Furthermore, the gain is higher for highly specular channels than for the i.i.d. channel.

Systems with CSI at the transmitter transmit over a channel with higher capacity, especially when the number of antennas grows large. Furthermore, providing the CSI at the transmitter limits the performance loss due to highly specular channels.

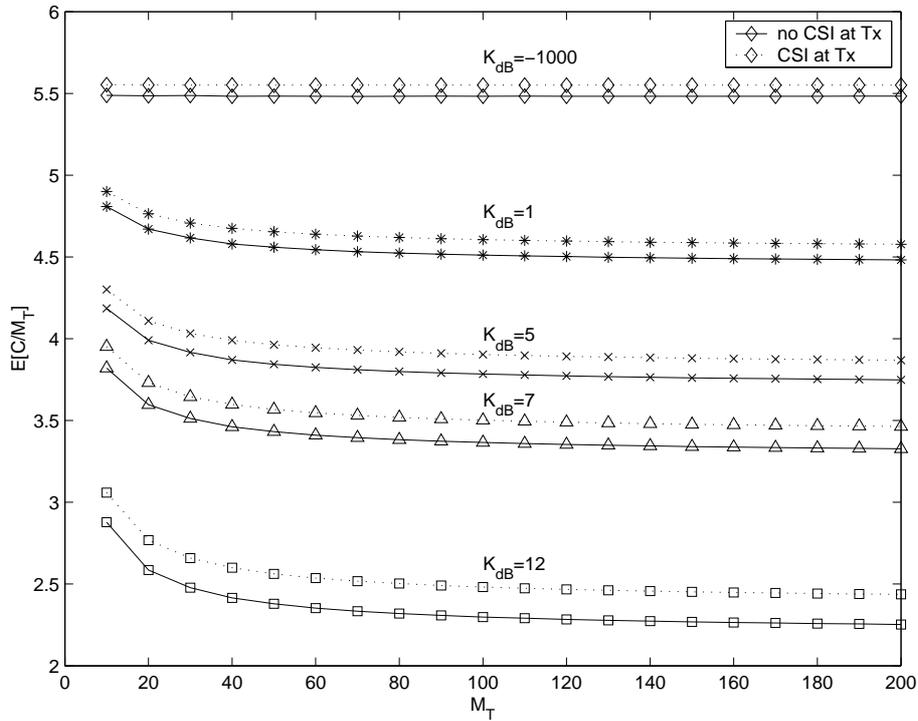


Figure III.12: Normalized capacity of Ricean channels with and without CSI at the transmitter.

### III.6 Conclusion

MIMO transmission techniques without CSI at the transmitter can be classified as uncoded or space-time coded. Their performances depend heavily on the trade-off chosen by system designers between diversity and multiplexing gains. Systems with diversity gain are robust, but are difficult to design for a large number of antennas. Specifically, a large constellation is required to achieve a non-negligible fraction of the capacity of the channel. Systems with multiplexing gain can achieve a non-negligible fraction of the capacity but suffer from performance loss when the channel is highly specular. This is due to the fact that highly specular channels have a lower capacity than i.i.d. channels. Specifically, the capacity of Ricean channels tends to the capacity of their i.i.d. component when the number of antennas grows large.

Finally, providing the CSI at the transmitter can simplify the system design when a large number of antennas is used, while mitigating the effect of channel correlation.



# Chapter IV

## Optimum structure: the SVD

The capacity of the MIMO channel is higher when CSI is available at the transmitter. A practical coding system is required to turn the promises of high capacity into improved performances. The new system has to take advantage of the knowledge of the channel at the transmitter. Information theory suggests an architecture based on the SVD of the channel matrix [56, 57]. This architecture decomposes the MIMO channel into SISO transmission eigenmodes and allocates power to the eigenmodes following a waterfilling algorithm.

Chapter IV introduces the SVD transmission architecture both theoretically and practically to demonstrate the relevance of this transmission technique.

The SVD structure corresponds to the optimal jointly designed linear precoder and decoder following the criterion of maximum capacity. The same structure can be modified to be optimal under other criteria by simply applying appropriate power allocation algorithms, as detailed in Section IV.2.1. The SVD structure is designed to be optimal over the flat-fading channel. Section IV.2.2 indicates that the SVD structure combined with OFDM is also an optimal space-time modulation in terms of information rate.

Practical SVD structures are introduced in Section IV.3, where coding is applied separately on each transmission eigenmode to reduce the complexity of the transmission system. The notion of system capacity is introduced to free the analysis from assumptions on coding.

The SVD transmission architecture combined with waterfilling is optimal, i.e. can achieve the channel capacity with perfect coding (as proven in [12]), when perfect CSI is available at both end of the link. Imperfect CSI can lead to noise enhancement, especially

at low SNR. Usually, CSI is obtained through channel estimation. Channel estimation requirements of the SVD structure are analysed in Section IV.4. Simulation results show that the estimation noise should be smaller than the noise in the transmission to avoid a loss of performance. These results have been partly published in [58].

The channel estimation is usually performed at the receiver. The receiver can feed the explicit precoding matrix or the complete CSI back to the transmitter. When the receiver feeds the complete CSI back to the transmitter, it is necessary to insure that the transmitter and the receiver use the same SVD of the channel matrix, which is possible as demonstrated in Section IV.4.2. These results are, to the knowledge of the author, new and unpublished.

Transmission architectures based on SVD have been proposed and studied extensively in the literature [59, 60, 61, 62]. Though SVD-based transmission devices have been proposed for multiuser channels [63], Section IV focuses on single user channels.

## IV.1 Introduction to the SVD structure

Section II.2.3.4 derived the capacity of the non-ergodic MIMO channel with CSI at the transmitter. The capacity is reached when the transmitted signal is circularly symmetric complex Gaussian with correlation  $\mathbf{Q}$  chosen so that

$$\tilde{\mathbf{Q}} = \mathbf{V}^* \mathbf{Q} \mathbf{V} \quad (\text{IV.1})$$

is diagonal (see (II.52)), where  $\mathbf{V}$  is derived from the SVD of the channel matrix

$$\mathbf{H} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*, \quad (\text{IV.2})$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are unitary and  $\mathbf{\Sigma}$  is diagonal real such that  $\Sigma_{1,1} \geq \Sigma_{2,2} \geq \dots \geq \Sigma_{M,M}$ , where  $M = \min(M_R, M_T)$ . IV.1 is equivalent to

$$\mathbf{Q} = \mathbf{V} \tilde{\mathbf{Q}} \mathbf{V}^*, \quad (\text{IV.3})$$

which is achieved by transmitting

$$\mathbf{x} = \mathbf{V} \tilde{\mathbf{x}} \quad (\text{IV.4})$$

where  $\tilde{\mathbf{x}}$  has independent real and imaginary Gaussian entries and

$$E[\tilde{\mathbf{x}}_i^2] = (P_{wf} - \Sigma_{i,i}^{-2})_+, \forall i \in \{1, \dots, M_T\} \quad (\text{IV.5})$$

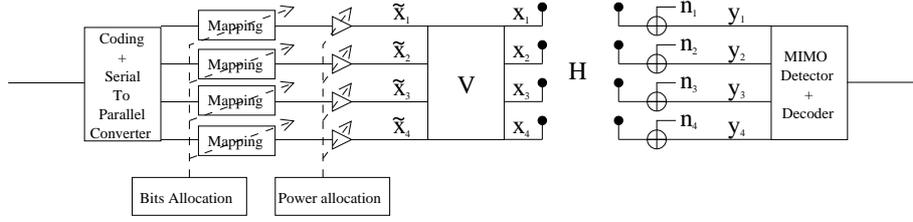


Figure IV.1: SVD transmission architecture

where  $P_{wf}$  is chosen to satisfy  $\sum_i E[\mathbf{x}_i^2] = P$ . This transmission architecture is presented in Fig. IV.1

Consider

$$\tilde{\mathbf{y}} = \mathbf{U}^* \mathbf{y}. \quad (\text{IV.6})$$

The relationship between  $\tilde{\mathbf{y}}$  and the transmitted symbols can be expressed as

$$\begin{aligned} \tilde{\mathbf{y}} &= \mathbf{U}^* \mathbf{y} \\ &= \mathbf{U}^* (\mathbf{H} \mathbf{x} + \mathbf{n}) \\ &= \mathbf{U}^* \mathbf{H} \mathbf{V} \tilde{\mathbf{x}} + \mathbf{U}^* \mathbf{n} \\ &= \mathbf{U}^* \mathbf{U} \mathbf{\Sigma} \mathbf{V}^* \mathbf{V} \tilde{\mathbf{x}} + \mathbf{U}^* \mathbf{n} \\ &= \mathbf{\Sigma} \tilde{\mathbf{x}} + \tilde{\mathbf{n}} \end{aligned} \quad (\text{IV.7})$$

where

$$\tilde{\mathbf{n}} = \mathbf{U}^* \mathbf{n}. \quad (\text{IV.8})$$

Since  $\mathbf{n}$  is assumed zero-mean Gaussian with i.i.d. real and imaginary entries and  $\mathbf{U}^*$  is unitary,  $\tilde{\mathbf{n}}$  and  $\mathbf{n}$  follow the same distribution.

Equation (IV.7) clearly indicates that the MIMO channel has been decomposed into parallel SISO virtual channels over which the power allocation is conducted. These SISO virtual channels are commonly referred to as transmission eigenmodes. The complete SVD transmission architecture and its equivalent model are presented in Fig. IV.2.

## IV.2 Optimality of the SVD

The SVD transmission structure is deduced from the analysis of the capacity achieving correlation of the transmitted signals over the flat-fading channel. However, this structure is optimal with other assumptions and other design criteria. The optimality of the SVD architecture is discussed in this section, with the following results:

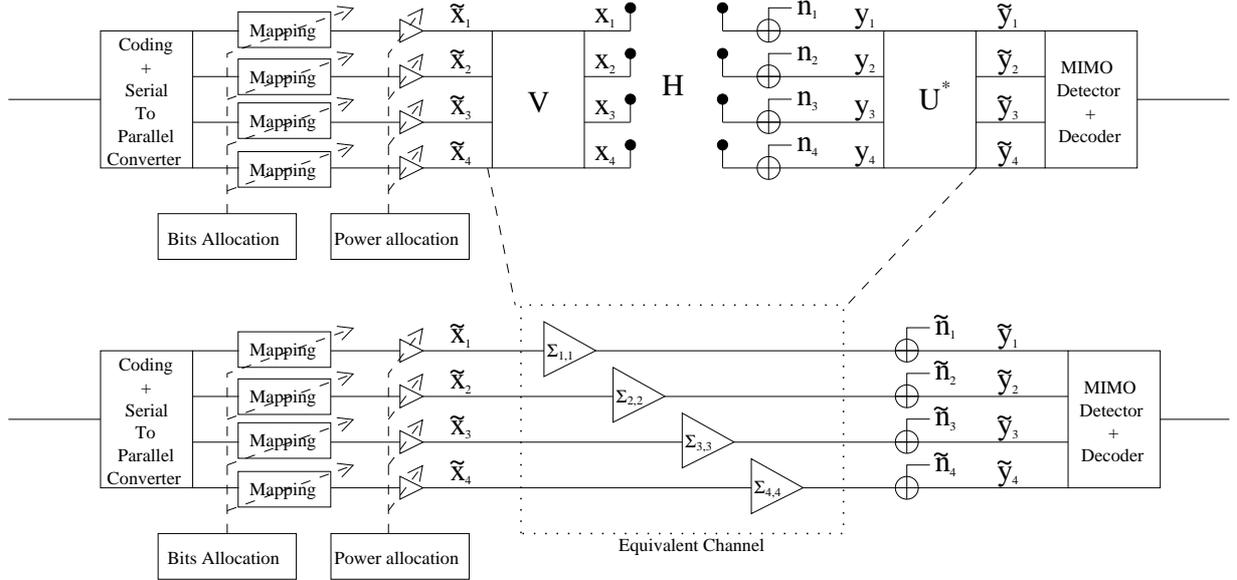


Figure IV.2: SVD transmission architecture with decoding matrix

- the SVD architecture is the optimal jointly designed precoder and decoder architecture over the flat-fading channel under several design criteria (Section IV.2.1),
- the SVD architecture combined with OFDM is equivalent to the optimal spatio-temporal coding over a general fading channel (Section IV.2.2).

### IV.2.1 Optimal linear precoder and decoder

Linear filters are relatively easy to implement and are well-studied. It is natural to try to determine the optimal, jointly designed, linear precoding and decoding filters for a given criterion.

The SVD architecture consists of a linear precoding filter ( $\mathbf{V}$ ) and a linear decoding filter ( $\mathbf{U}^*$ ). Furthermore, the SVD transmission architecture is potentially capacity achieving, i.e. the system is able to transmit information at a rate as close as desired to the capacity of the channel when the symbols on each transmission eigenmode follow a Gaussian PDF, waterfilling power allocation is applied and perfect coding is achieved. It is straightforward to deduce that the SVD architecture is the optimum linear precoder and decoder architecture under the criterion of data rate.

Surprisingly, the SVD architecture is also optimum under a variety of criteria through modification of the power allocation algorithm [64]. Some of the criteria for which the SVD is optimum are detailed in the following.

### IV.2.1.1 Maximum information rate

As discussed previously, the SVD architecture combined with waterfilling power allocation maximises the information rate [64].

### IV.2.1.2 Relative SNR design

It is possible to achieve any set of relative SNRs on streams of data transmitted in parallel (the number of streams cannot exceed the rank of the channel matrix) by assigning streams to transmission eigenmodes and applying adequate power allocation [64].

### IV.2.1.3 Equal error design

The system achieving equal error on each parallel stream can be considered as a special case of the previous criterion: the equal error rate criterion is fulfilled when the SNR is the same on each parallel substream [64].

### IV.2.1.4 Minimum mean square error design

The system achieving the minimum mean square error in the estimation of the symbols combines the SVD architecture with appropriate power allocation. The power allocation algorithm corresponding to this criterion does not guarantee equal mean square error on each eigenmode and might not transmit on the weaker eigenmodes [64].

## IV.2.2 Optimal space-time structure

The SVD architecture is potentially capacity achieving over the flat-fading channel, which corresponds to a subcarrier in an OFDM transmission (Section II.3). However, it is unclear whether the SVD-OFDM-MIMO architecture is optimal over a general fading channel. The optimal spatio-temporal coding architecture over a fading channel can be derived as follows.

Consider the SISO transmission of a block of  $L_t$  symbols over a time-invariant fading channel with impulse response  $\mathbf{h} = [h_0, h_1, \dots, h_\tau]$ , transmission equation (II.8) becomes

$$y(kT_s) = \left[ \sum_{i=0}^{\tau} h_i x((k-i)T_s) \right] + n(kT_s). \quad (\text{IV.9})$$

where  $T_s$  is the sampling period and is discarded in the following. Though (IV.9) is essentially the convolution of  $\mathbf{h}$  and  $\mathbf{x} = [x(0), x(1), \dots, x(L_t)]$ , it is possible to modify it

into a matrix multiplication (since the transmitted block is finite in time) as follows:

$$\begin{pmatrix} y(0) \\ y(1) \\ \vdots \\ y(L_t + \tau) \end{pmatrix} = \begin{pmatrix} h_0 & 0 & \dots & 0 \\ h_1 & h_0 & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_\tau & h_{\tau-1} & \ddots & h_0 \\ 0 & h_\tau & \ddots & \vdots \\ \vdots & \vdots & \ddots & h_{\tau-1} \\ 0 & 0 & \dots & h_\tau \end{pmatrix} \begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ x(L_t) \end{pmatrix} + \begin{pmatrix} n(0) \\ n(1) \\ \vdots \\ n(L_t + \tau) \end{pmatrix}. \quad (\text{IV.10})$$

This SISO model is naturally extended to MIMO transmission by defining

$$\begin{aligned} \mathbf{y} &= [y_1(0), \dots, y_1(L_t + \tau), y_2(0), \dots, y_{M_R}(L_t + \tau)]^\dagger, \\ \mathbf{x} &= [x_1(0), \dots, x_1(L_t), x_2(0), \dots, x_{M_R}(L_t)]^\dagger, \\ \mathbf{n} &= [n_1(0), \dots, n_1(L_t + \tau), n_2(0), \dots, n_{M_R}(L_t + \tau)]^\dagger, \end{aligned} \quad (\text{IV.11})$$

where  $(\cdot)^\dagger$  is the transpose non-conjugate operation and

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{1,1} & \dots & \mathbf{H}_{1,M_T} \\ \vdots & \ddots & \dots \\ \mathbf{H}_{M_R,1} & \dots & \mathbf{H}_{M_R,M_T} \end{pmatrix} \quad (\text{IV.12})$$

where  $\forall i < M_R, j < M_T$ ,  $H_{i,j}$  is a matrix of the Toeplitz form as appears in (IV.10), with entries the elements of the sampled channel impulse response from transmitting antenna  $j$  to receiving antenna  $i$ . The transmission equation of the MIMO space-time transmission can be expressed as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}. \quad (\text{IV.13})$$

The capacity of a system following this transmission equation has been studied extensively in Section II.2. Obviously, the capacity of this space-time MIMO channel with CSI at the transmitter can be found by decomposing the channel into space-time transmission eigenmodes and applying waterfilling power allocation. The block diagram of the suggested transmission architecture is represented in Fig. IV.3.

This architecture is quite complex since the SVD of a very large matrix has to be computed. The space and time dimensions are coupled, resulting in complex processing at both transmitter and receiver. This complexity has to be compared with that of the SVD-OFDM-MIMO architecture. As illustrated in its simplified block diagram (Fig. IV.4), the time processing and the space processing of the signal are separate since the space

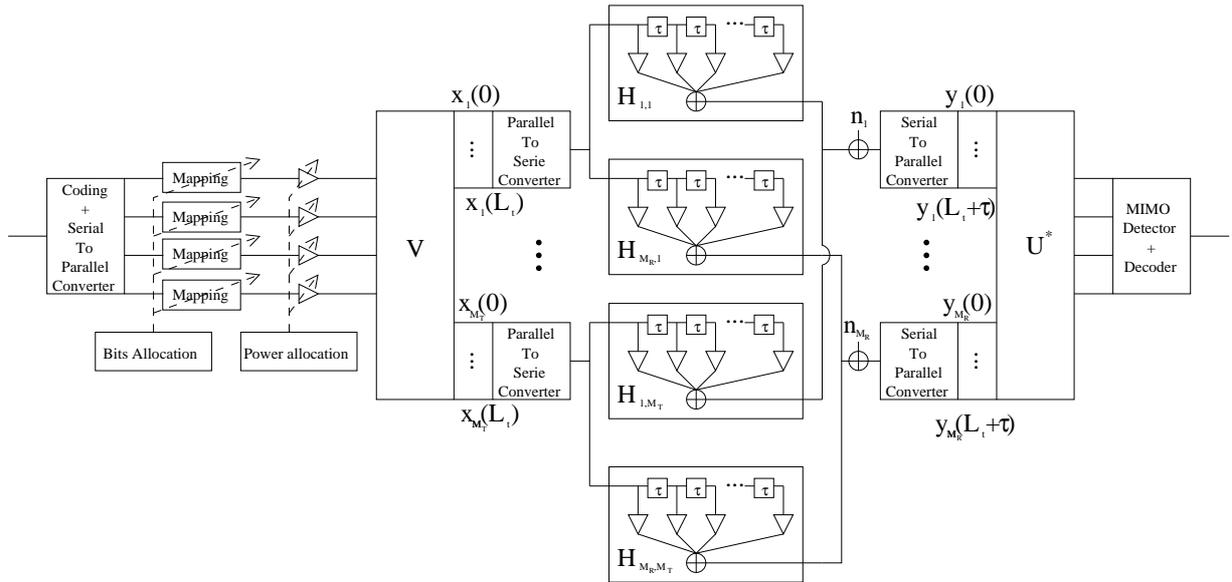


Figure IV.3: SVD Space-Time MIMO transmission architecture

dimension is exploited through the SVD approach, whereas the time dimension is handled through OFDM. This decoupling of the processing in space and time leads to a significant reduction in complexity since

- the OFDM modulation requires little adaptive processing (only the power allocation to the subcarriers) and can be efficiently implemented,
- computing  $L_t$  SVDs of  $(M_R, M_T)$  matrices is much simpler than computing the SVD of one  $(L_t \times M_R, (L_t + \tau) \times M_T)$  since the complexity of the SVD operation is much greater than linear.

Counter-intuitively, the reduction in complexity linked with choosing a SVD-OFDM-MIMO architecture rather than a SVD-Space Time-MIMO architecture is not obtained at the expense of performance. It is proved in [56], [65] that in the limit  $L_t \rightarrow \infty$ , the capacity of both the SVD-OFDM-MIMO architecture and the SVD-Space Time-MIMO architecture converge to the continuous frequency channel capacity. This result can be easily understood by noting that the basis functions of OFDM are complex exponentials, which are also (for infinite duration) the eigenfunctions of convolution. When the OFDM symbols are much longer than the delay spread, SVD-OFDM-MIMO provides orthogonal decomposition in space and time, and the decoupling arises naturally.

The benefits associated with the SVD-OFDM-MIMO architecture (low complexity and equal performance) justifies that the remaining of this thesis is focused solely on this

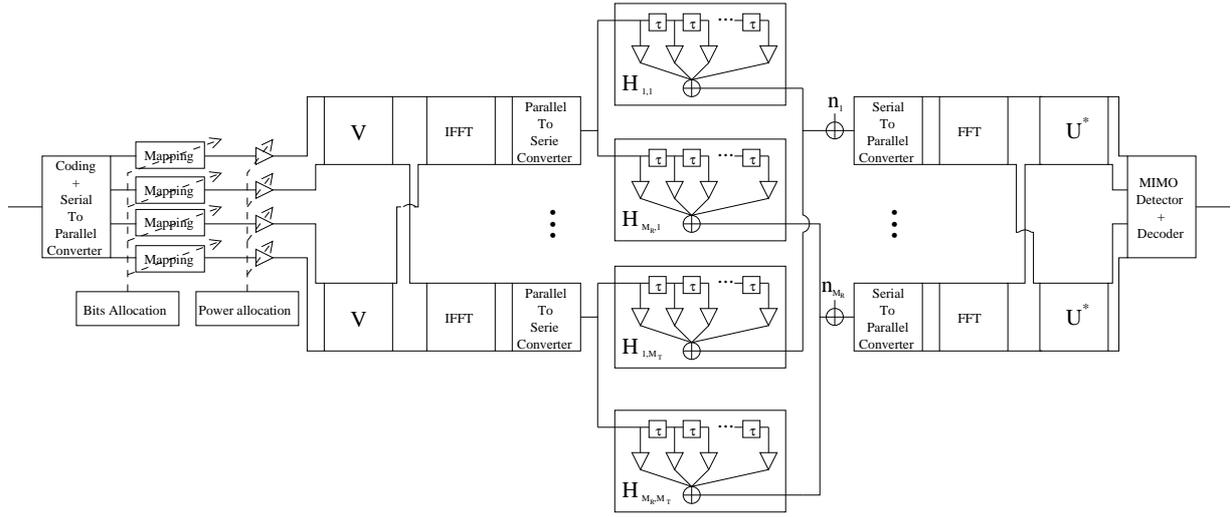


Figure IV.4: SVD-OFDM-MIMO transmission architecture

architecture. As a result, flat-fading is assumed for the remaining of the thesis.

### IV.3 Practical SVD architecture

It has been shown that the SVD architecture enables the decomposition of the MIMO channel into transmission eigenmodes, allowing the use of a separate SISO modulation on each transmission eigenmode without capacity penalty. However, coding across the transmission eigenmodes as well as joint demodulation and decoding of the received symbols across the eigenmodes are still assumed. The complexity of such a system is usually not acceptable. Therefore, a simplified SVD architecture uses the transmission eigenmodes as parallel Gaussian channels, as shown in Fig. IV.5.

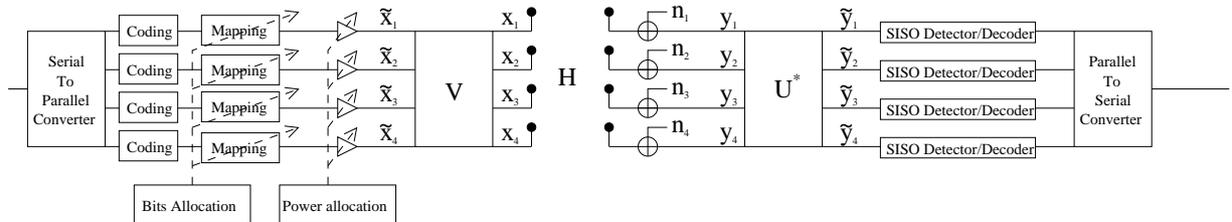


Figure IV.5: A practical SVD transmission architecture

### IV.3.1 Performance loss of the practical architecture

The loss of performance due to the practical structure is detailed in the following. Assuming a soft output demodulator, SISO demodulators can be applied without penalty, since modulation is applied independently on the transmission eigenmodes. The practical structure suffers from a loss of performance through the application of separate SISO coding. This is simply due to the fact that the capacity is only achieved in the asymptotic case of infinite length coding. The capacity of the channel is equal to the sum of the capacities of the eigenmodes. However, when separate coding is applied to each eigenmode, the length of the codes applied is approximately  $\min(M_T, M_R)$  times shorter than the length of a code that would be applied across all  $\min(M_T, M_R)$  eigenmodes. In practice, binary coding can be applied (before modulation) to each eigenmode, or to all eigenmodes. Using coding over all eigenmodes allows us to use longer binary codes for a given block size. It is well-known that the performance of good codes (codes transmitting close to the capacity) is linked to their length. For example, the performance of turbo-codes is closely related to the design of interleavers, and the performance of interleavers increases with their length [66]. Coding over all the eigenmodes leads to a faster convergence to the asymptotic channel capacity limit as the interleaving length is increased.

### IV.3.2 Notion of system capacity

The performance of the system depends heavily on the coding strategy, as does the robustness of the system to various impairments (e.g. imperfect channel estimation). However, the study of specific coding strategies is not the purpose of this thesis. Therefore, it is necessary to free the analysis from various assumptions related to coding strategies.

The practical SVD architecture considers each transmission eigenmode as a separate channel. It is possible to obtain the Signal to Interference and Noise Ratio (SINR) on each transmission eigenmode. From the SINR, it is straightforward to deduce the capacity of the eigenmode when no joint decoding is applied. The notion of capacity of a transmission system is then easily defined as the sum of the capacities of the transmission eigenmodes. The 'system capacity' corresponds to an upper bound on the data rate achievable when coding and decoding are applied separately on each transmission eigenmode.

The 'system capacity' is generally not equal to the capacity of the channel, as demonstrated by the following example: consider the SVD transmission systems where impair-

ments forced the decoding matrix to be mistakenly taken as equal to the identity matrix (Fig. IV.5 with  $\mathbf{U}^*$  replaced with the identity matrix). The output of the decoding matrix is exactly the signal on the receiving antennas. In most cases (with a probability equal to one), the signal on each receiving antenna is the weighted sum of signals from each transmission eigenmode plus Gaussian noise. Therefore, the SINR on the output of the eigenmodes of this system is poor and the 'system capacity' is low (always lower than the capacity of the channel). This is in sharp contrast to the maximum information rate supported by this architecture when coding and decoding is applied across the eigenmodes: the erroneous decoding matrix does not limit the maximum data rate achievable if a joint detector were applied on the output of the decoding matrix (Fig. IV.2 with  $\mathbf{U}^*$  replaced with the identity matrix), potentially allowing transmission at a rate equal to the capacity of the channel.

In the remainder of the thesis the influence of various practical impairments on the 'system capacity' of SVD systems is analysed and compared with the corresponding channel capacity.

## IV.4 Channel estimation and SVD architecture

Both the transmitter and receiver of SVD systems require, at least, the partial knowledge of the CSI: it is necessary for the transmitter to obtain  $\mathbf{V}$ , the precoding matrix, and for the receiver to obtain  $\mathbf{U}^*$ , the decoding matrix.

Usually, communication systems obtain the CSI through pilot based channel estimation, i.e. measurements of the channel response to the transmission of known symbols (pilot symbols). The insertion of pilot symbols corresponds to a loss of transmission slots, i.e. data rate, and a loss of transmission power. However, increasing the number of pilot symbols can lead to more accurate channel estimation. Determining the appropriate number of pilot symbols is a trade-off between the accuracy of the channel estimation and the loss related to pilot symbol insertion. This trade-off has been studied in the literature [67, 68].

Blind channel estimation [69, 70] can also be applied, where only the statistics of the symbols is known and CSI can be recovered without any overhead (transmission of pilot symbols).

Regardless of the specific channel estimation technique applied, the channel estimation

requirements are discussed in the following.

