Uppsala University
Signals and Systems

ON MIMO SYSTEMS AND ADAPTIVE ARRAYS FOR WIRELESS COMMUNICATION

Analysis and Practical Issues

Mattias Wennström

UPPSALA UNIVERSITY 2002

Dissertation for the degree of Doctor of Philosophy
in Signal Processing at Uppsala University, 2002.

ABSTRACT

Wennström, M., 2002. On MIMO Systems and Adaptive Arrays for Wireless Communications: Analysis and Practical Aspects, 291 pp. Uppsala. ISBN 91-506-1619-6.

This thesis is concerned with the use of multiple antenna elements in wireless communication over frequency non-selective radio channels. Both measurement results and theoretical analysis are presented. New transmit strategies are derived and compared to existing transmit strategies, such as beamforming and space time block coding (STBC). It is found that the best transmission algorithm is largely dependent on the channel characteristics, such as the number of transmit and receive antennas and the existence of a line of sight component. Rayleigh fading multiple input multiple output (MIMO) channels are studied using an eigenvalue analysis and exact expressions for the bit error rates and outage capacities for beamforming and STBC is found. In general are MIMO fading channels correlated and there exists a mutual coupling between antenna elements. These findings are supported by indoor MIMO measurements. It is found that the mutual coupling can, in some scenarios, increase the outage capacity. An adaptive antenna testbed is used to obtain measurement results for the SIMO channel. The results are analyzed and design guidelines are obtained for how a beamformer implemented in hardware shall be constructed. The effects of nonlinear transmit amplifiers in array antennas are also analyzed, and it is shown that an array reduces the effective intermodulation distortion (IMD) transmitted by the array antenna by a spatial filtering of the IMD. A novel frequency allocation algorithm is proposed that reduces IMD even further. The use of a low cost antenna with switchable directional properties, the switched parasitic antenna, is studied in a MIMO context and compared to array techniques. It is found that it has comparable performance, at a fraction of the cost for an array antenna.

Keywords: antenna array, calibration, mutual coupling, MIMO system, flat fading, nonlinear amplifier, switched parasitic antenna, analog beamformer, adaptive antenna testbed.

Mattias Wennström, Signals and Systems, Uppsala University, PO Box 528, SE-751 20 Uppsala, Sweden. Email: mattias.wennstrom@signal.uu.se.

This thesis has been prepared using LaTeX.

ISBN 91-506-1619-6

Printed in Sweden by Elanders Gotab AB, Stockholm, September 2002.
Distributed by Signals and Systems, Department of Material Science,
Uppsala University, Uppsala, Sweden.

To my parents

Contents

I may not have gone
where I intended to go,
but I think I have ended up
where I intended to be.

Douglas Adams

Acknowledgments

When I first stepped onto the unpaved path of research five years ago, I had an almost blank map in my hand, where only the largest tracks were drawn. I started walking, and during my journey I managed to draw the map myself. Today, the map is almost complete and I hope it can be useful for other travellers. I have been down many frustrating sidings, but I have also found new useful paths that lead to new discoveries, which is very rewarding and the source of my motivation. Several individuals have played an important role in the journey that led to this thesis. First of all, I would like to express my gratitude to my supervisors and co-authors, Professor Anders Rydberg and Dr. Tommy Öberg, for their support and for always believing in my capabilities, especially when I was lost for a while.

I am very grateful to Professor Anders Ahlén and Professor Mikael Sternad for all their help and encouraging words during the years and for reading and constructively criticizing the thesis manuscript.

And I especially want to acknowledge three of my co-authors during these years: First, Tekn.Lic. Jonas Strandell with whom I spent several weeks in a bunker under an airfield in Arboga, desperately trying to repair an unruly smart antenna, it was "träligt" but fun. We also enjoyed Chinese food at the local restaurant, every day, for weeks. Secondly, I acknowledge Dr. Erik Lindskog, who's diurnal rhythm always is synchronized to a GMT-7 hr Silicon Valley clock, independently of which country he is working in. Thanks also to Dr. Thomas Svantesson, I will never forget when we hatched the idea of using parasitic antennas during a pub visit in Boston.

I gratefully acknowledge all my former and present colleagues at the Signals and Systems Group for creating such a pleasant work environment and for helping me to solve many mathematical problems at the whiteboard during the "fika"-breaks. Especially the members of the Thursday club, consisting of some Ph.D. Students at the Signals and Systems Group plus our guest researchers Carmen Alonso and Dr. Monica Nicoli, are acknowledged. The Thursday club's weekly dinner endeavors at some "nation" has been a source of most welcome distraction. Especially our seemingly never-ending discussions about train movies and

the definition of strict altruism.

Two friends have had a special impact on me during these five years. I would like to thank Dr. Håkan Engqvist who is an excellent lunch partner who gladly discuss anything from tribology to Seinfeld episodes. Also Kristina Terenius is thanked for spreading her energy and pushing me to keep my heart and lungs in shape and also for being an invigorating discussion partner.

I also owe my thanks to my parents, Ann-Cathrin and Hans, who always encourage and support me in all my decisions. You have been patient.

Finally, I would like to express my deepest gratitude and appreciation to my fiancée Mian for her love and patience. You have been with me the whole journey. Thank you for all support, without you this thesis would never have been completed.

Mattias Wennström
Uppsala, October 2002.

Abbreviations and Acronyms

3G	Third Generation Mobile Telephony System
3GPP	3G-Partnership Project
4G	Fourth Generation Wireless Communication System
ACLR	Adjacent Channel Leakage Ratio
AMPS	Advanced Mobile Phone System
ABF	Analog Beam-Former
ADC	Analog to Digital Converter
AGC	Automatic Gain Control
AOA	Angle Of Arrival
AOD	Angle Of Departure
AP	Access Point
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BFN	Beam-Forming Network
BLAST	Bell Labs Layered Space-time architecture
BMT	Butler Matrix Transformer
BS	Basestation
BW	Beamwidth
CCI	Co-Channel Interference
CDMA	Code Division Multiple Access
CIMR	Carrier to Intermodulation plus interference Ratio
CINR	Carrier to Interference plus Noise Ratio
CIR	Carrier to Interference Ratio
CNR	Carrier to Noise Ratio

CRB	Cramér Rao Lower Bound
CSI	Channel State Information
CW	Continuous Wave
DAB	Digital Audio Broadcasting
DCS-1800	Digital Communication System - 1800 MHz
DBF	Digital Beam-Former
DF	Direction Finding
DFT	Discrete Fourier Transform
DMT	Discrete Multi-Tone modulation
DSP	Digital Signal Processor
DSL	Digital Subscriber Line
DOA	Direction Of Arrival
DVB	Digital Video Broadcasting
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FFT	Fast Fourier Transform
FRX	Feedback Receiver
GMSK	Gaussian Minimum Shift Keying
GSM	Global System for Mobile telecommunication
HFSS	High Frequency Structure Simulator
IBO	Input Back-Off
IM	Intermodulation
IMD	Intermodulation Distortion
IMP	Intermodulation Power
LD	Linear Dispersion
LMS	Least Mean Square
LOS	Line Of Sight
MCPA	Multi-Carrier Power Amplifier
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
ML	Maximum Likelihood
MLSE	Maximum Likelihood Sequence Estimator
MMSE	Minimum Mean Square Error
MQAM	Multilevel-QAM
MS	Mobile Station
MUSIC	Multiple Signal Classification
NLOS	Non-LOS
NPR	Noise Power Ratio
OBO	Output Back-Off
OFDM	Orthogonal Frequency Division Multiplex

PAE	Power Added Efficiency
PDF	Probability Density Function
PE	Parasitic Element
PEP	Pairwise Error Probability
PRBS	Pseudo Random Binary Sequence
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
RX	Receiver
SAF	Shimbo's Amplitude Function
SCPA	Single-Carrier Power Amplifier
SDMA	Spatial Division Multiple Access
SIMO	Single Input Multiple Output
SIR	Signal to Interference Ratio
SISO	Single Input Single Output
SLL	Side-Lobe Level
SMI	Sample Matrix Inversion
SML	Stochastic Maximum Likelihood
SNR	Signal to Noise Ratio
SPA	Switched Parasitic Antennas
SRX	Sampling Receiver
STBC	Space Time Block Code
STTC	Space Time Trellis Code
SVD	Singular Value Decomposition
SWR	Software Radio
TCM	Trellis Coded Modulation
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TRX	Transceiver
TX	Transmitter
UE	User Equipment
ULA	Uniform Linear Array
UMTS	Universal Mobile Telephony System
V-BLAST	Vertical-BLAST
WCDMA	Wide-band CDMA
WLAN	Wireless Local Area Network
ZF	Zero Forcing

Notation and used symbols

Scalar variables are in this thesis written as plain lower-case letters, vectors as bold face lower-case letters and matrices as bold-face upper-case letters. Some further notations used in the thesis include the following:

$A_{ij}, \mathbf{A}_{:,j}$	The ij:th element and j:th column of the matrix $\mathbf{A}$ respectively.
$\mathbf{A}^*, \mathbf{A}^T$	Complex conjugate transpose (Hermitian transpose) and transpose respectively of the matrix $\mathbf{A}$.
vec($\mathbf{A}$)	The vectorization operator, vec stacks the columns of $\mathbf{A}$ into a vector, i.e. when $\mathbf{A}$ has n columns: $\text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{A}_{:,1} \\ \vdots \\ \mathbf{A}_{:,n} \end{bmatrix}$
$\mathbf{A} \otimes \mathbf{B}$	The Kronecker product. When $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $r \times s$, then $\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}_{1,1}\mathbf{B} & \mathbf{A}_{1,2}\mathbf{B} & \cdots & \mathbf{A}_{1,n}\mathbf{B} \\ \mathbf{A}_{2,1}\mathbf{B} & \mathbf{A}_{2,2}\mathbf{B} & \cdots & \mathbf{A}_{2,n}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_{m,1}\mathbf{B} & \mathbf{A}_{m,2}\mathbf{B} & \cdots & \mathbf{A}_{m,n}\mathbf{B} \end{bmatrix}$ is of dimension $mr \times ns$.

$\lambda_{min}(\mathbf{A}), \lambda_{max}(\mathbf{A})$	The smallest and largest eigenvalue of the matrix $\mathbf{A}$ respectively
diag$[x_1, \ldots, x_n]$	A $n \times n$ diagonal matrix with diagonal elements $x_1, \cdots, x_n$.
rk($\mathbf{A}$)	The rank of $\mathbf{A}$.
$\mathbf{A}^\dagger$	The Moore-Penrose pseudo inverse of a $m \times n$ matrix $\mathbf{A}$. If $m \geq n$ and $\mathbf{A}$ is of full rank it is defined as $\mathbf{A}^\dagger = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$.
$\|\mathbf{A}\|_F$	The Frobenius norm of $\mathbf{A}$, $\|\mathbf{A}\|_F^2 = \text{Tr}(\mathbf{A}\mathbf{A}^*)$.
Re($\mathbf{A}$),Im($\mathbf{A}$)	The real and imaginary part of the complex matrix $\mathbf{A}$.
Tr($\mathbf{A}$)	The trace of the matrix $\mathbf{A}$, Tr($\mathbf{A}$)=$\sum_i \mathbf{A}_{ii}$.
$E\{\mathbf{A}\}$	Statistical expectation of the matrix $\mathbf{A}$. Sometimes $E_x\{\mathbf{A}\}$ for expectation with respect to x.
$\mathbf{I}$	The identity matrix, sometimes $\mathbf{I}_n$ for the $n \times n$ identity matrix.
$Q(x)$	The Q-function, or integrated normalized Gaussian density, $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\lambda^2/2} d\lambda$.
$\min_{\mathcal{S}}(V(\mathcal{S}))$	The minimium value of the function $V(\cdot)$ in the set $\mathcal{S}$.
$\arg\max_{\mathcal{S}}(V(\mathcal{S}))$	The maximizing argument of the function $V(\cdot)$ over the set $\mathcal{S}$.
$J_m(x)$	The m:th order Bessel function of the first kind.
$I_m(x)$	The m:th order modified Bessel function of the first kind.

$Ci(x)$	The cosine integral. $Ci(x) = -\int_x^\infty \frac{\cos y}{y} dy.$
$Si(x)$	The sine integral. $Si(x) = \int_x^\infty \frac{\sin y}{y} dy.$
$Ei(n, x)$	The exponential integral. $Ei(n, x) = \int_1^\infty t^{-n} e^{-xt} dt.$
$\Gamma(z, x)$	The incomplete gamma function. $\Gamma(z, x) = \int_0^x e^{-t} t^{z-1} dt.$
$\Gamma(z)$	The gamma function, $\Gamma(z) = \Gamma(z, \infty)$. When z is a positive integer $\Gamma(z) = (z-1)!$
$F(\alpha, \beta; \gamma; x)$	Gauss hypergeometric function. The hypergeometric series is $F(\alpha, \beta; \gamma; x) = 1 + \frac{\alpha \cdot \beta}{\gamma \cdot 1} x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1) \cdot 1 \cdot 2} x^2 + \cdots$
$(x)^+$	Is a short notation for $\max(0, x)$.
δ_{ij}	The Kronecker delta function.
sgn(x)	The sign of the real scalar x, sgn(x)=$x/\|x\|$.
q^{-1}	The unit delay operator. For any signal $x(n)$, $q^{-1}x(n) = x(n-1)$.
$\bar{s}(n), \tilde{s}(n)$	The real and imaginary part of the symbol $s(n)$. $s(n) = \bar{s}(n) + j\tilde{s}(n)$.
$\breve{s}(t)$	The complex lowpass equivalent of the signal $s(t)$, $s(t) = Re\left\{\breve{s}(t) \cdot e^{j\omega_c t}\right\}$.

Here follows a list of some commonly used symbols in this thesis:

f_c	Carrier frequency
f_D	Doppler frequency
f_s	Sampling frequency
B_c	Channel coherence bandwidth
C	Channel capacity
C_f	Total number of frequency channels in a FDMA system
c	The speed of electromagnetic waves in vacuum (299792458 m/s)
I	Mutual information
k_c	Wavenumber, $k_c = 2\pi/\lambda_c$
K	Ricean K-factor
L	The number of scatterers
M	Number of blocks in one super-codeword
M_d	The number of signal points of a modulation scheme
M_f	Number of frequency channels per BS
N	Block length in symbol intervals
N_I	The number of co-channel interferers
m_r	The number of receive modes
m_t	The number of transmit modes
n_p	The number of parasitic antenna elements
n_r	The number of receive antenna elements
n_s	The number of symbols per codeword
n_t	The number of transmit antenna elements
P_T	Total transmitted power
p	The path loss exponent
R	The code rate
R_{fb}	Bit rate in feedback channel
r_d	Number of bits per symbol, $r_d = \log_2(M_d)$
T_{coh}	Channel coherence time
T_s	Symbol time
T_{samp}	Sampling period
W	Transmission bandwidth
Z_A	Antenna impedance
Z_T	Transmitter impedance
Δ	Angular spread
δ	Inter-element spacing for uniform linear array
δ_{RX}	Receive array inter-element spacing
δ_{TX}	Transmit array inter-element spacing
Γ	Power protection ratio
λ_c	Carrier wavelength
η	Intrinsic impedance for free space
σ_n^2	Receiver thermal noise variance
ω_c	Angular frequency of the carrier

Chapter 1

Introduction

WIRELESS systems are now popular worldwide to help people and machines to communicate with each other irrespectively of their location. So far, using a cellular system is by far the most common wireless method to access data or to perform voice dialling. But in a near future, we will be surrounded by a numerous of options to set up an unwired connection over the radio interface. One of the slogans for the fourth generation wireless communications system (4G) is "always best connected", meaning that your wireless equipment should connect to the network or system that at the moment is the "best" for you. Various connections ranges from satellites that provides low bit rates but global coverage and cellular systems with continental coverage to high bit rate local area networks and personal area networks with a maximum range of a few to a hundred meters. If these systems should co-exist, then we would obtain a crowded frequency spectrum, since there are many different actors that wants their share of the limited frequency resource. To use a signalling strategy that is *spectrally efficient* is thus of utmost importance.

The current trend to achieve high spectral efficiency is by utilizing adaptivity in the ever changing radio environment and sources of interference. Adaptivity on the physical layer can be used in all possible dimensions: Time, frequency, power and space. Adaptivity can also be used on higher signalling layers to boost performance even further; an example is multi-user scheduling.

This thesis is devoted to the physical layer of wireless communication systems and will focus mainly on the adaptive utilization of the space dimension. Space utilization is possible through the use of multiple antenna elements arranged in an array, for the transmission and/or reception of the signals. Or, in some cases a single antenna element that has several polarizations or modes is used to obtain polarization or angle diversity. In the mid-1990s, the terms "smart antennas" and

"adaptive antennas" were introduced since through signal processing, the array antenna can be made adaptive, and in a smart manner change its transmission or reception characteristics when the radio environment changes.

Using an array of antenna elements to improve a wireless connection is an old technique, even used by Guglielmo Marconi in 1901 to increase the gain of the Atlantic transmissions of Morse codes [1]. Marconi used four 61 meter high tower antennas arranged in a circular array in Poldhu, England to transmit the Morse signal for the letter "S", the distance 3425 km to Signal Hill, St. John, Newfoundland. Today, array antennas in wireless communication systems are used to improve performance in several ways, not just for range improvement. Systems with array antennas in commercial operation have been reported for GSM networks [2, 3], fixed broadband wireless access networks (BWA) [4] and 3G CDMA networks [5]. A numerous of field trials with testbed antenna arrays have also been reported, see [6–12] among others.

This trend in wireless communications has been possible mainly through the advents of signal processing, digital signal processors and high speed ADC. The main arguments today, for using multiple antennas when transmitting over a wireless link are:

Array gain Due to the use of multiple antennas, the antenna gain is increased and this leads to an increased range and coverage. This is useful in remote areas with low population. A large area can thus be served with less basestations. Alternatively, the transmit power of the mobile units can be reduced due to the increased gain, or sensitivity, of the receiving basestation antenna array.

Interference suppression By using the spatial dimension provided by multiple antenna elements, it is possible to suppress interfering signals in a way that is not possible with a single antenna. Hence, the system can be tuned to be less susceptible to interference and the distance between basestations using the same time/frequency channel can be reduced, which is beneficial in densely populated areas. This leads to a *system capacity* improvement.

Spatial diversity Multiple antennas can also be used to counteract the channel fading due to multipath propagation. Sufficiently spaced multiple antennas at the receiver gives copies of the transmitted signal that has propagated through channels with different fading. The probability that all signal copies are in a deep fade simultaneously is small. Thus, spatial diversity increases the robustness of the wireless link and this can be utilized to obtain a higher data throughput or to decrease the transmission power. A *link capacity* improvement is thus obtained.

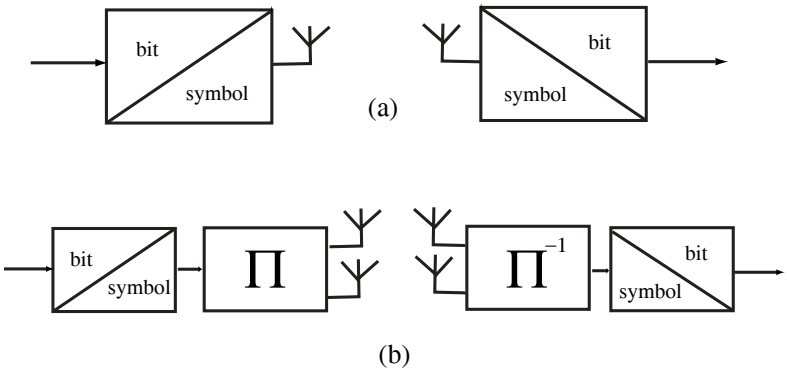


Figure 1.1: The concept of (a) SISO and (b) MIMO systems. A MIMO system has an additional mapping Π *and de-mapper* Π^{-1} *which can be optimized to decrease BER or increase spectral efficiency of the wireless link.*

Transmitter localization A receiver array antenna can be used to localize the transmitter, just as we can use our both ears to localize the source of a sound in a room without using our eyes. This has application in positioning services and emergency call localization.

Systems consisting of a transmitter, radio channel and receiver are often categorized by their number of inputs and outputs. Multiple inputs or outputs to a channel are realized as multiple antennas, polarizations or antenna modes. The trivial configuration is a single antenna at each side of the radio channel, hence a single input and single output system commonly abbreviated as a SISO system. With the use of multiple antennas on either or both sides of the wireless link, MISO, SIMO and MIMO systems can be defined in a similar manner.

The main difference between a SISO wireless communication system and a MIMO system with n_t transmit antennas and n_r receive antennas is the additional moment of mapping from a single stream of data symbols to n_t streams of symbols and the corresponding inverse operation at the receiver. This is illustrated in Figure 1.1 where the mappings[1] are denoted by the blocks Π and Π^{-1}.

The MIMO definition is the most general and contains the SISO, SIMO and

[1]The majority of MIMO systems can be described by the concept illustrated in Figure 1.1. However, there exists methods where the bit to symbol mapping and symbol to antenna mapping cannot be separated as in Figure 1.1(b).

MISO systems as special cases. The fundamental problem is, *how shall the mappings* Π *and* Π^{-1} *be designed to optimize the performance of the wireless link ?*. Performance might be expressed in the following entities

Bit rate The goal is to provide the highest possible bit rate per unit bandwidth. This is often defined as the spectral efficiency. The ultimate limit or highest possible bit rate that can be achieved, with arbitrarily low bit error probability in a Gaussian noise channel[2], was derived by Shannon and is sometimes used as a reference. In cellular systems, the spectral efficiency is sometimes defined as the number of transmitted bits per bandwidth per unit area of coverage.

Reliability Often conflicting with high bit rate is the reliability or robustness of the transmission which can be measured using the average bit error rate. Since the radio channel is time varying (unless the mobile is standing still), it is important to choose a communication strategy that can withstand the fading dips in the received SNR caused by the multipath propagation.

Complexity As one end of the wireless link might be battery powered, it is important to have low complexity mappings or inverse mappings since an algorithm with higher complexity will be more power consuming. Hence, it is often desirable to design the transmission/reception schemes so that the complexity is non-symmetric, locating the low complexity algorithms at the battery powered side. Hence, the mappings Π, Pi^{-1} are usually different on different sides of the channel in a duplex system.

The objectives above are often mutually conflicting and the system designer must carefully choose a trade-off, including the economic and the mobility aspects as well.

An important factor that have impact on the mapping Π in Figure 1.1 is knowledge of the instantaneous MIMO channel parameters at the transmitter side. This divides the mapping Π into two subclasses, one with channel state information (CSI) at the transmitter and one without. CSI at the transmitter requires in most cases a feedback channel, which consumes system bandwidth, but improves robustness and spectral efficiency. Hence, the possible use of a feedback channel is another tradeoff that must be considered. The mappings Π and Π^{-1} are sometimes implemented in hardware, which gives a system with lower computational complexity. This will degrade the performance somewhat compared to a software solution and the degradation depends on the channel scenario. This is further explored in Chapter 5.

[2]If the time delay is not a matter of concern.

A simple example is now given to illustrate how the mappings Π and Π^{-1} can be constructed in the case of a MIMO system.

EXAMPLE 1.1

Assume that a person is using a portable computer that has a wireless connection. The portable computer is assumed to be equipped with two antennas and it is currently connected to a base-station which also has two antennas, hence, we have a 2×2 MIMO system. Due to reflections and scattering of the transmitted signals from the base-station, the received signals at the computer antennas will be a superposition of many reflected signals that sometimes add constructively and sometimes destructively. This is called *multipath fading*. Furthermore, since the environment (cars, people,..) around this person is not stationary, the channel between a transmitter antenna and a receiver antenna will be time-varying. We will now compare different transmission and reception strategies (different Π and Π^{-1}), that can be used in this simple example. These are the spatial multiplexing, the space time block coding and the beamforming methods, where beamforming requires CSI at the transmitter.

Assume that the symbol to be transmitted from the base-station at time n is $s(n)$, where $|s(n)| = 1$. Write the received and transmitted signals as 2×1 vectors $\mathbf{y}(n)$ and $\mathbf{c}(n)$ respectively and $\mathbf{H}$ as the 2×2 channel matrix, which is assumed constant during the time it takes to transmit at least two transmit vectors $\mathbf{c}(n), \mathbf{c}(n+1)$. The channel matrix $\mathbf{H}$ is assumed to be known at the receiver through an estimation process, utilizing for instance a training sequence in the transmitted data. Then the MIMO input-output relation can be written as

$$\mathbf{y}(n) = \mathbf{H}\mathbf{c}(n) + \mathbf{v}(n) \tag{1.1}$$

where $\mathbf{v}(n)$ is the 2×1 receiver noise vector which is assumed to be white in space and time with covariance matrix $\sigma_n^2 \mathbf{I}$.

Spatial Multiplexing

If spatial multiplexing is used, then the mapping Π is obtained by first de-multiplexing the incoming data stream $s(n)$ into two data streams, $s_e(n)$ and $s_o(n)$. This can be made by for example directing all symbols with even index n into $s_e(n)$ and odd index n into stream $s_o(n)$. Then the transmission vector is created as

$$\mathbf{c}(n) = \sqrt{\frac{P_T}{2}} \begin{bmatrix} s_e(n) \\ s_o(n) \end{bmatrix} \tag{1.2}$$

where the factor $\sqrt{\frac{P_T}{2}}$ ensures that the total transmitted power is P_T. The two data streams are then transmitted on separate antennas and the received signals on the two receive antennas becomes

$$\begin{aligned} y_1(n) =& \sqrt{\frac{P_T}{2}} \{H_{11}s_e(n) + H_{12}s_o(n)\} + v_1(n) \\ y_2(n) =& \sqrt{\frac{P_T}{2}} \{H_{21}s_e(n) + H_{22}s_o(n)\} + v_2(n) \ . \end{aligned} \tag{1.3}$$

One way to recover the transmitted signals is to multiply the received vector, $\mathbf{y}(n)$, with the pseudo-inverse, $\mathbf{H}^\dagger \triangleq (\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*$, of the channel matrix. This is called zero-forcing (ZF) equalization and yields

$$\hat{\mathbf{c}}(n) = \mathbf{H}^\dagger \mathbf{y}(n) = \mathbf{H}^\dagger \mathbf{H}\mathbf{c}(n) + \mathbf{H}^\dagger \mathbf{v}(n) \ . \tag{1.4}$$

Since $\mathbf{H}$ in this example is 2×2, $\mathbf{H}^\dagger\mathbf{H} = \mathbf{I}$ and the channel impact on the transmitted vector is removed. Note that the receiver noise vector $\mathbf{v}(n)$ is multiplied by the pseudo inverse of the channel matrix. This leads to *noise amplification* which can destroy the possibility to detect $\mathbf{c}(n)$ correctly. The SNR for the ZF receiver is

$$\text{SNR}_k^{ZF} = \frac{P_T}{2\sigma_n^2 \left[\mathbf{H}^*\mathbf{H}\right]_{kk}^{-1}} \tag{1.5}$$

where $k = 1, 2$ gives the SNR for the two subchannels.

Noise amplification is a problem with spatial multiplexing transmission strategies, although better methods exists, such as the minimum mean square error (MMSE) receiver, which balance the error in channel equalization and noise amplification in an mean square error optimal way. Furthermore, additional antennas on the receiver side of the channel reduces the probability that the channel matrix $\mathbf{H}$ is ill conditioned. The benefits of spatial multiplexing with two transmit antennas is that two symbols are transmitted per transmitted vector, so the bit rate is high. In the general case with n_r receive and n_t transmit antennas, $\min(n_r, n_t)$ symbols can be transmitted per transmit vector $\mathbf{c}(n)$.

Space-time block coding

When space-time block codes (STBC) are used in this 2×2 example, the mapping, Π, of the data sequence is performed over two consecutive transmit vectors

$\mathbf{c}(n), \mathbf{c}(n+1)$ as

$$\begin{aligned}\mathbf{c}(n) =& \sqrt{\frac{P_T}{2}} \begin{bmatrix} s(n) \\ s(n+1) \end{bmatrix} \\ \mathbf{c}(n+1) =& \sqrt{\frac{P_T}{2}} \begin{bmatrix} -s^*(n+1) \\ s^*(n) \end{bmatrix} \quad .\end{aligned} \tag{1.6}$$

The reason for this mapping will be explained in Chapter 2. Since two symbols are transmitted per transmit vector, but twice, the total equivalent bit rate is 1 symbol per transmit vector $\mathbf{c}(n)$. The receive vector $\mathbf{y}(n)$ has the components

$$\begin{aligned} y_1(n) =& \sqrt{\frac{P_T}{2}} \left\{ H_{11} s(n) + H_{12} s(n+1) \right\} + v_1(n) \\ y_2(n) =& \sqrt{\frac{P_T}{2}} \left\{ H_{21} s(n) + H_{22} s(n+1) \right\} + v_2(n) \\ y_1(n+1) =& \sqrt{\frac{P_T}{2}} \left\{ -H_{11} s^*(n+1) + H_{12} s^*(n) \right\} + v_1(n+1) \\ y_2(n+1) =& \sqrt{\frac{P_T}{2}} \left\{ -H_{21} s^*(n+1) + H_{22} s^*(n) \right\} + v_2(n+1) \quad . \end{aligned} \tag{1.7}$$

With this particular structure on $\mathbf{c}(n), \mathbf{c}(n+1)$, also called the *Alamouti code* [13], the receiver (de-mapper Π^{-1}) takes the form

$$\begin{aligned} \hat{s}(n) =& H_{11}^* y_1(n) + H_{12} y_1^*(n+1) + H_{21}^* y_2(n) + H_{22}^* y_2^*(n+1) \\ \hat{s}(n+1) =& H_{12}^* y_1(n) - H_{11} y_1^*(n+1) + H_{22}^* y_2(n) - H_{21} y_2^*(n+1) \end{aligned} \tag{1.8}$$

which, if (1.1) is inserted becomes

$$\begin{aligned} \hat{s}(n) =& \sqrt{\frac{P_T}{2}} \left\{ |H_{11}|^2 + |H_{12}|^2 + |H_{21}|^2 + |H_{22}|^2 \right\} s(n) + v'(n) \\ \hat{s}(n+1) =& \sqrt{\frac{P_T}{2}} \left\{ |H_{11}|^2 + |H_{12}|^2 + |H_{21}|^2 + |H_{22}|^2 \right\} s(n+1) + v'(n+1) \end{aligned} \tag{1.9}$$

where $v'(n), v'(n+1)$ are noise terms that are linear combinations of the elements in $\mathbf{v}(n), \mathbf{v}(n+1)$. Note that the detection becomes completely decoupled, that is, the detection of $s(n)$ is independent of the detection of $s(n+1)$. It is possible to write the equations (1.8) as a matrix multiplication which is a linear operation and this yields a very low complexity receiver for the STBC. The

SNR:s for the symbols $s(n), s(n+1)$ are equal and found to be (see Chapter 2)

$$\begin{aligned} \text{SNR}^{STBC} =& \frac{P_T}{2\sigma_n^2}\left\{|H_{11}|^2 + |H_{12}|^2 + |H_{21}|^2 + |H_{22}|^2\right\} \\ =& \frac{P_T}{2\sigma_n^2}(\lambda_1 + \lambda_2) \end{aligned} \tag{1.10}$$

where $\lambda_1 \geq \lambda_2$ are the two eigenvalues to the matrix $\mathbf{HH}^*$. The benefits of using this transmission scheme is the reduced variability in the SNR, which gives a robust transmission channel. The SNR for each symbol is proportional to the squared magnitude of all 4 channels in the 2×2 MIMO system. Hence, this scheme is a *diversity* scheme. The drawback compared to spatial multiplexing is that the bit rate is halved. For larger systems than 2×2, the maximum obtainable bit rate is even smaller than a fraction $1/n_t$ of the corresponding spatial multiplexing rates.

Beamforming

If full CSI is available at the transmitter, for instance, if there is a feedback channel from the receiver to the transmitter, then beamforming can be utilized. Beamforming gives diversity, as the space time block code, but also an array gain, since the symbols can be transmitted so they combine coherently at the receiver antennas. Denote the transmitter and receiver beamforming vectors as $\mathbf{w}_T$ and $\mathbf{w}_R$ respectively, where the power constraint $|\mathbf{w}_T|^2 = P_T$ is applied to the transmit beamforming vector. Then the transmitted vector is found by the following mapping Π:

$$\mathbf{c}(n) = \mathbf{w}_T s(n) \tag{1.11}$$

and the input-output relation is

$$\hat{s}(n) = \mathbf{w}_R^* \mathbf{y}(n) = \mathbf{w}_R^* \mathbf{H} \mathbf{w}_T s(n) + \mathbf{w}_R^* \mathbf{v}(n) \ . \tag{1.12}$$

If $\mathbf{w}_R$ and $\mathbf{w}_T$ are chosen as the principal left and right singular vectors to $\mathbf{H}$ respectively, then the receiver SNR is maximized and this relation can be expressed as

$$\hat{s}(n) = \lambda_1^{1/2} s(n) + \mathbf{w}_R^* \mathbf{v}(n) \tag{1.13}$$

where λ_1 is the largest eigenvalue to the matrix $\mathbf{HH}^*$. Hence, the SNR for the beamforming method is

$$\text{SNR}^{BF} = \frac{P_T}{\sigma_n^2}\lambda_1 \ . \tag{1.14}$$

The bit rate for this beamforming method is one transmitted symbol per transmit vector, the same as for the space-time block coding system described above.

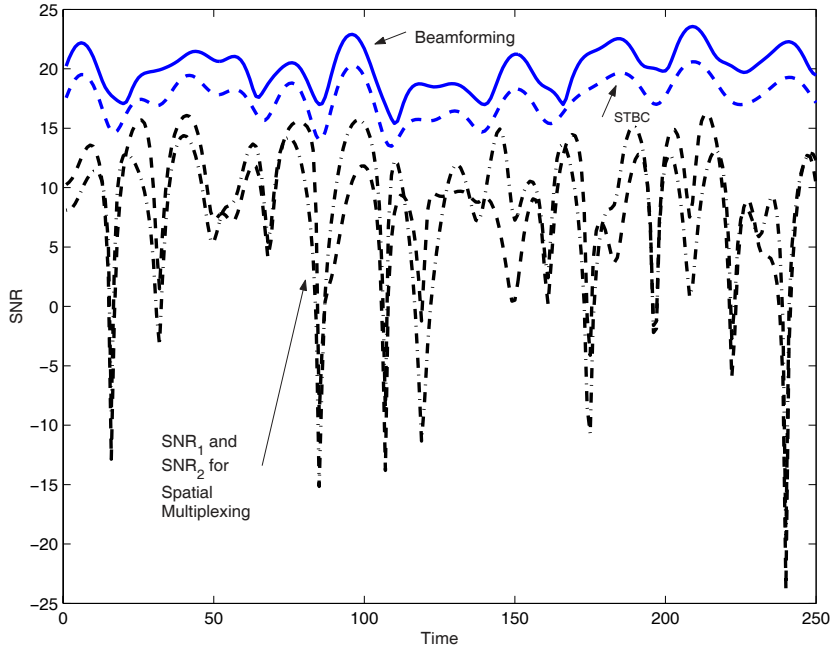


Figure 1.2: Comparisons of the received signal to noise ratios for a 2×2 MIMO system with different transmission/reception strategies. In the simulation, $P_T/\sigma_n^2 = 100$. The time axis has arbitrary scale units, depending on the product of the channel Doppler frequency f_d and the symbol time T_s, which are not specified in this example.

Comparisons

We shall now compare the three exemplified methods to construct the mappers Π and Π^{-1}, using a computer generated fading MIMO channel. The results are shown in Figure 1.2 where the time scale units on the x-axis depends on the Doppler frequency of the channel. We see that beamforming has an average 3 dB gain over space time block coding (STBC), due to combining gain, however it requires CSI at the transmitter. The SNR curve for beamforming and STBC shows a low variability compared to spatial multiplexing. This is due to the diversity obtained in these two methods, which "hardens" the equivalent channel and makes it more reliable for communication. Furthermore, spatial multiplexing with the ZF receiver shows the worst performance, but one must remember that it has the double bit rate compared to the two other schemes and better receivers for spatial multiplexing than the ZF receiver used in this example exists.

To conclude, the choice of transmission strategy depends on what is most

important for this particular wireless connection; reliability, hardware complexity or bit rate. STBC has low complexity since the receiver is linear, and it requires no CSI at the transmitter. It is reliable due to the diversity property but the bit rate is low. Spatial multiplexing provides high bit rate, especially for large systems, it is less reliable but can be made more robust (than shown in this example) against channel variations at the cost of increased complexity. Beamforming has diversity gain, hence it is robust against channel fading, but it requires CSI. It has also an average to high complexity since a singular value decomposition of the channel matrix must be performed. The bit rate for beamforming is low, equal to the STBC bit rate, in this example. However, for larger MIMO systems than 2×2, beamforming has a higher data rate than STBC. It should also be mentioned that beamforming and STBC can operate over MISO channels as well, so this gives an option to put all the complexity (several antennas) at one side of the wireless link, preferably the non-battery powered side.

It shall be noted that the mappings can be designed to take co-channel interference (CCI) from other users into account. If something is known about the CCI, then it can be utilized by the mappings Π, Π^{-1} to improve the link performance. The array antenna can then not only provide SNR gain, but also signal to interference (SIR) gain since the interfering signals are suppressed.

1.1 Modelling the wireless MIMO system

In this section the necessary prerequisites for the following chapters of this thesis are presented. To be able to analyze and improve a wireless communication system, models for signals, for hardware, and also for the channel are needed.

1.1.1 Signal models

Many digital information-bearing signals are transmitted using a modulated carrier over a bandwidth constrained channel. If the signal and channel (system) bandwidths are small compared to the carrier frequency, then the system is said to be *narrowband* and *bandpass*. When analyzing communication systems, it is often unnecessary to model the up- and down-conversion from the baseband to the carrier frequency, so one can choose to work with baseband models, or equivalent low pass signals and channels, which then becomes *complex valued*. If a signal has the form

$$x(t) = A(t) \cdot \sin\left(2\pi f_c t + \phi(t)\right) \tag{1.15}$$

where f_c is the carrier frequency and $A(t), \phi(t)$ are the amplitude and phase of the envelope respectively, then its complex baseband equivalent is $\breve{x}(t)$ where

$$x(t) = \text{Re}\left\{\breve{x}(t) \cdot e^{j2\pi f_c t}\right\} \tag{1.16}$$

and thus $\breve{x}(t) = A(t)e^{j\phi(t)}$. The rest of this thesis uses complex baseband representation of signals if not stated otherwise and the breve on the variables is dropped. A guide on how to represent bandpass signals, systems and stochastic processes can be found in e.g. [14, 15].

Furthermore, the models and signals are often assumed to be discrete in time, i.e. *sampled* with an uniform sampling rate, spaced T_{samp}, where $f_s = 1/T_{samp}$ is the sampling frequency. The continuous time signal $x(t)$ is then replaced by $x(nT_{samp})$ or just $x(n)$, where n is a non-negative integer. In Chapter 6, where non-linearities in wireless systems are investigated, continuous-time signal models are used, however, for computer simulations of non-linear systems the continuous-time signals will be sampled at a sampling rate much higher than the symbol rate. The choice of sampling frequency when simulating non-linear systems depends on the accuracy and the desired number of higher order harmonics in the output signal.

1.1.2 Channel models

A casual, linear[3], discrete-time multiple-input, multiple-output model is used to describe the multi-element antenna wireless system with n_t transmit antennas and n_r receive antennas. Due to multipath propagation from the transmitter to the receiver, the received signal at a certain antenna contains a weighted sum of previously transmitted symbols. Thus, in the most general case, the MIMO channel model can be written as the infinite series

$$\boldsymbol{\mathcal{H}}(q^{-1}) = \sum_{n=0}^{\infty} \mathbf{H}_n q^{-n} \tag{1.17}$$

where $\mathbf{H}_n$ are matrices with dimension $n_r \times n_t$ that contains complex valued elements that represent the attenuation and phase shift for the received signal with delay n. The i, j:th element of $\boldsymbol{\mathcal{H}}$ is the transfer function from transmitter j to receiver i. In practice, it is impossible to estimate the infinite number of channel matrices $\mathbf{H}_n$ from a finite number of data, so often a linear model with finite number of parameters is used, described by a rational function [17] or by limiting the

[3]This means that the modulation format must also be linear, however some non-linear modulation formats can be approximated by a linear modulation. See for example the linearization of the GMSK modulation used in GSM [16].

number of terms in (1.17) which results in a finite impulse response (FIR) channel model. The discrete time complex low pass channel model must also incorporate the effect of the pulse shaping and sometimes also the effect of receive and transmit filtering.

The most common configuration is that the transmitters and receivers are connected to different antennas. Other options include connections to different polarizations of a single antenna [18, 19], different beams in a multibeam antenna [20], a multimode antenna [21], a switched beam antenna [22], or even combinations of these [23]. Hence, there are many configurations that all are covered by the channel model (1.17). Degenerate cases of the MIMO channel model in (1.17) are the single-input multiple-output (SIMO), multiple-input, single-output (MISO) where $\boldsymbol{\mathcal{H}}(q^{-1})$ is a column or row vector respectively and the single-input, single-output (SISO) case where $\boldsymbol{\mathcal{H}}(q^{-1})$ is a sum of scalar terms.

Assume that we would like to model the communication between a terminal and a base-station or vice versa and $N_I + 1$ multiple users are active and sharing the common radio space in all dimensions; time, frequency and space, see Figure 1.3. Then N_I users are said to be *interfering* with the particular transmission under investigation (the *desired* user) and are termed *co-channel interferers* (CCI). Using the channel model (1.17), the MIMO communication system with co-channel interferers can be written as

$$\mathbf{y}(n) = \boldsymbol{\mathcal{H}}(q^{-1})\mathbf{c}(n) + \sum_{i=1}^{N_I} \boldsymbol{\mathcal{H}}_i(q^{-1})\mathbf{c}_i(n) + \mathbf{n}_T(n) \tag{1.18}$$

where $\mathbf{y}(n)$ is the received signal vector of dimension $n_r \times 1$ and $\mathbf{n}_T(n)$ is the thermal noise in the receivers. Furthermore, $\mathbf{c}(n)$ and $\mathbf{c}_i(n)$ are the transmitted vector from our desired user and the interferers respectively, and $\boldsymbol{\mathcal{H}}(q^{-1})$ is the channel for the desired user. The matrices $\left\{\boldsymbol{\mathcal{H}}_i(q^{-1})\right\}_{i=1}^{N_I}$ are the channels for the interfering users with row dimension n_r and a column dimension that matches the number of transmit antennas of the particular co-channel interferer, i.e. the row dimension of $\mathbf{c}_i(n)$.

Definitions for the flat fading model

The general MIMO model (1.17) is now restricted to the flat fading scenario. Also slow fading is assumed, that is, the channel coefficients are constant during the transmission of several symbols. Flat fading, or frequency non-selective fading, applies by definition to systems where the bandwidth of the transmitted signal is much less than the coherence bandwidth of the channel. Then all the frequency components of the transmitted signal undergoes the same attenuation and phase

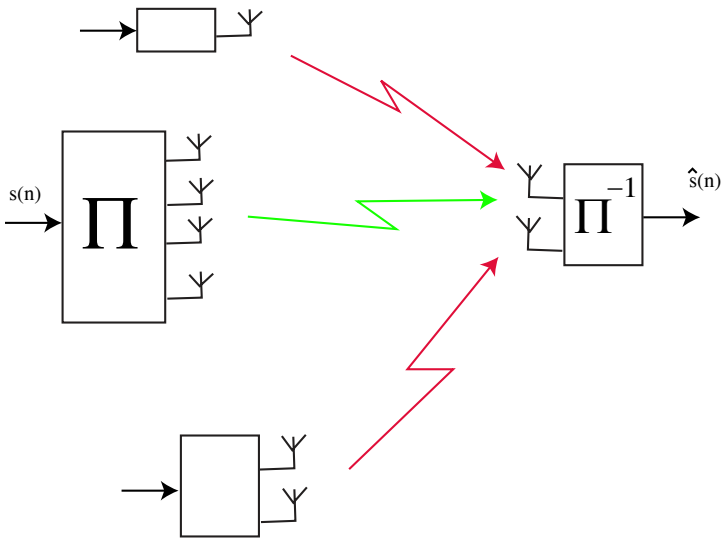


Figure 1.3: A 4×2 MIMO system with $N_I = 2$ co-channel interferers.

shift when propagating through the channel. In the time domain, flat fading corresponds to a channel delay spread which is much less than the symbol time, hence the channel induces no inter-symbol interference (ISI). This means that flat fading apply well to environments where there are significant scatterers close to the transmitter and/or the receiver, and no distant reflectors or scatterers are present. What is meant by "close" depend on the symbol time. An indoor channel measured at 1.8 GHz in an office environment at Uppsala University had an estimated coherence bandwidth of 2.8 MHz, see Section 3.4. So this is the upper limit on the transmission bandwidth in this particular location to be able to use a flat fading channel model.

In outdoor channels, the channel delay spread may become too large for the flat fading assumption to hold. Orthogonal Frequency Division Multiplexing (OFDM) is one method to overcome the ISI in these wide-band channels. In OFDM, the data is multiplexed over a large number of sub-carriers that are spaced apart at separate frequencies. This modulation scheme provides orthogonality between sub-carriers which simplifies the detection. Furthermore, the bandwidth of each sub-carrier is smaller than the channel coherence bandwidth so the frequency selective channel has been turned into a set of parallel flat fading channels, which makes channel equalization due to ISI superfluous[4]. OFDM is used in digital broadcasting of audio and television (DAB and DVB), and broadband indoor wireless systems [24].

[4]Channel equalization for the sub-carriers due to channel phase shifts and attenuation must still be performed.

For multiple antenna systems, OFDM is attractive since multiple received signals requires a multidimensional equalizer if the channel is frequency selective. This is very computationally demanding, especially if the number of receive antennas is large. The flat fading subchannels provided by OFDM simplifies the multi-antenna equalization dramatically [25]. OFDM is also a viable choice in proposals for 4G technologies, since it allows the transmitter to be adaptive in both power, frequency and modulation space [26].

The above discussion motivates the use of a flat fading channel model which will be used in the rest of this thesis. Hence, (1.17) can be simplified to

$$\boldsymbol{\mathcal{H}}(q^{-1}) = \mathbf{H}_0 \stackrel{\triangle}{=} \mathbf{H} \tag{1.19}$$

and by assuming that all CCI channels also can be described by a flat fading model the corresponding input-output relation can be written as

$$\mathbf{y}(n) = \mathbf{H}\mathbf{c}(n) + \sum_{i=1}^{N_I} \mathbf{H}_i \mathbf{c}_i(n) + \mathbf{n}_T(n) \stackrel{\triangle}{=} \mathbf{H}\mathbf{c}(n) + \mathbf{v}(n) \tag{1.20}$$

where the noise plus interference has been lumped together into a single vector $\mathbf{v}(n)$.

Array with plane wave input

If a plane wave is impinging on the array antenna (considering a SIMO system), then the channel vector (matrix) $\mathbf{h}$ has a particular structure as follows. Assume that a plane wave is impinging in the horizontal plane from direction θ and define the complex-valued vector function $\mathbf{h} = \mathbf{a}(\theta)$ as the *array response vector* or array manifold. This vector incorporates all the spatial characteristics of the array and is the response to a unit power signal from the direction θ. Sometimes, the number of plane waves, or paths is large, due to multipath reflections. Then the channel vector $\mathbf{h}$ can be modelled as

$$\mathbf{h} = \sum_{k=1}^{L_m} \alpha_k \mathbf{a}(\theta_k) \tag{1.21}$$

where L_m is the number of multipaths sharing the same time delay, i.e. the response due to a point reflector, α_k are the complex amplitude of path k and θ_k is the direction to reflector k.

A block fading model

The block fading model applies to systems where several adjacent symbols (*a block*) are subject to the same fading value [27]. This could be due to a slow fading

channel where the channel coherence time (T_{coh}) is much larger than the symbol time (T_s) so the channel is assumed to be constant over N symbol intervals.

Assume that block k has a length of N symbol intervals, and contains n_s transmitted symbols. Each symbol belongs to a finite complex constellation $\mathcal{X}$ with $M_d \triangleq |\mathcal{X}|$ symbols. Denote the n:th symbol in the k:th block by $s^{(k)}(n)$ and defined the set $\mathcal{S}$ of transmitted symbols in block k as $\left\{s^{(k)}(n)\right\}_{n=1}^{n_s}$. Next, form the $n_t \times N$ matrix $\mathbf{C}^{(k)}$ containing the n_s transmitted symbols $\left\{s^{(k)}(n)\right\}_{n=1}^{n_s}$ in block k. Furthermore, coding is assumed to be performed over M blocks, see Figure 1.4, so each group of M codeword contains Mn_s symbols, denoted a super-codeword. The blocks can be separated in time, as in a narrowband TDMA systems or in frequency [27]. An example is OFDM, where each of the M sub-carriers are assumed to be flat fading and independent. M can also be interpreted as the interleaving delay, so for strictly delay limited systems, M is small[5]. The input-output relation in

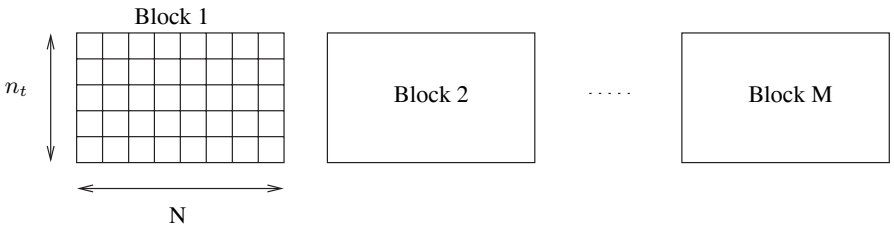


Figure 1.4: One super-codeword in a block-fading model consisting of a group of M codewords.

the block fading model can be written as

$$\mathbf{Y}^{(k)} = \mathbf{H}^{(k)}\mathbf{C}^{(k)} + \mathbf{V}^{(k)} \qquad k = 1, \ldots, M \tag{1.22}$$

where $\mathbf{Y}^{(k)}$ is the block received data matrix and $\mathbf{V}^{(k)}$ is the receiver noise plus interference matrix, both with dimension $n_r \times N$. The matrix $\mathbf{C}^{(k)}$ is sometimes called a (space-time) *code-word* and by proper design, diversity gain, multiplexing gain and/or coding gain can be achieved.

A set of n_s symbols, belonging to $\mathcal{S}$, uniquely defines a code-word matrix $\mathbf{C}^{(k)}$. In many coding schemes the code-word matrix is restricted to be a linear combination of the real and imaginary part of the input symbols $\left\{s^{(k)}(n)\right\}_{n=1}^{n_s}$. Then a code-word matrix can be written as

$$\mathbf{C}^{(k)} = \sqrt{\frac{P_T}{2n_t}} \sum_{n=1}^{n_s} \left[\mathbf{A}_n \bar{s}^{(k)}(n) + j\mathbf{B}_n \tilde{s}^{(k)}(n)\right] \tag{1.23}$$

[5] An example is the half rate GSM standard where $M = 4$ but $N \approx 100$ due to the slow fading. The mapping over $M = 4$ blocks is due to interleaving and channel coding.

where $s^{(k)}(n) = \bar{s}^{(k)}(n) + j\tilde{s}^{(k)}(n)$ and $\mathbf{A}_n, \mathbf{B}_n$ are n_s "elementary" matrices of dimension $n_t \times N$ that completely specifies the code. The notation $\bar{(\cdot)}$ and $\tilde{(\cdot)}$ denotes real and imaginary part respectively. Furthermore, P_T is the total transmitted power from the transmitting array antenna. For an example on how the space-time linear code-word matrix is used, see the expressions (2.68)-(2.70) in Section 2.3.2.

A codeword that can be expressed as (1.23) is called a *linear space-time code word* with rate $R = (n_s/N)\log_2(M_d)$. Note that $r_d \stackrel{\triangle}{=} \log_2(M_d)$ is the number of bits per symbol. The linear model (1.23) incorporates for instance the delay diversity [28], the Alamouti scheme [13], space time block codes [29] and the V-BLAST spatial multiplexing scheme [30] as special cases as well as beamforming and low complexity beamforming schemes such as switched beamforming. The linear coding techniques mentioned above all belong to the family of linear dispersion (LD) codes recently proposed by Hassibi and Hochwald [31]. LD codes have the form (1.23) and are designed to optimize the mutual information between the transmit and receive antennas, at the cost of increased complexity for the transmitter and the receiver.

Channel State Information

An important factor when deriving transmission strategies for MIMO systems is the availability of channel knowledge at the transmitter[6]. The transmitter is said to have *full* Channel State Information (CSI) if the instantaneous channel $\mathbf{H}$ is known at the transmitter. Full or partial CSI can be obtained by either a feedback channel or in a TDD system with time duplex distance shorter than the channel coherence time by direct applying the CSI from the receive channel estimate[7]. A feedback channel is already implemented in some systems for a fast power control[8] which provides the transmitter with partial CSI. In the W-CDMA standard the terms *open loop* and *closed loop* transmission is used for the cases without and with (partial) CSI at the transmitter respectively.

The feedback data of the channel matrix is often quantized. An interesting problem is then how to quantize the feedback data in the most efficient way, which is a problem related to vector quantization. This was further explored in [32] where the term transmitter *side information* was used instead of partial CSI at the transmitter. The feedback information can also be constituted of the long term behavior

[6] In this thesis, it is always assumed that CSI is known at the receiver through channel estimation.

[7] Also the time from the CSI becomes available at the transmit side to when it is used for the transmission must be shorter than the channel coherence time for the full CSI to be valid at the transmitter, to ensure causality.

[8] In the 3GPP Release 99 standard for W-CDMA, 1 bit per frame is allocated for feedback of channel information. This implies a feedback signalling rate of 1500 bits per second.

of the channel such as the channel covariance matrix [19, 33, 34]. Hence the transmitter can transmit in a way that on average is superior to a method that has no CSI at all. Another example of partial CSI is the trace of the matrix $\mathbf{H}\mathbf{H}^*$ [35]. Since this quantity is a sufficient statistic for the probability of error for orthogonal space-time block codes, it can be used to minimize the bit error rate for the connection.

The issue of what and how to feed back partial CSI is still an open problem and one approach is discussed in Section 3.7 of this thesis. An adaptive approach is likely to be useful here since the channel can vary between being static for a long time and be subject to rapid changes in the channel coefficients.

In this context it should be noted that by using differential modulation, the need for CSI at the receiver can be avoided as well. This was explored in e.g. [36] for space-time block codes and it was shown that the differential detector has a 3 dB penalty in signal to noise ratio compared to the coherent detector. Differential modulation will not be covered by this thesis.

Diversity advantage and coding advantage

Finally, two important measures are defined for transmission over a MIMO channel. These are the *diversity advantage*, p, and *coding advantage*, A_c. In [37] they were defined from the pairwise codeword error probability PEP (see Section 2.1) by finding an upper bound of the PEP.

Definition 1.1 *If the pairwise error probability (PEP), which is the probability that a transmitted codeword matrix $\mathbf{C}_1$ is detected in the receiver as another valid codeword matrix $\mathbf{C}_2$, as a function of the receiver signal to noise ratio (SNR) is upper bounded as*

$$PEP \leq A_c \cdot SNR^{-p} \tag{1.24}$$

in the region SNR$\gg 1$, then the system is said to have diversity advantage p and coding advantage A_c.

A system with a diversity advantage $p > 1$ is said to have *diversity gain*, since for a given PEP, a lower SNR is required if compared to a system with diversity advantage $p = 1$. The coding advantage shifts the upper bound of the PEP up or down and is the approximate measure of the gain over an uncoded system operating with the same diversity advantage. Later it will be shown that orthogonal space time block codes as well as beamforming achieves the maximum diversity advantage $p = n_r n_t$. For a SISO channel with binary modulation, the PEP is the same as the bit error rate (BER) of the transmission.

1.2 Outline and contributions of this thesis

The main contributions presented in this thesis are now briefly presented. Then, a summary of the publications follows.

One aim of this thesis is to find methods for the transmission over a channel with multiple antenna elements at either or at both sides. Hence, a novel transmission strategy, applicable when when CSI is available at the transmitter, is presented, called eigenmode transmission. It utilizes adaptive modulation to maximize the data throughput or to minimize the BER. The performance is evaluated theoretically, by assuming a Rayleigh fading channel, and by using the marginal pdf:s of the eigenvalues to the matrix $\mathbf{H}\mathbf{H}^*$, where $\mathbf{H}$ is the MIMO channel. These marginal pdf:s are derived in this thesis, to our knowledge, for the first time. These pdf:s are also used to analyze the performance of the traditional beamforming method over a Rayleigh fading channel. Some of the important conclusions of this analysis is (assuming that the transmitter has full CSI):

- Beamforming is capacity and SNR optimal in a MISO channel.
- Beamforming minimizes the BER in a MIMO channel if no constraints are put on data throughput rate.
- Eigenmode transmission minimizes BER in a MIMO channel under a data throughput constraint by assigning more power to "weak" subchannels, if the subchannels have the same bit rate. Hence, the BER of the subchannels is balanced.
- Eigenmode transmission maximizes data throughput in a MIMO channel under a target BER constraint by assigning more power to "strong" subchannels. For high SNR, the power distribution between subchannels becomes becomes approximately equal, which supports an constant-power variable-rate algorithm.
- Beamforming performance is enhanced in LOS channels.

An analysis of STBC is performed, assuming a correlated Nakagami-m fading MIMO channel with mutual coupling between antenna elements. It is shown, somewhat contrary to intuition, that in certain scenarios, the mutual coupling improves the performance of the STBC.

Furthermore, an investigation of the hardware limitations on the performance of an adaptive antenna array is also made in this thesis. The motivation for this investigation is that it is useful for engineers who design adaptive antenna arrays to know how different design choices affects the final performance of the system. The

design choices can consist of the number of bits used for the ADC or the accuracy of the phase and amplitude of the array antenna weights. Also effects from non-linearities in the transmit amplifiers for array antennas are analyzed and a novel method for reducing the intermodulation distortion is proposed.

An adaptive antenna testbed which was one of the first of its kind, designed specifically for a digital wireless communication system is used to validate some of the derived theories. Previously, adaptive array technology was mainly used by the military in radar applications and most of the work on hardware limitations was made with the military application in mind. We apply this theory to wireless communication systems. Also, two new algorithms for array calibration is derived.

In the last chapter, a switched parasitic antenna (SPA) is presented, which shows very promising results for direction finding (DF) applications. The SPA has only a single receiver/transmitter but can generate a set of radiation patterns with different directional properties. We investigate the performance of the SPA in a MIMO communication scenario and in a DF application.

Next, an overview of the thesis is given and the contributions are summarized, chapter by chapter.

1.2.1 MIMO systems

The first part of this thesis aims at analyzing and deriving transmission strategies for MIMO systems over flat fading channels. Both statistical models and measured channel models are used.

Chapter 2

In this chapter, the MIMO system is introduced and the maximum likelihood detector is derived for the flat fading channel. Performance is compared between MIMO systems with and without CSI at the transmitter. For the case of full CSI at the transmitter, an optimal transmission strategy is derived based on the singular value decomposition of the channel. This leads to an interpretation of the MIMO system as a set of independent subchannels.

For the case of a target BER, the bit rate is maximized using a Lagrange multiplier optimization method and similarly, for the case of a target bit rate, the BER is minimized. It turns out that maximization of the bit rate is achieved by allocating more transmit power to the best subchannels through a "water filling" approach. On the other hand, minimizing BER is obtained by balancing the subchannel BER:s, giving most power to the subchannel with the poorest quality. Other aspects such as power constraints and symbol constellation restrictions are also discussed.

Finally, beamforming is discussed as a transmission strategy, which can be used

when only partial CSI is available at the transmitter, or to reduce implementation complexity.

Chapter 3

In Chapter 3, the performance of MIMO systems is investigated for Nakagami-m and Rayleigh fading channels and for measured MIMO channels. It is observed that if the amplitude of the elements of the channel matrix $\mathbf{H}$ has a Rayleigh distribution, then the matrix $\mathbf{H}\mathbf{H}^*$ has a Wishart distribution. This result is utilized in the analysis. The measured MIMO channels and the comparisons between communication with and without CSI at the transmitter have in parts been presented in

> Mattias Wennström, Mats Helin, Anders Rydberg and Tommy Öberg, "On the Optimality and Performance of Transmit and Receive Space Diversity in MIMO Channels" *IEE Technical Seminar on MIMO Communication Systems: From Concept to Implementation.*, London, December 12, 2001.

Furthermore, space time block codes over flat Nakagami-m fading channels are examined. A general model is assumed that includes an arbitrary correlation between the signals at the receiver antennas which can consider both correlation due to multipath and due to mutual coupling. The result is submitted to

> Mattias Wennström, Tommy Öberg, and Anders Rydberg, "Performance Analysis of Space-Time Block Codes in Correlated Nakagami Fading Channels", Submitted to *IEEE Transactions on Wireless Communications*.

If partial CSI is available at the transmitter, due to a quantized feedback channel, then it can be used to improve the space time block code. It is further shown that knowledge of the receive covariance matrix of the interferers can be used to further improve the link performance. The transmitter can then avoid to transmit in the subspace spanned by the interfering signals. This was presented in parts at the conference

> Mattias Wennström and Tommy Öberg, "Transmit Antenna Diversity in Ricean fading MIMO channels with co-channel interference",*Nordic Radio Symposium 2001*, Nynäshamn.

1.2.2 Measurements and analysis of a SIMO system

The second part of the thesis emanates from the adaptive antenna testbed for the GSM-1800 system. Measurement results are presented and analyzed and suggestions for improved calibration algorithms are derived and compared.

Chapter 4

Chapter 4 handles results from measurements on an adaptive antenna testbed developed at Uppsala University in co-operation with Ericsson Radio Access AB. The testbed is briefly described and the interference suppression capability in outdoor field trials is determined to be more than 30 dB. The results were published in

> Jonas Strandell, Mattias Wennström, Anders Rydberg, Tommy Öberg, Olle Gladh, Leonard Rexberg, Erik Sandberg, Bengt-Victor Andersson and Mikael Appelgren, "Experimental Evaluation of an Adaptive Antenna for a TDMA Mobile Telephony System" *IEEE International Symposium on Personal Indoor and Mobile Radio Communications 1997 (PIMRC)*, Helsinki, September 1-4, 1997, pp.79-84.

and in the conference

> Jonas Strandell, Mattias Wennström, Anders Rydberg, Tommy Öberg, Olle Gladh, Leonard Rexberg and Erik Sandberg, "Design and Evaluation of a Fully Adaptive Antenna for Telecommunication Systems" *Antenn 97*, Gothenburg, Sweden, May 27-29, 1997, pp. 357-365.

Chapter 5

In Chapter 5, the hardware imperfections are modelled and bottlenecks in the performance of adaptive antenna arrays are identified. Design rules are presented on how to balance phase and amplitude accuracy, if hardware beamforming weights are used, for optimal performance. The results were published in

> Mattias Wennström, Tommy Öberg, and Anders Rydberg, "Effects of finite weight resolution and calibration errors on the performance of adaptive array antennas." *IEEE Transactions on Aerospace and Electronic Systems*, vol.37, no.2, April 2001, pp.549-562

and as a conference paper version in

> Mattias Wennström, Tommy Öberg, and Anders Rydberg, "Analysis of Quantisation Effects in Adaptive Array Antennas" *RVK'99, Radiovetenskap och Kommunikation*, Karlskrona, Sweden, June 14-17 1999, pp.451-455.

The analysis presented in these two publications showed that an accurate calibration of the antenna array is crucial for high performance. It is desirable to calibrate the adaptive antenna continuously, so two successful auto-calibration algorithms were derived and then published in

> Mattias Wennström, Tommy Öberg, Jonas Strandell, Anders Rydberg and Erik Lindskog, "Auto-Calibrating Adaptive Array for Mobile Telecommunications" *IEEE Transactions on Aerospace and Electronic Systems*, vol.36, no.2, April 2000, pp.729-736.

1.2.3 Non-linear transmit amplifiers in MISO systems

The third part of the thesis handles non-ideally linearized multi-carrier power amplifiers in MISO systems. The novelty is the analysis of the combination of transmit non-linearities and array antennas, which previously had only been briefly mentioned in the literature.

Chapter 6

In Chapter 6, the effect of non-linear power amplifiers in the transmit chain of array antenna systems are presented. An analysis and a novel method to perform channel allocation to reduce intermodulation distortion is presented. It utilizes the spatial dispersion of the intermodulation interference and is capable of significantly reduce the intermodulation distortion received by the mobile. The results will be published in

> Mattias Wennström, Tommy Öberg, and Anders Rydberg, "Effects of Nonlinear Distortion on Switched Multibeam FDMA Systems" *IEEE Transactions on Antennas and Propagation* , Accepted for publication, tentative issue: vol. 50, February 2003.

Furthermore, the case with intermodulation distortion in SISO systems was analyzed and for the first time, a closed form expression was derived for the outage probability in a log-normal fading channel. This result was published in

> Mattias Wennström, Tommy Öberg, and Anders Rydberg, "Analysis of Intermodulation Distortion on Log-Normal Shadowed WLAN Channels.", *IEE Electronic Letters* , vol.36, no.9, 2000, p.833-834.

Related contributions, although not included in this thesis are publications related to the analysis of the effects of antenna arrays in combination with non-linearities on a system level performance. These results were presented in the following conferences

Mattias Wennström, Anders Rydberg and Tommy Öberg "Intermodulation Distortion in Switched Multibeam Antennas for Cellular Radio Systems ", *IEEE International Symposium on Personal, Indoor and Mobile Radio Communication (PIMRC'2000).*, London, September 18-21, 2000, pp.1495-1499.

and in the conference

Mattias Wennström, "Considering Downlink Intermodulation Distortion in Switched Multibeam Antennas for Cellular Radio Systems" *IEEE Vehicular Technology Conference (VTC Fall)*, Boston, USA, September 24-28, 2000, pp.1858-1865.

and finally also at the conference

Mattias Wennström, Anders Rydberg and Tommy Öberg "Effects of Nonlinear Transmit Amplifiers in Smart Antennas in Wireless Systems" *European Wireless '99*, Münich, Germany, October 4-9, 1999, pp.119-124.

1.2.4 A novel antenna concept - The switched parasitic antenna

The last part of this thesis handles the switched parasitic antenna and an analysis of its performance in a MIMO context and in a direction finding application.

Chapter 7

In Chapter 7, a switched parasitic antenna is presented and analyzed. By using a single antenna element connected to a transceiver and surrounding passive antenna elements, which can be shorted to ground using pin-diodes, directional antenna patterns are obtained. It is shown that the correlation between patterns are sufficiently low to obtain diversity gain. The SPA uses only one transceiver and is attractive due to the low cost compared to the antenna array solution which uses one transceiver per antenna. It is shown that the capacity limit is comparable with the antenna array and by using a space time block code, the array and the SPA is compared in terms of bit error rates. The SPA requires a slightly higher SNR to achieve the same BER as the antenna array. The results was published in

Mattias Wennström and Thomas Svantesson, "An Antenna Solution for MIMO Channels: The Switched Parasitic Antenna ", *IEEE Symposium on Personal Indoor and Mobile Radio Communication (PIMRC) 2001*,pp.159-163, San Diego, USA, September 2001.

By exploiting the directional properties of the SPA, it can be used in direction finding (DF) applications. The DF performance is examined by calculating the lower bound on the DF accuracy, the Cramer-Rao Bound. It is shown that the SPA offers a low cost alternative to traditional array solutions of DF problems. The results of this analysis are presented in

> Thomas Svantesson and Mattias Wennström, "High Resolution Direction Finding using a Switched Parasitic Antenna", *IEEE Statistical Signal Processing Workshop 2001*, pp.508-511, Singapore, August 6-8, 2001.

1.3 Topics for future research

In this thesis, several derived communication methods required full CSI at the transmitter. This assumption holds in TDD systems, where the CSI is directly known from the estimated uplink channel. It is certainly an interesting problem to see how much can be gained in FDD systems, where only partial or delayed CSI is available at the transmitter due to a feedback channel with low bit rate. In [38], the open-loop and closed-loop performance were compared for a MISO system in a Rayleigh fading channel with feedback delay. It was shown that if the Doppler frequency of the channel is above a threshold, then an open-loop scheme is superior to a closed-loop (beamforming) scheme. With the results on marginal pdf:s of the eigenvalues to the matrix $\mathbf{H}\mathbf{H}^*$, derived in this thesis, these results could be extended to MIMO systems.

A related research topic is to investigate the robustness of the derived eigenmode transmission strategies to channel estimation errors at the receiver and also to delayed and quantized CSI at the transmitter. The eigenmode transmission strategies could also be combined with space time codes to improve the equivalent channel gain, especially for the weakest subchannels. For example, in a 4×4 MIMO system, two subchannels can be used for STBC and two can be used for spatial multiplexing transmission. Hence, a total rate of 3 symbols are transmitted over the MIMO channel. In a time-variant channel, the division ratio of the number of subchannels allocated to spatial multiplexing and to STBC, can be made adaptive, to always maximize the performance.

Which transmission strategy to use over a MIMO channel, depends on many varying factors, the CSI availability, the channel characteristics such as coherence time, LOS/NLOS, orientation of the array etc. The results in this thesis can be used as a starting point when further exploring this.

An identified problem with the average power constraints in the eigenmode transmission algorithms, is the large peak-to-mean ratios of the transmitted signal

power. Hence, to investigate the effect on the system performance, utilizing the developed theory for nonlinear amplifiers and array antennas in Chapter 6 is an interesting topic. To our knowledge, the effect of nonlinearities on MIMO systems has not yet been analyzed.

The pdf:s for the eigenvalues in the uncorrelated Rayleigh fading MIMO channel, derived in this thesis, has proven to be useful when analyzing the performance of different applications. An interesting research topic is to try to find the corresponding pdf:s for the Rice or Nakagami-m fading cases. With these at hand, it would then be possible to compare e.g. beamforming performance for different values of the Rice K-factor. We have tried to derive this for a Ricean MIMO channel, but not completely succeeded. If the elements of $\mathbf{H}$ have Ricean distributed amplitudes, then the matrix $\mathbf{H}\mathbf{H}^*$ has a non-central complex Wishart distribution [39]. The joint pdf of the eigenvalues to $\mathbf{H}\mathbf{H}^*$ contains hypergeometric functions with matrix arguments. These can be calculated through a series expansion in zonal polynomials [39], but the mathematics involved gets too intricate to pursue this any further. However, for 2×2 and 3×3 systems, there exists closed form expressions for the zonal polynomials and there might be more results from the mathematics of complex multivariate statistics that can be used on this problem.

In indoor environments, the channel can become static for quite some time, as observed in the MIMO channel measurements presented in this thesis. If a receiver or transmitter, say a laptop computer, has a single antenna, it might be located in a fading dip. Since the channel coherence time is so long, the user will experience a low bandwidth, if any, for a long time, which is undesirable. An expensive solution to this problem is to provide the computer with an additional diversity antenna. A cheaper solution could be to equip the laptop computer with parasitic antenna elements, perhaps parasitic patch antennas. Now, if the receiver/transmitter antenna is in a deep fading dip, the parasitic antenna pin diodes are switched until a setting is found where the fading dip is reduced. Switching parasitic antenna elements changes the radiation pattern of the antenna and thus, angle diversity is obtained. How well this idea work in indoor and outdoor environments is an interesting topic for future research. More work on how to optimally place the parasitic elements for maximum diversity gain is also needed.

Chapter 2

Multiantenna systems with general flat fading channels

IN this chapter the multiple input multiple output wireless communication system is discussed in general terms without making any prior assumptions on the channel except the basic assumptions of flat and quasi-stationary fading (the channel is constant during the transmission of N symbols). In Chapter 3, the analysis is extended by assuming more specific channels such as physical channel models, stochastic channel models, and by using measured channel data.

The chapter is organized as follows: The maximum likelihood receiver is derived in Section 2.1 and it will be used in the subsequent chapters. The fundamental limits in terms of mutual information and Shannon capacity are analyzed in Section 2.2. Some practical space and time coding methods are then presented in Section 2.3, for the case of no CSI at the transmitter, using results from the derivation of the ML-receiver in the preceding section. In Section 2.4, CSI is assumed to be known at the transmitter and a derivation of a novel transmit algorithm is presented that either maximizes the spectral efficiency at a given target BER or minimizes the average BER at a target spectral efficiency. Then beamforming is discussed as a special case of the derived method in Section 2.5. Finally, in Section 2.6 comparisons are made and the chapter is concluded.

2.1 The Maximum Likelihood detector for linear space time codes

The maximum likelihood (ML) detector for linear space-time codes used over a flat fading MIMO channel with spatially and temporally colored Gaussian noise is now derived. The derivation gives insight to the problems associated with spatial multiplexing transmission and the resulting detector is used in other sections of this thesis. Furthermore, the *pairwise error probability* (PEP) for the linear space time code is also studied. The PEP characterizes the performance of a system with coding over a finite number of blocks, M, and captures the diversity advantage of a code. In Section 3.7, it is shown how the PEP of an orthogonal space-time block code can be improved by the use of CSI.

Consider the channel model (1.22), with $M = 1$ so the block indices k on the matrices are dropped for notational convenience,

$$\mathbf{Y} = \mathbf{HC} + \mathbf{V} \ . \tag{2.1}$$

The block index is implicit throughout this chapter, if not stated otherwise. This means that only a single space-time codeword can be examined at a time and that the channel $\mathbf{H}$ is constant during the transmission of the valid space-time codeword $\mathbf{C}$. The elements of the interferer plus noise matrix $\mathbf{V}$ are dependent, i.e. the noise is correlated in both space and time. To make it possible to analyze this scenario, we introduce the following vectors and a matrix as follows

$$\mathcal{C} \triangleq \text{vec}(\mathbf{C}) \tag{2.2}$$
$$\mathcal{Y} \triangleq \text{vec}(\mathbf{Y}) \tag{2.3}$$
$$\mathcal{V} \triangleq \text{vec}(\mathbf{V}) \tag{2.4}$$
$$\mathcal{H} \triangleq \mathbf{I}_N \otimes \mathbf{H} \tag{2.5}$$

where $\otimes$ is the Kronecker product and $\mathbf{I}_N$ is the $N \times N$ identity matrix. With these definitions, (2.1) can be expressed as

$$\mathcal{Y} = \mathcal{HC} + \mathcal{V} \tag{2.6}$$

where $\mathcal{Y}, \mathcal{V}$ are $n_r N \times 1$ vectors, $\mathcal{C}$ is an $n_t N \times 1$ vector and the channel matrix is now described by an $n_r N \times n_t N$ block diagonal matrix. Form the covariance matrix of the noise vector as

$$\mathcal{R}_{\mathcal{VV}} = E\left\{\mathcal{VV}^*\right\} \tag{2.7}$$

where the * denotes complex conjugate transpose (Hermitian transpose). A special case occur, if the noise plus interference is temporally white. Then

$$\mathcal{R}_{\mathcal{VV}} = \mathbf{I}_N \otimes \mathbf{R}_{vv} \tag{2.8}$$

where $\mathbf{R}_{vv} \triangleq E\{\mathbf{v}(n)\mathbf{v}^*(n)\}$ is the positive definite unstructured spatial covariance matrix of the interferers plus noise. The temporally white assumption is customary, but clearly an approximation since the interference may consist of modulated signals that are not perfectly white.

The derivation of the ML detector begins by first forming a hypothesis test. Assume that the transmitter chooses between the codewords $\mathcal{C}_0, \ldots, \mathcal{C}_{P-1}$ where each codeword $\mathcal{C}_i$ is uniquely defined by its sequence $\{s_i(n)\}_{n=1}^{n_s} \in \mathcal{S}$ and where each symbol belongs to the used complex symbol constellation $\mathcal{X}$, $s_i(n) \in \mathcal{X}$. Note that a new index i has been introduced to separate different valid codeword matrices. Mathematically, we have the following multiple hypothesis problem at the receiver side:

$$\begin{aligned} H_0 : \mathcal{Y} =& \mathcal{H}\mathcal{C}_0 + \mathcal{V} \\ H_1 : \mathcal{Y} =& \mathcal{H}\mathcal{C}_1 + \mathcal{V} \\ & \vdots \\ H_{P-1} : \mathcal{Y} =& \mathcal{H}\mathcal{C}_{P-1} + \mathcal{V} \end{aligned} \tag{2.9}$$

where the codewords $\mathcal{C}_i$ are assumed to have equal a priori probabilities. Now, assume that the elements of the noise plus interference vector $\mathcal{V}$ are zero mean and have a circular complex Gaussian distribution with covariance matrix $\mathcal{R}_{\mathcal{V}\mathcal{V}}$. The optimal receiver is then a minimum distance receiver [14, 40]. It is found "after some calculations" that hypothesis H_i is chosen if the test statistic

$$\mathcal{T}_i(\mathcal{Y}) = -\frac{1}{2}\left(\mathcal{Y} - \mathcal{H}\mathcal{C}_i\right)^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\left(\mathcal{Y} - \mathcal{H}\mathcal{C}_i\right) + \frac{1}{2}\mathcal{Y}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{Y} \tag{2.10}$$

is the maximum statistic of $(\mathcal{T}_0(\mathcal{Y}), \ldots, \mathcal{T}_{P-1}(\mathcal{Y}))$. This is the maximum likelihood (ML) rule as we select the hypothesis with the largest conditional probability. Hence, the hypothesis whose signal vector after the channel is closest to the received vector $\mathcal{Y}$ in the Euclidian sense is chosen.