#### IV.4.1 Channel estimation accuracy

Consider the system presented in Fig. IV.6. The channel matrix is estimated through measurements. However, noise in the measurements leads to inaccurate CSI. The estimated CSI is decomposed through the SVD and the precoding and decoding matrices are modified accordingly.

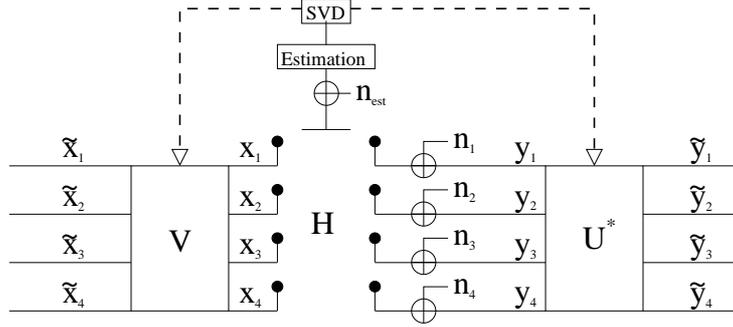


Figure IV.6: CSI estimation and SVD architecture.

The degradation of performance due to incorrect estimation of the CSI is analysed in the following. The channel matrix  $\mathbf{H}$  is incorrectly estimated as

$$\hat{\mathbf{H}} = \frac{\mathbf{H} + \mathbf{N}_{est}}{r}, \quad (\text{IV.14})$$

where  $\mathbf{N}_{est}$  is the channel estimation noise and the entries of  $\mathbf{N}_{est}$  are assumed i.i.d. complex Gaussian and  $r$  is a normalization factor defined as

$$r = \sqrt{(E[\|\mathbf{H}\|_F^2] + E[\|\mathbf{N}_{est}\|_F^2]) / (M_R \times M_T)}. \quad (\text{IV.15})$$

Normalization is required to maintain the equality between the total transmitted power and the received SNR. The following SVD applies:  $\hat{\mathbf{H}} = \hat{\mathbf{U}}\hat{\Sigma}\hat{\mathbf{V}}^*$ . The transmission equation becomes

$$\begin{aligned} \mathbf{y} &= \hat{\mathbf{U}}^*(\mathbf{H}\hat{\mathbf{V}}\mathbf{x} + \mathbf{n}) \\ &= \hat{\mathbf{U}}^*(r \times \hat{\mathbf{H}}\hat{\mathbf{V}}\mathbf{x}) - \hat{\mathbf{U}}^*\mathbf{N}_{est}\hat{\mathbf{V}}\mathbf{x} + \hat{\mathbf{U}}^*\mathbf{n} \\ &= r \times \hat{\Sigma}\mathbf{x} - \hat{\mathbf{U}}^*\mathbf{N}_{est}\hat{\mathbf{V}}\mathbf{x} + \hat{\mathbf{U}}^*\mathbf{n}. \end{aligned} \quad (\text{IV.16})$$

The received signal consists of the signal ( $r \times \hat{\Sigma}\mathbf{x}$ ), an interference term ( $\hat{\mathbf{U}}^*\mathbf{N}_{est}\hat{\mathbf{V}}\mathbf{x}$ ) and finally the noise term ( $\hat{\mathbf{U}}^*\mathbf{n}$ ). The analysis can be conducted by carefully examining the different terms.

- $\hat{\underline{\mathbf{H}}}$  is a complex i.i.d. Gaussian matrix as the sum of two Gaussian i.i.d. matrices. Furthermore  $\hat{\underline{\mathbf{H}}}$  is normalized to have entries with variance one. Therefore,  $\underline{\Sigma}$  and  $\hat{\underline{\Sigma}}$  follow the same distribution.
- $\hat{\underline{\mathbf{U}}}^* \underline{\mathbf{N}}_{est} \hat{\underline{\mathbf{V}}}$  follows the same distribution as  $\underline{\mathbf{N}}_{est}$  since  $\hat{\underline{\mathbf{U}}}$  and  $\hat{\underline{\mathbf{V}}}$  are unitary and  $\underline{\mathbf{N}}_{est}$  is an i.i.d. Gaussian matrix.
- $\hat{\underline{\mathbf{U}}}^* \underline{\mathbf{n}}$  follows the same distribution as  $\underline{\mathbf{n}}$  since  $\hat{\underline{\mathbf{U}}}^* \underline{\mathbf{n}}$  is unitary.

An approximation of the SINR is given by

$$\text{SINR} = \frac{(E[\|\underline{\mathbf{H}}\|_F^2] + E[\|\underline{\mathbf{N}}_{est}\|_F^2])E[\|\underline{\mathbf{x}}\|_F^2]}{E[\|\underline{\mathbf{N}}_{est}\|_F^2]E[\|\underline{\mathbf{x}}\|_F^2] + M_T \times E[\|\underline{\mathbf{n}}\|_F^2]}. \quad (\text{IV.17})$$

When the interference is negligible compared to both the noise and the signal, the SINR tends to the SNR of the system with no interference:

$$\text{SINR} \rightarrow \frac{(E[\|\underline{\mathbf{H}}\|_F^2])}{M_R \times M_T} \times \frac{E[\|\underline{\mathbf{x}}\|_F^2]}{E[\|\underline{\mathbf{n}}\|_F^2]/M_R}. \quad (\text{IV.18})$$

On the contrary, when the noise is negligible compared to the interference the SINR becomes

$$\text{SINR} = \frac{(E[\|\underline{\mathbf{H}}\|_F^2] + E[\|\underline{\mathbf{N}}_{est}\|_F^2])}{E[\|\underline{\mathbf{N}}_{est}\|_F^2]}. \quad (\text{IV.19})$$

Obviously, in this situation, the SINR does not depend on the SNR. Increasing the power of the signal does not improve the transmission.

Simulation results of the system capacity with varying estimation errors are presented in Fig. IV.7. Equal power and waterfilling power allocation differ only at low SNRs. Define  $\text{SNR}_{est}$  as the signal ( $\underline{\mathbf{H}}$ ) to estimation noise ( $\underline{\mathbf{N}}_{est}$ ) ratio. When  $\text{SNR}_{est} \gg \text{SNR}$ , the system capacity equals the channel capacity: the imperfect channel estimation does not reduce the performance of the system. On the contrary, when  $\text{SNR}_{est} \ll \text{SNR}$ , the performance plateaus with increasing SNR and the performance of the system depends only on the  $\text{SNR}_{est}$ .

#### IV.4.2 Uniqueness of the SVD

The architecture presented in Fig. IV.6 includes an estimation block and a SVD block. In practical systems, channel estimation is performed through measurements of pilot symbols. Therefore, the estimation block is usually located at the receiver. It is necessary to feedback the CSI to the transmitter. Two solutions are available:

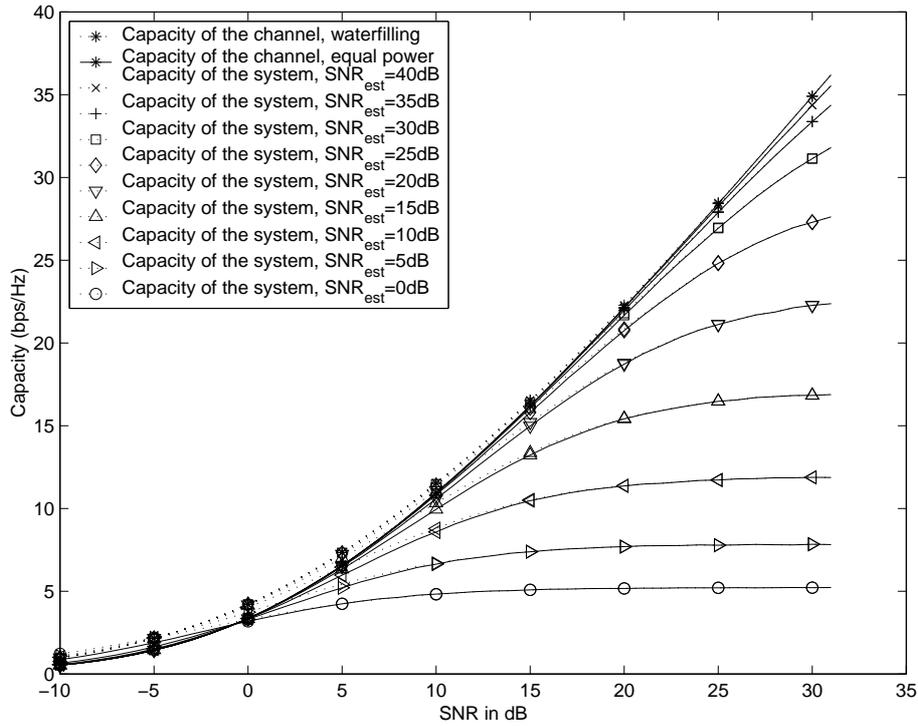


Figure IV.7: Capacity of the SVD system with imperfect channel estimation over the i.i.d. channel,  $M_R = M_T = 4$ . Plain lines and dotted lines denote respectively equal power and waterfilling power allocation.

- either the receiver performs the SVD and feeds the precoding matrix and the matrix of the singular values back to the transmitter,
- or the receiver feeds the CSI back to the transmitter.

In the second case, the SVD of the channel matrix is derived separately at the transmitter and at the receiver. This solution relies implicitly on the uniqueness of the SVD of the channel matrix: if the SVD is not unique, then the transmitter and receiver might decompose the channel in two different ways,  $\mathbf{H} = \mathbf{U}_T \mathbf{\Sigma}_T \mathbf{V}_T^*$  and  $\mathbf{H} = \mathbf{U}_R \mathbf{\Sigma}_R \mathbf{V}_R^*$ . The transmitter filters the transmitted signal by  $\mathbf{V}_T$  and the receiver filters the received signals by  $\mathbf{U}_R^*$ . The transmission equation becomes:

$$\mathbf{y} = \mathbf{U}_R^* (\mathbf{H} \mathbf{V}_T \mathbf{x} + \mathbf{n}) = \mathbf{U}_R^* \mathbf{U}_T \mathbf{\Sigma}_T \mathbf{x} + \mathbf{U}_R^* \mathbf{n}. \quad (\text{IV.20})$$

In such a situation, the MIMO channel is no longer decomposed into parallel channels.

**Theorem 6.** *The SVD of a matrix  $\mathbf{H} \in \mathbb{C}^{M \times M}$  is unique, up to a multiplication of the input and output eigenvectors by a complex number of norm 1, if the following assumptions hold:*

- the singular values are sorted in descending order,
- the singular values are of multiplicity 1, i.e.  $\forall i \neq j, \Sigma_{i,i} \neq \Sigma_{j,j}$ .

#### IV.4.2.1 Proof of the uniqueness theorem

It has been shown that the set of singular values of a matrix is unique [71]. Consider  $\mathbf{H} \in \mathbb{C}^{M \times M}$  such that:

$$\forall i \neq j, \Sigma_{i,i} \neq \Sigma_{j,j}, \quad (\text{IV.21})$$

with singular values of multiplicity one and singular values indexes chosen so that

$$\Sigma_1 > \Sigma_2 > \Sigma_3 \dots \Sigma_{M,M}. \quad (\text{IV.22})$$

Suppose  $\mathbf{H}$  has two SVD decompositions, then:

$$\exists(\mathbf{U}, \mathbf{V}, \Sigma, \mathbf{W}, \mathbf{Z}), \mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^* = \mathbf{W}\Sigma\mathbf{Z}^*,$$

with  $\mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{Z}$  unitary matrices.

Obviously  $\mathbf{V} \neq \mathbf{Z}$ , otherwise if  $\mathbf{V} = \mathbf{Z}$  and  $\mathbf{U} \neq \mathbf{W}$ , then  $\exists i$  with  $\mathbf{U}_{:,i} \neq \mathbf{W}_{:,i}$ . This is impossible for  $\Sigma_{i,i} \neq 0$  since

$$\mathbf{W}_{:,i} = \frac{1}{\Sigma_{i,i}} \mathbf{H}\mathbf{Z}_{:,i}^* = \frac{1}{\Sigma_{i,i}} \mathbf{H}\mathbf{V}_{:,i}^* = \mathbf{U}_{:,i} \quad (\text{IV.23})$$

Furthermore,  $\Sigma_{i,i} = 0$  is only possible for  $i = M$ . If  $\mathbf{U}_{:,i} = \mathbf{W}_{:,i}$  for all  $i \neq M$ , then  $\mathbf{U}_{:,M} = c\mathbf{W}_{:,M}$  where  $c$  is a complex number of norm one since  $\mathbf{U}$  and  $\mathbf{W}$  are unitary matrices.

Then  $\exists i$  with  $\forall j < i, \mathbf{V}_{:,j} = \mathbf{Z}_{:,j}$  and  $\mathbf{V}_{:,i} \neq \mathbf{Z}_{:,i}$ . Then  $\|\mathbf{H}\mathbf{V}_{:,i}^*\|_2 = \Sigma_{i,i}$ . Furthermore,  $\exists(c_1, \dots, c_M) \in \mathbb{C}^M$  such that

$$\mathbf{V}_{:,i} = \sum_{k=1}^M c_k \mathbf{Z}_{:,k} \quad (\text{IV.24})$$

since the vectors of  $\mathbf{Z}$  form a base of  $\mathbb{C}^M$ .

Since  $\mathbf{V}$  is unitary, and  $\forall j < i, \mathbf{V}_{:,j} = \mathbf{Z}_{:,j}$ , it is straightforward to show that

$$\mathbf{V}_{:,i} = \sum_{k=i}^M c_k \mathbf{Z}_{:,k}. \quad (\text{IV.25})$$

Then

$$\|\mathbf{H}\mathbf{V}_{:,i}^*\|_2 = \sum_{k=i}^n |c_k| \Sigma_{k,k}. \quad (\text{IV.26})$$

Remember that

$$\|\mathbf{H}\mathbf{V}_{:,i}^*\|_2 = \Sigma_{i,i}, \quad (\text{IV.27})$$

and

$$\sum_{k=i}^M |c_k|^2 = 1, \quad (\text{IV.28})$$

then  $|c_i| = 1$  since  $\forall i < j, \Sigma_{i,i} > \Sigma_{j,j}$ .

This implies that  $\mathbf{V}_{:,i} = \mathbf{Z}_{:,i}$  up to a complex scalar multiplication and concludes the proof.

#### IV.4.2.2 Application to practical systems

Under the previous assumptions, the SVDs of the matrix  $\mathbf{H}$  are all of the form  $\text{SVD}(\mathbf{H}) = (\mathbf{U}\mathbf{\Phi}, \mathbf{\Sigma}, \mathbf{V}\mathbf{\Phi})$ , where  $\mathbf{\Phi}$  is a diagonal matrix with complex entries of norm 1. A simple way to use the same SVD at both ends of the transmission chain is to choose  $\mathbf{\Phi}$  according to a given criterion, e.g. Matlab chooses  $\mathbf{\Phi}$  such that the first row of elements in  $\mathbf{V}$  are real numbers.

#### IV.4.2.3 Discussion of the assumptions

The assumptions of Theorem 6 are restrictive and do not always apply to practical systems. However, the channel matrix can be manipulated to verify the assumptions with a high probability.

The first assumption of Theorem 6 is that  $M_R = M_T$ . If  $M_R \neq M_T$ , both transmitter and receiver restrict the CSI to the  $\min(M_R, M_T)$  transmission eigenmodes, suppressing the requirement for the system to have the same number of antennas at both the transmitter and receiver sides of the link.

Therefore, the transmission system can operate when the singular values are of multiplicity 1. On the contrary, the system cannot operate when several singular values are equal to zero (highly correlated channel) or when two non-zero singular values are equal.

Matrices with several singular values equal to zero can be treated by removing the eigenmodes associated with the zero singular values. This is justified by the fact that eigenmodes with a gain equal to zero cannot transmit information (have a capacity of 0 bps/Hz), and therefore can be simply discarded.

Following the assumption that the channel matrix is i.i.d. Gaussian, the joint PDF of

the singular values is [72]

$$\text{pdf}_{\underline{\Sigma}} = \frac{2^{M_T}}{\prod_{j=1}^M (M_T - j)! (M_R - j)!} \exp(-\sum_{j=1}^M \Sigma_{j,j}^2) \times \left( \prod_{j=1}^M \Sigma_{j,j}^2 \right)^{\max(M_R, M_T) - M} \prod_{i < j} (\Sigma_{i,i}^2 - \Sigma_{j,j}^2)^2. \quad (\text{IV.29})$$

The last term in the expression implies the probability for two equal singular values is 0. The probability becomes finite when the singular value are quantized (see Section V.3).

This occurrence is rare but practical systems need to handle such cases. Possible solutions include

- considering that the event is so rare that the error bursts it creates can be tolerated,
- avoiding transmission on equal gain eigenmodes,
- transmitting additional pilots on the equal gain eigenmodes to provide the receiver with information on which SVD the transmitter selected.

#### IV.4.2.4 Conclusion

Though the SVD of a complex matrix is never unique, the SVD can be applied to wireless systems if its usage is restricted to the non-zero transmission eigenmodes. In such a case, the transmitter and the receiver can insure implicitly that they use the same SVD at both ends of the link.

The only difficulty arises with the case of a channel matrix where two eigenmodes have the same gain. In such a case, it is not possible to agree implicitly on a single SVD at both ends of the link but practical solutions exist to correct the effects of this problem.

## IV.5 Conclusion

The optimal structure, in terms of information rate, for MIMO transmission over a flat fading channel consists of a precoding matrix and a decoding matrix which decompose the MIMO channel into parallel transmission eigenmodes. Power allocation has to be applied on the transmission eigenmodes to obtain the optimal performance. The SVD transmission system actually corresponds to the optimal solution under a wide range of criteria, e.g. the SVD structure combined with OFDM is equivalent to the optimum space time coding technique in the limit of large OFDM blocks.

Practical SVD systems apply coding and decoding to each transmission eigenmode separately to reduce the complexity of the system. A loss of performance results. It is possible to define the 'system capacity' of such architectures and this definition allows analysis of various system impairments without placing assumptions on the coding techniques.

The precoding and decoding matrices, as well as the power allocation, need to be matched with the channel matrix. Errors in the estimation of the channel matrix induce a loss of performance which is negligible if the channel estimation inaccuracy is much smaller than the inaccuracy in the detection of the transmitted signal. Finally, practical systems can decompose the channel matrix separately at the transmitter and receiver, which is not straightforward since the SVD of a complex matrix is not unique. Problems, such as equal singular values, have already been flagged and resolved in the case of a static channel with perfect channel estimation. This problem is further discussed in Section V.3 under more realistic assumptions. Other problems include getting the CSI to the transmitter and handling the dynamics in the channel. These issues are discussed in the next chapter.



# Chapter V

## SVD architecture in TDD environment

The MIMO channel has a large capacity, however low complexity solutions without CSI at the transmitter do not take full advantage of that fact. The SVD architecture is optimal to transmit over the MIMO channel but requires the CSI at both the transmitter and the receiver. This might result in significant system overhead to transmit the CSI from the receiver back to the transmitter when the channel is time-varying. In practice, this overhead might be unacceptably large.

It is possible to suppress this overhead when the channel is reciprocal. In such a case the transmitter and the receiver can estimate the channel separately. TDD channels are practical examples of reciprocal channels. Chapter V introduces the SVD architecture over a reciprocal channel. The impact of the errors due to channel estimation (time delay and channel estimation noise) on the system capacity is analysed and practical solutions are proposed.

TDD channels are reciprocal in essence, but mismatched transmitter and receiver chains can remove the reciprocity of the channel. In such a case, hardware calibration has to be applied to recover the reciprocity property of the channel. The effect of mismatched transmitter and receiver chains is studied in Section V.2. A calibration procedure is proposed which completely corrects the effect of mismatched chains without requirement for additional hardware. The calibration procedure relies on a handshake at the beginning of the transmission, as well as the hypothesis that the impairments of the chains are static. These results have been partly published in [73] and the calibration procedure has been granted a provisional patent.

The channel estimation is usually conducted through measurement of pilot symbols.

Therefore, the channel estimation is not perfect. The effect of incorrect channel estimation on SVD-TDD systems is analysed in Section V.3 through the theory of matrix perturbation. An event named 'singular value crossing', which prevents the possibility of robust transmission, is exhibited. The probability of occurrence of this event is studied through matrix perturbation theory as well as system simulation. These results are, to the knowledge of the author, new and unpublished.

The effects of imperfect channel estimation on system capacity are extensive, as explained in Section V.4. However, a new architecture, with limited added complexity, obtains the benefit of the SVD architecture when CSI is precisely known at both ends of the link while seamlessly shifting to a non-precoded system when the channel estimation precision deteriorates at the transmitter. The new architecture transmits the pilot symbols through the transmitting matrix, which allows the receiver to gain knowledge of both the CSI and the transmitting matrix with a single set of pilot symbols. These results have been partly published in [58], [74] and [75]

Though the proposed architecture is robust to imperfect channel estimation at the transmitter, the system capacity is always higher when a better channel estimate can be obtained. It is possible to improve the channel estimation through increasing the density of pilot symbols. However, this results in additional overhead. Section V.5 proposes to reduce the errors due to the time-variation of the channel and noisy estimate through filtering of the CSI estimate in time. This proposed scheme consists simply in an extension to MIMO channels of Pilot Symbol Assisted Modulation (PSAM) [76]. The filtering can be applied to either the CSI or the precoding/decoding matrices, though superior performance is usually obtained by filtering the CSI (see Section V.5). These results have been partly accepted for publication in [77].

## V.1 SVD architecture on reciprocal channels

Consider a MIMO reciprocal flat-fading channel. The channel from antenna  $i$  of transceiver A to antenna  $j$  of transceiver B is represented by the complex coefficient  $\mathbf{H}_{j,i}$ . For a reciprocal channel, the channel from antenna  $j$  of transceiver B to antenna  $i$  of transceiver A is  $\mathbf{H}_{j,i}$ . Therefore, if the MIMO channel from A to B is the matrix  $\mathbf{H}$ , the channel from B to A is  $\mathbf{H}^\dagger$ .

Ideal SVD based systems have knowledge of the CSI at the transmitter and receiver

simultaneously. SVD systems over reciprocal channels can obtain the CSI at both ends of the link without explicit feedback of the CSI. Both transmitter and receiver estimate the channel through measurement of pilot symbols (Fig. V.1), as detailed in the following.

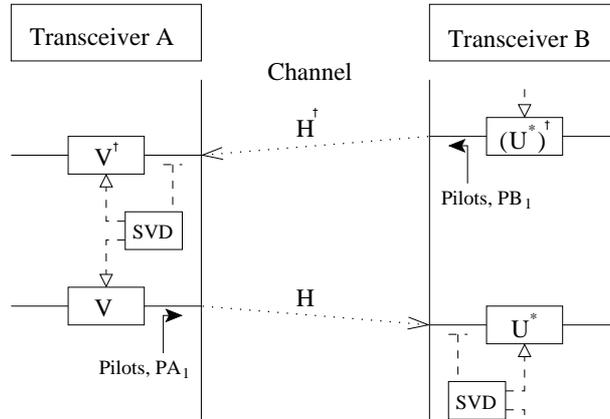


Figure V.1: SVD transmission over reciprocal channel

It is straightforward to obtain the CSI at the receiver. Pilot symbols  $PA_1$  are transmitted through the channel  $\mathbf{H}$  prior to data transmission (Fig. V.1).

To obtain the CSI at the transmitter, the reciprocity properties of the channel are used. If transceiver A wishes to send data to transceiver B, it requests B to send pilot symbols  $PB_1$  to obtain the CSI,  $\mathbf{H}^\dagger$ , and then transmits data through the  $\mathbf{V}$  matrix that has been derived from the CSI.

The necessity for pilot symbols can be suppressed if blind channel estimation is applied.

## V.2 Hardware calibration procedure

### V.2.1 Effect of hardware errors on channel reciprocity

For a TDD system, the wireless channel is reciprocal. However, the channel does not include the transmitter and receiver chains. So far, it has been assumed that the amplitude and phase shift of the transmitters' analog electronic (between the digital to analog converters and the antennas) is the same as that of the receivers' electronics (between the antennas and the analog to digital converters) for a given antenna. That is, the transmitter and receiver Radio Frequency (RF) chains are perfect. When the transmitter chain includes a phase or amplitude error that is not replicated in the receiver chain, the channel is not reciprocal.

The errors introduced by transmitter A, receiver A, transmitter B and receiver B are represented respectively by the matrices  $\mathbf{Er}^{T,A}$ ,  $\mathbf{Er}^{R,A}$ ,  $\mathbf{Er}^{T,B}$ ,  $\mathbf{Er}^{R,B}$ . Each of these matrices are diagonal, with complex elements on the diagonal, if it is assumed that there is no leakage between the hardware chains. In this case,  $\mathbf{Er}_{1,1}^{T,A}$  is the error introduced by the RF section of the transmitter connected to antenna 1 of transceiver A.

The overall channel from transceiver A to transceiver B is the matrix  $\mathbf{H}^{A \rightarrow B} = \mathbf{Er}^{R,B} \times \mathbf{H} \times \mathbf{Er}^{T,A}$  whereas the the channel from transceiver B to transceiver A is the matrix  $\mathbf{H}^{B \rightarrow A} = \mathbf{Er}^{R,A} \times \mathbf{H}^\dagger \times \mathbf{Er}^{T,B}$ . Clearly,  $\mathbf{H}^{A \rightarrow B}$  is not always the transpose of  $\mathbf{H}^{B \rightarrow A}$ . In most cases,  $\mathbf{H}^{A \rightarrow B}$  and  $\mathbf{H}^{B \rightarrow A}$  are uncorrelated.

Usually, transceiver B estimates the channel  $\mathbf{H}^{A \rightarrow B}$  through reception of pilot symbols. Transceiver B performs the SVD of  $\mathbf{H}^{A \rightarrow B}$ , and uses this results to transmit on the reverse link, i.e.

over the channel  $\mathbf{H}^{B \rightarrow A}$ , resulting in a loss of performance of the system. The overall transmission relationship becomes:

$$\mathbf{y} = (\mathbf{U}^{B \rightarrow A})^* (\mathbf{H}^{B \rightarrow A} \mathbf{V}^{A \rightarrow B} \mathbf{x} + \mathbf{n}). \quad (\text{V.1})$$

Simulation results of a  $M_T = M_R = 4$  SVD system are presented in the following to assess the effect of hardware errors on the system capacity. The MIMO channel is assumed to be i.i.d. Rayleigh distributed.

### V.2.1.1 Amplitude error

To measure the sensitivity of the system to an amplitude error in the hardware chains, the system was simulated over an i.i.d. Gaussian channel with  $\mathbf{Er}^{T,A}$ ,  $\mathbf{Er}^{R,A}$ ,  $\mathbf{Er}^{T,B}$  and  $\mathbf{Er}^{R,B}$  being diagonal matrices with real entries. The system capacity (defined in Section IV.3.2) was obtained.

A system with ideal RF chains corresponds to the system with  $\mathbf{Er}^{T,A} = \mathbf{Er}^{R,A} = \mathbf{Er}^{T,B} = \mathbf{Er}^{R,B} = \mathbf{I}$ . The amplitude error introduced by each hardware chain is considered to be uniform and is measured by the ratio of the maximum amplitude error introduced, with respect to the ideal amplitude, i.e. 1. Therefore, if the system suffers a 20% error, the entries of the diagonal of  $\mathbf{Er}^{T,A}$ ,  $\mathbf{Er}^{R,A}$ ,  $\mathbf{Er}^{T,B}$ ,  $\mathbf{Er}^{R,B}$  are random variables with an amplitude uniformly distributed between 0.8 and 1.2. Results are presented in Fig. V.2. Every transmission and reception chain has to be accurate in the limit of 4% (equivalent to 28 dB) to limit the capacity loss to a maximum of 2 bits.