Equation (2.10) can be expanded as

$$\begin{aligned} \mathcal{T}_i(\mathcal{Y}) =& -\frac{1}{2}\mathcal{Y}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{Y} + \frac{1}{2}\mathcal{Y}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\mathcal{C}_i + \frac{1}{2}\mathcal{C}_i^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{Y} \\ & - \frac{1}{2}\mathcal{C}_i^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\mathcal{C}_i + \frac{1}{2}\mathcal{Y}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{Y} \\ =& \frac{1}{2}\mathcal{C}_i^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{Y} + \left(\frac{1}{2}\mathcal{C}_i^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{Y}\right)^* \\ & - \frac{1}{2}\mathcal{C}_i^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\mathcal{C}_i \end{aligned} \tag{2.11}$$

and by using the relation $z + z^* = 2\text{Re}(z)$ and the fact that $\mathcal{R}_{\mathcal{V}\mathcal{V}}$ is Hermitian, (2.11) can be written compactly as

$$\mathcal{T}_i(\mathcal{Y}) = \text{Re}\left(\mathcal{C}_i^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{Y}\right) - \frac{1}{2}\mathcal{E}_i \tag{2.12}$$

where $\mathcal{E}_i = \mathcal{C}_i^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\mathcal{C}_i$.

Note that since $\mathcal{H}$ and $\mathcal{C}$ are deterministic, the test statistic is a linear combination of the complex Gaussian vector $\mathcal{V}$. Hence, the test statistic $\mathcal{T}_i(\mathcal{Y})$ has also a Gaussian distribution, since a linear combination of Gaussian distributed random variables is also Gaussian distributed.

Pairwise error probability

To determine the codeword error probability in general is difficult. Instead, a common measure for error probability, denoted the pairwise error probability (PEP), is used which is the probability that a transmitted codeword, say $\mathcal{C}_0$, is detected as $\mathcal{C}_1$ in the receiver. In other words, the sequence $\left\{s^{(0)}(n)\right\}_{n=1}^{n_s}$ that defines the space-time codeword $\mathcal{C}_0$ is errornously detected as another sequence $\left\{s^{(1)}(n)\right\}_{n=1}^{n_s}$. The PEP can be used in the codeword design, see Section 3.7, to maximize the Euclidean distance between the received signals under the finite energy constraint by minimizing the PEP between pairs of codewords. The PEP in this binary case is

$$P_{PEP} = \text{Pr}\left\{\mathcal{C}_0 \rightarrow \mathcal{C}_1 | H_0\right\} = \text{Pr}\left\{\mathcal{T}_1(\mathcal{Y}) > \mathcal{T}_0(\mathcal{Y}) | H_0\right\} \ . \tag{2.13}$$

Now, define

$$\mathcal{T}(\mathcal{Y}) = \mathcal{T}_1(\mathcal{Y}) - \mathcal{T}_0(\mathcal{Y}) = \text{Re}\left((\mathcal{C}_1 - \mathcal{C}_0)^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{Y}\right) - \frac{1}{2}(\mathcal{E}_1 - \mathcal{E}_0) \tag{2.14}$$

and since $\mathcal{T}_1(\mathcal{Y})$ and $\mathcal{T}_0(\mathcal{Y})$ are Gaussian distributed so is the difference $\mathcal{T}(\mathcal{Y})$. The conditional mean of $\mathcal{T}(\mathcal{Y})$ under the hypotheses H_0 and H_1 differ by sign

$$\begin{aligned} E\left\{\mathcal{T}(\mathcal{Y})|H_0\right\} &= -\frac{1}{2}\text{Tr}\left(\mathcal{O}\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\right) \\ E\left\{\mathcal{T}(\mathcal{Y})|H_1\right\} &= \frac{1}{2}\text{Tr}\left(\mathcal{O}\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\right) \end{aligned} \tag{2.15}$$

where

$$\mathcal{O} \stackrel{\triangle}{=} (\mathcal{C}_1 - \mathcal{C}_0)(\mathcal{C}_1 - \mathcal{C}_0)^* \ . \tag{2.16}$$

The derivation of (2.15) is shown in Appendix 2.A on page 68 together with derivation of the conditional variance, which becomes

$$\text{Var}\left\{\mathcal{T}(\mathcal{Y})|H_0\right\} = \frac{1}{2}\text{Tr}\left(\mathcal{O}\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\right) = \text{Var}\left\{\mathcal{T}(\mathcal{Y})|H_1\right\} \ . \tag{2.17}$$

Note that it obtains the same value as the expectation value under H_1.

To summarize, the test statistic $\mathcal{T}(\mathcal{Y})$ is Gaussian distributed with mean and variance given by (2.15) and (2.17) respectively. Under hypothesis H_0, the test statistic

$$\frac{\mathcal{T}(\mathcal{Y}|H_0) - E\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\}}{\sqrt{\text{Var}\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\}}} \tag{2.18}$$

is Gaussian distributed with zero mean and variance 1. Hence, the PEP, that is, the probability that the transmitted codeword $\mathcal{C}_0$ is detected as another codeword $\mathcal{C}_1$, can be written as

$$\begin{aligned}
\Pr\left\{\mathcal{C}_0 \to \mathcal{C}_1|H_0\right\} =& \Pr\left\{\mathcal{T}(\mathcal{Y}|H_0) > 0\right\} \\
=& \Pr\left\{ \frac{\mathcal{T}(\mathcal{Y}|H_0) - E\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\}}{\sqrt{\text{Var}\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\}}} > \frac{0 - E\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\}}{\sqrt{\text{Var}\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\}}} \right\} \\
=& Q\left\{ \frac{-E\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\}}{\sqrt{\text{Var}\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\}}} \right\} \\
=& Q\left\{ \sqrt{\frac{1}{2}\text{Tr}\left(\mathcal{O}\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\right)} \right\}
\end{aligned} \tag{2.19}$$

where $Q(x)$ is the Gaussian Q-function defined as [14]

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\lambda^2/2} d\lambda \ . \tag{2.20}$$

The same expression for the PEP is obtained under hypothesis H_1. Often, an upper bound for the Q-function is used, called the Gaussian tail approximation. It is derived from the Chernoff bound [15] and states

$$Q(x) \leq \frac{1}{2} e^{-x^2/2}, \qquad x \geq 0 \ . \tag{2.21}$$

The use of this bound in (2.19) gives

$$\Pr\left\{\mathcal{C}_0 \to \mathcal{C}_1|H_0\right\} \leq \frac{1}{2} \exp\left\{-\text{Tr}\left(\mathcal{O}\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\right)/4\right\} \ . \tag{2.22}$$

Even if orthogonal codeword matrices are chosen, that is $\mathcal{O} = \mathbf{I}$, the distance properties of the transmitted codewords are not preserved (unless $\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1/2}\mathcal{H}$ is unitary) after the channel. Hence, spatial multiplexing, which can be designed to use orthogonal data streams from the n_t transmit antennas, will have poor detection performance if the matrix $\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1/2}\mathcal{H}$ is ill conditioned. An example of this is when

the noise is uncorrelated in time and space, $\mathcal{R}_{\mathcal{VV}}^{-1} = \sigma_n^2\mathbf{I}$ and the channel has a strong line of sight (LOS) component. Then $\mathcal{H}$ become an ill conditioned matrix and spatial multiplexing will show bad performance.

The bound (2.22) can be further explored to calculate an upper bound on the error probability [41], by using the fact that $\text{Tr}\left(\mathcal{O}\mathcal{H}^*\mathcal{R}_{\mathcal{VV}}^{-1}\mathcal{H}\right) = |\mathcal{R}_{\mathcal{VV}}^{-1/2}\mathcal{H}(\mathcal{C}_0 - \mathcal{C}_1)|^2$ and the Rayleigh-Ritz theorem [42] for an arbitrary vector $\mathbf{e}$

$$\frac{\left|\mathcal{R}_{\mathcal{VV}}^{-1/2}\mathcal{H}\mathbf{e}\right|^2}{|\mathbf{e}|^2} \geq \lambda_{min}(\mathcal{H}^*\mathcal{R}_{\mathcal{VV}}^{-1}\mathcal{H}) \tag{2.23}$$

with equality if $\mathbf{e}$ is the minimum right singular vector to $\mathcal{R}_{\mathcal{VV}}^{-1/2}\mathcal{H}$. Using this inequality, the worst case PEP is given by

$$\Pr\{\mathcal{C}_0 \to \mathcal{C}_1|H_0\} \leq \frac{1}{2}\exp\left\{-\lambda_{min}(\mathcal{H}^*\mathcal{R}_{\mathcal{VV}}^{-1}\mathcal{H})|\mathcal{C}_1 - \mathcal{C}_0|^2/4\right\} . \tag{2.24}$$

So, the upper bound on PEP depends on the smallest eigenvalue to the matrix $\mathcal{H}^*\mathcal{R}_{\mathcal{VV}}^{-1}\mathcal{H}$ and if this matrix is rank deficient, the upper bound on PEP is 1/2. The expected value of λ_{min} is increased if the number of antennas on the receiver side is increased so that $n_r > n_t$. Hence more receive antennas give a reduced PEP and a more robust transmission.

In Section 2.3.2, orthogonal space time block codes (STBC) are analyzed and it is shown that STBC performance depends on the average of the eigenvalues instead of the smallest eigenvalue (under the condition that the noise is white; $\mathcal{R}_{\mathcal{VV}} = \sigma^2\mathbf{I}$). This unique property of the STBC makes them robust to channel conditions and therefore attractive although they have a bit rate which is lower than the bit rate for a spatial multiplexing transmission scheme [41].

Temporally white noise

To investigate the temporally white noise case, the relations [43, Sec. 2.2]

$$\text{Tr}(\mathbf{BCD}) = \text{vec}(\mathbf{B}^*)^*(\mathbf{I}\otimes\mathbf{C})\text{vec}(\mathbf{D}) \tag{2.25}$$

and

$$(\mathbf{I}\otimes\mathbf{C})^{-1} = (\mathbf{I}\otimes\mathbf{C}^{-1}) \tag{2.26}$$

will be used. When the noise is temporally white, the noise covariance matrix is given by (2.8), so the matrix product in (2.12) can be simplified using (2.25),(2.26) to

$$\begin{aligned}\mathcal{C}_i^*\mathcal{H}^*\mathcal{R}_{\mathcal{VV}}^{-1}\mathcal{Y} =& \left[\text{vec}\left[(\mathbf{HC}_i)^*\right]^*\right]^*\left[\mathbf{I}\otimes\mathbf{R}_{vv}\right]^{-1}\text{vec}(\mathbf{Y}) \\ =& \text{Tr}((\mathbf{HC}_i)^*\mathbf{R}_{vv}^{-1}\mathbf{Y}) \\ =& \text{Tr}(\mathbf{C}_i^*\mathbf{H}^*\mathbf{R}_{vv}^{-1}\mathbf{Y}) \ .\end{aligned} \tag{2.27}$$

Hence, the test statistic in (2.12) for hypothesis i is

$$\mathcal{T}_i(\mathbf{Y}) = \text{Re}\left[\text{Tr}\left(\mathbf{C}_i^*\mathbf{H}^*\mathbf{R}_{vv}^{-1}\mathbf{Y}\right)\right] - \frac{1}{2}E_i \tag{2.28}$$

where $E_i = \text{Tr}(\mathbf{C}_i^*\mathbf{H}^*\mathbf{R}_{vv}^{-1}\mathbf{H}\mathbf{C}_i)$. The maximum likelihood sequence estimate (MLSE) is

$$\{\hat{s}(n)\}_{n=1}^{n_s} = \arg\max_k \left\{\text{Re}\left[\text{Tr}\left(\mathbf{C}_k^*\mathbf{H}^*\mathbf{R}_{vv}^{-1}\mathbf{Y}\right)\right] - \frac{1}{2}E_i\right\} \tag{2.29}$$

where $\{s_k(n)\}_{n=1}^{n_s} \in \mathcal{S}$ defines a valid, space-time codeword matrix $\mathbf{C}_k$ and each symbol belongs to the set $\mathcal{X}$. The MLSE searches among all possible transmitted sequences (search over k) for the sequence with which the probability conditioned on the channel matrix and noise covariance matrix is maximized. This search can for general space-time codeword matrices $\mathbf{C}_k$ be prohibitively lengthy (search over $|\mathcal{X}|^{n_s}$ symbol combinations) and also ill-conditioned. However, in Section 2.3.2 it will be shown how orthogonal space time block codes imposes a special structure on $\mathbf{C}_k$ that *decouples* the search problem at the receiver into a scalar one (search over $n_s|\mathcal{X}|$ symbol combinations).

It is straightforward to show that the test statistic for the pairwise error probability, $\mathcal{T}(\mathbf{Y})$ becomes

$$\mathcal{T}(\mathbf{Y}) = \text{Re}\left[\text{Tr}\left((\mathbf{C}_1 - \mathbf{C}_0)^*\mathbf{H}^*\mathbf{R}_{vv}^{-1}\mathbf{Y}\right)\right] - \frac{1}{2}(E_1 - E_0) \ . \tag{2.30}$$

Remark 2.1 *Define the matrix*

$$\mathbf{W}_{bf}^* \stackrel{\triangle}{=} \mathbf{H}^*\mathbf{R}_{vv}^{-1} \tag{2.31}$$

in equation (2.28)*. It can be interpreted as the generalized matched spatial filter that operates on the received matrix* $\mathbf{Y}$ *before correlating with the codeword* $\mathbf{C}_i$ *and is sometimes called the beamforming vector or the weight vector for the receiving beamformer since in some cases it can be interpreted as the weighting coefficients that forms a pointing "beam" in the radiation pattern, towards the transmitter. Its significance will be seen in later chapters.*

The corresponding PEP will in the temporally white noise case be

$$\begin{aligned}\Pr\{\mathbf{C}_0 \to \mathbf{C}_1|H_0\} =& Q\left\{\sqrt{\frac{1}{2}\text{Tr}\left((\mathbf{C}_1 - \mathbf{C}_0)(\mathbf{C}_1 - \mathbf{C}_0)^*\mathbf{H}^*\mathbf{R}_{vv}^{-1}\mathbf{H}\right)}\right\} \\ =& Q\left\{\left\|\frac{1}{4}\mathbf{R}_{vv}^{-1/2}\mathbf{H}(\mathbf{C}_1 - \mathbf{C}_0)\right\|_F\right\}\end{aligned} \tag{2.32}$$

where in the second step, the relations $\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA})$ and $\text{Tr}(\mathbf{A}^*\mathbf{A}) = \|\mathbf{A}\|_F^2$ were used.

Temporally and spatially white noise

If the noise is temporally and spatially white, $\mathbf{R}_{vv} = \sigma_n^2\mathbf{I}$, then the detector bases the decisions on the statistics

$$\mathcal{T}_i(\mathbf{Y}) = \mathrm{Re}\left[\mathrm{Tr}\left(\mathbf{C}_i^*\mathbf{H}^*\mathbf{Y}\right)\right]/\sigma_n^2 - \frac{1}{2}E_i \tag{2.33}$$

where E_i is the received SNR when codeword $\mathbf{C}_i$ was transmitted. The PEP becomes

$$\Pr\left\{\mathbf{C}_0 \to \mathbf{C}_1|H_0\right\} = Q\left\{\sqrt{\frac{\|\mathbf{H}\left(\mathbf{C}_1 - \mathbf{C}_0\right)\|_F^2}{2\sigma_n^2}}\right\} \quad . \tag{2.34}$$

To summarize, the ML detectors for transmission over MIMO channels with space time block code-word matrices was derived, assuming a correlated circularly complex Gaussian noise model. As in the scalar case [15], the error probability is expressed using the Q-function, due to the Gaussian assumption. The PEP is a function of the distance between two code-word matrices and is used in the analysis and design of space-time codes [41, 44, 45]. It is a difficult problem to find a set of space-time code-word matrices that have a low average PEP, but if a linear structure is imposed on the code-word matrix, see equation (1.23), it is possible to find a set of matrices with attractive properties. This will further be explored in Section 2.3.2.

2.2 Mutual Information and Channel Capacity

Information-theoretic studies of wireless fading channels have relatively recently accelerated remarkably. The results inspire researchers to find new, interesting and better ways to transmit data over the wireless channel [46]. This renaissance has already lead to interesting results and new coding techniques, such as space time codes and linear dispersion codes.

The multi antenna wireless communication system is now discussed from an information theoretic perspective. This discussion motivates the use of specific transmission techniques, as space time codes, multi mode transmission and single mode transmission (beamforming) in subsequent chapters. The concept of mutual information gives a guideline to how well our design performs and how close the system operates to the ultimate *Shannon limit*.

For the analysis in this section, it is assumed that no co-channel interferers are present and that the noise is spatially and temporally white with variance σ_n^2. The transmitter is limited to a maximum output power of P_T.

Rewrite the general block fading MIMO channel model (1.22) vector by vector as

$$\mathbf{y}(n) = \mathbf{H}\mathbf{c}(n) + \mathbf{v}(n) \qquad n = 1, \ldots, N \ . \tag{2.35}$$

Recall that the block index k has been dropped for notational convenience. The vector $\mathbf{v}(n)$ contains i.i.d. circular symmetric Gaussian noise samples, $E\{\mathbf{v}(r)\mathbf{v}^*(s)\} = \delta_{rs}\sigma_n^2\mathbf{I}_{n_r}$. Define[1]

$$\mathbf{R}_{cc} = E\{\mathbf{c}(n)\mathbf{c}^*(n)\} \tag{2.36}$$

as the covariance matrix of the transmitted signal. Furthermore, we introduce the following average power constraint on the transmitted signal,

$$E\{\mathbf{c}^*(n)\mathbf{c}(n)\} = \mathrm{Tr}\{\mathbf{R}_{cc}\} \leq P_T \tag{2.37}$$

where P_T is the maximum transmitted power and define the signal to noise ratio[2] (SNR) as P_T/σ_n^2.

Given a deterministic flat fading channel matrix $\mathbf{H}$, define the *instantaneous mutual information* as [47, 48]

$$I = \max_{\mathbf{R}_{cc}>0} \log_2 \det\left\{\mathbf{I}_{n_r} + \frac{\mathbf{H}\mathbf{R}_{cc}\mathbf{H}^*}{\sigma_n^2}\right\} \tag{2.38}$$

measured in bit/s/Hz and where $\mathrm{Tr}(\mathbf{R}_{cc}) \leq P_T$. In words, the mutual information is the information of the event that the transmission of the particular signal vector $\mathbf{c}(n)$ has taken place by the reception of the signal vector $\mathbf{y}(n)$ given the model (2.35) [49].

Ergodic and Non-Ergodic Channel Capacity

So far, the channel matrix $\mathbf{H}$ has been treated as a deterministic constant matrix. Due to channel dynamics, this is not true for any longer period of time. The time-variant channel is due to movement of the receiver and transmitter but also due to scattering from moving objects in the channel. We shall in the sequel assume $\mathbf{H}$ to be time-variant. Let $\mathbf{H}$ be a stochastic channel matrix, where the elements of $\mathbf{H}$ are random variables. The quasi-stationary block fading model, defined in Section

[1] Recall that σ_n^2 stands for the noise variance in this thesis. The subscript n in σ_n^2 stands for "noise" and is not an index n.

[2] Note that this definition of the SNR is somewhat unconventional, since the transmit power is referred to the transmitter side and the noise power is referred to the receiver side. Hence, this is the SNR in a channel with channel gain normalized to one so it can be interpreted as the SNR at the receiver. This is however a common definition in the literature when dealing with information theory and will also be adopted here.

1.1.2 is assumed, hence, the channel matrix $\mathbf{H}$ randomly generated but constant during the transmission of one space-time code-word of length N. A new random channel matrix, independent of the previous ones, is then assumed generated for each new space-time code-word.

Due to the stochastic nature of the channel matrix $\mathbf{H}$, it becomes necessary to distinguish between ergodic and non-ergodic channel capacity. In the ergodic case, there are no time delay constraints on the transmission and we may allow the number of symbol blocks, M, which the coding is performed over to increase without bound. Specifically, the time it takes to transmit the M blocks is much larger than the coherence time of the channel. The ergodic capacity is equivalent with the capacity formula derived by Shannon.

Definition 2.1 *Ergodic channel capacity is the expected value of the mutual information,*

$$C = \max_{\mathbf{R}_{cc}>0, Tr(\mathbf{R}_{cc})\leq P_T} E_{\mathbf{H}} \left[\log_2 \det \left\{ \mathbf{I}_m + \frac{\mathbf{H}\mathbf{R}_{cc}\mathbf{H}^*}{\sigma_n^2} \right\} \right], \tag{2.39}$$

over all possible channel realizations, and is measured in bits per second per Hertz.

Hence, the ergodic channel capacity is normalized with the transmission bandwidth. It represents the long term achievable bit rate of the channel, averaged over the fading distribution, for which the error probability can be driven asymptotically to zero. Hence, it is the inherent bit rate limit of the channel, and if perfect CSI is available at the transmitter, it can be achieved by adapting transmission power and bit rate relative to the channel quality [50]. This will be further explored in Section 2.4. The capacity in (2.39) was derived assuming no co-channel interference (CCI). To take CCI into account, the channel matrix, $\mathbf{H}$, is exchanged for $\tilde{\mathbf{H}} = \mathbf{R}_{vv}^{-1/2}\mathbf{H}$, where $\mathbf{R}_{vv}$ is the covariance matrix of the noise plus interference term in (1.20)[51].

The ergodic assumption will not be satisfied in practical communication systems, especially not in delay constrained applications, such as speech transmission which must have a small value of M, where M is the number of blocks the coding is performed over. Hence, the capacity given by Definition 2.1 is too optimistic. In the non-ergodic case, when M is comparable with the channel coherence time, other information theoretic measures must be used, and we make the following definitions:

Definition 2.2 *The outage probability, P_0, is defined as the probability that the mutual information I, in equation* (2.38) *is less than some pre-determined value C_{P_0}. Hence*

$$P_0 = Pr(I < C_{P_0}) \ . \tag{2.40}$$

The value C_{P_0} is the outage capacity corresponding to the outage probability P_0.

Definition 2.3 *Zero outage capacity, C_0, or delay-limited capacity is the outage capacity with the outage probability $P_0 = 0$.*

The outage probability is closely related to the codeword error probability, as averaged over the ensemble of codewords and all channel realizations. For block lengths of around $N = 100$ symbols, the information outage probability predicts the word error probability well for good practical codes [45].

The above definitions are reasonable to use as capacity measures when the transmission is delay limited since, when M and N are small, the coding and eventually interleaving takes place over a channel matrix $\mathbf{H}$ that is just a "snapshot" of the underlying stochastic process. Hence, there is a probability P_0 that this particular channel realization is in such a deep fade that the communication system, operating with C_{P_0} bits/s/Hz will fail to transmit the bits without errors. The zero outage capacity can be interpreted as the lowest transmission bit rate that is invariant of the fading. For a SISO system, this corresponds to channel inversion, which then makes the observed channel independent of the fading. Since a SISO Rayleigh fading channel is not invertible with finite power, the zero outage capacity is zero in this case.

We now proceed by distinguishing between the two cases with and without full CSI available to the transmitter. In both cases, the receiver is assumed to have full CSI.

2.2.1 Mutual information with CSI at transmitter

Assume that in some way, the transmitter has knowledge of the channel matrix $\mathbf{H}$. To investigate some important features of the channel capacity, the singular value decomposition (SVD) of the channel and the transmit covariance matrix is used. This approach has previously been taken in e.g. [47]. Define the SVD of the channel matrix $\mathbf{H}$ as

$$\mathbf{H} = \mathbf{U}\Lambda\mathbf{V}^* \tag{2.41}$$

and for the covariance matrix of the transmit data

$$\mathbf{R}_{cc} = \mathbf{SDS}^* \tag{2.42}$$

where $\mathbf{U}$ and $\mathbf{V}$ are complex unitary matrices of dimensions $n_r \times n_r$ and $n_t \times n_t$ respectively and Λ is an $n_r \times n_t$ matrix containing the

$$m \stackrel{\triangle}{=} \min\{n_r, n_t\} \tag{2.43}$$

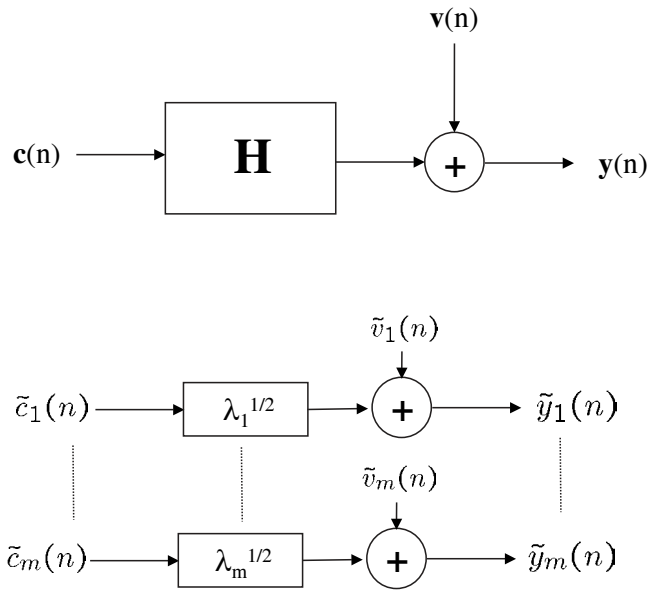


Figure 2.1: Channel model and equivalent channel model by the use of the singular value decomposition of the channel matrix $\mathbf{H}$.

real non-negative singular values denoted as $\lambda_1^{1/2}, \ldots, \lambda_m^{1/2}$ on its main diagonal. With this definition, $\lambda_i, i = 1, \ldots, m$ are the eigenvalues to the matrix $\mathbf{HH}^*$. Similarly, $\mathbf{S}$ is an $n_t \times n_t$ square matrix whereas $\mathbf{D}$ is a diagonal matrix containing the singular values of $\mathbf{R}_{cc}$, here denoted $\gamma_1, \ldots, \gamma_{n_t}$ respectively. By defining

$$\tilde{\mathbf{y}}(n) \triangleq \mathbf{U}^*\mathbf{y}(n) \tag{2.44}$$

$$\tilde{\mathbf{v}}(n) \triangleq \mathbf{U}^*\mathbf{v}(n) \tag{2.45}$$

$$\tilde{\mathbf{c}}(n) \triangleq \mathbf{V}^*\mathbf{c}(n) \tag{2.46}$$

the channel model (2.35) can be reformulated as

$$\tilde{\mathbf{y}}(n) = \Lambda\tilde{\mathbf{c}}(n) + \tilde{\mathbf{v}}(n) \tag{2.47}$$

where the statistics $E\{\tilde{\mathbf{v}}(r)\tilde{\mathbf{v}}(s)^*\} = \delta_{rs}\sigma_n^2\mathbf{I}_{n_r}$ and $\mathbf{R}_{\tilde{c}\tilde{c}} = \mathbf{R}_{cc}$ are preserved. Hence (2.35) and (2.47) are equivalent. Since Λ is diagonal with m nonzero diagonal elements, we have effectively m parallel and independent transmission channels. See Figure 2.1.

Using (2.41) and (2.42) in (2.38), the mutual information can be expressed as

[47]

$$I = \max_{\mathbf{S},\mathbf{D}} \log_2 \det \left\{ \mathbf{I}_{n_r} + \frac{\mathbf{U}\Lambda\mathbf{V}^*\mathbf{S}\mathbf{D}\mathbf{S}^*\mathbf{V}\Lambda^*\mathbf{U}^*}{\sigma_n^2} \right\} \tag{2.48}$$

under the average power constraint $Tr\{\mathbf{D}\} \leq P_T$. Note that if $\mathbf{S}$ is chosen as $\mathbf{S} = \mathbf{V}$, and since $\mathbf{U}^*\mathbf{U} = \mathbf{I}_{n_r}$,$\mathbf{V}^*\mathbf{V} = \mathbf{I}_{n_t}$ we can use the relation

$$\begin{aligned} \det(\mathbf{I} + \mathbf{U}\mathbf{A}\mathbf{U}^*) &= \det(\mathbf{U}\mathbf{U}^* + \mathbf{U}\mathbf{A}\mathbf{U}^*) \\ &= \det(\mathbf{U}(\mathbf{I} + \mathbf{A})\mathbf{U}^*) \\ &= \det((\mathbf{I} + \mathbf{A})\mathbf{U}^*\mathbf{U}) \\ &= \det(\mathbf{I} + \mathbf{A}) \end{aligned} \tag{2.49}$$

to rewrite the mutual information as

$$\begin{aligned} I &= \max_{\mathbf{D}, Tr\{\mathbf{D}\} \leq P_T} \log_2 \det \left\{ \mathbf{I}_m + \frac{\Lambda\mathbf{D}\Lambda^*}{\sigma_n^2} \right\} \\ &= \max_{\gamma_i} \log_2 \prod_{i=1}^{m} \left\{ 1 + \frac{\lambda_i \gamma_i}{\sigma_n^2} \right\} \\ &= \max_{\gamma_i} \sum_{i=1}^{m} \log_2 \left\{ 1 + \frac{\lambda_i \gamma_i}{\sigma_n^2} \right\} \end{aligned} \tag{2.50}$$

where $\lambda_i, i = 1, \ldots, m$ are the eigenvalues to the matrix $\mathbf{H}\mathbf{H}^*$ and γ_i are the n_t eigenvalues to $\mathbf{R}_{cc}$. Note that if $n_t > n_r$, only n_r out of the n_t singular values γ_i are used in calculation of the mutual information I. Hence, these are set to $\gamma_{m+1} = \ldots = \gamma_{n_t} = 0$ otherwise they will "consume" power that cannot reach the receiver.

Hence, if the transmitter has CSI, the transmitted signal can be designed so the covariance matrix of the transmitted signal has the same singular vectors, ($\mathbf{S} = \mathbf{V}$) as the channel. This is sometimes denoted as eigen-beamforming or eigenmode transmission since the eigenvectors of the channel is used as beamforming vectors. The MIMO system will, with this transmission strategy, be decoupled to a system with m parallel, non-interacting subchannels with corresponding gains $\lambda_i^{1/2}$ and the vector

$$\mathbf{c}(n) = \mathbf{V}\mathbf{D}^{1/2}\tilde{\mathbf{c}}(n) \tag{2.51}$$

is transmitted from the n_t antennas. Note that if $n_t > n_r$ then the $n_t - n_r$ last elements in $\tilde{\mathbf{c}}(n)$ must be zero, since only m subchannels (eigenvalues) have non-zero gain. Similarly, if $n_t < n_r$, the $n_r - n_t$ last elements in $\tilde{\mathbf{y}}(n)$ are zero. The power allocated to each parallel channel, given by the eigenvalues γ_i are free parameters which can be chosen to maximize the mutual information. This method

was derived by Gallager [49] and the optimal strategy is often referred to as *water-filling*. In this case, spatial water filling is employed as the power is distributed over spatial eigenmodes. For a given signal to noise ratio P_T/σ_n^2, the power γ_i is allocated to eigenchannel i, where

$$\gamma_i = \left(\mu - \frac{\sigma_n^2}{\lambda_i}\right)^+ \qquad i = 1, \ldots, m \tag{2.52}$$

and $(x)^+ = \max(0, x)$. The "water level" μ is given by the criterion

$$\sum_{i=1}^{m} \gamma_i = P_T \; . \tag{2.53}$$

This gives the mutual information of a flat fading MIMO system when CSI is available at the transmitter

$$I_{CSI} = \sum_{i=1}^{m'} \log_2 \left\{ \frac{\mu \lambda_i}{\sigma_n^2} \right\} \tag{2.54}$$

where m' is the number of non-zero power allocations γ_i. Note that the transmitter also needs knowledge of σ_n^2, the noise power at the receiver, to be able to calculate the mutual information optimal power allocation.

The mutual information in (2.54) can be achieved by generating random data sequences with Gaussian and i.i.d. components and having each codeword split into M blocks of N vectors with n_t components each. This gives an optimal encoder for the constant gain AWGN channel concatenated with an optimal beamformer, with weighting matrix $\mathbf{V}$, that generates these m parallel subchannels over the MIMO-channel $\mathbf{H}$ [27]. The m channels will have different SNR and hence, the modulation method should be adapted to the subchannel SNR to optimize the spectral efficiency or to minimize the BER. This technique will be further explored in Section 2.4 and in Section 3.6 for a special case of the MIMO channel where the elements of $\mathbf{H}$ are i.i.d. and have Rayleigh distributed amplitudes and uniformly distributed phases over $[-\pi, \pi[$.

If the distribution of the fading in the channel is known, then it is possible to calculate the ergodic (Shannon) capacity from Definition 2.1 as

$$C_{CSI} = E_{\mathbf{H}} \{I_{CSI}\} = E_\lambda \left[\sum_{i=1}^{m'} \log_2 \left\{ \frac{\mu \lambda_i}{\sigma_n^2} \right\} \right] \tag{2.55}$$

since the water-filling algorithm maximizes the mutual information for every possible fading state, it will also maximize the ergodic channel capacity. Note that in (2.55), μ depends on the eigenvalues.

2.2.2 Mutual information with no CSI at the transmitter

When CSI is not available at the transmitter,no directions in the channels will be preferred and hence, the transmission covariance matrix is chosen as

$$\mathbf{R}_{cc} = \frac{P_T}{n_t}\mathbf{I}\,. \tag{2.56}$$

This corresponds to transmitting independent and equal power signals on each antenna. Maximum capacity (if no CSI is available at the transmitter) is thus achieved by random codes with independent and circularly symmetric Gaussian distributed symbols [27]. To obtain this, a stream of data symbols can be scrambled with n_t orthogonal scrambling codes, to obtain the property in (2.56), and then be transmitted from the n_t antennas.

The mutual information becomes, from (2.38) (if $n_r < n_t$, otherwise $\mathbf{HH}^*$ is replaced by $\mathbf{H}^*\mathbf{H}$)

$$I_{noCSI} = \log_2 \det\left\{\mathbf{I}_m + \frac{P_T}{\sigma_n^2 n_t}\mathbf{HH}^*\right\} = \sum_{i=1}^{m}\log_2\left\{1 + \frac{P_T}{\sigma_n^2 n_t}\lambda_i\right\} \tag{2.57}$$

bit/s/Hz and the corresponding ergodic capacity is

$$C_{noCSI} = E_{\mathbf{H}}\left[I_{noCSI}\right] = E_\lambda\left[\sum_{i=1}^{m}\log_2\left\{1 + \frac{P_T}{\sigma_n^2 n_t}\lambda_i\right\}\right]\,. \tag{2.58}$$

For any channel distribution, $C_{noCSI} \leq C_{CSI}$. This is obvious, since by adding CSI at the transmitter, capacity cannot be reduced compared to the no CSI case because we could always choose not to exploit the CSI. The spatial multiplexing BLAST code [30] actually achieves the full channel capacity (2.57). The BLAST code imposes no restriction on the matrix $\mathbf{C}_k$ in (1.22) so it can be chosen to fulfill (2.56). However, BLAST suffers from bad performance unless $n_r > n_t$ and the channel is not ill conditioned as was discussed in Section 2.1.

2.2.3 MISO and SIMO systems

A SIMO, $\mathbf{H} = \mathbf{h}^*$, and a MISO channel, $\mathbf{H} = \mathbf{h}$, yields a matrix $\mathbf{HH}^*$ with only a single non-zero eigenvalue $\lambda = |\mathbf{h}|^2$. Hence, the mutual information for a MISO system where CSI is not available at transmitter is

$$I_{noCSI}^{MISO} = \log_2\left\{1 + \frac{P_T}{\sigma_n^2 n_t}\lambda\right\} \tag{2.59}$$

and

$$I_{CSI}^{MISO} = \log_2 \left\{ 1 + \frac{P_T}{\sigma_n^2} \lambda \right\} = I_{noCSI}^{SIMO} = I_{CSI}^{SIMO} \tag{2.60}$$

with CSI at transmitter. So, the difference between the open and closed loop system in the MISO case can be seen direct in loss of SNR, which can be substantial if n_t is large since then the available transmit power is distributed over many transmit antennas. In the SIMO case there is no difference in mutual information with or without CSI at the transmitter since there is only one transmit antenna.

2.3 Communication with no CSI at the transmitter

2.3.1 Spatial Multiplexing

In spatial multiplexing the incoming data stream is split into n_t data streams which are transmitted independently and simultaneously on the n_t antennas [30]. The receiver is able to remove the mixing effect of the MIMO channel and de-multiplex the n_t symbol streams if[3] $n_r \geq n_t$. A number of algorithms are possible for this operation, including e.g. the ML receiver derived in Section 2.1, the linear zero–forcing (ZF) receiver, the minimum mean square error (MMSE) receivers [52] and the successive cancellation receiver [30].

Spatial multiplexing is a linear code and can be expressed using (1.23). Since symbols are not dispersed in time, the space time block code word length, N, is $N = 1$. Thus, the code matrices become column vectors. If a real symbol alphabet is assumed, $s(n) = \bar{s}(n)$, then we obtain $n_s = n_t, N = 1$ and

$$\mathbf{A}_i = \sqrt{2} \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots 0 \end{bmatrix}^T, \tag{2.61}$$

c.f. (1.23), where the figure 1 is in the i:th position. Spatial multiplexing achieves a high bit rate by transmitting independent symbol streams on each antenna but suffers from sensitivity of the channel rank [41] and also from the requirement on the number of receive antennas, $n_r \geq n_t$ if we want to transmit independent data streams on the n_t antennas. Hence only MIMO systems can use spatial multiplexing[4]. The rank sensitivity was seen in (2.24) where the PEP bound was shown to depend on the smallest eigenvalue λ_{min} to the matrix $\mathbf{H}^* \mathbf{R}_{vv}^{-1} \mathbf{H}$. Furthermore, spatial multiplexing has no built-in spatial or temporal coding so more robust coding schemes have been found at the expense of reduced bit rate. This motivates the next section.

[3] The receiver can uniquely de-multiplex $\min(n_r, n_t)$ symbol streams.

[4] Sphere decoding finds the ML estimate even if $n_r < n_t$, but the computational complexity is exponential in $n_t - n_r$ [53].

2.3.2 Space-Time Coding

Space time coding is a *transmit diversity* technique that applies to both MISO and MIMO systems. Hence it is an attractive signalling scheme when the receiver only has one antenna, due to space and/or cost limitations. Space time coding introduces spatial and temporal correlation between the signals transmitted from different antennas in an intelligent manner, to provide diversity at the receiver and thereby a reliable reception of the transmitted symbols. Hence, a larger symbol constellation set $\mathcal{X}$ can be used to decrease the difference in bit rate somewhat, compared to spatial multiplexing. One member of the family of space time codes is the space time trellis code (STTC) that combines the diversity advantage with a coding advantage. The drawback of the STTC is the detection complexity, since a Viterbi detector is required and the number of states increases as $r_d^{n_t}$ when M_d-ary modulation is used. To avoid this disadvantage, orthogonal space-time block codes (STBC) have been proposed. They belong to the family of *linear codes* and the k:th codeword matrix can thus be described by

$$\mathbf{C}_k = \sqrt{\frac{P_T}{2n_t}} \sum_{n=1}^{n_s} \left[\mathbf{A}_n \bar{s}^{(k)}(n) + j\mathbf{B}_n \tilde{s}^{(k)}(n) \right] \quad . \tag{2.62}$$

It has been shown that STBC can be combined with an outer trellis code and actually outperform the STTC with the same number of trellis states (same computational complexity) and in the SNR region of interest when $n_r = 1$ or 2 [54].

Space time codes require CSI at the receiver, which in FDD systems has to be estimated using a training sequence. Another option is to use differentially detected STBC [36] which require no CSI at the receiver but has a performance penalty of 3 dB in SNR.

The ML-detector for the sequence of symbols $\{s(n)\}_{n=1}^{n_s}$, derived in Section 2.1, becomes linear and decoupled for orthogonal STBC, as will be shown below. These properties makes orthogonal STBC attractive since a linear detector implies a simple algorithm in the receiver. Also the channel estimation process using pilot symbols becomes trivial since the orthogonal STBC codewords are semi-unitary (that is, $\mathbf{C}\mathbf{C}^* = \mathbf{I}$). Hence, the ML estimate of the channel using training symbols involves the inverse of the matrix $\mathbf{C}\mathbf{C}^*$ [55] which then is trivial due to its diagonal form. The low complexity receiver makes it possible to put most of the complexity on one end of the wireless link. A portable unit might only have one antenna and limited battery power. The basestation can then use classical receive diversity and STBC transmit diversity since the detection and channel estimation in the mobile is computationally simple.

Orthogonal Space-Time Block Codes

We will now discuss how the underlying structure of the orthogonal STBC simplifies the expression (2.33), the test statistic for the ML-detector of linear space time block codes in temporally and spatially white noise. To simplify the discussion, assume that the symbol constellation $\mathcal{X}$ is unitary, i.e. $|s(n)|^2 = 1$, even though this is not strictly necessary. The ML-detector for the real part of the transmitted symbol $s(n)$ is then

$$\hat{\bar{s}}(n) = \arg \max_{s(n)\in\mathcal{X}} \text{ReTr}\left\{\bar{s}(n)\mathbf{A}_n^*\mathbf{H}^*\mathbf{Y}\right\} \tag{2.63}$$

and for the imaginary part of $s(n)$

$$\hat{\tilde{s}}(n) = \arg \max_{s(n)\in\mathcal{X}} \text{ReTr}\left\{-j\tilde{s}(n)\mathbf{B}_n^*\mathbf{H}^*\mathbf{Y}\right\} \tag{2.64}$$

where the details of these derivations are carried out in Appendix 2.B on page 69. The derivation relies on the special properties of the matrices $\mathbf{A}_k$ and $\mathbf{B}_k$ of the space time block code-word matrix (1.23):

$$\begin{aligned}
\mathbf{A}_j\mathbf{A}_j^* &= \mathbf{I} \qquad \forall j \\
\mathbf{B}_j\mathbf{B}_j^* &= \mathbf{I} \qquad \forall j \\
\mathbf{A}_j\mathbf{A}_k^* &= -\mathbf{A}_k\mathbf{A}_j^* \qquad j \neq k \\
\mathbf{B}_j\mathbf{B}_k^* &= -\mathbf{B}_k\mathbf{B}_j^* \qquad j \neq k \\
\mathbf{A}_j\mathbf{B}_k^* &= \mathbf{B}_k\mathbf{A}_j^* \qquad \forall j,k \ .
\end{aligned} \tag{2.65}$$

These relations were first presented by Tarokh et.al. in [29] using coding theory and then re-derived by Ganesan and Stoica in [56, 57] using maximum receiver SNR arguments. It is shown in Appendix 2.B that the symbol estimation of (2.63) and (2.64) decouples and an equivalent SISO model, as "seen" by each symbol $s(n)$ can be derived. The received symbol $r(n)$ at time instant n for the underlying SISO channel[5] can be written, for the real part as

$$\bar{r}(n) = \sqrt{\frac{P_T}{2n_t}}\,\|\mathbf{H}\|_F^2\,\bar{s}(n) + \text{ReTr}\left\{\mathbf{A}_n^*\mathbf{H}^*\mathbf{V}\right\} \tag{2.66}$$

and for the imaginary part

$$\tilde{r}(n) = \sqrt{\frac{P_T}{2n_t}}\,\|\mathbf{H}\|_F^2\,\tilde{s}(n) + \text{ReTr}\left\{-j\mathbf{B}_n^*\mathbf{H}^*\mathbf{V}\right\} \ . \tag{2.67}$$

[5]Note that a whole space time codeword of length N must be received before $r(n)$ can be calculated.

The problem of finding matrices that satisfy the relationships (2.65) is non-trivial and linked to the theory of Radon-Hurwitz matrices [58] and forms the basis of the design of orthogonal space time block codes. Before the existence of matrices that satisfy the relations above is discussed, a few definitions related to STBC is required.

The rate of the code is the number of transmitted symbols per symbol interval, n_s/N which is $n_s/N \leq 1$ for orthogonal STBC. The STBC is called *full rate* if $n_s/N = 1$, since then in N symbol intervals, a maximum of $n_s = N$ symbols are transmitted. Furthermore, the STBC is called *minimum delay* if $N = n_t$, since this is the minimum value of N for a code still to be full rate. A drawback of orthogonal STBC is that full rate code matrices does not exists for all values of n_s, N and n_t, i.e. matrices that fulfill the criterions (2.65) does not exist. For a complex symbol alphabet, eg. QPSK, a full rate, delay optimal STBC exists only for $n_t = 2$ and it is achieved by the *Alamouti scheme* [13]

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{A}_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{2.68}$$

and

$$\mathbf{B}_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \mathbf{B}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \tag{2.69}$$

The corresponding code-word matrix for the Alamouti scheme is

$$\mathbf{C} = \sqrt{\frac{P_T}{2}} \begin{pmatrix} s(n) & -s^*(n+1) \\ s(n+1) & s^*(n) \end{pmatrix} . \tag{2.70}$$

For $n_t = 3, 4$, codes with rate $3/4$ have been found [58]. The key property of orthogonal space time block codes is that the modulation matrices $\mathbf{A}_i, \mathbf{B}_j$ in (2.62) of the STBC are unitary and pairwise orthogonal, that is, for any weighting matrix $\mathbf{Q}$, $\mathrm{ReTr}(\mathbf{A}_k^* \mathbf{Q} \mathbf{A}_l) = 0$ for all $0 \leq k \neq l \leq n_s$. It was shown in [59] that modulation matrices with this property achieves the minimum upper bound on symbol error probability conditioned on the channel matrix $\mathbf{H}$, without any assumptions on the statistics of $\mathbf{H}$, or in other words without knowledge of CSI at the transmitter. This general approach to design orthogonal STBC:s was first presented in [29], although the Alamouti scheme for $n_t = 2$ antennas was discovered somewhat earlier.

The SISO equivalence of orthogonal STBC

We noted above that the input symbols get decoupled by the orthogonal space time block code, see also Appendix 2.B. For each transmitted symbol, $s(n)$, the MIMO

or MISO system can be treated as an equivalent SISO system

$$\mathbf{r} = \sqrt{\frac{P_T}{2n_t}} \|\mathbf{H}\|_F^2 \mathbf{s} + \mathbf{v} \ . \tag{2.71}$$

In the complex symbol alphabet case, $\mathbf{s}$ and $\mathbf{r}$ are the $2n_s \times 1$ input and output signal vectors and $\mathbf{v}$ is the $2n_s \times 1$ noise vector. It was shown in [60] that the noise vector has i.i.d Gaussian zero mean elements with variance $\sigma_n^2 \|\mathbf{H}\|_F^2 /2$. This means that the received power of the equivalent SISO system (2.71) is (per complex dimension) [60]

$$SNR_{eq} = \frac{E\left\{\|\mathbf{H}\|_F^2 (\sqrt{\frac{P_T}{2n_t}})^2 |s(n)|^2 \|\mathbf{H}\|_F^2\right\}}{\sigma_n^2 \|\mathbf{H}\|_F^2 /2} = \frac{P_T}{\sigma_n^2 n_t} \|\mathbf{H}\|_F^2 \ . \tag{2.72}$$

A SISO system with this SNR has the channel capacity[6]

$$C = E_{\mathbf{H}} \left\{ R \log_2 \left\{ 1 + \frac{P_T}{\sigma_n^2 n_t} \|\mathbf{H}\|_F^2 \right\} \right\} \tag{2.73}$$

where the factor $R = n_s/N$ is the (complex symbol) code rate. If this capacity is compared to the Shannon capacity of the channel, expression (2.58), then it is observed that orthogonal STBC equals the Shannon capacity if $R = 1$ and the channel has rank one. This can easily be shown by observing that for a rank one matrix $\mathbf{H}$, the Frobenius norm $\|\mathbf{H}\|_F^2 = \lambda_1$, so (2.58) and (2.73) are equivalent if R=1.

For a channel with arbitrary fading, and multiple transmit antennas, only a MISO system is guaranteed to have rank one. So orthogonal STBC with a single receiver antenna can achieve Shannon capacity. A code rate of R=1 can for complex symbol alphabets only be achieved with two transmit antennas. The Alamouti code used for a 2×1 MISO channel can thus attain the full Shannon capacity. This is the only orthogonal STBC over a MIMO (MISO) channel that has this property if CSI is not available at the transmitter.

2.4 Communication when CSI is available at the transmitter

Now, a novel way to transmit in a MIMO channel when CSI is available at the transmitter is derived. The aim is to approach the channel capacity as promised

[6]For real symbols $C = (R/2) \log_2 \left\{ 1 + \frac{2P_T}{\sigma_n^2 n_t} \|\mathbf{H}\|_F^2 \right\}$.

by Shannon, see the discussion in Section 2.2.1. However, the Shannon capacity places no restrictions on the complexity or delay of the transmission scheme which achieves this capacity. Therefore, a practical transmission scheme is derived using a combination of adaptive modulation and power waterfilling over the m parallel spatial eigenmodes of the channel. The idea is inspired by Section 2.2.1 where it was shown that the input output MIMO channel could be written as a set of m non-interfering and thus parallel channels, from (2.47)

$$\tilde{\mathbf{y}}(n) = \Lambda\tilde{\mathbf{c}}(n) + \tilde{\mathbf{v}}(n) \tag{2.74}$$

where $\tilde{\mathbf{c}}(n)$ is the $m \times 1$ vector of input symbols, $\tilde{\mathbf{y}}(n)$ the corresponding $m \times 1$ vector of output symbols and $\tilde{\mathbf{v}}(n)$ is the noise vector[7]. Note that the channel has been diagonalized using the singular value decomposition (SVD) of the channel and the left and right singular vectors as receive and transmit weight vectors respectively. The parallel channel power gains are given by the eigenvalues λ_i to the Hermitian matrix $\mathbf{W} \stackrel{\triangle}{=} \mathbf{H}\mathbf{H}^*$. The channel $\mathbf{H}$ is assumed to be quasi-stationary, hence (2.47) holds during a block of symbols of length N. See also the block fading model in Section 1.1.2. Thus the SVD of the channel that leads to Λ in (2.47) is valid for N symbols.

How the available transmit power P_T is distributed over the m channel eigenmodes depends on the particular design goal, such as e.g. high bit rate or low BER. Utilization of only one of the m eigenmodes is equivalent to beamforming (this eigenmode then receives all power P_T), see Section 2.5.

Compared to SISO channel adaptive modulation techniques [50, 61, 62], the m parallel channels of the MIMO channel gives us more degrees of freedom to adapt the bit rate and power to mitigate the channel fading. Although rate[8] adaption over the MIMO channel is new, the problem formulation is similar to bit-loading of discrete multitone modulation (DMT) over wired lines such as twisted pair connections [63, 64]. It is also similar to rate adaption in OFDM, see [65]. In [66], adaptive modulation over a Rayleigh fading SIMO channel was investigated for different diversity combining methods. An interesting conclusion was that adapting rate and power simultaneuously gives a small improvement over just rate adaption and constant power transmission.

The authors of [67] presented, independently of the work presented here, a joint optimal precoder and decoder for a MIMO channel where the CSI is available at the transmitter. In their work a minimization of the weighted sum of symbol errors over the subchannels under a peak power constraint and with continuous number of bits per symbol allocations (unrestricted modulation constellation sets)

[7] The $n_t - m$ zero elements in $\tilde{\mathbf{c}}(n)$ is removed so this vector is of dimension $m \times 1$.

[8] "Rate" in the meaning *bit rate*.

is performed. The work presented in this section extends parts of the work in [67] to include average power constraints, BER minimization and restricted modulation constellation sets as well. Furthermore, in [67], the Rayleigh fading channels were obtained through Monte Carlo simulations. In Chapter 3, expressions are derived for the pdf:s of the subchannel gains, assuming a Rayleigh fading MIMO channel. These are used here to analytically examine the performance instead of performing a Monte Carlo simulation.

Some practical constraints are linked to the design of rate adaptation systems. For instance, the channel must not change so fast that the CSI becomes invalid, which is determined by the maximum Doppler frequency of the channel[9]. This will also determine the block length N, which defines the length of the interval where the channel is assumed to be time-invariant.

In the following derivations we shall assume that the channel is known without errors at the transmitter and that the power and the bit rate can be adapted instantaneously[10]. Furthermore, it is assumed that the receiver and transmitter has agreed on a data transmission rate, that is, a modulation constellation for each utilized subchannel. Some of the derived algorithms require the probability density function (pdf) of the channel gains (eigenvalues to $\mathbf{HH}^*$). In a practical system, these must be estimated using some identification algorithm. In this section, it is assumed that the pdf:s are known without errors.

The used power constraint in the derivations is a constraint on the total transmitted power P_T, from the antenna array. The power distribution over different *antenna elements* will depend on the power distribution over channel eigenmodes and on the channel eigenvectors. Thus, the power amplifiers at each antenna element must be designed to have a maximum output power P_T. Although this scenario, where all power is transmitted from one out of n_t antenna elements is unlikely, it cannot be ruled out. It is however possible to impose an *element-wise power constraint* in addition to the total power constraint. In [57] it was shown that maximum SNR at the receiver with element-wise power constraints is achieved when $\mathbf{R_{cc}} \propto \mathbf{I}$ which is a restriction that leads to a reduction in the maximum data transmission rate.

Adaptive transmission provides many parameters that can be adjusted relative to the channel fading. The important factors for the optimization are the bit rate (or spectral efficiency), the transmit power and the BER. Different optimization problems occur if average rate[11], power and BER, or instantaneous rate, power and BER, or combinations of average and instantaneous constraints are considered.

[9] This validity time can be extended using channel prediction.

[10] The time it takes for the amplifiers to ramp up or down their transmitted power is thus neglected.

[11] In this section, we use the term rate in the meaning bit rate.

Also the issue of continuous rate adaption, where the set of signal constellations $\mathcal{X}$ is unrestricted, or a more practical scenario, where only a discrete and finite set of constellations are available[12], are handled in the derivation. Continuous rates lead to an optimization problem that can be solved analytically. Such a solution give more insight into the problem.

The optimization can be formulated as one of three objectives

1. Maximize the bit rate subject to constraints on bit error probability and transmission power.

2. Minimize bit error probability subject to constraints on bit rate and transmission power.

3. Minimize transmission power subject to constraints on bit rate and bit error probability.

The study in this thesis focus on the two first problems. The effect of channel coding is not considered. In Section 2.4.2, the bit rate is maximized under an instantaneous BER constraint for both unrestricted and restricted symbol constellations. In Section 2.4.3, the average BER is minimized under average and instantaneous power constraints and for both unrestricted and restricted constellations.

2.4.1 Preliminaries

Assume that M_d-ary modulation is used for the transmission over the MIMO subchannels. Hence, $r_d \triangleq \log_2(M_d)$ bits is mapped onto each symbol in the complex modulation constellation. We assume square constellations in the following, and denote the modulation as M_d-QAM. Square constellations are used due to their inherent spectral efficiency and ease of implementation [15].

A variable-rate and variable-power M_d-QAM technique for a flat fading MIMO channel with the transmission bandwidth W is now derived, inspired by the SISO results in [62]. Assume that subchannel i, where $i = 1, \ldots, m$, is allocated an M_i-QAM modulation alphabet and thus transmits $r_i \triangleq \log_2 M_i$ bits/symbol. The total number of bits transmitted during a symbol interval of time T_s is then

$$R \triangleq \sum_{i=1}^{m} r_i = \sum_{i=1}^{m} \log_2 M_i \tag{2.75}$$

and since T_s^{-1} symbols are transmitted per second, the total bit rate is the equal to RT_s^{-1} bits per second. The spectral efficiency equals the average bit rate per unit

[12]Normally BPSK, QPSK, 16-QAM, 64-QAM,... are the used constellations.

bandwidth, so we normalize (2.75) with the transmission bandwidth W and apply the expectation operator to obtain the average

$$\frac{R}{W} = E\left\{\sum_{i=1}^{m} \log_2 M_i\right\} \tag{2.76}$$

bits/Hz, where Nyquist data pulses are assumed ($W = 1/T_s$).

Assume that the transmitter has the option to use some or all of the $m = \min(n_r, n_t)$ subchannels (also called eigenmodes) with corresponding power gains $\lambda_i, i = 1, \ldots, m$. If the transmitted power in mode i is defined as γ_i, then the receiver SNR of mode i will be

$$\text{SNR}_i \triangleq \frac{\gamma_i \lambda_i}{\sigma_n^2} \tag{2.77}$$

where σ_n^2 is the receiver noise power. If M_i-QAM is used on subchannel i, then the BER for coherently detected M_i-QAM with Gray bit mapping is approximately given by [62]

$$\begin{aligned} \text{BER}_{M_i-\text{QAM}} \approx & 0.2 \exp\left(-1.6 \frac{\text{SNR}_i}{(M_i - 1)}\right) \\ = & 0.2 \exp\left(-1.6 \frac{\lambda_i \gamma_i}{\sigma_n^2 (M_i - 1)}\right) \end{aligned} \tag{2.78}$$

which is tight within 1 dB when $M_i \geq 4$ and BER$\leq 10^{-3}$. This approximation gives a BER expression that is invertible in its arguments and is differentiable which is required for the adaptive transmission design.

If a target BER, denoted BER$_t$, is desired for the particular transmission over the MIMO channel, modulation schemes and transmit powers for the set of m subchannels can now be assigned by choosing the $\{M_i\}_{i=1}^m$ and $\{\gamma_i\}_{i=1}^m$. Alternatively, if a total bit rate R_T for the MIMO transmission is desired, modulation schemes and powers are allocated to the different subchannels to minimize the resulting BER. How this shall be performed is derived in the next section.

2.4.2 Maximizing spectral efficiency at a target BER

Assume that all subchannels are set to have the same target BER, denoted BER$_t$. This is equivalent to stating that we minimize the maximum BER for any subchannel i [63]. That is

$$\text{BER}_t = \min_{\gamma_i, M_i} \max_{i=1,\ldots,m} \text{BER}_{M_i-\text{QAM}}(\gamma_i, M_i, \lambda_i) \ . \tag{2.79}$$

Hence the total target BER is equal to each subchannel:s target BER (note that the symbol error rate for the subchannels can vary). Under this constraint, we would like to maximize the spectral efficiency in (2.76).

If the number of bits per symbol for subchannel i fulfill $r_i \geq 2$ and the corresponding bit error rate fulfill $\text{BER}_i \leq 10^{-3}$, then (2.78) is valid. The number of bit/symbol for subchannel i, is then obtain by solving for r_i in expression (2.78)

$$r_i = \log_2 \left(1 + K_o \gamma_i \lambda_i\right) \tag{2.80}$$

where

$$K_o = \frac{-1.6}{\sigma_n^2 \ln(5\text{BER}_t)} \ . \tag{2.81}$$

Continuous rate allocation

The following theorem summarizes the result from the optimization of spectral efficiency.

Theorem 2.1 *Assume that a MIMO channel has been converted to a set of m non-interfering and parallel channels, according to* (2.74)*, and that* (2.78) *holds exactly. The marginal probability density function of the i:th subchannels gain (eigenvalue) is $p_i(\lambda_i)$. If the number of bits per symbol, r_i, used on the i:th subchannel, is real and non-negative and an average total transmitted power constraint of P_T is used, then the power control law for subchannel i that maximizes the spectral efficiency* (2.76) *under a target BER constraint is given by*

$$\gamma_i(\lambda_i) = \begin{cases} \frac{1}{\mu} - \frac{1}{K_o \lambda_i} & \gamma_i \geq 0, r_i \geq 0 \\ \\ 0 & \textit{otherwise} \end{cases} \tag{2.82}$$

where μ is a constant and K_o is given by (2.81)*. The constant μ is given by solving the average power constraint criterion*

$$\sum_{i=1}^{m} \int_0^{\infty} \left[\frac{1}{\mu} - \frac{1}{K_o \lambda_i}\right]^+ p_i(\lambda_i) d\lambda_i = P_T \ . \tag{2.83}$$

The resulting spectral efficiency is given by

$$\frac{R}{W} = \sum_{i=1}^{m} \int_{\mu/K_o}^{\infty} \log_2 \left(\frac{K_o \lambda_i}{\mu}\right) p_i(\lambda_i) d\lambda_i \ . \tag{2.84}$$

Proof: See Appendix 2.C on page 72. ■

Corollary 1 *If the same derivations as above are performed under the peak power constraint, then the same power control law* (2.82) *is found. However,* μ *is now given by the equation*

$$\sum_{i=1}^{m} \left[\frac{1}{\mu} - \frac{1}{K_o \lambda_i} \right]^+ = P_T \ . \tag{2.85}$$

Proof: Immediate result from the peak power constraint criterion

$$\sum_{i=1}^{m} \gamma_i(\lambda_i) = P_T \tag{2.86}$$

and by inserting (2.82). ■

Remark 2.2 *The power control law is a water filling solution. The threshold* μ^{-1} *is a function on the average power* P_T*, the target bit error rate* BER_t*, the receiver noise variance* σ_n^2*, and the fading distribution. It determines when a subchannel should be used. It is sometimes denoted the "water-level" of the water filling algorithm.*

Remark 2.3 *The resulting spectral efficiency is equal to the Shannon capacity with a power penalty* $1/K_o$*. If the water filling solution in Section 2.2.1, that achieves the Shannon capacity is compared with* (2.82)*, then it can be seen that the difference is in the constant* K_o *that depends on the target BER.* $1/K_o$ *is the maximum possible coding gain for an adaptive MQAM method [61]. This gap between the Shannon capacity and MQAM spectral efficiency was reported earlier in [61] for SISO systems and is here extended to a MIMO system.*

Remark 2.4 *There is no dependence on the fading distribution under the peak power constraint, since the transmission is optimized instantaneously, with no "memory" as opposed to the average power constraint case.*

Since the average power solution requires the pdf of the eigenvalues to $\mathbf{HH}^*$, a model for the channel fading statistics is required to further examine this case. However, the case with a peak power constraint is can be illustrated by an example, since it is independent on the fading distribution:

EXAMPLE 2.1

Assume that the minimum number of receive and transmit antennas is two, i.e. $m = 2$ and $\lambda_1 \geq \lambda_2$. The power allocation that maximize the spectral efficiency while maintaining a target bit error rate is sought. Assume that both subchannels are active, that is $\frac{1}{\mu} - \frac{1}{K_o \lambda_i} > 0$ for $i = 1, 2$, where K_o is given by (2.81). The constant μ can then be solved from (2.85) to be

$$\mu = \frac{2}{P_T + \frac{1}{K_o}\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right)}$$

and the corresponding power to the two subchannels are

$$\begin{aligned} \gamma_1 &= \frac{P_T}{2} + \frac{1}{2K_o}\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right) \\ \gamma_2 &= \frac{P_T}{2} - \frac{1}{2K_o}\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1}\right) \end{aligned} .$$

More power is allocated to the subchannel with the largest gain (λ_1) and from (2.80) it is seen that more power also implies the use of a symbol constellation with more bits allocated per symbol. Furthermore, if $\lambda_1 \gg \lambda_2$ as in a line of sight channel, much more power is allocated to subchannel 1 until subchannel 2 is shut off completely. A decrease in the target BER implies a decrease in K_o and hence, even more power is allocated to the channel with the highest gain, and eventually, the weakest channel is shut off and becomes unused.

For low target bit error rate applications, or in channels with low SNR, the use of only one channel eigenmode, i.e. beamforming, is thus often equal to the optimal transmission strategy. However, the dependence on the channel eigenvalues and available transmit power is strong, as will be illustrated in the next example. The power distributions over the channel eigenmodes as a function of available power P_T is calculated for two scenarios corresponding to a LOS and a NLOS channel.

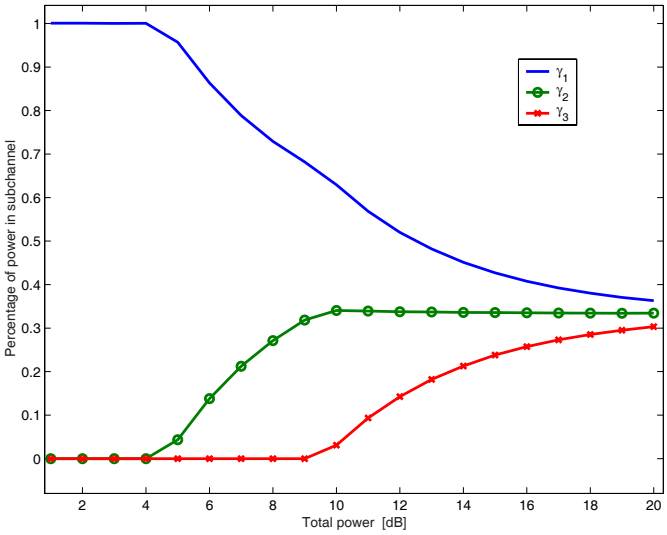


Figure 2.2: Power distribution over subchannels when $m = 3$ and the eigenvalue distribution is $30 : 2 : 1$. Peak power constraints are imposed and BER target is 10^{-5}. The noise power is set to $\sigma_n^2 = 1$.

EXAMPLE 2.2

First, assume a 3×3 MIMO channel that has a strong LOS component, hence one eigenvalue dominates. Figure 2.2 shows the calculated distribution of power when the relation between the eigenvalues are $30 : 2 : 1$. The target BER is set to $\text{BER}_t = 10^{-5}$ and the noise power $\sigma_n^2 = 1$. At low transmit power, only the principal eigenmode is used, which is equivalent to beamforming with adaptive modulation (see Section 2.5). When the transmit power is increased, the next subchannel becomes active and when the total output power P_T is large, the power is divided uniformly between the three channel modes. Hence, in the high SNR scenario, information of the relation between the eigenvalues becomes unnecessary. Figure 2.3 shows the corresponding distribution in the $1.2 : 1.1 : 1$ eigenvalue ratio case, corresponding to an NLOS channel. Note the expanded scale on the power axis. At high power, all three modes are asymptotically given an equal share of the available transmit power. Also, *more* power is given to the best channel, to maximize the number of bits allocated per symbol, as opposed to BER minimization at a target spectral efficiency (Section 2.4.3), where the most power is given to the weakest subchannel. This is a fundamental difference between rate optimization and BER optimization.

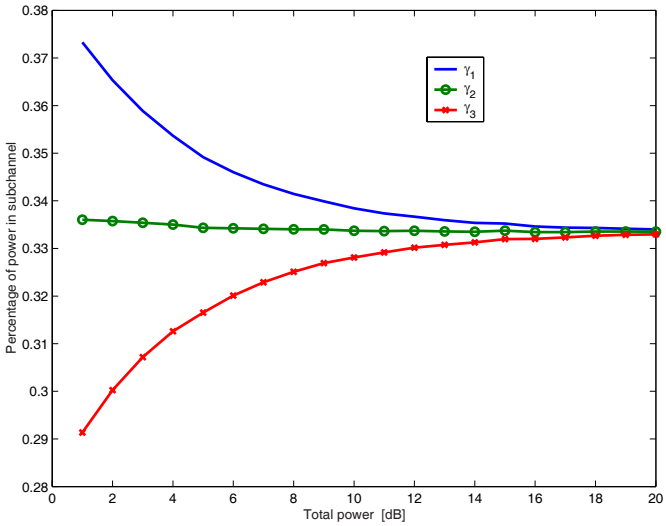


Figure 2.3: Power distribution over subchannels when $m = 3$ and the eigenvalue distribution is $1.2 : 1.1 : 1$. Peak power constraints are imposed and BER target is 10^{-5}. The noise power is set to $\sigma_n^2 = 1$.

Finite constellation set $\mathcal{X}$

The derivation is now restricted by assuming that the modulation uses M_d-QAM constellations. The number of bits allocated per symbol can thus only take certain integer values $r_i = 0, 1, 2, 4, 6, \ldots$ corresponding to BPSK, 4QAM, 16QAM, 64QAM an so forth and where zero rate means that the subchannel is unused. We are interested in the impact on spectral efficiency imposed by the finite constellation restriction. The problem of maximizing the bit rate subject to peak power and BER constraints is equivalent to the *discrete bit loading* techniques that have developed for DMT wired links. A number of algorithms have been proposed to solve the discrete bit loading problem and for large systems, a low complexity algorithm is crucial. Campello has presented a optimal discrete bit loading algorithm for DMT with $O(m)$ complexity [68]. In MIMO systems, m is usually small compared to DSL applications. Therefore, a search algorithm can be used despite its higher complexity, but it still finds the optimum since the search is exhaustive.

The boundaries of the optimal bit rate regions for all the m subchannels are required in the solution for the general case. They depend on both the fading distribution and the amount of power allocated to each subchannel. To find these

region boundaries involves optimization of a mixture of discrete and continuous variables for which no general method exists. Instead the problem is approached here by assuming that the fading state defined by $\lambda_1, \ldots, \lambda_m$ is fixed and then calculate the optimal transmission strategy by choosing bit per symbol allocations and power for the individual subchannels based on this particular fading state. An exhaustive search over all possible combinations of $\{r_i\}_{i=1}^m$ (which now are discrete) is performed and the one that yields the highest spectral efficiency is selected. The discrete optimization problem can thus be formulated as

$$\max \sum_{i=1}^{m} r_i \tag{2.87}$$

$$\text{when} \qquad \sum_{i=1}^{m} \int_0^{\infty} \gamma_i(\lambda_i) p(\lambda_i) d\lambda_i = P_T, \tag{2.88}$$

in the average power case and

$$\max \sum_{i=1}^{m} r_i \tag{2.89}$$

$$\text{when} \qquad \sum_{i=1}^{m} \gamma_i(\lambda_i) \leq P_T \tag{2.90}$$

in the peak power constrained case. The power allocation for subchannel i, to achieve the target bit error rate BER_t for a given r_i is found from (2.78) to be

$$\gamma_i(\lambda_i) = \frac{2^{r_i} - 1}{K_o \lambda_i} \ . \tag{2.91}$$

which is valid if $r_i \geq 2$ and with K_0 given by (2.81). In addition, to allow the solution to contain an one-dimensional modulation technique, the *exact* relation [15] between BER and SNR is BER=$Q(\sqrt{2\gamma_i\lambda_i/\sigma_n^2})$ so

$$\gamma_i(\lambda_i) = \frac{\sigma_n^2}{2\lambda_i} \left(Q^{-1}(BER_t)\right)^2 \tag{2.92}$$

is used for the case $r_i = 1$ (e.g. BPSK) where Q^{-1} is the inverse Q-function.

An example to the optimization problem is now given by using the search technique.

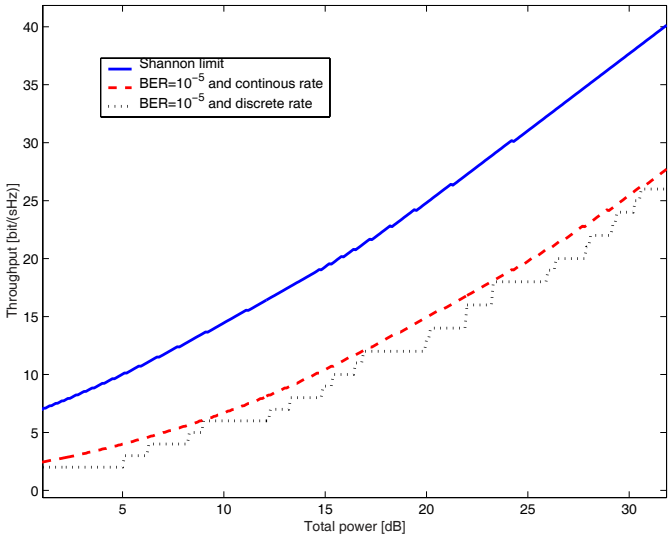


Figure 2.4: Throughput (spectral efficiency), that is, the sum of the number of bits allocated per symbol per subchannel, of a $m = 4$ system with peak power constraints and eigenvalue relations $30 : 8 : 2 : 0.1$. BER target is set to 10^{-5}.

EXAMPLE 2.3

Assume a 4×4 MIMO system. Figure 2.4 shows the maximum total bit rate, or spectral efficiency, of a channel with eigenvalue relations $30 : 8 : 2 : 0.1$ as a function of the peak power constraint and noise power set to $\sigma_n^2 = 1$. The discrete bit per symbol solution is within 3 dB of the continuous solution and there is an additional 9 dB to the Shannon limit.

Figure 2.5 shows the corresponding subchannel rates and how the optimal solution depends on the available peak power for this particular ratios of the subchannel gains (the eigenvalues). Sometimes a subchannel lowers its bit rate at an increased power to the benefit of an increased bit rate for a subchannel with better channel gain.

2.4.3 Minimizing the total BER at a target spectral efficiency

Now, suppose that the target bit rate R is fix and that the power and bit per symbol allocation that *minimizes the average bit error rate* at the receiver is sought. Define

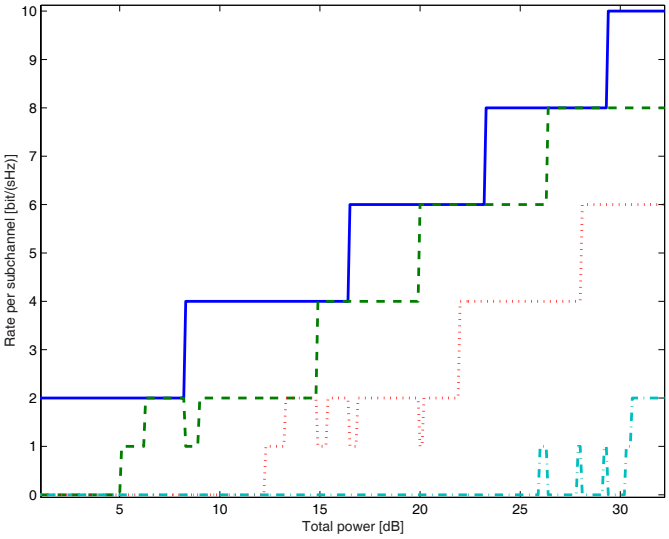


Figure 2.5: Subchannel rates (number of bits allocated per symbol) for $m = 4$ system with peak power constraints and eigenvalue relations $30 : 8 : 2 : 0.1$.

the target spectral efficiency R_T/W as

$$\frac{R_T}{W} = \sum_{i=1}^{m} r_i \ . \tag{2.93}$$

The average bit error rate is

$$\overline{\text{BER}} = \frac{1}{R_T} \sum_{i=1}^{m} \int_0^\infty r_i \text{BER}_i(\lambda_i, \gamma_i) p_i(\lambda_i) d\lambda_i \tag{2.94}$$

which can be minimized by allocating the correct continuous bit per symbol allocations r_i to the subchannels and jointly allocating the optimal distribution of power, γ_i, to the subchannels. The optimal allocations are given in the following theorem.

Theorem 2.2 *Assume that a MIMO channel has been converted to a set of m non-interfering and parallel channels, according to* (2.74)*. The marginal probability density function for the i:th subchannel gain (eigenvalue) is $p_i(\lambda_i)$. If the m subchannels bit per symbol allocations are real and non-negative and an average power constraint of P_T is used, then the power control law and the bit per symbol allocation for subchannel i that minimizes the average BER at a target spectral efficiency, are given by the two coupled equations*

$$\gamma_i = \left[\frac{2^{r_i} - 1}{c_1 \lambda_i} \ln \left(\frac{c_1 c_2 \lambda_i r_i}{(2^{r_i} - 1) R_T \mu_1} \right) \right]^+ \tag{2.95}$$

and

$$\int_0^\infty \left(BER(\lambda_i, \gamma_i) + r_i \frac{\partial BER_i(\lambda_i, \gamma_i)}{\partial r_i} \right) p_i(\lambda_i) d\lambda_i + \mu_2 R_T = 0 \ . \tag{2.96}$$

where $c_1 = 1.6/\sigma_n^2$, $c_2 = 0.2$ *and* $BER(\lambda_i, \gamma_i)$ *is given by* (2.78).

The Lagrange multipliers μ_1 *and* μ_2 *are found by solving the average power constraint*

$$\sum_{i=1}^{m} \int_0^\infty \gamma_i p_i(\lambda_i) d\lambda_i = P_T \tag{2.97}$$

and the target bit rate constraint

$$\sum_{i=1}^{m} r_i = R_T \ . \tag{2.98}$$

Proof: See Appendix 2.D on page 74. ■

Remark 2.5 *The power control law* $\gamma_i(\lambda_i, r_i)$ *and the bit rate allocation law* $r_i(\lambda_i, \gamma_i)$*, implicit in equation* (2.96) *are a set of coupled non-linear equations. Different subchannel rates and powers are also coupled through the Lagrange multipliers* μ_1, μ_2*. To solve this optimization problem, numerical methods are required.*

To achieve some insight into the power control law, assume for a moment that the subchannel rates are constants and equal $r_i = R_T/m$. The power control law can be expressed as the function $f(x) = [\ln(x/b)/x]^+$ and is plotted in Figure 2.6. By finding the maximum of the function $f(x)$ it can be seen that the power per channel is upper bounded by $\max(\gamma_i) = c_2 r_i/(R_T \mu_1 e^1)$ and this occurs when the channel gain λ_i is $\lambda_i = e\mu_1(2^{r_i} - 1)R_T/(c_1 c_2 r_i)$. If the parameter b is small, then the power control function will be sharply peaked around its maximum value which implies that weak subchannels will be given a large transmit power and strong subchannels a lower power. Under a certain threshold, the subchannel is unused. Hence, the algorithm balances the subchannel links, in a form of power inversion.

Corollary 2 *In the peak power constrained case*

$$\sum_{i=1}^{m} \gamma_i(\lambda_i, r_i) = P_T \tag{2.99}$$

is exchanged for (2.97) *and solved to obtain* μ_1.

Proof: Immediate since the total power of the m subchannels must be P_T. ■

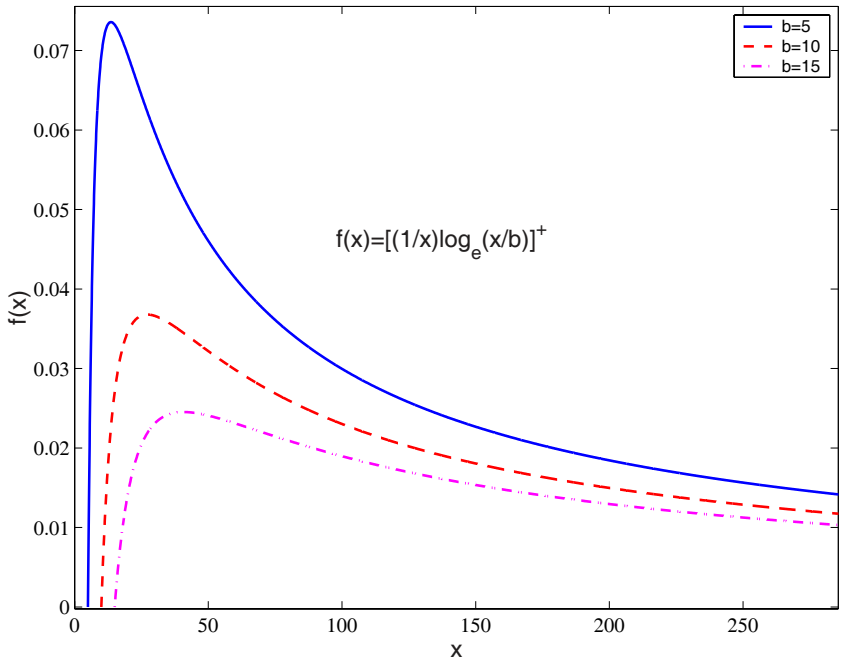


Figure 2.6: The power control law (2.95) *as a function of the subchannel gain. The function has it maximum at* be *with the value* $(be)^{-1}$.

Finite constellation set $\mathcal{X}$

If the subchannel rate allocations r_i are discrete and belonging to a known set $\mathcal{X}$, then the minimization of BER is performed by a search algorithm over all possible combinations of bit per symbol allocations that fulfills the total bit rate criterion (2.98). Then we select the allocations that yields the lowest BER. The Lagrange multiplier μ_1 in (2.95) is given in the average power case by solving

$$\sum_{i=1}^{m} \int_0^{\infty} \left[\frac{2^{r_i}-1}{c_1 \lambda_i} \ln \left(\frac{c_1 c_2 \lambda_i r_i}{(2^{r_i}-1) R_T \mu_1} \right) \right]^+ p_i(\lambda_i) d\lambda_i = P_T \tag{2.100}$$

and in the peak power case by

$$\sum_{i=1}^{m} \left[\frac{2^{r_i}-1}{c_1 \lambda_i} \ln \left(\frac{c_1 c_2 \lambda_i r_i}{(2^{r_i}-1) R_T \mu_1} \right) \right]^+ = P_T \ . \tag{2.101}$$

Examples are given for the Rayleigh fading channel in Section 3.6.

2.5 Beamforming

An important special case of the transmission strategies discussed in the previous section is beamforming. *Transmit beamforming* is, by definition, the transmit strategy with the covariance matrix $\mathbf{R}_{cc}$ of the transmitted data *a priori* chosen to have rank one. This can be achieved by forming

$$\mathbf{c}(n) = \mathbf{w}_T s(n) \tag{2.102}$$

where $\mathbf{w}_T$ is the $n_t \times 1$ transmit *beamforming vector* or simply the *weight vector* and $s(n)$ is the transmitted symbol. It is possible to implement the beamformer in hardware using phase shifters and attenuators [69]. It can also be implemented using a fixed beamforming network (consisting of passive components) and by switching between the fixed beams. The weight vector is a constant during transmission of a block of N symbols, that is, during the time where the channel $\mathbf{H}$ is time-invariant, and then re-calculated for each block.

With (2.102), the covariance matrix becomes

$$\mathbf{R}_{cc} = E\left\{\mathbf{c}\mathbf{c}^*\right\} = \mathbf{w}_T\mathbf{w}_T^* E\left\{|s(n)|^2\right\} \tag{2.103}$$

which clearly has rank one. Assume in the following that $E\left\{|s(n)|^2\right\} = 1$ and $|\mathbf{w}_T|^2 = P_T$, the total transmitted power.

Receive beamforming is defined as a linear combination of the n_r received signals at time instant n using the $n_r \times 1$ normalized receive beamforming vector, that is $|\mathbf{w}_R| = 1$. A MIMO system with beamforming at the receiver and transmitter has the input output relation

$$z(n) = \mathbf{w}_R^*\mathbf{y}(n) = \mathbf{w}_R^*\mathbf{H}\mathbf{w}_T s(n) + \mathbf{w}_R^*\mathbf{v}(n) \ . \tag{2.104}$$

Here, $\mathbf{v}(n)$ is the received noise plus interference. The received signal to interference plus noise ratio (SINR) for the system (2.104) is

$$\text{SINR} = \frac{E\left\{|\mathbf{w}_R^*\mathbf{H}\mathbf{w}_T s(n)|^2\right\}}{E\left\{|\mathbf{w}_R^*\mathbf{v}(n)|^2\right\}} = \frac{\mathbf{w}_R^*\mathbf{H}\mathbf{w}_T\mathbf{w}_T^*\mathbf{H}^*\mathbf{w}_R}{\mathbf{w}_R^*\mathbf{R}_{vv}\mathbf{w}_R} \tag{2.105}$$

where $\mathbf{R}_{vv} \triangleq E\left\{\mathbf{v}\mathbf{v}^*\right\}$ is the covariance matrix for the interference plus noise. To maximize (2.105) with respect to $\mathbf{w}_T$, we first select the transmit weight vector $\mathbf{w}_T$ as the scaled right principal eigenvector, $\mathbf{e}_1$, to $\mathbf{H}$, ($\mathbf{H}\mathbf{e}_1 = \lambda_{max}^{1/2}\mathbf{e}_1$), to obtain

$$\text{SINR} = P_T\frac{\mathbf{w}_R^*\mathbf{e}_1\mathbf{e}_1^*\mathbf{w}_R\lambda_{max}}{\mathbf{w}_R^*\mathbf{R}_{vv}\mathbf{w}_R} \tag{2.106}$$

where λ_{max} is the principal eigenvalue to the matrix $\mathbf{HH}^*$. Equation (2.106) is a generalized Rayleigh quotient [70] and for fixed $\mathbf{w}_T$ the maximizing $\mathbf{w}_R$ is found by solving the corresponding generalized eigenvalue problem

$$\mathbf{e}_1\mathbf{e}_1^*\mathbf{w}_R = \varepsilon\mathbf{R}_{vv}\mathbf{w}_R \tag{2.107}$$

for the largest generalized eigenvalue ε. The solution is [71, Sec 3.2]

$$\mathbf{w}_R = \mathbf{R}_{vv}^{-1}\mathbf{e}_1 \tag{2.108}$$

and the corresponding SINR is

$$\text{SINR}_{max} = P_T\lambda_{max}\mathbf{e}_1^*\mathbf{R}_{vv}^{-1}\mathbf{e}_1 \tag{2.109}$$

where $\mathbf{e}_1$ is the principal eigenvector to $\mathbf{HH}^*$. In the case of spatially white noise, $\mathbf{R}_{vv} = \sigma_n^2\mathbf{I}$, the solution maximizes the receive SNR. Equation (2.106) is then a Rayleigh quotient which is maximized by choosing $\mathbf{w}_R = \alpha\mathbf{w}_T$ where α is an arbitrary constant. The resulting maximized SNR is

$$SNR_{BF} = \frac{P_T}{\sigma_n^2}\lambda_{max} \ . \tag{2.110}$$

The corresponding ergodic capacity for a MIMO system with receive and transmit beamforming is

$$C = E_{\lambda_{max}}\left\{\log_2\left(1 + \frac{P_T}{\sigma_n^2}\lambda_{max}\right)\right\} \ . \tag{2.111}$$

The receive and transmit beamforming has transformed the MIMO system to an equivalent SISO system

$$z(n) = \lambda_{max}^{1/2}s(n) + v'(n) \tag{2.112}$$

where $v'(n) = \mathbf{w}_R^*\mathbf{v}(n)$. This noise has the variance σ_n^2 in the spatially white case if $\mathbf{w}_R^*\mathbf{w}_R = 1$, since then $E\left\{|v'(n)|^2\right\} = E\left\{|\mathbf{w}_R^*\mathbf{v}(n)|^2\right\} = \sigma_n^2\mathbf{w}_R^*\mathbf{w}_R = \sigma_n^2$.

The SNR for the equivalent SISO system is given by (2.110). In Section 2.3.2, it was shown how the STBC over a MIMO channel could be interpreted as an equivalent SISO system with $\text{SNR}_{STBC} = P_T\left\|\mathbf{H}\right\|_F^2/(\sigma_n^2 n_t)$. Comparing the STBC SNR with the beamformer SNR gives

$$\begin{aligned}\frac{\text{SNR}_{BF}}{\text{SNR}_{MIMO}} =& \frac{P_T\lambda_{max}}{\sigma_n^2}\frac{\sigma_n^2 n_t}{P_T\left\|\mathbf{H}\right\|_F^2} = \frac{\lambda_{max}n_t}{\left\|\mathbf{H}\right\|_F^2} \\ =& \frac{\lambda_{max}n_t}{\lambda_1 + \ldots + \lambda_m} \geq 1\end{aligned} \tag{2.113}$$

since $m = \min(n_r, n_t) \leq n_t$. Hence, the SNR for a beamforming system is never lower than the SNR of a STBC, however, beamforming requires CSI at the transmitter.

2.5.1 Receive beamforming in SIMO Channels

When transmitting with one antenna and receiving with an array, the channel matrix can be described by the column vector $\mathbf{h}$ and the received vector is $\mathbf{y} = \mathbf{h}s(n) + \mathbf{v}$. The SINR from (2.105) is then given by

$$\text{SINR} = \frac{E\left\{|\mathbf{w}_R^*\mathbf{h}s(n)|^2\right\}}{E\left\{|\mathbf{w}_R^*\mathbf{v}(n)|^2\right\}} = \frac{\mathbf{w}_R^*\mathbf{h}\mathbf{h}^*\mathbf{w}_R}{\mathbf{w}_R^*\mathbf{R}_{vv}\mathbf{w}_R}P_T \ . \tag{2.114}$$

This expression is a generalized Rayleigh quotient which is maximized by choosing $\mathbf{w}_R$ as

$$\mathbf{w}_R = \mathbf{R}_{vv}^{-1}\mathbf{h} \ . \tag{2.115}$$

Note the equivalence with the spatial matched filter (2.31) in the maximum likelihood detector for a SIMO system.

Sometimes a *training sequence* $d(n)$ is embedded in the received signal. This can be utilized to find the receive beamforming vector $\mathbf{w}_R$ as the vector that minimizes the mean square error

$$\min_{\mathbf{w}_R} E\left\{|d(n) - \mathbf{w}_R^*\mathbf{y}(n)|^2\right\} \ . \tag{2.116}$$

The solution is the Wiener-Hopf solution and is given by

$$\mathbf{w}_R = \mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{r}_{\mathbf{y}d} \tag{2.117}$$

where the cross-correlation vector $\mathbf{r}_{\mathbf{y}d} = E\left\{\mathbf{y}(n)d^*(n)\right\}$.

Use that $\mathbf{r}_{\mathbf{y}d} = \mathbf{h}$ if $\mathbf{y}(n) = \mathbf{h}d(n) + \mathbf{v}(n)$ where $d(n)$ is assumed to be uncorrelated with $\mathbf{v}(n)$ and have unit variance. Applying Woodbury's identity[13] [71, p.39] on the matrix $\mathbf{R}_{\mathbf{yy}}$, to write

$$\mathbf{R}_{\mathbf{yy}}^{-1} = \frac{1}{1 + \mathbf{h}^*\mathbf{R}_{vv}\mathbf{h}}\mathbf{R}_{vv}^{-1} = \beta\mathbf{R}_{vv}^{-1} \ . \tag{2.118}$$

Therefore, the optimum Wiener solution can be expressed as

$$\mathbf{w}_R = \beta\mathbf{R}_{vv}^{-1}\mathbf{h} \tag{2.119}$$

which is a scaled beamforming vector in the maximum SINR solution (2.115). Hence the Wiener-Hopf and the maximum SINR solutions yield the same output SINR, a result also shown in [71, Chapt.3].

[13] Woodbury's identity states that if $\mathbf{R} = \mathbf{h}\mathbf{h}^* + \mathbf{R}_{vv}$, then $\mathbf{R}^{-1} = \mathbf{R}_{vv}^{-1}/(1 + \mathbf{h}^*\mathbf{R}_{vv}^{-1}\mathbf{h})$.

2.5.2 Fixed multibeam antennas

Fixed multibeam antennas are a viable choice for the implementation of the downlink (base-to-mobile) in wireless cellular communication systems due to the attractive trade-off between system performance and the complexity and cost of implementation [2, 72, 73]. The switched multibeam antenna has thus been the scope of several testbed and field trial investigations for its feasibility in GSM/FDMA networks [3, 6, 7, 74]. A multibeam antenna consists of an array of antenna elements and a beam-forming network that generates multiple narrow beams together with a beam switching algorithm [75] of low-complexity which points (or switches) the downlink main beam in the direction of the strongest received signal [7, 76]. In some implementations the downlink beamforming is performed entirely in hardware, thereby relaxing the signal processing requirements even further [69].

The multibeam technique is a common downlink beamforming method used in frequency division duplex (FDD) systems, because the frequency duplex distance is typically significantly larger than the coherence bandwidth of the radio channel, which makes the uplink and downlink channel uncorrelated and only partial CSI is available at the transmitter. The partial CSI consists of the direction of arrival estimate of the uplink signal which still is a useful estimate of the direction to the mobile at the downlink frequency [77].

The low side-lobes of the radiation pattern suppress the co-channel interference from users in neighboring cells. A linear array antenna with n_t antenna elements can generate n_t beams with distinct spatial orientations in the horizontal plane, by the use of the BFN.

An FFT based BFN is often implemented as a Butler matrix [14, 20, 78], see also Appendix A. It has the property of a constant phase gradient $\Delta\theta$ over the antenna array aperture, yielding

$$\theta_n = (n-1)\Delta\theta \tag{2.120}$$

for $n = 1, \ldots, n_t$, where the n:th element of the weight vector has phase θ_n and unit amplitude. The phase gradient, $\Delta\theta$ belongs to the set Ω_S of n_t (or n_r) different phase gradients, unique for each beam the BFN can generate,

$$\Omega_S = \left\{ \frac{2\pi p}{n_t} \right\}_{p=0}^{n_t-1} . \tag{2.121}$$

The finite set Ω_S is a closed group under integer multiplication and addition, following *modulo*-2π algebra, due to the phase ambiguity of 2π. This property will be important in Chapter 6 when intermodulation products from multibeam antennas are studied.

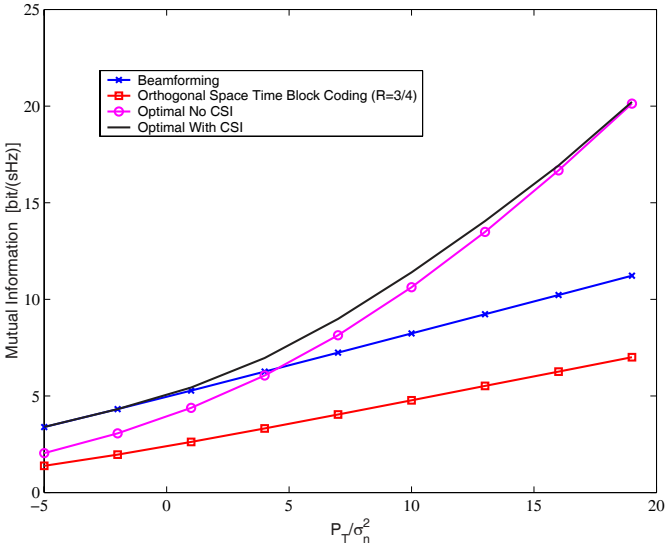
2.6 Concluding remarks

In this chapter it was shown that an important factor when choosing the transmission strategy is the relation between the eigenvalues to $\mathbf{H}\mathbf{H}^*$, which have a physical interpretation of how "directional" the channel is. As an example, two extreme cases were examined; when there is a dominating eigenvalue to the matrix $\mathbf{H}\mathbf{H}^*$ (corresponding to a channel with a LOS component) and when there is no dominating eigenvalue (NLOS case). Mutual information was discussed as it is a measure of the spectral efficiency of the wireless link. Figure 2.7 shows a comparison of the mutual information for a 4×4 MIMO system for different transmission strategies introduced in this chapter. When the channel has a dominating eigenvalue, Figure 2.7(a), beamforming is close or equal to the optimum mutual information given by the water-filling algorithm. Hence, in LOS channels, beamforming is a spectrally efficient transmission scheme, especially at low SNR. At high SNR, more channel eigenmodes can be used to fill the gap between beamforming capacity and the optimal water-filling capacity, as was demonstrated in Section 2.4.

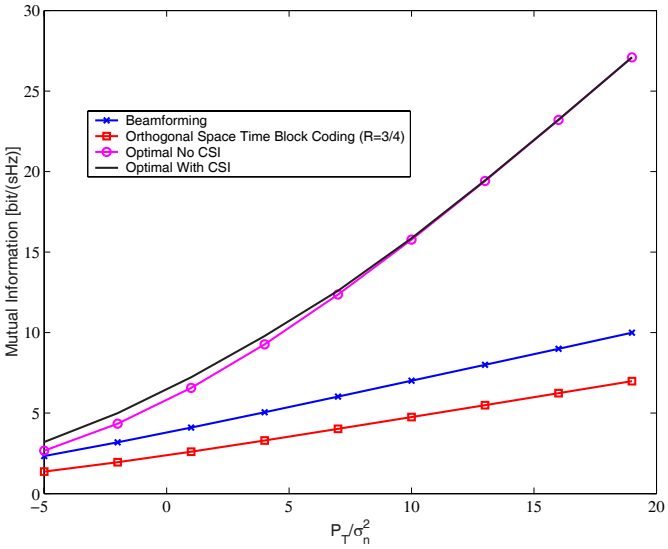
The open loop transmission scheme, used when CSI is not available at the transmitter, is clearly suboptimal, because energy is "wasted" in modes with a large attenuation. If the SNR is increased, by increasing the transmitter power P_T, then the "blind" transmission with uniformly allocated power over the channel eigenmodes is asymptotically approaching the water-filling solution and at a certain SNR it supersedes the beamforming capacity. Hence at high SNR, the gain from using CSI is small and the capacity is equal with or without CSI at the transmitter.

When the channel has rich scattering, the eigenvalues are approximately equal, and no eigenmode dominates. In this scenario, the "blind" method is expected to perform well, as it transmits an equal amount of energy in each mode, and Figure 2.7(b) shows that the "blind" transmission scheme is asymptotically close to the water-filling solution already at moderate SNR. Hence, in a rich scattering NLOS channel, close to an unitary channel, CSI at the transmitter gives only a small improvement in the mutual information. Beamforming is clearly suboptimal since it only utilizes one of the m channel modes. Only in an MISO channel, where $m = 1$ is beamforming optimal.

The optimal curves in Figure 2.7 are based on Shannon's capacity formula and gives the ultimate capacity limit. To compare these curves with a practical scheme, STBC is plotted in Figure 2.7 as well. The mutual information of the orthogonal STBC is always less than or equal to the mutual information of the beamforming due to the inequality (2.113) and the fact that the rate for orthogonal STBC $n_s/N \leq 1$. By comparing STBC (2.73) with beamforming (2.111) we see that equality occurs when $R = 1$ and all eigenvalues $\lambda_1, \ldots, \lambda_m$ are equal. This is the case in an unitary channel where $\mathbf{H}\mathbf{H}^* \propto \mathbf{I}$. In this case, the $1/n_t$ loss in power



(a) Mutual information with a dominating eigenvalue ($\lambda_1/\lambda_2 = 20$) to $\mathbf{HH}^*$. The power P_T/σ_n^2 is given in dB scale.



(b) Mutual information with approximately equal eigenvalues $\lambda_1/\lambda_2 = 1.2$ to $\mathbf{HH}^*$. The power P_T/σ_n^2 is given in dB scale.

Figure 2.7: Comparisons of the mutual information of the MIMO channel when $n_r = n_t = 4$.

of the STBC due to the unavailability of CSI at the transmitter is equal to the $1/m$ loss in power by the beamformer by only using one of the m orthogonal channels.

Although the gap between the optimal systems with and without CSI at the transmitter is small, at least for high SNR, the performance gap is in practical systems quite large [79] if the complexity should be kept low. Linear dispersion codes have the ability to reduce this gap [31], although at the expense over a highly complex optimization operation and also the requirement that the number of receive antennas must be known at the transmitter.

On the other hand, if CSI is available at the transmitter, the gap to the Shannon limit can be made small, if the transmission techniques developed in this chapter is used, together with channel coding techniques. In this view, TDD systems with no frequency duplex distance are preferable since then, full CSI is available at the transmitter. If highly broadband connections are required, then OFDM can be used together with the multiple antennas and on each OFDM subchannel, adaptive modulation is performed for the transmission over the channel eigenmodes [65]. This will lead to water-filling over both frequency and space subchannels.

A drawback with the proposed technique is that a TDD system requires accurate synchronization. This can be complicated to achieve for all mobile units and base-stations in a large geographical area. Another problems is the interference environment at the receiver, which is unknown and differ from the interference at the transmitter. This might reduce the gain of using adaptive modulation. A remedy for this is to use the array at the receiver to perform spatial interference cancellation. This is an interesting topic for further research in this field.

Appendix 2.A Derivation of the equation (2.15) and (2.17)

The difference between the test statistics is

$$\mathcal{T}(\mathcal{Y}) = \mathcal{T}_1(\mathcal{Y}) - \mathcal{T}_0(\mathcal{Y}) = \text{Re}\left((\mathcal{C}_1 - \mathcal{C}_0)^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{Y}\right) - \frac{1}{2}(\mathcal{E}_1 - \mathcal{E}_0) \quad (2.122)$$

which can, using the relations $Re(z) = (z + z^*)/2$ and $\mathcal{E}_i = \mathcal{C}_i^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H} \mathcal{C}_i$ be written as

$$\begin{aligned} \mathcal{T}(\mathcal{Y}) =& \frac{1}{2}(\mathcal{C}_1 - \mathcal{C}_0)^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{Y} \\ &+ \frac{1}{2} \mathcal{Y}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H}(\mathcal{C}_1 - \mathcal{C}_0) \\ &- \frac{1}{2}(\mathcal{C}_1^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H} \mathcal{C}_1 - \mathcal{C}_0^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H} \mathcal{C}_0) \ . \end{aligned} \quad (2.123)$$

Assume hypothesis H_0 and insert $\mathcal{Y} = \mathcal{H}\mathcal{C}_0 + \mathcal{V}$ into (2.123). This gives

$$\begin{aligned} \mathcal{T}(\mathcal{Y}|H_0) =& \frac{1}{2}(\mathcal{C}_1 - \mathcal{C}_0)^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} (\mathcal{H}\mathcal{C}_0 + \mathcal{V}) \\ &+ \frac{1}{2}(\mathcal{H}\mathcal{C}_0 + \mathcal{V})^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H}(\mathcal{C}_1 - \mathcal{C}_0) \\ &- \frac{1}{2}(\mathcal{C}_1^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H} \mathcal{C}_1 - \mathcal{C}_0^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H} \mathcal{C}_0) \ . \end{aligned} \quad (2.124)$$

Since $\mathcal{V}$ has zero mean, we get

$$\begin{aligned} E\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\} =& \frac{1}{2}(\mathcal{C}_1 - \mathcal{C}_0)^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H} \mathcal{C}_0 \\ &+ \frac{1}{2} \mathcal{C}_0^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H}(\mathcal{C}_1 - \mathcal{C}_0) \\ &- \frac{1}{2}(\mathcal{C}_1^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H} \mathcal{C}_1 - \mathcal{C}_0^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H} \mathcal{C}_0) \ . \end{aligned} \quad (2.125)$$

which can be more compactly written as

$$E\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\} = -\frac{1}{2}\left(\mathcal{C}_1 - \mathcal{C}_0\right)^* \mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H}\left(\mathcal{C}_1 - \mathcal{C}_0\right) \ . \quad (2.126)$$

Since this expression is a scalar and has the form $\mathbf{a}^* \mathbf{A} \mathbf{a}$, where $\mathbf{a}$ is a vector, we can use the property $\mathbf{a}^* \mathbf{A} \mathbf{a} = \text{Tr}(\mathbf{A} \mathbf{a} \mathbf{a}^*) = \text{Tr}(\mathbf{a} \mathbf{a}^* \mathbf{A})$ to write (2.126) as

$$E\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\} = -\frac{1}{2}\text{Tr}\left(\mathcal{O}\mathcal{H}^* \mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1} \mathcal{H}\right) \quad (2.127)$$

where $\mathcal{O} = \left(\mathcal{C}_1 - \mathcal{C}_0\right)\left(\mathcal{C}_1 - \mathcal{C}_0\right)^*$. Similarly, $E\left\{\mathcal{T}(\mathcal{Y}|H_1)\right\}$ can be found and it is seen that it differs from $E\left\{\mathcal{T}(\mathcal{Y}|H_0)\right\}$ only in sign.

To derive Var $\{\mathcal{T}(\mathcal{Y}|H_0)\}$, start with (2.124) and write

$$\begin{aligned}\text{Var}(\mathcal{T}(\mathcal{Y}|H_0)) =&\frac{1}{2}\text{Var}\left\{(\mathcal{C}_1-\mathcal{C}_0)^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}(\mathcal{H}\mathcal{C}_0+\mathcal{V})\right\}\\ &+\frac{1}{2}\text{Var}\left\{((\mathcal{H}\mathcal{C}_0+\mathcal{V})^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}(\mathcal{C}_1-\mathcal{C}_0)\right\}\\ &-\frac{1}{2}\text{Var}\left\{((\mathcal{C}_1^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\mathcal{C}_1-\mathcal{C}_0^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\mathcal{C}_0)\right\}\quad .\end{aligned} \tag{2.128}$$

Since $\mathcal{V}$ is zero mean vector, the relations $\text{Var}(\mathbf{A}\mathcal{V}) = \mathbf{A}E(\mathcal{V}\mathcal{V}^*)\mathbf{A}^*$ and $\text{Var}(\mathcal{V}^*\mathbf{A}) = \mathbf{A}^*E(\mathcal{V}\mathcal{V}^*)\mathbf{A}$ can be used together with $E(\mathcal{V}\mathcal{V}^*) = \mathcal{R}_{\mathcal{V}\mathcal{V}}$ to simplify (2.128) to

$$\begin{aligned}\text{Var}(\mathcal{T}(\mathcal{Y}|H_0)) =&\frac{1}{4}(\mathcal{C}_1-\mathcal{C}_0)^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{R}_{\mathcal{V}\mathcal{V}}\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}(\mathcal{C}_1-\mathcal{C}_0)\\ &+\frac{1}{4}(\mathcal{C}_1-\mathcal{C}_0)^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{R}_{\mathcal{V}\mathcal{V}}\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}(\mathcal{C}_1-\mathcal{C}_0)\\ =&\frac{1}{4}(\mathcal{C}_1-\mathcal{C}_0)^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}(\mathcal{C}_1-\mathcal{C}_0)\\ &+\frac{1}{4}(\mathcal{C}_1-\mathcal{C}_0)^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}(\mathcal{C}_1-\mathcal{C}_0)\\ =&\frac{1}{2}(\mathcal{C}_1-\mathcal{C}_0)^*\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}(\mathcal{C}_1-\mathcal{C}_0)\quad .\end{aligned} \tag{2.129}$$

Similarly, it can be shown that $\text{Var}(\mathcal{T}(\mathcal{Y}|H_1))$ yields the same value. Since this expression also is a scalar and has the form $\mathbf{a}^*\mathbf{A}\mathbf{a}$, where $\mathbf{a}$ is a vector, we can use the property $\mathbf{a}^*\mathbf{A}\mathbf{a} = \text{Tr}(\mathbf{a}\mathbf{a}^*\mathbf{A})$ to write (2.129) as

$$\text{Var}(\mathcal{T}(\mathcal{Y}|H_0)) = \text{Var}(\mathcal{T}(\mathcal{Y}|H_1)) = \frac{1}{2}\text{Tr}\left(\mathcal{O}\mathcal{H}^*\mathcal{R}_{\mathcal{V}\mathcal{V}}^{-1}\mathcal{H}\right) \tag{2.130}$$

where $\mathcal{O} = (\mathcal{C}_1-\mathcal{C}_0)(\mathcal{C}_1-\mathcal{C}_0)^*$.

Appendix 2.B Derivation of equation (2.63) and (2.64)

Assume that the noise is spatially and temporally white and that each symbol $s(n)$ belongs to a unitary constellation $\mathcal{X}$, that is $|s_i(n)|^2 = 1$. This implies that E_i can be removed from (2.33) as it does not affect the decisions. Assume that we transmit the n_s symbols $\{s_k(n)\}_{n=1}^{n_s}$, so the transmitted codeword is $\mathbf{C}_k$. The received matrix is $\mathbf{Y}$, where

$$\mathbf{Y} = \mathbf{H}\mathbf{C}_k + \mathbf{V}\quad . \tag{2.131}$$

We shall now analyze ML-detector for this case in detail and show that the symbol estimation problem decouples and that we can write the STBC-MIMO system as an equivalent SISO system, as seen by each transmitted symbol.

Insert the expression for linear block codes (1.23)

$$\mathbf{C}_k = \sqrt{\frac{P_T}{2n_t}} \sum_{n=1}^{n_s} [\mathbf{A}_n \bar{s}_k(n) + j\mathbf{B}_n \tilde{s}_k(n)] \tag{2.132}$$

into (2.131) and use the test statistic for hypothesis i in the ML-detector, equation (2.33), with E_i removed:

$$\begin{aligned} \mathcal{T}_i(\mathbf{Y}) =& \text{Re}\left[\text{Tr}\left(\mathbf{C}_i^* \mathbf{H}^* \mathbf{Y}\right)\right] / \sigma_n^2 \\ =& \text{Re}\left[\text{Tr}\left(\mathbf{C}_i^* \mathbf{H}^* \left(\mathbf{H}\mathbf{C}_k + \mathbf{V}\right)\right)\right] / \sigma_n^2 \ . \end{aligned} \tag{2.133}$$

Inserting (2.132) gives

$$\begin{aligned} \mathcal{T}_i(\mathbf{Y}) =& \frac{P_T}{2n_t} \sum_{n=1}^{n_s} \sum_{p=1}^{n_s} \text{ReTr}\left\{\mathbf{A}_n^* \mathbf{H}^* \mathbf{H} \mathbf{A}_p \bar{s}_k(n) \bar{s}_i(p) \right. \\ & - j\mathbf{A}_n^* \mathbf{H}^* \mathbf{H} \mathbf{B}_p \bar{s}_k(n) \tilde{s}_i(p) + j\mathbf{B}_n^* \mathbf{H}^* \mathbf{H} \mathbf{A}_p \tilde{s}_k(n) \bar{s}_i(p) \\ & \left. + \mathbf{B}_n^* \mathbf{H}^* \mathbf{H} \mathbf{B}_p \tilde{s}_k(n) \tilde{s}_i(p) \right\} \\ & + \sqrt{\frac{P_T}{2n_t}} \sum_{n=1}^{n_s} \text{ReTr}\left(\mathbf{A}_n \mathbf{H}^* \mathbf{V} \bar{s}_i(n) - j\mathbf{B}_n \mathbf{H}^* \mathbf{V} \tilde{s}_i(n)\right) \ . \end{aligned} \tag{2.134}$$

Now, we use the properties (2.65) together with $\text{Tr}(\mathbf{XY}) = \text{Tr}(\mathbf{YX})$ and $2\text{Re}(\mathbf{X}) = \mathbf{X} + \mathbf{X}^*$,$2\text{Re}(j\mathbf{X}) = \mathbf{X}^* - \mathbf{X}$ to find the useful relations

$$\begin{aligned} \text{ReTr}\left(\mathbf{A}_n^* \mathbf{H}^* \mathbf{H} \mathbf{A}_p\right) =& \text{ReTr}\left(\mathbf{H}^* \mathbf{H} \mathbf{A}_p \mathbf{A}_n^*\right) \\ =& \frac{1}{2}\text{Tr}\left(\mathbf{H}^* \mathbf{H} \mathbf{A}_p \mathbf{A}_n^* + (\mathbf{H}^* \mathbf{H} \mathbf{A}_p \mathbf{A}_n^*)^*\right) \\ =& \frac{1}{2}\text{Tr}\left(\mathbf{H}^* \mathbf{H} (\mathbf{A}_p \mathbf{A}_n^* + \mathbf{A}_n \mathbf{A}_p^*)\right) = \delta_{np} \text{Tr}\left(\mathbf{H}^* \mathbf{H}\right) \end{aligned} \tag{2.135}$$

and

$$\begin{aligned} \text{ReTr}\left(j\mathbf{A}_n^* \mathbf{H}^* \mathbf{H} \mathbf{B}_p\right) =& \text{ReTr}\left(\mathbf{H}^* \mathbf{H} \mathbf{B}_p \mathbf{A}_n^*\right) \\ =& \frac{1}{2}\text{Tr}\left((\mathbf{H}^* \mathbf{H} \mathbf{B}_p \mathbf{A}_n^*)^* - \mathbf{H}^* \mathbf{H} \mathbf{B}_p \mathbf{A}_n^*\right) \\ =& \frac{1}{2}\text{Tr}\left(\mathbf{H}^* \mathbf{H} (\mathbf{A}_n \mathbf{B}_p^* - \mathbf{B}_p \mathbf{A}_n^*)\right) = \delta_{np} \text{Tr}\left(\mathbf{H}^* \mathbf{H}\right) \end{aligned} \tag{2.136}$$

and similar results for $\text{ReTr}\left(\mathbf{B}_n^* \mathbf{H}^* \mathbf{H} \mathbf{B}_p\right)$ and $\text{ReTr}\left(\mathbf{B}_n^* \mathbf{H}^* \mathbf{H} \mathbf{A}_p\right)$. Using these,

(2.134) is simplified to

$$\begin{aligned}\mathcal{T}_i(\mathbf{Y}) =& \frac{P_T}{2n_t}\sum_{n=1}^{n_s}\{\text{Tr}(\mathbf{H}^*\mathbf{H})\bar{s}_k(n)\bar{s}_i(n) + \text{Tr}(\mathbf{H}^*\mathbf{H})\tilde{s}_k(n)\tilde{s}_i(n)\} \\ &+\sqrt{\frac{P_T}{2n_t}}\sum_{n=1}^{n_s}\text{ReTr}\left(\mathbf{A}_n\mathbf{H}^*\mathbf{V}\bar{s}_i(n) - j\mathbf{B}_n\mathbf{H}^*\mathbf{V}\tilde{s}_i(n)\right) \\ =& \frac{P_T}{2n_t}\text{Tr}(\mathbf{H}^*\mathbf{H})\sum_{n=1}^{n_s}\{\bar{s}_k(n)\bar{s}_i(n) + \tilde{s}_k(n)\tilde{s}_i(n)\} \\ &+\sqrt{\frac{P_T}{2n_t}}\sum_{n=1}^{n_s}\text{ReTr}\left(\mathbf{A}_n\mathbf{H}^*\mathbf{V}\bar{s}_i(n) - j\mathbf{B}_n\mathbf{H}^*\mathbf{V}\tilde{s}_i(n)\right) \quad .\end{aligned} \tag{2.137}$$

So, we see that the test statistic contains no cross products between symbols $\bar{s}_i(n)$ and $\bar{s}_i(p)$ and the real and complex parts are also decoupled. This is due to the properties of the underlying matrices $\mathbf{A}_n, \mathbf{B}_n$ in (2.65). Now the sequence i is picked, that is sequence $\{s_i(n)\}_{n=1}^{n_s}$, that maximizes $\mathcal{T}_i(\mathbf{Y})$. Recall that $|s_i(n)|^2 = 1$. Since it is a maximization of a sum, the product in each term must be maximized, which is obtained if $\bar{s}_k(n) = \bar{s}_i(n)$ and $\tilde{s}_k(n) = \tilde{s}_i(n)$ is chosen. Then the test statistics $\mathcal{T}_i(\mathbf{Y})$ become

$$\mathcal{T}_i(\mathbf{Y}) = \frac{P_T n_s}{2n_t}\text{Tr}(\mathbf{H}^*\mathbf{H}) + \sqrt{\frac{P_T}{2n_t}}\sum_{n=1}^{n_s}\text{ReTr}\left(\mathbf{A}_n\mathbf{H}^*\mathbf{V}\bar{s}_i(n) - j\mathbf{B}_n\mathbf{H}^*\mathbf{V}\tilde{s}_i(n)\right) \ . \tag{2.138}$$

If $\mathbf{C}_i$ in (2.133) is replaced by $\mathbf{A}_n$ then $\text{Tr}\left(\mathbf{A}_n^*\mathbf{H}^*\left(\mathbf{H}\mathbf{C}_k + \mathbf{V}\right)\right)$ can be simplified as

$$\begin{aligned}\text{ReTr}\left(\mathbf{A}_n^*\mathbf{H}^*\left(\mathbf{H}\mathbf{C}_k + \mathbf{V}\right)\right) =& \sqrt{\frac{P_T}{2n_t}}\text{Tr}\left(\mathbf{H}^*\mathbf{H}\right)\bar{s}_k(n) + \text{Tr}\left(\mathbf{A}_n\mathbf{H}^*\mathbf{V}\right) \\ =& \sqrt{\frac{P_T}{2n_t}}\left\|\mathbf{H}\right\|_F^2\bar{s}_k(n) + \text{ReTr}\left(\mathbf{A}_n\mathbf{H}^*\mathbf{V}\right)\end{aligned} \tag{2.139}$$

and if $-j\mathbf{B}_n$ is used instead, we get

$$\begin{aligned}\text{ReTr}\left(-j\mathbf{B}_n^*\mathbf{H}^*\left(\mathbf{H}\mathbf{C}_k + \mathbf{V}\right)\right) =& \sqrt{\frac{P_T}{2n_t}}\text{Tr}\left(\mathbf{H}^*\mathbf{H}\right)\bar{s}_k(n) + \text{Tr}\left(j\mathbf{B}_n\mathbf{H}^*\mathbf{V}\right) \\ =& \sqrt{\frac{P_T}{2n_t}}\left\|\mathbf{H}\right\|_F^2\bar{s}_k(n) - \text{ReTr}\left(j\mathbf{B}_n\mathbf{H}^*\mathbf{V}\right) \quad .\end{aligned} \tag{2.140}$$

So, we see that the transmitted symbol s_k is multiplied by $\sqrt{\frac{P_T}{2n_t}}\left\|\mathbf{H}\right\|_F^2$ which then can be interpreted as the gain of an equivalent SISO channel.

In the ML-detector, if $\mathbf{C}_i$ in (2.133) is replaced by $\bar{s}(n)\mathbf{A}_n$ or $j\tilde{s}(n)\mathbf{B}_n$, then an equivalent ML-detector that directly estimates $\hat{s}(n)$ is obtained. It becomes for the real part of the symbol $s(n)$

$$\hat{\bar{s}}(n) = \arg \max_{s(n)\in\mathcal{X}} \text{ReTr}\left\{\bar{s}(n)\mathbf{A}_n^*\mathbf{H}^*\mathbf{Y}\right\} \tag{2.141}$$

and for the imaginary part of $s(n)$

$$\hat{\tilde{s}}(n) = \arg \max_{s(n)\in\mathcal{X}} \text{ReTr}\left\{-j\tilde{s}(n)\mathbf{B}_n^*\mathbf{H}^*\mathbf{Y}\right\} \quad . \tag{2.142}$$

Appendix 2.C Proof of Theorem 2.1

Using the joint pdf of the eigenvalues to $\mathbf{HH}^*$, the spectral efficiency can be rewritten by using (2.76) and (2.80) as

$$\begin{aligned}\frac{R}{W} =& E\left\{\sum_{i=1}^{m} r_i\right\} = E\left\{\sum_{i=1}^{m} \log_2 M_i\right\} = \\ =& \int_0^\infty \cdots \int_0^\infty \sum_{i=1}^{m} \log_2\left(1 + K_o\gamma_i\lambda_i\right) p(\lambda_1, \ldots, \lambda_m) d\lambda_1 \cdots d\lambda_m \\ =& \sum_{i=1}^{m} \int_0^\infty \log_2\left(1 + K_o\gamma_i\lambda_i\right) p_i(\lambda_i) d\lambda_i\end{aligned} \tag{2.143}$$

where $p_i(\lambda_i)$ is the marginal pdf of the i:th largest eigenvalue. The optimization problem is to find the subchannel powers $\{\gamma_i\}_{i=1}^m$ as a function of the eigenvalues to $\mathbf{HH}^*$, $\{\lambda_i\}_{i=1}^m$ that maximizes the spectral efficiency (2.143) and meet the target BER under the average power constraint. We assume that the assigned number of bits per symbol, $\{r_i\}_{i=1}^m$, are real valued and non-negative. This can be formulated as

$$\begin{aligned}&\max_{\gamma_i} \sum_{i=1}^{m} \int_0^\infty \log_2\left(1 + K_o\gamma_i\lambda_i\right) p_i(\lambda_i) d\lambda_i \\ &\text{when} \qquad \sum_{i=1}^{m} \int_0^\infty \gamma_i p_i(\lambda_i) d\lambda_i = P_T \\ &\gamma_i \geq 0, r_i \geq 0, \qquad i = 1, \ldots, m \ .\end{aligned} \tag{2.144}$$

The optimization problem can be reformulated as an unconstrained optimization problem[14] by merging the number of bits per symbol and power through the La-

[14]This reformulation is equivalent provided that the rate is a convex function of power, which is the case here [64].

grange multiplier μ

$$\begin{aligned} J(\gamma_i, \mu) = &- \sum_{i=1}^{m} \int_0^{\infty} \log_2 \left(1 + K_o \gamma_i \lambda_i\right) p_i(\lambda_i) d\lambda_i \\ &+ \mu \left(\sum_{i=1}^{m} \int_0^{\infty} \gamma_i p_i(\lambda_i) d\lambda_i - P_T \right) \quad . \end{aligned} \tag{2.145}$$

For a fixed μ, the Lagrange cost is minimized when

$$\frac{\partial J(\gamma_i, \mu)}{\partial \gamma_i} = 0, \qquad i = 1, \ldots, m \ . \tag{2.146}$$

Differentiating (2.145), with respect to γ_k, we obtain

$$\begin{aligned} &- \frac{\partial}{\partial \gamma_k} \sum_{i=1}^{m} \int_0^{\infty} \log_2 \left(1 + K_o \gamma_i \lambda_i\right) p_i(\lambda_i) d\lambda_i \\ &+ \mu \frac{\partial}{\partial \gamma_k} \left(\sum_{i=1}^{m} \int_0^{\infty} \gamma_i p_i(\lambda_i) d\lambda_i - P_T \right) = 0 \ . \end{aligned} \tag{2.147}$$

And by exchanging the order of derivation and integration, which is allowed by Leibniz's rule since the boundaries of the integration interval are constants [80, Eq.12.211], we obtain

$$\begin{aligned} &\int_0^{\infty} \frac{\partial}{\partial \gamma_k} \left(- \log_2 \left(1 + K_o \gamma_k \lambda_k\right) + \mu \gamma_k \right) p_k(\lambda_k) d\lambda_k \\ &= \int_0^{\infty} \left(- \frac{K_o \lambda_k}{1 + K_o \lambda_k \gamma_k} + \mu \right) p_k(\lambda_k) d\lambda_k = 0 \, . \end{aligned} \tag{2.148}$$

And for this integral to be zero, for arbitrary non-zero eigenvalue distributions $p_k(\lambda_k)$, the expression in the parenthesis must be identically zero. Then the power control law $\gamma_k(\lambda_k)$ in (2.82) follows. Since power and bits per symbol by definition are non-negative, the constraints $\gamma_k \geq 0$ and $r_k \geq 0$ defines the region where (2.82) is applicable.

The constant μ is given by evaluating the power constraint criterion

$$\sum_{i=1}^{m} \int_0^{\infty} \gamma_i(\lambda_i) p_i(\lambda_i) d\lambda_i = P_T \tag{2.149}$$

with the power control law (2.82) inserted. The resulting spectral efficiency is

obtained by inserting the power control law (2.82) into (2.143)

$$\begin{aligned}\frac{R}{W} =& \sum_{i=1}^{m} \int_0^{\infty} \log_2 \left(1 + K_o \left[\frac{1}{\mu} - \frac{1}{K_o \lambda_i}\right]^+ \lambda_i\right) p_i(\lambda_i) d\lambda_i \\ =& \sum_{i=1}^{m} \int_{\mu/K_o}^{\infty} \log_2 \left(1 + \left(\frac{K_o \lambda_i}{\mu} - 1\right)\right) p_i(\lambda_i) d\lambda_i \\ =& \sum_{i=1}^{m} \int_{\mu/K_o}^{\infty} \log_2 \left(\frac{K_o \lambda_i}{\mu}\right) p_i(\lambda_i) d\lambda_i \ .\end{aligned} \tag{2.150}$$

Appendix 2.D Proof of Theorem 2.2

The optimization problem for the continuous bit per symbol allocation case is formulated as

$$\begin{aligned} &\min_{\gamma_i, r_i} \frac{1}{R_T} \sum_{i=1}^{m} \int_0^{\infty} r_i \mathrm{BER}_i(\lambda_i, \gamma_i) p_i(\lambda_i) d\lambda_i \\ \text{when} \quad &\sum_{i=1}^{m} \int_0^{\infty} \gamma_i p_i(\lambda_i) d\lambda_i = P_T \\ \text{and} \quad &\sum_{i=1}^{m} r_i = R_T \ . \end{aligned} \tag{2.151}$$

The problem can be reformulated to minimize the Lagrange cost function

$$\begin{aligned} J(r_i, \gamma_i, \mu_1, \mu_2) =& \frac{1}{R_T} \sum_{i=1}^{m} \int_0^{\infty} r_i \mathrm{BER}_i(\lambda_i, \gamma_i) p_i(\lambda_i) d\lambda_i \\ &+ \mu_1 \left\{ \sum_{i=1}^{m} \int_0^{\infty} \gamma_i p_i(\lambda_i) d\lambda_i - P_T \right\} \\ &+ \mu_2 \left\{ \sum_{i=1}^{m} r_i - R_T \right\} \end{aligned} \tag{2.152}$$

where μ_1 and μ_2 are the two Lagrange multipliers. Equation (2.152) is now minimized by fulfilling

$$\frac{\partial J}{\partial r_i} = 0 \qquad \text{and} \qquad \frac{\partial J}{\partial \gamma_i} = 0 \qquad i = 1, \ldots, m \ . \tag{2.153}$$

We begin with the first partial derivative, for a specific subchannel k

$$\begin{aligned}
&\frac{1}{R_T}\frac{\partial}{\partial r_k}\sum_{i=1}^{m}\int_0^\infty r_i \text{BER}_i(\lambda_i,\gamma_i)p_i(\lambda_i)d\lambda_i \\
&+\mu_1\left\{\frac{\partial}{\partial r_k}\sum_{i=1}^{m}\int_0^\infty \gamma_i p_i(\lambda_i)d\lambda_i - P_T\right\} \\
&+\mu_2\frac{\partial}{\partial r_k}\left\{\sum_{i=1}^{m} r_i - R_T\right\} = 0
\end{aligned} \tag{2.154}$$

by exchanging the order of the differentiation and the integration and merging the terms we obtain

$$\frac{1}{R_T}\int_0^\infty \left(\text{BER}_k(\lambda_k,\gamma_k) + r_k\frac{\partial \text{BER}_k(\lambda_k,\gamma_k)}{\partial r_k}\right)p_k(\lambda_k)d\lambda_k + \mu_2 = 0 \ . \tag{2.155}$$

and the result (2.96) follows. The second partial derivative, for a specific subchannel k, gives

$$\begin{aligned}
&\frac{1}{R_T}\frac{\partial}{\partial \gamma_k}\sum_{i=1}^{m}\int_0^\infty r_i \text{BER}_i(\lambda_i,\gamma_i)p_i(\lambda_i)d\lambda_i \\
&+\mu_1\frac{\partial}{\partial \gamma_k}\left\{\sum_{i=1}^{m}\int_0^\infty \gamma_i p_i(\lambda_i)d\lambda_i - P_T\right\} \\
&+\mu_2\frac{\partial}{\partial \gamma_k}\left\{\sum_{i=1}^{m} r_i - R_T\right\} = 0
\end{aligned} \tag{2.156}$$

which simplifies to

$$\begin{aligned}
&\frac{1}{R_T}\frac{\partial}{\partial \gamma_k}\int_0^\infty r_k \text{BER}_k(\lambda_k,\gamma_k)p_k(\lambda_k)d\lambda_k \\
&+\mu_1\frac{\partial}{\partial \gamma_k}\left\{\int_0^\infty \gamma_k p_k(\lambda_k)d\lambda_k - P_T\right\} = 0 \ .
\end{aligned} \tag{2.157}$$

Exchanging order of integration and differentiation and merging the terms gives

$$\int_0^\infty \left(\frac{r_k}{R_T}\frac{\partial \text{BER}_k(\lambda_k,\gamma_k)}{\partial \gamma_k} + \mu_1\right)p_k(\lambda_k)d\lambda_k = 0 \ . \tag{2.158}$$

For this to hold for arbitrary $p_k(\lambda_k) > 0$, the term in the parenthesis must be identically zero:

$$\frac{r_k}{R_T}\frac{\partial \text{BER}_k(\lambda_k,\gamma_k)}{\partial \gamma_k} + \mu_1 = 0 \qquad k = 1,\ldots,m \ . \tag{2.159}$$

Using this expression together with (2.78), and the fact that $r_i = \log_2 M_i$, the power control law becomes

$$\gamma_i(\lambda_i, r_i) = \left[\frac{2^{r_i} - 1}{c_1 \lambda_i} \ln \left(\frac{c_1 c_2 \lambda_i r_i}{(2^{r_i} - 1) R_T \mu_1} \right) \right]^+ \tag{2.160}$$

where $c_1 = 1.6/\sigma_n^2$ and $c_2 = 0.2$. The Lagrange multipliers can be solved for by using the power and bit rate constraints in (2.151) with (2.160) inserted.

Chapter 3

MIMO system performance

MIMO systems in flat fading channels are in this chapter analyzed and compared, with and without partial CSI at the transmitter, under the assumption of different fading statistics. It will be assumed that the signals from different receive and/or transmit antennas are correlated. Mutual coupling is also introduced between the antenna elements. Measurement results from a MIMO testbed is presented and it is verified that the i.i.d. Rayleigh fading assumption is valid in some cases. Finally, the gain from using feedback information about the channel state is discussed.

3.1 Introduction

Accurate and tractable channel modelling is critical to realize the full potential of antenna arrays. The chapter begins with a review of the channel models used in this thesis. Two subgroups of channel models for MIMO systems can be identified. One is based on the physical properties of the channel, and is a parametric model where the parameters are; the number of scatterers, angle of arrival of the signal from the scatterer, time delay, and power decay profiles. These parameters are often modelled as random variables from a given distribution. Such models can adapt to the multipath environment and the array geometry and its flexibility makes them attractive for computer simulations. For analytical derivations, however, a reductionistic approach is taken, where random matrix theory is used to define the channel with fewer parameters than in the physical model, although at the expense of lower accuracy in capacity prediction [81].

The chapter is organized as follows, in Section 3.2, the existing SISO channel models based on probability distribution functions of the channel gain coefficient

are reviewed and the Rayleigh fading model is then extended to the MIMO case. Section 3.3 introduces the physical channel models used in this thesis and in Section 3.4, results from MIMO channel measurements made in an office location at Uppsala University are presented. These measurements are used to validate the assumptions made in Section 3.2. Section 3.5 analyzes the performance of orthogonal STBC in a Nakagami-m fading channel and in Section 3.6, the performance of the eigenmode transmission method derived in Section 2.4 is investigated. This method requires full CSI at the transmitter which may not be the case if the channel coherence time is short and the feedback channel has low bandwidth. In Section 3.7 the performance of beamforming and orthogonal STBC is compared when the channel coherence time is varied and when the feedback channel has different bandwidths. Finally, the chapter is concluded in Section 3.8.

3.2 Stochastic channel models

The propagation scenario in a wireless communication system is very complex and the signal transmitted from an antenna will reach the receiving antenna after many path reflections. If the scattering is rich enough, then a stochastic method is suitable to model the channel. The well known stochastic models for the SISO channel are in this section extended to the MIMO channel. The elements of the $\mathbf{H}$ matrix are assumed to be random variables taken from a probability distribution function (pdf). The elements are in general correlated but sometimes the correlation is neglected in an initial analysis to make it tractable. The correlation depends on the scattering scenario, the antenna element radiation patterns, their configuration and separation distance. It is modelled using a general correlation model, where the covariance matrix of the channel matrix elements is defined. However, in the definition of the Rayleigh fading MIMO channel, a special structure will be imposed on the correlation matrix. This structure makes further analysis of the Rayleigh fading MIMO case possible, since the joint pdf of the eigenvalues to $\mathbf{H}\mathbf{H}^*$ for this case is known.

The phase of the channel matrix elements are assumed to be uniformly distributed between $[-\pi, \pi[$ and the amplitude is assumed to have a Nakagami-m, Rayleigh or Ricean fading distribution.

Let the elements of the channel matrix $\mathbf{H}$ be expressed as

$$H_{pq} = X_{pq} + jY_{pq} = A_{pq}e^{j\phi_{pq}} \qquad p = 1, \ldots, n_r \; q = 1, \ldots, n_t \tag{3.1}$$

in Cartesian and polar coordinates respectively.

3.2.1 The Nakagami-m fading distribution

The amplitude A_{pq} is Nakagami-m distributed if the probability density function of A_{pq} is

$$p_{A_{pq}}(A_{pq}) = \frac{2}{\Gamma(m_f)} \left(\frac{m_f}{\Omega}\right)^{m_f} A_{pq}^{2m_f-1} e^{-m_f A_{pq}^2/\Omega}, \qquad A_{pq} \geq 0 \tag{3.2}$$

where $\Omega = E\left\{A_{pq}^2\right\}$ is the second order moment (average power)[82]. The letter m in the name "Nakagami-m" is the parameter of the distribution called the fading figure. In this thesis we denote the fading figure as m_f and it is equal to

$$m_f = \frac{\Omega^2}{E\left\{\left(A_{pq}^2 - \Omega\right)^2\right\}}, \qquad m_f \geq \frac{1}{2} \ . \tag{3.3}$$

If m_f is an integer, then for m_f and $m_f/2$, the Nakagami random variable is the square root of the sum of squares of $2m_f$ independent Gaussian random variables [83]. The Nakagami distribution covers a wide range of distributions, when $m_f = 1/2$ it is the one-sided Gaussian distribution, if $m_f = 1$ then it is equal to the Rayleigh distribution and when m_f approaches infinity the channel becomes static and the pdf becomes an impulse function located at $\sqrt{\Omega}$.

The marginal pdf of the squared amplitude $\gamma_{pq} = A_{pq}^2$ of a Nakagami-m distributed variable is Gamma distributed with $2m_f$ degrees of freedom since it is composed of $2m_f$ independent Gaussian distributed random variables. In the case $\Omega = 1$

$$p_{\gamma_{pq}}(\gamma_{pq}) = \frac{1}{\Gamma(m_f)} \gamma_{pq}^{m_f-1} e^{-\gamma_{pq}} \ . \tag{3.4}$$

Nakagami-m fading MIMO channels

The Nakagami-m SISO channel is now extended to the MIMO case. Although the Nakagami-m distribution was originally empirically found, it has a physical interpretation as follows (from [83]). In a mobile radio environment, it is reasonable to assume that the received signal from transmit antenna q to receive antenna p is superimposed by a large number of multipath signals. All component signals can be treated as m_f independent groups and in each of these groups, there are $\eta_j, j = 1, \ldots, m_f$ unresolvable "subpath" signals $r_{pq}^{(j)}$ which have almost identical phase and amplitude. The sum of the "subpath" signals in each group forms the j:th resolved multipath signal

$$r_{pq}^{(j)} = \sum_{\{\eta_j\}} r_{pq}^{j,\eta_j} = A_{pq}^{(j)} e^{j\xi_{pq}^{(j)}} = X_{pq}^{(j)} + jY_{pq}^{(j)}, \qquad j = 1, \ldots, m_f \ . \tag{3.5}$$

Assume that the set of η_j is large, and that no line of sight (NLOS) between the receiver and transmitter exists. This NLOS assumption ensures that $r_{pq}^{(j)}$ has zero mean. By invoking the central limit theorem, both $X_{pq}^{(j)}$ and $Y_{pq}^{(j)}$ can be approximated by independent Gaussian distributed random variables with zero mean and with variance σ_H^2. We normalize the variance to $\sigma_H^2 = 1/2$ with no loss of generality. Thus, the amplitude $A_{pq}^{(j)}$ is Rayleigh distributed. The channel amplitude defined by the index pq can then be written as

$$A_{pq} = \sqrt{\sum_{j=1}^{m_f} |X_{pq}^{(j)} + jY_{pq}^{(j)}|^2} \stackrel{\triangle}{=} |\mathbf{r}_{pq}| \tag{3.6}$$

where $\mathbf{r}_{pq} = \left[r_{pq}^{(1)}, \ldots, r_{pq}^{(m_f)}\right]^T$. Hence, each element of this m_f-dimensional random vector is independent of the others and has a Rayleigh distributed envelope. So, when defining the correlation between different channels in the MIMO system (different values on p and q), it is only necessary to model the correlation between Gaussian component waves.

We are now ready to define the correlated Nakagami-m fading MIMO channel.

Definition 3.1 *Assume that the $n_r \times n_t$ channel matrix $\mathbf{H}$ has elements with Nakagami-m distributed amplitudes with fading figure m_f and independent and uniformly distributed phases on $[-\pi, \pi[$. The covariance matrix*

$$\mathbf{R}_H = E\left[vec(\mathbf{H})vec(\mathbf{H})^*\right] \quad . \tag{3.7}$$

calculated for $m_f = 1$, i.e. the Rayleigh fading case, together with the assumptions above defines the correlated Nakagami-m fading MIMO channel.

3.2.2 The Rayleigh and Ricean fading distributions

In the Rayleigh fading channel, the amplitude A_{pq} has a Rayleigh distribution

$$p_{A_{pq}}(A_{pq}) = \frac{A_{pq}}{\Omega} e^{-A_{pq}^2/2\Omega} \qquad A_{pq} \geq 0 \tag{3.8}$$

and if the amplitude A_{pq} has non-zero mean, then it has a Ricean distribution

$$p_{A_{pq}}(A_{pq}) = \frac{A_{pq}}{\Omega} e^{-(A_{pq}^2 + A_0^2)/2\Omega} I_0\left(\frac{A_{pq}A_0}{\Omega}\right), \qquad A_{pq} \geq 0 \tag{3.9}$$

where A_0 is the mean [84] and $I_0(\cdot)$ is the 0:th order modified Bessel function of the first kind. Ricean fading is a commonly used fading model for LOS channels. The

Ricean pdf is often characterized by the ratio of the power in the direct component (LOS) and the diffuse component as

$$K = \frac{A_0^2}{2\Omega} \ . \tag{3.10}$$

When $K = 0$, the Ricean distribution becomes Rayleigh. Note also that the Nakagami-m distribution can well be approximated by a Ricean distribution, especially for high values of the fading parameter m_f.

Rayleigh fading MIMO channels

The correlated Rayleigh fading MIMO channel is obtained from Definition 3.1 with the parameter m_f=1. However, a few more assumptions are needed regarding the channel covariance matrix $\mathbf{R}_H$ to enable an analysis of the Rayleigh fading case. These assumptions are:

Assumption 3.1 *The correlation between signals transmitted from antennas p and q and received by antenna k is independent of k.*

Assumption 3.2 *The correlation between signals received by antennas j and k transmitted from antenna p is independent of p.*

The assumptions 3.1 and 3.2 are quite accurate when the antenna elements are co-located in the same physical unit at the transmitter and receiver (i.e. a mobile phone unit or a basestation) [85]. Before proceeding, the following definition is made.

Definition 3.2 *The matrix $\mathbf{H}_w$ is defined as an $n_r \times n_t$ matrix with independent and identically distributed elements. Each element has zero mean and complex circular Gaussian distribution with variance $1/2$ per complex dimension.*

The matrix $\mathbf{H}_w$ is used as a hypothetical Rayleigh fading MIMO channel matrix where all the underlying subchannels (matrix elements) are independent and have the same fading statistics. In a rich scattering environment, where the antenna elements are sufficiently separated, the channel matrix can sometimes be approximated as $\mathbf{H}_w$. As will be seen later, under the assumption that the channel is $\mathbf{H}_w$, it is possible to derive several theoretical results for transmission over this channel, such as the bit error rate. The channel $\mathbf{H}_w$ can also be used to construct a new random channel matrix, with correlated matrix elements. Hence, we make the following definition [85, 86].

Definition 3.3 *Under the assumptions 3.1 and 3.2, the correlated Rayleigh fading MIMO channel is defined as the matrix $\mathbf{H}_r$, where $\mathbf{H}_r$ is given by*

$$\mathbf{H}_r \stackrel{\triangle}{=} \mathbf{C}_R \mathbf{H}_w \mathbf{C}_T \ . \tag{3.11}$$

The matrices $\mathbf{C}_T$ and $\mathbf{C}_R$ are of dimension $n_t \times n_t$, $n_r \times n_r$ respectively that models spatial fading correlation. The matrix $\mathbf{H}_w$ is given by Definition 3.2.

The model (3.11) is general for introducing correlation between elements in a random matrix. Hence, it can also be used to introduce mutual coupling between array antenna elements. In this case, the channel matrix $\mathbf{H}_w$ between the coupling matrices is not necessarily a "white" matrix according to Definition 3.2. Rather it can be a matrix that already contains the spatial fading correlation, for instance $\mathbf{H}_r$. Then (3.11) is used to add mutual coupling to the analysis by using $\mathbf{C}_R$ and $\mathbf{C}_T$ as mutual coupling matrices for the receiver and transmitter side arrays respectively. This approach is useful when a channel matrix is given from e.g. channel simulations and the additional impact from mutual coupling between antenna element is studied. This approach will be used in Section 3.5.1. For a more thorough analysis of mutual coupling in MIMO systems, see [87].

An alternative description is to use the transmit and receive covariance matrices with elements defined as

$$(\mathbf{R}_T)_{p,q} = E\left\{H_{k,p}H_{k,q}^*\right\} \tag{3.12}$$

$$(\mathbf{R}_R)_{l,m} = E\left\{H_{l,p}H_{m,p}^*\right\} \quad . \tag{3.13}$$

The matrices $\mathbf{C}_T$ and $\mathbf{C}_R$ are related to $\mathbf{R}_T$ and $\mathbf{R}_R$ as [88]

$$\left(\mathbf{R}_T^{1/2}\right)^T = \mathbf{C}_T \tag{3.14}$$

$$\mathbf{R}_R^{1/2} = \mathbf{C}_R \tag{3.15}$$

where for any matrix $\mathbf{R}$, $\mathbf{R}^{1/2}$ fulfills $\mathbf{R} = \mathbf{R}^{1/2}(\mathbf{R}^{1/2})^*$. The relation (3.11) implies that the covariance matrix of vec$(\mathbf{H})$ can be written as

$$\mathbf{R}_H \stackrel{\triangle}{=} E\left\{\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^*\right\} = \mathbf{R}_T \otimes \mathbf{R}_R \quad . \tag{3.16}$$

It can be shown that the properties (3.16) and (3.11) are equivalent in the sense that the channel matrix $\mathbf{H}$ has the same statistical properties [88].

In [89], the structure (3.16) was verified using measurements on an indoor channel at 5.2 GHz, using a least square rank one approximation method to minimize the Frobenius norm of the difference between $\mathbf{R}_H$ and $\mathbf{R}_T \otimes \mathbf{R}_R$. The error was less than 5% for a 3×3 MIMO system in an NLOS channel.

The matrix product $\mathbf{H}\mathbf{H}^*$ and its eigenvalues are of great interest for performance calculations in MIMO systems. An example where this matrix product occurs is the capacity derivations in Section 2.2.1. Results from random matrix theory is now used to derive probability density functions for the eigenvalues of

$\mathbf{HH}^*$. Under certain conditions in the Rayleigh and Ricean fading case, $\mathbf{HH}^*$ has a Wishart distribution [43]. These conditions are reviewed here for the Rayleigh fading case.

Define the matrix $\mathbf{W} \triangleq \mathbf{H}_r\mathbf{H}_r^*$ when $n_t \leq n_r$ and $\mathbf{W} \triangleq \mathbf{H}_r^*\mathbf{H}_r$ otherwise, where $\mathbf{H}_r$ is a correlated Rayleigh fading MIMO channel matrix given in Definition 3.3. In the following analysis, assume $n_t \leq n_r$ and using (3.11), write the matrix $\mathbf{H}_r\mathbf{H}_r^*$ as

$$\mathbf{W} \triangleq \mathbf{H}_r\mathbf{H}_r^* = \mathbf{C}_R\mathbf{H}_w\mathbf{C}_T\mathbf{C}_T^*\mathbf{H}_w^*\mathbf{C}_R^* \,. \tag{3.17}$$

If $\mathbf{C}_T\mathbf{C}_T^*$ is idempotent[1], then $\mathbf{W}$ is known to be *complex Wishart distributed* [39, 43]. Real Wishart matrices can also be defined, but in this application, we are more interested in the complex case and henceforth we shall refer to a complex Wishart matrix as simply Wishart. The requirement that $\mathbf{C}_T\mathbf{C}_T^*$ is idempotent is a restriction, but to be able to proceed with the analysis, this assumption must unfortunately be made. Proceed by assuming that $\mathbf{C}_T\mathbf{C}_T^*$ is idempotent and to simplify the analysis even further, we assume that[2] $\mathbf{C}_T = \mathbf{I}$, that is, the transmitter antennas is sufficiently spaced apart so the mutual coupling and the spatial correlation induced at the transmitter side can be neglected[3]. Under this assumption, the following definition is made [43]:

Definition 3.4 *If the $n_r \times n_r$ matrix is defined as $\mathbf{W} = \mathbf{H}_r\mathbf{H}_r^*$, where $\mathbf{H}_r$ is a correlated Rayleigh fading MIMO channel matrix given by Definition 3.3 with $\mathbf{C}_T = \mathbf{I}$, that is, with covariance matrix*

$$E\left\{vec(\mathbf{H}_r)vec(\mathbf{H}_r)^*\right\} = \mathbf{I}_{n_t} \otimes \mathbf{C}_R\mathbf{C}_R^* \tag{3.18}$$

and if $n_t \leq n_r$, then $\mathbf{W}$ is Wishart distributed with n_t degrees of freedom and covariance matrix $\mathbf{C}_R\mathbf{C}_R^*$. This is written as

$$\mathbf{W} \sim \mathcal{W}_{n_r}\left(n_t, \mathbf{C}_R\mathbf{C}_R^*\right) \tag{3.19}$$

where the subscript n_r denotes the size of the matrix $\mathbf{W}$.

Random matrix theory now gives the joint pdf of the ordered eigenvalues to a Wishart distributed matrix. The matrix $\mathbf{W}$ is Hermitian and self adjoint, hence its eigenvalues are non-negative and real. Since they are ordered, assume $\lambda_1 \geq \lambda_2 \geq$

[1] A matrix $\mathbf{A}$ is idempotent if $\mathbf{AA} = \mathbf{A}$

[2] This is a stricter assumption than $\mathbf{C}_T$ is idempotent.

[3] Note that if $n_t > n_r$ then the corresponding requirement is that the antenna elements at the receiver side should be sufficiently spaced to make $\mathbf{C}_R = \mathbf{I}$. Hence, the requirement applies to the side which has the fewest number of antennas.

$\cdots \geq \lambda_m$. The eigenvalues to a Wishart matrix plays an important role in principal component analysis and other multivariate techniques. The joint probability density function is [39]

$$p(\lambda_1, \ldots, \lambda_m) = \frac{K_{m,n}}{\det(\mathbf{C}_R \mathbf{C}_R^*)^n} {}_0F_0(-(\mathbf{C}_R \mathbf{C}_R^*)^{-1}, \mathbf{W}) \prod_{i=1}^{m} \lambda_i^{n-m} \prod_{i<j}^{m} (\lambda_i - \lambda_j)^2 \tag{3.20}$$

where

$$m \triangleq \min(n_r, n_t) \tag{3.21}$$

and

$$n \triangleq \max(n_r, n_t) \tag{3.22}$$

and $K_{m,n}$ is a scalar constant that depends on n and m

$$K_{m,n} \triangleq \frac{\pi^{m(m-1)}}{\Gamma_m(n)\Gamma_m(m)} \tag{3.23}$$

and where

$$\Gamma_m(a) = \pi^{m(m-1)/2} \prod_{i=1}^{m} \Gamma(a - i + 1) \tag{3.24}$$

is the complex multivariate gamma function. A useful property of Wishart matrices in this context is also [90]

$$E\left\{\det\left(\mathbf{I} + \psi \mathbf{W}\right)\right\} = m! \psi^m L_m^{n-m}(-1/\psi) \tag{3.25}$$

where L_k^{α} is the generalized Laguerre polynomial of order k [80, Eq.8.970.1].

The function ${}_0F_0(-(\mathbf{C}_R \mathbf{C}_R^*)^{-1}, \mathbf{W})$ is the hypergeometric function with matrix arguments [43]. Its computation is in general very difficult and involves a series expansion in zonal polynomials [43], which converges slowly. Fortunately, the special case of i.i.d. circularly complex Gaussian distributed elements of $\mathbf{H}$, i.e. $\mathbf{C}_R = \mathbf{I}$ or equivalently when the channel is $\mathbf{H} = \mathbf{H}_w$, gives a more simple expression for the ordered eigenvalues distribution as

$$p(\lambda_1, \ldots, \lambda_m) = K_{m,n} \prod_{i=1}^{m} e^{-\lambda_i} \lambda_i^{n-m} \prod_{i<j}^{m} (\lambda_i - \lambda_j)^2 \tag{3.26}$$

which follows from the property ${}_0F_0(-\mathbf{I}, \mathbf{W}) = \exp\left(-\mathrm{Tr}(\mathbf{W})\right)$ [39].

Before proceeding, this section is summarized. The Definition 3.3 of the correlated Rayleigh fading MIMO channel, which is a restriction on the correlation between the matrix elements that is equivalent to the covariance structure (3.16), makes the product $\mathbf{W} = \mathbf{H}\mathbf{H}^*$ Wishart distributed, if $\mathbf{C}_T\mathbf{C}_T^*$ is idempotent. For a Wishart distributed matrix, the joint pdf of the eigenvalues is known, equation (3.20), but is difficult to use, due to the involvement of the hypergeometric function with matrix arguments. For the case of a "white" channel matrix, $\mathbf{H}_w$, defined in Definition 3.2, the joint pdf of the eigenvalues is simplified to expression (3.26), which will be utilized in the following analysis.

The marginal pdf of λ_i

It is possible in the case of i.i.d. Rayleigh distributed elements of $\mathbf{H}$, that is $\mathbf{H} = \mathbf{H}_w$, to find the marginal distribution of any of the $m = \min(n_r, n_t)$ eigenvalues to the matrix $\mathbf{W}$. These pdf:s will be useful in later sections. We now present the following theorem.

Theorem 3.1 *Assume that* $\mathbf{W}$ *is Wishart distributed according to Definition 3.4 with* $\mathbf{C}_R\mathbf{C}_R^* = \mathbf{I}$*, that is,*

$$\mathbf{W} \sim \mathcal{W}_{n_r}(n_t, \mathbf{I}_{n_r}) \quad . \tag{3.27}$$

The marginal probability density function of the k*:th largest eigenvalue to* $\mathbf{W}$ *is then given by*

$$\begin{aligned} p_k(\lambda_k) = \mathcal{Q} \int_{\lambda_k}^{\infty} d\lambda_{k-1} \int_{\lambda_{k-1}}^{\infty} d\lambda_{k-2} \cdots \\ \cdots \int_{\lambda_3}^{\infty} d\lambda_2 \int_{\lambda_2}^{\infty} d\lambda_1 \int_{0}^{\lambda_k} d\lambda_{k+1} \cdots \int_{0}^{\lambda_{m-1}} d\lambda_m p(\lambda_1, \ldots, \lambda_m) \end{aligned} \tag{3.28}$$

where $\mathcal{Q}$ *is a constant that normalizes the pdf to unity, if desired, and* $p(\lambda_1, \ldots, \lambda_m)$ *is given by* (3.26)*.*

Proof: The joint pdf of the eigenvalues is given by (3.26). Since the eigenvalues are ordered, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$, the marginal pdf of an arbitrary eigenvalue λ_k is found by integrating out the other $m-1$ eigenvalues from (3.26) one by one in a certain order. Start by integrating out the smallest eigenvalue, λ_m. Since $0 \leq \lambda_m \leq \lambda_{m-1}$, these are the integration limits that should be used to find the marginal pdf $p(\lambda_1, \ldots, \lambda_{m-1})$. Hence

$$p(\lambda_1, \ldots, \lambda_{m-1}) = \int_0^{\lambda_{m-1}} p(\lambda_1, \ldots, \lambda_m) d\lambda_m \quad . \tag{3.29}$$

Now, since λ_m has been integrated out, the smallest eigenvalue is λ_{m-1} which is bounded by $0 \leq \lambda_{m-1} \leq \lambda_{m-2}$. Hence the joint pdf of the $m-2$ largest eigenvalues is found from

$$p(\lambda_1, \ldots, \lambda_{m-2}) = \int_0^{\lambda_{m-2}} p(\lambda_1, \ldots, \lambda_{m-1}) d\lambda_{m-1} \ . \tag{3.30}$$

This procedure is iterated until we reach the joint pdf of the k largest eigenvalues, $p(\lambda_1, \ldots, \lambda_k)$. Now, we integrate out the largest eigenvalue, which is bounded as $\lambda_2 \leq \lambda_1 < \infty$, to find

$$p(\lambda_2, \ldots, \lambda_k) = \int_{\lambda_2}^{\infty} p(\lambda_1, \ldots, \lambda_k) d\lambda_1 \tag{3.31}$$

and then proceed iteratively until only the marginal pdf for the eigenvalue $p_k(\lambda_k)$ remains. This iterative manner for finding $p_k(\lambda_k)$ can be written compactly as (3.28). Sometimes it is desirable to normalize the marginal pdf to have the property $\int_0^\infty p_k(\lambda_k) d\lambda_k = 1$, so the normalization constant Q is introduced in (3.28) for this purpose. ■

We demonstrate this theorem by an example.