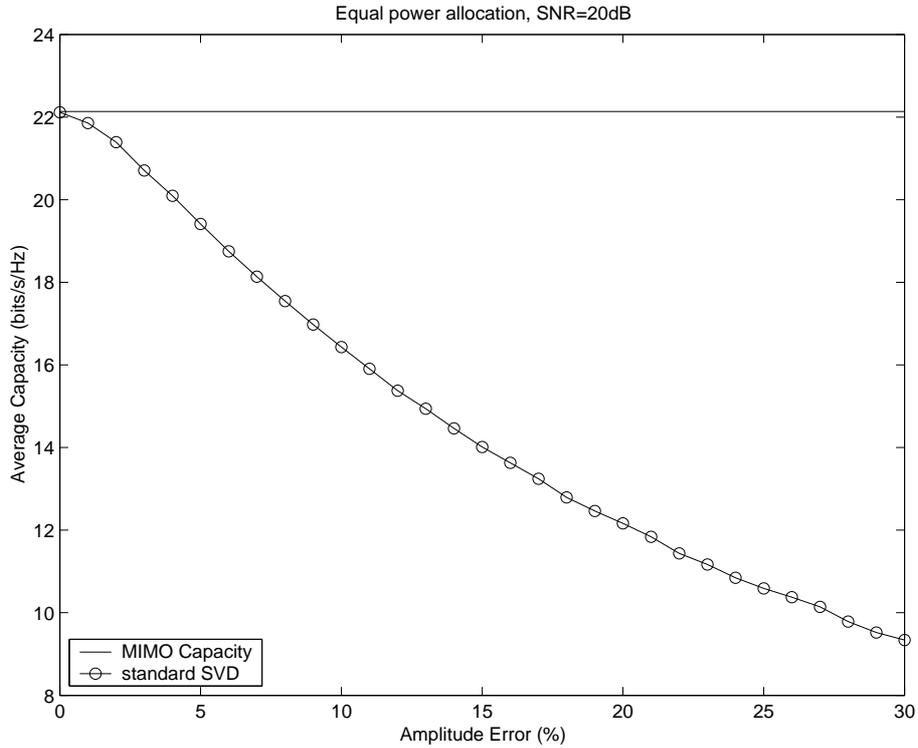


Figure V.2: System capacity of a SVD based system with hardware amplitude errors

### V.2.1.2 Phase error

To measure the sensitivity of the system to a phase error in the hardware chains, the system was simulated with  $\mathbf{E}r^{T,A}$ ,  $\mathbf{E}r^{R,A}$ ,  $\mathbf{E}r^{T,B}$  and  $\mathbf{E}r^{R,B}$  being diagonal matrices with complex entries with amplitude 1.

The phase error introduced by each hardware chain is considered to be uniform, and is measured by the maximum phase error introduced. Therefore, if the system suffers a 20 degrees error, the entries of the diagonal of  $\mathbf{E}r^{T,A}$ ,  $\mathbf{E}r^{R,A}$ ,  $\mathbf{E}r^{T,B}$ ,  $\mathbf{E}r^{R,B}$  are random variables with an amplitude of 1 and a phase uniformly distributed between  $-20$  and  $20$  degrees. System capacity (defined in Section IV.3.2) results are presented in Fig. V.3. The plot shows that the system is particularly sensitive to hardware phase errors. Considering the previous figure of 2 bits of loss of capacity, the phase error has to remain smaller than 5 degrees. This requirement is particularly unrealistic. For example OFDM systems experience flat fading channel on each tone. However the system is wideband, since OFDM uses a number of subcarriers, at different center frequencies, to reach commercially viable data rates. It would be difficult to design transmission and reception hardware chains with so small phase tolerance over the entire bandwidth.

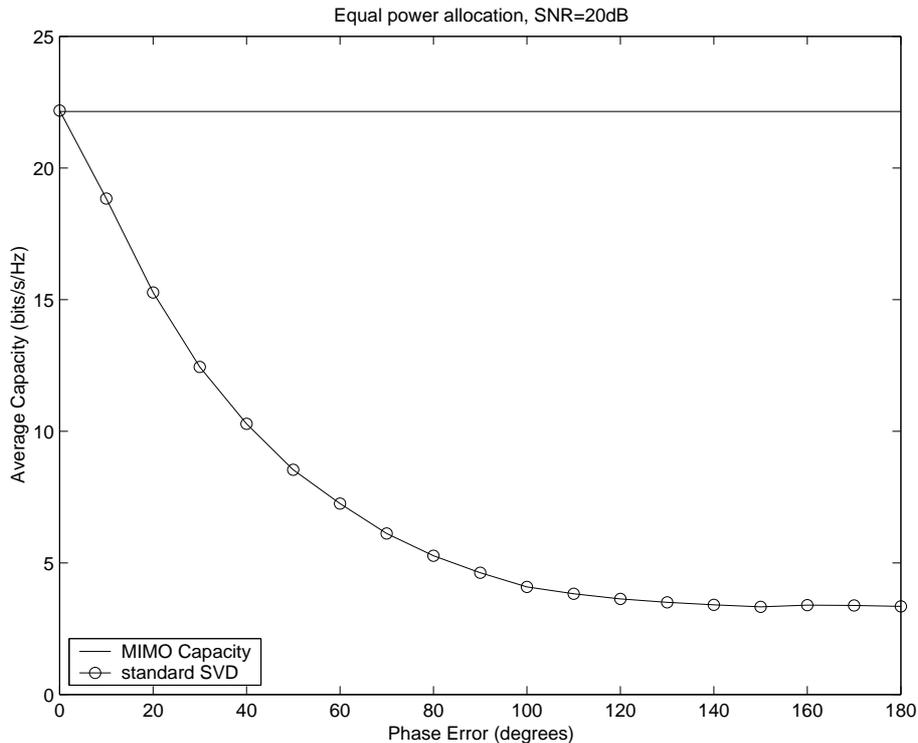


Figure V.3: System capacity of a SVD based system with hardware phase errors  
( $M_T = M_R = 4$ ,  $SNR = 20dB$ , i.i.d. Rayleigh channel).

## V.2.2 Calibration procedure

Considering the results aforementioned, it is unrealistic to rely on hardware design to provide RF transmission chains with performance close to the requirements of a SVD based MIMO system.

The most common solution presented in the literature to combat these effects in antenna arrays, is to conduct a calibration procedure. Calibration usually consists of measuring the distortion introduced on a test signal by the different RF chains taking part in the antenna array. However, this type of procedure usually relies on using a test receiving (transmitting) antenna, to measure (transmit) a signal emitted (received) by each of the transmitting (receiving) antennas [78]. This approach is only viable in the case of a base station, where only one end of the link is using an antenna array and where complexity can be added to the base station in order to increase the performance of the system. However, in the case of wireless modems and full MIMO systems (as opposed to Multiple Input Single Output (MISO) systems, e.g. antenna arrays), this solution is difficult, for obvious reasons of size and cost of the modems.

It is nevertheless possible to perform some type of calibration on the system. As in a standard calibration procedure, it is assumed that the hardware components have fixed amplitude and phase response in time, or that these characteristics drift very slowly, compared to an average transmission duration. Only sparse calibration runs are required in time. Consider a scenario where transceiver A and transceiver B initiate a data transfer by some kind of 'handshaking'. It is assumed that the wireless channel (including the RF chains to transmit and receive the signals) can be perfectly measured through transmission and reception of pilot signals. Also the channel is assumed fixed during the calibration procedure. At the start of a transmission from transceiver A to transceiver B, a calibration

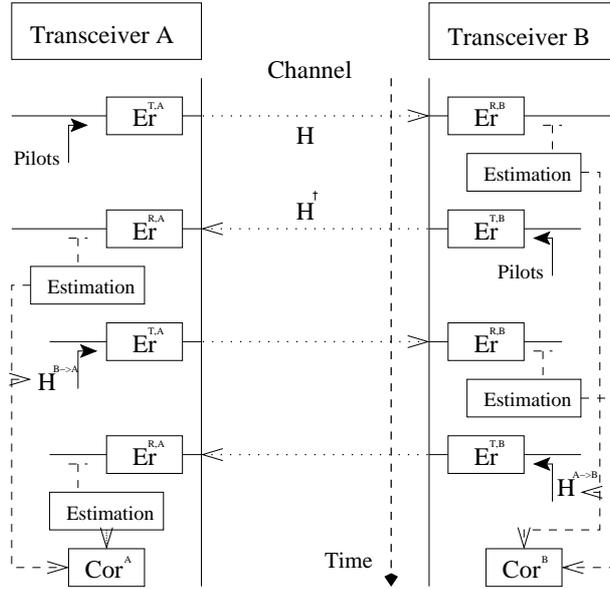


Figure V.4: Calibration procedure.

procedure takes place, as presented in Fig. V.4. Assume transceiver A (transceiver B) has  $M_A$  ( $M_B$ ) antennas. Transceiver A sends pilot symbols to transceiver B, enabling B to measure  $\mathbf{H}^{A \rightarrow B} = \mathbf{E}r^{R,B} \times \mathbf{H} \times \mathbf{E}r^{T,A}$ . Transceiver B immediately replies to A, sending pilot symbols to enable A to measure  $\mathbf{H}^{B \rightarrow A} = \mathbf{E}r^{R,A} \times \mathbf{H}^\dagger \times \mathbf{E}r^{T,B}$ . Transceiver B also transmits back to A the CSI of the forward link  $\mathbf{H}^{A \rightarrow B}$ . Likewise, A transmits back to B the CSI of the return link  $\mathbf{H}^{B \rightarrow A}$ .

The knowledge of both  $\mathbf{H}^{A \rightarrow B}$  and  $\mathbf{H}^{B \rightarrow A}$  is enough to deduce the correction matrices that can make the links symmetrical. The aim of the calibration procedure is to produce two matrices  $\mathbf{C}or^A$  and  $\mathbf{C}or^B$  such that  $\mathbf{H}^{A \rightarrow B} = \mathbf{C}or^B \times (\mathbf{H}^{B \rightarrow A})^\dagger \times \mathbf{C}or^A$ . It is also necessary that the reciprocity of the channel is maintained as the transmission channel changes in time. For example, it is not possible to choose the correction matrices as

$$\mathbf{Cor}^B = \mathbf{I} \text{ and } \mathbf{Cor}^A = ((\mathbf{H}^{B \rightarrow A})^\dagger)^{-1} \mathbf{H}^{A \rightarrow B}.$$

A possible choice to achieve correction is

$$\begin{aligned} \mathbf{Cor}_{i,i}^A &= \frac{\mathbf{H}_{1,i}^{A \rightarrow B}}{\mathbf{H}_{i,1}^{B \rightarrow A}} \quad \forall i \in (1, \dots, M_A) \\ \mathbf{Cor}_{i,j}^A &= 0 \quad \forall i \neq j \end{aligned} \quad (\text{V.2})$$

and

$$\begin{aligned} \mathbf{Cor}_{i,i}^B &= \frac{\mathbf{H}_{i,1}^{A \rightarrow B}}{\mathbf{H}_{1,i}^{B \rightarrow A}} \times \frac{\mathbf{H}_{1,1}^{B \rightarrow A}}{\mathbf{H}_{1,1}^{A \rightarrow B}} \quad \forall i \in (1, \dots, M_A) \\ \mathbf{Cor}_{i,j}^B &= 0 \quad \forall i \neq j \end{aligned} \quad (\text{V.3})$$

### V.2.3 Effects of the calibration

It is not possible to deduce from the two channels' matrices the parameters  $\mathbf{Er}^{T,A}$ ,  $\mathbf{Er}^{R,A}$ ,  $\mathbf{Er}^{T,B}$  and  $\mathbf{Er}^{R,B}$ . Therefore, it is not possible to correct explicitly the errors introduced by the hardware chains.

The first row of  $\mathbf{H}^{A \rightarrow B}$  consists of the first row of  $\mathbf{H}$ , with every element being multiplied by the error introduced by the first receiver chain of B, and the element in each column being multiplied by the error introduced by the corresponding transmitter chain of A. In a similar manner, the first row of  $(\mathbf{H}^{B \rightarrow A})^\dagger$  consists of the first row of  $\mathbf{H}$  with every element being multiplied by the error introduced by the first transmitter chain of B, and the element in each column being multiplied by the error introduced by the corresponding receiver chain of A.

Therefore, it is easily observed that, in both cases, dividing the entries of each column by the entry of the first column produce two effects:

- it cancels the influence of the errors introduced by transceiver B,
- it prevents us from measuring the errors introduced by each transmission (reception) chain of A on its own, but contains information on the ratio of the error introduced by a transmission (reception) chain, compared with the first transmission (reception) chain.

The calibration process ensures that the ratio of errors to signal introduced by transmitter  $i$  and  $j$  of transceiver A is identical to the ratio introduced by receiver  $i$  and  $j$  of transceiver A. The calibration does not correct the errors, it just ensures that they are symmetric on the transmission and reception chains of a transceiver. Therefore, it is self evident that using these fixed correction matrices, symmetry in the system remains, even when the

wireless channel varies in time. It is nonetheless crucial that the channel remains static during the calibration handshaking.

Finally, since the correction explicitly makes the transmission and reception chains symmetric for a transceiver, transceiver A does not need to recalibrate its hardware chains for a transmission to another completely different transceiver C.

#### V.2.4 Choice of calibration matrices

It is important to note that an infinite number of correction matrices that would correct  $(\mathbf{H}^{B \rightarrow A})^\dagger$  into  $\mathbf{H}^{A \rightarrow B}$  exists. Every matrix  $c \times \mathbf{Cor}^A$  and  $\frac{1}{c} \times \mathbf{Cor}^B$ , where  $c$  is any non-zero complex number, provides the adequate correction.

There is no advantage in choosing the phase of the correction matrices. However, there is a significant advantage in modifying the amplitude of the correction matrices: it may be interesting to transfer some workload from the transmitter to the receiver chains. This feature may be helpful to avoid saturation of either the transmitter or the receiver amplifiers.

### V.3 SVD-TDD system and singular values swapping

It is possible to obtain the CSI at both ends of the link when the channel is reciprocal as presented in Section V.1. This relies on the following assumptions:

- pilot symbols provide perfect channel estimation,
- the channel is static.

Neither of these assumptions applies to practical systems. The complete system, including errors in the channel estimation, over a TDD time varying channel is presented in Fig. V.5.

Consider the transmission from transceiver A to transceiver B at the time slot  $(k+1)t$ . The precoding matrix is derived from the CSI available at transceiver A, i.e. the CSI estimated from pilot symbols  $PB_1$  at time slot  $kt$ . This precoding matrix is denoted as  $\hat{\mathbf{V}}(kt)$ . The decoding matrix, denoted  $\hat{\mathbf{U}}^*((k+1)t)$ , corresponds to the CSI estimated from the reception at time slot  $(k+1)t$  of pilot symbols  $PA_1$ . The complete transmission equation becomes

$$\mathbf{y} = \hat{\mathbf{U}}^*((k+1)t)(\mathbf{H}\hat{\mathbf{V}}(kt)\mathbf{x} + \mathbf{n}), \quad (\text{V.4})$$

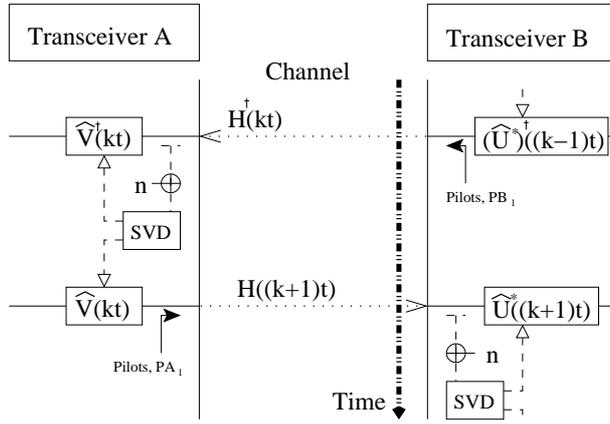


Figure V.5: SVD transmission over a TDD channel with channel estimation error.

where the precoding and decoding matrices are neither matched with the channel matrix nor with each other.

The impact of the estimation errors (errors due to noise in the estimation or errors due to the time-varying nature of the channel) can be analysed according to matrix perturbation theory, assuming that the channel matrix  $\mathbf{H}$  has been estimated incorrectly as  $\hat{\mathbf{H}} = \mathbf{H} + \mathbf{N}_{\text{est}}$ .

### V.3.1 Matrix perturbation theory, singular values

Several powerful results exist concerning the perturbation of singular values. The main result is probably the following theorem, stating that the singular values of a matrix are perfectly conditioned, i.e. no singular value can move more than the norm of the perturbations.

**Theorem 7.** *Mirsky* Let  $\mathbf{H}$  and  $\hat{\mathbf{H}}$  be matrices of the same dimensions with singular values

$$\begin{aligned} \Sigma_{1,1} &\geq \Sigma_{2,2} \geq \dots \geq \Sigma_{M,M}, \\ \hat{\Sigma}_{1,1} &\geq \hat{\Sigma}_{2,2} \geq \dots \geq \hat{\Sigma}_{M,M}. \end{aligned} \tag{V.5}$$

Then for any unitary invariant norm  $\|\cdot\|_u$ <sup>1</sup>

$$\|\text{diag}(\hat{\Sigma}_{i,i} - \Sigma_{i,i})\|_u \leq \|\hat{\mathbf{H}} - \mathbf{H}\|_u. \tag{V.6}$$

The proof of the theorem is given in [79]. Two immediate consequences are

<sup>1</sup>A unitary invariant norm  $\|\cdot\|_u$  is a norm such that for all matrix  $\mathbf{H}$  (dimension  $M \times N$ ) and all unitary matrices  $\mathbf{U}$  (dimension  $M \times M$ ) and  $\mathbf{V}$  (dimension  $N \times N$ ),  $\|\mathbf{UH}\|_u = \|\mathbf{H}\|_u$  and  $\|\mathbf{HV}\|_u = \|\mathbf{H}\|_u$ .

**Theorem 8.** Weyl

$$|\hat{\Sigma}_{i,i} - \Sigma_{i,i}| \leq \|\mathbf{N}_{est}\|_2, \quad i = 1, \dots, M, \quad (\text{V.7})$$

and

**Theorem 9.** Mirsky

$$\sqrt{\sum_i (\hat{\Sigma}_{i,i} - \Sigma_{i,i})^2} \leq \|\mathbf{N}_{est}\|_F \quad (\text{V.8})$$

as shown in [79], where  $\|\cdot\|_2$  is the matrix 2-norm<sup>2</sup> and  $\|\cdot\|_F$  is the Frobenius norm<sup>3</sup>.

It is notable that:

- there is no restriction on the size of the error,
- the ordering of the singular values by magnitude provides a natural pairing between the singular values of the channel matrix and the estimation of the channel matrix.

### V.3.2 Matrix perturbation theory, singular vectors

The perturbation of singular vectors is difficult to analyse or bound, due to several reasons:

- arbitrarily small perturbations can completely change singular vectors,
- it is difficult to define a meaningful distance between vectors.

#### V.3.2.1 Example of catastrophic perturbation

Consider the matrix

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 + \epsilon \end{pmatrix}, \quad (\text{V.9})$$

with precoding matrix

$$\mathbf{V} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (\text{V.10})$$

If  $\mathbf{H}$  is estimated as

$$\hat{\mathbf{H}} = \begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix}, \quad (\text{V.11})$$

---

<sup>2</sup>The matrix 2-norm is defined as the largest singular value of the matrix.

<sup>3</sup>The Frobenius norm of the matrix  $\mathbf{H}$  is defined as the square root of the sum of the absolute squares of its elements, i.e.  $\|\mathbf{H}\|_F = \sqrt{\sum_i \sum_j h_{i,j}^2}$ .

then the corresponding precoding matrix is

$$\hat{\mathbf{V}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad (\text{V.12})$$

without assumptions on the size of the perturbation  $\epsilon$ . Therefore it is impossible to derive perturbation bounds on singular vectors.

Consider such a catastrophic perturbation, the transmitter transmits the first data stream on the first antenna and the second data stream on the second antenna, according to (V.10). The receiver tries to receive the first stream by adding the symbol sent on the first antenna and the symbol sent on the second antenna since the receiver believes that both data streams were transmitted according to (V.12). Therefore, in the case of catastrophic perturbation, the smallest channel estimation error results in a bit error rate of 0.5. This example, combined with the results of Section IV.4, highlights the fact that the primary problem is mismatch of  $\mathbf{U}$  and  $\mathbf{V}$ , rather than mismatch of the channel to  $\mathbf{U}$  and  $\mathbf{V}$  when the latter are matched.

### V.3.2.2 Subspaces perturbation theory

Several results on subspaces perturbation theory presented in [79] are mentioned in the following.

To analyse the perturbation of singular subspaces, it is necessary to define a meaningful distance between two vectors. The distance usually used in perturbation theory is meaningless for singular vectors. E.g. consider  $\mathbf{V}$  a precoding matrix, estimated as  $\hat{\mathbf{V}} = -\mathbf{V}$ . Obviously the transmission eigenmodes have not been perturbed, whereas  $\|\mathbf{V} - \hat{\mathbf{V}}\|_F$  is large.

A meaningful distance between two singular subspaces is given by the canonical angles between the two subspaces [79]. Bounds on the perturbation of singular subspaces can be found but the accuracy of the bound depends on the inverse of the distance between singular values. The bound is not applicable when two singular values are equal.

An expansion of the perturbations of the singular value of a matrix is also derived in [79], however the expansion only applies to the subspaces with non-zero singular values of multiplicity one. It is clear, with this result, that the transmitter and receiver cannot unambiguously pair input and output singular subspaces simply on the basis of size ordering of measured singular values, and that significant mutual interference among such subspaces may result.

### V.3.3 Probability of robust transmission

From the previous results, it is clear that the SVD transmission system cannot operate properly when some of the channel matrix singular values are zero or of multiplicity higher than one.

If some of the singular values are zero or multiplicity higher than one:

- the SVD of the channel matrix cannot be determined uniquely,
- the perturbation of the corresponding singular subspaces cannot be bounded.

Therefore, it is necessary to determine the probability for the channel matrix to have singular values that are non-zero and of multiplicity one. In Section IV.4.2, it was mentioned that the probability of these events is one. This result is only true if it is assumed that the CSI is known with infinite precision. However, in practical systems, the CSI is only estimated through measurement of pilot symbols. Noise in the estimation, estimation delay and limitations of the system (e.g. quantization noise) can limit the channel estimation accuracy.

In such a case the singular values  $\Sigma_{i,i}$  are known up to a confidence interval. Two singular values are always distinct when their respective confidence intervals are not joined. Furthermore, a singular value whose confidence interval includes zero might not correspond to an actual transmission eigenmode. Therefore an SVD transmission system needs to determine the singular values of the channel matrix but also the confidence interval length, i.e. how precisely these singular values are known.

Applying Theorem 8, it is possible to obtain a confidence interval:

$$|\hat{\Sigma}_{j,j} - \Sigma_{j,j}| \leq \|\mathbf{N}_{\text{est}}\|_2. \quad (\text{V.13})$$

Furthermore, applying Theorems 8 and 9,  $\hat{\Sigma}_j$  and  $\hat{\Sigma}_i$  correspond to two non-equal singular values if

$$|\hat{\Sigma}_{j,j} - \hat{\Sigma}_{i,i}| \geq \min(2 \times \|\mathbf{N}_{\text{est}}\|_2, \sqrt{2} \times \|\mathbf{N}_{\text{est}}\|_F), \quad (\text{V.14})$$

where the second term is derived from Theorem 9 as explained in the following. From Theorem 9

$$\sqrt{(\hat{\Sigma}_{i,i} - \Sigma_{i,i})^2 + (\hat{\Sigma}_{j,j} - \Sigma_{j,j})^2} \leq \sqrt{\sum_k (\hat{\Sigma}_{k,k} - \Sigma_{k,k})^2} \leq \|\mathbf{N}_{\text{est}}\|_F \quad (\text{V.15})$$

and for all  $d_1, d_2$  real positive numbers

$$\sqrt{(d_1 + d_2)^2} \leq \sqrt{2}\sqrt{d_1^2 + d_2^2}. \quad (\text{V.16})$$

Therefore

$$|\hat{\Sigma}_{i,i} - \hat{\Sigma}_{j,j}| > |\hat{\Sigma}_{i,i} - \Sigma_{i,i}| + |\hat{\Sigma}_{j,j} - \Sigma_{j,j}| \quad (\text{V.17})$$

if

$$|\hat{\Sigma}_{i,i} - \hat{\Sigma}_{j,j}| > \sqrt{2}\|\mathbf{N}_{\text{est}}\|_F. \quad (\text{V.18})$$

Obviously the length of the confidence intervals depends on the realization of the estimation noise. One solution is to apply a probabilistic confidence interval: the interval that guarantees to contain the perturbed singular value  $x\%$  of the time. Another solution is to use the average confidence interval.

Given a confidence interval, it is straightforward to deduce the probability for all singular values to be non-zero and of multiplicity one, by just integrating  $\text{pdf}_{\underline{\Sigma}}$  over the domain corresponding to non-overlapping confidence intervals. Results are shown in Fig. V.6 where  $\text{SNR}_{\text{conf}}$  refers to the ratio between the power of an element of the channel matrix and the square of the size of the confidence interval. The case of distinct singular values with all singular values being non-zero corresponds to the curve 'well-separated'.

It is possible to reduce the requirements on the system following Section IV.4.2.3, i.e. non-zero eigenmodes are required to be of multiplicity one and zero eigenmodes are discarded for transmission. The curve labelled 'Zero eigenmode' corresponds to the probability for some eigenmodes to have a negligible gain while the other eigenmodes correspond to singular values of multiplicity one. Finally, the curve labelled 'Correct SVD' corresponds to the probability for the SVD transmission system to be robust, i.e. non-zero eigenmodes correspond to singular values with a multiplicity of one. The curves are worst case because of the inequalities in (V.13) and (V.18).

Wireless channels face fading in time, i.e. are time varying. If two singular values are equal to each other at one point in time, they are likely to separate later, due to channel fading. As explained earlier, when two singular values of the channel are equal the system is not robust since a small variation of the channel might result in a large variation of the precoding and decoding matrices. Therefore, it is of interest to determine the probability of occurrence of such an event which is referred to as 'singular value crossing' in the following. The probability of 'singular value crossing' is investigated through computer simulation in the next Section.

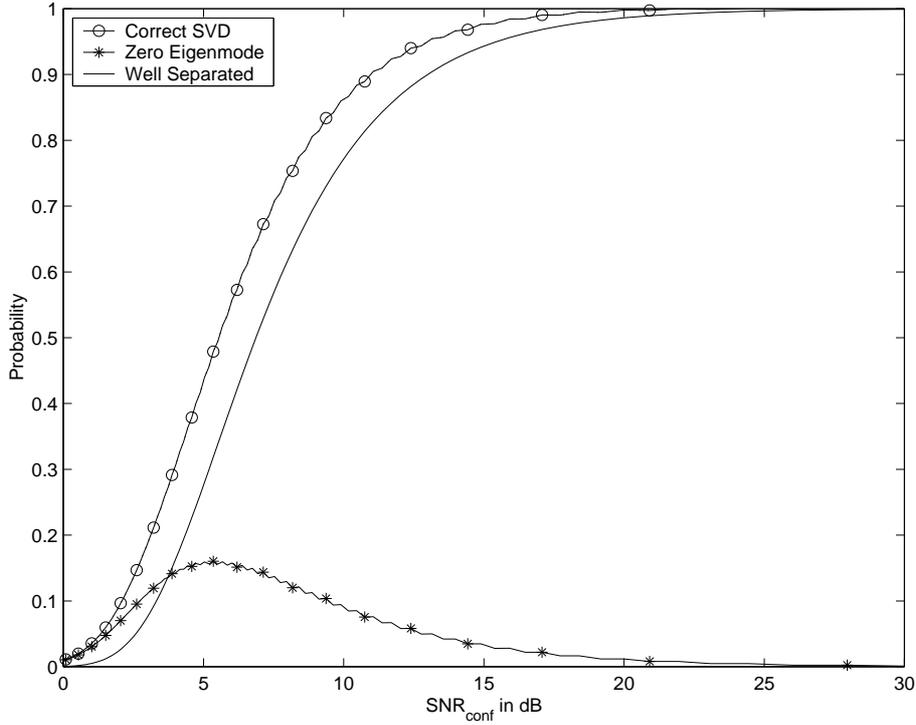


Figure V.6: Probability of robust transmission vs amplitude of the estimation confidence interval,  $M_T = M_R = 4$ , i.i.d. channel of complex Gaussian entries with mean 0 and variance 1.

### V.3.4 Singular value crossing

Consider a time-varying 4x4 MIMO channel treated as 16 independent SISO time-varying Rayleigh channels. Each SISO channel follows the Jake's time-varying channel model [80, 81], i.e. the autocorrelation of the time-varying channels is  $J_0(2\pi F_d \delta t)$ , where  $F_d$  denotes the Doppler frequency,  $\delta t$  the time delay and  $J_0(\cdot)$  is the zeroth-order Bessel function of the first kind. In simulations involving  $F_d \delta t$ , a value of  $F_d \delta t = 0.038$  was assumed. This represents a realistic value expected in the 802.11a and Hiperlan 2 standards with a Doppler frequency of 38 Hz (corresponding to a terminal velocity of  $2 \text{ m.s}^{-1}$  for a carrier frequency of 5.725 GHz) and a typical time-duplex delay of 1 ms. The amplitudes of the singular values of the time-varying MIMO channel are plotted versus time in Fig. V.7. A 'singular value crossing' is exhibited.