EXAMPLE 3.1

Assume that the 3×3 MIMO channel is given by $\mathbf{H}_w$ in Definition 3.2 on page 81. Then $\mathbf{W} = \mathbf{H}_w \mathbf{H}_w^*$ has a Wishart distribution according to Definition 3.4 and the joint pdf of the eigenvalues is given by Equation (3.26) with $m = n = 3$ as

$$p(\lambda_1, \lambda_2, \lambda_3) = \frac{1}{4} e^{-\lambda_1 - \lambda_2 - \lambda_3} (\lambda_1 - \lambda_2)^2 (\lambda_2 - \lambda_3)^2 (\lambda_3 - \lambda_1)^2 \ . \tag{3.32}$$

To find the marginal pdf of the smallest eigenvalue, we use Theorem 3.1 and integrate out the two largest as

$$p_3(\lambda_3) = \int_{\lambda_3}^{\infty} \left\{ \int_{\lambda_2}^{\infty} p(\lambda_1, \lambda_2, \lambda_3) d\lambda_1 \right\} d\lambda_2 = 3e^{-3\lambda_3} \tag{3.33}$$

The middle eigenvalue has the marginal pdf

$$\begin{aligned} p_2(\lambda_2) &= \int_{\lambda_2}^{\infty} \left\{ \int_0^{\lambda_2} p(\lambda_1, \lambda_2, \lambda_3) d\lambda_1 \right\} d\lambda_3 \\ &= \frac{1}{2} (12 - 12\lambda_2 + 6\lambda_2^2 + 2\lambda_2^3 + \lambda_2^4) e^{-2\lambda_2} - 12 e^{-3\lambda_2} \end{aligned} \tag{3.34}$$

and the largest eigenvalue

$$\begin{aligned} p_1(\lambda_1) = &\int_0^{\lambda_1} \left\{ \int_0^{\lambda_2} p(\lambda_1, \lambda_2, \lambda_3) d\lambda_3 \right\} d\lambda_2 \\ =&\frac{1}{4}(12 - 24\lambda_1 + 24\lambda_1^2 - 8\lambda_1^3 + \lambda_1^4)e^{-\lambda_1} \\ -&\frac{1}{2}(12 - 12\lambda_1 + 6\lambda_1^2 + 2\lambda_1^3 + \lambda_1^4)e^{-2\lambda_1} + 3e^{-3\lambda_1} \quad . \end{aligned} \tag{3.35}$$

The pdf for the largest eigenvalue of a Wishart matrix is related to the maximum achievable path gain in a beamforming system and to emphasize its importance we state that the pdf has a particular form in the following corollary

Corollary 3.1 *Assume that* $\mathbf{W}$ *is Wishart distributed according to Definition 3.4 with* $\mathbf{C}_R\mathbf{C}_R^* = \mathbf{I}$*. That is*

$$\mathbf{W} \sim \mathcal{W}_{n_r}(n_t, \mathbf{I}_{n_r}) \quad . \tag{3.36}$$

The marginal probability density function of the largest eigenvalue to $\mathbf{W}$ *has then the form*

$$p_\lambda(\lambda_{max}) = \sum_{k=1}^{m} \phi_k(\lambda_{max})\, e^{-k\lambda_{max}} \quad , \tag{3.37}$$

where, $m \triangleq \min(n_r, n_t)$ *and the functions* $\phi_k(x)$ *are polynomials. The polynomials for* $m = 2, 3, 4$ *and* $n \triangleq \max(n_r, n_t) = 2, 3, 4$ *are given in Table 3.1.*

Proof: See Appendix 3.A on page 144. ■

Conjecture 3.1 *The polynomials* $\phi_k(x)$ *are of maximum degree* $n_r + n_t - 2$*.*

This conjecture is discussed in Appendix 3.A and it is not included in the corollary since it is not formally proven. Furthermore, we have not been able to find a way to express the polynomials $\phi_k(x)$ explicitly, given arbitrary n_t and n_r, without having to resort to the $m - 1$-fold integration. However, we believe that this is possible and it is a topic for future research.

The largest eigenvalue of a matrix is called the spectral radius of the matrix and it is bounded by

$$\frac{1}{m}\sum_{i=1}^{m} \lambda_i \leq \lambda_{max} \leq \sum_{i=1}^{m} \lambda_i = \mathrm{Tr}(\mathbf{W}) \tag{3.38}$$

where m is the rank of $\mathbf{W}$, denoted $\mathrm{rk}(\mathbf{W})$. The upper bound is attained when $\mathrm{rk}(\mathbf{W}) = 1$ and the lower bound when all eigenvalues λ_i are identical. This property will be used later in the thesis.

	n=2	n=3	n=4
m=2	$\phi_1(x) = 2 - 2x + x^2$ $\phi_2(x) = -2$	$\phi_1(x) = 3x - 2x^2 + x^3/2$ $\phi_2(x) = -3x - x^2$	$\phi_1(x) = 2x^2 - x^3 + 1/6x^4$ $\phi_2(x) = -2x^2 - x^3 - 1/6x^4$
m=3	$-$	$\phi_1(x) = 3 - 6x + 6x^2 - 2x^3 + x^4/4$ $\phi_2(x) = -6 + 6x - 3x^2 - x^3 - x^4/2$ $\phi_3(x) = 3$	$\phi_1(x) = 6x - 8x^2 +$ $+ 9/2x^3 - x^4 + 1/12x^5$ $\phi_2(x) = -12x + 4x^2 + x^3 -$ $- x^4 - 1/12x^5 - 1/12x^6$ $\phi_3(x) = 6x + 4x^2 + 1/2x^3$
m=4	$-$	$-$	$\phi_1(x) = 4 - 12x + 18x^2 - 34/3x^3 + 7/2x^4 -$ $-1/2x^5 + 1/36x^6$ $\phi_2(x) = -12 + 24x - 24x^2 + 8/3x^3 - 4/3x^4 +$ $+ 4/3x^5 - 4/9x^6 + 1/18x^7 - x^8/72$ $\phi_3(x) = 12 - 12x + 6x^2 + 14/3x^3 +$ $+ 23/6x^4 + 5/6x^5 + 1/12x^6$ $\phi_4(x) = -4$

Table 3.1: The table shows the polynomials in the expression (3.37) *in Corollary 3.1 for the pdf:s of the largest eigenvalue* λ_{max} *to a MIMO channel matrix with* n_r *receive antennas and* n_t *transmit antennas. In the table,* $m = \min(n_r, n_t)$ *and* $n = \max(n_r, n_t)$.

To calculate the polynomials $\phi_k(x)$ for large m and n is difficult. Fortunately, as the matrix dimensions grow large, the pdf for the eigenvalues converges, and especially the extreme eigenvalues λ_{max} and λ_{min} converge to non-random quantities. If n and m increase without bound such that $m/n \to \beta \leq 1$, then it can be shown that [91–93]

$$\begin{aligned} \lambda_{max} &\to (1+\sqrt{\beta})^2 \\ \lambda_{min} &\to (1-\sqrt{\beta})^2 \quad . \end{aligned} \tag{3.39}$$

These relations will be used in this thesis for finding the asymptotic value of the channel capacity and bounds for bit error rates of Rayleigh fading MIMO channels.

3.3 Physical channel models

To generate signal correlation matrices $\mathbf{C}_R$ and $\mathbf{C}_T$, a model that resembles the physical scattering in the channel can be used. Many different approaches to model signal correlation, or directly the channel matrix for MIMO systems have been presented. The Saleh-Valenzuela model [94] is partially based on channel physics and characterizes the channel by representing the multipath components in terms of their amplitude, arrival time and angle of arrival/departure (AOA/AOD) and with specified pdf:s for the AOA/AOD, amplitudes and arrival times [95]. In this thesis, the one-ring model is used to generate correlation matrices and a scattering disc model is used to obtain channel matrices directly.

3.3.1 Scattering disc model

In [96], a simple model based on a circular disc of uniformly distributed scatterers placed around the receiving antenna array was presented. This model was then extended to a MIMO scenario with general antenna radiation functions in [21]. Since the model is based on fundamental physics, the correlations between the received signals when the antennas are arranged in different configurations or with different radiation patterns, are physically motivated. Different channel scenarios can easily be obtained by varying the radius of the scattering disc as well as the location of the transmitting array relative to the disc [21].

In the scattering disc model, the path gain can be written as

$$H_{pq} = \sum_{l=1}^{L} \alpha_l \exp\left\{-jk_c(D_{TX_q \to s_l} + D_{s_l \to RX_p})\right\} g_{t_q}(\phi_l^T) g_{r_p}(\phi_l^R), \tag{3.40}$$

where the number of scatterers is L and α_l represents the independent and complex Gaussian distributed reflection coefficients with zero mean and unit variance.

Furthermore, $D_{s_l \to RX_p}$ and $D_{TX_{q \to s_l}}$, denote the distance from receiving array antenna p to scatterer l and scatterer l to transmitting array antenna q respectively. Note that the wave number is denoted $k_c \triangleq 2\pi/\lambda_c$. Finally $g_{t_q}(\phi_l^T)$ and $g_{r_p}(\phi_l^R)$ are the complex voltage radiation patterns of the transmitting and receiving antennas respectively. This model is suitable for simulations where the L scatterers are randomly placed. The channel matrix is calculated using (3.40) and then normalized according to some criterion[4]. The model (3.40) will be used in Section 3.7 and Chapter 7.

This model can also be extended to generate channels where a LOS component exists between the transmitter and the receiver. This resembles the Ricean fading channel [84]. The channel matrix is then decomposed into the direct (LOS) component and the scattered, Rayleigh distributed, component. Similar to the Ricean fading model, the K-factor is introduced, and defined as the ratio of the deterministic (LOS) and the scattered power components that gives pure LOS or pure scattering as extreme cases ($K = \infty$ and $K = 0$). The channel matrix can thus be expressed as,

$$\mathbf{H} = \mathbf{H}_{LOS} + \mathbf{H}_{scat} = \sqrt{\frac{KG}{K+1}}\mathbf{a}^{TX}(\theta_{TX})(\mathbf{a}^{RX}(\theta_{RX}))^* + \sqrt{\frac{G}{K+1}}\tilde{\mathbf{H}}_{scat} \ . \tag{3.41}$$

Here θ is the respective angle from array broadside and $\mathbf{a}(\theta)$ is the normalized array response vector. G is the large scale path gain including the antenna element gains. It can be used to normalize $\mathbf{H}$. The matrix $\mathbf{H}_{scat}$ is calculated using the scattering disc model (3.40). By assigning the mobile a speed and a direction through the cloud of scatterers, a time varying channel for each transmit-receive antenna pair is obtained. Different characteristics of the fading and correlation between the received signals can be obtained by varying the disc radius, the speed of mobile, the LOS parameter K and/or the base to mobile distance.

3.3.2 One-ring model

For some derivations, the complex correlation coefficients between the MIMO channel elements are required. The one-ring model, first introduced by Jakes [97] and later also used in e.g. [88] is suitable for this. This model is in fact a ray-tracing model and is based on the following assumptions: The receiver antenna array is surrounded by a ring of scatterers of radius R_s, see Figure 3.1. The distance from the center of the ring to the transmitter is D. Since D and R_s are typically large compared to the antenna element spacing, the angular spread is $\Delta \approx \arcsin(R_s/D)$ for all transmitter antennas. The scatterers are assumed to be uniformly distributed

[4] A common normalization criterion for channel matrices is to set $\|\mathbf{H}\|_F^2 = 1$

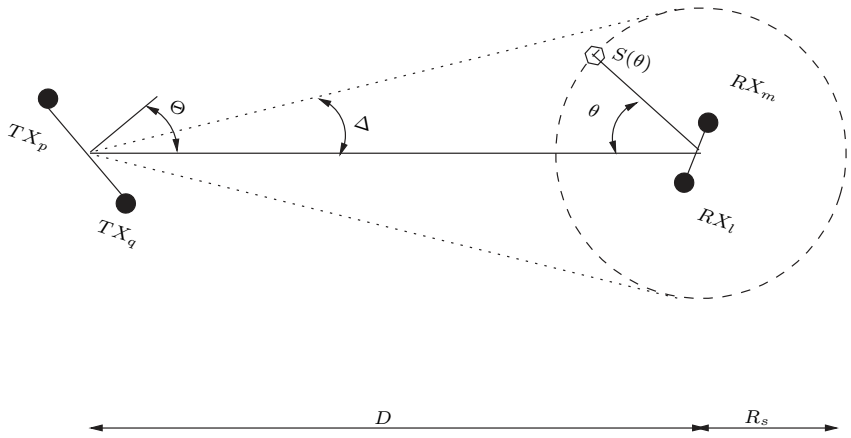


Figure 3.1: Description of the one-ring model. TX_i, RX_j are transmit and receive antenna i and j respectively. $S(\theta)$ is the scatterer at angle θ, Δ is the angular spread, D the transmit to receive antenna distance and R_s the radius of the scattering ring

on the scattering ring. Denote the scatterer located on the ring at angle θ, $S(\theta)$ and assume that rays that are scattered in $S(\theta)$ will change phase with the angle $\phi(\theta)$. Statistically, $\phi(\theta)$ is distributed uniformly in $[0, 2\pi[$ and i.i.d for every angle θ. Consider only rays that are reflected once and that all rays reaching the receiver antennas are equal in power. With these assumptions, the complex correlation coefficients are written as [88]

$$E\left\{H_{lp}H_{mq}^{*}\right\} = \frac{1}{2\pi}\int_{0}^{2\pi} \exp\{ -jk_c[\, D_{TX_p\to S(\theta)} - D_{TX_q\to S(\theta)} + \\ + D_{S(\theta)\to RX_l} - D_{S(\theta)\to RX_m}\,]\}\, d\theta \quad (3.42)$$

where $D_{TX_p\to S(\theta)}$ is the physical distance between transmit antenna p and the scatterer $S(\theta)$ on the ring and k_c is the wavenumber. The other distances in (3.42) are defined similarly. The elements defined by (3.42) are elements in the $n_r n_t \times n_r n_t$ channel covariance matrix $\mathbf{R}_H$ in (3.7). Recently, an analytical expression for the integral in (3.42) was given in [98], where a certain angular distribution of the angle of arrival (AOA) was included in the integral kernel, to be able to analyze scattering scenarios with non-uniform AOA distributions.

When Δ is small, the distances from the different transmitting array antennas via the scatterer to the receiver array are approximately equal, so the correlation between two paths from different transmit elements to one receive elements is expected to be high as compared to the large Δ case, where this correlation is smaller. Hence, by adjusting Δ, channel fading scenarios with weak or strong spatial cor-

relation is generated.

3.4 Measured MIMO channels

To collect real MIMO channel data, measurements were performed in an indoor environment at the Signals and Systems Group, Uppsala University. The aim of the measurements was to verify the assumptions made in the theoretical analysis regarding fading distribution, signal correlation and the flat fading assumption. In Chapter 4, outdoor measurements on a more advanced SIMO system are presented in detail.

3.4.1 The Measurement Setup

The measurements were performed in an indoor office environment using a 4×4 MIMO system at the frequency 1.8 GHz. A Vector Network Analyzer was used to measure the channel coefficients for the 16 channels using a switching method. Each complete "MIMO snapshot" of all the 16 channels were measured in less than 3 seconds, which is fast enough in indoor environments to ensure that the channel remains stationary during each measurement. The switching method was also used in [99] on an 8×8 MIMO system. Between each measurement, the receiving array was moved one eight of a wavelength in the broadside direction. The antenna elements were microstrip patch antennas placed in a linear array with an inter-element spacing (δ) of half a wavelength ($\delta = \lambda_c/2$) . The patch antennas had a half power beamwidth of 80° and a half power bandwidth of 170 MHz and the measurement SNR was set to 20 dB in all measurements. Two scenarios were investigated, one line of sight (LOS) and one non-LOS (NLOS) setup. In the LOS scenario, the two arrays were placed facing each other in an 8×6 meter laboratory room containing various instruments, tables and cabinets and $N_D = 146$ measurements were conducted. In the NLOS scenario, $N_D = 220$ measurements were conducted and the receive array was placed outside the laboratory room, centered in a long corridor with the array broadside parallel with the corridor. The transmit array was kept in the adjacent laboratory. To make comparisons with the theoretical models, each element of the measured $\mathbf{H}$ matrix was normalized as

$$H_{pq}(k) = \frac{H_{pq}^m(k)}{\sqrt{\frac{1}{N_D}\sum_{k=1}^{N_D} |H_{pq}^m(k)|^2}} \tag{3.43}$$

for $p = 1, \ldots, n_r$, $q = 1, \ldots, n_t$ and where $H_{pq}^m(k)$ is the measured channel matrix element in measurement k and $H_{pq}(k)$ is the normalized matrix element.

3.4.2 Coherence bandwidth

The initial measurements aimed to verify the flat Rayleigh fading assumption. Figure 3.2 shows the measured power spectrum in the NLOS case from one transmit antenna to the four receive antennas. The coherence bandwidth (at correlation coefficient 0.9) is estimated to $B_c = 2.8$ MHz so the flat fading assumption is valid if the signalling bandwidth is less than B_c. If a system with higher bitrate is required, then transmission over many subchannels can be used, where the bandwidth of each subchannel is less than B_c.

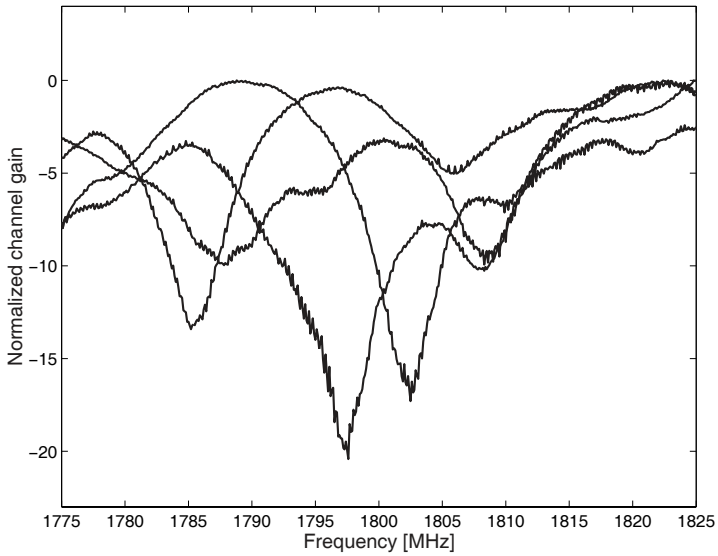


Figure 3.2: Power spectrum for NLOS channel. Each curve represents the received power in one out of four half-a-wavelength spaced antennas from one transmit antenna.

3.4.3 Fading statistics

Figure 3.3 shows the normalized channel amplitudes for all 16 channels in the NLOS case and it is immediately apparent that the channels fading patterns are different. The correlation between these subchannels is further explored below.

In Figure 3.4, the estimated probability density functions of the normalized amplitudes in the LOS and NLOS cases are shown. The curves are fitted to a Nakagami-m distribution using a moment based method [100]. The Nakagami-m distribution is given in (3.2). The m_f-parameter was estimated to $m_f = 1.07$ in the NLOS case and $m_f = 6.31$ in the LOS case. The measured data was also used

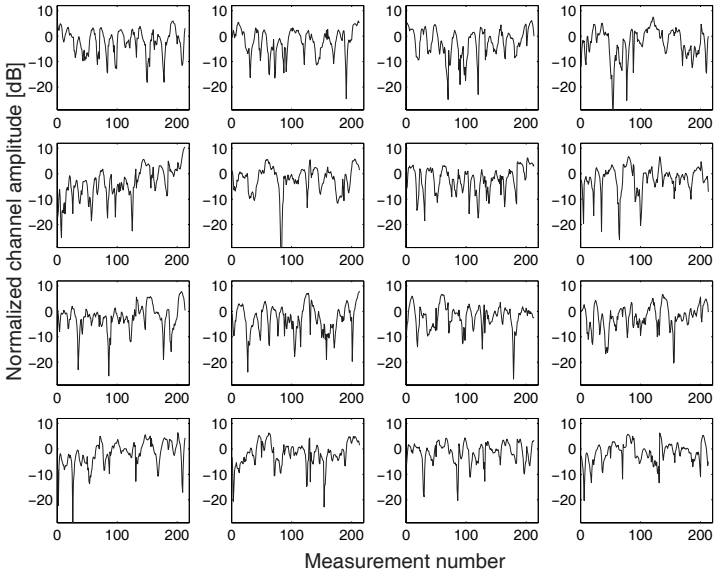


Figure 3.3: Channel amplitudes in a 4×4 MIMO NLOS channel as a function of measurement.

to estimate the Ricean K-factor (3.10) using a moment based estimator [101], and the value was found to be $K = 11.1$ in the LOS case and $K = 0.14$ in the NLOS case. Hence, the fading amplitude in the NLOS case is approximately Rayleigh distributed and the phase is close to uniformly distributed, see Figure 3.5.

3.4.4 Signal Correlation

An important property of the MIMO channel that essentially determines the channel capacity is the correlation between the channel coefficients. When multi antenna systems are analyzed, it is commonly assumed that the fading between pairs of transmit and receive antennas are independent and identically distributed random variables with a Rayleigh distribution, as in [48, 60]. However, in real propagation environments, the fades are dependent due to the local scattering environment or insufficient antenna element spacing [102]. On the other hand, a small antenna element spacing is often necessary to fit multiple antennas on a portable device. Small element spacing also introduces mutual coupling between the antenna elements which affects the achievable capacity of the system, although not necessarily in a negative way, as will be shown in Section 3.5. In [103] it was shown that under the model (3.11), the performance of a space-time block code having full diversity ($n_r n_t$) in the uncorrelated case, has its diversity order reduced to the product of the ranks of the receive and transmit correlation matrices, $\mathbf{C}_R$

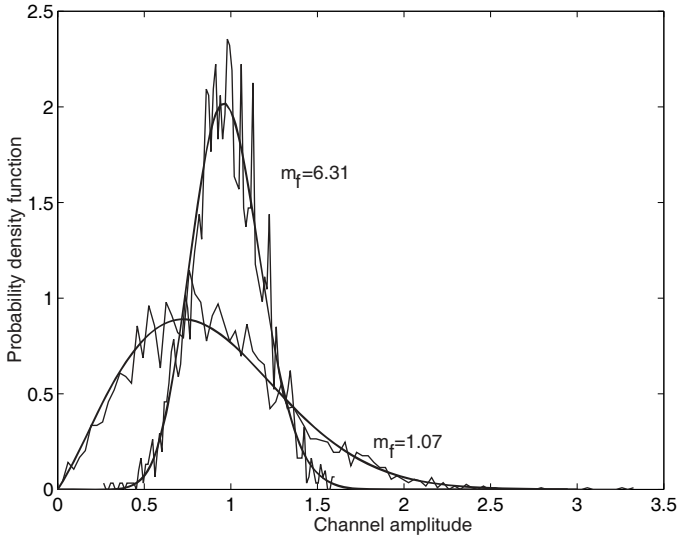


Figure 3.4: Estimated probability density functions of channel amplitudes, from measured data in LOS and NLOS channels. The dotted line shows the moment-based estimated Nakagami-m distribution. The m_f-parameters were estimated to 1.07 and 6.31 respectively. The Ricean K-factor was estimated to 11.1 in the LOS case and $K = 0.14$ in the NLOS case.

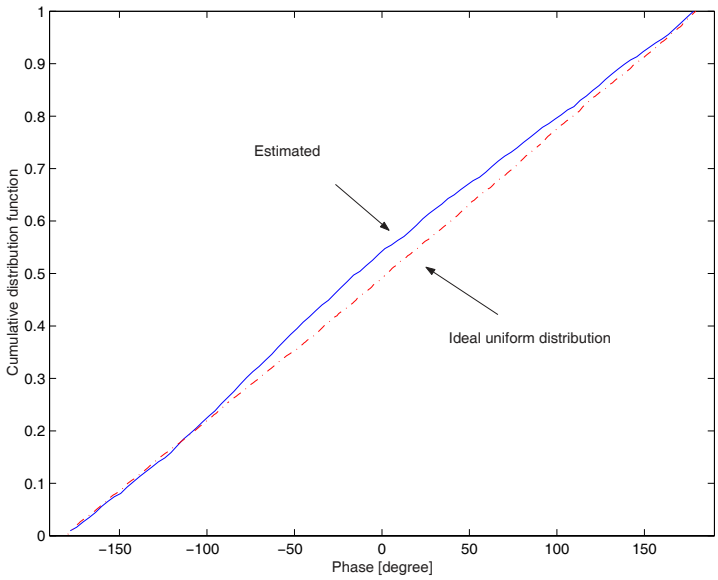


Figure 3.5: Cumulative distribution of phase in NLOS channel.

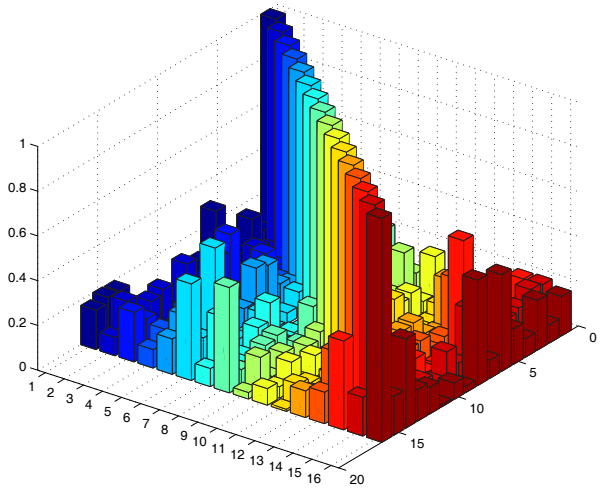


Figure 3.6: The magnitude of the envelope correlation matrix for a measured 4×4 MIMO channel. Non line of sight scenario.

and $\mathbf{C}_T$. Hence, in a LOS channel, where the received signals on the n_r antennas have a high correlation, diversity techniques are expected to perform poorly. On the other hand, if average CSI is known at the transmitter, in form of the channel covariance matrix $\mathbf{R}_H$, it was shown in [34] that a correlated channel has higher outage capacity and higher ergodic capacity than the uncorrelated one in the low SNR regime.

The envelope correlation coefficient can be approximated as [104]

$$\rho_{l,p}^{m,q} = \frac{E^2\left[H_{m,q}H_{l,p}^*\right]}{E\left[|H_{m,q}|^2\right] E\left[|H_{l,p}|^2\right]} \tag{3.44}$$

so the envelope correlation matrix is a normalization of the correlation matrix $\mathbf{R}_H$. This matrix is estimated by replacing the expectation operators in (3.44) by their sample mean equivalents. The 16×16 envelope correlation matrix can be seen in Figure 3.6 and 3.7 respectively. Clearly, the LOS channel has highly correlated channel coefficients and hence it is most suitable for beamforming, where coherent signals are desirable.

3.4.5 Eigenvalue analysis

In Figure 3.8 and 3.9, the four eigenvalues to the matrix $\mathbf{W} = \mathbf{H}\mathbf{H}^*$ is plotted as a function of the measurement snapshot. In the LOS case, the largest eigenvalue is dominant, 10-15 dB larger than the second largest eigenvalue, whereas in

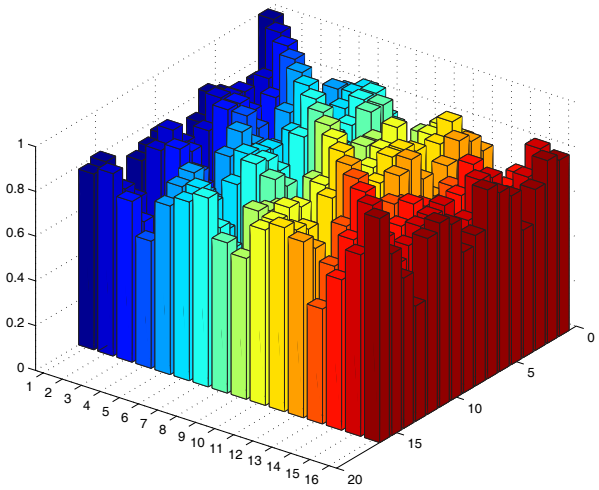


Figure 3.7: The magnitude of the envelope correlation matrix for a measured 4×4 *MIMO channel. Line of sight scenario.*

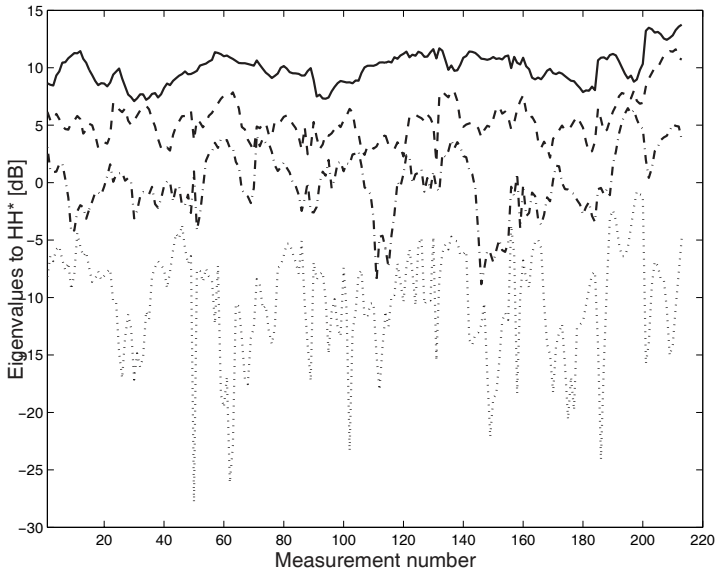


Figure 3.8: Eigenvalues for $\mathbf{W}$ *in a non line of sight (NLOS) MIMO channel as a function of measurement number.*

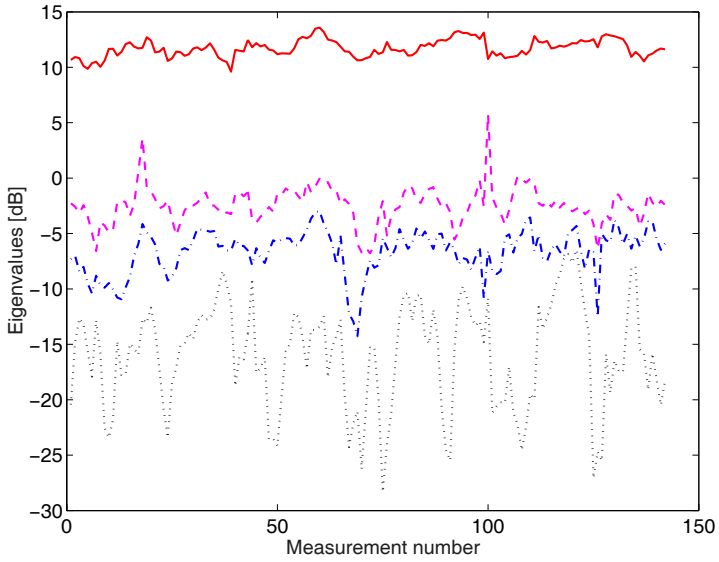


Figure 3.9: Eigenvalues for $\mathbf{W}$ *in a line of sight (LOS) MIMO channel as a function of measurement number.*

the NLOS case the difference is about 5 dB. Therefore, a beamforming method (which utilizes only one eigenmode of the channel) is expected to have the best performance in the LOS channel. Notice that changes in eigenvalues are somewhat dependent. The change in eigenvalues depends on how the array is moved relative to the channel eigenvectors. A move parallel to an eigenvector will only cause a change in its corresponding eigenvalue. For a discussion on temporal variations of MIMO channels, see [99]. In Figure 3.10, the cumulative distribution functions (cdf) for the four eigenvalues in the NLOS case are plotted. A Monte Carlo simulation was conducted to obtain the cdf:s for the four eigenvalues in the i.i.d. Rayleigh fading case. The largest eigenvalue is slightly larger than in the i.i.d Rayleigh case, which agrees with the observation that $m_f = 1.07$ was the best fit in the Nakagami-m distribution. A reasonable explanation is the measurement setup in a corridor, which should make the signals more focused along the corridor and hence the i.i.d. Rayleigh model with receiving signals uniformly distributed in arrival angle is not valid.

3.4.6 The capacity of the measured channels

The measured channel data is now used to calculate the mutual information, the capacity and the outage capacity. The data is used to investigate the performance with and without CSI at the transmitter. The average mutual information, or chan-

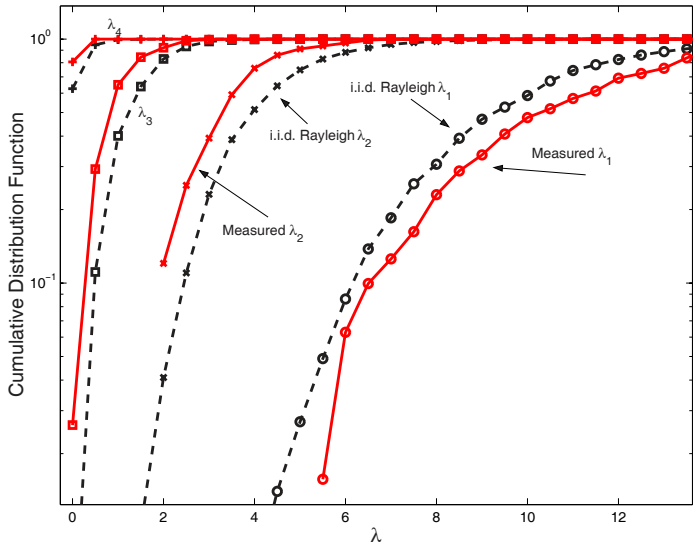


Figure 3.10: Cumulative distribution of the four eigenvalues in an NLOS channel. The i.i.d. Rayleigh fading channel CDF:s are also plotted for reference.

nel capacity $\hat{C}$, is estimated as

$$\hat{C} = \frac{1}{N_D}\sum_{i=1}^{N_D} \log_2 \det\left\{\mathbf{I}_m + \frac{1}{\sigma_n^2}\mathbf{H}_i\mathbf{R}_{cc}^i\mathbf{H}_i^*\right\} \tag{3.45}$$

where N_D is the number of measurements and $\mathbf{H}_i$ is the channel matrix from measurement i, and where $\mathbf{R}_{cc}^i$ is chosen to maximize mutual information for each channel (in the case when CSI is available at the transmitter). For the no CSI at the transmitter case, the capacity estimate

$$\hat{C} = \frac{1}{N_D}\sum_{i=1}^{N_D} \log_2 \det\left\{\mathbf{I}_m + \frac{P_T}{\sigma_n^2}\mathbf{H}_i\mathbf{H}_i^*\right\} \tag{3.46}$$

is used. In all results in this section, the measured channel matrices are normalized so that the rms value of all the channel matrix elements over all N_D measurements is one. When $N_D \to \infty$, this is equal to stating that $E\left\{|H_{pq}|^2\right\} = 1$. Furthermore, in this limit, $\hat{C}$ approaches the Ergodic (Shannon) capacity of the channel. The normalization ensures that path losses and antenna gains are removed from the measurements and only the spatial, scattering characteristics of the channel are investigated.

Figure 3.11 shows the estimated ergodic channel capacity for a 2×2 MIMO system where the element separation at the transmitter is $\lambda_c/2$ while for the receiver it varies between $\lambda_c/2$, λ_c and $3\lambda_c/2$. The LOS capacity is lower than the NLOS capacity. Note however that the normalization removes the gain the specular LOS channel has over the diffuse scattering NLOS channel. The gain in the LOS case is due to the unobstructed path between the two arrays which has low path loss. In [105], this was investigated using measurements and it was concluded that if the measured channel matrix is un-normalized, the capacity rose when moving from NLOS to LOS. An upper bound of this capacity can be found using the concavity of the $\log\det$ operator [106] together with Jensen's inequality[5]. Using (2.58)

$$\begin{aligned} C_{noCSI} =& E_{\mathbf{H}}\left[\log_2\det\left\{\mathbf{I}_m + \frac{P_T}{\sigma_n^2 n_t}\mathbf{H}\mathbf{H}^*\right\}\right] \\ \leq& \log_2\det\left\{\mathbf{I}_m + \frac{P_T}{\sigma_n^2 n_t}E_{\mathbf{H}}\left[\mathbf{H}\mathbf{H}^*\right]\right\} \\ =& \log_2\det\left\{\mathbf{I} + \frac{P_T}{\sigma_n^2 n_t}\mathbf{R}_R\right\} \end{aligned} \tag{3.47}$$

where $\mathbf{R}_R$ is the receive covariance matrix from (3.13). This upper bound is tight when the main part of the correlation is from the receive part of the system and when the channel is not rank deficient [106]. The ideal case $\mathbf{R}_R = \mathbf{I}$ is always satisfied with a semi-unitary channel matrix $\mathbf{H}$, so $\mathbf{H}\mathbf{H}^* = \mathbf{I}$. The reason for the gap between the measured curves and the ideal, unitary channel, is the correlation between the signals. The correlation seems to be approximately independent of the antenna element separation, which indicates that the correlation is not due to mutual coupling but from the scattering in the channel. The results in [106] also suggests that the angular spread of the incoming signals, Δ, in the measurements for the NLOS case, is larger than 1 rad$\approx 52^\circ$, since the channel capacity does not depend on the element separation (when it is larger than $\lambda_c/2$).

To analyze the performance gains from using CSI at the transmitter, the average mutual information (estimated Shannon capacity) when $\mathbf{R}_{cc}^i$ is chosen optimally, see Section 2.2.1, are compared in Figure 3.12 by expressing the ratio of the channel capacity with and without CSI. The antenna element spacing is $\lambda_c/2$ in both the 2×2 and the 4×4 cases. The benefits of CSI at the transmitter is decreased for increasing transmit power (or reduced receiver noise) and at an SNR around 25 dB, the benefits of having CSI is negligible. This is in accordance with the conclusions in Section 2.6. It is also seen that CSI gives a larger improvement in the

[5]Jensen's inequality states that if $f(X)$ is a concave function and X is a random variable, then $E\left[f(X)\right] \leq f(E\left[X\right])$.

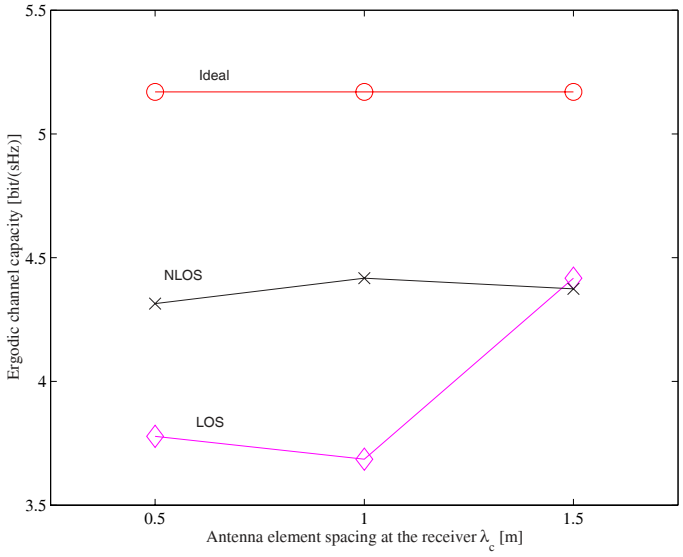


Figure 3.11: Estimated ergodic channel capacity when receiver antenna element distance is varied. Transmitter element distance fixed at $\lambda_c/2$ and no CSI at the transmitter. SNR=10 dB

LOS case, because there exists a strong mode in the channel which is more efficiently exploited by the water-filling algorithm. Furthermore, with more antennas, which implies a larger number of channel eigenmodes, CSI at the transmitter gives a larger advantage over the "blind" case in the NLOS scenario compared to the LOS case since a few or only one of the channel eigenmodes will be exploited by the water-filling algorithm.

Figure 3.13 shows the comparison between the beamforming and STBC. These two strategies are implementations of transmission with and without CSI. The estimated outage capacity is plotted at 10% outage probability. The estimated outage capacity, R, at 10% probability is defined as the median of the $N_D/5$ smallest values of the estimated channel capacity $\hat{C}$. This approach was used in [34] and is there called the sample-mean outage capacity. For two antenna element arrays, the difference between LOS and NLOS channels is small. At low SNR, knowledge of CSI at the transmitter gives a large advantage over the STBC in the 4×4 scenario.

The sub-optimality of beamforming

Figure 3.14 shows the sub-optimality of using beamforming when CSI is known at the receiver compared to the optimal waterfilling solution which utilizes all m

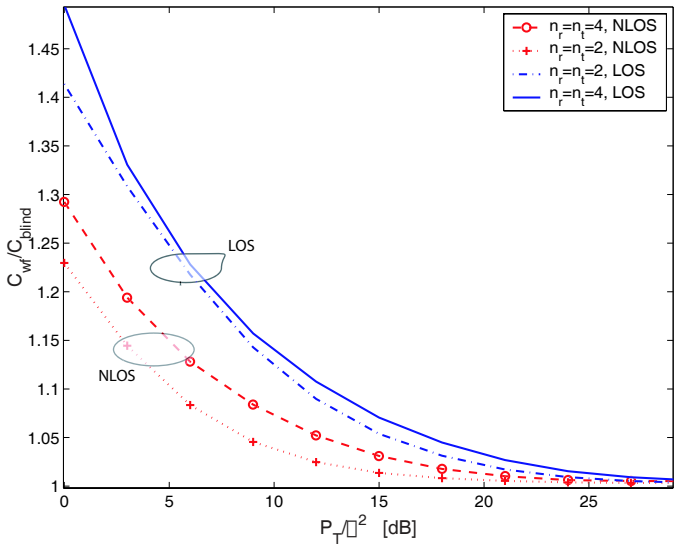


Figure 3.12: Relative channel capacity when using CSI (equation (3.45)*) compared to no CSI (equation* (3.46)*) at the transmitter as a function of transmitted power. Curves are estimated from measurements, using* (3.45)*. Water-filling based on CSI versus optimal blind transmission.*

subchannels. Recall that beamforming is capacity optimal in the MISO case but not in the general MIMO channel. Intuitively, a rank one MIMO channel would give a beamforming performance equal to the optimal waterfilling performance since the channel in this case has only one non-zero eigenvalue. This is confirmed in Figure 3.14, where for $n_t = n_r = 2$ MIMO systems at low SNR in the LOS case, the beamforming approach becomes equivalent to the waterfilling approach. The probability for beamforming to be optimal is given by the probability that no power is allocated to subchannel 2, that is, from (2.52):

$$\gamma_2 = \left(\mu - \frac{\sigma_n^2}{\lambda_2}\right)^+ = 0 \ . \tag{3.48}$$

If this expression is combined with the total power criterion, (2.53), we get the probability that beamforming is optimal as

$$\Pr\left\{\frac{1}{\lambda_2} - \frac{1}{\lambda_1} > \frac{P_T}{\sigma_n^2}\right\} \ . \tag{3.49}$$

In a LOS channel $\lambda_1 \gg \lambda_2$, so from (3.49) it is expected that beamforming is optimal at a larger SNR than in the NLOS case. In the 4×4 NLOS channel, the gain of using several subchannels is substantial and the method derived in Section 2.4 should be used to exploit these subchannels.

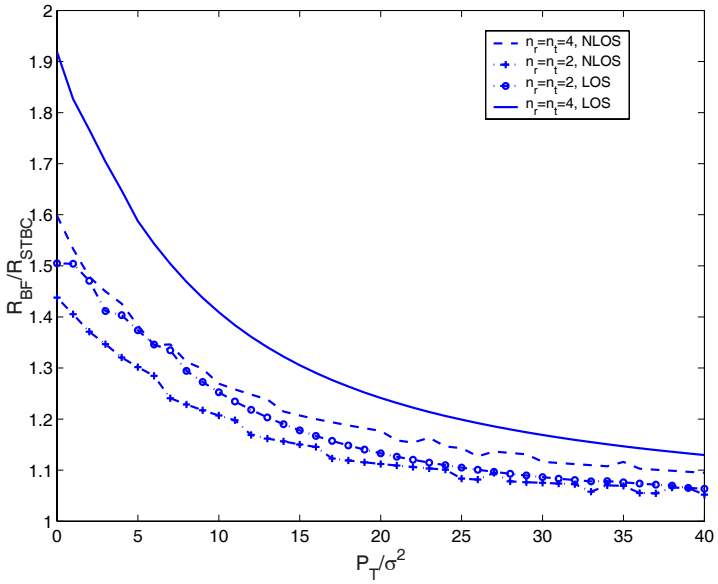


Figure 3.13: Relative outage capacity gain by using CSI as a function of transmitted power at $P_{out} = 0.1$ probability. Beamforming versus space time block coding with n_s=N, i.e. a rate one code.

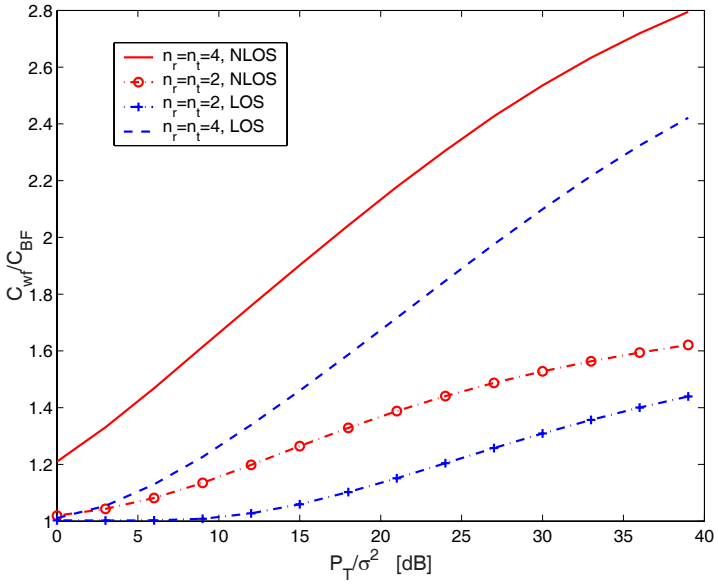


Figure 3.14: Relative estimated channel capacity by waterfilling versus beamforming as a function of SNR.

3.4.7 The bit error rate

Using the measured 2×2 channels, the BER assuming a coherently detected BPSK modulation is now calculated. Figure 3.15 shows the BER estimated using the measured channels along with the theoretical curves from (3.92) and (3.68) with $m_f = 1$, for beamforming and STBC respectively. Assuming coherently detected BPSK modulation, the BER, denoted $\hat{P}_{STBC}$ and $\hat{P}_{BF}$ for Space Time Block Coding and beamforming respectively, for the measured systems are estimated as

$$\hat{P}_{STBC} = \frac{1}{N_D} \sum_{i=1}^{N_D} Q\left(\sqrt{2 \frac{P_T}{\sigma_n^2 n_t} \|\mathbf{H}_i\|_F^2}\right) \tag{3.50}$$

and

$$\hat{P}_{BF} = \frac{1}{N_D} \sum_{i=1}^{N_D} Q\left(\sqrt{2 \frac{P_T}{\sigma_n^2} \lambda_{max}\left(\mathbf{H}_i \mathbf{H}_i^*\right)}\right) \tag{3.51}$$

where $\mathbf{H}_i$ are the channel matrices from the measurements. Due to the short measured data series, the measured curves cannot accurately estimate a BER below 10^{-2}. This is the reason for the deviation from the theoretical curves in Figure 3.15. For low SNR however, the number of bit errors are so large that the theoretical and measured curves coincide. The CSI at the transmitter gives a gain in the 2×2 case, of about 2 dB at a bit error rate of 10^{-2}. Note that the slope of the BER curves is equal, which indicates that beamforming and STBC has the same diversity advantage. This will be theoretically verified for the i.i.d. Rayleigh fading channel later in this chapter.

3.5 Transmission over fading channels without CSI at the transmitter

The aim of this section is to analytically investigate the performance of an orthogonal space time block coding system in a fading environment with correlated signals at the n_r receiver antennas. The amplitude of the flat fading channel gains are assumed to be Nakagami-m distributed [82] with integer parameter m_f, which includes the Rayleigh distribution as a special case ($m_f = 1$), thereby allowing both LOS and NLOS cases to be modelled. The correlated signals may arise from the scattering environment, closely spaced antenna elements and from mutual coupling between antenna elements. For the signal correlation generated by the scattering environment, the unstructured general model for $\mathbf{R}_H$, given by equation (3.7), is

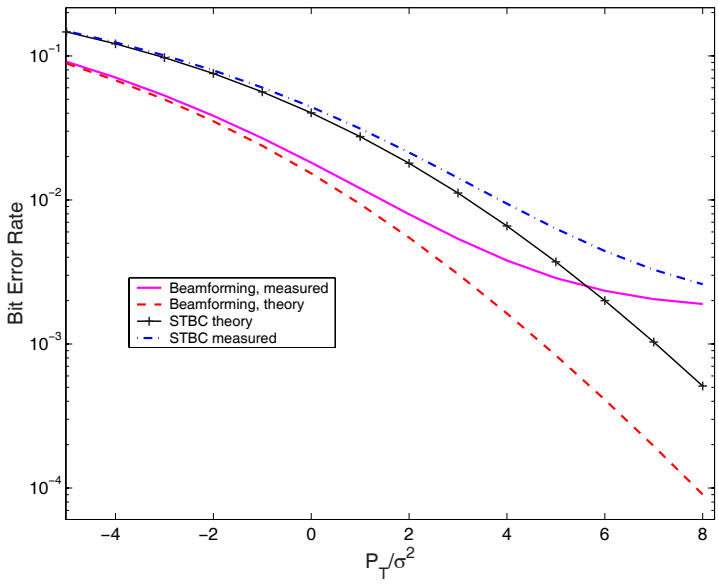


Figure 3.15: Bit error rates for beamforming and STBC ($n_r = n_t = 2$) as a function of SNR. The scale on the x-axis is in dB.

assumed where any real or imaginary part of two arbitrary elements of the channel matrix may be correlated with each other. The mutual coupling is modelled by a pre- and post-multiplication of the channel matrix $\mathbf{H}$ as $\mathbf{H}' = \mathbf{C}_R\mathbf{H}\mathbf{C}_T$. This is reasonable since the mutual coupling between an antenna element at the receiver side and one at the transmitter side is negligible.

The outage capacity is derived as well as the bit error rate for the STBC under the assumption of coherently detected BPSK modulation. The results are given in Theorem 3.2 and 3.3 respectively.

Previous works studied the impact of spatial fading correlation on multiple input multiple output (MIMO) capacity in a Rayleigh fading channel [88] by dividing the MIMO channel into subchannels using an eigen-decomposition. The corresponding subchannel gains are then determined by the fading correlation. Furthermore, in [86], the performance in terms of average pairwise codeword error probability (PEP) of a space time code system in a Rayleigh fading environment was investigated and it was further shown how the diversity order is dependent on the rank of the fading correlation matrices, $\mathbf{R}_R$ and $\mathbf{R}_T$.

The orthogonal STBC over a MIMO channel can be written as an equivalent SISO system, as was shown in Section 2.3.2. The SNR of the equivalent SISO system depends on the Frobenius norm of the channel matrix, see equation (2.71) and since the Frobenius norm is the sum of the magnitude of the matrix elements

squared, the statistical properties of this norm is derived using the theory of sums of correlated random variables and is presented in Lemma 3.1, which then is used to prove the Theorems 3.2 and 3.3. A previously studied and related problem is the performance of the maximum ratio diversity combiner in a correlated Nakagami-m fading channel, given complex correlation coefficients between the diversity branches [83] or envelope correlation coefficients [107].

The derived results are illustrated by a few examples in which the performance of a STBC system with fading correlation and mutual coupling is found. The fading correlation is modelled by the extended "one-ring model", see Section 3.3.2. Mutual coupling is modelled using half-wavelength dipoles, which are known to give a high coupling due to the omni-directional radiation properties [108]. The dipole model is convenient in the sense that it provides analytical expressions for the mutual coupling coefficients. Some concluding remarks of this section are presented in Section 3.5.6.

3.5.1 Analyzing orthogonal STBC performance

Assume that the space time block coding (STBC) encoder and decoder structure as presented in Section 2.3.2 is used over a wireless link using a transmitter with $n_t > 1$ antennas, a receiver with $n_r \geq 1$ antennas. Furthermore, the communication is carried out using bursts or packets of length N symbols. The channel is assumed to be quasi-static, i.e. the elements of $\mathbf{H}$ are fixed during the transmission of these N symbols. The model can then be formulated as the block fading model presented in Section 1.1.2. A key observation, made in [60], is that the input symbols become decoupled and the MIMO channel model (1.22) can be written as an equivalent SISO channel model (2.71),

$$\mathbf{y} = \sqrt{\frac{P_T}{\sigma_n^2 n_t}} \|\mathbf{H}\|_F^2 \, \mathbf{s} + \mathbf{v} \tag{3.52}$$

where the symbols $\mathbf{s}$ are defined as complex valued and unitary[6]. In the complex symbol case, $\mathbf{y}$ is the $2n_s \times 1$ vector after decoding of the received matrix $\mathbf{Y}$, $\mathbf{s}$ is a vector with two stacked $n_s \times 1$ blocks corresponding to the real and imaginary parts of the input symbols[7] and $\mathbf{v}$ is the noise vector with circular normal i.i.d elements and which variance can be shown to be $\|\mathbf{H}\|_F^2 \, \sigma_n^2$, [60] see also equation (2.72). So the effective SNR at the receiver is equal to $\frac{P_T}{\sigma_n^2 n_t} \|\mathbf{H}\|_F^2$, and to evaluate the STBC performance, the pdf of $\|\mathbf{H}\|_F^2$ is desired. It is given by the following lemma.

[6] If a symbol alphabet with non-equal powered symbols are used, such as 64QAM, then the assumption that the average power of the complex-valued symbols is 1 is used.

[7] This stacking is an abstract operation to be able to formulate the equivalent SISO system (3.52). In which order the n_s symbols are arranged in the two blocks in the vector s is of no importance.

Lemma 3.1 *Assume that* $\mathbf{H}$ *is a correlated Nakagami-m fading channel, according to Definition 3.1 on page 80. The probability density function of the squared Frobenius norm,*

$$\Upsilon \triangleq \|\mathbf{H}\|_F^2 \tag{3.53}$$

is then

$$p_\Upsilon(x) = \sum_{l=1}^{n_r n_t} \sum_{r=1}^{m_f} c_{l,r} p(x; l, r) \tag{3.54}$$

where $p(x; l, r)$ *is a Gamma distribution,*

$$p(x; l, r) = \frac{1}{\nu_l \Gamma(r)} \left(\frac{x}{\nu_l}\right)^{r-1} e^{-x/\nu_l} \tag{3.55}$$

and $\nu_k, k = 1, \ldots, m$, $m = \min(n_r, n_t)$ *are the distinct eigenvalues to the spatial covariance matrix* $\mathbf{R}_H$*. The coefficients* $c_{l,r}$ *are given by*

$$c_{l,r} = \left(\prod_{k=1, k\neq l}^{n_r n_t} (1 - s\nu_k)^{-r} \right) \Bigg|_{s=1/\nu_l} \tag{3.56}$$

for $r = m_f$ *and*

$$c_{l,r} = \frac{1}{(m_f - r)!(-\nu_l)^{m_f - r}} \frac{d^{(m_f - r)}}{ds^{(m_f - r)}} \left(\prod_{k=1, k\neq l}^{n_r n_t} (1 - s\nu_k)^{-m_f} \right) \Bigg|_{s=1/\nu_l} . \tag{3.57}$$

for $0 < r < m_f$.

Proof: See Appendix 3.B on page 145. ■

Note that the covariance matrix $\mathbf{R}_H$ in Lemma 3.1 has an arbitrary structure and is thus not restricted to have the form (3.16). Hence, the assumptions 3.1 and 3.2 on page 81 must not necessarily hold for Lemma 3.1 to be valid. Note also that it is the covariance matrix for the case of $m_f = 1$, i.e. the Rayleigh fading MIMO channel that should be used. Hence, even though $m_f > 1$, the covariance matrix $\mathbf{R}_H$ in Lemma 3.1 should be calculated *as if the channel was Rayleigh fading*. See the discussion prior to Definition 3.1. Furthermore, the eigenvalues must be distinct for the calculations of the coefficient $c_{l,r,}$ using the described procedure to be valid. The terms in the covariance matrix in Lemma 3.1 is most commonly due to the scattering in the channel. However, it can contain the mutual coupling if measurements or an electromagnetic software that takes mutual coupling into account has been used to obtain $\mathbf{R}_H$. Otherwise, coupling matrices from the array antenna design can be used to extend the Lemma. This method is described in the following section.

Mutual coupling

The theory derived above is now extended to take mutual coupling at the transmit and receive antenna array into consideration. It is a reasonable assumption that the coupling at the transmitter side is completely independent of the receiver side coupling and vice versa. The mutual coupling at the transmitter and the receiver is modelled by multiplying the channel matrix $\mathbf{H}$ with coupling matrices $\mathbf{C}_T$ and $\mathbf{C}_R$ for the transmit and receive side respectively to form a new channel matrix $\mathbf{H}' = \mathbf{C}_R\mathbf{H}\mathbf{C}_T$. This approach was also taken in [109]. It shall be noted that this model of mutual coupling, using coupling matrices, is a simplification, to make the following examples tractable. In [87], a more rigorous study of mutual coupling is performed, where the matching networks are incorporated.

As the correlation matrix $\mathbf{R}_H$ is defined in the $m_f = 1$ case, see Definition 3.1 on page 80, it is now shown in the following corollary how this correlation matrix should be updated when mutual coupling is present.

Corollary 3.2 *The pre- and post-multiplication of the correlated Nakagami-m MIMO channel* $\mathbf{H}$*, with the coupling matrices* $\mathbf{C}_R$ *and* $\mathbf{C}_T$ *under the condition* $m_f = 1$ *implies that the channel covariance matrix* $\mathbf{R}_H$ *in Lemma 3.1 is updated according to the following procedure.*

Make the separation of the original $\mathbf{R}_H$ *as*

$$\mathbf{R}_H = \mathcal{B} + j\mathcal{C} \tag{3.58}$$

where $\mathcal{B}$ *and* $\mathcal{C}$ *are real matrices. Define*

$$\Sigma = \begin{bmatrix} \mathcal{B} & \mathcal{C} \\ -\mathcal{C}^T & \mathcal{B} \end{bmatrix} \tag{3.59}$$

and

$$\mathcal{K} = \begin{bmatrix} Re\left(\mathbf{C}_T^T \otimes \mathbf{C}_R\right) & -Im\left(\mathbf{C}_T^T \otimes \mathbf{C}_R\right) \\ Im\left(\mathbf{C}_T^T \otimes \mathbf{C}_R\right) & Re\left(\mathbf{C}_T^T \otimes \mathbf{C}_R\right) \end{bmatrix}_{2n_rn_t \times 2n_rn_t}. \tag{3.60}$$

The channel covariance matrix with mutual coupling taken into account is then obtained by first forming $\Sigma' = \mathcal{K}\Sigma\mathcal{K}^T$ *and then using* (3.59) *to find* $\mathcal{B}'$,$\mathcal{C}'$*, and finally* (3.58) *to find the new channel covariance matrix* $\mathbf{R}'_H$.

Proof: See Appendix 3.C on page 149 . ■

The coupling matrices $\mathbf{C}_T$ and $\mathbf{C}_R$ depends on the type of antenna elements and their orientation, and is often obtained by a calibration measurement, see Chapter 5 or an electromagnetic modelling software, see Chapter 7. If the coupling is known, then it can be compensated for at the transmitter and the receiver [104], but requires

repeated calibrations due to drift in the coupling parameters over time. Algorithms for tracking these parameters are derived in Chapter 5 where also the effect of calibration errors are investigated.

Lemma 3.1 is now used to formulate the theorems that express the bit error rate and outage capacity of the orthogonal STBC system.

3.5.2 Outage capacity for STBC

Due to the rewriting of the MIMO system to an equivalent SISO system, (3.52), the outage capacity, according to Definition 2.2 on page 36, for a space time block coded system over a correlated Nakagami-m fading channel can be found and is presented in the following theorem.

Theorem 3.2 *The outage capacity C_{P_o} at outage probability P_o for the orthogonal STBC in a correlated Nakagami-m fading channel, see Definition 3.1, is given by the solution to the equation*

$$P_o = \sum_{l=1}^{n_r n_t} \sum_{r=1}^{m_f} \frac{c_{l,r}}{\Gamma(r)} \Gamma\left(r, \frac{\sigma_n^2 n_t}{P_T \nu_l}(2^{C_{P_o} N/n_s} - 1)\right) \tag{3.61}$$

where

$$\Gamma(\alpha, x) = \int_0^x e^{-t} t^{\alpha-1} dt \tag{3.62}$$

is the incomplete Gamma function [80, Sec. 8.351]. The eigenvalues ν_l and constants $c_{l,r}$ are those given in Lemma 3.1. If mutual coupling is considered, then the procedure in Corollary 3.2 is used to update the covariance matrix $\mathbf{R}_H$ before applying Lemma 3.1.

Proof: See Appendix 3.D on page 150. ■

The eigenvalues ν_l are in (3.61) multiplied with the SNR, $P_T/(\sigma_n^2 n_t)$, so the eigenvalues can be seen as the power gain of the corresponding mode of the channel.

3.5.3 The bit error rate of STBC

Using the pdf of Υ from Lemma 3.1, it is also possible to calculate the bit error rate. The result is summarized in this theorem.

Theorem 3.3 *The bit error rate, P_{STBC}, for orthogonal STBC using coherently detected BPSK modulation or Gray coded QPSK modulation, in a correlated*

Nakagami-m fading channel, according to Definition 3.1, is given by the expression

$$P_{STBC} = \sum_{l=1}^{n_r n_t} \sum_{r=1}^{m_f} c_{l,r} \left[\frac{1}{2} - \sqrt{\frac{P_T \nu_l}{n_t \sigma_n^2 \pi}} \frac{\Gamma(r+1/2)}{\Gamma(r)} F\left(\frac{1}{2}, \frac{1}{2}+r; \frac{3}{2}; -\frac{P_T}{n_t \sigma_n^2}\nu_l\right)\right] \tag{3.63}$$

The eigenvalues ν_l and constants $c_{l,r}$ are those given in Lemma 3.1. If mutual coupling is considered, then the procedure in Corollary 3.2 is used to update the covariance matrix $\mathbf{R}_H$ before applying Lemma 3.1. Here $F(\cdot)$ is the hypergeometric function which can be expressed as, [80, Sec. 9.100],

$$F\left(\frac{1}{2}, \frac{1}{2}+r; \frac{3}{2}; -x\right) = \frac{1}{(1+x)^{r-1/2}} \sum_{k=0}^{r-1} \frac{(1-r)_k}{(\frac{3}{2})_k} (-x)^k \tag{3.64}$$

where $(a)_k = a \cdot (a+1) \cdot (a+2) \cdots (a+k-1)$ is the hypergeometric coefficient (and $(a)_0 = 1$ by definition).

Proof: See Appendix 3.E on page 151. ■

Remark 3.1 *In the Rayleigh fading case, $m_f = 1$, and the property [80, Sec. 9.121]*

$$F(-n, \beta; \beta; -z) = (1+z)^n \qquad (\beta \text{ arbitrary}) \tag{3.65}$$

can be used together with $\Gamma(3/2) = \sqrt{\pi}/2$ to simplify (3.63) *as*

$$P_{STBC} = \sum_{l=1}^{n_r n_t} \frac{c_l}{2} \left[1 - \sqrt{\frac{\nu_l}{n_t \sigma_n^2 / P_T + \nu_l}}\right] \quad . \tag{3.66}$$

From (3.63) and (3.61), we see that the eigenvalues to the correlation matrix $\mathbf{R}_H$ is multiplied with the SNR, P_T/σ_n^2. Hence, the BER is a weighted sum of $n_r n_t m_f$ subchannels, each with an equivalent SNR of $P_T \nu_l / \sigma_n^2$ and with the weighting factors $c_{l,r}$. Similarly, the outage capacity is the linear combination of $m n_r n_t m_f$ subchannel capacities. Unfortunately, it is difficult to find the diversity advantage from (3.66) in the general case, since it depends on the coefficients $c_{l,r}$.

3.5.4 Independent fading and no mutual coupling

A commonly used assumption in the analysis of MIMO system is to assume i.i.d. fading of the $n_r n_t$ elements in the channel matrix $\mathbf{H}$. The following corollary handles this special case.

Corollary 3.3 *Assume that* $\mathbf{H}$ *is a Nakagami-m fading MIMO channel, according to Definition 3.1, with uncorrelated elements and fading figure* m_f*. The outage probability is then*

$$P_o = \frac{1}{\Gamma(m_f n_r n_t)} \Gamma\left(m_f n_r n_t, \frac{\sigma_n^2 n_t}{P_T}\left(2^{C_{P_o} N/n_s} - 1\right)\right) \tag{3.67}$$

and the BER in the coherently detected BPSK and QPSK case is

$$P_{STBC} = \frac{1}{2} - \sqrt{\frac{P_T}{n_t \sigma_n^2 \pi}} \frac{\Gamma(m_f n_r n_t + 1/2)}{\Gamma(m_f n_r n_t)} F\left(\frac{1}{2}, \frac{1}{2} + m_f n_r n_t; \frac{3}{2}; -\frac{P_T}{n_t \sigma_n^2}\right) \ . \tag{3.68}$$

Proof: See Appendix 3.F on page 151. ■

We can calculate the diversity advantage of the STBC from (3.68) by expand the expression (3.68) in the point $P_T/(n_t\sigma_n^2) = \infty$. From [110], the expansion of the hypergeometric function $\mathrm{F}(a, b; c; x)$ in $x = \infty$ is

$$\begin{aligned} \mathrm{F}(a, b; c; x) \propto & \frac{\Gamma(b-a)\Gamma(c)}{\Gamma(b)\Gamma(c-a)}(-x)^{-a}(1 + O(x^{-1})) \\ & + \frac{\Gamma(a-b)\Gamma(c)}{\Gamma(a)\Gamma(c-b)}(-x)^{-b}(1 + O(x^{-1})) \qquad |x| \to \infty, \forall a \neq b \ . \end{aligned} \tag{3.69}$$

If this result is used in (3.68) we get

$$P_{STBC} \approx \left(\frac{P_T}{n_t \sigma_n^2}\right)^{-m_f n_t n_r} \qquad \text{for } \frac{P_T}{n_t \sigma_n^2} \gg 1 \tag{3.70}$$

hence, the diversity advantage for a STBC in the channel $\mathbf{H}_w$ defined in Definition 3.2 is $m_f n_r n_t$. This is expected since we have $n_r n_t$ uncorrelated elements in $\mathbf{H}_w$ and in the Nakagami-m fading distribution, the amplitude consist of a sum of m_f independent terms. See Section 3.2.1.

In the SISO system with Rayleigh fading, $m_f n_r n_t = 1$ and by using the relation (3.65), we obtain

$$P_{STBC} = \frac{1}{2}\left[1 - \sqrt{\frac{P_T}{\sigma^2 + P_T}}\right] \tag{3.71}$$

which the observant reader identifies as the bit error rate for coherently detected BPSK in a Rayleigh fading channel [15].

3.5.5 Examples

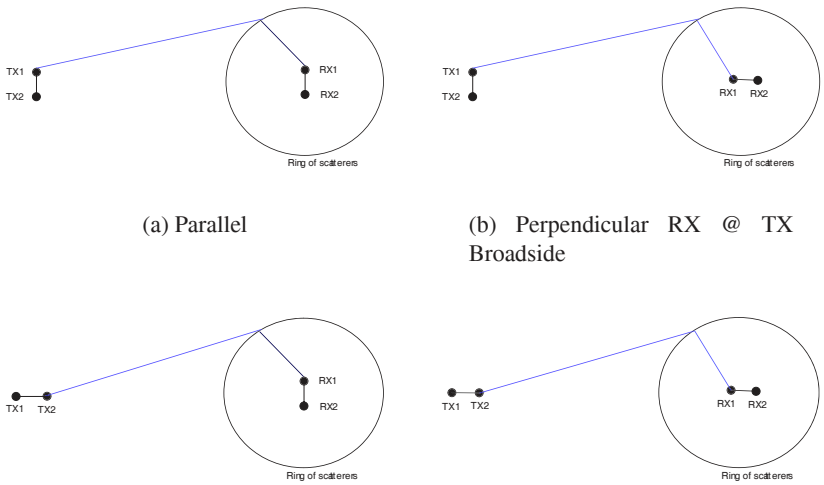
A series of numerical examples are presented using the theory derived in the previous sections. The theory is exemplified with the "one-ring" scattering model to introduce spatial correlation and mutual coupling, by assuming dipole antenna elements. In all examples, $E_b/N_0 = 10$ and λ_c is the wavelength of the transmitted signals. For a given number of transmit and receive antennas, the STBC that has the largest possible bit rate n_s/N is chosen from the table in [56].

The one ring model is used to calculate values for the complex correlation coefficients in the matrix $\mathbf{R}_H$ above. The model is introduced and discussed in Section 3.3.2 and here the integral in (3.42) is replaced by a summation over 180 equally spaced scatterers around the scattering ring. The spatial correlation in the one-ring model is adjusted by the parameter Δ, see Figure 3.1. Mathematically, a small Δ corresponds to an $\mathbf{R}_H$ matrix with a few dominant eigenvalues whereas a large Δ approaches the i.i.d. fading case which corresponds to all eigenvalues of $\mathbf{R}_H$ being equal and one.

The examples in this section are limited to uniformly spaced linear arrays and the antenna element spacing is defined as δ_{TX} at the transmit antenna array and δ_{RX} at the receive antenna array. The performance will depend on the mutual orientation of the transmitter and receiver arrays. In the examples in this chapter, the orientations presented in Figure 3.16 is used.

The first example shows the impact of the one-ring model on the STBC outage capacity. Figure 3.17 shows the $C_{0.1}$ outage capacity when the angular spread Δ is varied for a 2×2 MIMO system with rate $n_s/N = 1$ in a Rayleigh fading channel ($m_f = 1$). The probability that the channel capacity is less than $C_{0.1}$ is thus 0.1 and is calculated using Theorem 3.2. When Δ is increased, the outage capacity increases rapidly for the case when the two arrays are placed parallel to each other (arrays broadside facing each other), due to a reduced spatial correlation between signals from different transmit antennas. If the receiver array is placed orthogonal to the transmit antenna array, then the outage capacity reaches an upper limit, due to the inability of the array to separate the transmission from different antenna elements, which can also be interpreted as a smaller antenna aperture. In the inline case, when the arrays are placed end-fire to end-fire, the performance increase is very slow and it is concluded that the orientation of the antenna array have a large impact on the achievable outage capacity. The same conclusion was made in [88] where it also was shown that a hexagonal array structure was more robust to the array orientation.

The effect of the Nakagami-m fading parameter on the outage capacity is shown in Figure 3.18 for a fixed spreading angle of $\Delta = 5^\circ$ and also the i.i.d. channel curve is plotted for comparison. Rate $n_s/N = 1$ and $n_s/N = 3/4$ codes



(a) Parallel

(b) Perpendicular RX @ TX Broadside

(c) Perpendicular TX @ RX Broadside

(d) In-line

Figure 3.16: The four different mutual orientations of the transmitter (TX) and receiver (RX) antenna arrays, used in the examples. The circle is the ring of scatterers. The line represents one of the paths from a transmitter antenna to a scatterer and then to a a receiver antenna.

are used for the $n_r = n_t = 2$ and $n_r = n_t = 4$ cases respectively. It is clear that when the parameter m_f increases, the probability of a deep fade decreases and the channel becomes more reliable. Thus the outage capacity will increase.

Mutual coupling

In an array where no software compensation of the mutual coupling is performed, the outage capacity will strongly depend on the array element coupling as is shown in this example. The coupling matrix of an antenna array can be written as

$$\mathbf{C}_{co} = (Z_A + Z_T)\,(\mathbf{Z} + Z_T\mathbf{I})^{-1} \tag{3.72}$$

where Z_A is the antenna impedance, Z_T the receiver or transmitter impedance and $\mathbf{Z}$ is the mutual impedance matrix [111].

The mutual impedance can be calculated using the method of induced electromotive-force [108], but this usually leads to intractable calculations. One exception is the case of thin dipoles of finite length, which gives an expression for the elements in the mutual impedance matrix $\mathbf{Z}$ which can be found in [108] and modified

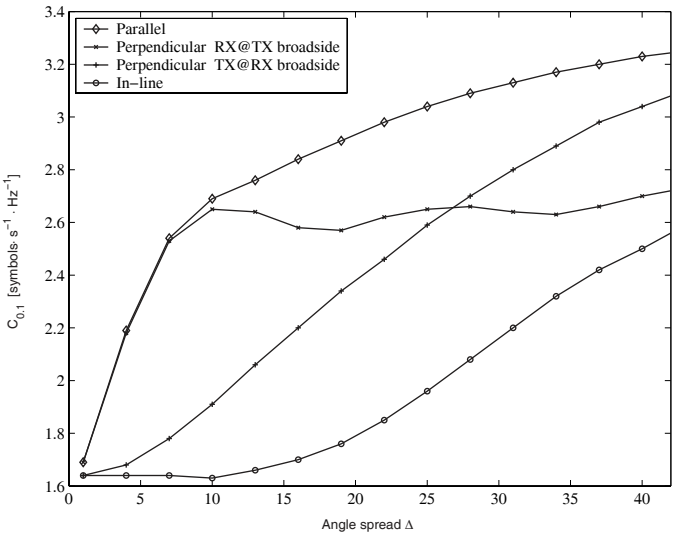


Figure 3.17: $C_{0.1}$ outage capacity versus angular spread Δ, see Figure 3.1. The number of antenna elements are $n_r = n_t = 2$ and the mutual orientation of the TX and RX arrays are varied. The Nakagami-m parameter is $m_f = 1$, i.e. Rayleigh fading. The antenna spacing at the transmitter $\delta_{TX} = 2\lambda_c$ and at the receiver $\delta_{RX} = 0.5\lambda_c$.

in [104] to

$$Z_{pq} = R_{pq} + jX_{pq} \tag{3.73}$$

$$R_{pq} = \frac{\eta}{4\pi \sin(\pi\delta/\lambda_c)} \left[2Ci(u_0) - Ci(u_1) - Ci(u_2)\right] \tag{3.74}$$

$$X_{pq} = -\frac{\eta}{4\pi \sin(\pi\delta/\lambda_c)} \left[2Si(u_0) - Si(u_1) - Si(u_2)\right] \tag{3.75}$$

where the constants u_0, u_1 and u_2 are defined as

$$u_0 = \frac{2\pi\delta}{\lambda_c}|p-q| \tag{3.76}$$

$$u_1 = \frac{2\pi}{\lambda_c}\left(\sqrt{\delta^2|p-q|^2 + l^2} + l\right) \tag{3.77}$$

$$u_2 = \frac{2\pi}{\lambda_c}\left(\sqrt{\delta^2|p-q|^2 + l^2} - l\right) \quad . \tag{3.78}$$

Here δ is the dipole spacing and l is the dipole length. The constant η is the intrinsic impedance which for free space is approximately 377 Ω. The functions $Ci(x)$ and

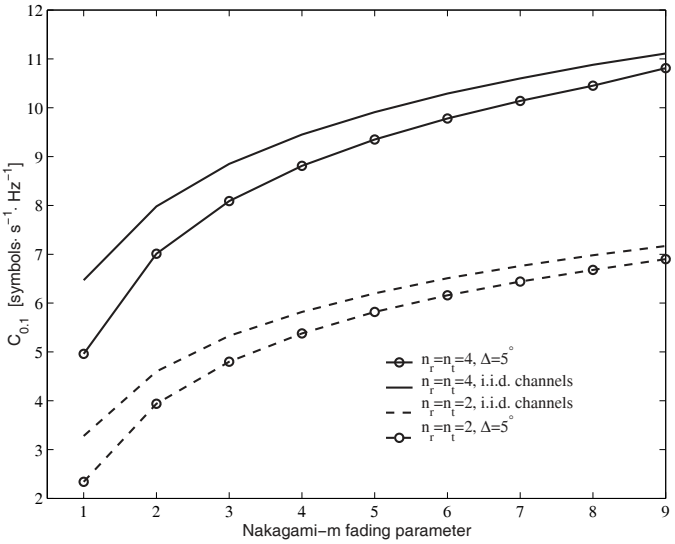


Figure 3.18: $C_{0.1}$ outage capacity versus Nakagami-m fading parameter. The mutual orientation of the TX and RX arrays is parallel and the antenna spacing at the transmitter is $\delta_{TX} = 2\lambda_c$ and at the receiver $\delta_{RX} = 0.5\lambda_c$. Δ is the angular spread, see Figure 3.1.

$Si(x)$ are the cosine and sine integrals

$$Ci(x) = -\int_x^\infty \frac{\cos y}{y} dy \tag{3.79}$$

$$Si(x) = \int_0^x \frac{\sin y}{y} dy \ . \tag{3.80}$$

Dipole antenna elements are assumed in the following analysis and due to their omnidirectional radiation properties (in the broadside plane), the coupling will be strong. The coupling for e.g. patch antennas are normally less, because of their directive property.