A 'singular value crossings' corresponds to the event where  $\Sigma_{i,i} - \Sigma_{i+1,i+1} = 0$ . Determining the probability of this event is similar to the well known issue of determining the zero-crossing rate of a random variable. Consider a real valued random variable  $\underline{x}(kT_s)$ , the zero crossing rate of  $\underline{x}$  for  $k \in [0, N]$  is defined as the number of indexes  $k$  such that

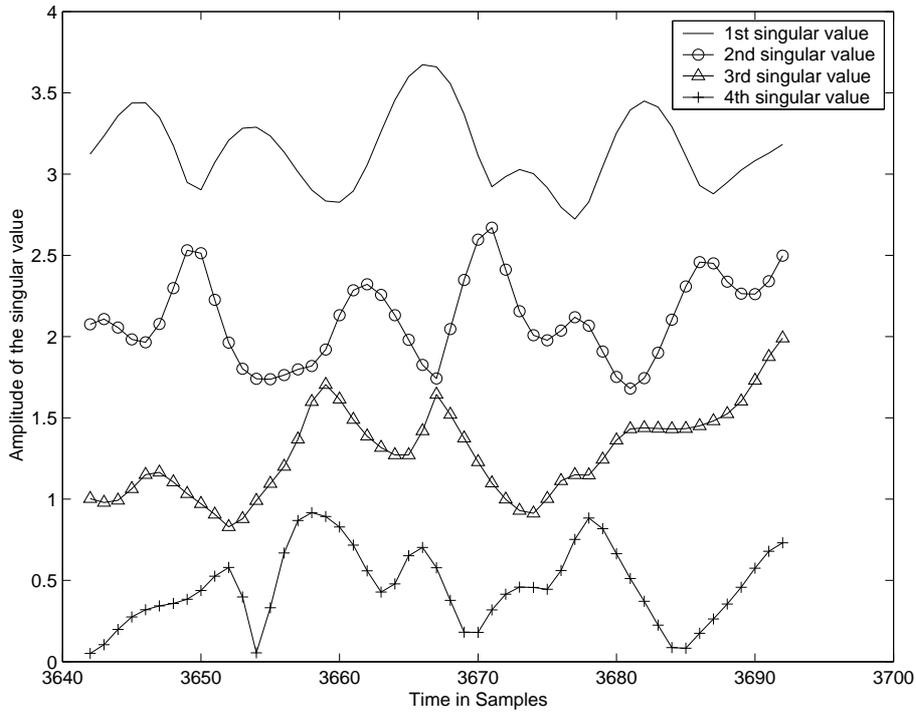


Figure V.7: A clear singular value crossing between the second and third singular values at time sample 3667.

$x(kT_s)x((k+1)T_s)$  is negative. Zero-crossing rate theory is used in communications, e.g. to determine the velocity of a mobile [82]. However, determining the level crossing rate of singular values of a matrix with random entries is difficult, due to the following reasons:

- The singular values of a matrix are highly non-linear and non-monotonous functions of the entries of the matrix. Singular values are determined, in practice, through recursive algorithms, not analytical functions.
- The SVD of the matrix involves a sorting function to sort the singular values in decreasing order. The sorting function implies that, strictly speaking,  $\Sigma_{i,i} > \Sigma_{i+1,i+1} \forall i$ , i.e. singular values are never crossing.
- Level crossing rate have been determined for Gaussian random variables and simple functions of Gaussian random variables such as monotone transformation of random processes [83] or mixtures and products of Gaussian processes [84]. However, to the knowledge of the author, the level crossing rate of the result of such a complex function as the SVD on Gaussian random variables has not been determined in the literature.

It is unclear whether the level crossing rate of the singular values of a matrix with Gaussian entries can be determined analytically. Due to the lack of results in the literature, an ad-hoc criterion is introduced in the following to estimate through simulations the level crossing rate of singular values.

The criterion chosen to define 'singular value crossings' follows.  $\Sigma_{i,i}$  and  $\Sigma_{i+1,i+1}$  are crossing each other at  $t_c$  if:

- $|\Sigma_{i,i}(t_c) - \Sigma_{i+1,i+1}(t_c)| \leq 0.1,$
- $|\frac{\partial \Sigma_{i,i}}{\partial t}(t < t_c) - \frac{\partial \Sigma_{i,i}}{\partial t}(t > t_c)| \geq |\frac{\partial \Sigma_{i,i}}{\partial t}(t < t_c) - \frac{\partial \Sigma_{i+1,i+1}}{\partial t}(t > t_c)|,$
- $|\frac{\partial \Sigma_{i+1,i+1}}{\partial t}(t < t_c) - \frac{\partial \Sigma_{i+1,i+1}}{\partial t}(t > t_c)| \geq |\frac{\partial \Sigma_{i+1,i+1}}{\partial t}(t < t_c) - \frac{\partial \Sigma_{i,i}}{\partial t}(t > t_c)|.$

This criterion corresponds to the intuitive idea of observing two curves that cross each other at one point: their value is the same at the crossing point and the gradient of both curves remains approximately constant around the crossing point. The choice of the value 0.1 is empirical and was chosen by the author through a trial and error process on the simulation results.

$50 \times 16384$  time samples of a  $M_R = 4, M_T = 4$  time-varying MIMO channel (16 i.i.d. SISO channels, Jake's fading,  $F_d \delta t = 0.038$ ) were simulated. 622 crossings were detected between the first and second singular values, 921 crossings between the second and the third and 1000 between the third and the fourth. This represents a total of 2543 crossings, for  $50 \times 16384$  time samples. At  $F_d \delta t = 0.038$ , there is less than 0.32% probability that a 'singular value crossing' occurs from one sample to the next. A 'singular value crossing' is very unlikely under the given assumptions.

Moreover, the frequency of 'singular value crossings' is independent of the sampling rate for high sampling rate, i.e. for a sampling rate much higher than the 'singular value crossing' rate. The results above can be restated as: at a doppler frequency of  $F_d = 0.038$ , if the channel is sampled every second ( $T_s = 1$ ), there will be on average a crossing approximately every 322 samples (seconds). The frequency of the crossings is  $F_{svc} = 1/322\text{Hz}$ . Therefore, an empirical general relationship can be deduced:  $F_{svc}/F_d = 0.08$ . This relationship is applicable for any  $M_T = 4, M_R = 4$  i.i.d. Rayleigh fading channel following Jake's fading.

The results of the simulation also indicate that the first and second singular values cross each other less often than the others. Due to bits (and possibly power) allocation,

the channel corresponding to the first singular value is the most important: it is the sub-channel carrying the most information. This sub-channel is less subject than the others to 'singular value crossings', offering a reliable, high SNR, channel for transmission.

### V.3.5 Subspace swapping

It has been shown in Section V.3.4 that 'singular value crossings' are rare, but do occur. 'Singular value crossings' can affect the throughput of a SVD-based system. If the transmitter measures the channel before the 'singular value crossing', and the receiver after the 'singular value crossing', the result of the SVD performed on their respective channel matrices might be very different, leading to a burst of errors in the transmission.

Therefore, it is of interest to determine the behaviour of the singular subspaces before and after a 'singular value crossing'. Following the SVD of the channel matrix, the first column of the  $\mathbf{V}$  matrix and the first column of the  $\mathbf{U}$  matrix are respectively the input subspace and output subspace corresponding to the first eigenvalue.

To determine the behaviour of the subspaces around a 'singular value crossing' point, the autocorrelation of the first input subspace (i.e. the autocorrelation of the the first column of the  $\mathbf{V}$  matrix) is plotted, as well as the cross-correlation between the first and second input subspaces (i.e. the cross-correlation between the first and second columns of the  $\mathbf{V}$  matrix). Results (corresponding to the 'singular value crossing' presented in Fig. V.7) are presented in Fig. V.8 for the input subspace, and Fig. V.9 for the output subspace.

The autocorrelation stays high until the crossing point and then drops suddenly. On the contrary the cross-correlation is very low, but increases suddenly after the crossing point. This clearly indicates that the subspaces are linked with the singular value, and when the first singular value becomes lower than the second, the corresponding columns of the  $\mathbf{V}$  matrix swap position at the same time as the singular values swap position.

The swapping of the subspaces is simply due to the ordering function performed during the SVD. If the transceiver measures the channel before the crossing and the receiver after the crossing, the information sent on sub-channel 1 is received on sub-channel 2, and the information sent on sub-channel 2 is received on sub-channel 1.

Supposing that it is possible to detect 'singular value crossings', either the transmitter or the receiver might be able to handle it by simply swapping the sub-channels at either the transmitter or the receiver. This result is of great interest for the design of a wireless SVD-

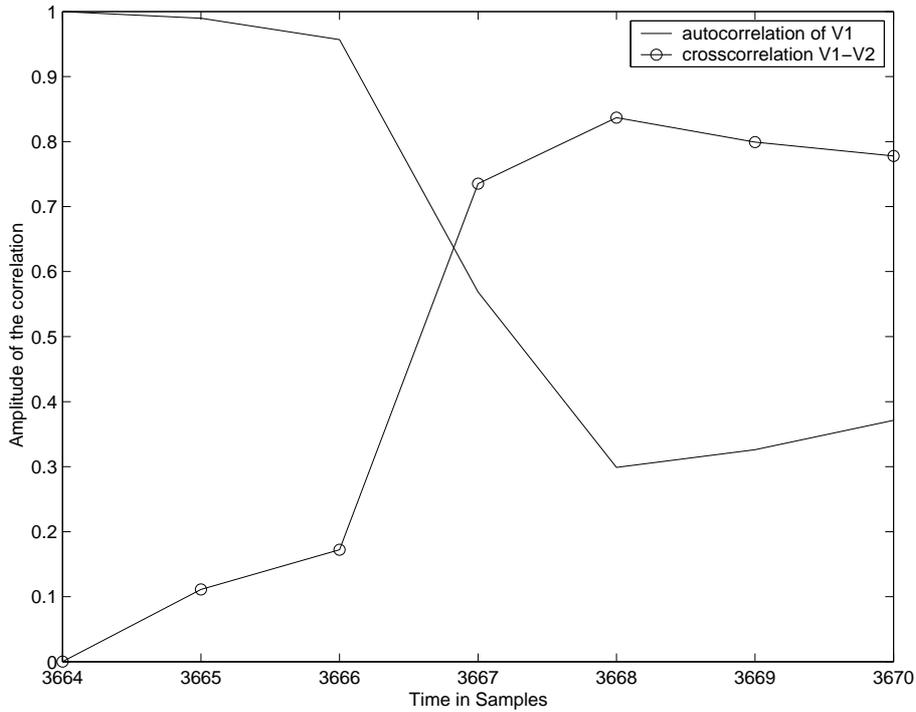


Figure V.8: Correlation of the input subspaces

based system: singular subspaces are related before and after a 'singular value crossing'.

'Singular subspace swapping' can be used as a criterion to detect a 'singular value crossing'. It is possible to determine whether two singular values become close to each other with or without crossing by checking the amplitude of the autocorrelation and cross correlation of both the input and output subspaces. If both of them display a crossing at a point where the singular values are close in amplitude, then a 'singular value crossing' is detected. This new criterion produces the following results:  $50 \times 16384$  time samples were simulated. 823 crossings were detected between the first and second singular values, 1149 crossings between the second and the third and 1065 between the third and the fourth. This represents a total of 3037 crossings, for  $50 \times 16384$  time samples and  $F_{svc}/F_d = 0.1$ . Obviously this criterion indicates a higher probability of singular value swapping. However it corresponds better to the reality of a MIMO-SVD transmission system and therefore is more representative of the expected behaviour of the system.

### V.3.6 Correction of singular value crossing

A practical way to correct the effect of singular value crossing involves transmitting a second set of pilots through the precoding matrix. Therefore the receiver can correct any

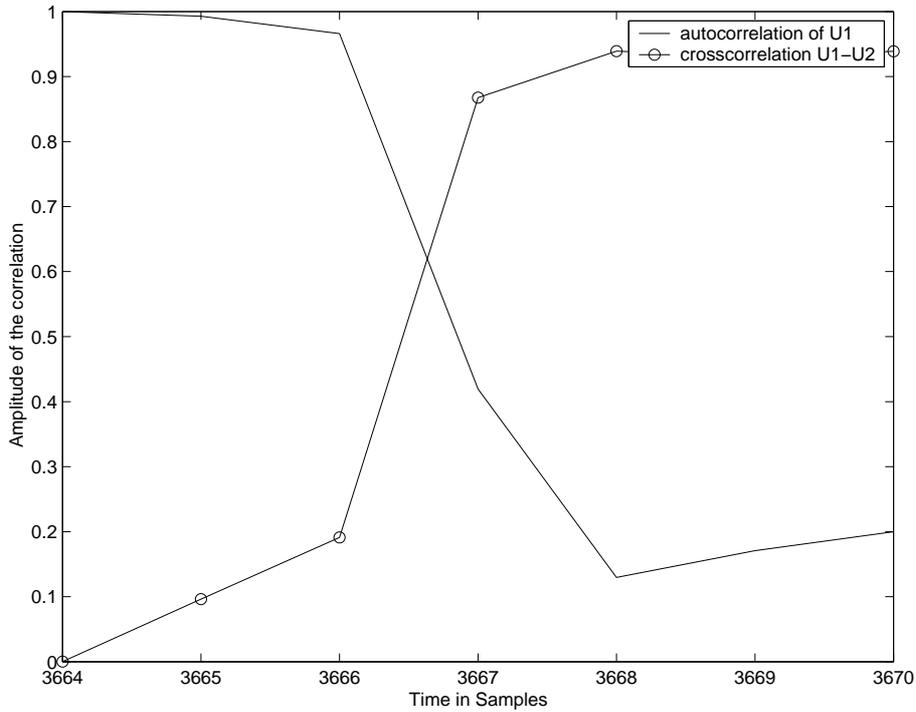


Figure V.9: Correlation of the output subspaces

mismatch between transmitting and decoding matrices. This method is, however, not limited to the correction of 'singular value crossings' as explained in the following Section.

## V.4 SVD architecture modified for TDD environment

As presented in Section V.3, the precoding and decoding matrices can be completely modified when the channel is perturbed. However, these catastrophic events, referred to as 'singular value crossings' are rare.

Regardless of 'singular value crossings', the precoding matrix of SVD-TDD systems is generally not matched with the channel matrix (see (V.4)). This leads to a degradation of the system capacity (defined in Section IV.3.2) of SVD-TDD systems. A severe <sup>4</sup> drop in capacity results.

The delay between channel measurement in one direction and transmission in the other may vary due to time slot allocation in the TDD protocol. Furthermore, this delay may not be identical for symbols inside a frame, depending on whether the symbols are at the

---

<sup>4</sup>For a SVD-MIMO system with 4 transmitting and 4 receiving antennas, over a flat fading i.i.d. channel, at SNR=20dB, assuming Jake's fading, the capacity drops from more than 21 bits/s/Hz for perfect CSI to less than 8 bits/s/Hz for  $F_d\delta t > 0.1$ , see Section V.4.3.2

start or the end of the frame. A conservative approach in assessing the performance of the system, identifies  $\delta t$  with the worst case delay. In the following, the capacity of a system with fixed  $F_d\delta t$  is considered and then the capacity is analysed for varying  $F_d\delta t$ . It is considered in Section V.4 that pilot symbols provide perfect CSI, i.e. the channel estimation noise is considered negligible.

## V.4.1 Correction of the CSI impairment

### V.4.1.1 Proposed Architecture

System capacity drops when the transmitter filters data through a  $\mathbf{V}(t - \delta t)$  matrix, which corresponds to an outdated CSI. However, the receiver can mitigate the effect of this impairment through processing of the received signals. The receiver can be provided with information on the outdated  $\mathbf{V}$  matrix ( $\mathbf{V}(t - \delta t)$ ) if the transmitter sends an additional set of pilot signals  $PA_2$  through  $\mathbf{V}(t - \delta t)$ , the channel and  $\mathbf{U}(t)$  (Fig. V.10). A correction

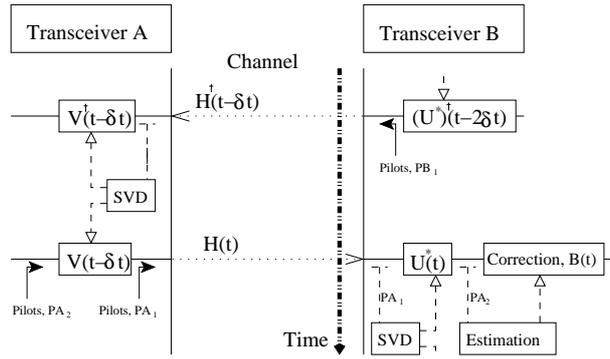


Figure V.10: SVD transmission over a TDD channel with channel estimation error recovery pilot symbols  $PA_2$ .

matrix,  $\mathbf{B}(t)$  can then be derived to partially compensate the effect of the outdated  $\mathbf{V}(t - \delta t)$  matrix.

$$\mathbf{y} = \mathbf{B}(t)\mathbf{U}(t)(\mathbf{H}(t)\mathbf{V}(t - \delta t)\mathbf{x} + \mathbf{n}) \quad (\text{V.19})$$

The requirement of two sets of pilots in the one direction ( $PA_1$  to calculate  $\mathbf{U}(t)$  and  $PA_2$  to calculate  $\mathbf{B}(t)$ ) adds to the signaling overhead in the system. A new architecture is proposed that does not use the set of pilot signals  $PA_1$ . In the new architecture, one set of pilot symbols ( $PA_2$ ) is sent from the transmitter through  $\mathbf{V}(t - \delta t)$  and the channel  $\mathbf{H}(t)$ , to the receiver (Fig. V.11). In this situation, the matrix  $\mathbf{H}(t)\mathbf{V}(t - \delta t)$  is known

at the receiver and linear processing can be applied:

$$\mathbf{y} = \mathbf{C}(t)(\mathbf{H}(t)\mathbf{V}(t - \delta t)\mathbf{x} + \mathbf{n}) \quad (\text{V.20})$$

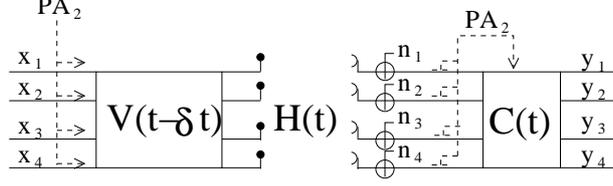


Figure V.11: Correction of the CSI Impairment

where  $\mathbf{C}(t)$  is the matrix of the linear correction. Since  $\mathbf{U}(t)$  is unitary, the performance of equations (V.19) and (V.20) is identical for a given selection criteria of  $\mathbf{C}(t)$  and  $\mathbf{B}(t)$ . Therefore, from here on, we only consider the system that uses one set of pilots. Two criteria for the selection of  $\mathbf{C}(t)$  are considered: zero-forcing (ZF) and minimum mean square error (MMSE). The ZF correction is applied by setting

$$\mathbf{C}(t) = (\mathbf{H}(t)\mathbf{V}(t - \delta t))^+. \quad (\text{V.21})$$

This method forces the interference to zero, but can lead to noise enhancement. It is referred to in the remainder of the paper as a 'SVD+ZF' system.

The MMSE correction is applied by setting

$$\mathbf{C}(t) = \mathbf{Q}(\mathbf{H}(t)\mathbf{V}(t - \delta t))^* (\mathbf{I}_{M_R} + (\mathbf{H}(t)\mathbf{V}(t - \delta t))\mathbf{Q}(\mathbf{H}(t)\mathbf{V}(t - \delta t))^*)^{-1}. \quad (\text{V.22})$$

It is referred to in the remainder of the paper as a 'SVD+MMSE' system. It is assumed that  $\mathbf{Q}$  is a real diagonal matrix, i.e. the input symbols are independent and the entries of  $\mathbf{Q}$  represent the power allocated to the sub-channels. The receiver needs to have prior knowledge of the power allocation applied by the transmitter to perform MMSE decoding.

From an implementation perspective the new architecture only requires the set of pilots  $PA_2$ . This applies to the reverse link (transceiver B to transceiver A), where  $(\mathbf{U}(t)^*)^\dagger$  is required for transmission because  $(\mathbf{U}(t)^*)^\dagger$  can be obtained from the SVD of  $\mathbf{H}(t)\mathbf{V}(t - \delta t)$ , measured by pilots  $PA_2$  only. The fact that  $\mathbf{V}(t - \delta t)$  does not match the channel has no effect on the calculated value of  $\mathbf{U}(t)$ , since any error in  $\mathbf{V}(t - \delta t)$  only affects the input subspaces ( $\mathbf{V}$  is a unitary matrix).

It should also be noted that the so-called 'SVD transmission' system over a reciprocal channel performs the eigenvalue decomposition of the channel, but does not need to perform a full SVD. Transceiver B requires knowledge of  $\mathbf{U}(t)$  to receive data transmitted over  $\mathbf{H}(t)$ , and  $(\mathbf{U}(t)^*)^\dagger$  to transmit data over  $(\mathbf{H}(t + \delta t))^\dagger$ , but does not require knowledge of  $\mathbf{V}(t)$ .

The new architecture solves some implementation issues of the SVD architecture while keeping the desirable features of SVD architectures that have been outlined in Section IV.2.1.

## V.4.2 Comparison with reference systems

The system capacity (see Section IV.3.2) of the 'SVD+ZF' and 'SVD+MMSE' systems was compared with the following reference systems:

- The theoretical capacity of the channel (named 'MIMO Capacity'), which provided the upper bound of the systems capacities.
- A standard SVD-based system (named 'Standard SVD') which provided a reference for the capacity gains provided by both 'SVD+ZF' and 'SVD+MMSE' systems.
- A MIMO system with no precoding which included filtering of the received symbols (Fig. V.12).

### V.4.2.1 Theoretical capacity

The theoretical capacity of the channel is independent of  $F_d \delta t$ . It should be noted that two different capacity formulas exist, depending on whether the CSI is available at the transmitter or not. Systems implementing waterfilling were compared with the theoretical capacity of the channel with CSI at the transmitter [12]. Systems with equal power allocation were compared with the theoretical capacity of the channel without CSI at the transmitter [15].

### V.4.2.2 Standard SVD system

A standard SVD system refers to a system based on (V.4) (named 'Standard SVD') which included CSI impairment.

### V.4.2.3 MIMO systems without precoding

A MIMO system with no precoding and receiver filtering (Fig. V.12), can be represented by the transmission equation:

$$\mathbf{y} = \mathbf{D}(t)(\mathbf{H}(t)\mathbf{x} + \mathbf{n}). \quad (\text{V.23})$$

MIMO systems with no precoding have a low complexity: the SVD does not have to be

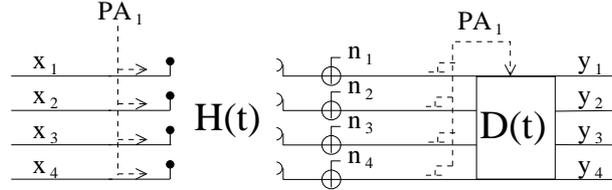


Figure V.12: Transmission system with no precoding

performed and the pilots  $PB_1$  are not needed. The ZF linear receiver (system named 'ZF Alone', [85]) was simulated by setting

$$\mathbf{D}(t) = (\mathbf{H}(t))^+ \quad (\text{V.24})$$

This method forces the interference to zero, but can lead to noise enhancement. The MMSE linear receiver (system named 'MMSE Alone') was simulated by setting

$$\mathbf{D}(t) = \mathbf{Q}\mathbf{H}(t)^* (\mathbf{I}_{M_R} + \mathbf{H}(t)\mathbf{Q}\mathbf{H}(t)^*)^{-1}. \quad (\text{V.25})$$

In both cases (ZF and MMSE), equal power allocation was used ( $\mathbf{Q} = \mathbf{I}$ ), because the CSI was not available at the transmitter. Their capacity is unaffected by  $F_d\delta t$ .

### V.4.3 Simulation and results

To validate the proposed architecture described in Section V.4.1 a  $M_T = M_R = 4$  MIMO system was simulated (16 i.i.d. SISO channels, Jake's fading,  $F_d\delta t = 0.038$ ). The time varying MIMO channels were simulated for independent variables SNR and  $F_d\delta t$ .

The simulated systems and their corresponding transmission equations are shown in Table V.4.3. The 'system capacity' of each system, for a given channel matrix  $\mathbf{H}(t)$ , is obtained from its transmission equation as explained in IV.3.2. The average 'system capacity' is obtained by repeating the process over a large number of realizations of the channel and averaging the results.

System	Transmission Equation
'SVD+MMSE'	(V.20)
'SVD+ZF'	(V.20)
'Standard SVD'	(V.4)
'MMSE Alone'	(V.23)
'ZF Alone'	(V.23)

Table V.1: Simulated systems equation references

#### V.4.3.1 System capacity with equal power allocation for varying SNR

The system capacity (see Section IV.3.2) of the systems against SNR with  $F_d\delta t = 0.038$  are shown in (Fig. V.13).

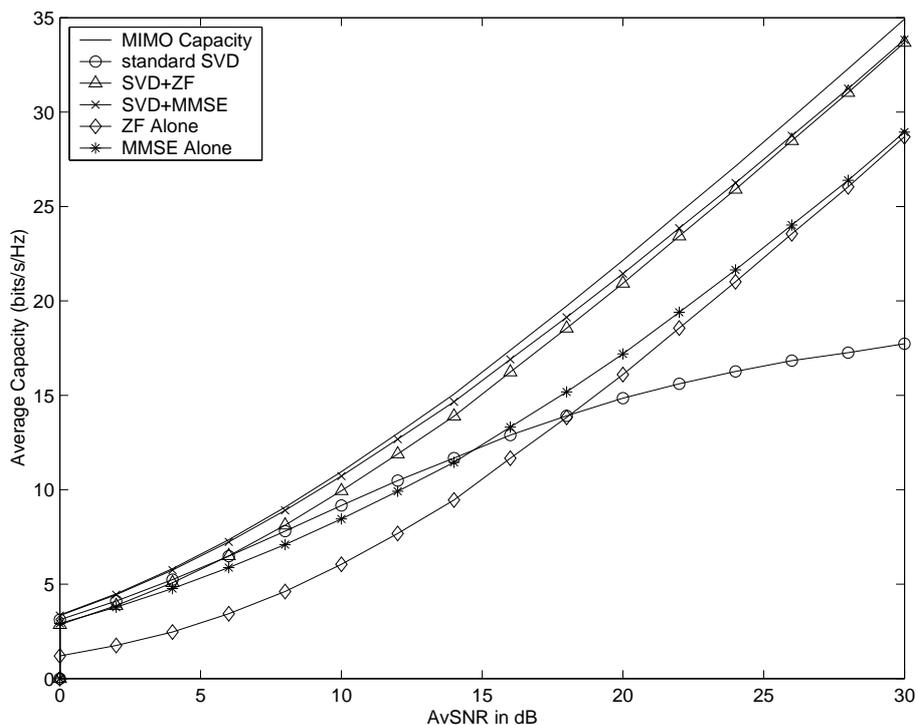


Figure V.13: System capacity of equal power systems with delayed CSI ( $F_d\delta t = 0.038$ ) at the transmitter

All systems collapse at low SNRs. However, at high SNRs, the capacity of the MIMO channel increases linearly with the SNR when expressed in dB, this has been demonstrated in [15].

The systems employing linear processing at the receiver ('SVD+ZF', 'SVD+MMSE',

'ZF Alone' and 'MMSE Alone') exhibit a linear increase in system capacity. However, the 'SVD+ZF' and 'SVD+MMSE' systems display a system capacity several bits higher than the 'ZF Alone' and 'MMSE Alone' systems. The difference is approximately 5 bits/s/Hz at high SNRs. The 'MMSE Alone' outperforms the 'ZF Alone' at low SNRs, and equals 'ZF Alone' performance at high SNRs. This trend is identical when coupled with the SVD ('SVD+MMSE' and 'SVD+ZF' systems).

The 'Standard SVD' system with the outdated  $\mathbf{V}(t - \delta t)$  matrix does not benefit from the MIMO effect, its system capacity reaches a ceiling at high SNRs. This is due to the cross talk between the sub-channels, which dominate the SINR. The cross talk creates an interference noise floor, that grows proportionally to the signal power increase. This interference noise floor does not fall even if the receiver noise is lowered.

#### V.4.3.2 System capacity with equal power allocation

The system capacity (see Section IV.3.2) of the three systems against  $F_d\delta t$  with SNR=10dB is shown in Fig. V.14.