In Figure 3.19 the outage capacity is shown as a function of receiver antenna element separation for a 2×2 MIMO system in a Rayleigh fading environment, using the one-ring model to generate spatial correlation. In the calculations, the $\mathbf{R}_H$ matrix was normalized to $Tr(\mathbf{R}_H) = n_r n_t$ to make the comparison fair. In the i.i.d. case, where the channel matrix is given by $\mathbf{H}_w$ in Definition 3.2 on page 81, the mutual coupling can only degrade the outage capacity, especially at strong mutual coupling, which is present when the element separation on the receiver side $\delta_{RX} < 0.5\lambda_c$. On the other hand, when the spatial correlation is strong, at $\Delta = 6°$, the mutual coupling actually improves the outage capacity at small

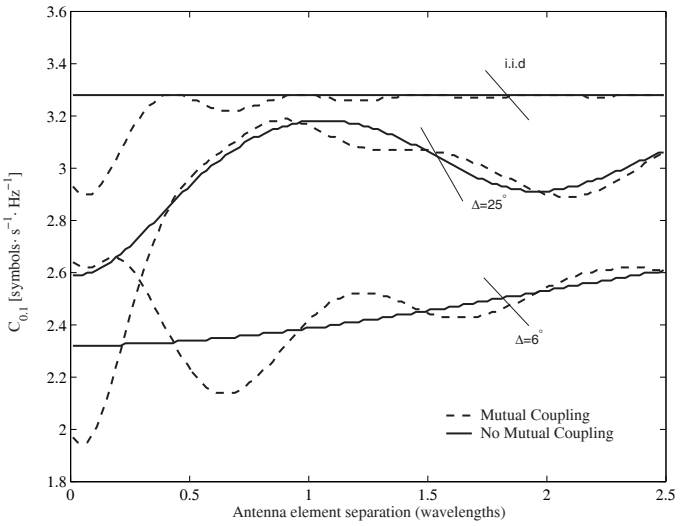


Figure 3.19: $C_{0.1}$ outage capacity versus receiver antenna element separation δ_{RX}. The number of antenna elements are $n_r = n_t = 2$ and the mutual orientation of the TX and RX arrays is parallel. The Nakagami-m parameter is $m_f = 1$, i.e. Rayleigh fading. The antenna spacing at the transmitter $\delta_{TX} = 2\lambda_c$ and Δ is the angular spread, see Figure 3.1.

element spacing, as compared to a system without mutual coupling. The reason is that the mutual coupling de-correlates the signals and in the case where the spatial correlation is high, a few eigenvalues of $\mathbf{R}_H$ is dominating and the mutual coupling will actually make the eigenvalues of $\mathbf{R}_H$ "more equal" in magnitude. Figure 3.20 shows this further by plotting the outage capacity when the angular spread is varied for a small ($\delta_{RX} = 0.2\lambda_c$) receive antenna element spacing, as might be necessary in a small portable handset. When the angular spread is less than $\Delta = 8°$, corresponding to a highly correlated fading between different channel paths, or mathematically, only one or two dominating eigenvalues of the $\mathbf{R}_H$-matrix, the mutual coupling dominates over the spatial correlation and improves the outage capacity. But when Δ is increased, the mutual coupling degrades the outage capacity. Results similar to this have recently been reported in [83] where the BER of a Nakagami fading channel actually improved when the mutual coupling increased and also in [109] where an increase of ergodic capacity of a MIMO system was observed when mutual coupling was introduced.

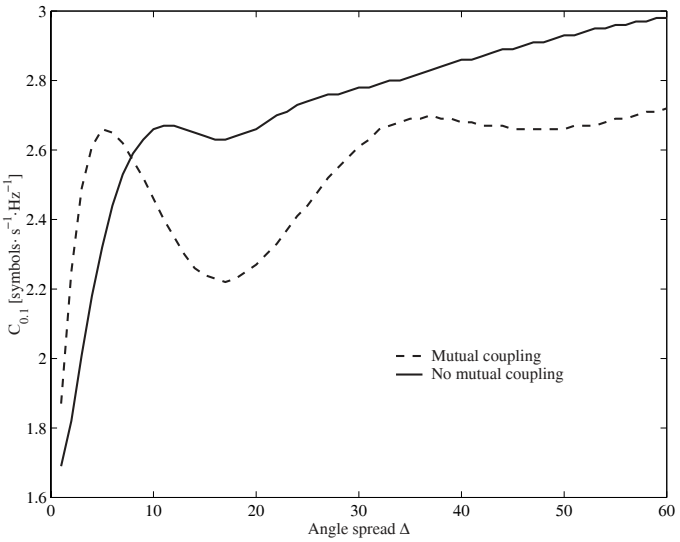


Figure 3.20: $C_{0.1}$ outage capacity for parallel linear arrays versus different angular spread Δ. The number of antenna elements are $n_r = n_t = 2$ and the Nakagami-m parameter is $m_f = 1$, i.e. Rayleigh fading. The antenna spacing at the transmitter $\delta_{TX} = 2\lambda_c$ and at the receiver $\delta_{RX} = 0.2\lambda_c$.

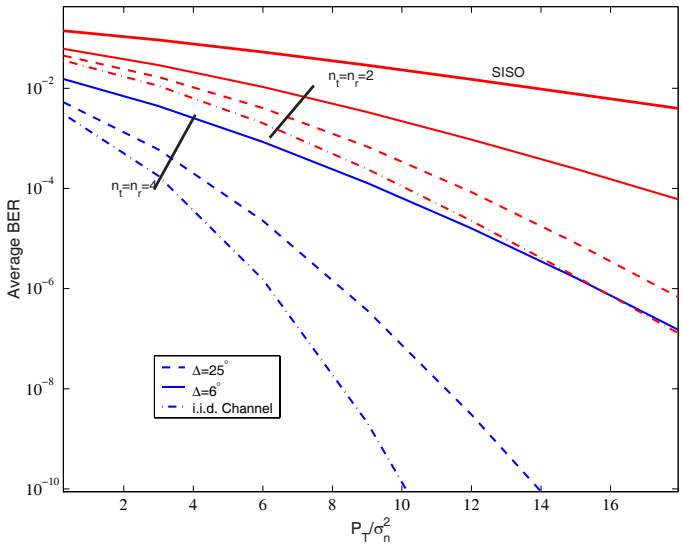


Figure 3.21: BER for STBC in a Rayleigh fading MIMO channel ($m_f = 1$) as a function of SNR P_T/σ_n^2 in decibel. The antenna spacing at the parallel transmitter and receiver arrays are $\delta_{TX} = 2\lambda_c$ and $\delta_{RX} = 0.5\lambda_c$ respectively and the calculations include mutual coupling.

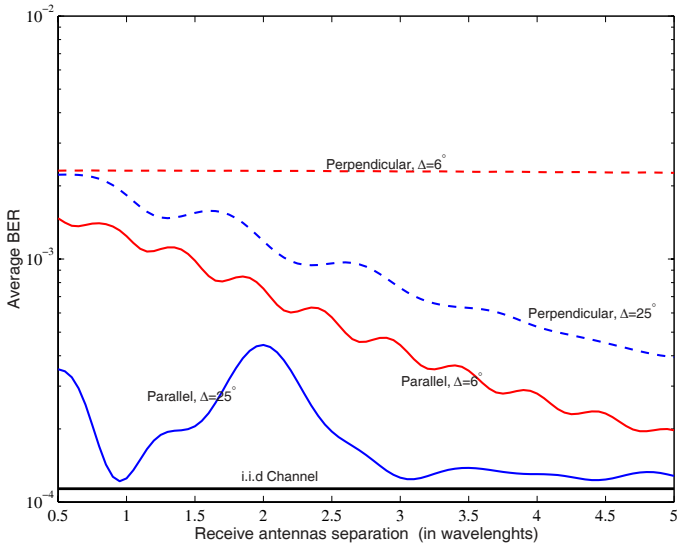


Figure 3.22: BER for STBC in a 2×2 Rayleigh fading MIMO channel ($m_f = 1$) as a function of the receive antenna spacing δ_{RX}. The antenna spacing at the transmitter is $\delta_{TX} = 2\lambda_c$. Δ is the angular spread and $P_T/\sigma_n^2 = 10$.

BER calculations

In Figure 3.21, the calculated BER is shown as a function of SNR for a 4×4 and a 2×2 system in a Rayleigh fading channel with dipole antennas that introduce mutual coupling. At BER=10^{-3}, the loss for the 2×2 channel is 1.3 dB in SNR between the i.i.d. channel and when the one-ring scattering model is introduced with $\Delta = 25°$. This is the penalty the correlated signals impose on the STBC. The loss in the 4×4 case is approximately the same but in the $\Delta = 6°$ channel, the signals are highly correlated and the diversity advantage, which is equal to the slope of the BER curves (equation (1.24)) in the 4×4 case is approximately equal to the 2×2 curves. This indicates that correlation implies that the STBC looses its diversity advantage, as expected.

Figure 3.22 shows the calculated BER using (3.66) for a 2×2 Rayleigh fading MIMO channel ($m_f = 1$) as a function of the receive antenna spacing δ_{RX}. Mutual coupling using the dipole model is used. The orientation of the receive array is varied between perpendicular and parallel to the transmitter array. For low angular spread and the arrays perpendicular to each other, the received signals have high correlation independently of the antenna separation due to the small antenna aperture. Hence, the BER decreases very slowly when increasing δ_{RX}. The lowest BER is achieved by the largest angular spread and parallel array, since this scenario

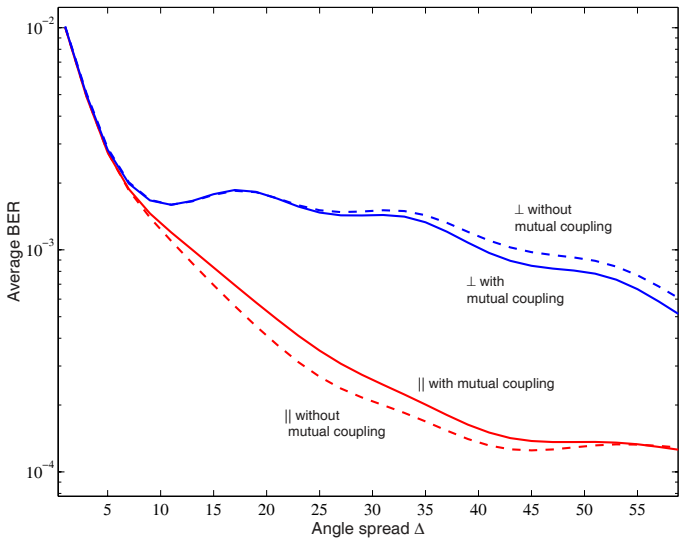


Figure 3.23: BER for STBC in a 2×2 Rayleigh fading MIMO channel ($m_f = 1$) as a function of angular spread Δ. The antenna spacing at the transmitter and receiver is $\delta_{TX} = 2\lambda_c$ and $\delta_{RX} = 0.5\lambda_c$ respectively and $P_T/\sigma_n^2 = 10$.

gives the lowest signal correlation. However, when $\delta_{TX} = \delta_{RX}$ the BER increases. An explanation is a symmetry in the one-ring scattering model which gives signal cancellations that reduces the received SNR. Hence, the peak should be attributed to an anomaly in the underlying model. The oscillations in the BER curves are due to the coherence and incoherence of the mutual coupling contribution.

Figure 3.23 shows the dependence on the bit error rate for the 2×2 Rayleigh fading MIMO channel ($m_f = 1$) as a function of the angular spread Δ. First, notice that mutual coupling reduces BER in the perpendicular receive array case. When the arrays are perpendicular, the signals are highly correlated and the mutual coupling decouples the signals somewhat, but the effect on the BER is very small.

3.5.6 Conclusions of STBC performance in correlated Nakagami channels

The characteristic function of a sum of correlated random variables was used to find the outage capacity of a coherent STBC system in a Nakagami-m flat fading environment. The derived theory gives the outage capacity as a sum of incomplete gamma functions, weighted by coefficients that depends on the correlation between the elements in the channel matrix. Hence, given measurements or simulations of the antenna element mutual coupling and a model for the spatial correlation due to multipath scattering, it is possible to calculate the effects on the outage capacity of

the STBC.

The numerical examples clearly shows the effect of mutual coupling on the performance. Mutual coupling degrades performance when the spatial correlation is low, i.e. when the channel fading is close to independent, but actually increases performance for a two element receive array in an environment with high spatial correlation, i.e. in the case of a line of sight channel. It is hard to utilize this channel capacity improvement in the general case because the spatial correlation is a random variable. A more important factor is the orientation of the array, which can, in the worst case degrade a rich scattering channel into a low rank, line of sight channel with poor performance. This can be mitigated by using several antennas and use only the best subset of them [112, 113], or to place the antennas so the array performance becomes robust to array orientation [88, 114].

It was also noticed how the performance increased with increasing m_f, due to a reduced probability of simultaneous fades. The channel thus becomes more robust for increasing m_f. However, often channels with high m_f correspond to channels with line of sight between transmitter and receive antenna array, and in these cases it is expected that the spatial correlation increases. More work is needed on a model for spatial correlation in multi antenna Nakagami fading channels. Further work should also investigate the case of non-integer m_f and a covariance matrix $\mathbf{R}_H$ with eigenvalues of (algebraic) multiplicity larger than one, which is not covered in this thesis.

3.6 Transmission over a Rayleigh fading channel with CSI at the transmitter

In Section 2.4, a novel method to transmit over a MIMO channel when CSI is available at the transmitter was derived, denoted the eigenmode transmission. It relied on the singular value decomposition of the channel matrix and was inspired by information theoretic results. As the subchannel gains vary, the transmitter adapts the modulation rate and power allocation to each of the subchannels to satisfy a chosen goal. The goals are, as in Section 2.4, set to either maximize the throughput (total bit rate) at a target BER or to minimize the BER at a target throughput.

In this section, the pdf of the subchannel gains, given by Theorem 3.1 on page 85, are used to investigate the performance of eigenmode transmission in the special case of a Rayleigh fading channel. In [115], the MMSE and symbol error rate, of a MIMO system in a Rayleigh fading channel, was investigated in the asymptotic case of an infinite number of transmit and receive antennas. Here, the analysis cover the case with a finite, and small number of antennas which often is the case in practically realizable MIMO systems.

3.6.1 Maximizing the spectral efficiency at a target BER

Under the Rayleigh fading assumption, the marginal pdf for each of the eigenvalues to the matrix $\mathbf{H}_w\mathbf{H}_w^*$ is given by Theorem 3.1. The throughput maximization problem at a given target BER under average power constraints was derived in Section 2.4.2. The result was summarized in Theorem 2.1 on page 51. Equation (2.82) gives the power control law where the constant μ is found by solving the average power constraint

$$\sum_{i=1}^{m} \int_0^\infty \left[\frac{1}{\mu} - \frac{1}{K_o\lambda_i}\right]^+ p_i(\lambda_i)d\lambda_i = P_T \ . \tag{3.81}$$

For a given μ (which is fixed when target BER and average transmit power is fixed), the probability that subchannel i is *not* used is equivalent that the power allocated to subchannel i is zero. Hence, from (2.82)

$$\gamma_i = [1/\mu - 1/K_o\lambda_i]^+ = 0 \tag{3.82}$$

and the probability that this occurs can be calculated as

$$\Pr\left\{\frac{1}{\mu} - \frac{1}{K_o\lambda_i} < 0\right\} = \Pr\left\{\lambda_i < \frac{\mu}{K_o}\right\} = \int_0^{\mu(K_o)/K_o} p_i(\lambda_i)d\lambda_i \tag{3.83}$$

where in the last step, the dependence of μ on K_o has been emphasized. This is now illustrated by an example.

EXAMPLE 3.2

Assume that $n_r = n_t = 2$ and that the channel is i.i.d. Rayleigh fading according to Definition 3.2 on page 81. The noise power is set to $\sigma_n^2 = 1$ without loss of generality. The pdf for λ_1 and λ_2 where $\lambda_1 \geq \lambda_2$ is calculated using Theorem 3.1 on page 85 and is found to be

$$p_1(\lambda_1) = (2 - 2\lambda_1 + \lambda_1^2)e^{-\lambda_1} - 2e^{-2\lambda_1} \tag{3.84}$$

$$p_2(\lambda_2) = 2e^{-2\lambda_2} \ . \tag{3.85}$$

Inserting the pdf:s $p_1(\lambda_1)$ and $p_2(\lambda_2)$ into (3.81) gives the following equation for μ

$$(\frac{2}{\mu} + \frac{1}{K_o})\exp\left(-\frac{\mu}{K_o}\right) - \frac{2}{K_o}\mathrm{Ei}\left(1, \frac{\mu}{K_o}\right) = P_T \tag{3.86}$$

where $\mathrm{Ei}(n, x) = \int_1^\infty t^{-n}e^{-xt}dt$ is the exponential integral. In Figure 3.24, the

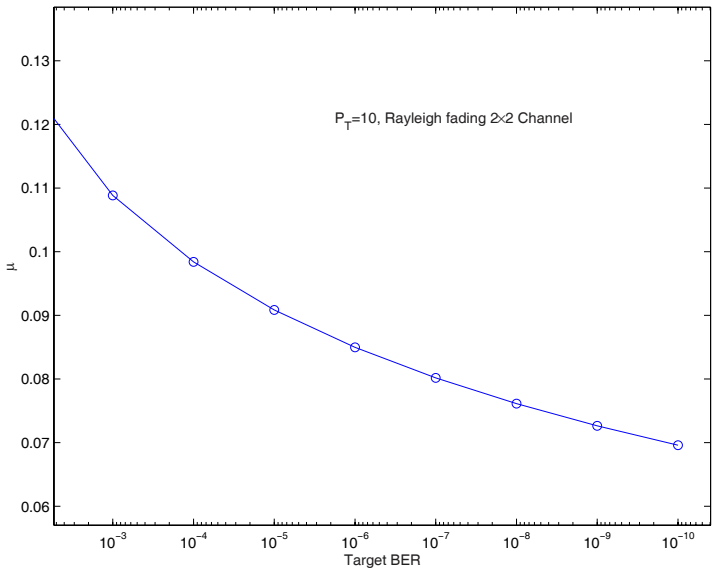


Figure 3.24: The variation of μ as a function of target BER setting. The average power $P_T = 10$.

solution for μ is shown as a function of target bit error rate for an average power of $P_T = 10$. Note that for lower target BER, the constant μ becomes smaller so it is more likely that only one mode is used or perhaps none of them. The probability that the channel is not used at all can be calculated from (3.83)

$$\Pr\left\{\frac{1}{\mu} - \frac{1}{K_o\lambda_1} < 0\right\} = 1 - e^{-\mu/K_o}\left(2 + (\mu/K_o)^2\right) + e^{-2\mu/K_o} \tag{3.87}$$

and the probability that the mode with lowest gain (λ_2) is not used as

$$\Pr\left\{\frac{1}{\mu} - \frac{1}{K_o\lambda_2} < 0\right\} = 1 - e^{-2\mu/K_o} \quad . \tag{3.88}$$

Figure 3.25 shows these probabilities for this example. These probabilities can also be interpreted as the amount of time the transmission is idle. If the target BER is very low (10^{-10}), then it is seen in Figure 3.25 that the transmission is idle for 2.5% of the time. This level is of course dependent on the available transmit power. In Figure 3.26 the corresponding spectral efficiency (2.84) is plotted, here given by equation

$$\frac{R}{W} = \left(\frac{\mu}{K_o} + 1\right) e^{-\mu/K_o} + 2\mathrm{Ei}\left(1, \frac{\mu}{K_o}\right) \quad . \tag{3.89}$$

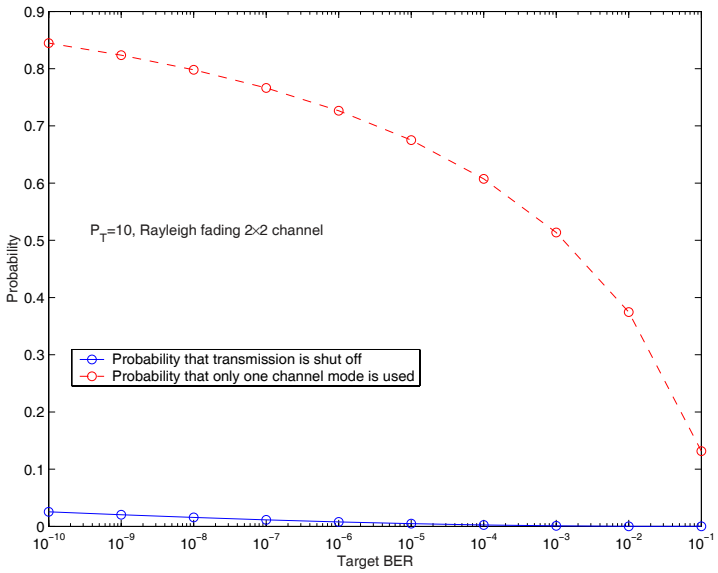


Figure 3.25: Probability that no or only one channel eigenmode is used. Average power is $P_T = 10$.

This example clearly shows the tradeoff between target BER and throughput of the MIMO system. When the target BER is decreased, μ must be increased to maintain the data throughput of the system. An increase in μ is obtained by increasing the transmit power P_T.

3.6.2 Minimizing BER at given throughput

The results from Section 2.4.3, summarized in Theorem 2.2 on page 58, are now used to analyze the eigenmode transmission in the special case of an i.i.d. Rayleigh fading MIMO channel. The aim is to minimize the BER of the transmission while maintaining a specified spectral efficiency of R_T/W bits/seconds per Hz of transmission bandwidth. The power control law is given by (2.95) which together with the power constraints (2.100) and (2.101) gives the Lagrange multiplier μ_1. The difference between the peak power and average power constraints is illustrated by a 2×2 MIMO example. Furthermore, a restricted constellation set is assumed, that is, an integer number of bits are allocated per symbol, so the calculations is performed over all possible set of bits per symbol $\{r_i\}_{i=1}^m$ that meet the target throughput $\sum r_i = R_T$. When the system is implemented, an algorithm selects the

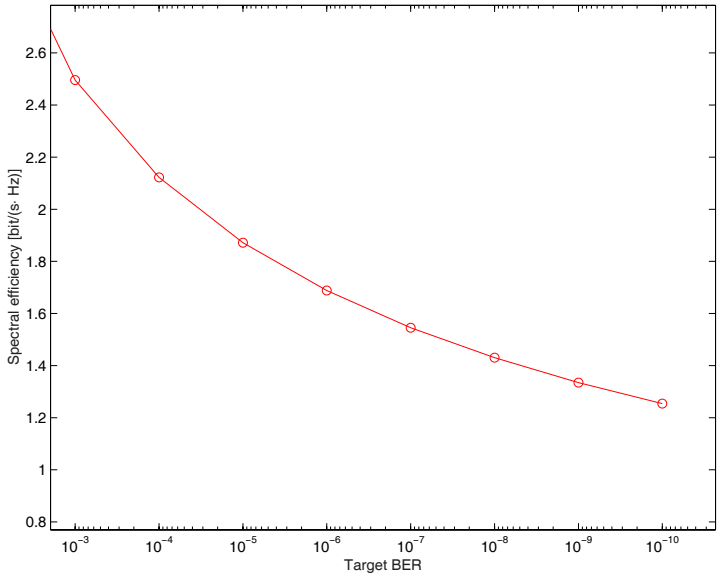


Figure 3.26: Spectral efficiency for a 2×2 MIMO system in an i.i.d. Rayleigh fading channel as a function of target BER. Average power is $P_T = 10$.

rate assignment that minimizes the BER depending on the required throughput.

EXAMPLE 3.3

Assume that $n_r = n_t = 2$ and that the channel is i.i.d. Rayleigh fading according to Definition 3.2. The noise power is set to $\sigma_n^2 = 1$ without loss of generality. The pdf for λ_1 and λ_2 where $\lambda_1 \geq \lambda_2$ is calculated using Theorem 3.1 and is found to be

$$\begin{aligned} p_1(\lambda_1) =& (2 - 2\lambda_1 + \lambda_1^2)e^{-\lambda_1} - 2e^{-2\lambda_1} \\ p_2(\lambda_2) =& 2e^{-2\lambda_2} \quad . \end{aligned} \tag{3.90}$$

By inserting into the pdf:s $p_1(\lambda_1)$ and $p_2(\lambda_2)$ into (2.100) an equation for μ_1 is obtained that can be solved numerically, based on subchannel rates r_1 and r_2 that fulfill the target throughput $r_1 + r_2 = R_T$. With this μ_1, the power allocations for the two subchannels γ_1 and γ_2 can be calculated. To calculate the resulting average BER from (2.94) does not easily lend itself to an analytical solution since γ_i depends on the eigenvalues λ_i and also the average power P_T through μ_1. Hence, this integral is solved by Monte Carlo integration by first generating a series of i.i.d. Rayleigh MIMO channel matrices and their corresponding eigenvalues. Then, the Lagrange multiplier μ_1 is calculated analytically from

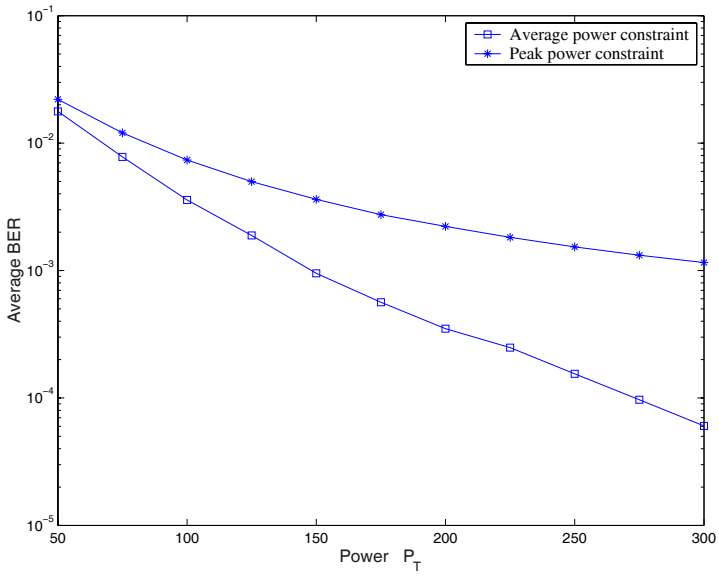


Figure 3.27: Average bit error rate versus total transmission power P_T for a 2×2 MIMO system in i.i.d. Rayleigh fading channel. Total rate target $R_T = 8$ bps/Hz. For all power levels in the shown range, the optimal (discrete) rate allocation is $r_1 = 6$ bps/Hz and $r_2 = 2$ bps/Hz.

(2.100) and then the BER is obtained from (2.78). This is repeated 1000 times and the resulting average BER is plotted in Figure 3.27 for $R_T = 8$ bps/Hz while varying the total transmitted power P_T. Both the peak constrained power and the average constrained power is shown. Given a certain rate allocation (r_1, r_2) and transmit power P_T, there is a probability that the power of one of the subchannels is zero. This is denoted as a rate allocation failure and the corresponding probability for this to occur is estimated in the $P_T = 50$ case and average power constraint to 5% and under peak power constraint to 3%. When the power is increased to 100, these failure rates has decreased to 1 and 0.5 $\%$ respectively. For even larger P_T, these rate allocation failures becomes negligible. However, these cases are removed from the calculation of the BER. This also implies that the actual achievable bit rate is less then 8 bps/Hz at low transmit power. First, note that the average power constraint always outperforms the peak power constraints in terms of BER. The reason is that in the average power constraint case the algorithm can economize with the transmitter power and utilize a higher output power than in the peak power constrained case when the channel is bad.

For all shown points in Figure 3.27, the rate allocation that minimizes the average BER is the $(r_1, r_2) = (6, 2)$ allocation (out of the possible (4,4), (6,2),

(2,6), (8,0), (0,8)). Hence, the algorithm balances the rate allocation to the average gain in the channel, since, from (3.90), we can calculate the expected values of the channel gains

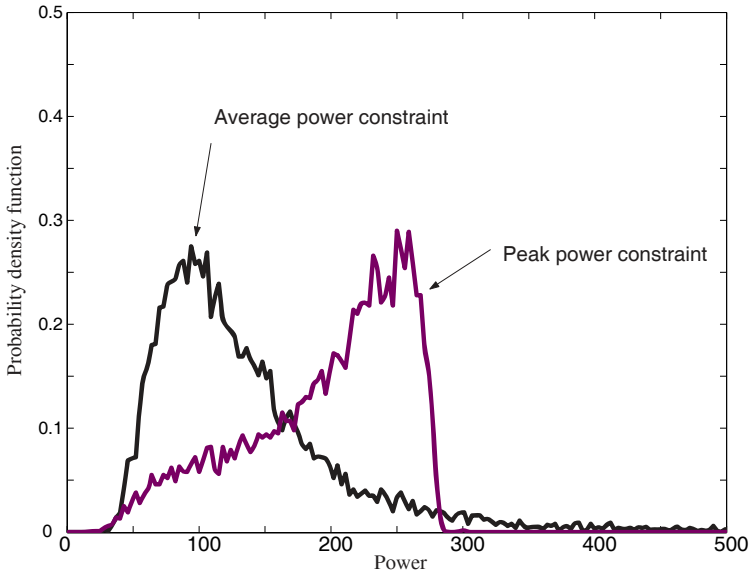
$$\begin{aligned} E\{\lambda_1\} &= 3.5 \\ E\{\lambda_2\} &= 0.5 \ . \end{aligned} \tag{3.91}$$

Thus, the allocation (6,2) matches the average subchannel gains better than the (4,4) allocation. Figure 3.28 shows estimated pdf:s on how the power is distributed between the two subchannels in the peak and average power constrained case for the (6,2) subchannel rate allocation and P_T=300. The most notable difference is in the power allocation for the subchannel with the largest path gain. The average power allocation algorithm is much more conservative with the use of "power" and sometimes peaks with very large power allocations, to power equalize deeply faded channels (the maximum value of γ_1 in this average power case was found to be $P_T > 1000$ in these simulations). Furthermore, in this example, the subchannel with the highest gain, is allocated the most power since it uses the modulation alphabet with the largest constellation and thus require a high SNR.

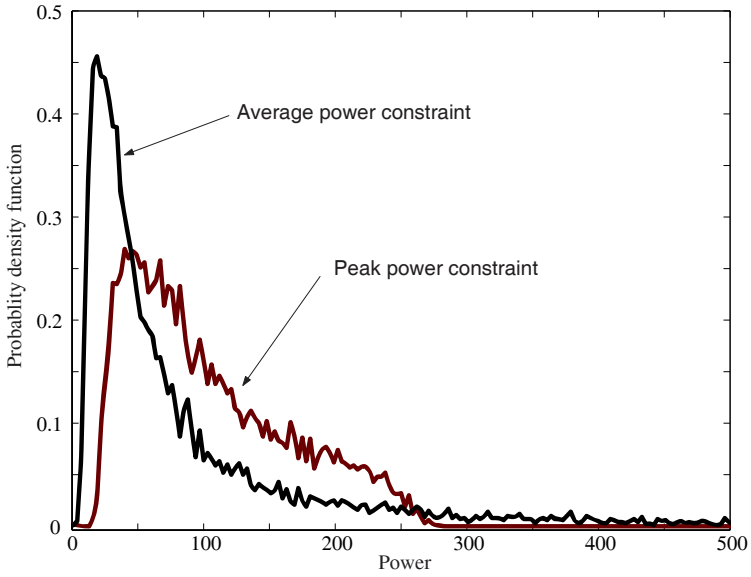
In conclusion, the average power constrained algorithm performs on average better than the peak power constrained algorithm. It has however some drawbacks. Most importantly, all amplifiers have a peak power rating, and if the transmit amplifier must be capable to sometimes transmit with 4-5 times the average output power, it must be designed to operate with a very large back-off from its saturation point. This leads to an amplifier design with a very low DC power efficiency. Furthermore, the pdf:s of the subchannel gains must be known in the average power constrained case, to be able to calculate the Lagrange multiplier μ_1. These are time variant and must be estimated during operation. Hence, the peak power constrained algorithm has a simpler implementation but the price is a worse average BER performance. The transmit amplifiers in the peak power constrained case can operate with smallest possible back-off, since the maximum output power is know. For more discussions about back-off in transmit amplifiers, see Chapter 6.

3.6.3 Beamforming

In this section the performance of beamforming in the Rayleigh fading channel is investigated. Beamforming is by the definition utilization of only one of the m channel eigenmodes for transmission. From (2.110), observe that the largest eigenvalue λ_{max} to the Wishart matrix $\mathbf{W}$ is a factor in the SNR after the maximum ratio combiner in the receiver. The pdf of λ_{max} is now used to calculate the BER and



(a) The pdf of allocated subchannel transmission power, γ_1, for the strongest eigenmode channel.



(b) The pdf of allocated subchannel transmission power, γ_2, for the weakest eigenmode channel.

Figure 3.28: Comparisons of the transmission power allocation of the i.i.d. Rayleigh fading MIMO channel when $n_r = n_t = 2$, total power $P_T = 300$ and subchannel rates $(r_1, r_2) = (6, 2)$.

channel capacity of beamforming. In the case with correlated fading and mutual coupling, the pdf of λ_{max} is unavailable, so an upper and a lower bound of the BER is derived. Also the asymptotic case when the number of antennas goes to infinity is discussed. One must note that we assume that the Rayleigh fading characteristics of the channels is preserved when beamforming is applied. Beamforming will in some cases remove the effect of some scatterers and this might change the channel fading to a Ricean channel. Here it is assumed that the scattering is rich enough to preserve the Rayleigh amplitude distribution.

Bit error rates in i.i.d. Rayleigh fading channels

The BER for a multiple antenna matched filter receiver system with transmit beamforming assuming a coherent BPSK system is now summarized in this theorem.

Theorem 3.4 *Assume a i.i.d. Rayleigh fading MIMO channel, according to Definition 3.2 and that the number of antennas $n_r, n_t = 1, 2, 3$. Then the average probability of bit error, P_{BF}, using beamforming with transmit and receive weights matched to the channel, see Section 2.5, and with coherently detected BPSK modulation, or Gray coded QPSK modulation, is*

$$P_{BF} = \frac{1}{2}\left(1 - \sum_{k=1}^{m} \sqrt{\frac{\bar{\gamma}}{\bar{\gamma}+k}}\varphi_k^{(m,n)}(\bar{\gamma})\right) \tag{3.92}$$

where $\bar{\gamma} = P_T/\sigma_n^2$ and $\varphi_k^{(m,n)}(x)$ is a rational function of two polynomials, shown in Table 3.2 in Appendix 3.I for $n = \max(n_r, n_t) = 1, 2, 3$ and $m = \min(n_r, n_t) = 1, 2, 3$.

Proof: See Appendix 3.G on page 152. ■

We have not been able to prove Theorem 3.4 for arbitrary n_r, n_t nor have we found a way to directly calculate the rational functions $\varphi^{(m,n)}(x)$. The reason is our inability to express a general formula for the pdf of the largest eigenvalue for arbitrary n_r, n_t in Corollary 3.1.

Note that in the SISO case, the classical expression for BER of coherently detected BPSK modulation in a Rayleigh fading channel is obtained:

$$P_{SISO} = \frac{1}{2}\left(1 - \sqrt{\frac{\bar{\gamma}}{\bar{\gamma}+1}}\right) \tag{3.93}$$

Figure 3.29 compares the average BER for different number of transmit and receive antennas. Note that because of the assumption that the channel is known at the transmitter, there is no difference in BER for an $n_t > n_r$ or $n_t < n_r$ system as

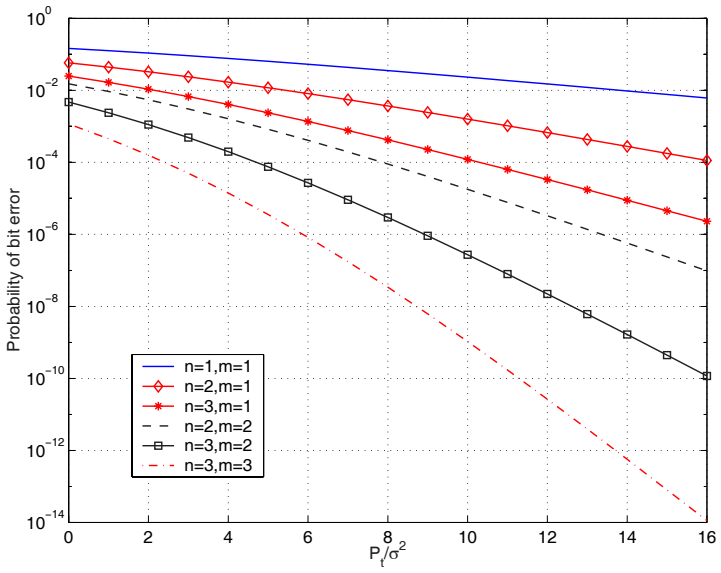


Figure 3.29: Average bit error rate versus SNR (P_T/σ_n^2) in decibel for maximum SNR receiver and coherent detected BPSK modulation.

long as n, m are fixed. It can also be noted that a (m, n)=(3,1) system has a larger BER than the (m, n)=(2,2) system, although they have the same total number of antennas. So, balancing the number of antennas between transmitter and receiver gives an advantage in terms of BER, under the given assumptions. This can be explained by the fact that a (2,2) system has four independently fading channels whereas the (3,1) system only have three. Hence, the diversity advantage for the (2,2) system is larger.

Diversity advantage

To formulate a general theorem for the diversity advantage for arbitrary n_r and n_t in a beamforming system is prohibitively difficult. The reason is linked to the fact that the pdf of the largest eigenvalue in Corollary 3.1 is not explicitly known for arbitrary n_r, n_t. For certain values of n_r and n_t it is however possible to derive tractable expressions. We illustrate this with a 2×2 MIMO and a 4×1 MISO example next.

EXAMPLE 3.4

Assume first a 2×2 MIMO system in a i.i.d. Rayleigh fading channel, i.e. the channel is $\mathbf{H}_w$ from Definition 3.2. The BER for a beamforming system is then given by equation (3.92). In the high SNR limit $P_T/\sigma_n^2 \gg 1$, it is possible to show that the BER in (3.92) is equal to

$$P_{BF}^{(2\times 2)} = \frac{35}{8}\left(\frac{P_T}{\sigma_n^2}\right)^{-4} + O\left(\left(\frac{P_T}{\sigma_n^2}\right)^{-5}\right) \quad . \tag{3.94}$$

To obtain (3.94), MAPLE 7.0 was used. Hence, a 2×2 beamforming system achieves an diversity advantage of $p = 4$ according to equation (1.24).

It is interesting to compare this result with the 4×1 MISO system (or equivalently a 1×4 SIMO system). The BER is in this case obtained as, for $P_T/\sigma_n^2 \gg 1$

$$P_{BF}^{(4\times 1)} = \frac{35}{16}\left(\frac{P_T}{\sigma_n^2}\right)^{-4} + O\left(\left(\frac{P_T}{\sigma_n^2}\right)^{-5}\right) \quad . \tag{3.95}$$

Hence, this MISO system obtains half the BER of the 2×2 system, although the diversity advantage is $n_r n_t = 4$ in both cases. The reason is that in the MISO system, the four transmitted signals are combined coherently at the receiver antenna, which leads to maximum combining gain as well as maximum diversity advantage. For the MIMO system, some combining gain is lost due to the simultaneous transmission from the transmitting array to two receiver antennas. Note however that the MISO system utilizes a total of $n_r + n_t = 5$ antenna elements and the MIMO system $n_r + n_t = 4$ antenna elements.

It seems reasonable that a diversity advantage of $nm \equiv n_r n_t$ is achieved for beamforming in a i.i.d. Rayleigh fading $n_r \times n_t$ MIMO channel. Hence, the following conjecture is stated.

Conjecture 3.2 *The diversity advantage for beamforming over an i.i.d. Rayleigh fading MIMO channel* $\mathbf{H}_w$ *according to Definition 3.2, with dimensions* $n_r \times n_t$*, is* $n_r n_t$*.*

The conjecture is supported by the discussion and example above.

Bounds on BER in correlated fading channels

If the channel coefficients are correlated, then the pdf of the largest eigenvalue is difficult to derive. The joint pdf of the eigenvalues are given by equation (3.20).

It involves hypergeometric functions of matrix arguments, which must be numerically calculated in the general case. Still, bounds on the BER performance of beamforming can be found by using the upper and lower bounds on the largest eigenvalue (3.38). It is thus possible to formulate the following Corollary.

Corollary 3.4 *The beamforming BER, stated in Theorem 3.4 is upper bounded by the orthogonal STBC BER from Theorem 3.3, in a Rayleigh fading channel with arbitrary correlation and mutual coupling. If the number of receive antennas n_r, is smaller than the number of transmit antennas n_t, then beamforming BER is strictly lower than the orthogonal STBC BER.*

Proof: See Appendix 3.H on page 152. ■

These results were also obtained independently in [116]. Furthermore, these results can also be obtained by studying the relation between orthogonal STBC SNR and beamforming SNR in equation (2.113). In conclusion, beamforming is always better, in a BER sense, than orthogonal STBC in a Rayleigh fading channel, although both methods have the diversity advantage $n_r n_t$. The beamformer uses CSI to optimally combine the transmitted signals at the receiver to maximize the SNR. In the unitary channel, and when $n_r \geq n_t$, the beamformer and STBC BER coincide. The unitary channel has no preferred "direction" that the beamformer can utilize, since all channel eigenvalues are identical.

An asymptotic result

The pdf of λ_{max} for the general n_r, n_t case is difficult to calculate but in the limit of $n_r, n_t \to \infty$ where $n_r/n_t \to \beta \geq 1$, $E\{\lambda_{max}\}$ converges to $\left(\sqrt{n_r}+\sqrt{n_t}\right)^2$, see equation (3.39). By using Jensen's inequality and the concavity of the $\log_2$ function, the ergodic capacity is *asymptotically* upper bounded by

$$\begin{aligned} C =& E\left\{\log_2\left(1+\frac{P_T\lambda_{max}}{\sigma_n^2}\right)\right\} \\ \leq& \log_2\left(1+\frac{P_T E\{\lambda_{max}\}}{\sigma_n^2}\right) \\ \to& \log_2\left(1+\frac{P_t}{\sigma_n^2}\left(\sqrt{n_r}+\sqrt{n_t}\right)^2\right) \qquad \text{when } n_r, n_t \to \infty \ . \end{aligned} \tag{3.96}$$

It is seen that this bound is maximized when $n_r = n_t$, i.e. when the distribution of the number of antennas is balanced. This supports Conjecture (3.2), since this balancing maximizes the diversity advantage. It shall be noted, as indicated in Example 3.4 and in [116], that a lower BER is obtained using beamforming, for the same diversity advantage, if all the antennas are placed on one side of the channel

(MISO/SIMO). If this is the case, the SIMO system has the advantage over MISO systems, since it has the capability to suppress interferers at the receiver, using spatial interference suppression. However, a MISO/SIMO system requires a larger total number of antennas than the balanced MIMO system, for the same diversity advantage.

Optimality of beamforming in Rayleigh fading channels

The condition for beamforming to be optimal (all eigenmodes except the principal one are in the water filling algorithm assigned zero power) in capacity sense, was given in equation (3.48) in Section 3.4.6. The probability for this to occur was given in equation (3.49) and Figure 3.30 shows this probability versus the SNR, P_T/σ_n^2. When the number of antennas is increased, it becomes less likely that the beamforming solution is equal to the Shannon capacity of the MIMO channel. The SNR on the x-axis can be interpreted as the water level. If it is increased, then it becomes more likely that more than one mode is active. The $m = 2, n = 4$ system needs a slightly higher SNR than the $m = n = 3$ system for beamforming to be optimal. So if the system is balanced with an equal number of transmit and receive antennas, then for beamforming to achieve Shannon capacity is less probable than for an unbalanced system with the same total number of antennas. For practical systems, it is common that $P_T/\sigma_n^2 > 10dB$ so only $m = 2$ systems will have a probability larger than 0.01 to achieve Shannon capacity using beamforming. Figure 3.30 shows that at $P_T/\sigma_n^2 = 0$ dB, beamforming is optimal in a 2×2 MIMO system 75% of the time. From the measurements in Figure 3.14 it can be seen that there is a negligible difference between the beamforming and full Shannon capacity for this MIMO system at $P_T/\sigma_n^2 = 0$ dB.

3.7 Improving STBC with quantized transmitter CSI

The issue of providing the transmitter with channel state information can be approached in several ways; a simple way is to design the system to use time-division duplex (TDD) with a duplex distance smaller that the channel coherence time. Hence, the uplink channel is valid as an estimate of the downlink channel and knowledge of the downlink channel is then used to optimize the downlink transmission. If a frequency division duplex (FDD) system is used (or a TDD system with duplex delay larger than the coherence time). Then the duplex distance is commonly much larger than the channel coherence bandwidth and the uplink data cannot be used to optimize the downlink transmission. Then the system must be designed with a feedback channel, as in the currently developing W-CDMA stan-

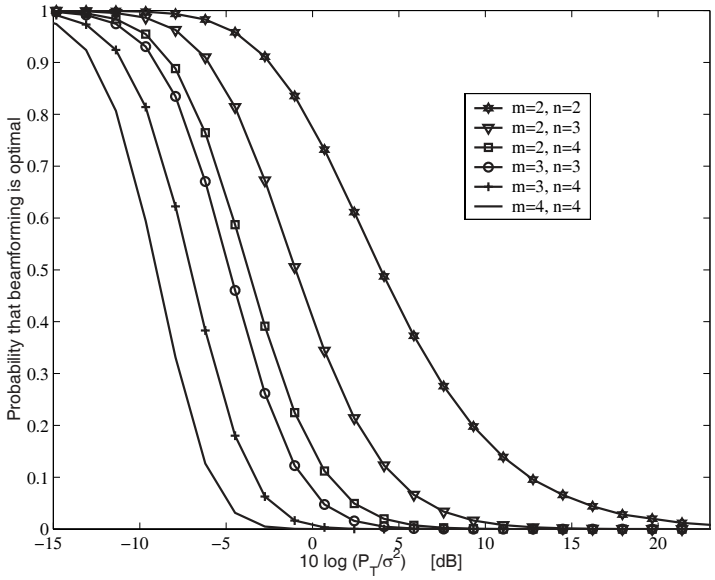


Figure 3.30: The probability that beamforming is the capacity optimal transmit strategy as a function of P_T/σ^2.

dard. The feedback signalling provides the transmitter with information which is used to minimize the bit error rate or maximize the throughput of the transmission. The feedback channel has a low bandwidth so sometimes only partial CSI can be used by the transmitter. To decide what information to send on the feedback channel for best performance is an open issue. The reliability of the CSI depends on the bit rate of the feedback channel and how fast the channel changes, i.e. the channel time-frequency product. The time-frequency product is $T_s f_D$, where T_s is the symbol time and f_D is the Doppler frequency.

In [117], partial CSI was modelled using a purely statistical approach and it was shown how a space-time code (the open-loop approach) can be improved when the transmitter has partial knowledge of the channel.

In this section, the performance of the open loop space-time block codes is compared with a feedback closed-loop transmit diversity system where the feedback bit rate is varied. Comparisons are made with a TDD system as well, which corresponds to a system with full CSI. It is of interest how the performance of these systems vary in different radio environments, and how the required bit rate of the feedback channel varies with the temporal characteristics of the channel. The effect of a line of sight (LOS) component in the channel and also the effect of a co-channel interferer (CCI) on the performance is studied. The scattering disc model described in Section 3.3.1 will be used in all cases.

3.7.1 The pairwise error probability

Assume that the noise is temporally white and colored in space. The pairwise error probability (PEP) for a STBC system using the ML-detector derived in Section 2.1 is given by equation (2.32). Begin by assuming that the transmitter has full CSI and also full knowledge of the noise covariance matrix at the receiver. Using the Chernoff bound (2.21), write the PEP as

$$Pr\left\{\mathbf{C}_k \rightarrow \mathbf{C}_l|\mathbf{H},\mathbf{R}_{vv},H_k\right\} \leq \frac{1}{2}\exp\left(-\|\mathbf{R}_{vv}^{-1/2}\mathbf{H}(\mathbf{C}_k-\mathbf{C}_l)\|_F^2/4\right) \quad (3.97)$$

where H_k denotes hypothesis k, see Section 2.1. The conditioning on the hypothesis is dropped from the following equations. Now, our aim is to minimize the error probability and to find the optimal set of codeword matrices that minimizes (3.97), given $\mathbf{R}_{vv}$ and $\mathbf{H}$. An exhaustive search over all possible codewords $\mathbf{C}_k, \mathbf{C}_l$ must then be performed. This is a prohibitively large search that has to be made for each new channel matrix $\mathbf{H}$ and noise covariance matrix $\mathbf{R}_{vv}$. Instead, the codewords are fixed to be orthogonal STBC codewords from Section 2.3.2[8]. The Euclidean distance between two orthogonal STBC codeword matrices is equal for all codeword matrix pairs. In [117], it was shown that the suboptimal transmitted codeword matrix $\mathbf{C}$ can be linearly transformed by a weighting matrix $\mathbf{J}$,

$$\tilde{\mathbf{C}} = \mathbf{JC} \quad (3.98)$$

to improve the error probability (3.97). When orthogonal STBC is used, the codeword matrices $\mathbf{C}_i$ are proportional to semi-unitary matrices[9]. Since the difference between two semi-unitary matrices is a new semi-unitary matrix, $\mathbf{C}_k - \mathbf{C}_l$ is also semi-unitary. From (2.70) it can be seen that the proportionality constant is $\sqrt{P_T/n_t}$. Now, as $\|\mathbf{AQ}\|_F^2 = \|\mathbf{A}\|_F^2$, when $\mathbf{Q}$ is unitary or semi-unitary, equation (3.97) can be rewritten in the case of orthogonal STBC with transmit weighting as

$$Pr\left\{\tilde{\mathbf{C}}_k \rightarrow \tilde{\mathbf{C}}_l|\mathbf{H},\mathbf{R}_{vv},\mathbf{J}\right\} \leq \frac{1}{2}\exp\left(-\frac{P_T}{n_t}\|\mathbf{R}_{vv}^{-1/2}\mathbf{HJ}\|_F^2/4\right) \quad (3.99)$$

which is to be minimized by proper choice of $\mathbf{J}$, under the power conservation constraint $\|\mathbf{J}\|_F^2 = 1$. The following lemma is then needed.

Lemma 3.2 *If $\mathbf{A}$ is a non-negative definite $r \times r$ matrix and $\mathbf{J}$ is an $r \times r$ matrix with the property $\|\mathbf{J}\|_F^2$=1, then $\|\mathbf{AJ}\|_F^2$ is maximized by choosing*

$$\mathbf{J} = \begin{bmatrix}\alpha_1\mathbf{a} & \alpha_2\mathbf{a} & \cdots & \alpha_r\mathbf{a}\end{bmatrix} \quad (3.100)$$

[8] Remark that the STBC is a suboptimal transmission scheme if $\mathbf{H}$ and $\mathbf{R}_{vv}$ are known at the transmitter

[9] An $n_r \times n_t$ matrix $\mathbf{A}$ is semi-unitary if $\mathbf{AA}^* = \mathbf{I}$ but $\mathbf{A}^*\mathbf{A} \neq \mathbf{I}$

where $\mathbf{a}$ *is the principal eigenvector to* $\mathbf{A}$ *and* α_i *are non-negative constants with the property* $\sum_{i=1}^{r} \alpha_i^2 = 1$.

Proof: Define $\mathbf{J} = \begin{bmatrix} \mathbf{j}_1 & \mathbf{j}_2 & \cdots & \mathbf{j}_r \end{bmatrix}$ where $\mathbf{j}_k$ is the k:th column of $\mathbf{J}$. Write

$$\begin{aligned} \|\mathbf{AJ}\|_F^2 &= \sum_{i=1}^{r} |\mathbf{A}\mathbf{j}_i|^2 \\ &= \sum_{i=1}^{r} \mathbf{j}_i^* \mathbf{A}^* \mathbf{A} \mathbf{j}_i \end{aligned} \quad . \tag{3.101}$$

Since the terms $\mathbf{j}_i^* \mathbf{A}^* \mathbf{A} \mathbf{j}_i$ are Hermitian forms, they are maximized by choosing $\mathbf{j}_i$ proportional to the principal eigenvector, $\mathbf{a}$, to the matrix $\mathbf{A}^*\mathbf{A}$. Hence, choose $\mathbf{j}_i = \alpha_i \mathbf{a}$. The maximizing value is $\lambda_{max}(\mathbf{A}^*\mathbf{A})\alpha_i^2$. With this choice, (3.101) becomes

$$\|\mathbf{AJ}\|_F^2 = \lambda_{max}(\mathbf{A}^*\mathbf{A})(\alpha_1^2 + \ldots + \alpha_r^2) \ . \tag{3.102}$$

The constraint $\|\mathbf{J}\|_F^2$=1 implies that $\sum_{i=1}^{r} \alpha_i^2 = 1$. ■

Since e^{-x} is a monotone decreasing function, equation (3.99) is minimized, by the matrix choice (choosing $\alpha_1 = 1, \alpha_i = 0 \forall i > 1$ for simplicity)

$$\mathbf{J}_o = \begin{bmatrix} \mathbf{v}_{max} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \tag{3.103}$$

according to Lemma 3.2. Here $\mathbf{v}_{max}$ is an eigenvector corresponding to the largest eigenvalue of the matrix $\mathbf{H}^H \mathbf{R}_{vv}^{-1} \mathbf{H}$.

The result (3.103) can be interpreted as beamforming using $\mathbf{v}_{max}$ as the beamforming vector [117]. Note that the codeword matrix now reduces to a rank one matrix and if binary modulation alphabet is used, the PEP and BER becomes equivalent, since only one symbol is transmitted at a time (weighted by $\mathbf{v}_{max}$). In conclusion, when the channel is perfectly known at the transmitter, beamforming is the optimal transmission strategy that minimizes the upper bound of the PEP (BER). When there is no knowledge of the channel at the transmitter, the so called open loop scenario, it was shown in [32] that $\mathbf{J} = \mathbf{I}$ minimizes (3.99) so the STBC is used without modifications.

In the example in Section 3.6.2, the BER of a MIMO system was minimized under a power and a throughput constraint and it was seen that beamforming, i.e. utilizing only one of the channel eigenmodes did not minimize the BER. Here, the PEP under a power constraint is minimized but there is no throughput constraints and no adaptive modulation schemes and we see that under these particular conditions, beamforming will always minimize the PEP when full CSI is available at the transmitter. Hence, the throughput constraint forces the use of several eigenmodes and without this constraint, the lowest possible throughput (beamforming) is obtained and the SNR at the receiver os maximized (also obtained by beamforming).

Feedback of partial CSI

Assume that the feedback channel has a capacity of R_{fb} bits per second and that the number of transmit antennas $n_t = 2$. Furthermore, CSI and the covariance matrix of the noise plus interference are assumed to be perfectly estimated at the receiver. The receiver then calculates the eigenvector $\mathbf{v}_{max}$, in (3.103), corresponding to the largest eigenvalue. Because of limited feedback rate, the exact weight vector $\mathbf{v}_{max}$ with unlimited precision can not be fed back to the transmitter. Instead,

$$\mathbf{v'}_{max}(n) = \begin{bmatrix} 1 \\ e^{j\theta(n)} \end{bmatrix} \frac{1}{\sqrt{2}} \tag{3.104}$$

is used, where $\theta(n)$ is a uniform quantization of the ideal value of the difference in the arguments of the elements of $\mathbf{v}_{max}$. Now, only $\theta(n)$ is sent to the transmitter through the feedback channel. This approach is similar to the one used in the W-CDMA standard. The transmitter averages the received feedback signals over two consecutive received bits while maintaining the transmitted power in the both antennas. This can be described as [118]

$$\theta(n) = \arg\left\{ i^{n \diamond 2} \text{sgn}\left(z(n)\right) + i^{(n-1) \diamond 2} \text{sgn}\left(z(n-1)\right) \right\} \tag{3.105}$$

where $z(n)$ is the feedback bit received at time n and $n \diamond 2$ is n taken modulus 2. The argument $\theta(n)$ has four states similar to QPSK. In [118], a more advanced method is proposed, where more than the two recent samples are used in updating $\theta(n)$. This is beneficial in environments where the fading is varying slowly so the channel is approximately constant over a longer time than it takes to receive two bits in the feedback channel. For optimal performance, the filter length should be updated adaptively to match the channel coherence time. This will not be investigated further in this thesis.

3.7.2 Simulation results

The following parameters are common for all simulations. The number of transmit and receive antennas $n_t = n_r = 2$, antenna separation transmitter is $2\lambda_c$ and at the receiver $0.5\lambda_c$. The transmitter to receiver distance is $600\lambda_c$ and the radius of the scattering disc is set to $250\lambda_c$. 30000 samples are collected in each simulation and Alamouti's scheme for the STBC described in Section 2.3.2 is used. The carrier frequency is set to 2.15 GHz.

The effect of feedback bit rate

The effect of feedback bit rate in channels with different coherence time is investigated. As suggested in [118], the system performance could be improved in slowly

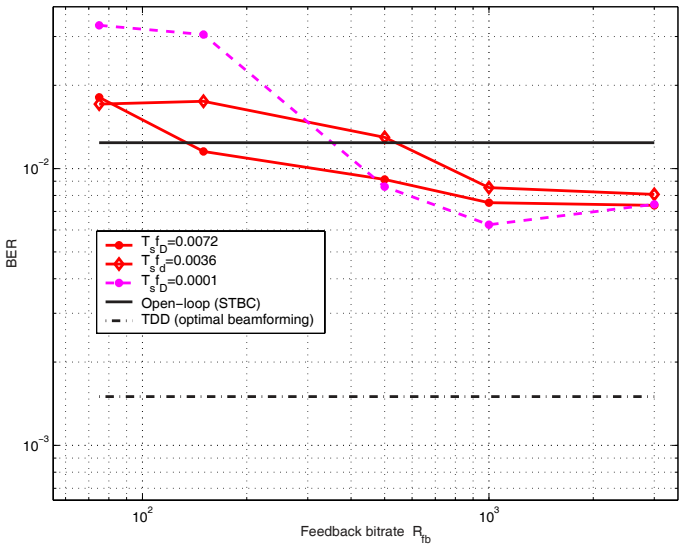


Figure 3.31: BER when bit rate of feedback channel is varied in channels with different time-frequency products. The TDD and STBC levels are also indicated. SNR=4 dB. No LOS component (K=0).

varying channels by filtering the feedback bits with a filter which length is adaptive. The length depends on the doppler frequency of the channel.

The simulations show, for BPSK modulation and a given feedback rate, at which time-frequency product the BER is higher than the open-loop, or STBC BER, i.e. when the feedback data should not be taken into consideration. In these simulations there is no co-channel interferer, hence $\mathbf{R}_{vv} = \sigma_n^2\mathbf{I}$, where σ_n^2 is the spatially and temporally white noise variance. Figure 3.31 shows how the BER decreases when the feedback bit rate increases. For the slowly fading channel, i.e. $T_sf_D = 0.0001$, the performance is better than the open-loop STBC at a modest feedback rate of 150 bps. When the channel coherence time is reduced, the feedback rate has to be increased to perform better than the open loop STBC. In the figure the TDD mode is also shown, where it is assumed that the channel state information is perfectly known at the transmitter and hence, the beamforming weights are optimal. It is also noticeable that for low feedback rates, the fast varying channel ($T_sf_D = 0.0072$) shows best performance. An explanation is that the quantization error of the vector $\mathbf{v}_{max}$ is large, and if the channel is slowly varying, then these errors will be large for a longer period of time than in the fast varying channel.

In Figure 3.32, there exists a strong line of sight component between the base-station and the mobile with Rice factor $K = 10000$, see (3.10). In this case, a low

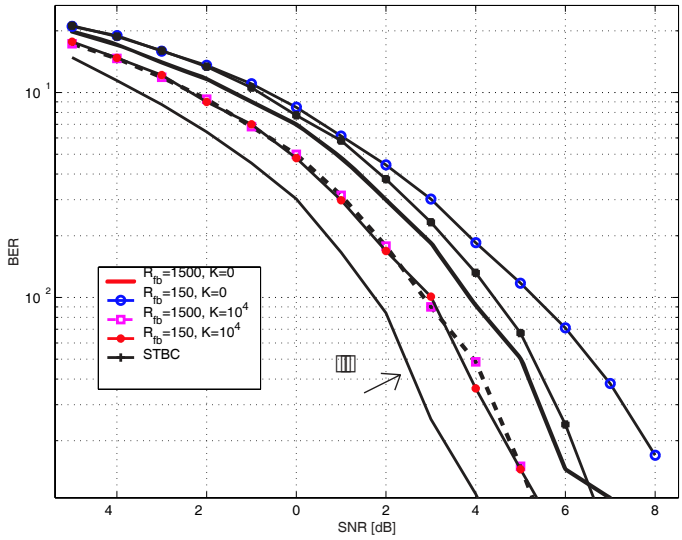


Figure 3.32: BER as a function of SNR per receive antenna. The LOS component is varied and also the feedback bit rate. Note that in a strong LOS environment, there is no improvement by increasing the feedback bit rate above 150 bps. In the figure, $T_s f_D = 0.0036$. The arrow points at the curve for the TDD case.

feedback bit rate will be sufficient as the channel has a clear directional property. Once the direction to the mobile has been accommodated, the transmission can lock on to the mobile and beam-form in that direction. Hence, there is no need for a high bit rate feedback channel.

Effects of co-channel interference

Another MS is now added at the location $800\lambda_c$ from the primary base station. Assume that this MS transmits at the same physical channel as our mobile under study. This implies that the covariance matrix of the received noise vector $\mathbf{R}_{vv}$ is a full rank matrix. We want to investigate the effect of taking the non-spatially white noise into account before transmission, i.e. when calculating $\mathbf{J}$ using equation (3.99). The channel is generated using the same model as above, and the noise covariance matrix is estimated using the received data.

A semi-analytical approach is used to find the upper bound of the error probability using (3.99). The resulting BER is compared with the BER calculated using $\mathbf{R}_{vv} = \sigma_n^2\mathbf{I}$ in (3.99), that is, by neglecting the information about the colored noise. In Figure 3.33 the "instantaneous" BER averaged over 3 seconds is shown as a function of SNR when SIR is held at 20 dB. The use of the correct covariance matrix corresponds to an equivalent 2.2 dB gain in SNR at BER=10^{-3} and the use

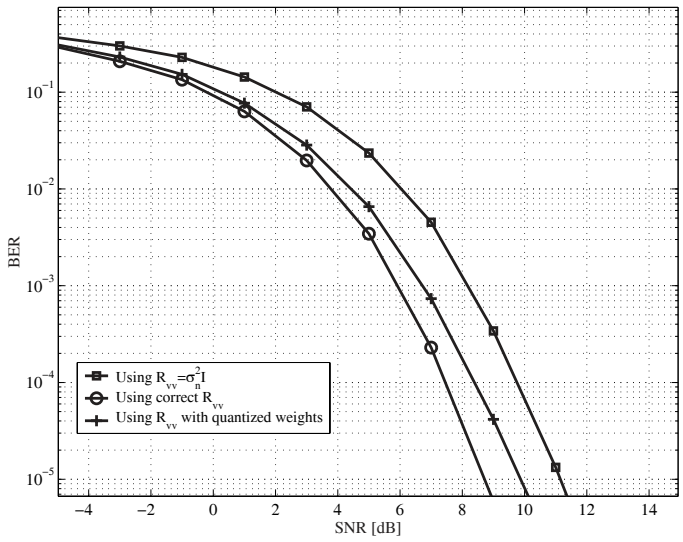


Figure 3.33: BER as a function of SNR with and without the use of the information about the covariance matrix of the spatially non-white noise

of quantized phase reduces this gain to 1.4 dB, still an improvement.

Discussion

It has been shown how the feedback information (CSI+$\mathbf{R}_{vv}$) in a transmit diversity system can be utilized to better deal with non-white noise or interferers at the receiver array. This is a common situation in systems with co-channel interference. It was shown that in a spatially colored, but temporally white noise environment, 2.2 dB could be gained in SNR for a 2×2 MIMO system, when a 20 dB weaker interfering base station is present. A simple feedback scheme, similar to the scheme in W-CDMA was compared to an open-loop solution, using space time block codes and it was shown how the required feedback bit rate depends on the channel time-frequency product. In LOS channels, the demands on the feedback channel capacity becomes smaller. In future work, the feedback signalling can be made adaptive, to provide the transmitter with the best possible information, independent on the mobility of the mobile user. Further studies should also deal with more than 2 transmitting antennas, as the demands on the feedback bit rate then increases.

3.8 Concluding remarks

This chapter is finalized by a concluding discussion.

Coherence time, CSI and the feedback rate

Full CSI at the transmitter always gives an advantage over open loop transmit diversity systems. The advantage increases with an increasing number of transmit and receive antennas and with decreasing SNR. The advantage is also larger in LOS channels where a dominating principal channel eigenvalue exists compared to NLOS channels. However, one must remember that more antennas and lower SNR requires more and longer pilot sequences to estimate all the parameters of the channel matrix. Hence, a larger system also requires a channel with longer coherence time, to be able to benefit from CSI at the transmitter, if the bitrate on the feedback channel is constant.

A channel with short coherence time requires a high rate feedback channel which would consume a large portion of the available bandwidth. In this scenario, transmit diversity, which is an open loop method, gives comparable performance to closed loop methods at no overhead signalling cost, especially in relatively small systems (2-3 transmit antennas) and in NLOS channels at SNR$>$10 dB.

In Section 3.7 it was shown that there exists a threshold of the channel time-frequency product when an open-loop system outperforms the closed loop system, given a certain feedback rate. This threshold was further explored in [38] where it was given analytically. In [38] it was concluded that if n_t is increased but n_r is fixed, then the gain of closed loop transmission increases over open loop transmission and alternatively, if n_t is fixed and n_r is increased, then the gain decreases. Recently, some interesting methods have been proposed to improve initially open-loop methods with a partial channel feedback. In [41], the authors use a feedback channel to switch the transmission between STBC and spatial multiplexing transmission based on the minimum Euclidian distance of the receive constellation. The gain is however small, about 1 dB in SNR and the spatial multiplexing suffers from the constraint that $n_r \geq n_t$. In [119], STBC and adaptive modulation are combined. Since the STBC transforms the MIMO channel to an equivalent SISO channel, the Frobenius norm of the channel matrix is a sufficient statistic used to adapt the modulation constellation. However, STBC makes the equivalent SISO channel "harder", meaning that the variance in the SNR reduces. This will reduce the gain of using adaptive modulation.

Mutual coupling and the robustness of STBC

The mutual coupling between dipole antenna elements has only a small effect on the outage capacity of the orthogonal STBC MIMO system unless the antenna element spacing is reduced below $\lambda_c/2$ or if the angular spread of the received signal components is small. This shows the extreme robustness to channel conditions of the orthogonal STBC. The STBC is also of very low decoding complexity since the ML-detector is linear and also channel estimation is trivial due to the orthogonal properties of the space time block code. The penalty is the loss in channel capacity compared to the optimal open loop capacity. Only in the MISO transmit diversity case, orthogonal STBC does not destroy the possibilities to achieve the Shannon capacity limit. In a channel which has a strong LOS component, the channel capacity after using STBC is close to the Shannon limit since the channel matrix can be well approximated by a rank-one matrix. Again, the robustness of STBC is demonstrated as it works well over LOS and NLOS channels compared to spatial multiplexing which require a channel with rich scattering and which gives low correlation between the different paths.

Validation of the i.i.d. Rayleigh fading assumption

Through measurements it was demonstrated that the assumption of i.i.d. Rayleigh fading amplitudes and uniformly distributed phases of the channel matrix elements is valid in a NLOS scenario. The paths had a maximum envelope correlation coefficient around 0.5 and the amplitudes fitted to a Nakagami-m distribution with $m_f = 1.07$ where $m_f = 1$ equals the Rayleigh distribution. Under the above assumption, the joint probability density function of the m eigenvalues to the matrix $\mathbf{W} = \mathbf{H}\mathbf{H}^*$ was found in the literature and it was possible to derive the marginal pdf for the eigenvalues. These pdf:s are useful in deriving the performance of a beamforming or a eigenmode transmission system.

Performance of closed loop systems

Using the novel expressions for the marginal pdf:s of the eigenvalues to an i.i.d. Rayleigh fading channel, the performance of the proposed eigenmode transmission algorithms was explored and compared, both in the average and the peak power constrained designs. When the average transmitted power is constrained, a smaller BER can be achieved although higher demands are put on the transmit amplifiers, since the transmitted power has a larger peak to average ratio. For high SNR, it was shown in an example that all the subchannels is assigned approximately the same power, hence a constant-power and variable rate allocation algorithm is expected to perform well in this region.

In the spectral efficiency maximization case, eigenmode transmission outperforms beamforming, unless the target BER is very low. Then beamforming spectral efficiency becomes comparable or even equal to eigenmode transmission. The usefulness of beamforming for minimizing BER or maximizing spectral efficiency should be even more pronounced in a LOS scenario, since beamforming utilizes only the principal channel eigenmode which dominates in LOS channels.

Co-channel interferers

When co-channel interferers are present in the closed loop scenario, transmitting in the subspace spanned by the interference can be avoided. This was demonstrated by the simulations in Section 3.7 and it was seen that a few dB could be gained by using this method.

Conclusions for MIMO and MISO systems

If the results from Chapter 2 and 3 are summarized, the following comprehensive list can serve as a guideline when designing MIMO and MISO systems.

No CSI + MISO system

Spatial multiplexing is not possible, since only a single data stream can be transmitted from the n_t antennas. Hence this case corresponds to beamforming with a random weight vector.

Orthogonal Space Time Block Coding can achieve the Shannon capacity in the case $n_t = 2$, where full rate codes exists in the complex modulation alphabet case. If $n_t > 2$ the capacity loss compared to the Shannon capacity is large, although if the channel is ill-conditioned (rank-deficient) as is the case in LOS channels, the loss compared to the Shannon capacity is smaller.

No CSI + MIMO system

Spatial Multiplexing can transmit $m = \min(n_r, n_t)$ independent data streams from the n_t transmitter antennas. This scheme can achieve the Shannon capacity. If $n_r > n_t$, the extra dimensions at the receiver can be used for interference suppression and/or to robustify the detector.

Orthogonal Space Time Block Coding has full diversity advantage $n_r n_t$. The capacity is strictly less than the Shannon capacity and the difference is nominally large. If the channel is ill-conditioned (rank-deficient) as is the case in LOS channels, the loss compared to the Shannon capacity is however reduced. If $n_r > n_t$, the extra dimensions at the receiver can be used for interference suppression and/or to improve the performance through an increased diversity advantage.

CSI + MISO system

Beamforming is optimal in the Shannon capacity sense. The receiver SNR is maximized and the diversity advantage is n_t. Beamforming has array gain from coherent combining of the signals from the n_t transmitter antennas at the receiver antenna.

CSI + MIMO system

Beamforming can achieve Shannon capacity in "small" systems (2×2) and for low SNR. In LOS channels, the loss to the Shannon capacity decreases compared to the NLOS case. Beamforming has diversity advantage $n_r n_t$ (conjecture). If no constraints are set on the data throughput, beamforming minimizes the BER and beamforming outperforms STBC in terms of SNR.

Eigenmode transmission has, for a target BER, a power penalty of $1/K_o$ compared to the Shannon capacity limit. The constant K_o depends on the target BER and the receiver noise variance. If the target BER decreases, the power penalty increases. Bit rate maximization is obtained by assigning the most power to the "strongest" subchannel.

Eigenmode transmission minimizes the BER at a target bit rate by assigning more power to weak subchannels (if the subchannels have the same rate), hence a subchannel gain balancing strategy.

Appendix 3.A Proof of Corollary 3.1 and a discussion of Conjecture 3.1

By applying Theorem 3.1 on page 85, $p_1(\lambda_1) \equiv p_\lambda(\lambda_{max})$ is found as the pdf for the largest eigenvalue to the Wishart matrix $\mathbf{W}$. The marginal pdf for the largest eigenvalue is found from the integral

$$p_k(\lambda_k) = \int_0^{\lambda_1} d\lambda_2 \int_0^{\lambda_2} d\lambda_3 \cdots \int_0^{\lambda_{m-1}} d\lambda_m p(\lambda_1, \ldots, \lambda_m) \ . \tag{3.106}$$

The joint pdf of the eigenvalues has the form

$$p(\lambda_1, \ldots, \lambda_m) = f_{11}(\lambda_1, \ldots, \lambda_m) e^{-\lambda_1 - \lambda_2 \cdots - \lambda_m} \ , \tag{3.107}$$

where $f_{11}(\lambda_1, \ldots, \lambda_m)$ is a polynomial in the variables $\lambda_1, \ldots, \lambda_m$ of order $n - m + 2(m-1)$. Now, from [80, Sec. 2.321], the integration formula follows

$$\begin{aligned} &\int_0^z x^n e^{ax} dx \\ &= e^{az} \left(\frac{z^n}{a} + \sum_{k=1}^{n} (-1)^k \frac{n(n-1)\cdots(n-k+1) z^{n-k}}{a^{k+1}} \right) - \frac{(-1)^n n!}{a^{n+1}} \\ &\triangleq g(z) e^{az} + c \end{aligned} \tag{3.108}$$

where $g(z)$ is a polynomial in z of order n and c is a constant. Hence, it is seen that the integration of $x^n e^{ax}$ preserve the order n of the variable x in the polynomial in front of the exponential function, plus an addition of a constant. By using (3.108) in (3.106), after the first integration, the result has the form

$$h_1(\lambda_1, \cdots, \lambda_{m-1}) e^{-\lambda_{m-1}} + c \tag{3.109}$$

where $h_1(\cdot)$ is a polynomial in which λ_1 has the order $n - m + 2(m-1)$ (since the order is preserved) and c is a constant. The integrand for the next integration has the form

$$\begin{aligned} p(\lambda_1, \ldots, \lambda_{m-1}) = &f_{21}(\lambda_1, \ldots, \lambda_{m-1}) e^{-\lambda_1 - \lambda_2 \cdots - 2\lambda_{m-1}} \\ &+ f_{22}(\lambda_1, \ldots, \lambda_{m-1}) e^{-\lambda_1 - \lambda_2 \cdots - \lambda_{m-1}} \ , \end{aligned} \tag{3.110}$$

where $f_{21}(\lambda_1, \ldots, \lambda_{m-1})$ and $f_{22}(\lambda_1, \ldots, \lambda_{m-1})$ are polynomials in the variables $\lambda_1, \ldots, \lambda_{m-1}$. After $m-1$ integrations, the result has the form

$$p_\lambda(\lambda_{max}) = \sum_{k=1}^{m} \phi_k(\lambda_{max}) e^{-k\lambda_{max}} \tag{3.111}$$

where $\phi_k(x)$ is a polynomial of maximum degree $n-m+2(m-1) = n+m-2 \equiv n_r+n_t-2$. This is found by performing the integrations and saving the monomial with the highest order in each step. Since this is not formally proven, we state this as a conjecture. Furthermore, we have not been able to find a way to directly calculate the polynomials $\phi_k(x)$, given n_r and n_t. However, by calculating the largest eigenvalue pdf for $n_r = 2, 3, 4$ and $n_t = 2, 3, 4$ we have found that all the resulting pdf:s can be described by the formula (3.111) and the polynomials $\phi_k(x)$ are given in Table 3.1.

Appendix 3.B Proof of Lemma 3.1

Write the Frobenius norm of the channel matrix $\mathbf{H}$ as

$$\Upsilon = \|\mathbf{H}\|_F^2 = \sum_{p=1}^{n_r}\sum_{q=1}^{n_t}|H_{pq}|^2 \quad . \tag{3.112}$$

The pdf $p_\Upsilon(\Upsilon)$, of Υ when the elements of the channel matrix are correlated is now sought. The problem is equivalent to finding the pdf of a maximal ratio combining diversity system in a correlated Nakagami fading channel [83]. The solution is obtained by utilizing the principles of characteristic functions of random variables. Following the method in [83], the characteristic function of the stochastic variable Υ, denoted Φ is found and it is defined as

$$\Phi(s) = E\left[e^{js\Upsilon}\right] \quad . \tag{3.113}$$

Then, the sought pdf, $p_\Upsilon(\Upsilon)$, is obtained as the inverse Laplace transform of $\Phi(s)$.

In the SISO case, if Z is a Nakagami-m distributed random variable with integer fading parameter m_f, then it can be constructed as

$$Z = \sqrt{\sum_{i=1}^{m_f}|X_i + jY_i|^2} = \sqrt{\sum_{i=1}^{m_f}X_i^2 + Y_i^2} \tag{3.114}$$

where X_i, Y_i are independent and real Gaussian random variables with zero mean and variance 1/2 [83]. Since the $|H_{pq}|$:s in (3.112) are Nakagami-m distributed, Υ can, using the relation (3.114), be rewritten as

$$\Upsilon = \sum_{p=1}^{n_r}\sum_{q=1}^{n_t}|H_{pq}|^2 = \sum_{i=1}^{n_r}\sum_{j=1}^{n_t}\sum_{l=1}^{m_f}X_{ijl}^2 + Y_{ijl}^2 \tag{3.115}$$

where X_{ijl}, Y_{ijl} are real valued Gaussian random variables with zero mean and variance 1/2. These variables are dependent, but according to definition of a Nakagami-m random variable in (3.114), the following holds

$$E\{X_{ijl}X_{pqk}\} = 0 \qquad \text{if } l = k \tag{3.116}$$
$$E\{Y_{ijl}Y_{pqk}\} = 0 \qquad \text{if } l = k \tag{3.117}$$
$$E\{X_{ijl}Y_{ijk}\} = 0 \qquad \forall l, k \ . \tag{3.118}$$

Hence, for a given index l, the real and imaginary parts of the elements in $\mathbf{H}$ are mutually correlated, excepts the real and imaginary part of a specific element in $\mathbf{H}$, which are independent. The independence between variables with different index l, implies that (3.115) can be regrouped to

$$\Upsilon = \sum_{l=1}^{m_f} \left\{ \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} X_{ijl}^2 + Y_{ijl}^2 \right\} \ . \tag{3.119}$$

Hence, the random variable Υ consists of a sum of $2m_f n_r n_t$ Gaussian random variables that can be partitioned into m_f independent groups with $2n_r n_t$ dependent elements in each group. The characteristic function for the sum of $2n_r n_t$ correlated Gaussian random variables can now be derived and the extension to a general integer m_f is a simple operation of adding m_f independent sums.