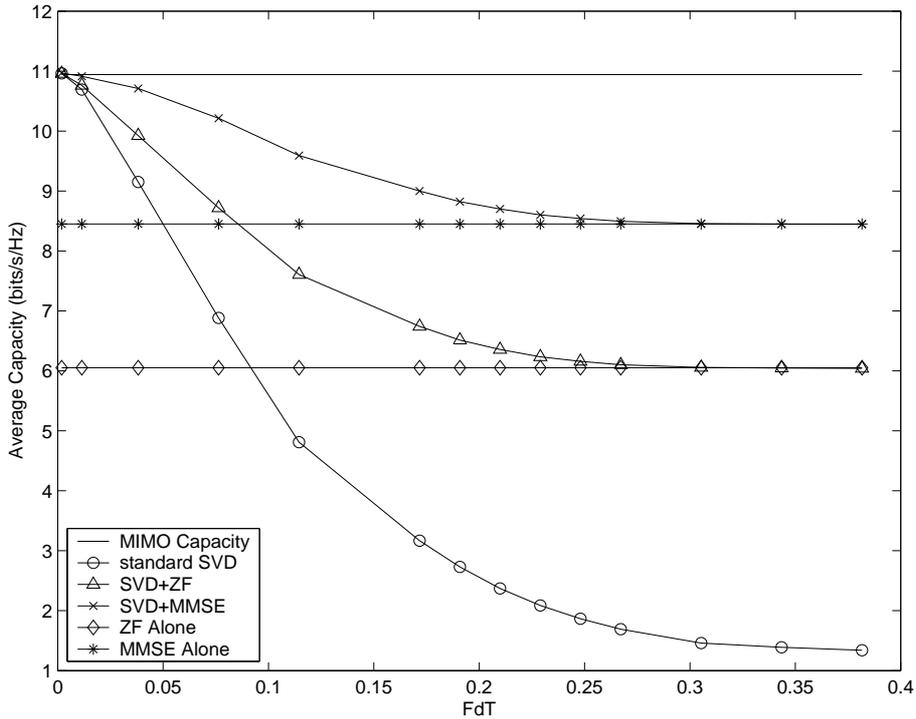


Figure V.14: System capacity of equal power systems with delayed CSI at the transmitter, average SNR=10dB

When  $F_d\delta t$  is small, capacities for all systems implementing SVD converge towards the theoretical capacity of the channel. This is to be expected, since the channel varies

slowly and the transmitter can estimate CSI accurately, in this situation SVD systems allow transmission at a rate equal to the theoretical capacity of the channel [12].

The 'MMSE Alone' system is 2.5 bits below the theoretical capacity of the channel whereas the 'SVD+MMSE' system approaches the theoretical bound. This demonstrates the effectiveness of the SVD approach and the importance of the processing by the  $\mathbf{V}$  matrix at the transmitter. The 'ZF Alone' system suffers an additional 2.5 bits degradation in system capacity, due to noise enhancement when the channel matrix is ill-conditioned.

The system capacity of the 'Standard SVD' system deteriorates rapidly and its system capacity at  $F_d\delta t = 0.038$  is nearly 2 bits below the theoretical capacity of the channel. The 'SVD+ZF' system reduces the effect of the CSI impairment at the transmitter, but leads to noise enhancement, hence outperforming the 'Standard SVD' system by less than a bit at  $F_d\delta t = 0.038$ . On the other hand, the 'SVD+MMSE' system mitigates the effects of CSI impairment without suffering from noise enhancement. This results in a system capacity drop by only 0.3 bits at  $F_d\delta t = 0.038$ .

When  $F_d\delta t$  is large (fast fading), the estimate of the CSI at the transmitter is incorrect. The 'SVD+ZF' and 'SVD+MMSE' systems compensate for this incorrect processing at the transmitter, by achieving similar system capacity to the 'ZF Alone' and the 'MMSE Alone' systems respectively. The precoding by a completely incorrect matrix does not change the capacity of the channel, since the precoding matrix is unitary. The 'Standard SVD' system collapses due to the incorrect CSI at the transmitter, and has a system capacity (1.3 bits/s/Hz) lower than the capacity of a SISO channel at SNR=10dB (3.4594 bits/s/Hz).

The system capacity of the three systems against  $F_d\delta t$  with SNR=20dB is shown in Fig. V.15. When  $F_d\delta t$  is small or large, the conclusions presented above are applicable, excluding the fact that the capacities are higher, since SNR=20dB.

At high SNRs, the ZF solution becomes closer to the MMSE solution, since the noise becomes negligible when compared with the interference.

The range of  $F_d\delta t$ s over which precoding provides an improvement ranges from zero to approximately 0.25. The range of improvement is therefore similar at high or low SNR.

#### V.4.3.3 System capacity of systems with waterfilling power allocation

The system capacity (see Section IV.3.2) of the three systems against  $F_d\delta t$  with SNR=10dB is shown in Fig. V.16.

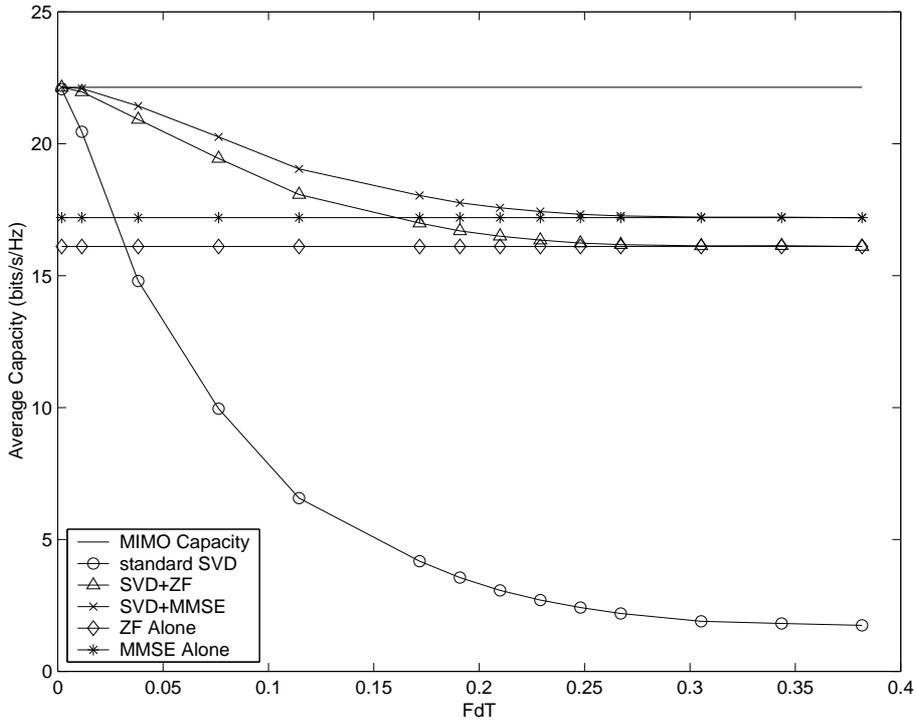


Figure V.15: System capacity of equal power systems with delayed CSI at the transmitter, average SNR=20dB

As discussed in Section V.4.1, the receiver requires prior knowledge of the power allocation applied by the transmitter to perform MMSE decoding. This is derived by the transmitter using its best estimate of the channel ( $\mathbf{H}(t - \delta t)$ ) and therefore is not the optimal power allocation. It was assumed for the 'SVD+MMSE' system that the receiver had access to this power allocation information.

When the receiver does not have access to this information it can estimate the applied power allocation using the current CSI ( $\mathbf{H}(t)$ ) and applying the waterfilling power allocation algorithm. A drop in system capacity will occur, since the power allocation applied by the transmitter, will differ from that calculated by the receiver to derive the MMSE filter. The new system, named 'SVD+MMSE+I', is compared with the 'SVD+MMSE' system in Fig. V.16.

At low  $F_d \delta t$ s (slow fading) the capacities of 'SVD+MMSE' and 'SVD+MMSE+I' are similar. As  $F_d \delta t$  increases the capacities diverge. However, the difference remains minimal, particularly over the  $F_d \delta t$  values the 'SVD+MMSE' scheme is likely to be used. The small system capacity variation indicates the high stability of the channel singular values (corresponding to the eigenmodes' gains) over time. This is in marked contrast to the  $\mathbf{V}$  matrix which fluctuates rapidly as is evident by the sudden reduction in system capacity

of the 'standard SVD' curve.

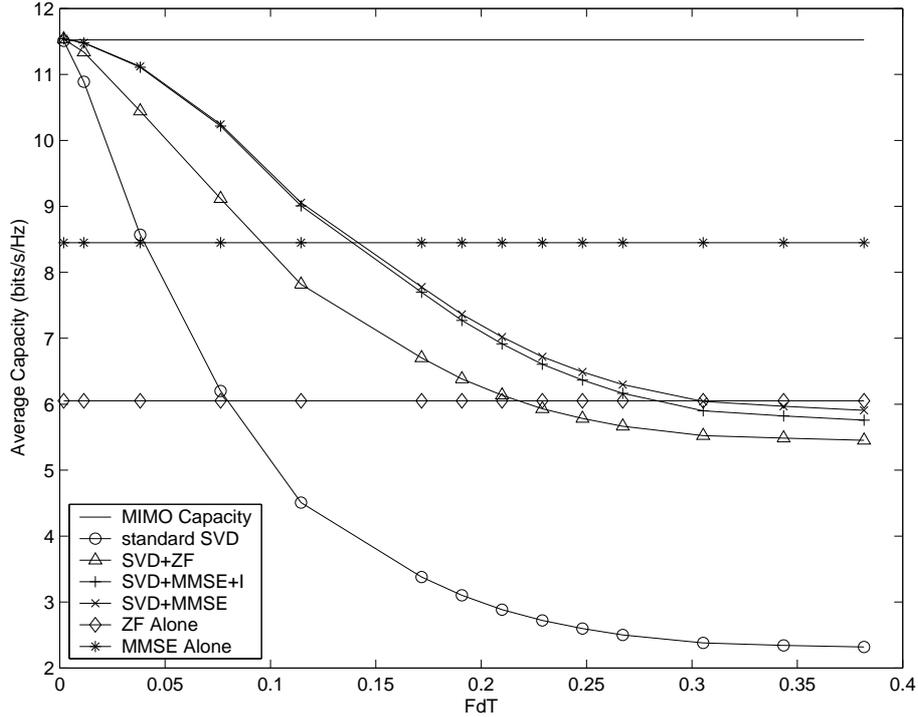


Figure V.16: System capacity of waterfilling systems with delayed CSI at the transmitter, average SNR=10dB

The observations reported in Section V.4.3.2 at low  $F_d \delta t$  are relevant to this section. However the capacity of the channel is higher with waterfilling power allocation. At high  $F_d \delta t$ , the 'SVD+ZF' and 'SVD+MMSE' systems not only suffer from CSI impairment but also from incorrect power allocation. This is confirmed in Fig. V.16, where the 'SVD+ZF' and 'SVD+MMSE' system capacity curves drop below the 'ZF Alone' and 'MMSE Alone' system capacity boundaries. The 'SVD+MMSE' system is particularly sensitive to power allocation errors, with performance dropping below the 'MMSE Alone' line at  $F_d \delta t = 0.15$ . The crossover point for the 'SVD+ZF' and 'ZF Alone' systems is at  $F_d \delta t = 0.22$ .

At  $F_d \delta t = 0.038$ , the capacities of both the 'SVD+ZF' and 'SVD+MMSE' using waterfilling are marginally better to the corresponding values with equal power allocation. However, the system capacity of the 'Standard SVD' system degrades faster with waterfilling compared to equal power allocation, with system capacity half a bit lower (8.5 compared to 9 bits/s/Hz) at  $F_d \delta t = 0.038$  at SNR=10dB.

In an ideal situation (low  $F_d \delta t$ , slow fading), the system capacity of waterfilling power allocation systems is only marginally higher than the system capacity of their equivalent equal power allocation systems. However, the simulation was based on i.i.d. MIMO

channels. The i.i.d. channel corresponds to a high scattering environment (numerous multipaths), and the gain of the sub-channels seldom becomes small. Other MIMO channels would produce different results. For example, in a line of sight (LOS) channel, a single positive singular value for the LOS path and null singular values for the other eigenmodes would result. Hence an equal power allocation system would only use a transmitting power of  $\frac{P}{M_T}$ , whereas the waterfilling power allocation system would allocate all the available power to the single non zero transmission eigenmode. Caution should be applied when comparing equal power allocation systems with waterfilling power allocation systems.

#### V.4.4 Conclusion

MIMO systems based on SVD algorithms produce transmission rates close to theoretical capacities, provided the transmitter has an accurate estimate of the CSI. When a TDD system is considered, the CSI can be obtained by sending pilot symbols from the receiver to the transmitter. However, frequent CSI updates are required at the transmitter, since the performance of an SVD based system severely degrades when the CSI is incorrect. The 'Standard SVD' system was shown to be unsuitable for  $F_d\delta t$ s greater than 0.03 (a system capacity loss of approximately 5 bits at SNR=20dB). The loss gets even larger at higher SNRs because the performance plateaus rather than linearly increasing. There is no benefit in implementing an SVD algorithm alone if the  $\mathbf{V}$  matrix is outdated.

A new architecture was proposed to counter the effects of incorrect CSI at the transmitter. It uses the outdated  $\mathbf{V}$  matrix of the 'Standard SVD' system in combination with linear filtering at the receiver. Firstly, pilot tones were sent through the  $\mathbf{V}(t - \delta t)$  matrix and channel prior to transmission and secondly, ZF or MMSE processing was implemented at the receiver for good MIMO performance. Finally, the prefiltering matrix,  $(\mathbf{U}(t)^*)^\dagger$ , for transmission in the reverse direction is obtained by taking the SVD (or eigen decomposition) of  $\mathbf{H}(t)\mathbf{V}(t - \delta t)$ . This architecture was shown to mitigate the effects of incorrect CSI at the transmitter, without increasing pilot overhead. Problems with respect to non-uniqueness of the SVD, subspace swapping, and sub-channel cross talk are all suppressed by linear filtering based on a channel that includes the outdated pre-coding matrix.

Results showed that the 'SVD+MMSE' system always outperforms the 'MMSE Alone' (no prefiltering) system with equal power allocation. Its performance is bounded by the channel capacity at  $F_d\delta t = 0$  and approaches the 'MMSE Alone' performance at large

$F_d\delta t$ s. In fact at a useable  $F_d\delta t$  of 0.038 and SNR=20dB the scheme has an operating limit less than 1 bit below the channel capacity. This compares to a 5 bit loss for the 'MMSE Alone' system and a 7 bit loss for the 'SVD Alone' system. In addition there is no performance plateau with increasing SNR. Similar trends apply to the 'SVD+ZF' system and this confirms that improved system capacity is possible by combining SVD prefiltering and linear post processing.

When waterfilling power allocation is added the performance of the 'SVD+linear filtering' schemes is no longer lower bounded by the linear filtering alone limit at large  $F_d\delta t$ s. Waterfilling should therefore not be used when power allocation is based on an overly outdated channel. At low  $F_d\delta t$ s there is some system capacity improvement, but for i.i.d. channels this is small and again waterfilling would not be recommended.

In conclusion systems using both SVD transmission and linear processing at the receiver outperform systems with no CSI at the transmitter by several bits at reasonable  $F_d\delta t$  ( $<0.1$ ), whether waterfilling is used or not. The addition of some precoding by the  $\mathbf{V}$  matrix, regardless if it is erroneous (outdated), was shown to be the source of major capacity gains.

## V.5 CSI improvement through channel tracking

It is shown in Section V.4 that the SVD architecture can be modified to see its performance degrade gracefully to the performance of an uncoded system as the coding matrix  $\mathbf{V}$  becomes less accurate. It is assumed in Section V.4 that pilot symbols provide perfect CSI, i.e. the channel estimation noise is negligible. In practice, the channel estimation noise is usually not negligible and degrades the accuracy of the precoding and decoding matrices.

The architecture presented in Section V.4 insures that a catastrophic perturbation of the precoding matrix does not have a catastrophic result on the performance of the system. However, the proposed architecture does not prevent errors in the decoding matrix to significantly reduce the performance of the system.

Decoding matrices are obtained through applying the SVD on the CSI. However, the SVD is a non-linear function and a slight error in the estimated CSI can result in a large variation of the decoding matrix. Therefore channel estimation has to be more accurate than in a standard MIMO system.

In a slow fading channel, tracking of the CSI in time can improve the accuracy of the channel estimation. Another solution consists of tracking the precoding and decoding matrices. These two solutions are compared in the following when tracking is performed using linear Finite Impulse Response (FIR) Wiener filters.

### V.5.1 Tracking of the CSI

The estimation of Rayleigh fading channels through PSAM has been analysed in [76]. Estimation is achieved by linearly combining measured pilot symbols that have been time-multiplexed with data.

This technique cannot be applied when the channel is quasi-static, i.e. the channel is fixed for one frame and consecutive frames are subject to independent fading. In such a case, the CSI estimated on previous frames cannot be used to improve the channel estimation for the current frame.

In indoor wireless LAN systems, such as modems following the 802.11a standard, the situation is likely to be a combination of both previous cases. The channel can be considered constant over one frame and correlated with channels at previous time slots.

#### V.5.1.1 Forward estimation

In a slow fading environment, the pilot symbols measured at previous time-slots can be used to improve the accuracy of the CSI estimate. The channel estimate measured from pilot symbols can be defined as:

$$\mathbf{P}(t) = \mathbf{H}(t) + \mathbf{N}_{est}(t). \quad (\text{V.26})$$

Defining  $\delta t$  as the time between two received frames, using an FIR Wiener filter (the optimum linear FIR filter) leads to:

$$\hat{\mathbf{H}}_{i,j}(l\delta t) = \mathbf{w}_{opt}^{\mathbf{H}_{i,j}} \times \vec{\mathbf{P}}_{i,j}(l\delta t), \quad (\text{V.27})$$

where  $l + 1$  is the length of the filter, and

$$\begin{aligned}
\vec{\mathbf{P}}_{i,j}(l\delta t) &= \begin{pmatrix} \mathbf{P}_{i,j}(0) \\ \mathbf{P}_{i,j}(\delta t) \\ \vdots \\ \mathbf{P}_{i,j}(l\delta t) \end{pmatrix} \\
\mathbf{w}_{opt}^{\mathbf{H}_{i,j}} &= (\mathbf{R}_p^{-1} \mathbf{c}_p)^* \\
\mathbf{R}_p &= E[\vec{\mathbf{P}}_{i,j}(l\delta t) \vec{\mathbf{P}}_{i,j}^*(l\delta t)] \\
\mathbf{c}_p &= E[\mathbf{H}_{i,j}(l\delta t) \vec{\mathbf{P}}_{i,j}^*(l\delta t)].
\end{aligned} \tag{V.28}$$

From this definition, the optimum filter is different for all  $M_R \times M_T$  paths of the MIMO channel. However, if all paths have the same statistics, the filters are identical, i.e.  $\mathbf{w}_{opt}^{\mathbf{H}_{i,j}} = \mathbf{w}_{opt}^{\mathbf{H}}$ . It is assumed the channels time variation follows Jake's model, i.e.  $\forall i, j \ E[\mathbf{H}_{i,j}(0) \mathbf{H}_{i,j}^*(\delta t)] = J_0(2\pi F_d \delta t)$ . Then obviously  $(\mathbf{R}_p)_{i,j} = J_0(2\pi F_d(i - j)\delta t) + E[nn^*]\delta(i, j)$  and  $(\mathbf{c}_p)_i = J_0(2\pi F_d(l - i)\delta t)$ , where  $\delta(i, j)$  is the Kronecker function.

### V.5.1.2 Forward-backward estimation

The previous channel estimation method only accounts for pilot symbols in the past. Channel estimation using both past and future pilot symbols is likely to provide a better channel estimation. In such a case, the system needs to buffer the received frame and process it later (when the pilot symbols required for CSI estimation have been received). Storing  $l/2$  frames is only possible when a large memory buffer is available at the receiver. Furthermore, delay sensitive applications might forbid such a solution.

However, in most fading scenarios (such as the Jake's fading assumed here), most of the improvement in the channel estimation is due to the pilots immediately preceding and following the desired frame. This is because the channel time correlation tends to decrease with increasing delay. Therefore, waiting for the next pilot before processing the frame might not only provide a significant improvement of the channel estimation but also keep the hardware requirements reasonable and satisfy the delay constraints of the system.

## V.5.2 Tracking of the precoding and decoding matrices

Instead of tracking the channel, it is possible to track the precoding and decoding matrices.

With  $(\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}) = \text{SVD}(\mathbf{H})$ , define the functions  $f^u(\cdot)$  and  $f^v(\cdot)$  as

$$\begin{aligned} f_{i,j}^u(\mathbf{H}) &= \mathbf{U}_{i,j}, \\ f_{i,j}^v(\mathbf{H}) &= \mathbf{V}_{i,j}. \end{aligned} \quad (\text{V.29})$$

For each received frame (at transceiver A or transceiver B), pilot symbols are measured, and the SVD of the resulting channel matrix is calculated.

The decoding matrices  $f^u(\mathbf{P}(t))$  can be filtered in time to improve their accuracy. Defining

$$\vec{f}_{i,j}^u(\mathbf{P}(l\delta t)) = \begin{pmatrix} f_{i,j}^u(\mathbf{P}(0)) \\ \vdots \\ f_{i,j}^u(\mathbf{P}(l\delta t)) \end{pmatrix}, \quad (\text{V.30})$$

the elements of this vector are combined in the following way:

$$\hat{\mathbf{U}}_{i,j}(l\delta t) = \mathbf{w}_{opt}^{U_{i,j}} \times \vec{f}_{i,j}^u(\mathbf{P}(l\delta t)) \quad (\text{V.31})$$

where  $\mathbf{w}_{opt}^{U_{i,j}}$  is the FIR Wiener (optimum) filter.

$$\begin{aligned} \mathbf{w}_{opt}^{U_{i,j}} &= (\mathbf{R}_{f_{i,j}^u(\mathbf{P})}^{-1} \mathbf{c}_{u_{i,j}})^* \\ \mathbf{R}_{f_{i,j}^u(\mathbf{P})} &= E[\mathbf{f}_{i,j}^u(\mathbf{P}(l\delta t)) \mathbf{f}_{i,j}^u(\mathbf{P}(l\delta t))^*] \\ \mathbf{c}_{u_{i,j}} &= E[\mathbf{U}_{i,j}(l\delta t) \vec{f}_{i,j}^u(\mathbf{P}(l\delta t))^*]. \end{aligned} \quad (\text{V.32})$$

Similarly, the precoding matrices  $f^v(\mathbf{P}(t))$  can be filtered in time. In the following, there is no attempt to correct for subspace swapping before the interpolation. Additional performance gain may be obtained by correcting for subspace swapping.

### V.5.2.1 Correlation of the decoding matrix

Derivation of the Wiener filter presented in the previous section implies explicit knowledge of the correlation in time of the elements of the decoding matrix. This correlation, unlike the correlation of the elements of the channel matrix, is not well known.

Consider  $\mathbf{H}(1)$  and  $\mathbf{H}(2) = r\mathbf{H}(1) + \sqrt{1-r^2}\mathbf{N}_{tvc}$ , where  $\mathbf{H}(1)$  and  $\mathbf{N}_{tvc}$  are i.i.d. complex Gaussian matrices with zero mean and unit variance. The mathematical derivation of the correlation between  $\mathbf{U}(1) = f^u(\mathbf{H}(1))$  and  $\mathbf{U}(2) = f^u(\mathbf{H}(2))$ , is still an open problem, to the author's knowledge. The expectancy of the correlation between the elements of  $\mathbf{U}(1)$  and the elements of  $\mathbf{U}(2)$  is presented in Fig. V.17. As expected, the

elements of a column of  $\mathbf{U}$  have the same expected correlation when perturbed, since the channel is i.i.d. and a permutation of the label of the antennas at the receiver permutes the lines of  $\mathbf{U}$  accordingly. All correlations have a maximum value of 0.25 since the elements of  $\mathbf{U}$  are equally likely and the matrix is unitary. The correlation of the channel is always higher in absolute value than the correlation of the singular vectors: as mentioned in [58], a small perturbation of the channel may result in a large change in the singular vectors. The singular vectors corresponding to the largest singular value shows the highest correlation.

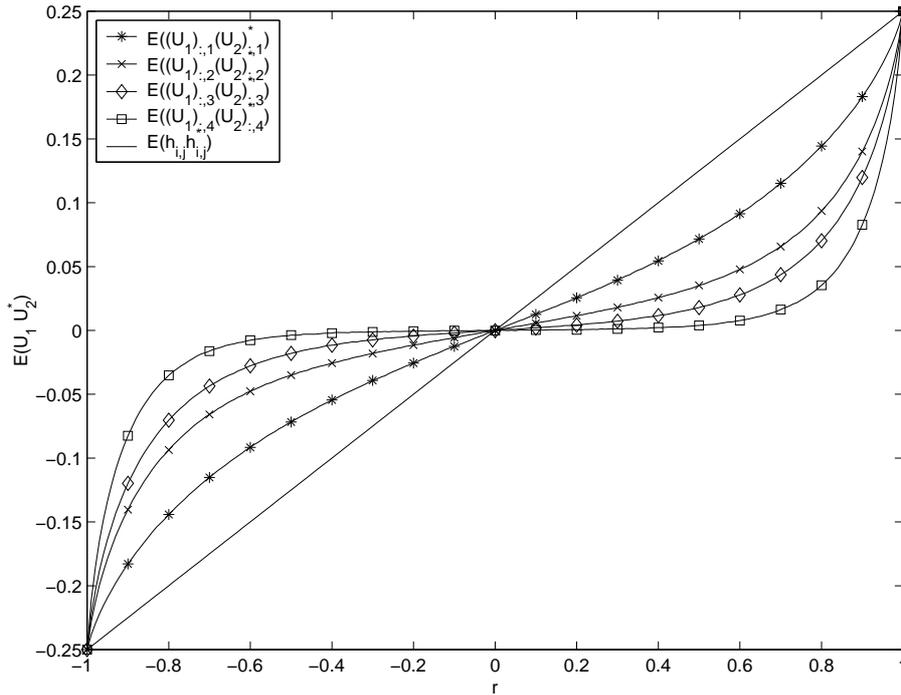


Figure V.17: Correlation of the decoding matrix:

$$E(f^u(\mathbf{H}(1))f^u(r\mathbf{H}(1) + \sqrt{1-r^2}\mathbf{N}_{tvc})^*)$$

### V.5.2.2 Correlation of the precoding matrix

Fig.V.18 shows the correlation between elements of a column of  $\mathbf{V}$  under similar assumptions as in the previous section. The lower curves represent the correlation between elements of the precoding matrix not on the main diagonal<sup>5</sup>. They have the same correlation when perturbed, since the channel is i.i.d. and a permutation of the labels between

<sup>5</sup>The SVD in this sense is a modified version to standard literature where columns of the  $\mathbf{V}$  matrix are rotated such that the main diagonal are positive real values. Columns of  $\mathbf{U}$  must be rotated by the same amount. This does not affect generalisation.

antennas at the transmitter permutes the columns of  $\mathbf{V}$  accordingly. The upper curves of Fig.V.18 are the correlation between elements along the main diagonal. They have different correlations since each correlation corresponds to a different singular value (sorted in descending order). All correlations have a maximum value of 0.25 since the elements of  $\mathbf{V}$  are equally likely and the matrix is unitary.

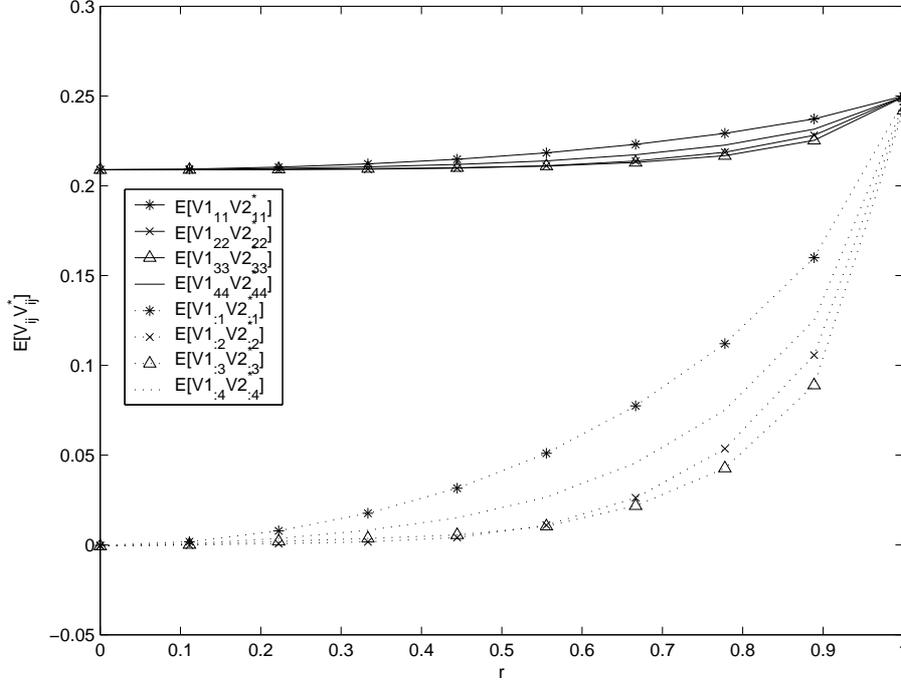


Figure V.18: Correlation of the Precoding Matrix:

$$E[f^v(\mathbf{H}(1))f^v(r\mathbf{H}(1) + \sqrt{1-r^2}\mathbf{N}_{tvc})^*]$$

### V.5.2.3 General SVD correlation results

Two further theoretical results are required to obtain the coefficients of the filter of the decoding matrix.