Proceed by modelling the correlation between the $2n_r n_t$ Gaussian distributed variables. Define the $n_r n_t \times 1$ vector, for a fixed l,

$$\mathbf{h}^{(l)} = \text{vec}(\mathbf{X}_{:,:,l}) + j\text{vec}(\mathbf{Y}_{:,:,l}) \tag{3.120}$$

where $\mathbf{X}_{:,:,l}$ is the matrix that is obtained for a fixed l and $i = 1, \ldots, n_r$,$j = 1, \ldots, n_t$. Consider the vector, for a fixed l,

$$\mathcal{H}_l = \left[\text{Re}(\mathbf{h}^{(l)})^T \quad \text{Im}(\mathbf{h}^{(l)})^T\right]^T \qquad l = 1, \ldots, m_f \tag{3.121}$$

and assume the cross-correlation coefficients to be given by (the index l is omitted for clarity)

$$b_{pq} = E\left\{h_p^R h_q^R\right\} = E\left\{h_p^I h_q^I\right\} = b_{qp} \tag{3.122}$$
$$\beta_{pq} = E\left\{h_p^R h_q^I\right\} = -E\left\{h_p^I h_q^R\right\} = -\beta_{qp} \tag{3.123}$$

where $h_p = h_p^R + jh_p^I$ is the p:th element in the vector $\mathbf{h}^{(l)}$. The vector $\mathcal{H}_l$ in (3.121) has a zero-mean multivariate Gaussian pdf with covariance matrix

$$\Sigma = E\left\{\mathcal{H}_l \mathcal{H}_l^T\right\} = \begin{bmatrix} \mathcal{B} & \mathcal{C} \\ -\mathcal{C}^T & \mathcal{B} \end{bmatrix} \tag{3.124}$$

which is assumed the same for each l, and where

$$\mathcal{B} \stackrel{\triangle}{=} \begin{bmatrix} 1 & b_{12} & \cdots & b_{1,n_r n_t} \\ b_{21} & 1 & \cdots & b_{2,n_r n_t} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n_r n_t,1} & b_{n_r n_t,2} & \cdots & 1 \end{bmatrix} \tag{3.125}$$

$$\mathcal{C} \stackrel{\triangle}{=} \begin{bmatrix} 0 & \beta_{12} & \cdots & \beta_{1,n_r n_t} \\ \beta_{21} & 0 & \cdots & \beta_{2,n_r n_t} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n_r n_t,1} & \beta_{n_r n_t,2} & \cdots & 0 \end{bmatrix} \tag{3.126}$$

which completely defines the spatial correlation due to multipath scattering. If the channels $\mathbf{H}_{pq}$ are independent, then Σ will be the unity matrix, and if channels emanating from the same transmit antenna are independent, then the matrices $\mathcal{B}$ and $\mathcal{C}$ are block-diagonal with blocks of size $n_r \times n_r$. This corresponds to the case when the transmitter antennas are placed much farther apart than the coherence length of the channel.

To find the pdf of Υ, recall that $\mathcal{H}_l$ is a real Gaussian distributed vector with zero mean and covariance matrix Σ. Next, construct the sum

$$\mathbf{S} = \sum_{l=1}^{m_f} \mathcal{H}_l \mathcal{H}_l^T \sim \mathcal{W}_{2n_r n_t}(m_f, \Sigma) \tag{3.127}$$

which has a central Wishart distribution with m_f degrees of freedom and with a covariance matrix Σ [43, p.86]. The characteristic function for $\mathbf{S}$ is [43, p.87]

$$\Phi \stackrel{\triangle}{=} E\left[\exp(jTr(\mathbf{ST}))\right] = \det\left(\mathbf{I}_{2n_r n_t} - 2j\mathbf{T}\Sigma\right)^{-m_f/2} \tag{3.128}$$

where $\mathbf{T}$ is an arbitrary symmetric matrix. By setting the non-diagonal elements of $\mathbf{T}$ to zero, the characteristic function of the multivariate Gamma distribution is obtained as a special case of the Wishart distribution and by setting all the diagonal $t_k = t, k = 1, \ldots, n_r n_t$ we obtain

$$\Phi = E\left[\exp\left(jt\sum_{k=1}^{n_r n_t} g_k\right)\right] = E\left[\exp\left(jt\Upsilon\right)\right] \tag{3.129}$$

where $g_k = |h_k|^2$. This is the characteristic function of the sum in (3.112) and it can be evaluated as [83]

$$\Phi(s) = \det\left(\mathbf{I}_{n_r n_t} - s\mathbf{R}_H\right)^{-m_f} \tag{3.130}$$

where $s = jt$. In obtaining (3.130), the property that Σ is a positive definite matrix was used, hence it holds that $\det\left(\mathbf{I}_{2n_rn_t} + z\Sigma\right)^{-1/2}$ can be simplified [120] to $\det\left(\mathbf{I}_{n_rn_t} + z\mathbf{R}_H\right)^{-1}$ where

$$\begin{aligned}\mathbf{R}_H &= E\left\{\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^*\right\} \\ &= \begin{bmatrix} 1 & B_{12}^* & B_{13}^* & \cdots & B_{1,n_rn_t}^* \\ B_{12} & 1 & B_{23}^* & \cdots & B_{2,n_rn_t}^* \\ B_{13} & B_{23} & 1 & \cdots & B_{3,n_rn_t}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{1,n_rn_t} & B_{2,n_rn_t} & B_{3,n_rn_t} & \cdots & 1 \end{bmatrix}_{n_rn_t \times n_rn_t}.\end{aligned} \tag{3.131}$$

The elements $B_{pq} = b_{pq} + j\beta_{pq} = E\left(\tilde{h}_p^{(l)}\tilde{h}_q^{(l)*}\right)$ are the complex component correlation coefficients between *two pairs of transmit/receive antennas* and depends on the scattering scenario, antenna element positions and separation distance and polarization. The elements B_{pq} have no index l, since Σ is the same for all l, see the discussion prior to Definition 3.1 on page 80.

Conclusively, the characteristic function of the pdf with arbitrary correlation is given by equation (3.130). The correlation is defined by the covariance matrix $\mathbf{R}_H$ in (3.131).

The pdf of Υ is now derived by writing the characteristic function of $\Phi(s)$ as

$$\Phi(s) = \det\left(\mathbf{I}_{n_rn_t} - s\mathbf{R}_H\right)^{-m_f} = \prod_{k=1}^{n_rn_t}(1 - s\nu_k)^{-m_f} \tag{3.132}$$

where ν_k are the eigenvalues of $\mathbf{R}_H$. Assume that the eigenvalues are distinct and expand (3.132) in a partial sum

$$\Phi(s) = \prod_{k=1}^{n_rn_t}(1 - s\nu_k)^{-m_f} = \sum_{l=1}^{n_rn_t}\sum_{r=1}^{m_f}\frac{c_{l,r}}{(1 - s\nu_l)^r} \tag{3.133}$$

where the partial fraction expansion coefficients $c_{l,r}$ are given by [107]

$$c_{l,r} = \left(\prod_{k=1,k\neq l}^{n_rn_t}(1 - s\nu_k)^{-r}\right)\Bigg|_{s=1/\nu_l} \tag{3.134}$$

for $r = m_f$ and for $0 < r < m_f$ the coefficients are

$$c_{l,r} = \frac{1}{(m_f - r)!(-\nu_l)^{m_f - r}}\frac{d^{(m_f-r)}}{ds^{(m_f-r)}}\left(\prod_{k=1,k\neq l}^{n_rn_t}(1 - s\nu_k)^{-m_f}\right)\Bigg|_{s=1/\nu_l}. \tag{3.135}$$

The terms in the summation (3.133) is of the form $(1 - s\nu_l)^{-r}$ which inverse Laplace transform is the Gamma distributed pdf [107]

$$p(x; l, r) = \frac{1}{\nu_l \Gamma(r)} \left(\frac{x}{\nu_l}\right)^{r-1} e^{-x/\nu_l} \quad . \tag{3.136}$$

So the inverse Laplace transform of the sum in (3.133) is equal to the sum over the inverse Laplace transform of each term. This yields finally the pdf of Υ as

$$p_\Upsilon(x) = \sum_{l=1}^{n_r n_t} \sum_{r=1}^{m_f} c_{l,r} p(x; l, r) \tag{3.137}$$

and the result follows.

Appendix 3.C Proof of Corollary 3.2

First, define $\mathcal{H}_l$ as in the proof of Lemma 3.1 in Appendix 3.B. The matrix $\mathbf{R}_H$ is defined for the complex Gaussian channel case, that is, when $m_f = 1$, according to Definition 3.1 on page 80. Hence, we proceed by assuming that $m_f = 1$. The vector $\mathcal{H}_l$ has the covariance matrix Σ. Rewrite the Frobenius norm of the channel as

$$\begin{aligned} \left\|\mathbf{H}'\right\|_F^2 =& |\text{vec}(\mathbf{H}')|^2 \\ =& |\text{vec}(\mathbf{C}_R \mathbf{H} \mathbf{C}_T)|^2 \\ =& |(\mathbf{C}_T^T \otimes \mathbf{C}_R)\text{vec}(\mathbf{H})|^2 \\ =& |(\mathbf{C}_T^T \otimes \mathbf{C}_R)\mathbf{h}^{(l)}|^2 \\ \triangleq& |\mathcal{K}\mathcal{H}_l|^2 \\ \triangleq& |\mathcal{H}'|^2 \end{aligned} \tag{3.138}$$

where the identity [43, p.74]

$$vec(\mathbf{BXC}) = (\mathbf{C}^T \otimes \mathbf{B}) vec(\mathbf{X}) \tag{3.139}$$

and the equations (3.120),(3.121) have been used and where

$$\mathcal{K} = \begin{bmatrix} \text{Re}\left(\mathbf{C}_T^T \otimes \mathbf{C}_R\right) & -\text{Im}\left(\mathbf{C}_T^T \otimes \mathbf{C}_R\right) \\ \text{Im}\left(\mathbf{C}_T^T \otimes \mathbf{C}_R\right) & \text{Re}\left(\mathbf{C}_T^T \otimes \mathbf{C}_R\right) \end{bmatrix}_{2n_r n_t \times 2n_r n_t} \quad . \tag{3.140}$$

The vector $\mathcal{H}'$ is a linear combination of the elements in $\mathcal{H}$. Results from multivariate probability theory, see e.g. [43, p.6], then gives that $\mathcal{H}'$ will also be a multivariate, zero mean, Gaussian distributed vector with covariance matrix $\Sigma' = \mathcal{K}\Sigma\mathcal{K}^T$, and the result follows.

Appendix 3.D Proof of Theorem 3.2

Start with the mutual information of the equivalent SISO system

$$I = \frac{n_s}{N} \log_2 \left(1 + \frac{P_T}{\sigma_n^2 n_t} \|\mathbf{H}\|_F^2 \right) \tag{3.141}$$

symbols per seconds per Hz of bandwidth for complex valued modulation symbols. The outage probability for a given rate C_{P_o} is the probability P_o that the mutual information falls below the given rate C_{P_o}, and the outage capacity is the minimum rate C_{P_o} that can be supported by the channel with probability $(1 - P_o)$. So, recall that the random variable is $\Upsilon = \|\mathbf{H}\|_F^2$ and write

$$\begin{aligned} P_o =& Pr(I < R_o) \\ =& Pr\left(\frac{n_s}{N} \log_2 \left(1 + \frac{P_T}{\sigma_n^2 n_t} \Upsilon \right) < C_{P_o} \right) \\ =& \int_0^{\left(\frac{P_T}{\sigma_n^2 n_t}\right)^{-1} (2^{C_{P_o} N/n_s} - 1)} p_\Upsilon(\Upsilon) d\Upsilon \ . \end{aligned} \tag{3.142}$$

The pdf $p_\Upsilon(\Upsilon)$ is given by Lemma 3.1 on page 107 as a sum of Gamma distributions. The integral

$$\int_0^t \frac{1}{\nu_l \Gamma(r)} \left(\frac{x}{\nu_l} \right)^{r-1} e^{-x/\nu_l} dx \tag{3.143}$$

can, by doing the substitution $y = x/\nu_l$ be written as

$$\frac{1}{\Gamma(r)} \int_0^{t/\nu_l} y^{r-1} e^{-y} dy \tag{3.144}$$

which is identified as proportional (with proportionality constant $1/\Gamma(r)$) to the incomplete Gamma function [80, Def. 8.350] $\Gamma(r, t/\nu_l)$.

To find the outage probability, (3.54) is used in (3.142) and the integrations are carried out using (3.143),(3.144), and this results in

$$P_o = Pr(I < R_o) = \sum_{l=1}^{n_r n_t} \sum_{r=1}^{m_f} \frac{c_{l,r}}{\Gamma(r)} \Gamma\left(r, \frac{\sigma_n^2 n_t}{P_T \nu_l} (2^{C_{P_o} N/n_s} - 1) \right) \tag{3.145}$$

which is the result.

Appendix 3.E Proof of Theorem 3.3

For coherently detected BPSK and Gray coded QPSK modulation alphabet, the BER is given by [15]

$$P_b(x) = Q\left(\sqrt{2x}\right) \tag{3.146}$$

where x is the SNR. The BER is then obtained by averaging (3.146) over the SNR distribution for the equivalent SISO system,

$$P_{STBC} = \int_0^\infty P_b\left(\frac{P_T\Upsilon}{n_t\sigma_n^2}\right) p_\Upsilon(\Upsilon)d\Upsilon \tag{3.147}$$

where the equivalent SNR for a STBC system is $P_T\Upsilon/(n_t\sigma_n^2)$, see Section 3.5.1. If the integrations are carried out, using [80], then the following result follows

$$P_{STBC} = \sum_{l=1}^{n_r n_t}\sum_{r=1}^{m_f} c_{l,r}\left[\frac{1}{2} - \sqrt{\frac{P_T\nu_l}{n_t\sigma_n^2\pi}}\frac{\Gamma(r+1/2)}{\Gamma(r)}\mathrm{F}\left(\frac{1}{2},\frac{1}{2}+r;\frac{3}{2};-\frac{P_T}{n_t\sigma_n^2}\nu_l\right)\right] \tag{3.148}$$

where $F(\cdot)$ is the hypergeometric function.

Appendix 3.F Proof of Corollary 3.3

If no mutual coupling is present, and $\mathbf{R}_H = \mathbf{I}$, then this corresponds to a covariance matrix $\Sigma = \mathbf{I}_{2n_rn_t}$ (refer to the proof of Lemma 3.1 in Appendix 3.B) and the characteristic function (3.130) can thus be written as

$$\Phi(s) = (1-s)^{-m_f n_r n_t} \tag{3.149}$$

and hence, the pdf of $\Upsilon = \|\mathbf{H}\|_F^2$ becomes, by calculating the inverse Laplace transform of (3.149)

$$p_\Upsilon(x) = \frac{x^{m_f n_r n_t - 1}}{\Gamma(m_f n_r n_t)}e^{-x} \tag{3.150}$$

which is the gamma distribution with $2m_f n_r n_t$ degrees of freedom.

The expressions for the outage capacity and BER follows by applying the same calculations as in equation (3.142) and (3.147) but using the distribution (3.150) instead.

Appendix 3.G Proof of Theorem 3.4.

The SNR for beamforming is from Section 2.5 given by $P_T\lambda_{max}/\sigma_n^2$ and the BER for coherently detected BPSK or coherently detected Gray coded QPSK in an additive white Gaussian channel is given by [15]

$$P_b(x) = Q\left(\sqrt{2x}\right) \tag{3.151}$$

where x is the SNR. The average probability of bit error, P_{BF}, is then given by averaging (3.151) over the pdf of the principal eigenvalue, which is given in Corollary 3.1 on page 87:

$$P_{BF} = \int_0^\infty Q\left(\sqrt{\frac{2P_T}{\sigma_n^2}\lambda_{max}}\right) p_\lambda(\lambda_{max})d\lambda_{max} \tag{3.152}$$

Since the pdf $p_\lambda(\lambda_{max})$ consists of terms $x^p e^{-kx}$, according to (3.37), we obtain by using the symbolic integration software package MAPLE 7.0

$$\begin{aligned}\int_0^\infty & x^p Q(\sqrt{\alpha x})e^{-kx}dx = \\ &= \sqrt{\frac{\alpha}{2\pi}}\frac{1}{k^{p+3/2}}F\left(\frac{1}{2},\frac{3}{2}+p;\frac{3}{2},-\frac{\alpha}{2k}\right)\Gamma\left(\frac{3}{2}+p\right)+\frac{\Gamma\left(1+p\right)}{k^{1+p}} \quad .\end{aligned} \tag{3.153}$$

Hence, the integral in (3.152) is solved, term by term, by using (3.153). The simplifying formula for this particular hypergeometric function, expressed in (3.64) is the used, and after some lengthy calculations, the result (3.92) follows.

Appendix 3.H Proof of Corollary 3.4.

The pdf of the trace of the matrix $\mathbf{W}$ is given by (3.54), since $\|\mathbf{H}\|_F^2 = \text{Tr}(\mathbf{H}\mathbf{H}^*) = \text{Tr}(\mathbf{W})$. So, the following holds:

$$\frac{1}{m}\text{Tr}(\mathbf{W}) \leq \lambda_{max} \leq \text{Tr}(\mathbf{W}) \tag{3.154}$$

where $m = \min(n_r, n_t)$. From the correlated Nakagami-m fading channel with the parameter m_f set to $m_f = 1$, the upper bounded pdf in the Rayleigh fading case is equal to the pdf for $\|\mathbf{H}\|_F^2$, obtained in the analysis of STBC. Hence, from (3.54) and (3.55) we get the pdf as

$$p_{\lambda_{max}}^{Upper}(\lambda_{max}) = \sum_{l=1}^{n_r n_t}\frac{c_l}{\nu_l}e^{-\lambda_{max}/\nu_l} \quad . \tag{3.155}$$

To find the pdf for $\|\mathbf{H}\|_F^2/m$, we make the variable substitution $\lambda = \lambda_{max}/m$ in (3.155). A result in [121] states that if $p_x(x)$ is a pdf of x, then $y \stackrel{\triangle}{=} ax$ has the pdf $p_y(y) = p_x(y/a)/|a|$. Using this and (3.154) we get

$$p_{\lambda_{max}}^{Lower}(\lambda_{max}) = \sum_{l=1}^{n_r n_t} \frac{m c_l}{\nu_l} e^{-m\lambda_{max}/\nu_l} \quad . \tag{3.156}$$

The eigenvalue λ_{max} attains the lower bound in an unitary channel where all eigenvalues ν_l to the channel covariance matrix $\mathbf{R}_H$ are equal. The upper bound is attained in a rank-1 channel where only one eigenvalue is non-zero. If the BER is calculated, using (3.152), then the following expression is obtained using the lower bound for the largest eigenvalue pdf,

$$P_{BF} = \sum_{l=1}^{n_r n_t} \frac{c_l}{2} \left[1 - \sqrt{\frac{\nu_l}{m\sigma_n^2/P_T + \nu_l}} \right] \tag{3.157}$$

which should be compared with (3.66), the BER of STBC in a correlated Rayleigh fading channel. The only difference is in the scaling factor m and n_t respectively, of the SNR (P_T/σ_n^2). If $m = \min(n_r, n_t) \equiv n_t$ then the upper bound of beamforming BER is identical to the BER of STBC in a correlated Rayleigh fading channel. That is, the number of receive antennas is equal to or larger than the number of transmit antennas. If $n_t > n_r$, then $m = n_r < n_t$ and the STBC BER expression (3.66) has a larger SNR penalty then the upper bound of the beamforming BER in expression (3.157). Hence, for $n_r < n_t$, beamforming BER is strictly lower than orthogonal STBC BER.

Appendix 3.I Table 3.2 in Theorem 3.4.

The table is given on the next page.

	m=1	m=2	m=3
n=1	$\varphi_1^{(1,1)}(x) = 1$	-	-
n=2	$\varphi_1^{(1,2)}(x) = \frac{x+3/2}{x+1}$	$\varphi_1^{(2,2)}(x) = \frac{2x^2+4x+11/4}{(x+1)^2}$ $\varphi_2^{(2,2)}(x) = -1$	-
n=3	$\varphi_1^{(1,3)}(x) = \frac{x^2+5x/2+15/8}{(x+1)^2}$	$\varphi_1^{(2,3)}(x) = \frac{2x^3+7x^2+61x/8+57/16}{(x+1)^3}$ $\varphi_2^{(2,3)}(x) = -\frac{x^2+5x+51/8}{(x+2)^2}$	$\varphi_1^{(3,3)}(x) = \frac{3x^4+12x^3+81x^2/4+117x/8+321/64}{(x+1)^4}$ $\varphi_2^{(3,3)}(x) = -\frac{3x^4+24x^3+297x^2/4+855x/8+3993/64}{(x+2)^4}$ $\varphi_3^{(3,3)}(x) = 1$

Table 3.2: Table of the polynomials $\varphi_k^{(m,n)}(x)$ in equation (3.92) for the BER of STBC in Rayleigh fading channels. The quantities m and n are defined as $\min(n_r, n_t)$ and $\max(n_r, n_t)$ respectively.

Chapter 4

Measurements and analysis of a SIMO system

THIS chapter presents measurement results from an adaptive array antenna testbed for the GSM-1800 system developed at Uppsala University. The testbed is used in laboratory measurements and in outdoor field trials in both LOS and NLOS environments. The testbed is able to suppress an interfering signal by 31 dB using null steering and simultaneously have a 2 dB array gain towards a desired user. This give a sufficient improvement to allow two users to share a time-frequency channel served by a basestation and still maintain a low BER for both users. A qualitative demonstration where two simultaneous and channel sharing phone calls were handled by the testbed is presented. Furthermore, the issue of regularization of the SMI algorithm is discussed and demonstrated in this chapter.

4.1 Introduction

Using a multiple antenna array at the basestation in a cellular system gives a number of benefits, see Chapter 1. This chapter deals with the uplink of a cellular system, hence a SIMO configuration in an interference limitied environment. Contrary to the previous chapter, the multiple antennas are primarily not used to provide diversity to combat fading, but to provide additional degrees of freedom to supress interferers. Note also that a narrower beamwidth will suppress delayed signals and a reduction in the delay spread is expected. The interferer suppression is possible since the array gives the receiver an ability to discriminate in space or angle. The ability of the receiver to suppress interferers can directly be related to an increase in the spectral efficiency of a cellular system, measured in the number of users per

unit area and per bandwidth [2]. Hence, it is important to quantify the amount of interference suppression that is possible and how it is related to the physical design of the basestation.

To investigate these capabilities of a SIMO system in the uplink of an cellular system, an adaptive antenna testbed was designed and realized at Signals and Systems Group, Uppsala University. The adaptive antenna testbed project commenced 1994 at Uppsala University in a cooperation with Ericsson Radio Access AB. The aim was to develop and evaluate adaptive antenna technologies for mobile communication systems and to obtain experience of using this technology in base stations. Due to limited resources, only the uplink was implemented. Furthermore, receive beamforming was used, see Section 2.5.1 and the weighting was implemented in hardware using digitally controlled phase shifters and attenuators. This is sometimes called an *analog beamformer* (ABF) [69]. The benefits with the ABF is that it can be used as an add-on system on existing base station receivers, where the output from the ABF is connected to the ordinary base station receiver. In [3], field trial results showed that by upgrading 5% of the basestations to adaptive antennas yielded a 25% network capacity increase. Hence, basestations in traffic "hot-spots" can be upgraded with array antennas and thereby boost the capacity significantly.

Another implementation option is the fully digital beamformer (DBF) [71], where both the adaptive algorithm and the weighting and summation of the signals is performed in DSP software. The DBF is flexible and allows for system upgrades by changing the DSP software, but requires completly new BS equipment. However, for broadband systems, the limitations on the maximum speed of the I/O bus in the DSP reduces the ability to sample and process data from many antennas, as is required in the DBF configuration. In [122], it was reported that the maximum data throughput to the DSP is currently around 10 Mbit/s. Hence, congestion in the I/O bus occurs for an $n_r = 4$ element antenna system at about 2.5 Mbit/s throughput. If higher data speeds are desired, ABF becomes an interesting choice, since then, the beamforming takes place in hardware which is throughput insensitive. To calculate the beamformer weights, the received signals are sampled at a speed that is chosen based on the coherence time for the channel (as compared to the symbol time in DBF applications). Hence, the system is separated into a throughput extensive part and a computation intensive part [122]. This further motivates the study of ABF systems.

This chapter contains only a brief description of the hardware of the testbed since a more detailed description can be found in the Licentiate thesis of Andersson and Landing [123]. Some other testbed projects for GSM/DCS 1800 have also been reported by Ericsson [7], the TSUNAMI project [11] and others, see [124] for an overview of adaptive antenna testbeds. This chapter also contains measurement results from the conducted field trials. In Chapter 5, these measurements are

analyzed in more detail and the performance bottlenecks are identified.

This chapter is organized as follows. It starts with a description of the adaptive antenna architecture in Section 4.2 followed by an analysis of the performance in the ideal case in Section 4.3. Section 4.4 presents results from measurements performed with the testbed in a laboratory, where a Butler matrix was used to emulate the front end. Section 4.5 describes measurements performed at an outdoor antenna measurement range and Section 4.6 describes a demonstration with voice transmission using the testbed. Finally, this chapter ends with some conclusions in Section 4.7.

4.2 Adaptive antenna architecture

The receiver antenna array is designed for integration with an existing base station, using the GSM-1800 standards, thus the radio interface frequency is 1.7-1.8 GHz. A photo of the antenna is shown in Figure 4.1 and in Figure 4.2 a schematic outline of the system is shown. The front-end consists of ten antenna elements, mounted in a circular array configuration. The antennas are connected to directional couplers to allow injection of a calibration signal. The front end also contains, for each antenna element, cavity filters to remove out of operating band interference and a low noise amplifier. In the receiver rack, each of the ten signals from the front-end are split in three replicas where one is connected to a sampling receiver and the two other to two independent beamformers. Hence, it is possible to simultaneously weight and sum the ten antenna signals in two independent ways. The two beamformers enable trials with the SDMA (Spatial Division Multiple Access) method. The two signal sources (e.g. mobiles) are separated by the receiver algorithm by assigning them different training sequences (pilots).

A summary of some characteristics of the adaptive array system is shown in Table 4.1. The beam-forming is performed on the received RF signals at 1.721 GHz and the received vector signal is reduced to a scalar signal, on the RF frequency. This enable the use of an ordinary base station as a receiver. The weights are calculated using the algorithm in the digital signal processor (DSP) and this type of weighting is denoted a hybrid-analog beamformer in [69] due to the use of both digital and analog signal processing.

The basestation is a support system used in the measurements to demodulate the signals and provide the adaptive antenna with synchronization signals for correct timing of the sampling. Furthermore, the transmitted data is generated by the basestation which contains software to compare the transmitted data bits with the received data, for bit error rate (BER) calculations and for estimation of other parameters that characterize the transmission. The basestation has many features but

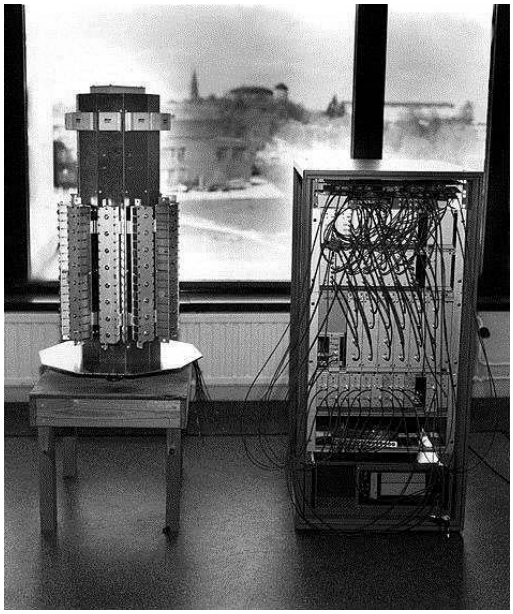


Figure 4.1: The adaptive antenna, to the left the front end and to the right the rack with receivers, DSP-system, and weighting units.

in these measurements it runs in a limited mode and only traffic channel frames are transmitted.

The sampling receivers (SRX) down-convert the RF signal to the baseband and separate it into I and Q channels using double-down-conversion receivers. The sampling in the analog to digital converters (ADC:s) is performed at 270 kbit/s which is equal to the symbol rate in the GSM standard [125]. The receiver gain is set so the quantization noise is equal to the thermal noise for maximal use of the receiver dynamic range. The ADC:s use 8 bits and the DSPLINK2 bus that connects to the DSP is a 32 bit bus. Thus four channels can be read from the ADC:s to the DSP simultaneously. The possible increase in performance by using ADC:s with more bits is investigated in Section 5.4.3. The DSP consists of seven TMS320C40 signal processors which can be programmed to run a number of different weight calculation algorithms. In the measurements presented here, the sample matrix inversion (SMI) algorithm [126] is used, see further Section 5.3.2. SMI was chosen for its rapid convergence compared to other algorithms [127], and to avoid the need for calibration of the hardware channel between the antennas and the SRX. Many existing algorithms rely on direction of arrival (DOA) estimates of the signals and thus require a calibrated front end. This is avoided using SMI, but calibration of the weights is still needed, since hardware weights are used. The SMI algorithm,

Item	Specification
Front end and radio parameters	
Center Frequency	1721 MHz
Modulation	GMSK (BT_s=0.3)
Number of antenna elements	10
Antenna configuration	Circular
Antenna element spacing	$0.56\lambda_c$ @ 1721 MHz
Antenna polarization	Vertical
Antenna element beam-width	80°
Receiver and DSP system	
DSP Processors	Seven TMS320C40
Sampling frequency per I- and Q-channel	270 kHz
ADC resolution	8 bit
ADC dynamic range	-32 dBm to -80 dBm
Algorithm for weight calculation	SMI
Measured noise figure for receiver	11 dB
Receiver maximum voltage gain	75 dB
Input IP3	+2 dBm
IQ isolation	75 dB
Weighting units	
Class	Analog
Phase shift resolution	1°
Amplitude attenuation resolution	1 dB
Dynamic range for weight setting	50 dB
Noise figure	6 dB
IP3	+6dBm
Phase stability	1°/hour
Amplitude stability	0.1 dB/hour
Calibration	
Phase accuracy	1°
Amplitude accuracy	0.75 dB

Table 4.1: Characteristic parameters of the adaptive antenna testbed

which is a reference signal type of algorithm, see Section 2.5.1, use the training sequence of 26 bits in the mid-amble of each GSM-1800 traffic channel burst as a reference signal to separate users. For a survey of adaptive antenna algorithms, see [128, 129].

Although SMI does not require a calibration of the front end, the signals that are used in the DSP to calculate the weights are not exactly the same as the signals

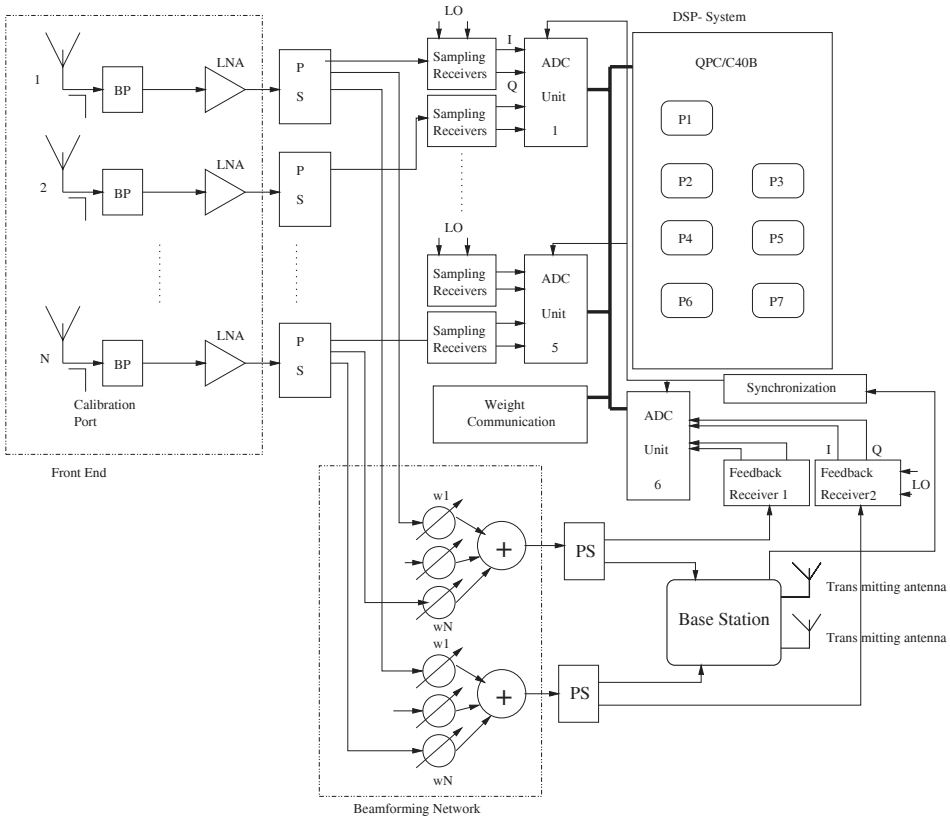


Figure 4.2: Adaptive antenna architecture. To the left the front end, to the right the DSP system consisting of seven digital signal processors. The basestation (bottom right) provides the system with synchronization signals and transmits and receives data bits for BER calculations. BP= Bandpass Filter, LNA= Low Noise Amplifier, LO= Local Oscillator, PS= Power Splitter.

used in the beam-formers (the hardware weighting units). Rather they are a phase shifted and attenuated replica of them, so there is a need for calibration of these weighting units. The phase shift and attenuation is dependent on the length of the cables, their losses, and also the characteristics of the weighting units. They are built from active components, as phase shifters, attenuators and amplifiers, which suffer from temperature drift. This will soon make a calibration invalid and it is therefore necessary to recalibrate often. Calibration algorithms are presented in Sections 5.2.1 and 5.5.

A feedback receiver is used to down-convert and sample the summed beam-former output signal which is then utilized by the calibration algorithms. The calibration takes place off-line, when the adaptive antenna is not adapting to any

external signal sources, since the calibration algorithm needs full access to the weight controller. The calibration requires a CW (continuous wave) signal injected through directional couplers as close to the antenna elements as possible. The data from the calibration is then stored as a look-up table in the DSP:s memory. Off-line calibration implies that the base-station must be taken out of service but can be avoided by more advanced calibration methods. Two proposals for on-line calibration methods that are transparent to normal operation are presented in Chapter 5.5.

4.3 Performance in a LOS channel

In this section, the expression for the carrier to interference plus noise ratio (CINR)[1] in the output signal of the beamformer is derived. The channel model is a LOS channel with no angular spread, to make the derivation simple and yield an analytical result that lends itself to interpretation. Furhtermore, the measurements are made in a LOS channel and the derived CINR can thus be compared with the measured results.

In Chapter 2, the flat fading SIMO channel was described by the column vector $\mathbf{h}$. If the channel is LOS and with negligible angular spread, then the channel vector $\mathbf{h}$ can be parameterized using the direction of arrival (DOA) θ of the impinging signal. This is denoted $\mathbf{a}(\theta)$ and is commonly called the *array response vector*. Define the vectors $\mathbf{a}_i(\theta_i), \mathbf{a}_d(\theta_d)$ as the array response vectors towards the interfering and the desired[2] mobile respectively. Furthermore, assume that the array response vectors are normalized, i.e. $\mathbf{a}_d^*(\theta_d)\mathbf{a}_d(\theta_d) = 1$. If the training sequence is assumed to have infinite length, then effects from finite sample estimation such as unconverged covariance matrices and cross-correlation estimates can be neglected, and estimated covariances can be replaced by their true values. The Wiener-Hopf solution for the receive weight vector $\mathbf{w}_R$ can then be expressed as, from (2.117)

$$\mathbf{w}_R = \mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{r}_{\mathbf{y}d} \quad . \tag{4.1}$$

Here $\mathbf{R}_{\mathbf{yy}}$ is the covariance matrix of the received signal $\mathbf{y}(n)$ and $\mathbf{r}_{\mathbf{y}d}$ is the cross-correlation between the received signal and the training sequence $d(n)$. The covariance matrix of the received signal vector is, in the LOS with no angular spread case given by

$$\mathbf{R}_{\mathbf{yy}} = \sigma_d^2\mathbf{a}_d(\theta_d)\mathbf{a}_d^*(\theta_d) + \sigma_i^2\mathbf{a}_i(\theta_i)\mathbf{a}_i^*(\theta_i) + \sigma_n^2\mathbf{I} \tag{4.2}$$

[1] Also known as the signal to intereference plus noise ratio (SINR). Carrier and signal will be used interchangeably in this thesis

[2] The MS which data we want to receive.

where σ_i^2, σ_d^2 and σ_n^2 is the interferer, the desired signal and the noise power respectively, as measured by a virtual unit gain omnidirectional antenna at the base-station[3].

By applying the matrix inversion lemma [130] twice in equation (4.2) and separating the interferer, the desired signal and the noise parts of the output signal, it is possible to show that the beamformer output carrier to interferer plus noise ratio can be written [130, 131] as

$$\text{CINR} = \frac{n_r \sigma_d^2}{\sigma_n^2} \left(1 - \frac{\sigma_i^2 |\rho|^2}{(\sigma_n^2 + n_r \sigma_i^2) n_r} \right) \tag{4.3}$$

where

$$\rho = \mathbf{a}_i^*(\theta_i) \mathbf{a}_d(\theta_d) \tag{4.4}$$

is the spatial correlation coefficient between the interfering and desired signal's array response vectors. When the interferer to noise power (INR) is high and the spatial correlation is low, $\sigma_n^2/\sigma_i^2 \ll n_r(1 - |\rho|^2)$, we have

$$\text{CINR} \simeq \frac{n_r \sigma_d^2}{\sigma_n^2} \left(1 - |\rho|^2 \right) \quad . \tag{4.5}$$

Equation (4.5) shows how the CINR at the output of the adaptive array with Wiener-Hopf weights is the input signal to noise ratio (σ_d^2/σ_n^2), amplified n_r times by the array, as in the absence of an interferer, and then reduced by the factor $(1 - |\rho|^2)$. The loss in CINR, $1 - |\rho|^2$, is due to the pattern distortion when forming a null in the interference direction. Note that the interferer power is not present in the expression (4.5). The interfering signal is successfully suppressed to the noise floor, independent of its power in this idealized model.

If the desired signal and interferer are very close in azimuth angle θ, then the spatial correlation is $|\rho| \simeq 1$. In this case the noise appears to be "amplified" by the spatial correlation and $\sigma_n^2/\sigma_i^2 \gg n_r(1 - |\rho|^2)$ holds. This will lead to the asymptotic CINR

$$\text{CINR} \simeq \frac{n_r \sigma_d^2}{\sigma_n^2 + n_r \sigma_i^2} \tag{4.6}$$

which is the signal to interferer plus noise ratio of an n_r element omnidirectional receiver with maximum ratio combining of the desired signal. Thus, the improvement by using the array antenna is in this case only due to a reduction of $1/n_r$ of the noise power. A more thorough derivation of the output CINR for the minimum variance beamformer is presented by Wax and Anu [132] which takes the finite sample size effect when estimating the covariances and also the correlation between the desired and the interfering signal into account.

[3]The desired signal power σ_d^2 corresponds to the transmitted power P_T times the path loss through the channel.

4.4 Laboratory measurements

To verify the adaptive antenna performance, laboratory measurements were perfprmed, using a flat fading LOS channel with no angular spread. This was accomplished by replacing the front end by a 8×8 Butler matrix. See Appendix A for a description of the Butler matrix. Two signal generators were connected to two different input ports of the Butler matrix to emulate a scenario corresponding to a 8-element linear array of isotropic antenna elements spaced $\lambda_c/2$ with two impinging signals with DOA:s -14.5° and 14.5° respectively, relative to broadside. The 8 outputs of the Butler matrix are then connected to the adaptive antenna. Ideally, the array response vectors from the signals generated by a Butler matrix are orthogonal, i.e. $\mathbf{a}(\theta_k)^*\mathbf{a}(\theta_l) = \delta_{kl}$, where k and l represents ports of the Butler matrix, generating signals from direction θ_k and θ_l. If $k \neq l$, then the spatial correlation coefficient will be $|\rho| = 0$, due to the structure of a Butler matrix, and the performance of the interferer suppression will be excellent, see (4.5). However, imperfections in the manufacturing of the Butler matrix makes the array response vectors non-orthogonal, which corresponds to a more realistic signal environment. The manufacturer gives the maximum errors as 8.5° in phase and 0.8 dB in magnitude.

The performance of the adaptive antenna was examined by measuring the output BER and output power (after the beamformer) of the desired signal, called the carrier (C_{out}) and output interferer signal (I_{out}) for different settings of input power (C_{in}, I_{in}). The input power of the two signals were measured at the input of the Butler matrix. With this setup, it is possible to measure and distinguish the suppression of the interferer and the gain of the desired signal.

The desired signal is GMSK-modulated with GSM-1800 traffic channel data containing a 26 bit training sequence in the middle of each time slot. The interferer is also a GMSK-modulated signal transmitting pseudo random (PRBS) data or traffic data with another training sequence, depending on if both beamformers are used simultaneously. Ten measurements series were performed and the diagrams presented show the mean of these series, see Figure 4.3, 4.4 and Figure 4.6.

Figure 4.3 shows the average suppression of the interferer by the adaptive antenna system. The standard deviation in the measurements is 7 dB. The large standard deviation is due to the randomness in the estimate of the steering vector $\mathbf{r}_{\mathbf{y}d}$ which is calculated as the correlation between the training sequence and the received data. Due to the short training sequence, the correlation between the training sequence and the PRBS data transmitted by the interferer will be nonzero and vary randomly. If the correlation is small, then the interferer suppression will be excellent but a large correlation will result in a poor suppression. This variation is the reason of the large standard deviation in the measurements. This effect was

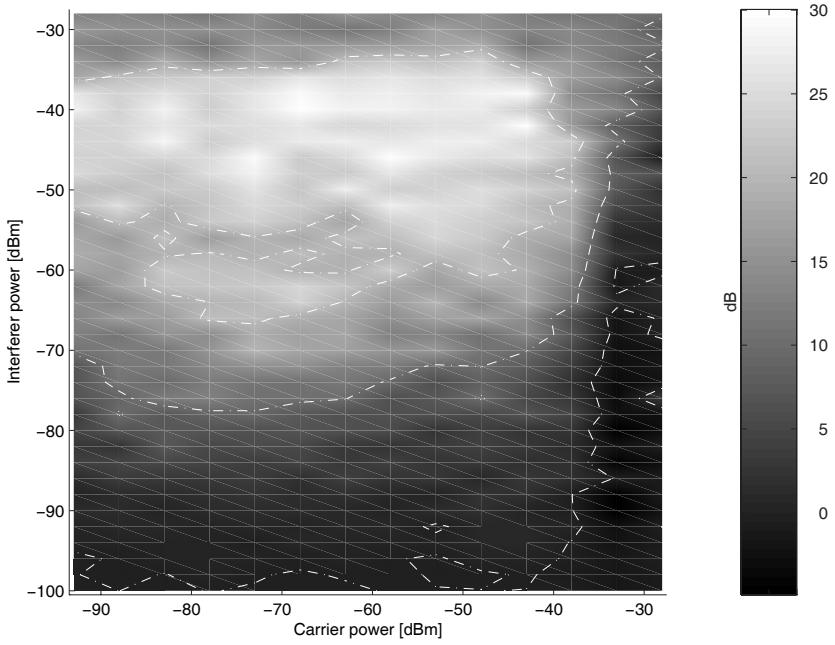


Figure 4.3: Suppression of interfering signal, in dB. Measurements performed in lab, on an 8 element array antenna, mean values over 10 measurements.

also observed in the adaptive antenna testbed in the TSUNAMI project [133]. The adaptive antenna is able to suppress the interfering signal when the power of the interferer and the desired signal is in the dynamic range of the ADC:s. The ADC use 8 bits for conversion and this gives a dynamic range of approximately 48 dB. The receiver gain was adjusted so the quantization noise power and the thermal noise power was approximately equal. The noise floor in the receiver was -80 dBm. An interferer is suppressed to the quantization noise floor if the signal sources are sufficiently separated in angle, or if the spatial correlation is sufficiently low, which can be assumed in the Butler matrix measurements. Compare with equation (4.5) which expresses the output CINR for the optimal antenna which is independent of the interferer power. Thus, the interferer suppression in Figure 4.3 and 4.4 should be proportional to I_{in}. The dashed line surrounding the brightest area in Figure 4.3 is the 20 dB iso-suppression line, and the maximum suppression is about 30 dB.

If the ratio of the output carrier to interference ratio (CIR_{out}) to the input CIR (CIR_{in}) is plotted, then we obtain the CIR improvement, presented in Figure 4.4 and 4.5. The CIR_{in} was measured using one of the antennas in the array so the improvement will be referred to a single antenna. The main contribution to the CIR improvement is from the interferer suppression. The gain of the desired signal

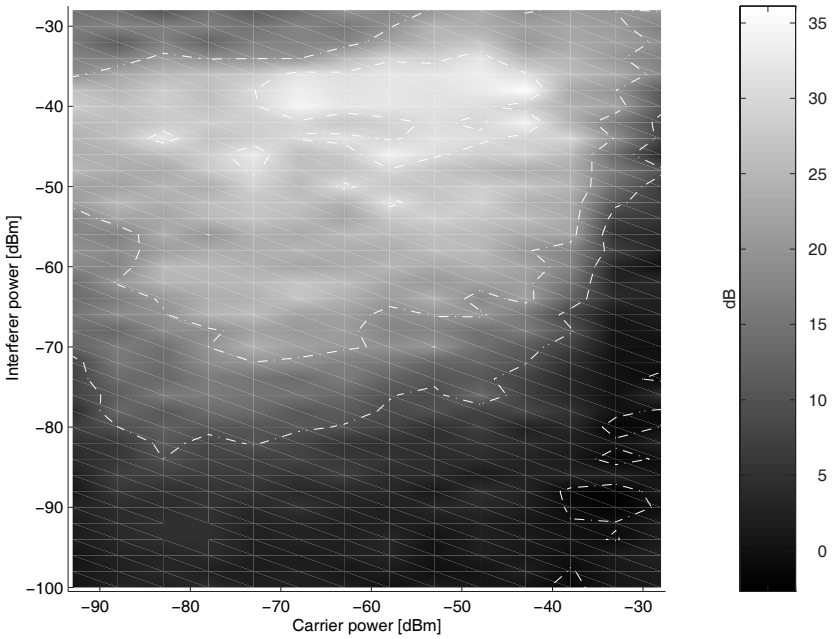


Figure 4.4: Improvement in CIR, in dB relative to one element antenna. Measurements performed in lab, on an 8 element array antenna, mean values over 10 measurements.

was approximately constant at about 2-6 dB for all carrier and interferer powers in the ADC dynamic range. With an interferer power at -40 dBm and the desired input signal power between -70 dBm and -40 dBm (CIR_{in} $\leq$0 dB), the adaptive antenna is improving the CIR more than 30 dB. When either of the signal levels (I_{in},C_{in}) exceeds the ADC dynamic range the CIR improvement drops abruptly to 0 dB or less. The reason is that with saturated receivers, the calculation of correct weights is not successful, and the beamforming is not performed correctly.

Figure 4.5 shows a vertical cut of the plot in Figure 4.4 at C_{in}=-73 dBm. The carrier to interference improvement increases linearly (on dB scale) between interferer power between -90 and -70 dBm in accordance with equation (4.5). At an interference level above -70 dBm the improvement saturates and levels out between 25 and 30 dB. The reason for this is the hardware limitations, which will be furhter investigated in Chapter 5.

BER improvement

Another illustrative way of presenting the measurements is to plot the improvement of the BER for different settings of C_{in} and I_{in}, see Figure 4.6. The presented data

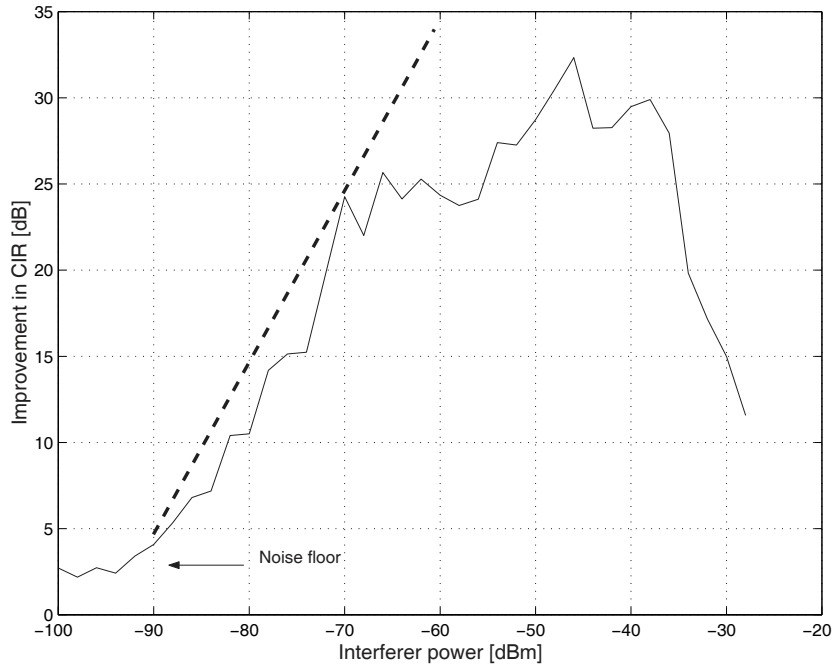


Figure 4.5: Improvement in CIR at carrier power -73 dBm. Note how the improvement levels out for high interferer powers and when the power is below the noise floor

is post-FEC (forward error correction) decoding of class Ib data bits in the base station. The improvement in BER is measured relative to one of the Butler matrix outputs, i.e. using a single antenna element compared to the 8 element adaptive antenna array. Hence, if BER is measured in $\%$, the BER improvement is defined as

$$\text{BER}_{\text{improvement}} = \text{BER}_{\text{one antenna}} - \text{BER}_{\text{adaptive antenna}} \qquad [\% \text{ units}] \ . \tag{4.7}$$

The BER improvement is zero for CIR_{in} $\geq$9 dB because the BER is 0% for the single antenna base station receiver in this region. This is the threshold level in GSM-1800 for 0 % BER. Above the lower right triangular area in Figure 4.6, is the region where there are benefits from using the adaptive antenna compared to using an ordinary single element base station antenna. Here the BER improvement is 50 percentage units, giving a BER out from the adaptive antenna of 0%. A CIR_{in} above -20 dB is thus necessary to give a BER of 0% at this setup, using the Butler matrix. Clearly, the separation of the signal sources affects the minimum CIR_{in} to still maintain 0% BER at the beamformer output. This is further investigated in

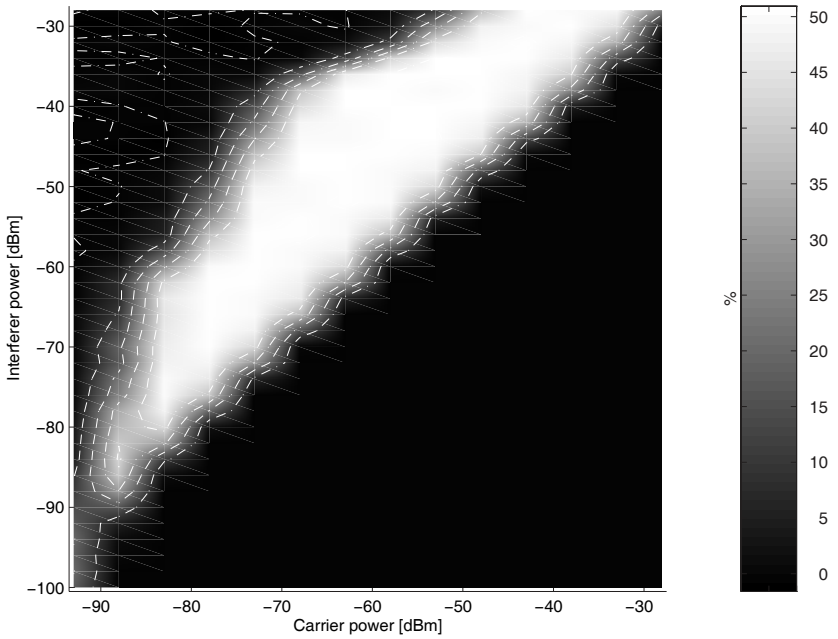


Figure 4.6: Improvement in BER using the adaptive antenna, in percentage units relative to one element antenna. Measurements performed in lab, on an 8 element array antenna, mean values over 10 measurements.

Section 4.5.3.

A deeper theoretical insight into what limits the suppression of the interfering signals is given in Chapter 5. There, it is shown that the weight quantization, both in the phase and the magnitude, limits the maximum achievable interferer suppression. For this particular testbed, it is the 1 dB step size in weight magnitude that dominates the performance degradation. The quantization in the weights also limits the calibration accuracy which have an impact on the performance. The number of bits of the ADC limits the range of received signal power where an improvement with the adaptive antenna is possible.

4.5 Outdoor measurements

4.5.1 The measurement range setup

The purpose of the outdoor field trial was first to characterize the system and its performance in a controlled flat fading environment with few variable parameters. In the next stage, measurements in a more complicated signal environment with the additional complexity of multipath propagation, frequency selective fading,

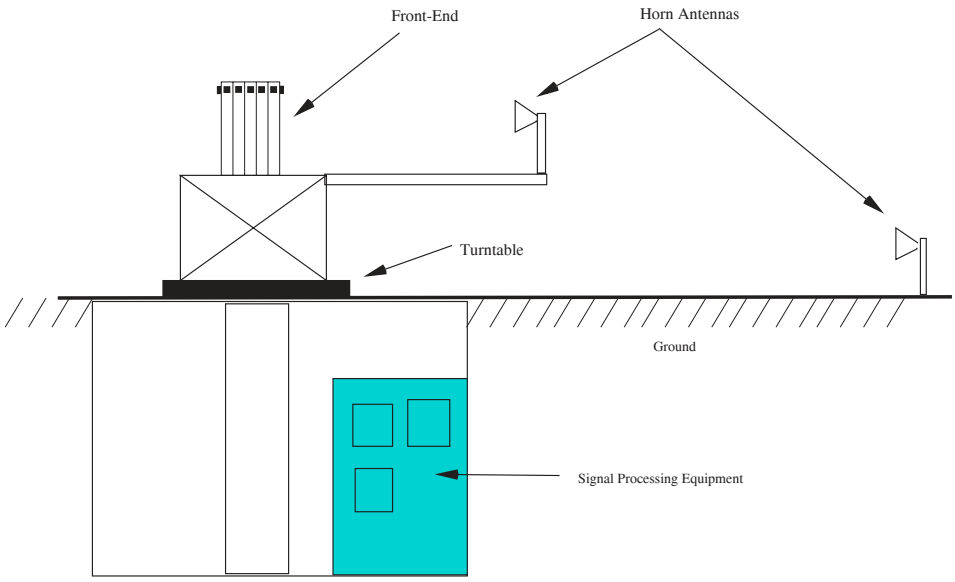


Figure 4.7: Antenna range setup. The front-end is placed on a turntable and is illuminated by two horn antennas. Below the turntable, a small room for the signal processing equipment.

Doppler spread etc. were to be performed. The first stage was performed at the outdoor antenna measurement range at FFV Aerotech in Arboga, Sweden.

The front-end was mounted on a turntable enabling 360 degrees azimuth rotation, see Figure 4.7. The circular array was then approximately three meters above the ground level. A five meter horizontal bar was mounted on the rotational turntable on which a horn antenna was placed. A second horn antenna was placed 30 cm above ground, 20 meters away from the turntable. The two horns were used as "mobiles" and illuminated the antenna from two directions. The separation is then adjustable by rotating the turntable. All equipment was stored in a room below the turntable. The distance to the closest obstacles (trees) is about 500 meters providing a rather echo free environment and a LOS channel. All experiments were made at the frequency of 1721 MHz.

4.5.2 Measurement of radiation patterns

Two signals with GSM standard GMSK modulation were generated and connected to their respective horn antennas. The separation angle between the interfering and desired signal was set by rotating the turntable. The adaptive antenna was then allowed to adapt on the signal environment. The adaption was then stopped and the weights were held constant. The two modulated signal sources were turned off

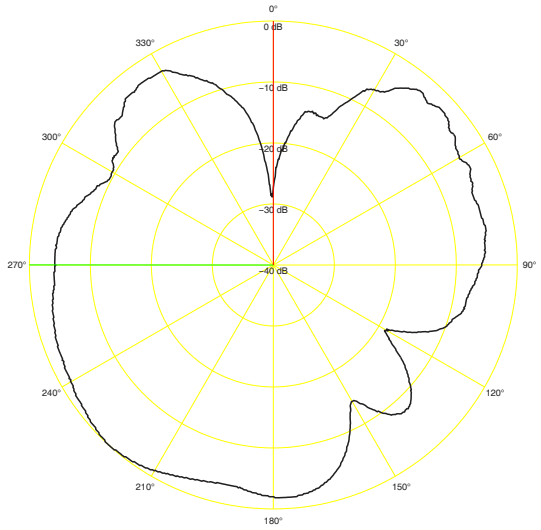


Figure 4.8: Adaptive antenna radiation pattern. Desired and interfering signals of equal strength with DOAs of 270° and 0° respectively. The interfering signal is suppressed 25 dB.

and a CW signal at 1721 MHz was connected to the field antenna.

During a complete rotation of the turntable, the output power from the adaptive antenna was measured at angles from 0 to 360 degrees in steps of half a degree, using a spectrum analyzer. This gives the receive "radiation" pattern[4] of the adapted beamformer. This measurement was then repeated for different angle separations and different power levels of the desired and interfering signals. Figure 4.8 shows the adapted radiation pattern of one of the beamformers, where $\theta_d = 270°$ and $\theta_i = 0°$. The separation angle between the desired and interfering mobile is thus 90 degrees and the power levels were set to C_{in}=I_{in}, thus CIR_{in}=0 dB. The SMI algorithm calculates a weight vector $\mathbf{w}_R$ the creates a null towards the interfering signal. The interfering signal is suppressed 25 dB relative to the desired signal.

Diagonal loading

As seen in Figure 4.8, there is no well defined main-lobe against the desired signal. This is due to the short data sequence used to estimate the covariance matrix and the steering vector. The covariance matrix $\mathbf{R_{yy}}$ and steering vector $\mathbf{r}_{\mathbf{y}d}$ in (4.1) are estimated based on the 26 samples in the training sequence. A study on how the number of samples affects the performance of the SMI algorithm can be found

[4]The term "radiation pattern" is used for both transmitting and receiving arrays.

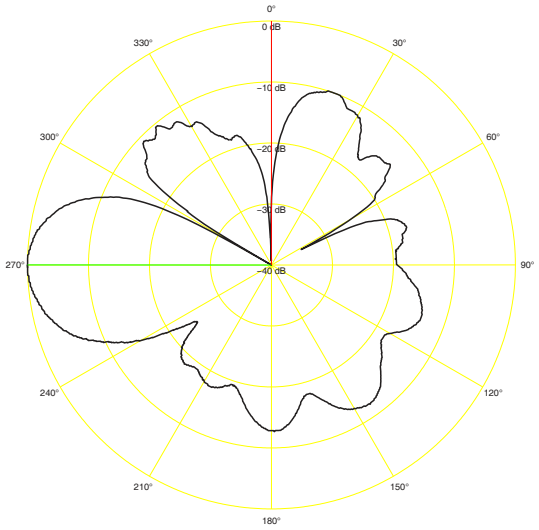


Figure 4.9: Effect of diagonal loading. The covariance matrix is regularized and the radiation patterns show lower sidelobes and a more stable radiation pattern between different noise realizations.

in [134]. With the desired signal present in the received signal, as is the case in a communication systems, the SMI algorithm converges slower than without its presence. One way of dealing with the slow convergence problem is to use the technique of diagonal loading, i.e. the adding of a small number to the diagonal elements of the covariance matrix. Another is to use bootstrapping [16], where the covariance matrix and steering vector is first estimated using the training sequence, and then the symbols in the whole time slot are detected. Then the symbols of the whole time slot is used as a "training sequence"to re-estimate $\mathbf{R_{yy}}$ and $\mathbf{r}_{\mathbf{y}d}$.

If we assume that the noise is spatially white, then the perfectly estimated (converged) covariance matrix would have all noise eigenvalues identical and equal to the noise variance σ_n^2. An unconverged estimate gives non-identical noise eigenvalues resulting in a radiation pattern with high side-lobes and the shape of the pattern changing between each batch of 26 samples, although the signal environment is stationary. By choosing the loading value larger than the noise eigenvalues but smaller than the eigenvalues of the desired and interfering signal acts as adding artifical noise and the overall equivalent noise level is increased. This results in almost identical noise eigenvalues [135]. The loading value L_0 was therefore chosen so that $L_0/\sigma_n^2 \sim 100$, where σ_n^2 is the noise power.

The diagonal loading will decrease the CIR_{out} somewhat, but increase the carrier to noise ratio CNR_{out} because of the decrease in side-lobe levels [134]. Figure 4.9 shows the measured radiation pattern where the signal environment is the same

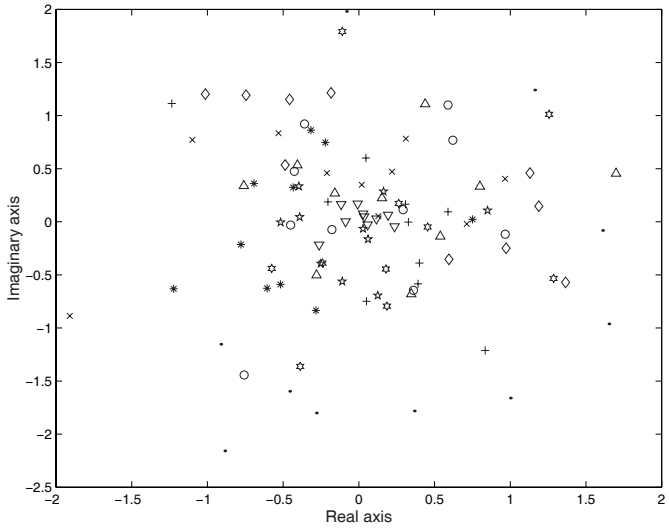


Figure 4.10: Weights in the complex plane. No diagonal loading and ten successive measurements on the same signal environment. Weights corresponding to the same channel is marked using the same symbol. Note how the solution changes completely between different batches of 26 training sequence symbols, due to the short sequence.

as in Figure 4.8 but with a diagonally loaded covariance matrix. The radiation pattern retains its null in the direction of the interfering signal, but now the main lobe points towards the desired signal with an estimated 3dB beamwidth of 30° and the side-lobe level is less than -10 dB.

In Figure 4.10 and Figure 4.11, the corresponding weights for ten successive adaptions are presented and can be compared, and in Figure 4.12, the effect on the radiation pattern is shown. It can be seen that almost the same weights makes successive radiation patterns very similar, hence the robustifying property of diagonal loading has been verified.

The dependence on the DOA separation

Figure 4.13(a) and 4.13(b) show radiation patterns of the two beam-formers where the desired signal and interferer are separated 2.5°. Note that the two signals impinging on the array play opposite roles in the two beam-formers. Despite the small angle separation, the interfering signal is suppressed more than 20 dB relative to the desired signal in both plots. Note also the similarity of the two radiation patterns. This indicates that apart from the opposite treatment of the two signals the beam-formers experience approximately the same noise environment.

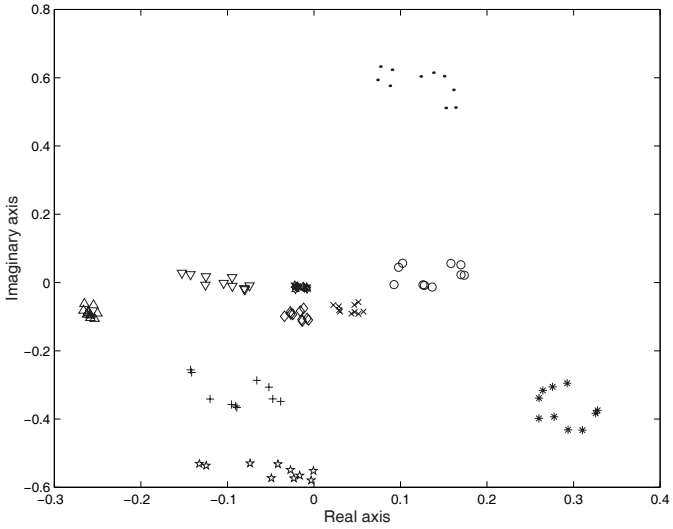


Figure 4.11: Weights in the complex plane when diagonal loading is used. Ten successive measurements are presented. Almost a similar weight solution arise every time, up to a constant phase shift on all weights, which has been removed in the figure.

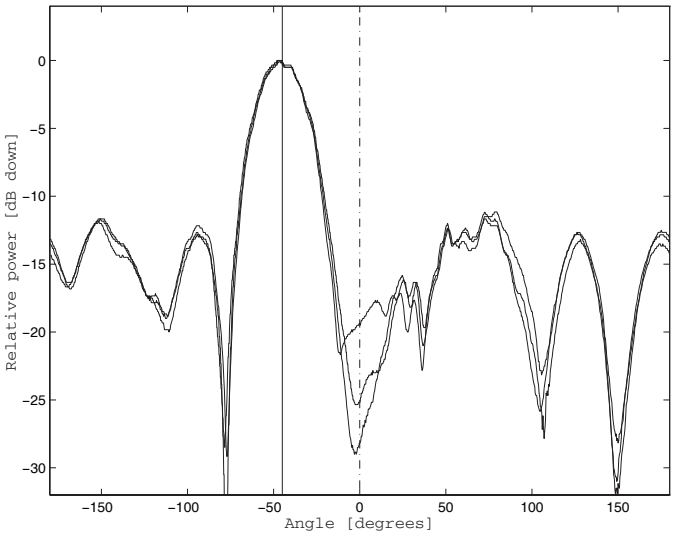
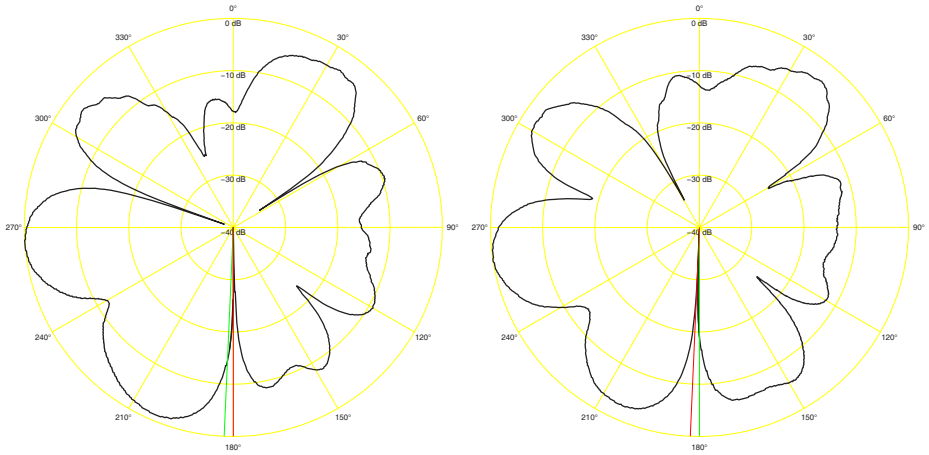


Figure 4.12: Three successive radiation patterns when diagonal loading is introduced. The DOAs and signal levels are identical for all three cases. Desired signal from $\theta_d = -48^\circ$ and interferer from $\theta_i = 0^\circ$.



(a) Beamformer one (b) Beamformer two

Figure 4.13: Desired and interfering signals of equal strength with DOA:s of 182.5° and 180°, respectively

For every measured radiation pattern, the suppression of the interferer and the amplification of the desired signal was estimated. The results for CIR_{in}=-20 dB is presented in Table 4.2. It can be seen that the suppression of the interferer is, within the error margin, independent of the DOA separation whereas the amplification of the desired signal decreases with a decreasing DOA separation. This is a natural consequence of the limited beam-width of the array, i.e. the most narrow beam possible with the actual array configuration is not narrow enough to provide the main-lobe peak towards the desired signal when a null is directed in the DOA of the interferer. The results of Table 4.2 is in agreement with equation (4.5) and (4.6). When the DOA separation is large, the spatial correlation $|\rho|$ is small and the output CINR (CIR) is approximately constant, as is the improvement in CIR. When the DOA separation angle becomes small, the spatial correlation approaches one and the performance degrades.

4.5.3 BER from outdoor measurements

The BER was logged from the supporting base station system while the antenna array was rotated. This gave the BER as a function of DOA separation between interfering and desired signal sources. The measurements were made for different settings of CIR_{in}. Figure 4.14 shows the BER after the decoding of class Ib bits by the forward error correcting (FEC) code. The class Ib bits are coded with a rate 1/2

DOA separation	Interferer suppression	Carrier gain	Improvement in CIR
180°	31 dB	2 dB	32 dB
135°	30 dB	4 dB	34 dB
90°	30 dB	4 dB	34 dB
45°	32 dB	2 dB	34 dB
10°	29 dB	2 dB	31 dB
5°	27 dB	-10 dB	17 dB
2.5°	26 dB	-8 dB	18 dB

Table 4.2: CIR_{out} improvement in dB for different DOA separation. CIR_{in}=-20 dB. The improvement in CIR is maintained above 30 dB until the DOA separation angle is smaller than 10°.

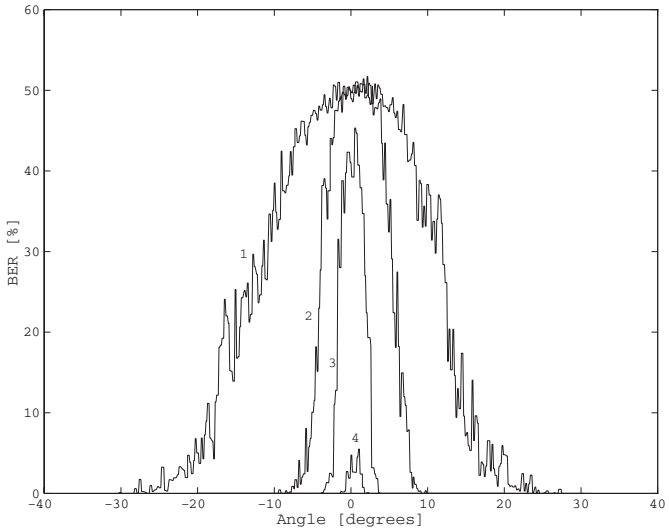


Figure 4.14: BER (after FEC decoding of class Ib data) as a function of interferer and desired signal DOA separation in the horizontal plane. CIR_{in}=-25 dB(curve 1), -20 dB(2), -15 dB(3), -10 dB(4) and 0 dB(5)

convolutional code and class II bits have no error correcting coding. Figure 4.15 presents the BER of unprotected, class II data. The transition between a BER of 0% and BER of 50% is very sharp when the CIR drops below a threshold value as is characteristic for a digital communication system. For CIR_{in}=-10 dB, the minimum angle separation for 0% BER in class Ib bits is less than 4 degrees and for decreasing CIR_{in} this minimum angle increases. The FEC improves the adaptive array antenna ability to separate users in the spatial dimension.

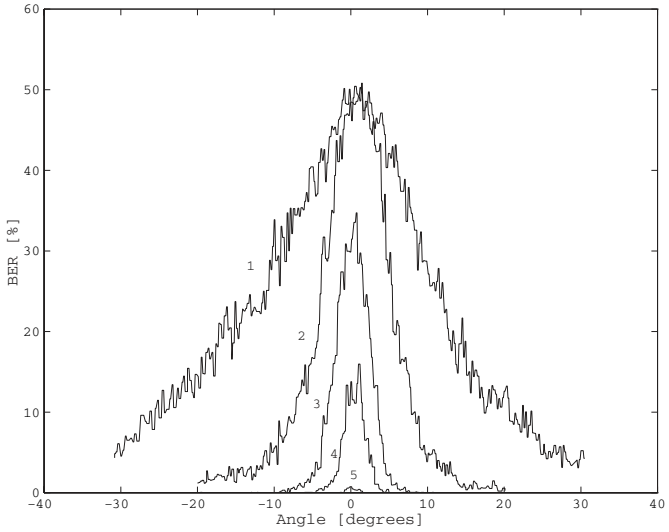


Figure 4.15: BER (Class II unprotected data) as a function of interferer and desired signal DOA separation in the horizontal plane. CIR_{in}=-25 dB(curve 1), -20 dB(2), -15 dB(3), -10 dB(4) and 0 dB(5)

For a CIR_{in}=0 dB a BER of 0% was measured although the plane waves impinging on the array only differed in elevation angle (same azimuth angle). This can be explained by the vertical beamforming of the circular array, i.e. plane waves with the same azimuth angle but with different elevation angles give rise to different spatial signatures or array response vectors, which makes it possible to separate them.

For practical reasons the two illuminating antennas were positioned at different heights as can be seen in Figure 4.7, giving a separation in the elevation plane of approximately eight degrees. Different spatial signatures and low correlation between the training sequences of the two signal sources are enough for the adaptation algorithm to separate the signals, and thus provide a beamforming property in elevation angle as well. This is a property that could be exploited in urban environments by designing arrays with volume extent, to separate users and interference in both azimuthal and horizontal angle.

4.6 Qualitative test in a multipath environment

The testbed has also been used in a multipath environment. The venue was a demonstration of the adaptive antenna using speech data transmission with “real” GSM-1800 cellular phones. The place was in central Kista outside of Stockholm

and the occasion was a demonstration of the project in cooperation with Ericsson Radio Access AB. The location can be classified as "typical urban". The demonstration gave qualitative results of the performance although no quantitative measurements was conducted.

In the demonstration, the adaptive antenna array was placed on the ground in a yard with high building surrounding it, thus a signal environment with severe multipath components due to many wall reflections. We successfully managed to connect two GSM-1800 cellular phones transmitting on the same frequency channel and timeslot using the testbed array antenna as a receiver. The voice transmission using the two cellular phones progressed perfectly although the physical separation between the two users was reduced to only 0.2 meters. This small separation distance corresponded to an angle separation at the basestation site of less than 1°, which is too small for bit error free transmission in the line of sight scenario.

Conclusively, the demonstration showed that with an array antenna, a multipath channel between the mobiles and the basestation actually helps the basestation to separate the users. The time delay spread for the multipath components must be considered small as the extent of the yard was less than 75×75 meters, so the channel can be approximated by a frequency flat channel[5]. The received signal vector was first space processed by the beamformer to a scalar signal which then was feed to the Viterbi detector in the base-station. Lindskog [16], presented a similar reduced rank MLSE equalizer that consisted of a beamformer followed by a temporal detector. Applied to measurement data, the performance degradation compared to a full rank vector MLSE equalizer was shown in [16] to be very small.

4.7 Conclusions

The laboratory tests show that the adaptive antenna presented here is capable of improving the CIR ratio more than 30 dB when the desired signal is weak and in the presence of a strong interferer. When the angle between the desired and the interfering signal is decreased the CIR improvement is reduced, mainly due to a loss in carrier gain. With a CIR=-20 dB the minimum separation between the desired and the interfering mobile for an bit-error-free transmission is 10 degrees. By adding forward error correction (FEC) to the transmitted symbols, the ability for the adaptive antenna array to separate users in the spatial dimension was shown to improve. This is due to the less susceptibility of a FEC coded data stream to interference, thereby the adaptive antenna interference suppression and the FEC combined leads to a system that is more robust against interference.

Furthermore, it was verified that diagonal loading is a way to improve the es-

[5]The equivalent length of a GSM symbol is approximately 1100 meters.

timation of the covariance matrix, giving lower sidelobes and a distinct main-lobe towards the desired signal. This is a method that also should be used in applications where the DOA to the transmitting signal is to be estimated using the covariance matrix of the received data.

This chapter demonstrated the feasibility of using adaptive antennas at base stations for mobile communication networks. Although the quantitative measurements were conducted in the simplest signal scenario with line of sight communication between the mobile transmitters and the base station array antenna, the results are not optimal in the sense that the interferer is not fully suppressed to the noise floor for all interferer power levels. This was expected by the theory for received power levels well inside the dynamic range of the receivers. The reason is due to the limited accuracy in the weights of the analog beamformer and the calibration errors. This is more thoroughly studied in Chapter 5.

A demonstration of the testbed in a multipath scenario showed that the weight updating frequency was sufficient to track the changes in the multipath channel. Although no qualitative measurements were performed, the voice quality test was satisfactory. Hence, the SMI algorithm is useful in this environment with multipath with short time delay spread. In [16] it was demonstrated that this type of a reduced rank equalizer has performance similar to the full rank space-time equalizer.

Chapter 5

Finite Weight Resolution and Calibration Issues

HARDWARE imperfections and related limitations on the interference suppression performance on a basestation array antenna are investigated in this chapter. It is shown how the phase and magnitude accuracy of the analog beamforming weights can be balanced for best performance and also which requirements lie on the accuracy of the calibration algorithm. Furthermore, two calibration algorithms are proposed that calibrate the antenna array simultaneously and without disturbing the normal antenna operation. The performance of the calibration algorithms are compared and it is concluded that they both can mitigate the temperature drift in antenna array hardware components.

5.1 Introduction

To be able to analyze the limitations of the testbed presented in the previous chapter, a theoretical model of the hardware channel is derived. The model can be used by future ABF designers to balance the resolution in phase and magnitude of the weighting units and to find the requirements on the calibration accuracy. Furthermore, the effects of finite word length in the DSP and the sampling receivers (SRX) that introduces quantization errors and limits the dynamic range of the receiver are also handled in this chapter. Due to the nonlinear nature of dynamic range limiting, it is not included in the derived theoretical hardware channel model, hence this chapter's aim of completely describe the hardware in the array antenna is only partly accomplished.

However, to extend the simplified (linear) model, both computer simulations

and practical measurements using the adaptive antenna testbed was performed. The testbed measurement results are used to validate the theoretical model and find in which ranges of operation the linear model is applicable. For a description of the testbed, see Section 4.2. The simulation model includes the saturation effects from the limited dynamic range of the SRX:s, to give a more accurate prediction of the adaptive antenna performance, especially at high input signal levels, where the SRX:s are close to saturation.

Implementation errors on adaptive antenna arrays has previously been studied by several authors. The previous work can be separated into the following four categories, which also serves as a division of the modeling problem into sub-areas:

Sampling receivers

If the received signal amplitude exceeds the maximum allowed amplitude level of the analog-to-digital converter (ADC) in the SRX, then nonlinear signal distortion will occur, and performance will quickly degrade. Due to the finite word length, the ADC will generate quantization noise which defines the lower limit of the SRX dynamic range. Previous studies by Hudson [136] and Takahashi *et al.* [137] investigated the necessary number of ADC bits to achieve a certain level of interference suppression. Hudson studied the quantization of received signals and concluded that eight bits in the ADC was necessary to give 40 dB interference suppression with a ten element array. The number of ADC bits must be chosen to cover the whole dynamic range of the received signals, which can be substantial due to the near-far ratio and the fading in the radio channel.

DSP

The finite word length in the digital signal processor affects the numerical stability and accuracy of matrix inversions used by many algorithms. Nitzberg studied the required word-length to achieve desirable performance [138]. It was showed that the spread of required precision is large and depends on the signal environment and as an example, the case with a single interference source requires the highest precision in the DSP (the largest DSP word-length). Performance limiting errors in the DSP depends also on the choice of algorithm. For instance, many algorithms use the covariance matrix to estimate the beamformer weights. Due to the time variant mobile channel the weights have to be updated frequently, and only the most recent samples are reliable for estimating the covariance matrix. The number of samples used in the estimation is an important parameter for the performance. Reed, Mallet and Brennan studied this [139], and showed how the carrier to interference ratio on the adaptive antenna output depends on the number of used samples and the

number of antennas (size of covariance matrix).

Weighting units

The weighting units have finite accuracy determined by the type of weight used, and the number of control bits from the DSP. Analog weights can be implemented in various ways, as two phase shifters in parallel, as two multipliers on the in-phase and quadrature branch respectively or as a phase shifter and a multiplier in series connection [140]. The effect of quantization of weights have previously been studied in [141, 142] and for phase weights only in [143] where it was concluded that the interference suppression capability was inversely proportional to the product of impinging signal power and weight error variance.

Calibration

A calibration must be performed to match the phase and amplitude of the different hardware channels. The calibration must track the variations in time due to temperature, humidity, etc and also be transparent and have no or a small effect on the normal operation of the adaptive array antenna [144]. Depending on the calibration method used, there will be a limited resolution in the calibration process and the residual calibration error will degrade the performance. Calibration errors was studied by Tsoulos *et al.* [12, 145], where measurements on an adaptive array using a calibration algorithm was presented. In Section 5.5, two on-line calibration algorithms that are transparent to main antenna operation are presented that continuously tracks the changes in antenna channels. The algorithms use a feedback of the output RF signal, and thus allows the use of e.g. the LMS algorithm for updating the weights to minimize the error between the desired and actual output.

The derivations in this chapter does not depend on the actual implementation of the weighting units or the choice of the algorithm for calibrating the ABF. A generic error model is assumed and the differences in weighting units enter through the variance of the weight errors and maximum value of the calibration error in phase and amplitude. The obtained results can be used to find the effects of different implementation techniques on the total performance of an adaptive antenna array of the ABF type in use in a mobile communication system. The performance bottle-necks are identified to help the designer to achieve a balanced dimensioning of the adaptive array antenna.

The chapter is organized as follows, in Section 5.2 the hardware channel model is derived, Section 5.3 describes the adaptive antenna testbed and Section 5.4 the results from measurements and simulations. Two on-line calibration algorithms for the ABF is presented in Section 5.5 and finally, the chapter ends with a discussion

in Section 5.6.

5.2 The hardware channel model

In this section, the hardware channel model is derived. Figure 5.1 shows the ABF with the analog weighting path and the digital path which connects the antenna elements with the sampling receivers. It is assumed that all signals are represented by their complex baseband equivalents. The noise generated by the low noise amplifiers in the receiver front-end and the noise received by the n_r antennas are modelled as an equivalent $n_r \times 1$ noise vector, $\mathbf{n}_T(t)$, which is assumed to be spatially and temporally white[1]. Hence, $\mathbf{n}_T(t)$ is a vector with zero mean variables and covariance matrix $\mathbf{R}_{\mathbf{n}_T} = \sigma_n^2\mathbf{I}$.

To isolate the performance evaluation to the hardware channel, the simplest possible radio propagation channel is assumed; two narrowband signals impinge on an uniform linear array (ULA) antenna from two distinct azimuthal directions θ_d and θ_i. Hence the channel is LOS and with zero angular spread. The channel vector $\mathbf{h}$ can the be replaced by array response vectors, see the discussion in Section 4.3.

The continous time signal received by the n_r antennas is then described by the $n_r \times 1$ vector $\mathbf{y}(t)$ as

$$\begin{aligned}\mathbf{y}(t) =& \mathbf{y}_d(t) + \mathbf{y}_i(t) + \mathbf{n}_T(t) = \\ =& \mathbf{a}(\theta_d)s_d(t) + \mathbf{a}(\theta_i)s_i(t) + \mathbf{n}_T(t)\end{aligned} \tag{5.1}$$

where $\mathbf{a}(\theta_d)$ and $\mathbf{a}(\theta_i)$ is the complex valued array response vector in azimuth direction θ_d and θ_i respectively, including antenna element gain and polarization. The baseband signals $s_d(t)$ and $s_i(t)$ denote the desired and interferer signal respectively.

In the model, the sampling receivers (SRX) are replaced by limiters and a noise source $\mathbf{n}_r(t)$. This noise represents the internal noise generated in the receivers and the quantization noise generated in the sampling process. The corresponding signal vector in the DSP can thus be written as

$$\mathbf{y}'(t_k) = \text{sat}\left[\mathbf{a}(\theta_d)s_d(t_k) + \mathbf{a}(\theta_i)s_i(t_k) + \mathbf{n}(t_k)\right] \tag{5.2}$$

where t_k represents the sampling instants and $\mathbf{n}(t_k)$ is the sum of front end thermal noise $\mathbf{n}_T(t)_{t=t_k}$, quantization noise and receiver noise $\mathbf{n}_r(t)_{t=t_k}$:

$$\mathbf{n}(t_k) = \mathbf{n}_T(t_k) + \mathbf{n}_r(t_k) \ . \tag{5.3}$$

[1]This assumption means that no other interference sources are present which have a directional property. Furthermore, the white noise generated in each antenna branch in the front end are assumed mutually independent. Thus the noise is both temporally and spatially white.

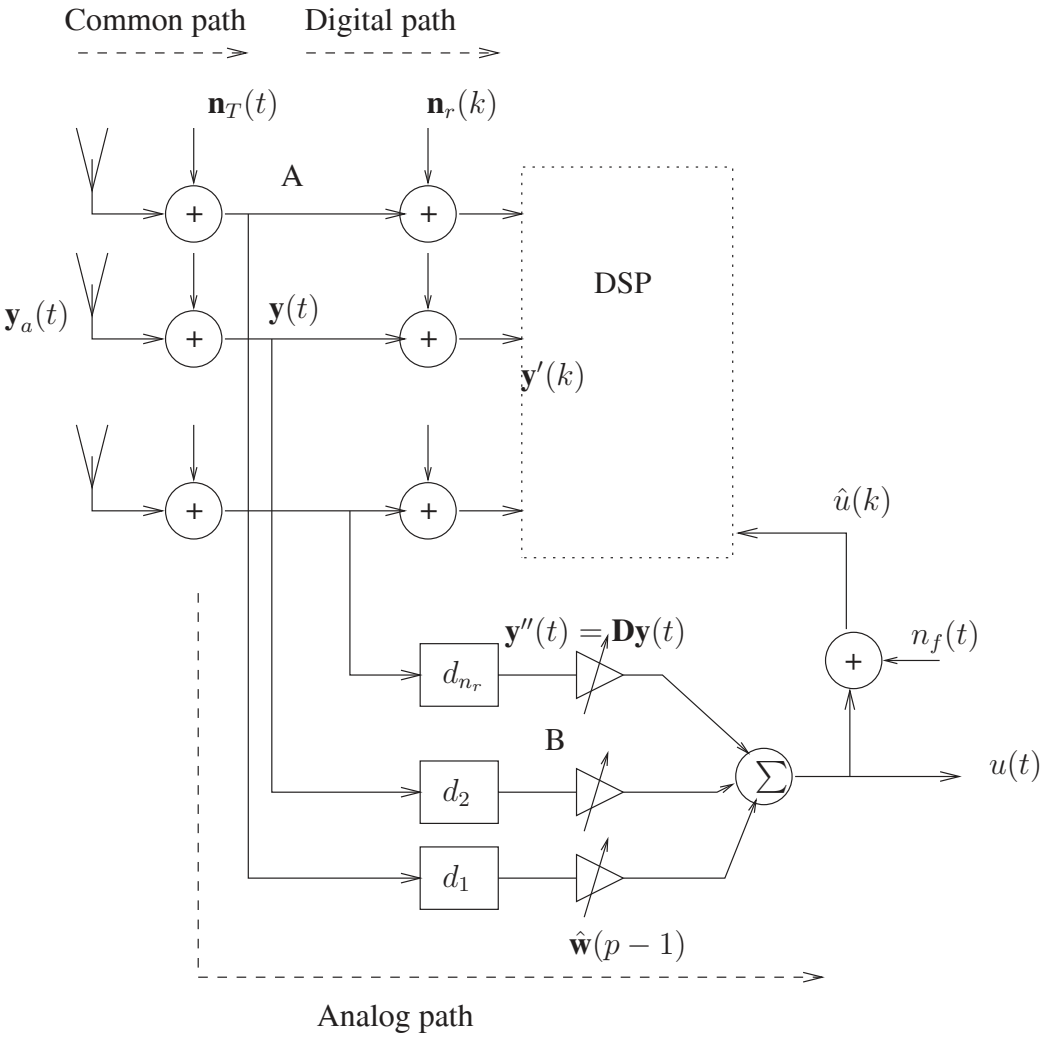


Figure 5.1: Block diagram of the adaptive antenna. Showing only one set of the beamformer weights, $\hat{\boldsymbol{w}}(p-1)$, and $u(t)$ is the corresponding output signal from the adaptive antenna.

The saturation operator, sat$[\cdot]$, hard-limits the quadrature signals above a certain amplitude level. The maximum amplitude level in each quadrature branch before saturation is dependent on the automatic gain control (AGC) setting of the receivers and the dynamic range of the ADC. If the signal voltage amplitude for any of the quadrature branches exceeds the maximal allowed amplitude, then the linear model does not apply and the signal is distorted. From here on, the nonlinear effects from signal saturation are not considered, to make the analysis analytically tractable.

5.2.1 Off-line Calibration

The aim of the calibration is to estimate the transfer function of the hardware channel between the SRX (point A in Figure 5.1) and the summation point after the weighting units (point B). The estimated transfer functions for each hardware channel is used to compensate the corresponding weight in software before it is multiplied with the signal. This compensation is important because the weights are calculated based on the signals from the SRX but applied to the signals entering the weighting units. Assume that the channel between the SRX and the weighting units is wideband compared to the received narrowband signals and that the transfer function is frequency flat over the passband of interest. This implies that the transfer function can be represented by a complex number representing the gain/attenuation and the phase shift of the channel. Note that this assumption might not be valid in a wideband system such as W-CDMA, if the transfer function cannot be assumed to be flat over the whole system bandwidth.

By introducing the complex matrix $\mathbf{D} = \text{diag}\,[d_1, \ldots, d_{n_r}]$, the relation of the signal at the weighting units, $\mathbf{y}''(t)$, and the SRX, $\mathbf{y}(t)$, can be expressed as

$$\mathbf{y}''(t) = \mathbf{D}\mathbf{y}(t) \ . \tag{5.4}$$

The off-diagonal elements in $\mathbf{D}$, which here are assumed negligible, represent the mutual coupling between the channels in the beamformer. Assume that the algorithm calculates the weight vector $\mathbf{w}_0$, based on the received signals $\mathbf{y}'(t_k)$. To compensate for the differences in the receiving channels, the weights are pre-adjusted to

$$\mathbf{w} = (\mathbf{D}^{-1})^*\mathbf{w}_0 \ . \tag{5.5}$$

The analog beamformer output signal $u(t)$ will then be

$$u(t) = \mathbf{w}^*\mathbf{y}''(t) = \mathbf{w}_0^*\mathbf{D}^{-1}\mathbf{D}\mathbf{y}(t) = \mathbf{w}_0^*\mathbf{y}(t) \ . \tag{5.6}$$

Thus the undesired effect of the transfer function $\mathbf{D}$ is cancelled.

When the calibration is performed, the weights are adjusted to measure the transfer function for different weight settings. This implies that the accuracy in the weight settings will also have an impact on the accuracy of the calibration. Also, due to temperature drift, humidity variations and component aging, the calibration correction matrix has a time variant residual error.

Assume that the array has been calibrated by some arbitrary method and let the matrix $\hat{\mathbf{D}}$ be the estimated calibration correction matrix that is stored in the DSP. If the calibration is error-free, then $\hat{\mathbf{D}} = \mathbf{D}$. Now, write $\hat{\mathbf{D}}^{-1}\mathbf{D} = \mathbf{I}+\delta\mathbf{D}$, where $\delta\mathbf{D}$ is a diagonal matrix with complex elements, representing the relative calibration errors. Writing the diagonal elements of $\delta\mathbf{D}$ as $\delta d_l = a_l e^{j\phi_l}$, the relative calibration errors

are separated into an amplitude error a_l and a phase error ϕ_l. The amplitude a_l is assumed to be bounded in the range $[\pm\epsilon_a]$ and the phase error ϕ_l is assumed to be in the range $[\pm\epsilon_\phi]$. As the phase and magnitude errors are orthogonal, see Figure

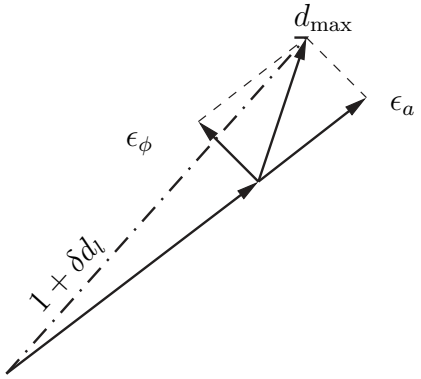


Figure 5.2: The squared calibration error is bounded by $d_{max}^2 = \epsilon_a^2 + \epsilon_\phi^2$

5.2, the maximum calibration error squared, d_{max}^2 is given as

$$d_{max}^2 = \epsilon_a^2 + \epsilon_\phi^2 \ . \tag{5.7}$$

5.2.2 The output signal power

As a performance measure for the adaptive antenna, the carrier to interference plus noise ratio (CINR) in the beamformer output signal is used. An expression for the output power is derived and the different terms are identified as the desired (carrier) terms and the interferer and noise terms. These are then used to calculate the CINR. Some simplifying assumptions are made to make the analysis possible. The saturating effect of the SRX (5.2), is not taken into consideration, and the sources of the two signals, $s_d(t)$ and $s_i(t)$, are assumed to be sufficiently separated in azimuth to make the spatial correlation close to zero. Due to the finite step size in the hardware weighting units, the weights $\mathbf{w}$ will be quantized and an error vector $\delta_{\mathbf{w}}$ is introduced. The total weight error covariance matrix is $E\{\delta_{\mathbf{w}}\delta_{\mathbf{w}}^*\} = \sigma_w^2\mathbf{I}$. An example of how σ_w^2 can be derived for a specific type of weighting units is presented in Section 5.3.3. The signal on the beamformer output can be written as

$$\begin{aligned} u(t) &= (\mathbf{w}+\delta_{\mathbf{w}})^*\hat{\mathbf{D}}^{-1}\mathbf{D}\mathbf{y}(t) = (\mathbf{w}+\delta_{\mathbf{w}})^*(\mathbf{I}+\delta\mathbf{D})\mathbf{y}(t) = \\ &= \mathbf{w}^*\mathbf{y}(t) + \text{error terms} \end{aligned} \tag{5.8}$$

and the power of the beamformer output signal is then written as

$$E\left\{|u(t)|^2\right\} = E\left\{\left[(\mathbf{w}+\delta_{\mathbf{w}})^*\hat{\mathbf{D}}^{-1}\mathbf{D}\mathbf{y}(t)\right]\left[(\mathbf{w}+\delta_{\mathbf{w}})^*\hat{\mathbf{D}}^{-1}\mathbf{D}\mathbf{y}(t)\right]^*\right\} \ . \tag{5.9}$$

Using equation (5.8) and assuming that the elements of the stochastic vectors $\delta_{\mathbf{w}}$ and the matrix $\delta\mathbf{D}$ are mutually uncorrelated, and also independent of the received vector $\mathbf{y}(t)$, equation (5.9) is simplified to

$$\begin{aligned} E\left\{|u(t)|^2\right\} = \mathbf{w}^* E\left\{\mathbf{y}(t)\mathbf{y}^*(t)\right\}\mathbf{w} + E\left\{\delta_{\mathbf{w}}^*\mathbf{y}(t)\mathbf{y}^*(t)\delta_{\mathbf{w}}\right\} + \\ + \mathbf{w}^* E\left\{\delta\mathbf{D}\mathbf{y}(t)\mathbf{y}^*(t)\delta\mathbf{D}^*\right\}\mathbf{w} + E\left\{\delta_{\mathbf{w}}^*\delta\mathbf{D}\mathbf{y}(t)\mathbf{y}^*(t)\delta\mathbf{D}^*\delta_{\mathbf{w}}\right\} \quad . \end{aligned} \tag{5.10}$$

If the SRX:s are saturated, then elements of the weight error vector $\delta_{\mathbf{w}}$ will become mutually correlated and also the independence assumption between the weight errors $\delta_{\mathbf{w}}$ and the signal $\mathbf{y}(t)$ becomes invalid. Note that $\mathbf{y}(t)$ is the analog RF signal (5.1) which is not saturated.

The terms in (5.10) can be identified as the ideal output term, the weight error term, the calibration error term and the combined weight-calibration error term respectively. The combined weight-calibration error term contains the product of the two error variables and is neglected in the following. The weight and calibration error term are quadratic forms with the error terms $\delta_{\mathbf{w}}$ and $\delta\mathbf{D}$ and are assumed to be sources of interference. This is a pessimistic assumption, because these terms also include some signal which are correlated with the desired signal and might add coherently to the output desired power.

The ideal output term

The first term in (5.10) is the output signal from a beamformer without calibration or weight quantization errors. This term is dominant in the DBF type of antenna, where weighting of the signals is performed in the DSP itself, so weight errors are negligible and calibration errors of the type considered here is not at all present.

The performance of the error-free beamformer is dependent on the algorithm used for calculating the weights $\mathbf{w}$. In a temporal reference algorithm or a direction finding algorithm, the number of samples N_S used in the estimation of the covariance matrix compared to the number of antennas is an important factor [139]. Due to the finite length of the training sequence, the temporal correlation between the two signal sources, $s_d(t)$ and $s_i(t)$, will be nonzero. Also the azimuthal separation angle between the two sources, quantified as the spatial correlation, will have impact on the performance of the algorithms, c.f. Chapter 4.

For the analog signal $\mathbf{y}(t)$, the following holds under the LOS assumption

$$\begin{aligned} \mathbf{R}_{\mathbf{yy}} = E\left\{\mathbf{y}(t)\mathbf{y}^*(t)\right\} = \sigma_d^2\mathbf{a}(\theta_d)\mathbf{a}^*(\theta_d) + \sigma_i^2\mathbf{a}(\theta_i)\mathbf{a}^*(\theta_i) + \sigma_n^2\mathbf{I} = \\ = \mathbf{R}_d + \mathbf{R}_i + \sigma_n^2\mathbf{I} \end{aligned} \tag{5.11}$$

where $\mathbf{R}_d$ and $\mathbf{R}_i$ are the desired and interfering signal's spatial covariance matrix respectively. Also, σ_d^2 and σ_i^2 is the desired and interfering signal power, respec-

tively. The ideal term in (5.10) can now be written and upper bounded as

$$\begin{aligned}\mathbf{w}^* E\left\{\mathbf{y}(t)\mathbf{y}^*(t)\right\}\mathbf{w} &= \mathbf{w}^*\mathbf{R}_d\mathbf{w} + \mathbf{w}^*\mathbf{R}_i\mathbf{w} + \sigma_n^2\left|\mathbf{w}\right|^2 \\ &\leq \mathbf{w}^*\mathbf{R}_d\mathbf{w} + \mathbf{w}^*\mathbf{R}_i\mathbf{w} + n_r\sigma_n^2 w_{max}^2\end{aligned} \tag{5.12}$$

where w_{max} is the maximum allowed weight magnitude, determined by the hardware and σ_n^2 is the thermal noise variance. The first term in equation (5.12) represents the power in the desired, or useful signal. The second term is the interference output power, which also is dependent on the beamformer weights $\mathbf{w}$. Clearly, the adaptive beamformer tries to calculate the weight $\mathbf{w}$ so that the Hermitian form $\mathbf{w}^*\mathbf{R}_i\mathbf{w}$ is small, to suppress the interferer. This term will be nonzero if the algorithm has not converged, which can be the case for recursive algorithms in a time variant radio channel or for block algorithms if a finite number of samples are used in the weight estimation. Note that the weight error $\delta_{\mathbf{w}}$ is due to the use of hardware weighting units, and will be present even if the exact, interference cancellation weight $\mathbf{w}$ has been estimated. To mitigate this residual error, a feedback of the beamformer output signal $u(t)$ could be used to iteratively tune the weights so the weight and calibration errors decreases. A method for this is presented in section 5.5.

The third term in (5.12) represents the sum of noise power from n_r uncorrelated noise sources with same power $w_{max}^2\sigma_n^2$.

The weight-error term

The second term in (5.10) is due to weight quantization and is expanded as

$$\begin{aligned}E_{y,\delta_{\mathbf{w}}}\left\{\delta_{\mathbf{w}}^*\mathbf{y}(t)\mathbf{y}^*(t)\delta_{\mathbf{w}}\right\} &= E_{y,\delta_{\mathbf{w}}}\left\{\mathrm{Tr}\left\{\mathbf{y}(t)\mathbf{y}^*(t)\delta_{\mathbf{w}}\delta_{\mathbf{w}}^*\right\}\right\} \\ &= \mathrm{Tr}\left\{E_{\mathbf{y}}\left\{\mathbf{y}(t)\mathbf{y}^*(t)\right\}E_{\delta_{\mathbf{w}}}\left\{\delta_{\mathbf{w}}\delta_{\mathbf{w}}^*\right\}\right\} \\ &= \sigma_w^2\mathrm{Tr}\left\{\mathbf{R}_{\mathbf{yy}}\right\} \\ &= \sigma_w^2(\sigma_d^2|\mathbf{a}(\theta_d)|^2 + \sigma_i^2|\mathbf{a}(\theta_i)|^2 + n_r\sigma_n^2)\ .\end{aligned} \tag{5.13}$$

The property that the trace of the covariance matrix is equal to the sum of the received signal power and that the signal and the weight errors are mutually independent was used.

Note that this only is valid under the assumption of non-saturated SRX:s since otherwise, $\delta_{\mathbf{w}}$ and $\mathbf{y}(t)$ are dependent. The weight errors, $\delta_{\mathbf{w}}$, causes signal power to leak through the beamformer, with a power proportional to the total impinging power. The amount of "leaking" power is determined by weight error variance σ_w^2.

The calibration-error term

The third term in (5.10) includes the diagonal matrix $\delta\mathbf{D}$ that models the calibration errors. The structure is a Hermitian form and the upper bound is found as

$$\begin{aligned}\mathbf{w}^* E_{\mathbf{y}} \left\{ \delta\mathbf{D}\mathbf{y}(t)\mathbf{y}^*(t)\delta\mathbf{D}^* \right\} \mathbf{w} &= \mathbf{w}^*\delta\mathbf{D}\mathbf{R}_{\mathbf{yy}}\delta\mathbf{D}^*\mathbf{w} \\ &\leq \lambda_{max}|\mathbf{w}^*\delta\mathbf{D}|^2 \\ &\leq \mathrm{Tr}\left\{\mathbf{R}_{\mathbf{yy}}\right\}|\mathbf{w}^*\delta\mathbf{D}|^2\end{aligned} \tag{5.14}$$

where λ_{max} is the largest eigenvalue of the Hermitian matrix $\mathbf{R}_{\mathbf{yy}}$. The bound cannot be tighter, since the bound is attained when the error vector $\mathbf{w}^*\delta\mathbf{D}$ is co-linear with the eigenvector to $\mathbf{R}_{\mathbf{yy}}$ associated with the largest eigenvalue. The largest eigenvalue has to be found on a case-by-case basis so the worst possible case is examined. Since $\mathbf{R}_{\mathbf{yy}}$ is Hermitian positive definite, all eigenvalues are real and positive and the largest eigenvalue is always less or equal to the trace of $\mathbf{R}_{\mathbf{yy}}$.

Furthermore, the vector norm expression (5.14) cannot be evaluated, because it depends on the magnitude of the weights and the calibration errors. An upper bound can be found by assuming that all weight magnitudes are smaller than w_{max}, which is the maximum possible weight that can be steered out by the hardware weighting units. Similarly, the upper bound of the calibration errors d_{max} is defined. Thus by setting $|w_l|^2 = w_{max}^2$ for all weights and $|d_l|^2 = d_{max}^2$ for all calibration paths, the vector norm in expression (5.14) can be written as

$$\begin{aligned}|\mathbf{w}^*\delta\mathbf{D}|^2 = |w_1^* d_1|^2 + \cdots + |w_{n_r}^* d_{n_r}|^2 \leq |w_1|^2|d_1|^2 + \cdots + |w_{n_r}|^2|d_{n_r}|^2 \leq \\ \leq n_r w_{max}^2 d_{max}^2\end{aligned} \tag{5.15}$$

where the Cauchy-Schwartz inequality was used. Now, use that the trace of $\mathbf{R}_{\mathbf{yy}}$ is equal to the total impinging power on the array and rewrite (5.14) as

$$\mathbf{w}^* E_{\mathbf{y}} \left\{ \delta\mathbf{D}\mathbf{y}(t)\mathbf{y}^*(t)\delta\mathbf{D}^* \right\} \mathbf{w} \leq n_r w_{max}^2 d_{max}^2 (\sigma_d^2|\mathbf{a}(\theta_d)|^2 + \sigma_i^2|\mathbf{a}(\theta_i)|^2 + n_r\sigma_n^2). \tag{5.16}$$

Similar to the weight error term (5.13), the total impinging power, described by the trace of $\mathbf{R}_{\mathbf{yy}}$ is weighted by a proportionality constant. The constant in this case depends on the maximum calibration error d_{max} and is bounded by the squared root of the sum of the squared magnitude and phase errors (5.7).

The output CINR

Using the equations above, the expression for the beamformer output CINR can now be formulated. The power of the desired, or useful signal in the beamformer

output is described by the first term in equation (5.12). The two other terms in equation (5.12) and the equations (5.13) and (5.16) are assumed to be interferer and noise power terms. Thus, the beamformer output CINR lower bound (worst case) can be expressed as:

$$\text{CINR} \geq (\mathbf{w}^* \mathbf{R}_d \mathbf{w}) \, / \\ \left(\mathbf{w}^* \mathbf{R}_i \mathbf{w} + n_r \sigma_n^2 w_{max}^2 + (\sigma_w^2 + n_r d_{max}^2 w_{max}^2)(\sigma_d^2 |\mathbf{a}(\theta_d)|^2 + \sigma_i^2 |\mathbf{a}(\theta_i)|^2 + n_r \sigma_n^2) \right) . \tag{5.17}$$

To maximize this expression, the weight calculation algorithm should make $\mathbf{w}^* \mathbf{R}_i \mathbf{w}$ as small as possible to yield an antenna radiation pattern with "nulls" in the direction of the interfering sources.

The expression (5.17) was derived using several simplifications, but it illustrates how the desired and interfering signal power "leaks" through the beamformer, due to calibration errors and weight quantization. The system designer should choose the implementation of the weighting units so that the weight error variance σ_w^2 has the same order of magnitude as the term $n_r d_{max}^2 w_{max}^2$ that represents the calibration errors. In the limit when calibration errors and the weight quantization errors approaches zero, i.e. $d_{max} \rightarrow 0$ and $\sigma_w^2 \rightarrow 0$, the CINR approaches the well known expression for CINR of a digital beamformer,

$$CINR_{DBF} \simeq \frac{\mathbf{w}^* \mathbf{R}_d \mathbf{w}}{\mathbf{w}^* \mathbf{R}_i \mathbf{w} + n_r \sigma_n^2 w_{max}^2} \ . \tag{5.18}$$

How the different terms in the denominator of (5.17) dominates in different cases is investigated by the simulations and the measurements presented in Section 5.4.

5.3 The adaptive antenna testbed

To validate the theoretical expression for the CINR (5.17), measurements using the adaptive antenna testbed described in Section 4.2 were performed. The weight error variance for the testbed specific type of weighting units is derived, followed by a discussion of the measurement setup.

5.3.1 Hardware

The main parameters of the testbed hardware can be found in Table 4.1. The testbed has two independent sets of weights with ten weighting units per set. Each digitally controlled weighting units consists of two 180° phase-shifters and a 50 dB logarithmic attenuator connected in series. The reason for choosing logarithmic attenuators, was that in the design process, it was believed that a high dynamic range

of the weights were beneficial. The logarithmic attenuator was thus chosen in favor of the linear one. The logarithmic attenuator has a range of attenuation of 50 dB but it was later discovered that a 50 dB range is over-abundant. During operation it was observed that the beamformer seldom uses weight magnitudes below -15 dB. The weighting units have a calculated noise figure of 6 dB and a measured third order intercept point of +6 dBm. The temperature drift was measured to 0.1 dB and 1° per weighting unit and hour of operation. Hence, frequent re-calibration or on-line calibration is necessary to maintain the highest level of performance.

Calibration of the antenna array is performed prior to normal operation of the testbed. A continuous wave (CW) signal is injected at one antenna element at a time by directional couplers. The received signal is compared to the beamformer output signal and the phase and magnitude of the transfer functions $a_l e^{j\phi_l}$ can be measured. The calibration has an accuracy of 1° in phase and 0.75 dB in magnitude [123]. Using equation (5.7), the relative calibration error constant d^2_{max} can be calculated as $d^2_{max} = 8.4 \cdot 10^{-3}$.

5.3.2 The SMI algorithm

The DSP calculates the weights by using the sample matrix inverse (SMI) algorithm [127]. As a reference signal, the 26 bit training sequence in the mid-amble of a GSM-1800 timeslot is used. The covariance matrix of the received signals and the cross-correlation between the received signal vector $\mathbf{y}(t_k)$ and the reference signal $d(t_k)$ is estimated using $N = 26$ samples. The signal covariance matrix estimate is then

$$\hat{\mathbf{R}}_{\mathbf{yy}} = \frac{1}{N_S} \sum_{k=1}^{N_S} \mathbf{y}'(t_k)\mathbf{y}'^*(t_k) \ . \tag{5.19}$$

The cross-correlation is estimated as

$$\hat{\mathbf{r}}_{\mathbf{y}d} = \frac{1}{N_S} \sum_{k=1}^{N_S} \mathbf{y}'(t_k)d^*(t_k) \ . \tag{5.20}$$

The reference signal, $d(t_k)$, is a modulated, pre-recorded copy of the training sequence of the desired signal, stored in the DSP memory.

The mean square optimal receive weight vector is obtained by solving the Wiener-Hopf equations, see Section 2.5.1

$$\hat{\mathbf{w}}_0 = \alpha \hat{\mathbf{R}}^{-1}_{\mathbf{yy}} \hat{\mathbf{r}}_{\mathbf{y}d} \tag{5.21}$$

where the factor α scales the weights for full utilization of the dynamic range of the hardware weights. The weight vector is adjusted by the estimated calibration matrix $\hat{\mathbf{D}}$ and applied to the RF signals $\mathbf{y}''(t)$, see Figure 5.1.

The SMI algorithm has a fast convergence rate as compared to other known algorithms, e.g. the LMS algorithm. Approximately $2n_r$ samples in the block are required to obtain weights that give a CINR within 3 dB of the optimum achievable CINR [139]. Here, n_r=8 antennas is used, so $N_S = 26 > 2n_r = 16$.

5.3.3 The weight error variance

Given the phase and magnitude resolution of these hardware parts, the weight error variance is derived for logarithmic amplitude and linear phase weights. Weight error variances for other weight implementations can be found in the paper by Davis and Sher [140], especially for the linear amplitude weighting technique.

An arbitrary weight with magnitude A and phase θ with magnitude accuracy $\pm\epsilon_A$ and phase accuracy $\pm\epsilon_\theta$ can be written as

$$w = Ae^{j\theta} \quad . \tag{5.22}$$

Assuming that the magnitude and phase errors are small, and by differentiation of both sides of (5.22) and collecting terms gives

$$|dw|^2 = dA^2 + A^2 d\theta^2 \tag{5.23}$$

where dA is the magnitude error and $d\theta$ is the phase error. Assume that dA and $d\theta$ are independent random variables and that dA is uniformly distributed in the interval $\pm\epsilon_A$ and $d\theta$ uniformly distributed in the interval $\pm\epsilon_\theta$. Taking the expectation value of (5.23) then gives

$$\sigma_w^2 = \sigma_A^2 + A^2\sigma_\theta^2 \tag{5.24}$$

where $\sigma_A^2 = \epsilon_A^2/3$ and $\sigma_\theta^2 = \epsilon_\theta^2/3$.

The total error variance (5.24) is thus dependent on the weight magnitude A. An upper bound on the weight error variance is given by replacing A with the maximum weight magnitude w_{max}, thus

$$\sigma_w^2 \leq \sigma_A^2 + w_{max}^2\sigma_\theta^2 = \frac{1}{3}\left(\epsilon_A^2 + w_{max}^2\epsilon_\theta^2\right) \quad . \tag{5.25}$$

The contribution from the phase error to the sum in equation (5.25) is thus dependent on the maximal weight magnitude. Note for large attenuator settings, say -15 dB, the difference to the next weight setting of -16 dB is very small if measured on a linear scale. Hence small values of the weight amplitudes makes the phase error part of the weight quantisation noise small. The largest error variance occurs when the weight amplitude is set at its maximum value.

The weighting units in the testbed has a phase resolution of 1° and uses logarithmic attenuators with 1 dB magnitude step size and a 50 dB range between [-2

dB,-52 dB]. Thus, the constant w_{max}=-2 dB. The weight magnitude variance is then calculated using (5.25) as

$$\begin{aligned} \sigma_A^2 &= \frac{(0.5 \cdot (10^{\frac{1}{20}} - 1))^2}{3} = 1.24 \cdot 10^{-3} \\ \sigma_\theta^2 &= \frac{(\frac{0.5 \cdot \pi}{180})^2}{3} = 0.025 \cdot 10^{-3} \\ w_{max}^2 &= 10^{-\frac{2}{10}} \\ \Rightarrow \sigma_w^2 &\leq 1.25 \cdot 10^{-3} \end{aligned} \tag{5.26}$$

where a conversion from dB scale to the linear scale is performed. This implies that the proportionality constant $n_r d_{max}^2 w_{max}^2$ in equation (5.17) is an order of magnitude larger than the weight error constant σ_w^2. Thus, the calibration errors limits the performance for the testbed, which is also observed in the simulation results. It should also be noted that the main contribution to the weight error variance of the testbed arise from magnitude errors (the step attenuators). The phase error contribution is two orders of magnitude smaller than the magnitude error contribution. Noticeable is that the weight magnitude variance is calculated for the worst case using logarithmic weighting units, i.e. at minimum attenuation. When the weight magnitude is smaller the steps are smaller, due to the logarithmic nature of the weights. A simple calculation yields that at a weight magnitude of -16dB, the phase error variance and the magnitude error variance are equal.

5.3.4 Comparison of measurements and theory

Some modifications to the theory presented in Section 5.2.2 must be made to be able to compare with measurements. This is due to the procedure used when measuring the output CIR. Measurements were performed in a laboratory by replacing the front end by an 8×8 Butler matrix as previously was described in Section 4.4. Two signal generators were connected to the Butler matrix to emulate signals impinging ideally from -61° and -7.2°. In practice, the CINR in equation (5.17) cannot be measured, due to the inability to separate the desired signal from the interfering and noise signal. Instead, a modified carrier to interference ratio on the adaptive antenna output was measured, denoted CIR_{out}. In fact, the carrier plus noise to interferer plus noise ratio, where "noise" includes the thermal noise, calibration error noise and the weight quantization noise was the measured quantity.

The measurement procedure were as follows; the levels of the two signals where set and the adaption of the beamformer weights was started. To measure CIR_{out} for an arbitrary weight realization, the adaption was stopped at an arbitrary time instant. When the adaption was stopped, the weight update also stopped and

the CIR in the beamformer output could be measured. This was made in a two step procedure where the interferer power and the desired signal power was measured separately. The signal generator emulating the desired signal was turned off and the ABF interferer plus noise output power was measured using a spectrum analyzer. Then the desired signal's generator was turned on and the interfering signal's generator was turned off and the ABF desired signal plus noise output power was measured. By assuming that the system is linear and the superposition principle holds, the measured CIR_{out} can be expressed by the use of the theory in Section 5.2.2 as

$$\text{CIR}_{out} \leq \frac{\mathbf{w}^*\mathbf{R}_d\mathbf{w} + n_r\sigma_n^2 w_{max}^2 + (\sigma_w^2 + n_r d_{max}^2 w_{max}^2)(\sigma_d^2|\mathbf{a}(\theta_d)|^2 + n_r\sigma_n^2)}{\mathbf{w}^*\mathbf{R}_i\mathbf{w} + n_r\sigma_n^2 w_{max}^2 + (\sigma_w^2 + n_r d_{max}^2 w_{max}^2)(\sigma_i^2|\mathbf{a}(\theta_i)|^2 + n_r\sigma_n^2)} . \tag{5.27}$$

The second and third term in the numerator of equation (5.27) makes the measured CIR_{out} estimate of the true beamformer CIR_{out} too optimistic. Therefore the expression (5.27) is larger or equal to the actual CIR_{out}. The term $\mathbf{w}^*\mathbf{R}_i\mathbf{w}$ is the interferer signal power that leaks through the correct beamformer weights, and is usually very small, if the DOA separation is large enough, which is the case in our measurements. It will therefore be neglected in the following. The term $\mathbf{w}^*\mathbf{R}_d\mathbf{w}$ expresses the received power from the desired signal. If the DOA separation is large enough, then it can be approximated as a constant κ multiplied with the desired signal power, σ_d^2, where κ is the array gain, including amplifiers in the front end and in the weighting units, degraded by the spatial and temporal correlation of the two signals. By making these simplifications, equation (5.27) can be rewritten as

$$\text{CIR}_{out} \leq \frac{\kappa\sigma_d^2 + n_r\sigma_n^2 w_{max}^2 + (\sigma_w^2 + n_r d_{max}^2 w_{max}^2)(\sigma_d^2|\mathbf{a}(\theta_d)|^2 + n_r\sigma_n^2)}{n_r\sigma_n^2 w_{max}^2 + (\sigma_w^2 + n_r d_{max}^2 w_{max}^2)(\sigma_i^2|\mathbf{a}(\theta_i)|^2 + n_r\sigma_n^2)} . \tag{5.28}$$

Clearly, if the desired signal power σ_d^2 is large compared to the interfering signal power σ_i^2 and noise power σ_n^2, then the CIR_{out} in equation (5.28) will differ from the CINR_{out}, due to the large term in the numerator. If the interferer to carrier ratio is high, say 20 dB, then equation (5.28) will resemble equation (5.17) well. Conclusively, the theory and the measurements are expected to correspond well for high interference to noise power ratios.

5.4 Simulation and measurement results

5.4.1 Simulation setup

The simulation model was intended to imitate the testbed as much as possible. The simulation model is shown in Figure 5.1. Two Gaussian minimum shift keying (GMSK) modulated signals were generated with 8 samples per symbol. The Gaussian filter had an impulse response length of 6 symbols and the normalized bandwidth was 0.3. As training sequences, the training sequences 0 and 4 given by the GSM-1800 standard were used. The received signal was quantized and the calculated weights were used with finite accuracy. Furthermore, calibration errors were introduced, to emulate the testbed.

The ADCs used the sampling frequency of 270 kHz, the same as in the adaptive antenna testbed. The covariance matrix (5.19) and the cross-correlation vector (5.20) were estimated using the $N_S = 26$ complex samples. The direction of arrivals (DOA) for the two signals were equal to the DOA:s in the measurements, described in Section 5.3. The array response vectors were slightly distorted to emulate the 0.8 dB magnitude errors and 8.5° phase errors in the Butler matrix and to make the spatial correlation nonzero.

The weight vector was calculated and normalized using (5.21) and the calculated weight vector was quantized to the desired accuracy in magnitude and phase and applied to the signals at the weighting units after multiplication with the matrix $(\hat{\mathbf{D}}^{-1})^*$ as described by equation (5.5).

5.4.2 Validation of simulation model

After adaption, the procedure to measure CIR_{out} as described in Section 5.3.4 for the measurements was used to measure the CIR_{out} in the simulations, to make the comparisons fair. A GMSK modulated PRBS signal using $L_S = 200$ symbols was used to estimate the CIR_{out}. To verify the theoretical model and the simulation results, a comparison is made in Figure 5.3. The simulation parameters are set equal to the parameters used in the adaptive antenna testbed, e.g. 8 bit ADCs and 1° and 1 dB weight accuracy. Figure 5.3 shows the measured output CIR, denoted CIR_{out} as a function of interfering signal power, when the desired signal power was held constant at two different levels, -44 dBm and -54 dBm respectively. The figure also shows the corresponding simulation results and the theoretical curve, described by equation (5.28). The measured curves constitute mean values over ten measurements and the simulated are averaged over 100 simulations. The standard deviation is 7 dB and 0.8 dB in measurements and simulations respectively. The theoretical expression does not consider the limited dynamic range of the ADC, so

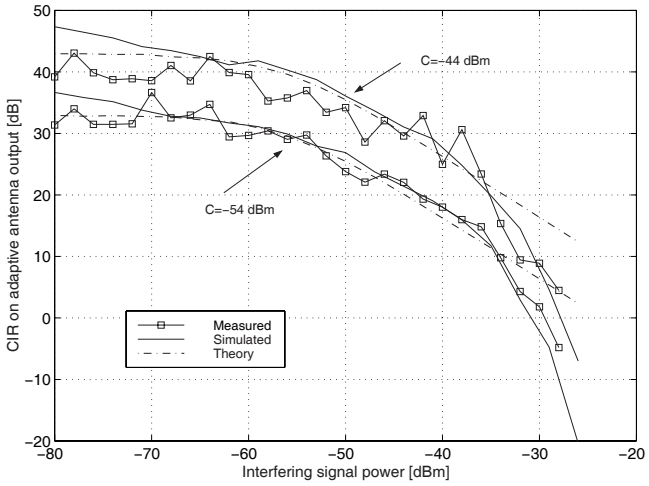


Figure 5.3: CIR_{out} as a function of interferer power. Comparisons between measurements, simulation and theory (equation (5.28)*). Carrier power constant at -44 dBm and -54 dBm respectively. Noise level at -76 dBm.*

the theoretical CIR_{out} is larger than the measured and simulated CIR_{out} when the interfering signal saturates the ADC.

The theory, equation (5.28), predicts the CIR_{out} to reach a constant level when the interferer is decreased below the thermal noise level, i.e. when the dominating term in the denominator of equation (5.28) is the noise term $n_r\sigma_n^2 w_{max}^2$. This is verified by the measurements and the level is determined by the desired signal power and the weight and quantization error variances. The simulated curve does not fit into this level at the low interference situation. An attempt to explain this is that the noise level was not correctly set in the simulations, so the noise term $n_r\sigma_n^2 w_{max}^2$ does not dominate over the interference power σ_i^2 even for the lowest interference power.

Noticeable is however the agreement of theory and measurements for high carrier to noise ratios. It was in this region were the CIR_{out} approximation, equation (5.28), was less accurate.

5.4.3 Number of ADC bits

By increasing the number of bits used in the ADC, the dynamic range of the SRX is increased, as shown in the simulation results of Figure 5.4. Here the same calibration errors and weight errors as in the testbed is used. With a larger dynamic range,

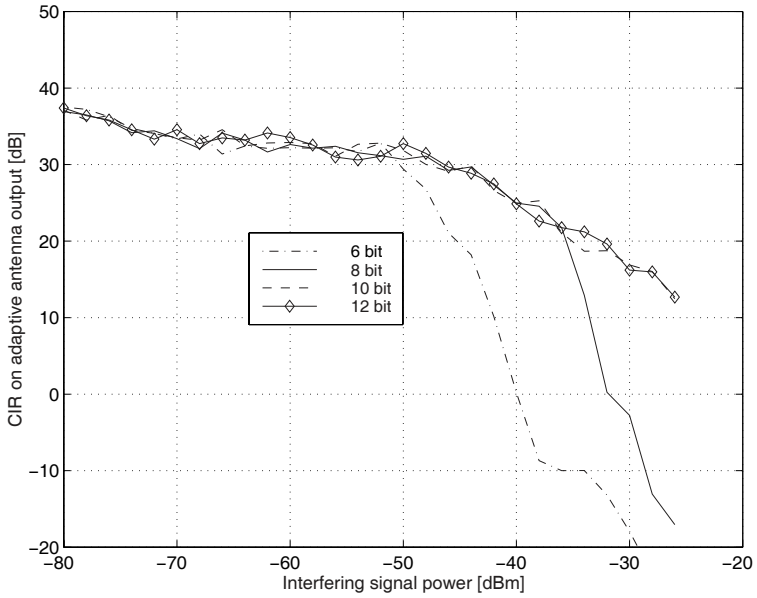


Figure 5.4: Simulated CIR$_{out}$ [dB] when the number of ADC bits is varied as a function of CIR. Carrier constant at CNR=42 dB and interferer power varied.

the SRX can receive a signal with higher power without saturating. Increasing the number of bits from 10 to 12 will not improve the output CIR in this scenario, because the SRX is not saturated for these input levels. The limiting factor in the 10 and 12 bit case is the weight and calibration errors.

5.4.4 Weight accuracy

To solely investigate the impact of weight accuracy on the performance of the adaptive antenna array, a perfect calibration is assumed, i.e. $\hat{\mathbf{D}}^{-1}\mathbf{D} = \mathbf{I}$. Also, an 8 bit ADC was used to make comparisons with the adaptive antenna testbed possible. The CIR$_{out}$ was measured for different settings of the range of the relative weight errors ϵ_A and ϵ_θ.

The CIR$_{in}$ that a single virtual omnidirectional antenna would measure was set to -15 dB and the carrier to noise ratio (CNR) was 42 dB. The CIR$_{out}$ on the adaptive antenna output was estimated and the results are presented in Figure 5.5.

Figure 5.5 shows that a 30 dB CIR$_{out}$ can be achieved if the phase and magnitude quantization steps are less than 1° and 0.3 dB. Thus, the maximal CIR improvement in this scenario is 45 dB and decreasing the weight quantisation steps further will not improve the antenna performance for this particular DOA and signal levels. To achieve a larger improvement, more antennas can be used, which will

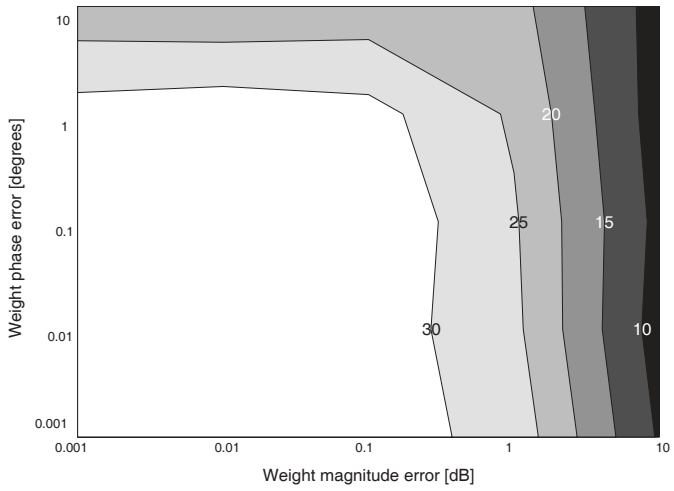


Figure 5.5: Simulated CIR_{out} [dB] as a function of quantization of phase and magnitude of weights, CIR_{in}=-15 dB, CNR=42 dB. The shades refer to different CIR_{out} levels with white and black being the largest and smallest respectively.

make the spatial correlation smaller and the ability to suppress interferers larger.

5.4.5 Calibration errors

Finally, it is investigated how the calibration accuracy affects the CIR_{out}, see the results in Figure 5.6. This figure displays CIR_{out} as a function of the weight error variance, for four different calibration accuracies. A weight error variance less than 10^{-5} will not further improve the CIR because the calibration errors limit the maximum achievable CIR[2]. To compare with the testbed, with a weight variance of $1.25{\cdot}10^{-3}$ the maximum achievable CIR_{out} in Figure 5.6 is approximately 24 dB if the calibration is performed without errors. This should be compared to the measured 18 dB from the testbed in the same conditions. Thus, to improve the testbed performance, effort should be put to improve the calibration algorithm, in favor for improving the weight accuracy. Improving weight accuracy will raise CIR_{out} only a few dB.

When the weight errors are negligible, an amplitude calibration error of 1% gives a CIR_{out} degradation of 7-8 dB from the ideal case, with no errors. This is a

[2]This weight error variance could for example be achieved with 1° phase step size and 0.1 dB weight amplitude step size.

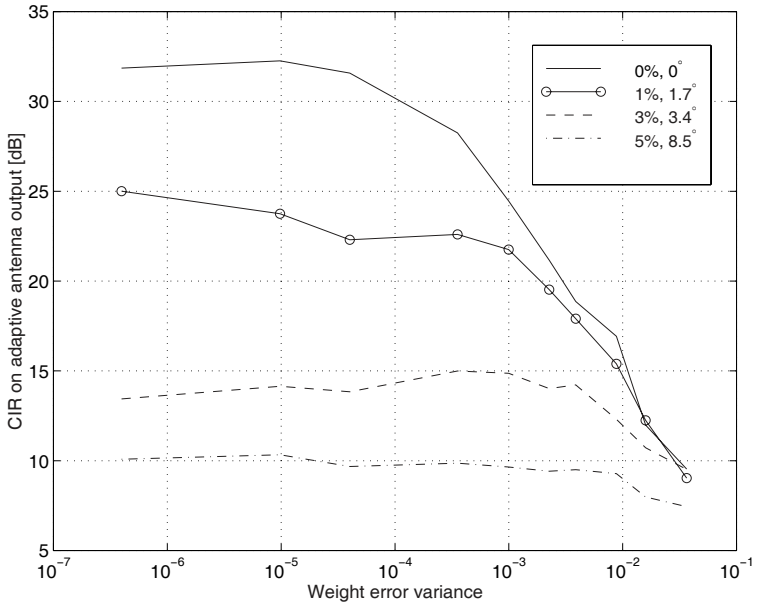


Figure 5.6: CIR_{out} as a function of the weight quantization error for different calibration errors, CIR_{in}=-15 dB, CNR=42 dB.

severe degradation and the conclusion is that the calibration must have a high accuracy and also be performed frequently to maintain the adaptive antenna's interferer suppression performance over a long time of operation.

The effect of calibration errors when the CIR_{in} is varied by varying the interference to noise ratio (INR) is presented in Figure 5.7. The carrier to noise ratio (CNR) was held constant at 22dB. A comparison with the adaptive antenna testbed was made, and the CIR_{out} was measured for different calibration errors and different CIR_{in}. The weight quantization steps were set to 1 dB and 1°. The measured curve fits the curve with calibration error of 1% in relative magnitude and 1° in phase. Furthermore, when INR $<$ 22 dB, i.e. when CIR$>$ $0dB$, the calibration errors have an negligible effect on the antenna performance. When the interferer power gets large, the output noise, described in equation (5.13) and (5.14), is increasing and the CIR decreases. When the total input power saturates the ADC, the CIR decreases abruptly and the antenna cannot maintain a CIR above 0 dB on the output. Note that the RF analog signal on the output from the adaptive antenna is not saturated. But the signal used in the DSP to calculate the weights is saturated and the CIR_{out} drops.

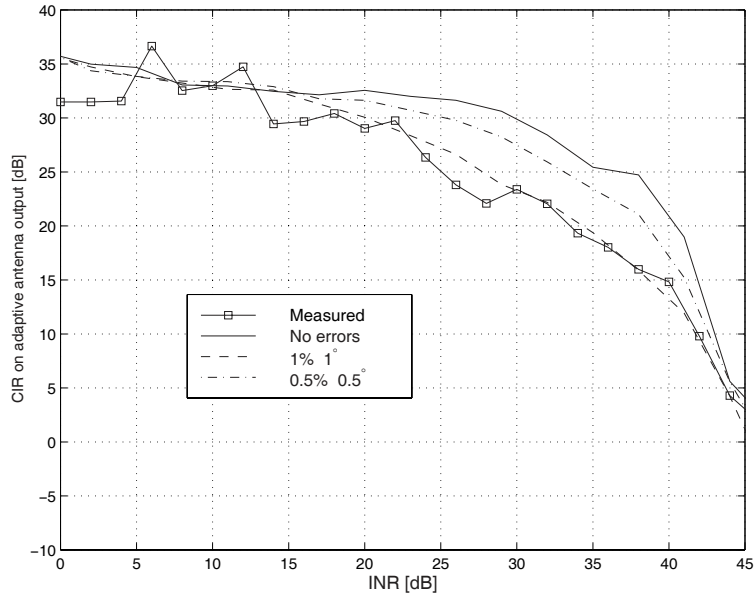


Figure 5.7: CIR_{out} as a function of interferer to noise ratio (INR) for different calibration errors. CNR=22 dB

5.5 Two Auto-Calibration Algorithms

In this section, an important issue in multiple antenna technology is addressed, namely the calibration of the antenna array. The results from the previous sections showed that the performance degradation of the adaptive antenna can be severe if the calibration is inaccurate. Furthermore, it is desirable to calibrate the antenna array during normal operation, to avoid having the basestation taken out of service. Two algorithms are proposed in this section, that calibrates an array antenna, on-line, during normal operation. Calibration of antenna arrays for mobile communication systems has earlier been presented in [12] where the sensitivity of a digital beamforming system to calibration errors were studied and a method to calibrate the array prior to operation was presented. In [10] an auto-calibration algorithm for the transmitter part of a digital beamformer was presented.

In this section, the uplink, or receiving part of the array antenna in a TDMA system, as for example GSM, is considered. It is also assumed that an analog beamformer (ABF) is used, where the beamformer weights are calculated in a digital signal processor (DSP), but the weighting is performed by hardware weighting units on the RF signals.

As discussed in Section 5.2.1, when using the ABF, it is important to know the transfer function between the input sampling receivers and the point in the signal path where the weights are applied to the signals, to be able to compensate for this difference. Prior to start-up, this transfer function is measured using an off-line calibration algorithm. However, there are active components in the weighting units and the receivers, which are sensitive to temperature variations and which characteristic also will change due to aging. This will soon make the off-line calibration data invalid.

In Figure 5.8, measurements of the temperature drift in amplitude and phase for the hardware weighting unit used in an ABF testbed described in this thesis is presented. The measurements were performed over a period of ten hours after a cold start-up. It can be seen that the drift is approximately 0.1 dB and 1° per hour of operation and it demonstrates the need for frequent re-calibrations according to the theory in the previous section. It is therefore desirable if the calibration can be performed simultaneous with normal antenna operation.

Two auto-calibration, or *on-line* calibration algorithms are presented in this section. They are transparent to normal antenna operation and have low complexity. The first algorithm is derived from the least mean square (LMS) algorithm and is a non-parametric solution. The second algorithm is derived using a parametric approach, where the transfer function is identified and the temperature drift is tracked. To study the performance of the proposed algorithms, a simulation was performed, modelling a simple signal environment with two signal sources

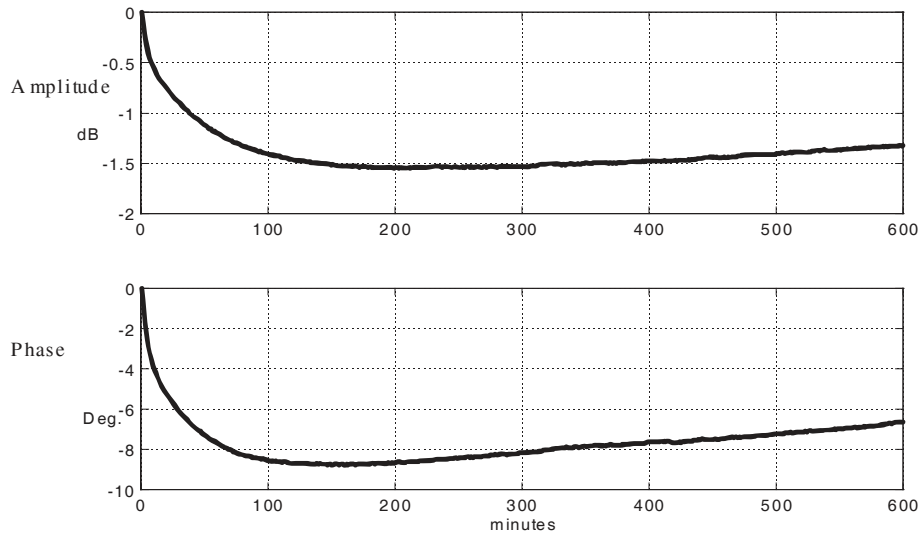


Figure 5.8: Temperature drift of hardware weight. The drift is 1° and 1 dB per hour of operation.

(mobiles) and no multipath propagation. The time variations in the hardware was simulated by using a parameter drift similar to the measured one. Both algorithms succeed to maintain the output signal to interference ratio (SIR) at the same level as with perfect knowledge of the transfer functions. The LMS-like algorithm requires a slowly or non-varying signal environment, while the second algorithm actually benefits from a rapidly varying signal environment, which is characteristic for fading in a multipath scenario.

5.5.1 Problem formulation

The block diagram of the adaptive antenna is shown in Figure 5.1, where the feedback channel signal $\hat{u}(k)$ now will be utilized. The signals arriving at the n_r antennas are described by the $n_r \times 1$ column vector $\mathbf{y}_a(t)$. The noise generation in the front end is modelled as a zero mean white Gaussian noise vector, $\mathbf{n}_T(t)$, with covariance matrix $\sigma_n^2\mathbf{I}$. The noise in the input down-converting and sampling receivers, $\mathbf{n}_r(t)$ is also modelled as a zero mean white Gaussian noise vector, but with covariance matrix $\sigma_r^2\mathbf{I}$. The signals from the antennas are split into a digital path and an analog path. The analog path consists of cables and the hardware weighting units. The digital part consists of digital signal processing for weight computation and calibration.

Assuming all weights are set to unity, the transfer function measured between the input sampling receiver at point A in Figure 5.1 and the beamformer weighting

units, point B, is described by the complex diagonal matrix $\mathbf{D} = \text{diag}\,[d_1, \ldots, d_{n_r}]$. The signals at the beamformer is denoted $\mathbf{y}''(t)$ and is defined as $\mathbf{y}''(t) = \mathbf{D}\mathbf{y}(t)$. Each factor d_l describes the attenuation and the phase shift of the analog path l relative to the corresponding digital path, except for the noise $\mathbf{n}_r(t)$.

The factors d_l will in general be nonlinear functions of the calculated weights, $\mathbf{w}$, due to coupling between the phase shifters and the attenuators. The matrix $\mathbf{D}$ will thus in general be a nonlinear function of $\mathbf{w}$, i.e. $\mathbf{D} = \mathbf{D}(\mathbf{w})$. The elements in $\mathbf{D}$ are assumed to be constant over the receiver passband, but they will change during operation due to temperature drift in the weighting units, as seen experimentally in Figure 5.8.

The output from the beamformer is down-converted to the baseband and sampled giving $\hat{u}(k) = u(k) + n_f(k)$ where $u(k)$ denote the time continuous signal sampled at the time instant $t = t_k$. The term $n_f(t)$ is the noise added by the feedback receiver, assumed to be zero mean and complex Gaussian distributed with variance σ_f^2. The signal $\hat{u}(k)$ will be used in the calibration algorithms, to be presented below.

The example antenna presented here uses the Sample Matrix Inversion (SMI)-algorithm to calculate the adaptive antenna weights, see Section 5.3. The mean square optimal receive weight vector is given by (5.21)

$$\hat{\mathbf{w}}_0 = \hat{\mathbf{R}}_{\mathbf{yy}}^{-1}\hat{\mathbf{r}}_{\mathbf{y}d} \tag{5.29}$$

where the sample-mean covariance matrix and cross-correlation vector are given by (5.19) and (5.20).

Calibration

Compensation of phase and magnitude differences between the digital path and analog path must be performed for each antenna element, since the adaptive antenna weights are calculated based on the sampled signals of the digital paths, $\mathbf{y}'(k)$, whereas the beamforming is performed on the RF-signals of the analog paths $\mathbf{y}''(t)$.

Prior to operation, the matrix $\mathbf{D}$ is computed by performing a calibration. This calibration step is called an off-line calibration and we denote the corresponding matrix as $\mathbf{D}_0$. A continuous wave (CW) signal at the carrier frequency is injected at one antenna element at a time by directional couplers. The attenuation and the relative phase of that specific analog path compared to the digital path can be measured by using the signal $\hat{u}(k)$. The calibration process is then repeated for all n_r antenna branches.

If the magnitude of a weight is adjusted, then the weight phase will also change, and vice versa, due to non-ideal isolation between the phase shifter and the attenua-

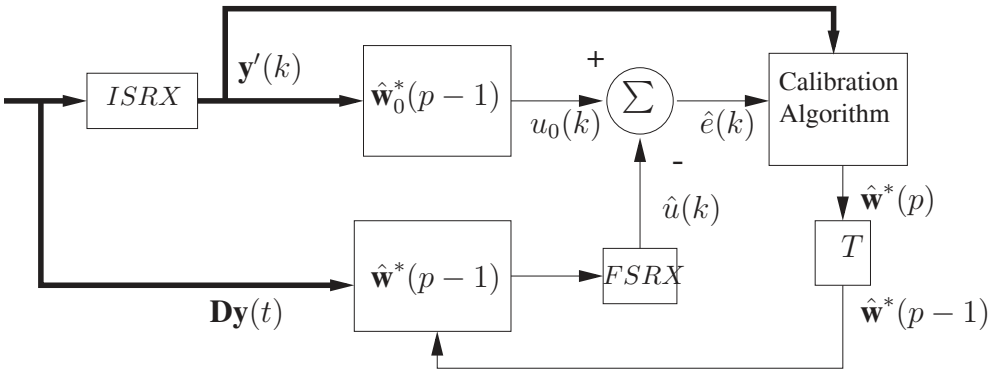


Figure 5.9: Block diagram of the direct approach. $FSRX$=Feedback Sampling Receiver, $ISRX$=Input Sampling Receiver. T represents one frame delay

tor in the hardware weights. The relation between the desired and the actual weight used for weight control is stored in a look-up table with all possible weight settings as entries. Since full access to the control of the weighting units is required in the calibration, the off-line calibration can only be performed prior to operation of the adaptive antenna.

The requirement on the calibration algorithm is a low complexity and yet accurate enough to maintain a good performance of the adaptive antenna. Furthermore, the calibration should not interfere with the main operation of the antenna.

5.5.2 The algorithms

The direct approach

This section outlines the auto-calibration algorithm using an LMS-approach and presents the assumptions made concerning the off-line calibration and the thermal drift of the involved hardware.

A signal-flow graph of the proposed algorithm is shown in Figure 5.9. The signals at the weighting units, $\mathbf{y}''(t)$, are multiplied with the weight-vector $\hat{\mathbf{w}}^*(p-1)$, based on the calculations in the previous frame $p-1$, where p is the frame index, i.e. $u(t) = \hat{\mathbf{w}}^*(p-1)\mathbf{y}''(t)$. Since adaptive arrays utilizing analog beamforming is studied, the weights are assumed to be applied to the data of the next TDMA frame. This is necessary since otherwise the weights have to be calculated and steered out during a time period that is much shorter than a frame. This also implies that the time step in the algorithm is one TDMA frame.

The sampled beamformer output signal, $\hat{u}(k)$, is compared with the corresponding beamformer output signal in the DSP, $u_0(k)$, and the error signal $e(k) =$

$u_0(k) - \hat{u}(k)$ is formed. The signal $u_0(k)$ is calculated by using the SMI-weights from the previous frame as $u_0(k) = \hat{\mathbf{w}}_0^*(p-1)\mathbf{y}'(k)$ However, only one sample k is used and it can be taken arbitrarily from the whole frame p.

For calculation of the necessary adjustments an LMS-like (least mean square) algorithm is proposed. The aim of the algorithm is to minimize the mean squared magnitude of the error signal $e(k)$.

Assume that the signal vector $\mathbf{y}''(t_k)$ relates to $\mathbf{y}'(k)$ according to

$$\mathbf{y}''(t_k) = \mathbf{D}\mathbf{y}(t_k) = \mathbf{D}\left\{\mathbf{y}'(k) - \mathbf{n}_r(k)\right\} \quad . \tag{5.30}$$

The relationship between $\mathbf{D}\mathbf{y}(t_k)$ and $\mathbf{y}'(k)$ in (5.30) is an approximation, since $\mathbf{D} = \mathbf{D}(\mathbf{w})$ and $\mathbf{w}$ in turn depends on $\mathbf{y}'(k)$. It is thus assumed here that the matrix $\mathbf{D}$ is independent of the weights. It is possible to accomplish this by designing the hardware weights with high isolation between the phase shifter and the attenuator.

To adjust the weights to compensate for the temperature drift, the well known LMS-approach would use $\mathbf{y}''(t_k)$ and $e(k)$ to update the weight vector as:

Desired algorithm: *(p≥1)*

$$\begin{aligned} u_0(k) &= \hat{\mathbf{w}}_0^*(p-1)\mathbf{y}'(k) \\ e(k) &= u_0(k) - u(t_k) \\ \hat{\mathbf{w}}(p) &= \hat{\mathbf{w}}(p-1) + \mu\mathbf{y}''(t_k)e^*(k) \end{aligned} \tag{5.31}$$

where the initialization of the algorithm has been omitted.

The constant μ is the step-size parameter in the algorithm and $u(t)$ is the output of the analog beamformer. The algorithm (5.31) cannot be used since $\mathbf{y}''(t_k)$ is not measurable and only a noisy estimate of $u(t_k)$ is available. The approach used here is therefore to use $\mathbf{y}'(k)$, i.e. the noisy measurement of $\mathbf{y}(k)$, and the matrix $\mathbf{D}_0$ obtained from the off-line calibration process to estimate $\mathbf{y}''(t_k)$, using (5.30). The proposed algorithm can then be stated as follows:

Proposed algorithm:

Initialization: (p=0)

$$\hat{\mathbf{w}}(0) = (\mathbf{D}_0^{-1})^*\hat{\mathbf{w}}_0(0) \tag{5.32}$$

Algorithm: (p≥1)

$$\begin{aligned} u_0(k) &= \hat{\mathbf{w}}_0^*(p-1)\mathbf{y}'(k) \\ \hat{e}(k) &= u_0(k) - \hat{u}(k) \\ \mathbf{y}''(k) &= \mathbf{D}_0\mathbf{y}'(k) \\ \hat{\mathbf{w}}(p) &= \hat{\mathbf{w}}(p-1) + \mu\mathbf{y}''(k)\hat{e}^*(k) \quad . \end{aligned} \tag{5.33}$$

The fact that the calibration data from the off-line calibration is used will only slightly affect the performance of the gradient method since the LMS-approximation of the gradient is in itself very crude. The correct instantaneous gradient direction is given by $\mathbf{y}''(t_k)e^*(k)$ but here the approximation $\mathbf{D}_0\mathbf{y}'(k)\hat{e}^*(k)$ is used instead. An exact analysis of the influence of the properties of this error in gradient estimate on the convergence of the algorithm remains to be investigated. Simulation studies presented in Section 5.5.3 show that the performance of the proposed algorithm will be satisfactory in a relatively stationary signal environment, i.e. with a slowly varying SMI-weight vector. When the signal environment is non stationary, the long convergence time inherent in LMS-like algorithms can be expected to create problems. The error term $\hat{e}(k)$ will be large if the SMI-weight vector $\mathbf{w}_0$ is not changing slowly. The performance will then be degraded.

The indirect approach

An alternative on-line calibration method is to track the temperature drift of the transfer functions d_l, $l = 1, \ldots, n_r$, to form the estimated row-vector $\hat{\mathbf{d}}(p) = \left[\hat{d}_1(p)\hat{d}_2(p)\cdots\hat{d}_n(p)\right]$. The idea is to consider the output signal $u(t_k)$ as a linear combination of the signals $\mathbf{z}(t_k)$, after the weights, $\hat{\mathbf{w}}$, as in Figure 5.10. Note that the order of the weights and the transfer function coefficients d_l has been switched. However, only $\mathbf{y}'(k)$ and $\hat{u}(k)$ can be measured, and not $\mathbf{z}(t_k)$ and $u(t_k)$. By using the known weights and the identified transfer function vector from the previous frame, $\hat{\mathbf{d}}(p-1)$, it is possible to estimate $\mathbf{z}(t_k)$. The estimate is denoted by $\mathbf{z}(k)$ and element l in the column vector $\mathbf{z}(k)$ is defined as:

$$\begin{aligned} z_l(k) =& y'_l(k)\hat{w}^*_l(p-1) \\ =& y'_l(k)\frac{\hat{w}^*_{0,l}(p-1)}{\hat{d}_l(p-1)} \quad . \end{aligned} \tag{5.34}$$

This regressor vector is now used in the identification of $\mathbf{d}$ by forming a least-squares identification problem utilizing the sampled output $\hat{u}(k)$. This approach is possible since the system is assumed linear. The output $u(t_k)$ is a linear combination of the regressor vector $\mathbf{z}(t_k)$:

$$u(t_k) = \mathbf{d}(p)\mathbf{z}(t_k) \quad . \tag{5.35}$$

By multiplying with $\mathbf{z}^*(t_k)$ and applying the expectation operator

$$\mathbf{r}_{u\mathbf{z}} = \mathbf{d}(p)\mathbf{R}_{\mathbf{zz}} \quad . \tag{5.36}$$

Now, estimate the least-squares estimated temperature drift in frame p by using

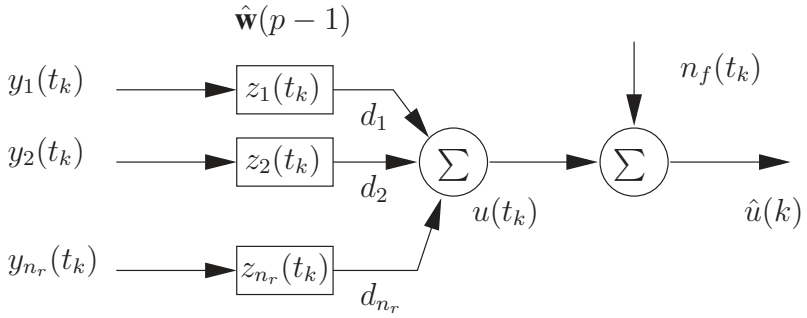


Figure 5.10: Identification model

the estimated covariance and cross-correlation matrices

$$\begin{aligned} \hat{\mathbf{R}}_{\mathbf{zz}} &= \frac{1}{N_S}\sum_{k=1}^{N_S}\mathbf{z}(k)\mathbf{z}^*(k) \\ \hat{\mathbf{r}}_{\hat{u}\mathbf{z}} &= \frac{1}{N_S}\sum_{k=1}^{N_S}\hat{u}(k)\mathbf{z}^*(k) \end{aligned} \tag{5.37}$$

and calculate the diagonal elements of $\hat{\mathbf{D}}(p)$ as

$$\hat{\mathbf{d}}(p) = \hat{\mathbf{r}}_{\hat{u}\mathbf{z}}\hat{\mathbf{R}}_{\mathbf{zz}}^{-1} \quad . \tag{5.38}$$

Note that the samples used in (5.37) are not necessarily from the same frame p. The performance will be improved if samples from several frames are used, as discussed below.

The elements in $\hat{\mathbf{d}}(p)$ are low-pass filtered in order to introduce a memory in the algorithm. A first order filter with a pole at 0.9 is used. This will reduce the variance of the regressor vector and therefore improve the tracking of $\mathbf{d}$. Experiences from the simulations show that this improves the tracking ability. When using this indirect approach instead of the direct method described in Section 5.5.2, samples from the whole frame is used in the identification process, instead of the utilization of only one sample as in the LMS-approach. Furthermore, mobile communication systems are interference limited and not noise limited. Thus the noise level in $\mathbf{y}(t)$ (and also in $\mathbf{z}$(k)) will be low compared to the signal levels. Therefore $\hat{\mathbf{R}}_{\mathbf{zz}}$ will be ill conditioned if the number of incoherent rays impinging on the array is less than n_r. The calculation of the inverse $\hat{\mathbf{R}}_{\mathbf{zz}}^{-1}$ will create problems when computing $\hat{\mathbf{d}}(p)$ in (5.38) and the tracking of the temperature drift may not succeed. This can however be mitigated by using a few number of samples from several frames to estimate the covariance matrices. This will in general make $\hat{\mathbf{R}}_{\mathbf{zz}}$ better conditioned. The idea is

that the fading will cause the signal components in $\mathbf{z}(k)$ to vary over the frames, creating a sequence that is persistently exciting with a well conditioned covariance matrix $\hat{\mathbf{R}}_{\mathbf{zz}}$. This approach will thus benefit from a rapidly time-varying signal environment and as opposed to the LMS-approach presented in Section 5.5.2, which requires a stationary or slowly time-varying signal environment. The reason for this is that only $\mathbf{D}$ is tracked which varies more slowly than the weight vector $\mathbf{w}$.

5.5.3 Simulation study

The aim of the simulation is to study the performance of the two auto-calibration algorithms presented above. A simple signal environment with two mobiles transmitting binary phase shift keying (BPSK) modulated signals of equal power but with different training sequences was used. The two signals were transmitting on the same frequency and the wavelength was λ_c. The desired and the interfering signals are assumed to impinge on an eight element uniform linear array (ULA) from distinct directions θ_d and θ_i respectively. No multipath propagation or fading is assumed. The antenna array inter-element spacing was $\lambda_c/2$. The length of the training sequence was 26 symbols to comply with the GSM-1800 standard.

The received signal can be expressed as

$$\mathbf{y}_a(t) = \mathbf{a}(\theta_d)s_d(t) + \mathbf{a}(\theta_i)s_i(t) \tag{5.39}$$

where $\mathbf{a}(\theta_d)$ and $\mathbf{a}(\theta_i)$ are the array response vectors in direction θ_d and θ_i and $s_d(t), s_i(t)$ are the desired and interfering signals respectively and the noise is neglected.