Consider  $\mathbf{H}(1)$ ,  $\mathbf{H}(2)$  two matrices.  $\forall r \geq 0$ ,  $f^u(r\mathbf{H}(2)) = f^u(\mathbf{H}(2))$ . Therefore

$$E((f_{i,j}^u(\mathbf{H}(1)))(f_{i,j}^u(\mathbf{H}(2)))^*) = E((f_{i,j}^u(\mathbf{H}(1)))(f_{i,j}^u(r\mathbf{H}(2)))^*). \quad (\text{V.33})$$

The correlation between the decoding matrices of two channel matrices are not affected when the power of the channel matrices is modified.

Consider  $\mathbf{N}_{tvc}$  a perturbation matrix,  $\mathbf{H}(1)$  and  $\mathbf{H}(2)$  are i.i.d. complex Gaussian matrices with zero mean and same variance. Then

$$E(f_{i,j}^u(\mathbf{H}(1) + \mathbf{N}_{tvc})f_{i,j}^u(\mathbf{H}(2))^*) = E(f_{i,j}^u(\mathbf{H}(1))f_{i,j}^u(\mathbf{H}(2) + \mathbf{N}_{tvc})^*). \quad (\text{V.34})$$

This is simply because  $\forall c_1, c_2 \in \mathbb{C}, E(c_1 c_2^*) = E(c_2 c_1^*)^*$  and  $E(f_{i,j}^u(\mathbf{H}(1)) f_{i,j}^u(\mathbf{H}(2))^*)$  is real. Exchanging the labels  $c_1$  and  $c_2$  completes the proof.

The two previous results can be obtained in a similar manner for the precoding matrix.

### V.5.3 Correlation of pilots

From the general results presented in Sections V.5.2.1 and V.5.2.3, the correlation between pilot symbols at time 0 and at time  $\delta t$  can be derived for all  $\delta t$ .

Then,

$$\begin{aligned} & E(f_{i,j}^u(\mathbf{P}(0)) f_{i,j}^u(\mathbf{P}(\delta t))^*) \\ &= E(f_{i,j}^u(\mathbf{H}(0) + \frac{\mathbf{N}(0)}{\sqrt{\text{SNR}_{\text{est}}}}) f_{i,j}^u(\mathbf{H}(\delta t) + \frac{\mathbf{N}(\delta t)}{\sqrt{\text{SNR}_{\text{est}}}})^*) \\ &= E(f_{i,j}^u(\mathbf{H}(0)) f_{i,j}^u(r\mathbf{H}(0) + \sqrt{1-r^2}\mathbf{N}_{eq})), \end{aligned} \quad (\text{V.35})$$

where  $\mathbf{N}(0)$ ,  $\mathbf{N}(t)$  and  $\mathbf{N}_{eq}$  are complex i.i.d. Gaussian random matrices with zero mean and unit variance,

$$r = \frac{J_0(2\pi F_d \delta t)}{1 + \frac{1}{\text{SNR}_{\text{est}}}}. \quad (\text{V.36})$$

Combining (V.35) and the results in Fig. V.17, it is straightforward to obtain the correlation between the elements of the decoding matrices derived from pilot symbols at different time slots, for varying SNRs. Finally, the taps of the FIR Wiener filter for each element of the precoding matrix are deduced from the correlation, as in Section V.5.1.

Similar results can be obtained for the precoding matrix. The prediction on the precoding matrix results in a matrix that, due to some predictive error, is not necessarily unitary. The precoding matrix is required to be unitary in a double pilot architecture (Section V.4). The predicted matrix,  $\mathbf{V}^{\text{pred}V}(\delta t)$ , can be projected onto an orthonormal basis. Based on the correlation curves of Fig. V.18, the first column  $\mathbf{V}_{:,1}^{\text{pred}V}(\delta t)$  is normalized and then made orthogonal to  $\mathbf{V}_{:,4}^{\text{pred}V}(\delta t)$ , the column with the next strongest correlation. The resultant matrix  $\mathbf{V}^{ON}(\delta t)$  is the orthonormal basis matrix of the predicted precoding matrix,  $\mathbf{V}^{\text{pred}V}(\delta t)$ .

### V.5.4 Simulation results, decoding matrix

To validate the proposed filtering methods, a  $M_T = M_R = 4$  MIMO system was simulated (16 i.i.d. SISO channels, Jake's fading). The time varying MIMO channels were simulated using independent variables of SNR and  $F_d \delta t$ . The influence of the filter length on the system capacity was also investigated.

Fig. V.19 presents the block diagrams of the various simulated systems. The precoding matrix is assumed perfect.

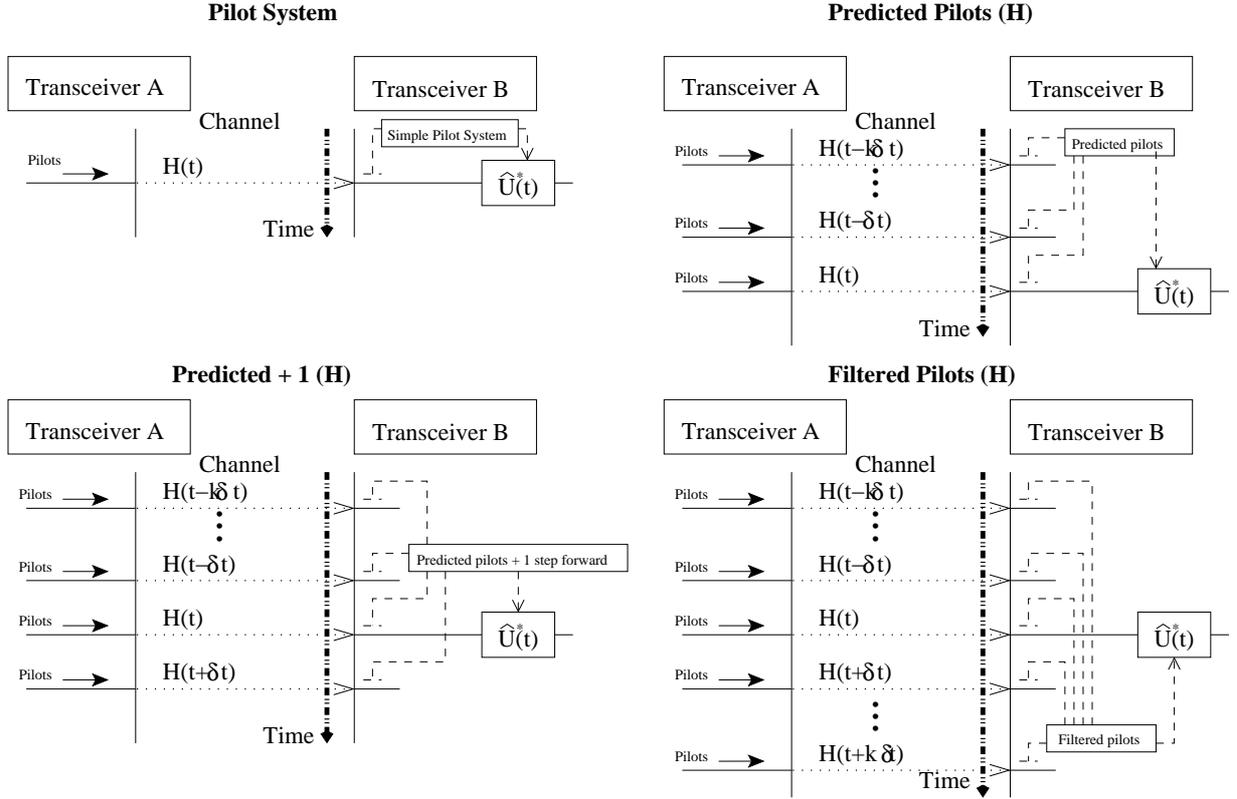


Figure V.19: The CSI at the receiver can be improved using previous and/or future pilot symbols to improve the accuracy of the channel estimation.

#### V.5.4.1 System capacity for varying $F_d\delta t$

The system capacity (defined in Section IV.3.2) of the systems against  $F_d\delta t$  with an SNR=20dB, is shown in Fig. (V.20).

The capacity of the MIMO channel (curve 'MIMO Capacity') as well as the capacity achieved by a using only the pilots of the current frame (curve 'Pilot System') are given as a reference. The capacity of the MIMO channel is not affected by  $F_d\delta t$ . Neither is the system capacity of the 'Pilot System' under the block fading assumption.

Obviously the performance of these two systems is not affected by  $F_d\delta t$ .

At all  $F_d\delta t$ , filtering the CSI (curve 'Filtered Pilots (H)') or filtering the decoding matrix (curve 'Filtered Pilots (U)') leads to a significant improvement of the capacity of the system.

Filtering the  $\mathbf{H}$  matrix rather than the  $\mathbf{U}$  matrix with the same filtering technique

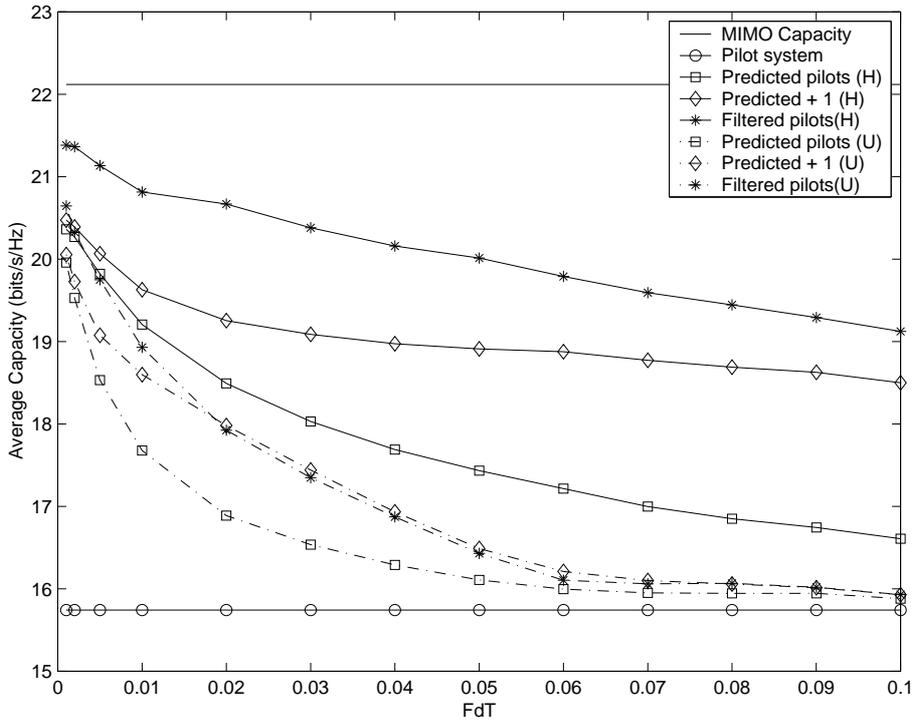


Figure V.20: Performance of MIMO-PSAM systems, SNR=20dB, Equal Power, filter length of 31 taps

(e.g. curve 'Filtered Pilots (H)' rather than curve 'Filtered Pilots (U)') leads to a better estimation of the decoding matrix for all  $F_d\delta t$ . Filtering the decoding matrix directly is effective at low  $F_d\delta t$  but does not perform well at high  $F_d\delta t$ . The correlation between consecutive decoding matrices decreases rapidly with  $F_d\delta t$ .

For both methods (filtering the CSI or filtering the decoding matrix), a balanced filter (curves 'Filtered Pilots'), which filters past and future channel estimates, provides the best performance. However, a balanced filter requires a large memory buffer and introduces large delay in the reception chain, as discussed in Section V.5.1. An unbalanced filter buffering for only one time-slot (curves 'Predicted + 1'), can achieve a similar performance with less stringent hardware requirement at high  $F_d\delta t$ . This can be explained by observation of the correlation curves of Fig. V.17. At high  $F_d\delta t$  little correlation exists beyond the first time-slot and hence minimal performance improvement is obtained by waiting for further pilots. However, using only the information about past pilots (curves 'Predicted Pilots') leads to a severe degradation in performance at all  $F_d\delta t$

#### V.5.4.2 System capacity for varying SNR

The capacity of the systems against SNR with a  $F_d\delta t = 0.04$  is shown in Fig. V.21.

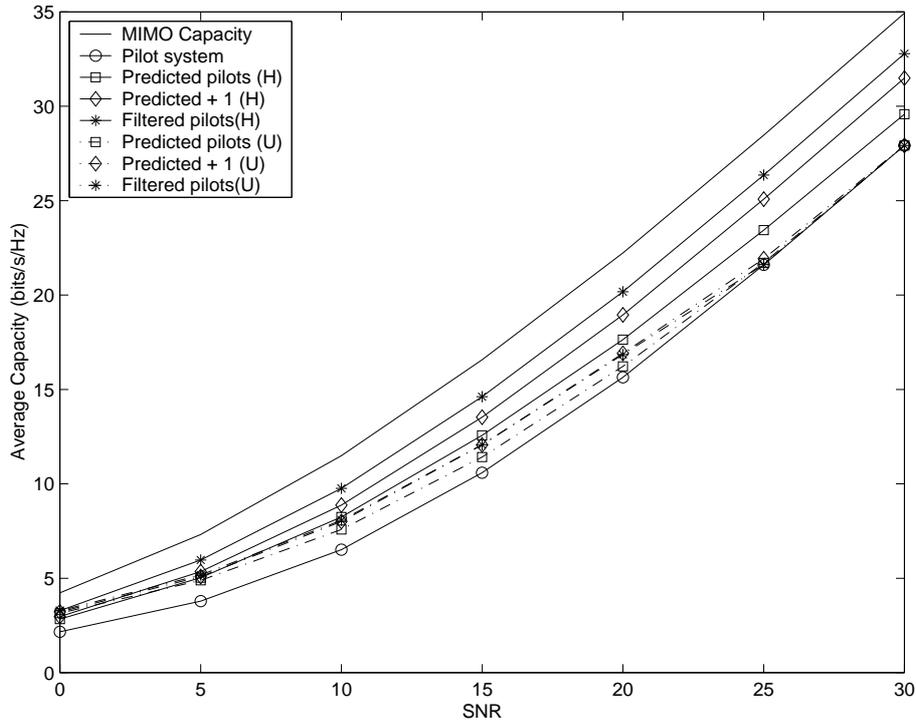


Figure V.21: Performance of MIMO-PSAM systems,  $F_d T = 0.04$ , equal power allocation, filter length of 31 taps

Filtering on  $\mathbf{H}$  provides better performance than filtering on  $\mathbf{U}$ , at all SNR.

At low SNR, the noise dominates the time variation of the channel and therefore the filters mainly average out the noise. Therefore the performance of the filters do not depend on their position.

On the contrary, at high SNR, the filters are disregarding the noise and focus on predicting the variations of the channel. In such a case, a balanced filter performs better than a filter with a delay of one or a predicting filter.

### V.5.4.3 System capacity for varying filter length

The capacity of the systems against the length of the filter with SNR=20dB and  $F_d \delta t = 0.04$ , is shown in Fig. V.22.

For a filter of length 3, the balanced (curves 'Filtered pilots') and unbalanced (curves 'Predicted + 1') filters are the same, which explains why they provide exactly the same system capacity.

Filters of reasonable length ( $L = 3$ ) achieve a large part of the system capacity gain achievable through filtering. This is mainly due to the fact that the time-slots most correlated with the time-slot of interest are its immediate neighbours. This is especially

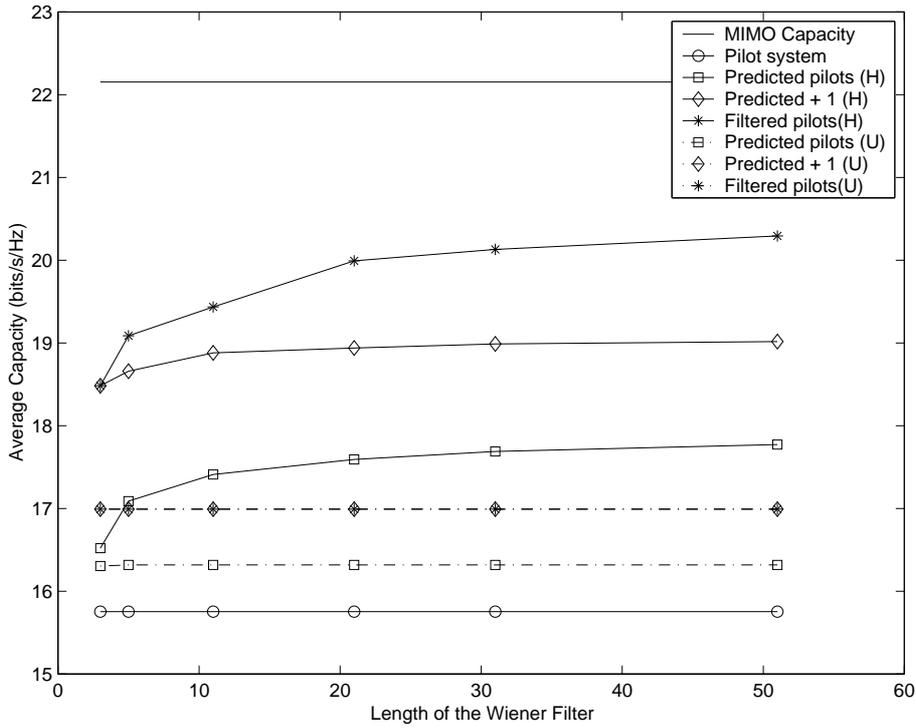


Figure V.22: Performance of MIMO-PSAM systems, SNR=20dB,  $F_d\delta t = 0.04$ , equal power allocation

crucial when filtering is performed on the decoding matrix. The correlation of the decoding matrix is so weak that the decoding matrices more than two frames away provide little information about the desired frame. Using them provides negligible system capacity improvement.

#### V.5.4.4 Orthonormalization of the filtered decoding matrix

When filtering the decoding matrix directly, the filtering process alters the unicity of the decoding matrix. It is possible to follow the filtering with a Gram-Schmidt orthonormalization, to retrieve the interesting unicity properties of the decoding matrix. Simulation results are presented in Fig. (V.23).

At all  $F_dT$ , for all three filters, the orthonormalization provides a slight improvement of the performance. Orthonormalizing actually corresponds to further filtering, since it is simply a projection of the estimation of the decoding matrix on the unitary matrices ensemble.

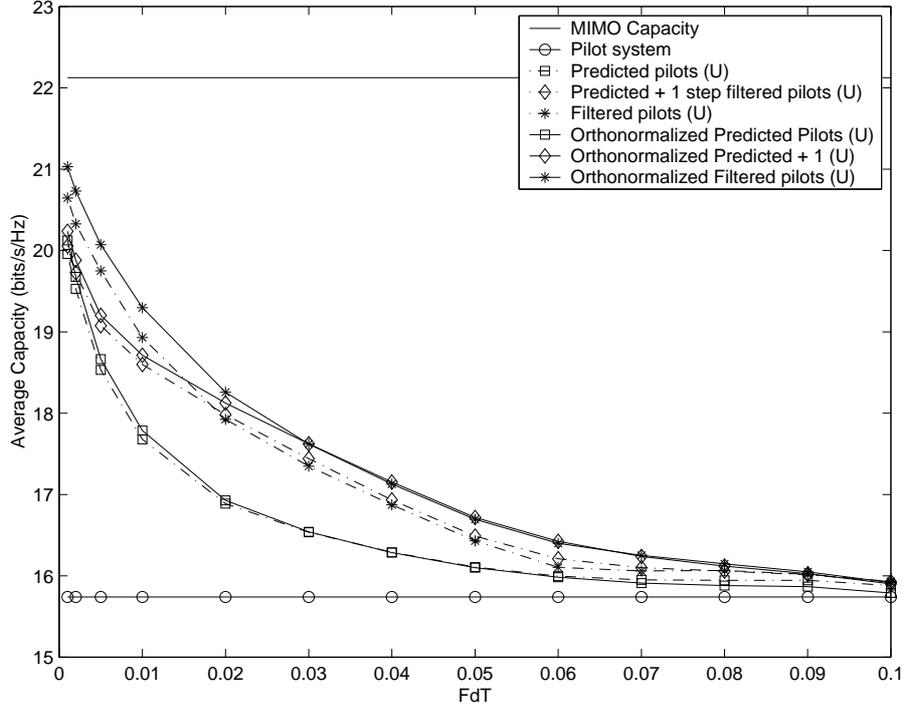


Figure V.23: Performance of PSAM systems, SNR=20dB, Equal Power

### V.5.5 Simulation results, precoding matrix

The system capacity of SVD system with filtering of the CSI and filtering of the precoding matrix at the transmitter is presented in this section through simulation results, with identical parameters as in Section V.5.4. A perfect decoding matrix ( $\mathbf{U}$ ) is assumed. The precoding matrix can only be estimated using past pilot symbols.

#### V.5.5.1 System capacity vs. $F_d\delta t$

The system capacity against  $F_d\delta t$  is determined through simulations (Fig.V.24). As expected, the system capacity performance reduces for faster moving channels for all cases. The curve labelled 'Prev V' corresponds to a system where no prediction is performed and the precoding matrix  $\mathbf{V}^P$  is used. Significant performance improvement results when  $l$  previous CSI estimates are linearly combined to estimate the CSI on the downlink as  $\hat{\mathbf{H}}$  (Section V.5.1). The corresponding system capacity curve is labelled 'Pred H'. The performance improvement (when compared with a system using no prediction) offered through prediction of the precoding matrix, is about half that of the gain offered from prediction of the CSI. A further small performance increase can be achieved by projecting the predicted precoding matrix onto an orthonormal basis,  $\mathbf{V}^{ON}$ , curve 'Pred V+ON'.

For a static channel channel,  $F_d\delta t = 0$ , all prediction filters become averaging filters, i.e. all taps are equal. It is seen that noise dominates when no prediction is performed. System performance is almost ideal when averaging is performed on the CSI depending on the length of the filter. Filtering the precoding matrix has a similar system capacity as filtering the CSI on very slow fading channels.

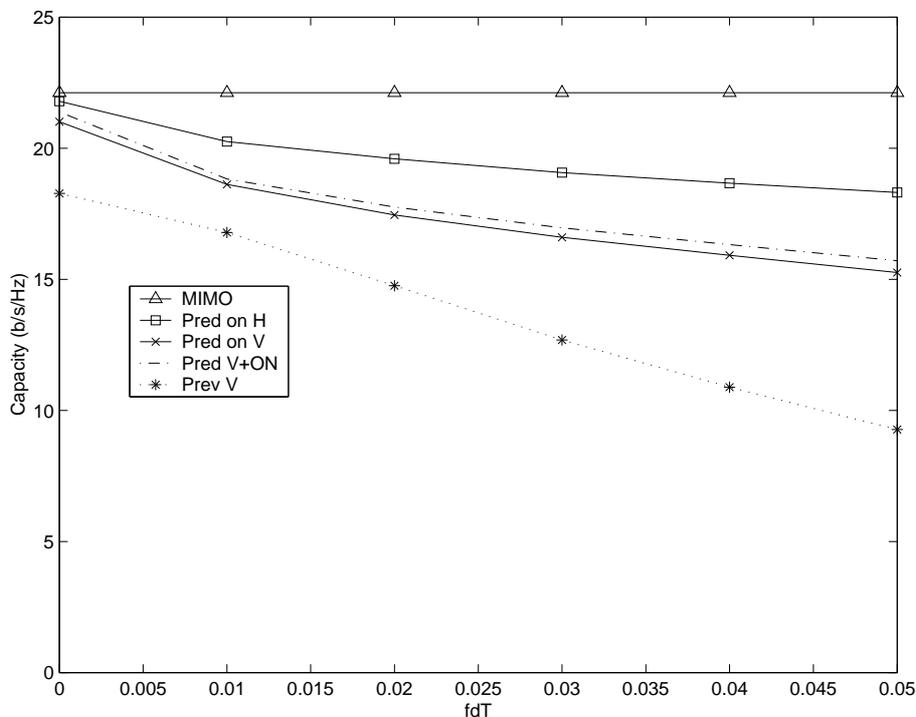


Figure V.24: System capacity vs.  $F_d\delta t$  (SNR=20dB,  $L = 30$  taps)

### V.5.5.2 System capacity vs. SNR

Higher SNR means that the pilot measured CSI,  $\mathbf{P}$  is more accurate. Therefore less interference occurs as a result of incorrectly matched precoding and decoding matrices. Higher system capacity is obtained for all systems as shown in Fig. V.25. It can be seen that when no filtering is applied (curve 'Prev V'), an improvement in SNR increases the system capacity but an error floor exists due to the channel fading. For a system capacity of 10 bps/Hz, a gain of slightly more than 7dB can be achieved by filtering the CSI (curve 'Pred H'). When prediction is performed on the precoding matrix (curve 'Pred V'), a gain of about 5dB can be achieved. An additional gain of almost 1dB can be achieved by projecting the predicted precoding matrix onto an orthonormal basis (curve 'Pred V+ON').

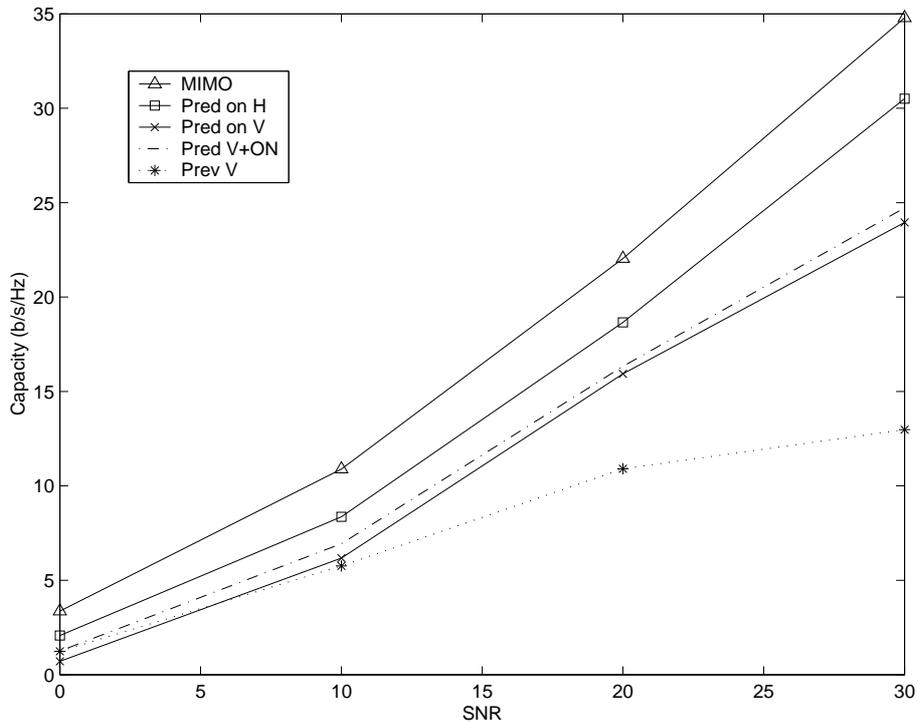


Figure V.25: System capacity vs. SNR ( $F_d\delta t = 0.040$ ,  $L = 30$  taps)

### V.5.6 Conclusion

Most practical communication systems deduce the CSI through measurement of pilot symbols multiplexed with the data. In a MIMO-SVD system, the channel needs to be estimated accurately since the SVD is a non-linear function. Therefore, estimation of the channel on a frame by frame basis might not provide sufficient accuracy in the estimation of the channel. The channel correlation from frame to frame can be exploited to improve the channel estimation without additional pilot overhead. This is achieved through FIR linear Wiener filtering. Filtering can be applied on the CSI or directly on the precoding and decoding matrices.

Properties of the correlation of the precoding and decoding matrices when the channel is perturbed were presented, as well as a generic method to derive the optimal filter corresponding to the precoding and decoding matrices.

Estimation and filtering of the precoding and decoding matrices were shown to be possible through simulation results. However, the best performance was achieved when estimation and filtering was performed on the CSI. At the receiver side, when delays are tolerable and hardware complexity is not an issue, the best overall performance is obtained with a balanced filter. For most practical scenarios it was shown that only a small number

of taps is necessary to achieve most of the performance gain. Most importantly, it is not necessary to wait for more than one additional pilot symbol before estimating the channel.

Filtering the precoding and decoding matrices should be limited to cases where the  $F_d\delta t$  is small. In such a case, orthonormalization of the estimated precoding and decoding matrices provides additional filtering and leads to performance improvement.

## V.6 Conclusion

The SVD transmission architecture requires knowledge of the CSI at both the transmitter and receiver. It is possible to provide the transmitter with the CSI without additional signaling overhead when the channel is reciprocal. However, the CSI is usually obtained through pilot symbols. Therefore the CSI is imperfect at both the transmitter and the receiver.

The errors due to imperfect CSI can create a large loss of system capacity. Specifically, when catastrophic events named 'singular value crossings' occur, a small perturbation in the channel (however small) can create a large perturbation of the precoding and decoding matrices. In such a case, the transmission is no longer robust to noisy channel estimates. The analysis of 'singular value crossings' can be conducted through the theory of matrix perturbation. However the time correlation of the channel prevents the direct application of the theory to fading channels, to the knowledge of the author. Simulation results indicate that 'singular value crossings' create 'singular subspace swappings' in Rayleigh fading: two singular values vary in amplitude to the point where they are crossing each other, but their corresponding subspaces are stable through the process. Results show that the probability of 'singular value crossings' is small and the effects of 'singular value crossings' can be corrected.

Incorrect CSI reduces the system capacity of SVD systems. When the precoding matrix is in error, the capacity of the system can drop below the system capacity of uncoded systems. In a TDD environment where the channel is affected by Doppler spread, the SVD system was shown to be unsuitable for  $F_dT$ s greater than 0.03 (a system capacity loss of approximately 5 bits at SNR= 20dB). The loss gets even larger at higher SNRs because the performance plateaus rather than linearly increases. There is no benefit in implementing an SVD algorithm alone if the precoding matrix is outdated. A new architecture is proposed which allows the system to benefit from the high system capacity

of SVD systems when the channel estimation is correct and seamlessly transmit with the performance of an uncoded system when the precoding matrix is in error. The new architecture considers the precoding matrix as part of the channel for channel estimation purposes. Therefore, errors in the precoding matrix can be corrected in the decoding matrix using a simple zero-forcing or MMSE process. This new architecture does not require additional pilots and improves the useable  $F_d T$  from 0.03 to infinity: even at very high  $F_d T$  the performance of the system is better or equal to the performance of systems without linear precoding.

Finally, the estimation of the channel on a frame by frame basis might not provide sufficient accuracy in the estimation of the channel. The channel correlation from frame to frame can be exploited to improve the channel estimation without additional pilot overhead. This is achieved through FIR linear Wiener filtering. Filtering can be applied on the CSI or directly on the decoding matrix. Properties of the correlation of the decoding matrix when the channel is perturbed were demonstrated, as well as a generic method to derive the optimal filter corresponding to the decoding matrix. Estimation and filtering of the decoding matrix was shown to be possible through simulation results. The best performance was achieved when estimation and filtering was performed on the CSI. When delays are tolerable and hardware complexity is not an issue, the best overall performance is obtained with a balanced filter. However for most practical scenarios it was shown that only a small number of taps (i.e. 3) is necessary to achieve most of the performance gain. Particularly, it is not necessary to wait for more than one additional pilot symbol before estimating the channel.

# Chapter VI

## Conclusion

### VI.1 Summary

Wireless LANs are both user-friendly and cost-effective. These advantages explain the current interest in Wireless LAN telecommunications equipment. Transmission techniques based on transceivers with multiple element antennas enable higher spectrum efficiency than is currently obtained using existing wireless LAN products. Therefore, it is of interest to apply MIMO transmission techniques to the wireless LAN environment.

MIMO telecommunications theory indicates that the capacity of a wireless link increases linearly with the number of antennas at both ends of the channel for constant bandwidth and transmitted power (Chapter II) when the channel is uncorrelated. Furthermore, a series expansion of the capacity of the ergodic channel with no CSI at the transmitter indicates that symmetric antenna allocation (same number of antennas at the transmitter and the receiver) maximizes the capacity of a MIMO channel with a given total number of antennas.

The capacity of correlated MIMO channels is lower than that of uncorrelated channels. Specifically, the capacity of Ricean channels tends to the capacity of their i.i.d. component when the number of antennas grows large (Chapter III). Theoretical bounds of the normalized capacity of the Ricean channel are proven in Chapter III. These bounds allow the capacity of the Ricean channel to be estimated in the asymptotic limit of a large number of antennas without recourse to simulation.

Practical transmission techniques are required to turn the promises of high capacity into high performance transmission devices. Chapter III presents an overview of some

well-known techniques to transmit over the MIMO channel and detect the transmitted symbols. Robust systems use diversity to obtain reliable transmission. However, they require large constellations to achieve a non-negligible fraction of the capacity of the channel. Systems with a large multiplexing gain are bandwidth efficient but only transmit well on uncorrelated channels.

Therefore, transmission systems with no CSI at the transmitter are designed either for the correlated channel or for the uncorrelated channel. Wireless LANs experience a large variety of propagation environments and so suffer from performance loss when the CSI is not available at the transmitter. CSI at the transmitter increases the capacity of the MIMO channel and is especially beneficial when the channel is highly correlated as demonstrated in Section III.5. The capacity gain with CSI at the transmitter, grows linearly with the number of antennas.

Information theory suggests a transmission architecture based on the SVD of the channel matrix to benefit from the CSI at the transmitter. The SVD architecture decomposes the MIMO channel into SISO transmission eigenmodes and allocates power to the eigenmodes following a waterfilling algorithm. This architecture is the optimal linear precoder and decoder under a variety of criteria, as detailed in Section IV.2.1. Furthermore, the SVD structure combined with OFDM is also an optimal space-time modulation in terms of information rate (Section IV.2.2).

The complexity of the SVD structure can be reduced by considering each transmission eigenmode as a separate channel, which leads to the notion of system capacity (Section IV.3). However, this reduction of complexity is obtained through a reduction of the robustness of the system to various impairments. E.g. the channel estimation noise should be smaller than the noise in the transmission to avoid a loss of performance.

As explained earlier, SVD systems benefit from the availability of CSI at the transmitter. The estimation of an accurate CSI at the transmitter is the key challenge faced by SVD system designers. TDD channels enable the estimation of the CSI at the transmitter without overhead data transmission from the receiver to the transmitter. This is due to the fact that TDD channels are theoretically reciprocal. In practical TDD channels, a loss of accuracy in the estimation of the CSI at the transmitter occurs, due to several imperfections of the channel and the transmission system:

- The SVD of the channel matrix is not unique. Matched transmitting and decod-

ing matrices are required. The transmitter and the receiver can implicitly choose matched transmitting and decoding matrices as presented in Section IV.4.2.

- Mismatched transmitter and receiver RF chains cancel the reciprocity of the channel. A calibration procedure is presented in Section V.2 to cancel the effects of mismatched transmitter and receiver RF chains. The calibration procedure relies on a handshake at the beginning of the transmission, as well as the hypothesis that the impairments of the chains are stationary.
- The performance of practical channel estimation algorithms is limited and the CSI is quantised in practical systems. The theory of matrix perturbation highlights the effect of imperfect channel estimation on SVD systems (Section V.3). A small perturbation of the CSI can result in a large performance loss when a 'singular value crossing' occurs, i.e. the channel matrix has two equal singular values. Simulation results indicate that 'singular value crossings' create 'singular subspace swappings' in Rayleigh fading: two singular values vary in amplitude to the point where they are crossing each other, but their corresponding subspaces are stable through the process. Results show that the probability of 'singular value crossings' is small and the effects of 'singular value crossings' can be corrected.
- Practical channels are time-varying. In Section V.4, the SVD system is shown to be unsuitable for  $F_d T$ s greater than 0.03 (a capacity loss of approximately 5 bits at SNR= 20dB). The loss gets even larger at higher SNRs because the performance plateaus rather than linearly increases. There is no benefit in implementing an SVD algorithm alone if the precoding matrix is outdated.

It is possible to mitigate the effects of both imperfect channel estimation and channel fading: FIR linear Wiener filters can exploit the channel correlation from frame to frame to improve the channel estimation without additional pilot overhead, as presented in Section V.5. The correlation of the CSI degrades rapidly in time for a Rayleigh fading channel. Therefore, filters of reasonable length (3 taps) achieve near optimum performance. At the receiver, the estimated CSI should include knowledge of the future pilot symbols when allowed by the hardware, i.e. when large memory buffers are available and the application is not time-delay sensitive. The CSI is usually more accurate at the receiver than at the transmitter since the CSI of the current time slot (and possibly future time slots) is

available. This is not the case for the transmitter, when the CSI for the current time slot is not available and must be predicted from past estimates.

It is possible to mitigate the effects of imperfect CSI at the transmitter if the system includes the precoding matrix in the channel estimation. This allows the receiver to mitigate the effect of the incorrect precoding matrix. This new architecture, with limited added complexity, obtains the benefit of the SVD architecture when CSI is precisely known at both ends of the link while seamlessly shifting to a non-precoded system (such as ZF or MMSE linear decoder) when the channel estimation precision deteriorates at the transmitter. The SVD architecture adapts dynamically to the propagation environment to obtain very high spectrum efficiency under varying channel conditions.

## VI.2 Future work

The results presented in the previous Chapters create several research opportunities in the following key areas: pilot symbol theory, error correction codes, wireless channel models and complexity issues.

### VI.2.1 Pilot symbols

Section V.5 introduces a proposal to improve the performance of pilot symbol assisted systems. The proposed systems combines the noisy channel estimate at several points in time to obtain an accurate estimate of the current channel. Practically, the solution is implemented through filtering of the noisy channel estimate. It is implicitly assumed that the time samples of the channel (the transmission slots) are evenly spaced in time. This assumption is realistic for Time Division Multiple Access (TDMA) systems such as Hiperlan 2. However, wireless LAN standards such as 802.11 (a, b and g) are Carrier Sense Multiple Access - Collision Detection (CSMA-CD) systems where the assumption is no longer realistic. The current proposal requires to derive the coefficients of the filter for each set of time-intervals between the transmission time slots. Therefore, it is unrealistic to expect the current proposal to be implemented as is. Further research is required to adapt the proposal to CSMA-CD systems.

The proposed system takes advantage of the time-correlation properties of the time-varying wireless channel. Further performance gains are expected through adequate con-

sideration of the frequency and space correlation properties of the channel. Further research is required to determine the most effective pilot allocation in time/space/frequency. Finally, a novel algorithm is required to combine the estimates derived from each pilot transmission (i.e. at a precise time slot, from a single transmitting antenna, on a given subcarrier) into a global estimate (i.e. the channel estimate at every time slot, from each transmitting antenna to each receiving antenna, on every subcarrier).

## VI.2.2 Coding

The notion of 'system capacity' (the sum of the capacity of each transmission eigenmode for equal power systems) has been introduced to free the analysis from assumptions on coding. However, the coding strategy is an important part of any practical transmission system. To achieve the promises offered by system capacity results, further research needs to be conducted in the following areas:

- Power and bit allocation. Several algorithms are required to allocate power to each transmission eigenmode on each subcarrier. Rate adaptation (bit allocation to each transmission eigenmode on each subcarrier) is likely to improve the performance of the system.
- Level crossing rates. In practical systems, the capacity of each transmission eigenmode on each subchannel is time-varying. Practical systems require an estimate of the speed of capacity variation to determine the required update frequency of the bit and power allocation. Further research is required to determine the level crossing rate of the eigenmode capacities.
- Practical coding schemes. It is necessary to choose or develop adequate coding schemes for SVD-MIMO-OFDM modems over the time-varying channel.

## VI.2.3 System level design

The results introduced in this thesis focus on the optimization of a single wireless link. Further research is required to determine:

- the possibilities offered by Spatial Division Multiple Access. Multiple point to point links can possibly transmit simultaneously at the same frequency when each point

to point link has a transmission eigenmode with limited interference from the other point to point links.

- the impact of multiple antenna modems on multiple access techniques. Current OFDM systems allow multiple access to the transmission medium through TDMA (Hiperlan 2) or CSMA-CD (802.11 a and g). Code Division Multiple Access (CDMA) is also a multiple access technique candidate for OFDM systems. Multiple antenna modems are likely to modify the respective performance of these multiple access techniques. E.g., multiple antenna modems require a large overhead when pilot symbol assisted channel estimation is implemented. Therefore, multiple access techniques allowing longer frames are likely to perform better with multiple antenna modems.

#### **VI.2.4 Study of the wireless channel**

The results presented in this thesis were derived under precise assumptions on the wireless channel. Further research is required to improve the MIMO wireless channel models. The time, frequency and space correlation of the channel is of particular interest. These properties of the channel can be obtained through extensive channel measurement campaigns of typical propagation environments.

#### **VI.2.5 Implementation**

The advances in the communications theory offer a wide range of possibilities to the transmission system designer. As presented throughout this thesis, the introduction of modems with multiple antennas allows an increase in the transmission data rate at the expense of complexity (further processing). However, wireless transmission devices should remain small in size and power efficient. Further research is required to:

- estimate and reduce the complexity of MIMO transmission algorithms,
- identify and solve the implementation issues linked with multiple antenna wireless modems.

# Appendix A

## Ricean channel capacity bounds

### I.1 Study of the Eigenvalues of $\mathbf{F}$

#### I.1.1 Ricean channel

The Ricean channel is defined as

$$\mathbf{H} = a\mathbf{H}^{sp} + b\mathbf{H}^{sc} \quad (\text{A.1})$$

where  $a = \sqrt{\frac{10^{K/10}}{1+10^{K/10}}}$  and  $b = \sqrt{\frac{1}{1+10^{K/10}}}$ .

$\mathbf{F}$  is defined by

$$\frac{P}{M_T}\mathbf{H}\mathbf{H}^* = b^2\mathbf{H}^{sc}(\mathbf{H}^{sc})^* + \frac{P}{M_T}\mathbf{F}. \quad (\text{A.2})$$

Therefore,

$$\mathbf{F} = ab(\mathbf{H}^{sc}(\mathbf{H}^{sp})^* + \mathbf{H}^{sp}(\mathbf{H}^{sc})^*) + a^2\mathbf{H}^{sp}(\mathbf{H}^{sp})^*. \quad (\text{A.3})$$

$\mathbf{F}$  consists of two parts: a cross product term due to the specular and scattering channel gains and a specular term. The former is itself a sum of two terms. The eigenvalues of  $\mathbf{F}$  provide useful insights to understanding the MIMO Ricean capacity. It is assumed throughout this analysis that  $K \neq -\infty$ .

#### I.1.2 Singular Value Decompositions

The matrices  $\mathbf{H}^{sp}(\mathbf{H}^{sp})^*$  and  $\mathbf{H}^{sc}(\mathbf{H}^{sp})^*$  can be written,

$$\mathbf{H}^{sp}(\mathbf{H}^{sp})^* = M_T \times (\mathbf{1})_{M_R, M_R}. \quad (\text{A.4})$$

and

$$\mathbf{H}^{sc}(\mathbf{H}^{sp})^* = \left( \sum_{k=1}^{M_T} h_{i,k} \right)_{i=1..M_R, j=1..M_R}. \quad (\text{A.5})$$

Both are rank one matrices and have the following singular value decompositions,

$$\mathbf{H}^{sp}(\mathbf{H}^{sp})^* = (\vec{v}_1) M_R M_T (\vec{v}_1^*) \quad (\text{A.6})$$

$$\mathbf{H}^{sc}(\mathbf{H}^{sp})^* = (\vec{u}_1) \sigma (\vec{v}_1^*). \quad (\text{A.7})$$

The singular values are  $M_R M_T$  and  $\sigma$  and the singular vectors are  $\vec{v}_1$  and  $\vec{u}_1$ . These are defined below,

$$\vec{v}_1 = \frac{1}{\sqrt{M_R}} (\mathbf{1})_{M_R,1}. \quad (\text{A.8})$$

The singular vector  $\vec{u}_1$  is given by  $\vec{u}_1 = \vec{x}_1 / \|\vec{x}_1\|$ , where

$$\vec{x}_1 \left( \sum_{k=1}^{M_T} h_{1,k}, \sum_{k=1}^{M_T} h_{2,k}, \dots, \sum_{k=1}^{M_T} h_{M_R,k} \right)^\dagger, \quad (\text{A.9})$$

$$\sigma = \sqrt{M_R \times \sum_{i=1}^{M_R} \left\| \sum_{k=1}^{M_T} h_{i,k} \right\|^2}, \quad (\text{A.10})$$

and  $(.)^\dagger$  denotes the transpose.

### I.1.3 Eigenvalues of $\mathbf{F}$

Using the singular value decompositions above,  $\mathbf{F}$  in (III.19) can be written as  $\mathbf{F} = a^2 M_R M_T (\vec{v}_1) (\vec{v}_1^*) + ab\sigma((\vec{v}_1) (\vec{u}_1^*) + (\vec{u}_1) (\vec{v}_1^*))$ . Hence  $\text{rank}(\mathbf{F}) \leq 2$  since  $\mathbf{F}$  is the sum of two rank one matrices,  $a^2 M_R M_T (\vec{v}_1) (\vec{v}_1^*) + ab\sigma(\vec{v}_1) (\vec{u}_1^*)$  and  $ab\sigma(\vec{u}_1) (\vec{v}_1^*)$ . By construction, it follows that any eigenvector,  $\vec{k}$ , of  $\mathbf{F}$ , associated with the non-zero eigenvalue  $\kappa$  satisfies the following,

$$\begin{cases} \exists \beta_1, \beta_2 \text{ such that } \vec{k} = \beta_1 \vec{v}_1 + \beta_2 \vec{u}_1 \\ \mathbf{F} \vec{k} = \kappa \vec{k}, \end{cases} \quad (\text{A.11})$$

Substituting for  $\mathbf{F}$  and  $\vec{k}$  and equating coefficients in (A.11) gives:

$$\begin{cases} \beta_1 ab\sigma(\vec{u}_1^* \vec{v}_1) + \beta_2 ab\sigma + \beta_1 a^2 M_R M_T + \beta_2 a^2 M_R M_T (\vec{v}_1^* \vec{u}_1) = \kappa \beta_1 \\ \beta_1 ab\sigma + \beta_2 ab\sigma(\vec{v}_1^* \vec{u}_1) = \kappa \beta_2 \end{cases} \quad (\text{A.12})$$

Defining  $o = \vec{v}_1^* \vec{u}_1$ , and solving (A.12) for  $\kappa$  gives,

$$\kappa = \frac{a^2 M_R M_T + ab\sigma(o+o^*)}{2} \pm \frac{\sqrt{(a^2 r t + ab\sigma(o+o^*))^2 + 4(ab\sigma)^2(1-oo^*)}}{2}. \quad (\text{A.13})$$

which defines the two possibly non-zero eigenvalues of  $\mathbf{F}$ .

### I.1.4 Asymptotic eigenvalues of $\mathbf{F}$

Equation (A.10) indicates that  $\sigma \geq 0$  and  $E(\sigma^2) = r^2 \times t$ . Furthermore,  $\|o\| \leq 1$ , so for  $M_R, M_T \rightarrow \infty$ ,  $(a^2 M_R M_T + ab\sigma(o+o^*))^2 \gg 4(ab\sigma)^2(1-oo^*)$  and  $a^2 M_R M_T \gg \|ab\sigma(o+o^*)\|$  with probability 1. Hence, one solution of (A.13) is positive and the other negative. Since, all other eigenvalues are zero we have the ordered eigenvalues denoted by  $\lambda_{M_R}(\mathbf{F}) < 0 = \lambda_{M_R-1}(\mathbf{F}) = \dots = \lambda_2(\mathbf{F}) < \lambda_1(\mathbf{F})$ . Taking the positive square root in (A.13) gives  $\lambda_1(\mathbf{F}) \sim a^2 M_R M_T + ab\sigma(o+o^*)$  and in the limit

$$\lambda_1(\mathbf{F})/(M_R M_T) \rightarrow a^2. \quad (\text{A.14})$$

From equation (A.13),  $\mathbf{F}$  is a matrix of maximum rank two, with one negative and one positive eigenvalue. The positive eigenvalue, denoted  $\lambda_1(\mathbf{F})$ , behaves as below

$$\lambda_1(\mathbf{F})/(M_T M_R) \rightarrow a^2. \quad (\text{A.15})$$

Hence the positive eigenvalue of  $\mathbf{F}$  grows quadratically with the number of antennas (when  $M_R = M_T$ ). Despite this,  $\mathbf{F}$  is expected to have a negligible effect in (III.18) for large numbers of antennas, since  $\mathbf{F}$  only has two eigenvalues whereas the scattering term has  $\min(M_R M_T)$  with a probability of one. The two eigenvalues of  $\mathbf{F}$  are shown in Fig. A.1 and Fig. A.2.

Now the following further observations can be made:

- The positive eigenvalue is several orders of magnitude larger than the magnitude of the negative eigenvalue. Its growth with  $M_T, M_R$  ( $M_T = M_R$ ) is quadratic.
- This disparity between the positive and negative eigenvalues increases even further when the  $K$  value is such that the channel is effectively a LOS channel.

Results not reported here show that the eigenvalues of the sum of the cross product terms in (III.19) is a matched pair of positive and negative terms. The effect of the

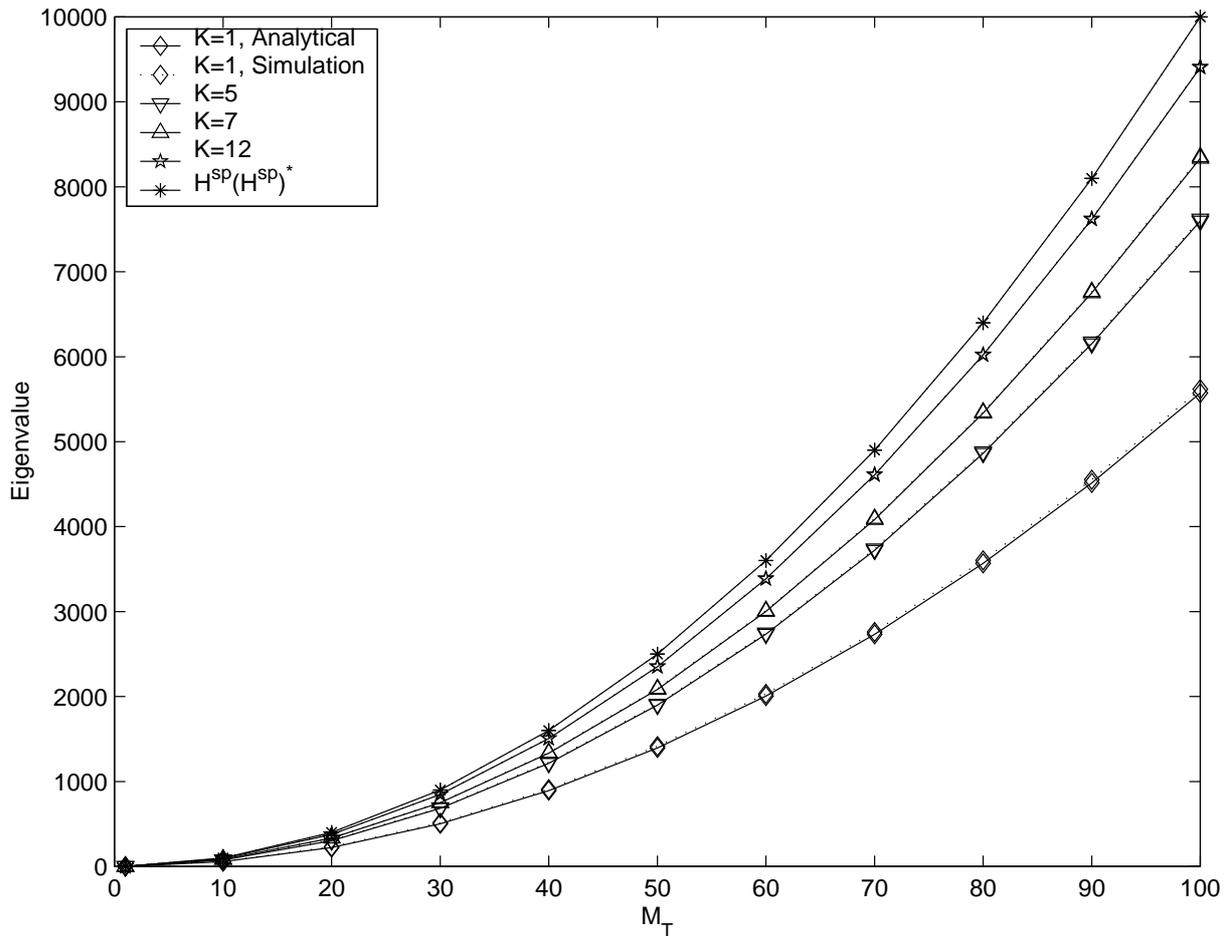


Figure A.1: Positive eigenvalue of  $F$  versus antenna numbers for various  $K$  values

term  $\mathbf{H}^{SP}(\mathbf{H}^{SP})^*$  in (III.19) is to present a large increase to the positive eigenvalue that is obvious in A.1.

Whilst the positive eigenvalue may seem quite large, its contribution to capacity is relatively small due to the logarithmic operation.

## I.2 Capacity Lower bound

We now derive the capacity lower bound.

Since  $\mathbf{H}\mathbf{H}^*$  is a non-central complex Wishart matrix we can use Bartlett's decomposition [86] to give

$$\mathbf{H}\mathbf{H}^* = b^2 \mathbf{L}^* \mathbf{L} \quad (\text{A.16})$$

where  $\mathbf{L}$  is upper triangular with diagonal elements denoted  $L_1, L_2, \dots, L_r$  which are independent of all other elements. It is assumed that  $M_R \leq M_T$  but the proof can easily be adapted to  $M_R > M_T$ . The distribution of  $L_1^2$  is non-central chi-squared,  $L_1^2 \sim \chi_{2M_T}^2(\delta)$

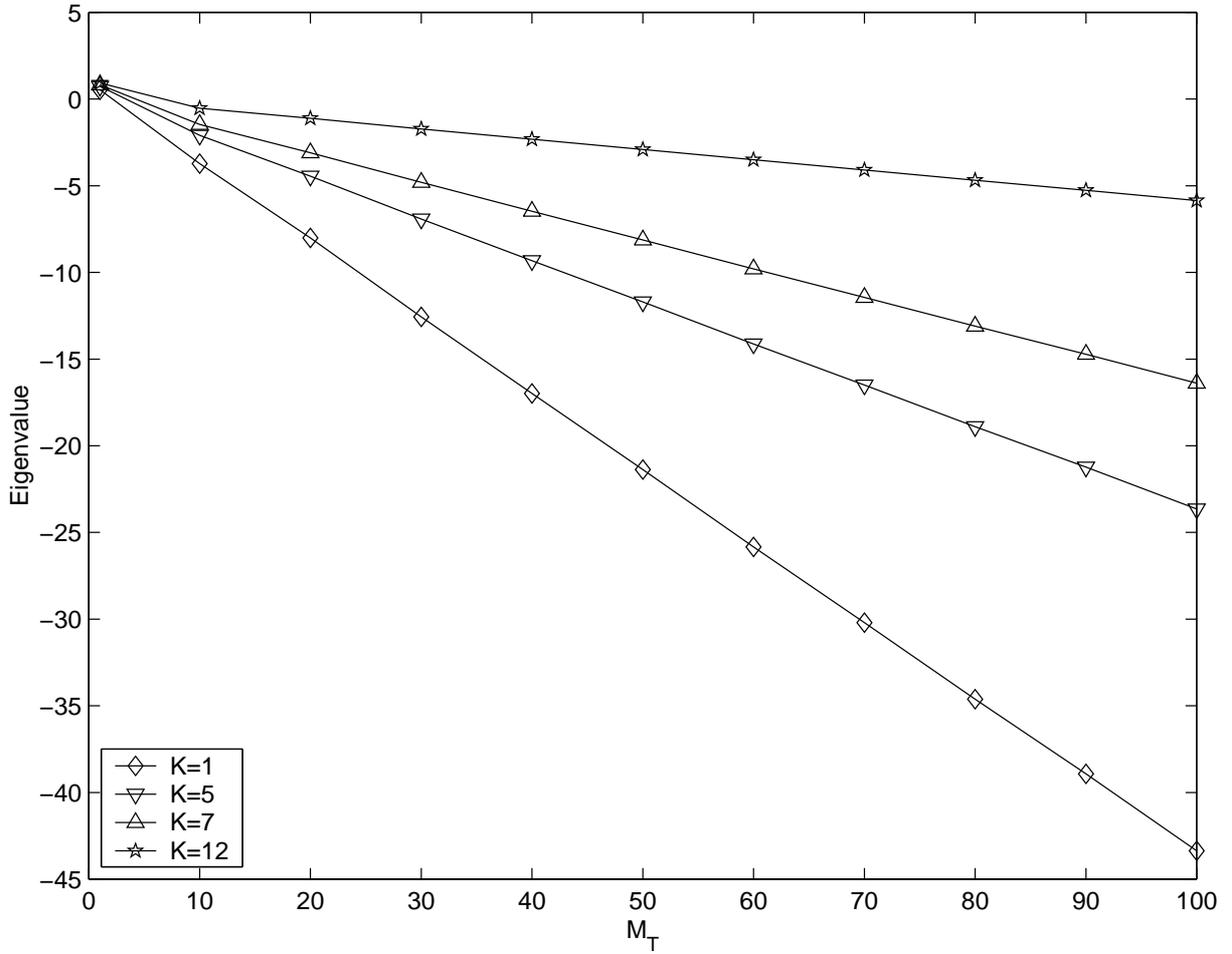


Figure A.2: Negative eigenvalue of  $F$  versus antenna numbers for various  $K$  values

with  $\delta = (a^2/b^2)\text{trace}(\mathbf{H}^{sp}(\mathbf{H}^{sp})^*)$ . For  $j > 1$  the distributions are central chi-squared,  $L_j^2 \sim \chi_{2M_T-2j+2}^2$ . Hence,

$$D = |\mathbf{I}_{M_R} + \frac{\rho}{M_T} \mathbf{H} \mathbf{H}^*| = \left| \left[ \mathbf{I}_{M_R} \sqrt{\frac{b^2 \rho}{M_T}} \mathbf{L}^* \right] \left[ \begin{array}{c} \mathbf{I}_{M_R} \\ \sqrt{\frac{b^2 \rho}{M_T}} \mathbf{L} \end{array} \right] \right| \quad (\text{A.17})$$

Using the Cauchy-Binet theorem gives

$$D = \sum_{\gamma} |\mathbf{A}_{\gamma}| |\mathbf{A}_{\gamma}|^* = \sum_{\gamma} |\mathbf{A}_{\gamma}|^2, \quad (\text{A.18})$$

where  $\mathbf{A}_{\gamma}$  is an  $M_R \times M_R$  submatrix of  $[\mathbf{I}_{M_R} \sqrt{\frac{b^2 \rho}{M_T}} \mathbf{L}^*]$  and  $\gamma$  is a subset of  $M_R$  columns from  $(1, 2, \dots, 2M_R)$ .

Now the summation is split into two parts, over  $\gamma_1$  where the determinants  $|\mathbf{A}_{\gamma_1}|$  do not involve  $L_1$  and over  $\gamma_2$  where the determinants  $|\mathbf{A}_{\gamma_2}|$  do involve  $L_1$ . Hence

$$D = \sum_{\gamma_1} |\mathbf{A}_{\gamma_1}|^2 + \sum_{\gamma_2} |\mathbf{A}_{\gamma_2}|^2. \quad (\text{A.19})$$

The only choice of columns which gives determinants involving  $L_1$  are those where column  $M_R + 1$  is selected and column 1 is omitted. Hence the  $\mathbf{A}_{\gamma_2}$  matrices are of the form

$$\mathbf{A}_{\gamma_2} = \begin{bmatrix} 0 & \dots & 0 & \sqrt{\frac{b^2 P}{M_T}} L_1 & 0 & \dots & 0 \\ & & & 0 & & & \\ & & D_{\gamma_{21}} & \vdots & & D_{\gamma_{22}} & \\ & & & 0 & & & \end{bmatrix}. \quad (\text{A.20})$$

Hence  $|\mathbf{A}_{\gamma_2}|^2 = \frac{b^2 P}{M_T} L_1^2 |D_{\gamma_2}|^2$  where  $D_{\gamma_2} = [D_{\gamma_{21}} D_{\gamma_{22}}]$  and

$$\begin{aligned} D &= \sum_{\gamma_1} |\mathbf{A}_{\gamma_1}|^2 + \frac{b^2 P}{M_T} L_1^2 \sum_{\gamma_2} |D_{\gamma_2}|^2 \\ &= X + L_1^2 Y. \end{aligned} \quad (\text{A.21})$$

The exact same analysis holds for the Rayleigh case, except  $L_1^2 \sim \chi_{2M_T}^2$ .

To summarize,

$$\begin{aligned} D^{\text{Ricean}} &= X + \chi_{2M_T}^2(\delta) Y \\ D^{\text{Rayleigh}} &= X + \chi_{2M_T}^2 Y \end{aligned} \quad (\text{A.22})$$

where  $X, Y$  are positive random variables with  $X, Y$  independent of the  $\chi^2$  variables.

Hence

$$\begin{aligned} E[C(\mathbf{H})] &= E[\log_2(X)] + E[\log_2(1 + \chi_{2M_T}^2(\delta) Y/X)] \\ E[C(b\mathbf{H}^{\text{sc}})] &= E[\log_2(X)] + E[\log_2(1 + \chi_{2M_T}^2 Y/X)]. \end{aligned} \quad (\text{A.23})$$

Now  $\chi_{2M_T}^2(\delta)$  is stochastically greater than  $\chi_{2M_T}^2$ . Hence  $E[f(\chi_{2M_T}^2(\delta))] \geq E[f(\chi_{2M_T}^2)]$  for any increasing function  $f$  and  $E[C(\mathbf{H})] \geq E[C(b\mathbf{H}^{\text{sc}})]$  as required. A lower bound of the capacity can now be written  $\forall M_R, M_T, P, K$ ,

$$E \left[ \frac{C(K, M_T, M_R, P)}{\min(M_T, M_R)} \right] \geq E \left[ \frac{C(K = -\infty, M_T, M_R, b^2 P)}{\min(M_T, M_R)} \right]. \quad (\text{A.24})$$

### I.3 Capacity Upper bound

The capacity upper bound is derived in this Section. Defining

$$\mathbf{A} = \mathbf{I}_{M_R} + \frac{b^2 P}{M_T} \mathbf{H}^{\text{sc}} (\mathbf{H}^{\text{sc}})^*, \quad (\text{A.25})$$

the normalized capacity becomes

$$\frac{C}{M_T} = \frac{1}{M_T} \log_2(|\mathbf{A} + \tilde{\mathbf{F}}|) = \frac{1}{M_T} \log_2 \left( \prod_{i=1}^{M_R} \lambda_i(\mathbf{A} + \tilde{\mathbf{F}}) \right), \quad (\text{A.26})$$

where  $\tilde{\mathbf{F}} = \frac{P}{M_T}\mathbf{F}$  and  $\lambda_i(\mathbf{A} + \tilde{\mathbf{F}})$  are the eigenvalues of the hermitian positive definite matrix  $\mathbf{A} + \tilde{\mathbf{F}}$ , ordered so that  $0 \leq \lambda_{M_R}(\mathbf{A} + \tilde{\mathbf{F}}) \leq \dots \leq \lambda_1(\mathbf{A} + \tilde{\mathbf{F}})$ . Combining Weyl's theorem [87] and results from Appendix I.1 leads to

$$\begin{aligned}\lambda_{M_R}(\mathbf{A} + \tilde{\mathbf{F}}) &\leq \lambda_{M_R-1}(\mathbf{A}) + \lambda_2(\tilde{\mathbf{F}}) = \lambda_{M_R-1}(\mathbf{A}) \\ \lambda_{M_R-1}(\mathbf{A} + \tilde{\mathbf{F}}) &\leq \lambda_{M_R-2}(\mathbf{A}) + \lambda_2(\tilde{\mathbf{F}}) = \lambda_{M_R-2}(\mathbf{A}) \\ &\vdots \\ \lambda_2(\mathbf{A} + \tilde{\mathbf{F}}) &\leq \lambda_1(\mathbf{A}) + \lambda_2(\tilde{\mathbf{F}}) = \lambda_1(\mathbf{A}) \\ \lambda_1(\mathbf{A} + \tilde{\mathbf{F}}) &\leq \lambda_1(\mathbf{A}) + \lambda_1(\tilde{\mathbf{F}})\end{aligned}\tag{A.27}$$

Therefore,

$$\begin{aligned}\frac{C}{M_T} &\leq \frac{1}{M_T} \log_2(\lambda_{M_R-1}(\mathbf{A}) \dots \lambda_1(\mathbf{A})(\lambda_1(\mathbf{A}) + \lambda_1(\tilde{\mathbf{F}}))) \\ &= \frac{1}{M_T} \log_2(\prod_{i=1}^{M_R} \lambda_i(\mathbf{A})) + \frac{1}{M_T} \log_2\left(\frac{\lambda_1(\mathbf{A}) + \lambda_1(\tilde{\mathbf{F}})}{\lambda_{M_R}(\mathbf{A})}\right).\end{aligned}\tag{A.28}$$

Now write

$$\Delta = \frac{1}{M_T} \log_2\left(\frac{\lambda_1(\mathbf{A}) + \lambda_1(\tilde{\mathbf{F}})}{\lambda_{M_R}(\mathbf{A})}\right),\tag{A.29}$$

and, since  $\lambda_j(\mathbf{A}) \geq 1$  for any  $j$ ,

$$\Delta \leq \frac{1}{M_T} \log_2(\lambda_1(\mathbf{A}) + \frac{P}{M_T} \lambda_1(\mathbf{F})).\tag{A.30}$$

It is known that the eigenvalues of  $\mathbf{A}$  are bounded as  $M_R, M_T \rightarrow \infty$  [72]. Therefore,  $\exists M$  such that  $\lambda_1(\mathbf{A}) \leq M$  and when  $M_T \rightarrow \infty$ ,

$$\Delta \leq \frac{1}{M_T} \log_2(M + PM_R(\lambda_1(\mathbf{F})/(M_R M_T))) \rightarrow 0,\tag{A.31}$$

since  $\lambda_1(\mathbf{F})/(M_R M_T) \rightarrow a^2$ . This concludes the demonstration. From [72], it is known that

$$\lambda_1(\mathbf{A}) \rightarrow 1 + b^2 P(1 + \sqrt{\min(M_T, M_R)/\max(M_T, M_R)})^2,\tag{A.32}$$

as  $M_T, M_R \rightarrow \infty$  with  $M_T/M_R = \alpha$ . This provides the smallest value for  $M$  that can be used and gives the bound that is used in the simulations.

Therefore the capacity upper bound as  $M_T, M_R \rightarrow \infty$  can be written,

$$E \left[ \frac{C(K, M_T, M_R, P)}{\min(M_T, M_R)} \right] \leq E \left[ \frac{C(K = -\infty, M_T, M_R, b^2 P)}{\min(M_T, M_R)} \right] + \Delta,\tag{A.33}$$

where  $\Delta \rightarrow 0$  as  $M_T, M_R \rightarrow \infty$ .

Hence, for Ricean channels that are not pure LOS ( $K \neq +\infty$ ), the normalized ergodic capacity tends to the normalized ergodic capacity of the scattering component. Hence,

$$E \left[ \frac{C(K, M_T, M_R, P)}{\min(M_T, M_R)} \right] \rightarrow E \left[ \frac{C(K = -\infty, M_T, M_R, b^2 P)}{\min(M_T, M_R)} \right].\tag{A.34}$$



# Appendix B

## IEE Electronics Letters

G. Lebrun, T. Ying, M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *IEE Electronics Letters*, vol. 37, no. 32, pp. 1363-1364, Oct. 2001.



# Appendix C

## IEEE Global Telecommunications Conference 2002

G. Lebrun, T. Ying and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *Proc. IEEE Global Telecommunications Conference, (Globecom '02)*, Nov. 17-21, 2002 pp. 414-418.



# Appendix D

## IEEE Vehicular Technology Conference 2003

T. Ying, G. Lebrun and M. Faulkner, "An adaptive channel SVD tracking strategy in time-varying TDD system," *Proc. IEEE Vehicular Technology Conference 2003, (VTC 2003-Spring)*, Apr. 22-25, 2003 Vol. 1, pp. 769-773.



# Appendix E

## IEEE International Conference on Communications 2004

G. Lebrun, M. Faulkner, P. J. Smith and M. Shaafi, "MIMO Ricean channel capacity," in *Proc. IEEE International Conference on Communications (ICC 2004)*, 20-24 Jun. 2004, pp. 2939-2943.



# Appendix F

## IEEE International Conference on Communications 2004

G. Lebrun, S. Spiteri and M. Faulkner, "Channel estimation for an SVD-MIMO System," in *Proc. IEEE International Conference on Communications (ICC 2004)*, 20-24 Jun. 2004, pp. 3025-3029.



# Appendix G

## IEEE Transactions on wireless communications

G. Lebrun, J. Gao and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *IEEE Transactions on wireless communications*, vol. 4, no. 2, pp. 757-764, Mar. 2005.



# Appendix H

## IEEE Transactions on wireless communications

G. Lebrun, M. Faulkner, P. J. Smith and M. Shaafi, "MIMO Ricean channel capacity: an asymptotic analysis," *IEEE Transactions on wireless communications*, Vol. 5, No. 5, May 2006,



# Bibliography

- [1] H. Nyquist, “Certain factors affecting telegraph speed,” *Bell System Technical Journal*, vol. 3, p. 324, 1924.
- [2] ———, “Certain topics in telegraph transmission theory,” *AIEE transactions*, vol. 47, pp. 617–644, 1928.
- [3] R. V. L. Hartley, “Transmission of information,” *Bell System Technical Journal*, vol. 7, p. 535, 1928.
- [4] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, July 1948.
- [5] ———, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 623–656, Oct. 1948.
- [6] A. N. Kolmogorov, “Sur l’interpolation et extrapolation des suites stationnaires,” *Comptes rendus de l’academie des sciences*, vol. 208, p. 2043, 1939.
- [7] N. Wiener, *The extrapolation, interpolation and smoothing of stationary time series with engineering applications*. New York: Wiley, 1949.
- [8] J. M. Whittaker, “Interpolatory function theory,” *Cambridge Tracts in Mathematics and Mathematical Physics*, no. 33, 1935.
- [9] C. E. Shannon, “Communication in the presence of noise,” *Proceedings of the IRE*, vol. 37, pp. 10–21, Jan. 1949.
- [10] R. W. Hamming, *Coding and information theory*. Englewoods Cliffs, New Jersey 07632: Prentice-Hall, 1986.
- [11] C. Berrou and A. Glavieux, “Near optimum-error correcting coding and decoding: turbo-codes,” *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

- [12] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
- [13] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY, 10020: McGraw-Hill, Inc., 2001.
- [14] J. B. Andersen, "Array gain and capacity for known random channels with multiple elements arrays at both ends," *IEEE J. Select. Areas Commun.*, vol. 18, no. 11, pp. 2172–2178, Nov. 2000.
- [15] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
- [16] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," in *Proc. IEEE International Conference on Communications (ICC2002)*, New York, April 28-May 2 2002, pp. 406–410.
- [17] E. Biglieri and G. Tarico, "Large-system analysis of multiple-antennas system capacities," *Journal of Communications and Networks*, vol. 5, no. 2, pp. 96–103, June 2003.
- [18] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.
- [19] W. Rhee and J. M. Cioffi, "On the capacity of multiuser wireless channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2580–2595, Oct. 2003.
- [20] G. Lebrun, M. Faulkner, P. J. Smith, and M. Shafi, "MIMO Ricean channel capacity," in *Proc. of the 2004 IEEE International Conference Communication (ICC2004)*, 20-24 Jun. 2004, pp. 2939–2943.
- [21] ———, "MIMO Ricean channel capacity: an asymptotic analysis," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, May 2006.
- [22] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1639–1642, July 1999.

- [23] J.-C. Belfiore, M. O. Damen, and K. Abed-Meraim, "Generalized sphere decoder for asymmetrical space-time communication architecture," *IEE Electronic Letters*, vol. 36, no. 2, pp. 166–167, Jan. 2000.
- [24] A. M. Chan and I. Lee, "A new reduced-complexity sphere decoder for multiple antenna systems," in *Proc. IEEE International Conference on Communications (ICC2002)*, New York, April 28-May 2 2002, pp. 460–464.
- [25] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-Blast space-time communication architecture," *IEE Electronics Letters*, vol. 35, no. 1, pp. 14–16, Jan. 1999.
- [26] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-Blast: an architecture for realizing very high data rates over the rich scattering wireless channel," in *Proc. IEEE International Symposium on Signals, Systems and Electronics (ISSSE 98)*, Pisa, 29 Sept.-2 Oct. 1998, pp. 295–300.
- [27] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
- [28] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
- [29] M. Stoytchev and H. Safar, "Statistics of the MIMO radio channel in indoor environments," in *Proc. IEEE Vehicular Technology Conference (VTC 2001 Fall)*, 7-11 Oct. 2001, pp. 1804–1808.
- [30] P. Kyritsi and D. C. Cox, "Correlation properties of MIMO radio channels for indoor scenarios," in *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, 4-7 Nov. 2001, pp. 994–998.
- [31] D. McNamara, M. Beach, and P. Fletcher, "Spatial correlation in indoor MIMO channels," in *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRIC 2002)*, 15-18 Sept. 2002, pp. 290–294.
- [32] J. P. Kermoal, L. Schumacher, P. E. Mogensen, and K. I. Pedersen, "Experimental investigation of correlation properties of MIMO radio channels for indoor picocell

- scenarios,” in *Proc. IEEE Vehicular Technology Conference (VTC 2000 Fall)*, 24-28 Sept. 2000, pp. 14–21.
- [33] P. Kyritsi, D. C. Cox, R. A. Valenzuela, and P. W. Wolniansky, “Correlation analysis based on MIMO channel measurements in an indoor environment,” *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 713–720, June 2003.
- [34] D. McNamara, M. Beach, P. Fletcher, and P. Karlsson, “Capacity variation of indoor multiple-input multiple-output channels,” *IEE Electronics Letters*, vol. 36, no. 24, pp. 2037–2038, Nov. 2000.
- [35] D. Chizhik, F. Rashid-Farrokhi, J. Ling, and A. Lozano, “Effect of antenna separation on the capacity of BLAST in correlated channels,” *IEEE Commun. Lett.*, vol. 4, no. 11, pp. 337–339, Nov. 2000.
- [36] S. L. Loyka, “Channel capacity of MIMO architecture using the exponential correlation matrix,” *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sept. 2001.
- [37] S. L. Loyka and G. Tsoulos, “Estimating MIMO system performance using the correlation matrix approach,” *IEEE Commun. Lett.*, vol. 6, no. 1, pp. 19–21, Jan. 2002.
- [38] V. V. Veeravalli, A. Sayeed, and Y. Liang, “Asymptotic capacity of correlated MIMO Rayleigh fading channels via virtual representation,” in *Proc. IEEE International Symposium on Information Theory (ISIT 2003)*, Yokohama, 29 Jun.-4 Jul. 2003, p. 247.
- [39] V. Raghavan and A. M. Sayeed, “MIMO capacity scaling and saturation in correlated environments,” in *Proc. IEEE International Conference on Communications (ICC ’03)*, 2003, pp. 3006–3010.
- [40] Z. Hong, K. Liu, R. W. J. Heath, and A. M. Sayeed, “Spatial multiplexing in correlated fading via the virtual channel representation,” *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 856–866, June 2003.
- [41] C.-N. Chuah, J. M. Kahn, and D. Tse, “Capacity of multi-antenna array systems in indoor wireless environment,” in *Proc. IEEE Global Telecommunications Conference (GLOBECOM 98)*, 8-12 Nov. 1998, pp. 1894–1899.

- [42] D.-S. Shiu, M. J. G. Gerard J. Foschini, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.
- [43] C.-N. Chuah, D. N. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.
- [44] C. Martin and B. Ottersten, "Analytic approximations of eigenvalue moments and mean channel capacity for MIMO channels," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02)*, 13-17 May 2002, pp. III2389–III2392.
- [45] A. Giorgetti, M. Chiani, M. Shafi, and P. J. Smith, "Level crossing rates and MIMO capacity fades: Impacts of spatial/temporal channel correlation," in *Proc. IEEE International Conference on Communications (ICC '03)*, 2003, pp. 3046–3050.
- [46] S. Loyka and A. Kouki, "On the use of Jensen's inequality for MIMO channel capacity estimation," in *Canadian Conference on Electrical and Computer Engineering, 2001*, 13-16 May 2001, pp. 475–480.
- [47] ———, "New compound upper bound on MIMO channel capacity," *IEEE Commun. Lett.*, vol. 6, no. 3, pp. 96–98, Mar. 2002.
- [48] ———, "The impact of correlation on multi-antenna system performance: correlation matrix approach," in *Proc. IEEE Vehicular Technology Conference (VTC 2001 Fall)*, 7-11 Oct. 2001, pp. 533–537.
- [49] M. Khalighi, J.-M. Brossier, G. Jourdain, and K. Raouf, "On capacity of Rician MIMO channels," in *Proc. 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRIC 2001)*, 30 Sept.-3 Oct. 2001, pp. A150–A154.
- [50] P. Driessen and G. Foschini, "On the capacity formula for multiple input-multiple output wireless channels: a geometric interpretation," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 173–176, Feb. 1999.

- [51] P. J. Smith and L. Garth, "Exact capacity distribution for dual MIMO systems in Ricean fading," *IEEE Commun. Lett.*, vol. 8, no. 1, pp. 18–20, Jan.
- [52] A. Tulino, A. Lozano, and S. Verdu, "Capacity of multi-antenna channels in the low-power regime," in *Proc. of the 2002 IEEE Information Theory Workshop*, 20-25 Oct. 2002, pp. 192–195.
- [53] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. Valenzuela, "Link-optimal space-time processing with multiple transmit and receive antennas," *IEEE Commun. Lett.*, vol. 5, no. 3, pp. 85–87, Mar. 2001.
- [54] E. Jorswieck, G. Wunder, V. Jungnickel, and T. Haustein, "Inverse eigenvalue statistics for Rayleigh and Rician MIMO channels," in *Proc. IEE Seminar on MIMO: Communications Systems from Concept to Implementations*, 12 Dec. 2001, pp. 3/1–3/6.
- [55] J. Ayadi, A. Hutter, and J. Farserotu, "On the multiple input multiple output capacity of Rician channels," in *Proc. 5th International Symposium on Wireless Personal Multimedia Communications*, 27-30 Oct. 2002, pp. 402–406.
- [56] G. Raleigh and J. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.
- [57] B.-G. Song and J. A. Ritcey, "Prefilter design using the singular value decomposition for MIMO equalization," in *Conference record of the thirtieth Asilomar conference on signals, systems and computers*, 3 Nov.-6 Nov. 1996, pp. 34–38.
- [58] G. Lebrun, T. Ying, and M. Faulkner, "MIMO transmission over a time-varying TDD channel using SVD," *IEE Electronics Letters*, vol. 37, no. 22, pp. 1363–1364, Oct. 2001.
- [59] W. Y. Tao, R. S. Cheng, and K. B. Letaief, "Adaptive space-time coding system in fading channels," in *Proc. IEEE Vehicular Technology Conference (VTC 2001)*, 6 May-9 May 2001, pp. 103–107.
- [60] B. N. Getu, J. B. Andersen, and J. R. Farserotu, "MIMO systems: optimizing the use of eigenmodes," in *IEEE Proc. on Personal, Indoor and Mobile Radio Communications (PIMRC 2003)*, 7 Sept.-10 Sept. 2003, pp. 1129–1133.

- [61] B. N. Getu and J. B. Andersen, "BER and spectral efficiency of a MIMO system," in *Proc. IEEE International Symposium on Wireless Personal Multimedia Communications*, 27 Oct.-30 Oct. 2002, pp. 397–401.
- [62] H. Zhiqiang, T. Baoyu, and W. Weiling, "Performance analysis of multiple antenna diversity in transmitter and receiver known channels," in *Proc. International Conference on Communication Technology (ICCT)*, 9 Apr.-11 Apr. 2003, pp. 1903–1906.
- [63] K. Kim, S.-K. Lee, and K. Chang, "An efficient multiuser access scheme combining the transmit diversity with the modified SVD methods for MIMO channels," in *Conference record of the thirty-sixth Asilomar conference on signals, systems and computers*, 3 Nov.-6 Nov. 2002, pp. 1719–1721.
- [64] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.
- [65] G. Raleigh and J. Cioffi, "Spatio-temporal coding for wireless communications," in *Proc. IEEE Global Telecommunications Conference (Globecom'96)*, November 18-22 1996, pp. 1809–1814.
- [66] P. Robertson, "Illuminating the structure of code and decoder of parallel concatenated recursive systematic (turbo) codes," in *Proc. IEEE Global Telecommunications Conference (Globecom'94)*, vol. 3, 28 Nov.-2 Dec. 1994, pp. 1298–1303.
- [67] Y. Kegen, J. Evans, and I. Collings, "Performance analysis of pilot symbol aided QAM for Rayleigh fading channels," in *Proc. IEEE International Conference on Communications (ICC 2002)*, 28 Apr.-2 May 2002, pp. 1731–1735.
- [68] F. Tufvesson and T. Maseng, "Pilot assisted channel estimation for OFDM in mobile cellular systems," in *Proc. IEEE Vehicular Technology Conference (VTC 1997)*, 4 May.-7 May 1997, pp. 1639–1643.
- [69] R. Heath and G. Giannakis, "Blind channel estimation for multirate precoding and OFDM systems," in *Proc. International Conference on Digital Signal Processing (DSP 1997)*, 2 Jul.-4 Jul. 1997, pp. 383–386.

- [70] C. Ming-Xian and Y. Su, "Blind joint channel and data estimation for OFDM signals in Rayleigh fading," in *Proc. IEEE Vehicular Technology Conference (VTC 2001)*, 6 May-9 May 2001, pp. 791–795.
- [71] G. H. Golub and C. F. V. Loan, *Matrix computations*. Baltimore, London: The Johns Hopkins University Press, 1989.
- [72] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, Massachusetts Institute of Technology, 1989.
- [73] S. Spiteri, G. Lebrun, and M. Faulkner, "Capacity performance of hardware erred MIMO systems in slowly fading Rayleigh channels," in *Proc. Australian Communication Theory Workshop (AusCTW)*, 2004.
- [74] G. Lebrun, T. Ying, and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," in *Proc. IEEE Global Telecommunications Conference (Globecom'02)*, 17-21 Nov. 2002, pp. 414–418.
- [75] G. Lebrun, J. Gao, and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, Mar. 2005.
- [76] J. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.
- [77] G. Lebrun, S. Spiteri, and M. Faulkner, "Channel estimation for an SVD-MIMO system," in *Proc. of the 2004 IEEE International Conference Communication (ICC2004)*, 20-24 Jun. 2004, pp. 3025–3029.
- [78] K. R. Dandekar, H. Ling, and G. Xu, "Smart antenna array calibration procedure including amplitude and phase mismatch and mutual coupling effects," in *Proc. IEEE International Conference on Personal Wireless Communications*, 17-20 Dec. 2000, pp. 293–297.
- [79] G. W. Stewart and J. guang Sun, *Matrix perturbation theory*. San Diego, CA 92101: Academic Press, Inc., 1990.
- [80] T. S. Rappaport, *Wireless Communications*. New Jersey: Prentice Hall PTR, 1996.

- [81] W. C. J. Jakes, *Microwave Mobile Communications*. New York: J. Wiley and Sons, 1974.
- [82] M. D. Austin and G. L. Stüber, “Velocity adaptive handoff algorithms for micro-cellular systems,” in *Record of the International Conference on Universal Personal Communications*, 12 Oct.-15 Oct. 1993, pp. 793–797.
- [83] J. T. Barnett and B. Kedem, “Zero-crossing rates of functions of Gaussian processes,” *IEEE Trans. Inform. Theory*, vol. 37, no. 4, pp. 1188–1194, July 1991.
- [84] ———, “Zero-crossing rates of mixtures and products of Gaussian processes,” *IEEE Trans. Inform. Theory*, vol. 44, no. 4, pp. 1672–1677, July 1998.
- [85] V. Jungnickel, V. Pohl, U. Kruger, C. von Helmolt, T. Haustein, and S. Stanczak, “A radio system with multi-element antennas,” in *Proc. IEEE 53rd Vehicular Technology Conference*, 2001, pp. 167–170.
- [86] R. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: John Wiley & Sons Inc., 1982.
- [87] R. Horn and C. Johnson, *Matrix Theory*. Cambridge: Cambridge University Press, 1985.