In the simulations the temperature drift in magnitude and phase was generated as independent integrated random-walk processes to obtain a "smooth" drift. The elements of the diagonal matrix $\mathbf{D}$ can be written as

$$d_l = A_l e^{j\varphi_l} \qquad l = 1, \ldots, n_r \tag{5.40}$$

where A_l and φ_l are generated as integrated random walk processes:

$$\begin{aligned} A_l(p+1) &= A_l(p) + \frac{1}{1-q^{-1}} v_{A,l}(p) \\ \varphi_l(p+1) &= \varphi_l(p) + \frac{1}{1-q^{-1}} v_{\varphi,l}(p) \end{aligned} \tag{5.41}$$

with $v_{A,l}$(p) and $v_{\varphi,l}$(p) being white noise sequences of appropriate variance. A_l and φ_l are initialized by the values from the off-calibration. This gives a temperature drift with statistical properties similar to the measurements presented in Figure 5.8.

To reduce the conditional number (or eigenvalue spread) of the covariance matrix $\hat{\mathbf{R}}_{\mathbf{yy}}$, regularization is employed to obtain a condition number approximately equal to 10^2. More details of this regularization method can be found in [146]. The covariance matrix can therefore be inverted without numerical problems, and the weight vector calculated by the SMI-algorithm will be approximately the same from frame to frame, under steady state conditions. This will support the one frame delay in the weight settings. Since the LMS-algorithm proposed in Section 5.5.2 is a closed loop algorithm, it is not possible to update the weights more frequently than once per frame; the "response" of the old weights is used to calculate the new weights. This means that the time step of the algorithm is one TDMA frame (or 4.615 ms in GSM/DCS-1800).

The noise variances of the integrated random walk processes (5.41), that model the temperature drift of the magnitude and the phase, sets the time scale of the simulation and can be increased to reduce the simulation time. It is therefore not necessary to simulate more than 500 frames corresponding to 2.3 seconds in GSM-1800.

To measure the performance of the adaptive antenna, the SIR on the beamformer output was estimated as

$$\widehat{SIR} = \frac{E(|\hat{\mathbf{w}}^*\mathbf{a}(\theta_d)s_d(t)|^2)}{E(|\hat{\mathbf{w}}^*\mathbf{a}(\theta_i)s_i(t)|^2)} = \frac{|\hat{\mathbf{w}}^*\mathbf{a}(\theta_d)|^2}{|\hat{\mathbf{w}}^*\mathbf{a}(\theta_i)|^2} \quad . \tag{5.42}$$

Here, the noise levels are assumed to be well below the signal levels.

The SIR of the adaptive array utilizing the auto-calibration algorithm presented in Section 5.5.2 is plotted as a function of frame number in Figure 5.11. In this particular simulation the angles of the two mobiles have been chosen to $\theta_d = 15^\circ$ and $\theta_i = 43^\circ$ relative to the broadside direction of the ULA. The step size μ used in the LMS-like algorithm was 0.005. The variances of the noise sources σ_n^2,σ_r^2 and σ_f^2 were chosen 20 dB, 40 dB and 30 dB below the signal level, respectively. The SIR of the adaptive array only utilizing the off-line calibration is also plotted in Figure 5.11 for comparison. It is evident that the auto-calibration algorithm is able to maintain the SIR on a high level, whereas the performance of the traditional SMI adaptive array is severely degraded as the drift in the magnitude and phase of the weights is introduced. The variations of the SIR is also lower for the LMS-method. This is due to the recursion involved in the algorithm that smoothes the SIR.

Figure 5.12 shows the SIR when using the identification approach presented in Section 5.5.2, compared to the SMI algorithm with the off-line calibration data only. All simulation parameters are identical to those in the LMS-study. Also in this case the SIR of the auto-calibration algorithm is maintained on a high level. However, the variations of the SIR is higher in this case compared to the LMS-approach, and more similar to the SMI without drift compensation. This can be

explained by the batch oriented approach in the tracking as opposed to the LMS method where the feedback signal smoothes the weight adjustments and the variation of CIR is slower. In the identification approach it is also possible to study the tracking ability of the algorithm. The tracking of the temperature drift for the fifth antenna path is presented in Figure 5.13, and it can be seen that the tracking of both magnitude and phase is successful.

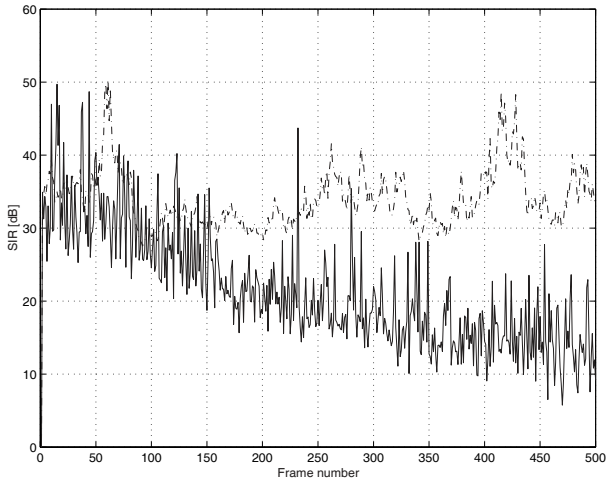


Figure 5.11: SIR as a function of frame number. Dash-dotted: auto-calibration using LMS, solid: SMI with off-line calibration data only

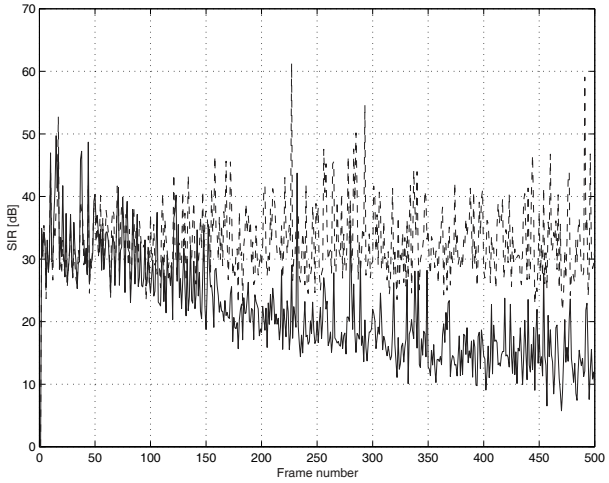


Figure 5.12: SIR as a function of frame number. Dashed: auto-calibration using the identification approach, solid: SMI with off-line calibration data only

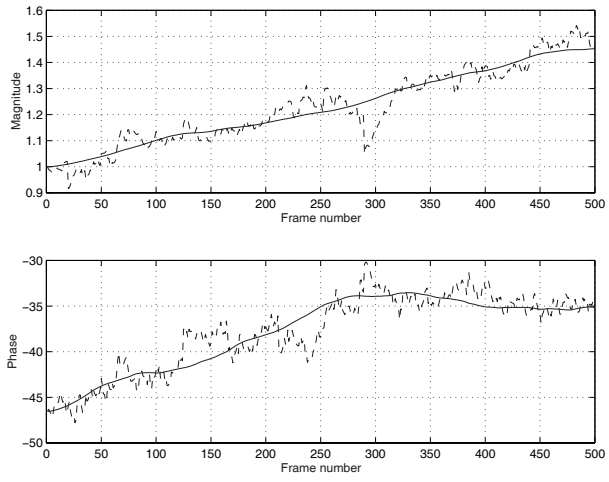


Figure 5.13: Tracking of variations in magnitude and phase for the fifth antenna path obtained from the identification method. Dashed: tracked drift, solid: true drift

5.6 Conclusions

To study the performance degradation when implementing an adaptive antenna array algorithm in hardware, using an ABF type of beamformer, an expression for the output CINR was derived. Due to the complexity of the problem, involving nonlinear saturation effects, the derivation was only partly accomplished and some simplifications were made. The study showed how the weight quantization errors and calibration errors increase the output noise power, thus decreasing the output CINR. An important observation is that the decrease in CINR is proportional to the total impinging power on the array.

The theoretical results were verified and extended using an ABF type adaptive antenna testbed with ten array elements, working in the receive mode and designed to partly follow the GSM-1800 standard. To extend the testbed to support the GSM-1800 standard, protocol issues and random access channels must be handled for call set-up and handover situations [6]. Also, synchronization of sampling instants must be automatically handled, an algorithm for this is described in [147]. The fact that the GSM-1800 standard is not fully implemented does not affect the results in this chapter as long as the received signals are within the dynamic range of the ADC:s, which for the testbed is not sufficiently large to comprise an actual GSM-1800 radio channel with fading variations and large near far ratios. The weight resolution and calibration accuracy have an impact on the interference suppression capability of the adaptive antenna which affects the link budget only. Hence, the conclusions in this chapter can be applicable to an ABF basestation antenna and especially a GSM-1800 standard system with the reservation of the number of ADC bits.

The theoretical expressions were also verified using a simulation model of the hardware channel and allowed for an extension of the analytical model. The balance between weight accuracy and calibration accuracy showed that with a certain calibration accuracy, the output CINR could not improve above a certain limit, regardless of weight accuracy. Thus, the system designer should balance these two sources of error. For the testbed, it can be concluded that improving the calibration accuracy would gain more in output CIR as compared to improving the weight accuracy, the magnitude steps in the weighting units (1 dB) were too coarse, as compared to the phase accuracy (1°). Thus the bottleneck in the testbed is identified as the coarse magnitude steps. The system designer should put effort into making the phase and magnitude error variances equal to minimize a large overhead in either the phase-shifter or attenuator accuracy. Note that it was assumed that the number of bits in the ADC:s is sufficiently large, i.e. chosen so that the dynamic range is not a limitation of the system performance.

The extension of these results to other wireless standards depends on the com-

plexity of the receivers. Using the ABF in a W-CDMA system is not feasible, even if the beamforming is carried out prior to the code correlator as in the multi-dimensional RAKE receiver [148]. Usually the channel model (5.1) is too simple to model a CDMA channel, although Naguib, Paulraj and Kailath used it in one of their early derivations of the capacity improvement in using antenna arrays in CDMA systems [149]. In the multidimensional RAKE receiver, a set of weighting units for each RAKE finger and each user is required and the number of weighting units becomes very large, which is unpractical if they are implemented in hardware. Also, using a spread spectrum standard, the system bandwidth is increased and the transfer function of the hardware channel, defined as the matrix $\mathbf{D}$, could become frequency dependent. This makes the calibration more difficult, and the complexity of the compensation algorithm increases.

Two algorithms are proposed in this chapter for mitigating the temperature drift in adaptive antenna arrays using hardware weights. The two methods utilize a feedback signal from the summed beamformer output. The first method used the initially calculated weights to create a reference signal in the DSP. An LMS-like algorithm then adjusts the hardware weights to make the feedback signal follow the reference. The second method attempts to track the drift in the transfer functions, and use this tracking information to adjust the weights calculated in the DSP. Simulations shows that the output SIR is unaffected for both methods when a realistic temperature drift generated as an integrated random walk process is introduced. The SMI-algorithm without compensation for the temperature drift however suffers a considerable performance degradation. The variance of output SIR of the LMS-approach is lower due to the smoothing effect introduced by the recursive algorithm. The LMS-approach requires a slowly time-varying signal environment. Otherwise the algorithm will lose track of the weight-vector $\mathbf{w}$. The tracking approach does not demand a slowly varying or quasi-static signal environment as the LMS-approach. It actually benefits from rapid time variations as in a multipath environment. Such variations in the signal environment will tend to make the input signal to the algorithm persistently exciting. As long as the problem of a persistently exciting input signal can be handled, it will therefore be the more attractive approach. It is however more computationally complex.

Conclusively, the results in the chapter applies to systems where ABF is possible to implement, such as in the demonstrated narrowband GSM-1800 system and in systems where the spatial beamforming and the temporal equalization are separated. The results from the theoretical model affects the link budget only, in terms of the received CINR at the BS, as long as the received signals are within the dynamic range of the receivers. The results are also independent of the choice of weight calculation algorithm.

Chapter 6

Nonlinearities in transmit amplifiers for MISO systems

NONLINEAR distortion generated by non-ideally linearized multicarrier amplifiers in the transmit chain of a basestation operating over a MISO channel, is investigated in this chapter. It is shown, both experimentally and analytically, that it is possible to reduce the effective intermodulation distortion (IMD) by a proper allocation of users to new frequencies in a cell. Finally, an analytical expression for the blocking probability in a SISO system with IMD is derived which has application in indoor WLAN systems. Although we have assumed that the receiver has a single receiver antenna, some results in this chapter are applicable to MIMO systems as well, especially the eigenmode algorithms developed in Section 2.4.

6.1 Introduction

Introducing multiple antennas at the basestation in a cellular system increases the size and complexity of the basestations due to additional radio transceivers, amplifiers, filter combiners and so forth. If a basestation is operating on M_f frequency channels, then it needs M_f transceivers and M_f amplifiers if the conventional single carrier power amplifier (SCPA) technology is applied. By introducing an n_t element array antenna, the number of SCPA transmit amplifiers increases to $n_t M_f$ which will consume more power and space at the basestation site. Thus, if it is possible, it is highly desirable to integrate the base station hardware these factors.

One proposed solution, applicable on the transmitting link, is to sum the signals prior to amplification and co-amplify them in a multicarrier amplifier (MCPA)

[12]. This eliminates the need for combining signals of high power using bulky cavity filters combiners. With this technology, every group of M_f SCPA:s in a conventional basestation site is replaced by a single MCPA. The number of amplifiers is then reduced from $n_t M_f$ to n_t, which is a considerable reduction if the number of frequency channels M_f is large. Another benefit of using MCPA:s is the possibility to move the summation of the M_f signals to the digital domain. This reduces the size of the basestation even further, but requires high dynamic range and wideband ADC:s. The trend is to move more and more functionality into the digital domain to ultimately approach the full software radio architecture (SWR) [150]. The combination of MCPA and array antennas is not new, for satellite systems where small equipment size and low power consumption is of high importance, it has been utilized for a long time [151–153].

Unfortunately, using MCPA:s has a major drawback in terms of power added efficiency (PAE). The reason is that when several constant-envelope signals are combined in an MCPA, the envelope of the composite signal becomes non-constant and have a large peak to mean ratio. Hence, to avoid amplifier saturation, a large margin must be introduced and then the amplifier cannot be operated in the highly-efficient nonlinear region since this would introduce too much intermodulation distortion (IMD). So, to improve the PAE, the margin must be reduced by moving the amplifier operating point closer to the 1-dB compression point of the amplifier, which will require linearization. MCPA linearization has been the subject of considerable research effort using for example feed-forward [154] and pre-distortion [155] techniques. These results are promising and increase the amplifier efficiency although still the PAE is limited to 15% for MCPA amplifiers as opposed to SCPA:s which have a PAE above 40% [156].

The generated IMD is a source of interference and must be kept low for proper system operation. Intermodulation products are commonly referred to as second-, third-, fourth-, ... , nth-order products depending on the nonlinearity that give rise to the products. The most general form of third-order interference occurs when channels at frequencies $\omega_i, \omega_j, \omega_k$ intermodulate in such fashion as to produce interference on a channel operating at frequency ω_l. In this case

$$\omega_i + \omega_j - \omega_k = \omega_l \ . \tag{6.1}$$

If ω_l belongs to the set of used frequencies in the system, then the generated IMD may act as a serious source of impairment. Another form of third-order interference is of the type

$$2\omega_i - \omega_j = \omega_l \ . \tag{6.2}$$

These two types of third order IMD are commonly denoted as type B and type A respectively.

Even if it is possible to construct an MCPA which meets the stringent linearity requirements in terms of IMD, its peak power requirement is much greater than the individual power of the M_f carriers. For example, a $M_f = 16$ channel basestation requiring 10 Watts per channel, would require a peak power in the MCPA as high as 2.56 kW [157]. This amplifier has to be biased with a very large back-off from the saturating point to avoid that the peak power rating of the MCPA is exceeded resulting in a low PAE. On the other hand, it has been shown that it is possible to limit the peak-to-mean ratio of the input signal (in software or hardware) without introducing too much intermodulation distortion that violates the overall system specifications. The signal is simply limited in amplitude prior to amplification and a reduction in peak power rating of the MCPA [157] is thereby allowed. The level of the amplitude limiter for a specific bit error rate is dependent on peak to mean ratio of the used modulation scheme. This was investigated in [158].

Some definitions are required for the following analysis. The back-off is defined by either the input back-off (IBO) or the output back-off (OBO), measured as

$$IBO = 10\log_{10}\left(\frac{P_{sat,in}}{P_{avg,in}}\right) \qquad \text{[dB]} \tag{6.3}$$

$$OBO = 10\log_{10}\left(\frac{P_{sat,out}}{P_{avg,out}}\right) \qquad \text{[dB]} \tag{6.4}$$

where $P_{avg,in}, P_{avg,out}$ is the average input/output power and $P_{sat,in}$ is the input power when the output power reaches saturation and $P_{sat,out}$ is the corresponding output power. A smaller OBO/IBO implies a higher amplifier PAE.

However, reducing the OBO/IBO increases the amount of generated intermodulation interference and the signal to intermodulation interference ratio of the wireless link decreases. Thus, there is a trade-off between amplifier efficiency and link quality which will be studied in this chapter for the case of a combination of nonlinear MCPA:s with multiple element antenna arrays, i.e. MISO systems. This combination gives some peculiar results which will be presented and analyzed.

Previous and related work

The radiation pattern provided by the beam-forming network will generate IMD that, in general, are radiated in directions different from the principal beam directions as was discussed in [159–163]. These IMD lobes have also been called "phantom lobes" [71, Chapt.4] as they are not generated purposely, but are created by the nonlinearity. The array antenna spatially filters the IMD and this effect can in some circumstances be used to reduce the signal to intermodulation ratio of the mobiles in the system, as was previously shown in [152] for a satellite sys-

tem. The required circumstances to achieve this advantageous property in a mobile communication system will be investigated in this chapter.

Linearity considerations for adaptive antennas in mobile communications has previously been studied by Beach *et.al.* [164] who showed that the intermodulation distortion (IMD) created in the nonlinear amplifier was shown to degrade the null depth and increase the side-lobe levels of the antenna radiation pattern. Another study by Litva and Lo [71, Chapt.4], investigates the receive radiation patterns from a digital beamforming (DBF) array under nonlinear distortion. Fixed and adaptive beamforming was compared and it was shown that adaptive beamforming in the uplink can reduce the effect of nonlinearities, by suppressing the *phantom interferers* created by the nonlinearity.

Modelling nonlinearities

Nonlinearities in communication systems are commonly characterized by the amplitude to amplitude modulation (AM/AM) distortion characteristics and amplitude to phase modulation (AM/PM) distortion characteristics, which is the input amplitude dependent gain and phase conversion of the amplifier.

When simulating multicarrier communication systems operating over nonlinear channels, the required sampling rate for a correct representation of the IMD in the output signal contributes significantly to long execution times. If the bandwidth of the input signal is W, then the sampling theorem says that to accurately reproduce a continuous signal faithfully, the sampling frequency must be at least twice the bandwidth, i.e. $f_s \geq 2W$ [165] to avoid aliasing distortion. However, the output of a nonlinear device contains energy at frequencies that does not have energy in the input signal. For a power series nonlinearity

$$y(t) = \sum_{k=0}^{N_p} a_k c^k(t) \tag{6.5}$$

it is easy to show that the k:th component is bandlimited to kW since the Fourier transform of $c^k(t)$ is the k-fold convolution of $C(f)$, where $C(f)$ is the frequency spectrum of $c(t)$. So the sampling rate of the simulation of this system must be $f_s \geq 2N_pW$, which can be prohibitively high and lead to very long simulation times. There are ways to reduce the simulation time by performing some approximations. This is possible since the magnitude spectrum of $y(t)$ is not uniformly distributed over the bandwidth $2N_pW$, rather decreasing toward zero at the band edges. So by choosing a sampling frequency less than $2N_pW$, only the relatively low valued tails of the spectrum are aliased, and this contribution is small.

Furthermore, if the input signal is bandpass around some center frequency

$\pm\omega_0$, output signals around $\pm\omega_0$ are of greatest interest since output signal components around $\pm n\omega_0$ for $n > 1$ can easily be filtered out. This partial signal (around $\pm\omega_0$) is called the *first-zone* output. It can easily be shown that only odd-order terms in (6.5) contribute to the first zone output so it is common to model an amplifier using only odd order terms in (6.5) ($a_0 = a_2 = a_4 = \ldots = 0$).

A circumstance that can dramatically reduce the simulation times of a nonlinear system is if the input signal is a multicarrier signal with constant envelope subcarriers and the number of carriers M_f is large, as is common in OFDM systems. The intermodulation distortion can then be modelled by an additive Gaussian noise source [166]. This simplifications is not possible in multiple element antennas basestations, since the signals that enter the MCPA are few, and have a time varying envelope due to the near far ratio and the transmit weight adaption. The situation was improved by the work of Schneider et.al. [166] who presented a method to simulate a multicarrier digital communication system over a band limited channel with a nonlinear device. The method is based on the work by Shimbo [167] and reduces the simulation time significantly. The idea is to use Shimbo's analytical expression for the amplitudes of the intermodulation products that falls in the first zone output. Shimbo's notation is used in this chapter to describe the output from a non-linear amplifier as a sum over all the intermodulation products. This allows us to calculate the radiation pattern for each individual intermodulation product.

Effects of nonlinearities in array antenna systems

By introducing MCPA:s in MISO systems and assuming that they are not perfectly linearized and have finite peak power ratings, three effects can be observed:

1. Signal power from intermodulation distortion will be emitted from the antenna array. As opposed to a conventional basestation with one antenna, the array makes the radiated intermodulation distortion dependent on the azimuthal angle. Thus in some direction the intermodulation products from all n_t antennas adds coherently and a mobile on that frequency and in that particular direction will experience a maximal interference from IMD.

2. The dependence of the gain and phase shift of the MCPA on the input signal envelope will make the transmitting antenna branches unmatched. An intended "null" in the radiation pattern towards a known co-channel user will be shifted, and the interference for that particular user will increase. For a two element array and a phase error of 1° and an amplitude error of 0.1 dB, the maximum null depth is limited to 30 dB for interference cancellation [12]. Using MCPA:s the situation gets more complicated, since the phase

and amplitude shifts are then a function of the power of every signal entering the MCPA, due to the interaction between the input signals.

3. Due to the large signal envelope variations of the combined signal on the input of the MCPA, the peak-to-average ratio of the input signal is often limited prior to amplification, allowing for a reduction in the peak power rating of the amplifier. The limiting is often performed in software, prior to D/A conversion. This leads to considerable benefits in terms of power and cost savings in the design of the MCPA, on the other hand, the limiter will also generate intermodulation distortion that, unfortunately, will decrease the signal to interference ratio for the mobiles.

Although excellent linearization techniques have been proposed, there is a trade-off between linearity and cost, and if many basestations are to be equipped with array antennas, it is important that the employed linearization technique is inexpensive. Thus, a certain degree of nonlinearity must be tolerated. The required degree of linearity is investigated in this chapter and how it affects the performance of a cellular system. In Section 6.2 the signal and amplifier models are defined. Section 6.3 investigates how the transmitted IMD depends on the number of carriers M_f and number of transmit antennas n_t for a switched beam basestation antenna. A novel frequency allocation algorithm is proposed that reduces the effect IMD and thereby allows for an increase of the amplifier PAE. Finally, the SISO case is studied in 6.4 which has applications in WLAN systems and then the chapter is concluded in Section 6.5.

6.2 Signal and amplifier models

The models defined in Section 1.1 are now extended to the multicarrier case and the non-linear amplifier model is introduced. Note that continuous time representation of signals are used to be able to accurately describe the effects of the nonlinearities.

Signal Model

Assume a basestation (BS) that provides service in one sector in a tri-sector cellular communication system. An n_t element uniformly spaced linear antenna array is used for reception and transmission of data to the users, although the study here considers the downlink transmission only. Furthermore, assume that M_f FDMA frequency channels are used in the particular sector, each with a bandwidth of W Hz and equally separated by Δ_ω, see Figure 6.1.

Prior to, or integrated with each antenna, is an MCPA, see Figure 6.5. It provides the signals with necessary gain to overcome the path loss and the fading in

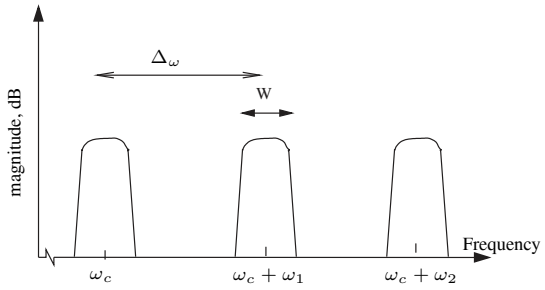


Figure 6.1: Input signal spectrum to the multicarrier amplifier. The sub-carriers of width W Hz are separated Δ_ω Hz.

the radio channel. The signal that enters MCPA l (at antenna l) can be written in the form [69]

$$c_l(t) = \sum_{m=1}^{M_f} \sqrt{P_m} s_m(t) w_{ml} \cos\left[(\omega_c + \omega_m)\, t + \phi_m(t) + \theta_{ml}\right] \tag{6.6}$$

where ω_c is the RF channel frequency reference and ω_m is the m:th sub-carrier frequency offset from this reference where $\omega_c \gg \omega_m$. Assume that the subcarrier's center frequencies, ω_m are equally spaced, as is common in FDMA systems

$$\omega_m = (m-1)\Delta_\omega + \omega_1 \ . \tag{6.7}$$

Furthermore, $s_m(t)$ is the m:th sub-carrier envelope, after modulation and pulse shape filtering, P_m is the corresponding scaling to set the correct average transmitted power for sub-carrier m and $\phi_m(t)$ is the m:th sub-carrier phase. The transmitted information is contained in $\phi_m(t)$ for the phase modulation of the signal and in $s_m(t)$ for the amplitude modulation. Furthermore, w_{ml} is the amplitude weight and θ_{ml} is the phase of the beamforming weight for sub-carrier m and antenna l. In some implementations the amplitude weight w_{ml} is independent of the sub-carrier and only phase adjustment is performed to shape the beam for a particular user signal $s_m(t)$. The antenna aperture can still be amplitude tapered using w_l to get the desired side lobe levels (SLL) or beamwidth (BW) of the main-beam [168].

Using complex baseband signal notation, (6.6) can be expressed as

$$c_l(t) = Re\left\{\breve{c}_l(t) \cdot e^{j\omega_c t}\right\} \tag{6.8}$$

provided that the bandwidth W is much less than the carrier frequency ω_c; $\breve{c}_l(t)$ is the complex envelope of $c_l(t)$ and is a low-pass baseband signal. The complex envelope can be written as

$$\breve{c}_l(t) = \sum_{m=1}^{M_f} \sqrt{P_m} s_m(t) w_{ml} e^{j(\omega_m t + \phi_m(t) + \theta_{ml})} \ . \tag{6.9}$$

Multicarrier Power Amplifier Model

Assume that the amplifier is memoryless, i.e. the AM/PM and AM/AM conversion characteristics are not functions of frequency. Stated differently, the total bandwidth of the amplified signal is much less than the inherent bandwidth of the amplifier. The memoryless baseband nonlinearity has a simple functional relationship between the input and output signal as

$$y(t) = F\left[c(t)\right] \tag{6.10}$$

where $F(\cdot)$ is the instantaneous transfer function. Using (6.10) directly leads to long simulation times as was discussed in the introduction 6.1. The use of complex low-pass equivalents in simulating communication systems makes the simulation more computationally efficient. Thus, the complex low-pass equivalent of the nonlinearity is required which is related to the instantaneous transfer function $F(\cdot)$ as follows. Consider a bandpass input

$$c(t) = A(t)\cos\left\{2\pi f_0 t + \phi(t)\right\} \tag{6.11}$$

and make the substitution $\alpha = 2\pi f_0 t + \phi(t)$. Now, write (6.10) as

$$y(t) = F\left[A\cos\alpha\right] \tag{6.12}$$

and expand it in a Fourier series

$$y(t) = a_0 + \sum_{k=1}^{\infty}\left(a_k\cos k\alpha + b_k\sin k\alpha\right) \quad . \tag{6.13}$$

Now, the output contains terms at many different frequencies, of which only the *first-zone* is of interest, since these will be centered at f_0, the center frequency of the input signal. This means that $k = 1$ in (6.13) and the DC-term is removed by setting $a_0 = 0$. The output is then

$$y(t) = a_1\cos\left\{2\pi f_0 t + \phi(t)\right\} + b_1\sin\left\{2\pi f_0 t + \phi(t)\right\} \quad . \tag{6.14}$$

The function $a_1 \equiv f_1(\cdot)$ is known as the (first-order) Chebychev transform [165] of $F(x)$

$$a_1 \equiv f_1(A) = \frac{1}{\pi}\int_0^{2\pi} F\left(A\cos\alpha\right)\cos\alpha d\alpha \tag{6.15}$$

and the function $f_2(\cdot)$ is a similar integral calculated as

$$b_1 \equiv f_2(A) = \frac{1}{\pi}\int_0^{2\pi} F\left(A\cos\alpha\right)\sin\alpha d\alpha \quad . \tag{6.16}$$

These integrals are by necessity, often calculated by using numerical methods and the function $(f_1(A) + jf_2(A))/A)$ is known in control theory literature as the *describing function* [169] of the nonlinearity. We now are ready to make the following definition:

Definition 6.1 *The function defined as*

$$g(A) \stackrel{\triangle}{=} \sqrt{f_1(A)^2 + f_2(A)^2} \tag{6.17}$$

is the AM/AM conversion characteristic of the amplifier and

$$f(A) \stackrel{\triangle}{=} \arctan\left\{\frac{f_2(A)}{f_1(A)}\right\} \tag{6.18}$$

is the AM/PM conversion characteristic[1]*. The functions $f_1(A)$ and $f_2(A)$ are given in* (6.15) *and* (6.16) *respectively and where $F(\cdot)$ is the instantaneous transfer function of the nonlinearity.*

Now, using (6.9), note that the amplitude is $A(t) \stackrel{\triangle}{=} |\breve{c}(t)|$. Using these relationships, the analytic signal input/output relation for the MCPA at antenna l can now be written as

$$\breve{y}_l(t) = g\left(|\breve{c}_l(t)|\right) \exp\left[j\phi_{c_l}(t) + jf(|\breve{c}_l(t)|)\right] \tag{6.19}$$

where $\phi_{c_l}(t)$ is the phase of the input signal $\breve{c}_l(t)$.

To model the amplifier nonlinearity, the memoryless envelope limiter model, proposed by Cann [170] is often used. It is described by the *baseband* transfer characteristic

$$y(t) = \frac{Dsgn(c(t))}{\left[1 + \left(\frac{l}{|c(t)|}\right)^s\right]^{1/s}} = F(c(t)) \tag{6.20}$$

where $y(t)$ is the instantaneous output, $c(t)$ the instantaneous input, D the asymptotic output level as the input amplitude $|c(t)| \to \infty$, $sgn(c(t))$ the signum function, l the input limit level and s is the "knee sharpness". Cann's model is used for its parametric form as opposed to polynomial amplifier models (6.5) which are better suited when fitting experimental data to a model. In Cann's model the degree of linearity is easily adjusted by varying the parameter s. Note that (6.20) must be converted to a low-pass equivalent of the first-zone output using (6.15) if it is to be used in efficient simulations. Recently, Loyka [171] showed that the use of Cann's model to predict levels of intermodulation products should be used with extreme care. The reason is that the derivatives of Cann's model (6.20) does not exist at

[1] Laboratory measurements of nonlinear amplifiers give the $g(A)$ and $f(A)$ functions directly, so by working with measured amplifier characteristics, the integrals (6.15),(6.16) need not be evaluated.

$c(t) = 0$ for many choices of the parameter s. Since the work on non-linearities in this thesis are of qualitative type and not quantitative, the drawn conclusions in this chapter are not affected by this late discovery. However, the actual IMD levels from simulations should be handled with caution.

The nonlinear amplifier model is the used in the simulation at a certain IBO or OBO from (6.3) or (6.4). As the input signal is a multicarrier signal, the average input power $P_{avg,in}$ is estimated in the simulations during a pre-run phase and then the desired IBO/OBO for the amplifier is set by scaling the amplitude of the input signals.

A rewriting of the MCPA output signal

The multicarrier input signal (6.9) to the MCPA generates a large number of intermodulation products (IMP)[2] in the output signal and a combinatorial problem follows to determine the frequency and relative phase of each of the IMP:s. Adding the complexity of an array antenna, the spatial dispersion of the IMP:s must also be considered. Each IMP will have a certain radiation pattern and it is possible to calculate the direction of the maximum of each of these IMP:s by first rewriting equation (6.19) for the output of MCPA l as

$$\breve{y}_l(t) = \sum_{\mathcal{K}} S\left[\mathcal{K}, \mathbf{A}(t)\right] \exp\left(j \sum_{m=1}^{M_f} k_m \left(\omega_m t + \phi_m(t) + \theta_{ml}\right) \right) \tag{6.21}$$

where $\mathbf{A}(t) = \left[A_1(t), A_2(t), \cdots, A_{M_f}(t)\right]$ are the envelopes of the transmitted signals and $\mathcal{K}$ is the following set of integer indices

$$\mathcal{K} = \left\{ \left\{k_1, k_2, \ldots, k_{M_f}\right\} : \sum_{m=1}^{M_f} k_m \omega_m = \omega_p \right\} \tag{6.22}$$

to consider the output in the frequency channel with center frequency ω_p only. If ω_p is one of the sub-carriers center frequencies $(\omega_1, \ldots, \omega_{M_f})$, then the summation in (6.21) will be restricted to terms that lie in the zone of the first harmonic (fundamental) of the output signal. To study the fundamental and third order IMP solely, $\mathcal{K}$ in (6.21) is easily exchanged for the subset $\mathcal{K}_{1,3}$, defined as

$$\mathcal{K}_{1,3} = \mathcal{K} \cap \left\{ \left\{k_1, k_2, \ldots, k_{M_f}\right\} : \sum_{m=1}^{M_f} |k_m| = 1, 3 \right\} \quad . \tag{6.23}$$

[2] A note on the nomenclature: Intermodulation distortion (IMD) is the total output power generated from the non-linearity, whereas intermodulation products (IMP) is defined as the particular intermodulation distortion product generated by the intermodulation between two tones on distinct frequencies on the input of the non-linearity. Hence, the union of all output IMP:s is denoted IMD.

Furthermore, in equation (6.21), $A_m(t) = \sqrt{P_m}s_m(t)$ is the sub-carrier m:s envelope and unity amplitude weights $w_{ml} = 1$ has been assumed. The complex valued function $S[\cdot]$ is known as the Shimbo amplitude function (SAF)[166, 167], and gives the amplitude and relative phases of the output intermodulation products with indices $\{k_i\}_{i=1}^{M_f}$. Shimbo's approach is described below and we start with the following definition.

Definition 6.2 *The function $S\left[\mathcal{K}, \mathbf{A}(t)\right]$, called the Shimbo Amplitude Function (SAF), is defined as*

$$S\left[\mathcal{K}, \mathbf{A}(t)\right] \triangleq \int_0^\infty r \prod_{m=1}^{M_f} J_{k_m}\left[A_m(t)r\right] \int_0^\infty \rho g(\rho) e^{jf(\rho)} J_1(r\rho)\, d\rho\, dr \quad (6.24)$$

where $A_m(t)$ is the envelope of the m:th input signal component, $J_p(x)$ is the pth order Bessel function of the first kind and $f(A), g(A)$ are given by Definition 6.1.

The SAF is a complex valued function and it gives the complex baseband amplitude of the signal component in the output signal corresponding to the set of indices $\{k_m\}_{m=1}^{M_f}$. The calculation of SAF involves the amplifier characteristics $f(\cdot)$, $g(\cdot)$ and the envelope of the sub-carriers in the multicarrier input signal. Note that the SAF depends on the envelopes and not the phases of the input sub-carrier signals and if the amplifier does not exhibit any AM/PM conversion ($f(A) = $ const.), the SAF will be real. In general, the calculation of (6.24) is not simple, but Shimbo found that the Bessel function series expansion of the nonlinearity $g(A)e^{jf(A)}$, yields a simple and useful expression for the SAF [172]. With this approach it is possible to find an analytical expression for the amplitude and phase of the desired signals and each of the IMP:s in the output of a nonlinear amplifier with a multitone input.

If $g(A)$ and $f(A)$ are the AM/AM and AM/PM characteristics of the amplifier, then the coefficients α, β_s in a Bessel series expansion are found by solving

$$\{\alpha, \beta_1, \ldots, \beta_{M_S}\} = \arg \min_{\alpha,\beta_1,\ldots,\beta_{M_S}} \left\{ \left| g(x)e^{jf(x)} - \sum_{s=1}^{M_S} \beta_p J_1(\alpha s x) \right|^2 \right\} \quad (6.25)$$

for all amplitudes x in the input amplitude interval. Here J_1 is the 1st order Bessel function of the first kind and note that the coefficients β_s are complex. Using the coefficients from the least square fit in equation (6.25), the complex voltage gain for the p:th component in the output at frequency ω_p can be approximately written as [167]

$$S\left[\mathcal{K}, \mathbf{A}(t)\right] \approx \sum_{s=1}^{M_S} \beta_s \prod_{m=1}^{M_f} J_{k_m}\left(\alpha s A_m(t)\right) \quad (6.26)$$

where J_{k_m} is the Bessel function of order k_m. Note that the output amplitude of the desired, linearly amplified carrier at frequency ω_p is given by $S\left[\mathcal{K}, \mathbf{A}(t)\right]$ with $\mathcal{K} = \{\ldots, 0, \underset{p:th}{1}, 0, \ldots\}$.

6.2.1 The far field radiation pattern

Assume that the angle Θ is measured from broadside of the array. The complex envelope of the received signal $z(t, \Theta)$ at a hypothetical user in the far field and in direction Θ on frequency ω_p is represented by a discrete complex sequence $z(n, \Theta)$ [108]

$$z\left(n, \Theta\right) = G_T(\Theta) \sum_{l=1}^{n_t} \breve{y}_l(n) e^{-j\delta_{TX}\omega_p(l-1)\sin(\Theta)/c} \tag{6.27}$$

where δ_{TX} is the array inter-element spacing of an assumed uniform linear array, c the speed of light and $G_T(\Theta)$ is the individual antenna element gain in direction Θ which is assumed equal for all n_t antenna elements. The far field effective isotropic radiated power (EIRP) in direction Θ from broadside direction is

$$EIRP(\Theta) = \left|\frac{z(n, \Theta)}{n_t}\right|^2 = \left|\frac{G_T(\Theta)}{n_t} \sum_{l=1}^{n_t} \breve{y}_l(n) e^{-j\delta_{TX}\omega_p(l-1)\sin(\Theta)/c}\right|^2 . \tag{6.28}$$

By inserting (6.21) into equation (6.28) one finds that for a particular set of indices, $\{k_m\}_{m=1}^{M_f}$, which uniquely labels any IMP, the far field radiation pattern as

$$EIRP(\Theta, \mathcal{K}) = \left|\frac{G_T(\Theta)}{n_t} S\left[\mathcal{K}, \mathbf{A}(t)\right] \sum_{l=1}^{n_t} \prod_{m=1}^{M_f} e^{jk_m(\omega_m t + \phi_m(t) + \theta_{ml}) - j\delta_{TX}\omega_p(l-1)\sin(\Theta)/c}\right|^2 \tag{6.29}$$

where it was assumed that the amplitudes of the M_f subcarriers does not depend on the antenna number. This is the case of phase only beamforming without amplitude tapering. If amplitude weights or tapering is used, then $\mathbf{A}(t)$ will be different for different antennas and the SAF will depend on n well. In this case, the SAF cannot be moved outside the summation in equation (6.29).

Switched Multibeam Antenna Arrays

An important sub-class of beamforming methods is now introduced which further simplify the expression for the far field radiation pattern (6.29), namely the switched multibeam antenna array, see Section 2.5.2. This beamforming method is

often used for the downlink beamforming in FDD systems, where full knowledge of the downlink channel state information (CSI) is unknown. From the uplink data, some partial CSI can be extracted such as the average angle of arrival (AOA) of the signals. It was shown in [173] that in a NLOS environment the AOA change at the BS for the strongest path was less than 3° for a mobile movement of $2\lambda_c$. So, one choice for the downlink beamformer is to form a beam with low side-lobes and transmit in the direction where the largest average power was received from the particular user in the uplink direction. The low side-lobes of the radiation pattern decreases the co-channel interference to users in neighboring cells. A linear array antenna with n_t antenna elements can generate n_t beams with distinct spatial orientations in the horizontal plane, by use of a beam forming network (BFN), often implemented using a Butler matrix [20, 78], see Appendix A. It has the property of a constant phase gradient $\Delta\theta_m$ over the antenna array aperture so for antenna l and signal m, the phase of the beamformer weight is

$$\theta_{ml} = (l-1)\Delta\theta_m \tag{6.30}$$

for $l = 1, \ldots, n_t$. The phase gradient, $\Delta\theta_m$ belongs to the set Ω_S of n_t different phase gradients, unique for each beam the BFN can generate,

$$\Omega_S \triangleq \left\{ \frac{2\pi p}{n_t} \right\}_{p=0}^{n_t-1} . \tag{6.31}$$

The set of phases, Ω_S in (6.31) is a closed group under addition, subtraction and integer multiplication, following modulo-2π algebra, due to the phase angle ambiguity of 2π. This property implies that the main-lobe direction of the desired signals are also main-lobe directions for the IMD, as will be shown below.

By using (6.30) for the FFT based BFN, in equation (6.29), it can be simplified as

$$EIRP(\Theta, \mathcal{K}) = \left| \frac{G_T(\Theta)}{n_t} S\left[\mathcal{K}, \mathbf{A}(t)\right] \right|^2 \frac{\sin^2\left[n_t \xi(\Theta)\right]}{\sin^2\left[\xi(\Theta)\right]} \tag{6.32}$$

where

$$\xi(\Theta) = \frac{1}{2} \sum_{m=1}^{M_f} \left(k_m \Delta\theta_m - \delta_{TX}\omega_p \sin(\Theta)/c\right). \tag{6.33}$$

Note that the $EIRP$ in (6.32) is maximized when $\xi(\Theta) = 0$. Hence the radiation pattern for the linearly amplified signal $s_1(t)$ is given by setting $\mathcal{K} = \{1, 0, 0, \ldots, 0\}$ and thus $\omega_p = \omega_1$ from (6.22). Hence the maximum power is radiated in direction Θ_{max}, where, from (6.33):

$$\Delta\theta_1 - \delta_{TX}\omega_p \sin(\Theta_{max})/c = 0 \ . \tag{6.34}$$

Solving for Θ yields

$$\Theta_{max} = \arcsin\left(\frac{\Delta\theta_1}{\pi}\right) \tag{6.35}$$

where $\Delta\theta_1$ is the phase gradient for user 1 and where it was assumed that the elements are spaced $\delta_{TX} = \lambda_c/2$. Furthermore, the direction of the maximum radiated IMP for a third order intermodulation product at frequency $\omega_p = 2\omega_j - \omega_k$ can be found, if $\delta_{TX} = c\pi/\omega_p$ (half a wavelength spaced antenna elements) as

$$\Theta_{max,IMP} = \arcsin\left(\frac{2\Delta\theta_j - \Delta\theta_k}{\pi}\right). \tag{6.36}$$

It is easily verified that (6.36) maximizes (6.32) for the $2\omega_j - \omega_k$ intermodulation product.

Equation (6.32) allows for introducing an equivalent phase gradient for the particular IMP. It is calculated as

$$\Delta\theta_{eq} = \sum_{m=1}^{M_f} k_m \Delta\theta_m \tag{6.37}$$

which will maximize the P_{EIRP} in a direction given by the angle

$$\Theta = \arcsin\left(\frac{c\Delta\theta_{eq}}{\delta_{TX}\omega_p}\right) \quad . \tag{6.38}$$

Since Ω_S is closed under integer multiplication and addition, $\Delta\theta_{eq}$ will also belong to Ω_S and this implies that the IMP directions will coincide with the directions of the linearly amplified signals. We illustrate this with an example:

EXAMPLE 6.1

Assume that two signals for two users, $\breve{c}_1(t)$ and $\breve{c}_2(t)$, separated in frequency according to (6.7) are transmitted in the same beam. They will then use equal phase gradients ($\Delta\theta_1 = \Delta\theta_2$) and from (6.37) they will generate third order IMP ($k_1 = 2, k_2 = -1$ and vice versa) with equal phase gradients $\Delta\theta_{eq} = \Delta\theta_1 = \Delta\theta_2$, i.e. IMP at the frequencies $\omega_3 = 2\omega_2 - \omega_1$ and $\omega_0 = 2\omega_1 - \omega_2$ are radiated in the same beam directions as the amplified original signals $\breve{c}_1(t)$ and $\breve{c}_2(t)$.

If, however, the two signals are transmitted in distinct beams, $\Delta\theta_1 \neq \Delta\theta_2$, then the third order IMP are radiated with phase gradients $\Delta\theta_{eq} = 2\Delta\theta_2 - \Delta\theta_1$ and $\Delta\theta_{eq} = 2\Delta\theta_1 - \Delta\theta_2$ at the frequencies $\omega_3 = 2\omega_2 - \omega_1$ and $\omega_0 = 2\omega_1 - \omega_2$ respectively. Note that ω_0 is outside the defined transmitter band (since $\omega_0 < \omega_1$

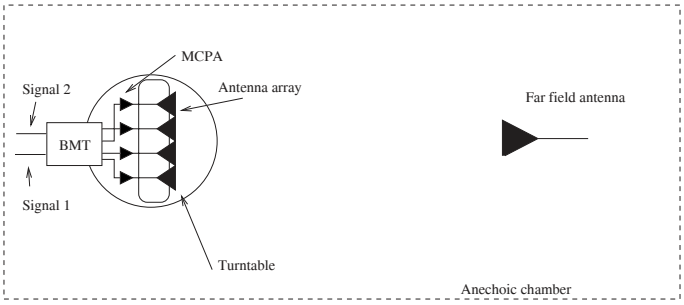


Figure 6.2: Measurement setup in the anechoic chamber for intermodulation measurements. A four element antenna array was used and a BMT for the beamforming.

and ω_1 is the lowest used frequency of the system), and can thus be removed by the transmit filter. The structure of the FFT beamformer implies that the equivalent phase gradients of the IMP, $\Delta\theta_{eq}$ belongs to the set Ω_S as well.

6.2.2 Measurements of the far field radiation pattern

To verify the results of the previous Section 6.2.1, measurements on a switched beam array antenna was performed in an anechoic chamber using a 4-element linear antenna array and an 8-by-8 analog Butler matrix. See Appendix A for a overview of the Butler matrix. Four standard microwave amplifiers were used, and they were biased in the nonlinear region to generate IMD. The setup is shown in Figure 6.2. The Butler matrix was connected to two GSM signals with center frequencies 1.8000 GHz and 1.8004 GHz in two different input ports. The input power of the two signals were adjusted to the appropriate input back-off and the radiation pattern was measured using the anechoic chamber with a turntable as shown in Figure 6.2. The angle was swept in 1° steps and for each angle the power density spectrum was measured with a spectrum analyzer.

Frequency-angle power spectral density measurements

The measurements give an intuitive picture of the spatial distribution of the intermodulation distortion. Input port number 6 was used for the f_1=1.8000 GHz signal and port number 4 was used for the f_2=1.8004 GHz signal. According to the theory in the previous section, the main-lobes will in this case be in the directions 22.0° and -7.2° respectively, while the third order intermodulation beams are in direction 61° for the $2f_1 - f_2$ product at 1.7996 GHz and $2f_2 - f_1$ product at 1.8008

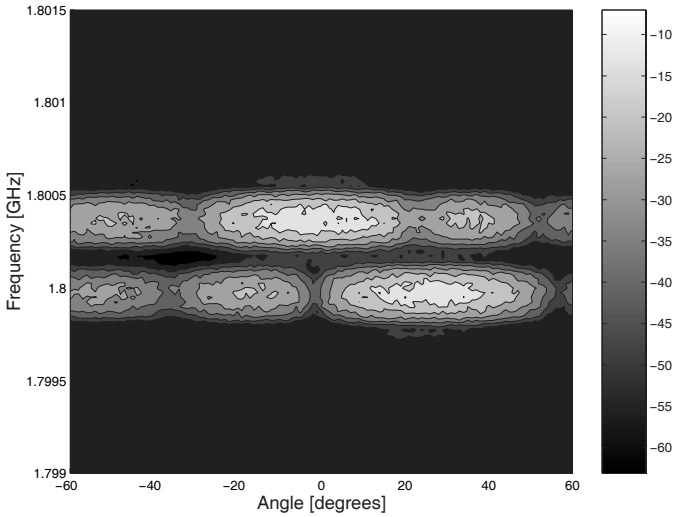


Figure 6.3: Frequency-angle plot in decibel of the power spectral density with large back off. The gray-scale shows different power levels, normalized so that maximum received power is 0 dB. The two beam-formed signals has main-lobes in 22° and −7° respectively. The intermodulation products are weakly visible at -39° and 1.8004 GHz.

GHz in $-39°$. This is easily verified by studying Figure 6.3 and Figure 6.4 for a large and small IBO respectively. These figures are normalized to 0dB maximum. In the small IBO case, see Figure 6.4, the fifth-order intermodulation product (e.g. $k_1 = 3, k_2 = -2$) is weakly visible at the DOA angles $61°$ and $-39°$. Noteworthy is also the orthogonal array response vector properties of the beams from the Butler matrix. In direction $-7.2°$, where beam 2 has its maxima, the other beams have nulls, giving a high level of isolation. This includes the intermodulation, which also has a null in that direction. So, by using a Butler matrix, the spatial dimension can be used to reduce the intermodulation distortion at the mobile user. However, the third order nonlinearity adds up coherently in another beam which increases the interference level of a potential co-channel user located in that direction. This property of spatial intermodulation distortion filtering by the array antenna has earlier been described in [159] for satellite mounted array antennas and also in a recent paper by Hemmi [161].

The analog Butler matrix used in this experiment have a manufacturing error of $0.8dB$ and $8.5°$ which destroys the orthogonality properties of the Butler matrix array response vectors. These errors degrades the isolation and explains the not perfect positioned nulls in the measurements. This also demonstrate the need for calibration.

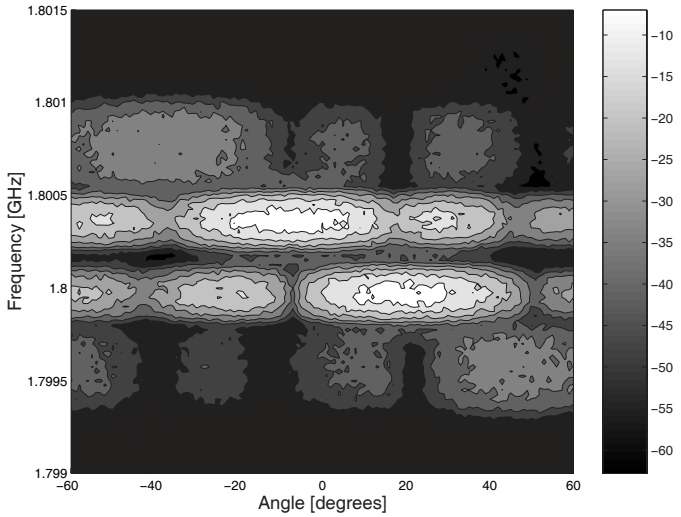


Figure 6.4: Frequency-angle plot in decibel of the power spectral density with saturated amplifiers (back-off decreased 8 dB compared to Figure 6.3). The gray-scale shows different power levels, normalized so that maximum received power is 0 dB. The third order intermodulation products are visible and also the fifth order intermodulation products are weakly visible.

6.3 Nonlinear Distortion in Switched Multibeam FDMA Systems

The results in the previous section showed that the IMD power level is dependent on the azimuthal angle from the basestation antenna array. This encourages the idea of a smart frequency allocation method for the users in the served cell to reduce the carrier to intermodulation distortion ratio (CIMR). A novel frequency planning strategy is proposed in this section which reduces the effective interference from the third order IMP. Figure 6.5 shows an overview of the assumed transmitter architecture using fixed multibeam beamforming. The motivation for using multibeam antennas was discussed in Section 2.5.2.

In some systems, the IMD is a substantial source of interference and must be included in the interference budget, although it has been shown that the IMD can be reduced by different linearization techniques [164]. Some residual IMD power will however still be emitted by the antennas, and it is important to be able to predict this level to assure an acceptable system performance and system designers should be aware of how the array can amplify the IMD.

Not all IMD are harmful however, since some parts of the distortion power can easily be removed by filtering. If $\{\omega_1, \ldots, \omega_{M_f}\}$ is the set of M_f center fre-

quencies of the modulated carriers that enters the MCPA, then it was shown in the previous section that some third order intermodulation products (IMP) in the output signal fall onto other frequency channels that are used in the FDMA system and thereby cannot be removed by filtering. In this section, only the third order IMP are considered since the amplifiers operate well below the 1 dB compression point and in this region, the third order IMP dominates over higher order IMP so the distortion power from higher order IMP can safely be neglected. This is a common assumption when studying effects of nonlinearities in communications systems.

The degree of a non-linearity is often measured by a two tone test, where two continuous wave (CW) tones with distinct frequencies are combined and connected to the input of the amplifier. The power of the IMP are measured in the output signal and the ratio to the power of the desired signal is calculated as a measure of the non-linearity. In the third generation Universal Mobile Telephony Systems (UMTS), the input signal has a 5 MHz bandwidth and a nonlinear amplifier will then introduce spectrum regrowth. The CW characterization method will then fail to give an accurate measure of the effects of the nonlinearity, instead, the Adjacent Channel Leakage power Ratio (ACLR) is defined as the ratio of the transmitted power to the power measured after a receiver filter in the adjacent channels [174]. The receiver filter has a -3 dB bandwidth equal to the chip rate (3.84 MHz) in the 3GPP-standard of UMTS.

In a cellular system, the dominant IMD will come from the MCPA in the same cell, if the generated IMD falls onto other frequency channels used in that cell. Other examples of problems related to MCPA generated IMD are in multi-layer cellular network structures, where IMD from an MCPA-equipped micro-cell interfere with a macro-cell connected mobile [175], or as was investigated in [176], the mutual interference between an AMPS system BS using MCPA:s and a CDMA system mobile.

The aim of this section is to show how the use of multibeam antennas reduces the downlink IMD for the mobiles in one cell and in the next section the effect on the system level outage probability is investigated by considering more than one basestation equipped with switched multibeam arrays. Furthermore, the spatially filtering of IMP:s can be utilized by the frequency channel allocation algorithm in the BS that assigns new mobiles to one of the empty frequency channels in the particular cell sector. If a non-frequency hopping system is assumed, then each new mobile can be allocated to the frequency channel that generates least IMD to other mobile users in the same sector. This was briefly mentioned by Sandrin [159], and here an algorithm for this allocation procedure is proposed, and the performance is compared to the random frequency channel allocation algorithm. The solution is related to the recently proposed techniques for reducing IMD in a conventional one-antenna BS [177, 178] by spacing the used frequency channel unequally, to

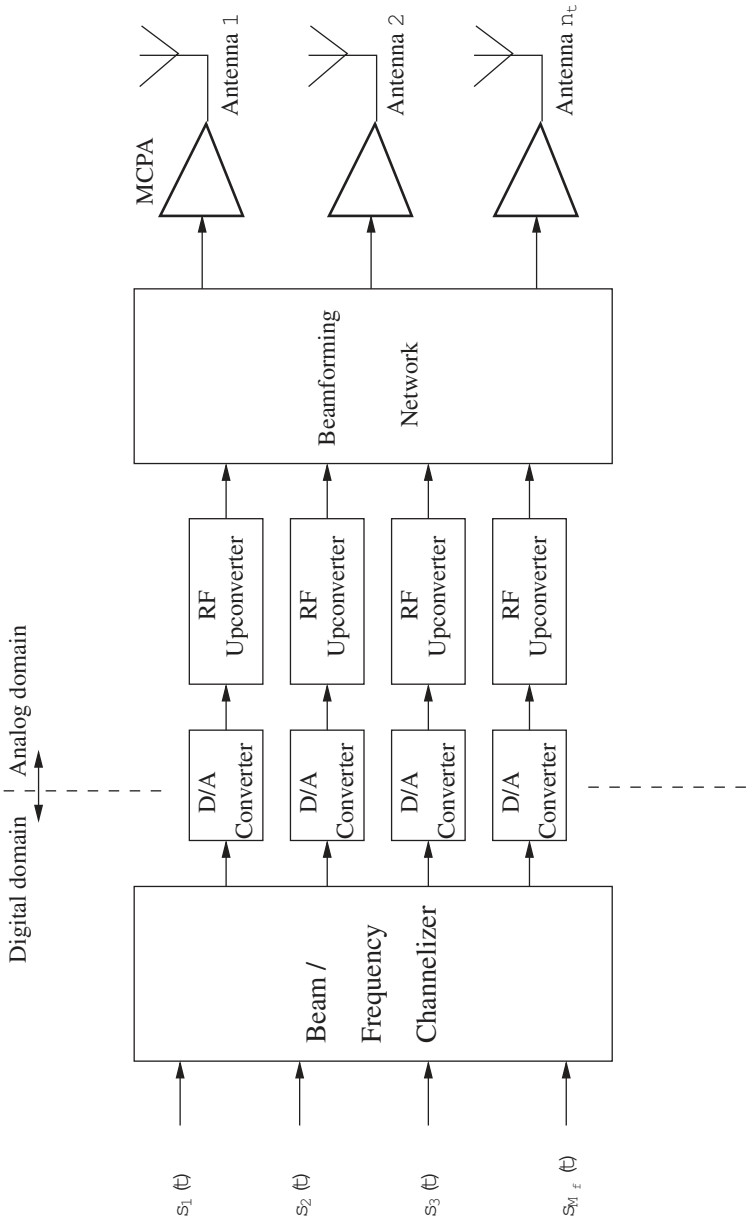


Figure 6.5: Switched multibeam array architecture. The beam/frequency channelizer directs the input signals to their respective FDMA frequency slots and selects the "best" downlink beam using information from the uplink.

make the IMP fall onto unused frequency channels in between.

6.3.1 Utilizing the Combined Beam-Frequency Scheme

In Section 6.2.1, an expression for the far field radiation pattern of the array antenna was derived. The result included both the radiation patterns for the desired signals and the intermodulation products. Assume that the frequency channels are equally spaced with frequency channel m having a center frequency ω_m, according to (6.7). It is then possible drop the actual frequency and proceed in terms of frequency channel numbers. The frequency channel with center frequency ω_k has by definition the frequency channel number k. Hence, the frequency channel numbers of the third order in-band IMP generated by mixing the channels with number j and l respectively are given as

$$k_{IMP1} = 2j - l \tag{6.39}$$

$$k_{IMP2} = 2l - j \tag{6.40}$$

which then corresponds to the frequency channels with center frequencies $\omega_{k_{IMP1}}$ and $\omega_{k_{IMP2}}$ respectively. Note that these generated frequencies might fall outside the frequency band of interest, if for example $\omega_{k_{IMP1}} < \omega_1$. They can then easily be removed by the transmit filters and will thus not act as a source of interference in the system.

The number of IM products

The number of generated IMP in the output signal is required to calculate the total IMD emitted from the array antenna. We now define the two types of third order IMP.

Definition 6.3 *Assume that M_f equally spaced sub-carriers at frequencies $\{\omega_i\}_{i=1}^{M_f}$ are co-amplified in an MCPA. The third order IMP in the output are categorized into two different groups. Type A IMP is defined as the cross-modulation term $2\omega_i - \omega_j$ and type B IMP as the composite triple-beat terms $\omega_i + \omega_j - \omega_k$.*

The following analysis is thus restricted to third order IMP but an extension to higher order IMP is straightforward but tedious.

Type B IMP becomes more deleterious when the total bandwidth is less than an octave because most of the type A IMP falls fall outside the band of interest. The center frequency channel in the transmit band, receives the highest number of IMP, and we denote the number of IMP that falls on this center channel as ν_c. It is shown in [177] that if all subchannels are active, this number is

$$\nu_c = \frac{3}{8}\left(M_f^2 - 2M_f + 1\right) \quad . \tag{6.41}$$

The ν_c IMP in the center channel can be divided into M_A type A as [178]

$$M_A = \begin{cases} (q-1) & \text{if } M_f\text{=2q} \\ 2\lfloor q/2 \rfloor & \text{if } M_f\text{=2q+1} \end{cases} \tag{6.42}$$

and the number of type B IMP is

$$M_B = \nu_c - M_A \ . \tag{6.43}$$

If the center channel is left empty, as common in noise power ratio (NPR) measurements, then all IMP involving the carrier in the center channel most be subtracted from (6.41). This gives, after some calculations, the number of IMP in the center channel if the center channel is empty, denoted as ν_c^e

$$\nu_c^e = \frac{1}{8}\left(3M_f^2 - 10M_f + 7\right) \tag{6.44}$$

where it is assumed that M_f is an odd number.

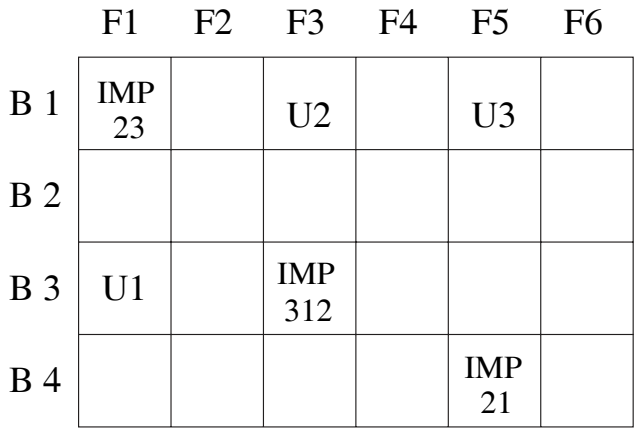
Beam-frequency scheme

Using the above definitions, given a set of M_f equally spaced frequency channels, and a set of n_t beams, the beam-frequency scheme can be defined, which can be used to reduce the IMP levels in the system. We now define the concept of beam-frequency slots.

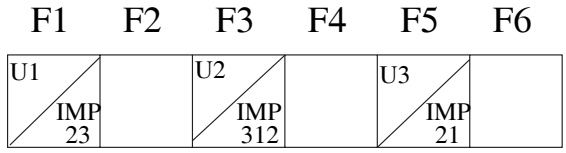
Definition 6.4 *Assume that transmission takes place from a switched multibeam array using one out of the n_t beams. Denote its number Bk, where $k \in \{1, \ldots, n_t\}$. The transmission takes place on a frequency channel with center frequency ω_l and thus the frequency channel number is $l \in \{1, \ldots, M_f\}$. Then the transmission is said to take place on beam-frequency slot (Bk,Fl).*

Main beam IMP is defined as the direction of the maximum radiation intensity of the particular IMP. Certainly, IMP will be radiated in all directions, but suppressed by the side-lobe of the array radiation pattern in other directions than the main beam direction.

Furthermore, the FFT based BFN is used, so the properties of closeness and finiteness of the *modulo* 2π addition and integer multiplication of the phase gradients (6.37) makes the main beam direction of the IMP coincide with the set of main beam directions for the desired (linearly amplified) signals. Figure 6.6 shows an example of the beam-frequency slot scheme where three users have been randomly assigned (in frequency) to three slots. The figure shows the slots where main beams of type A and type B IMP will appear (type B IMP312, and two type A IMP23,



(a)



(b)

Figure 6.6: Beam-frequency scheme for (a) n_t=4, (b) n_t=1, where M_f=6. The figure shows three active users (U1,U2,U3) occupying three FDMA slots. Only third order IMP that falls in the frequency channels of interest ($F1$-$F6$) are considered. Type A IMP is denoted "IMP ab" when user a and b generates the IMP. User 1,2,3 generates type B IMP denoted "IMP abc".

IMP21). The n_t=4 element antenna array will certainly reduce the amount of IM distortion at the mobile as compared to the reference n_t=1 antenna BS in Figure 6.6(b) because main beam IMP falls onto other beam slots, or stated differently, in other directions than the user main beams. Hence, the IMP will be suppressed by the side-lobes of the antenna array radiation pattern.

Reducing the IMD by using an antenna array

To calculate the reduction in IMD due to the use of a multibeam antenna, statistical methods must be used, as the activity of the users as well as their angle Θ seen

from the BS, are stochastic processes. We now state the following:

Theorem 6.1 *Assume a fully loaded system (all M_f frequency channels are occupied) and that M_f is an odd number. The position in angle of the mobiles in the served sector are assumed independent and uniformly distributed in the range $[-\pi/3, \pi/3]$ and the number of beams is n_t, which are assumed to have equal beam-widths. The number of type A and type B intermodulation products, denoted ν_A and ν_B, with the main beam pointing at the mobile that is active on the center channel, that is, the frequency channel number q, where $M_f = 2q+1$, is binomially distributed Bin(M_x,1/n_t):*

$$P_{M_x}(\nu_x) = \left(\begin{array}{c} M_x \\ \nu_x \end{array} \right) \left(\frac{1}{n_t} \right)^{\nu_x} \left(1 - \frac{1}{n_t} \right)^{M_x - \nu_x} \tag{6.45}$$

where $x = A, B$ and M_A,M_B are given in (6.42) *and* (6.43).

Proof: We want to derive the distribution of the number of type x IMP, denoted ν_x, that point its main beam on the mobile that is active on the center channel. The total number of IMP falling on this channel is given by M_A and M_B from (6.42). Since the positions of the mobiles on other frequency channels than the center one, are assumed to be uniformly distributed, the IMP that falls on the center frequency channel will have its main beams pointing in directions uniformly distributed over the n_t beam directions. Assume that the mobile under study is located in a specific beam direction Bk. The probability that a certain IMP falling on the center frequency channel has it main beam pointing in the direction Bk is $1/n_t$. Since there are M_x independent IMP:s falling on the center frequency channel and each has a probability $1/n_t$ to interfere with the main beam pointing towards the mobile under study, the number ν is binomially distributed *Bin*(M_x,1/n_t). ■

Remark that we study the mobile in the center frequency channel, since it receives the largest amount of IMP, hence this is a worst case analysis. The expectation value of (6.45) is M_A/n_t and M_B/n_t for type A and B IMP respectively and the variance is $M_A(n_t-1)/n_t^2$ and $M_B(n_t-1)/n_t^2$ respectively. If the number of IMP in a specific beam slot is less than the mean value of (6.45), then the IMD power in this direction is less than average. If the number of frequency channels M_f is increased, and keeping the number of beam directions n_t fixed, then the quotient of the number of IMP in a beam compared to any of the $n_t - 1$ other beams is on average approaching unity. By defining the normalized variance, or *variability*, υ_x, of the distribution (6.45), as

$$\upsilon_x = \frac{E\left\{\nu_x^2\right\}}{E\left\{\nu_x\right\}^2} = \frac{n_t - 1}{M_x} \tag{6.46}$$

for $x = A, B$, it is possible to study this. The variability indicates that when the number of IMP of type x, M_x, is increased, as when the number of frequency channels M_f is increased, the IMP are evenly distributed over all beams, and the total IMP radiation pattern, which is the sum of all IMPs, is approaching the radiation pattern of a single antenna element. Note that the individual IMP maintain their ideal array pattern, but their sum has a "smoothing" effect on the radiated IM power. This was also observed in [153]. By increasing the number of antennas n_t, the variability (6.46) is increased, and the total IMP radiation pattern directivity is increased, i.e. lobes are formed is some directions. For higher order IMPs, the number of IMPs, M_x, is a large number compared to n_t and the small variability leads to a radiation pattern close to the single antenna element pattern.

Hence, it is expected that by using an n_t beam array antenna, the IM power experienced by the users, compared to a reference $n_t = 1$ conventional BS, are on average reduced $10\log_{10}(n_t)$ dB, as the number of main beam IMP towards the mobile are on average reduced by the factor $1/n_t$. Note also that the transmitted power from each antenna is reduced by $20\log_{10}(n_t)$ dB compared to the n_t=1 antenna case, to yield the same received power at the mobile.

If the number of active frequency channels is less than M_f, which is the normal case[3], then it is shown below how it is possible to reduce the amount of IMD further, by utilizing the extra degree of freedom the n_t beams provide.

IM-reducing frequency channel allocation

Assume in the following a non-frequency hopping system, or a system with deterministic frequency hopping. In the example of Figure6.6, a new user can be placed in any of the empty frequency channels $F2$,$F4$ or $F6$. Depending of the beam allocation for the new user (which is out of the basestations control, it depends on the spatial position of the new user), the basestation should allocate the new user to a frequency channel that produces the lowest number of "collisions", where a collision has taken place when the generated main beam IMP falls into a slot that is occupied by a user. The basestation performs a search over all free slots and calculates the number of collisions in the $M_f \times n_t$ beam-frequency scheme for each case. When a call is dropped, a new frequency channel is available and added to the set of searched slots for the next new user.

To compare the proposed algorithm, the expected value of the number of collisions is calculated and compared with the random frequency channel allocation algorithm. The number of slot collisions depends on the number of available beams and on the number of active users in the particular sector. Define the probability

[3] M_f active frequency channels corresponds to a blocking probability of one for a new user, and is a rare case if the system is properly planned.

p_{ch} as the probability that a given frequency channel is in use in a cell, which is a function of the offered traffic A_o (in Erlang), the blocking probability P_B and the number of frequency channels M_f in the sector [73],

$$p_{ch} = \frac{A_o}{M_f}(1 - P_B) \ . \tag{6.47}$$

The number of active users in the sector is a random variable varying from zero to M_f and has a binomial distribution $P_a(m) \sim \textit{Bin}(M_f, p_{ch})$ [73]. If blocked calls are cleared (calls arriving when all the channels are found to be busy are lost), then the blocking probability is given by the Erlang-B formula

$$P_B = \frac{A_o^{M_f}/M_f!}{\sum_{k=0}^{M_f} A_o^k/k!} \ . \tag{6.48}$$

To derive the expected number of collisions C_o, the conditional expectation

$$E\left\{C_o\right\} = \sum_{m_u=0}^{M_f} E\left\{C_o|m_u\right\} P_a(m_u) \tag{6.49}$$

is used, where E is the expectation operator. The conditional expectation value $E\left\{C_o|m_u\right\}$ of C_o collisions conditioned on m_u active users involves summation over all possible beam-frequency slot allocations and its calculation is a formidable task. It depends on the chosen frequency channel allocation algorithm, the number of beams n_t, the number of frequency channels M_f and on the blocking probability P_B. To illustrate how the algorithms performs, an example is presented, where a Monte Carlo simulation method is used to estimate $E\left\{C_o|m_u\right\}$ which then is used to calculate the expected value of C_o.

The system is simulated with M_f=9 frequency channels and blocking probability $P_B = 2\%$. This gives, by using (6.48), the offered traffic A_o=4.34 Erlang per sector and hence a channel activity p_{ch} of 47.3%. The simulation is performed as follows. In each simulation step, a mobile is activated in a random direction or a random mobile drops its call. This is simulated by using a generalized birth-death process with $M_f + 1$ states [179]. For each new mobile, all unoccupied frequency channels are searched and the number of collisions are counted. The slot allocation that gives the fewest collisions is chosen for the new user. If two or more slots give the same number of collisions, then the total number of generated IMP in the $n_t \times M_f$ beam-frequency scheme is calculated for the slots that gave the same number of collisions, and the algorithm choose the slot that generates the fewest IMP. 10 000 steps were performed to estimate $E\left\{C_o|m_u\right\}$ for each case of n_t=8,4,2 and also n_t=1 for comparison. Table 6.1 shows the reduction in number of collisions

Antennas n_t	Channel Allocation	Estimated $E\{C_o\}$
8	Best	0.41
8	Random	2.15
4	Best	1.42
4	Random	4.02
2	Best	4.24
2	Random	8.12
1	Best	10.70
1	Random	16.19

Table 6.1: Estimated number of collisions when using proposed algorithm for frequency channel allocation compared to random allocation.

when the proposed algorithm is used as compared to random frequency channel allocation. It is interesting to see that four beams and using the proposed algorithm, has on average fewer collisions than an eight beam system using a random frequency channel allocation. This can also be seen in Figure 6.7 where random/n_t=8 have a higher collision count than the best/n_t=4 algorithm, when the number of active users are less than seven.

When the number of beams (antennas) is increased, the beam-width of the main beam is narrower and the IMP is also concentrated into narrower beams. Hence, it is less likely that a user collides with the main beam of an IMP, and as seen in Table 6.1, the number of collisions is reduced.

6.3.2 Computer Generated Results

Due to the random positions of the users and the stochastic nature of the number of active users, the IMD power as experienced by a mobile user in the system will be a random variable. Hence, to study the improvement in carrier to intermodulation ratio by using a multibeam antenna, as discussed in the previous sections, the probability distribution function (pdf) of the IMD power at the mobile is estimated using simulations. A channel separation and modulation that resemble the GSM system is assumed. The variation of the shape and position of the pdf with the parameters n_t and M_f are discussed and conclusions is drawn.

Modulation and Sampling Rates

Assume that each sub-carrier is GMSK modulated with a bit rate T_b=271 kbit/s, together with a Gaussian low pass pulse-shaping filter with a normalized bandwidth WT_b=0.3 and filter length of six bits. A minimum sampling frequency of 50 sam-

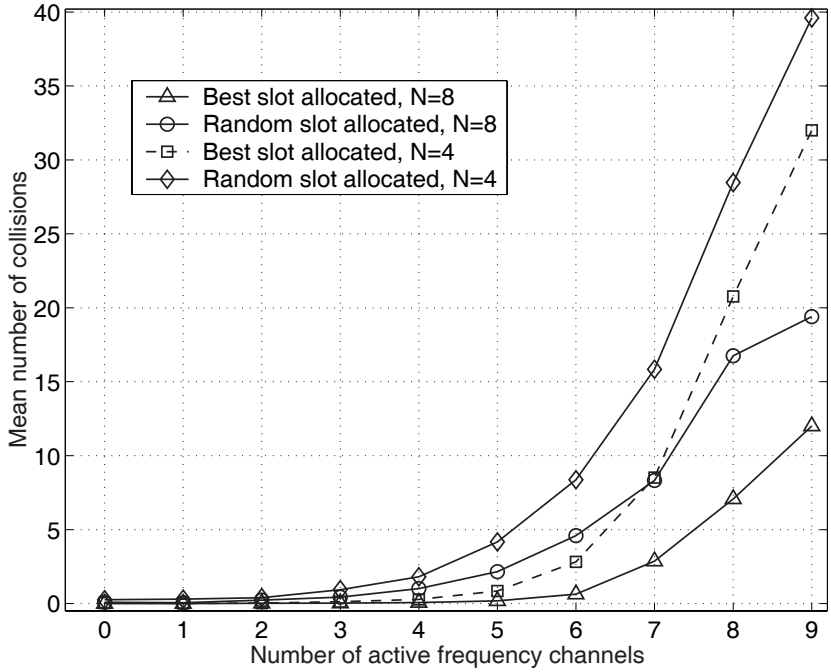


Figure 6.7: Estimated value of $E\{C_o|m\}$, the mean number of collisions given m active users using the random frequency channel allocation algorithm and the proposed algorithm for n_t=4,8 (denoted as N in the figure). The proposed algorithm reduces the number of collisions and hence, reduces the IMD power at the mobiles.

ples per symbol is used in the simulations for the sub-carrier with largest ω_m to ensure an accurate representation of the generated intermodulation distortion. Using M_f=9 channels, this yields a multichannel sampling rate of 67.75 MHz. The length of a data burst is 156 bits and consists of randomly generated data for each user.

Adjacent Channel Distortion Simulation

A common method to measure the effect of nonlinearities in multicarrier communication systems, is to use a measurement channel, centered in the frequency channel plan. This unoccupied center channel will, on the MCPA output contain IMP, and the effect of distorting the input signal can directly be related to the output power in the unoccupied channel [158]. To measure the distortion introduced by the nonlinearities in the MCPA and the effect of beamforming for the spatial distribution of this distortion, a simulation, illustrated in schematic form in Figure6.8 is per-

formed. The method measures the interference in an unoccupied center channel of the transmitted frequency band. Hence, the input frequency channels centered at $\omega_1, \omega_2, \ldots, \omega_{q-1}, \omega_{q+1}, \ldots, \omega_{M_f}$ are occupied, while the center channel at ω_q is unoccupied.

Since most communication systems are designed to meet a certain level of adjacent channel interference, the output power in this unoccupied channel can directly be related to the performance of the system. Note that the center channel contains contributions from all other channels and the power measured there is therefore the worst possible case of interference. The channel spacing Δ_ω is set to $3W$ to assure that the spectral truncation noise, due to finite impulse response filtering, of the channels immediately below and above the unoccupied channel does not increase the spectral content in the unoccupied channel.

The $L_S = 156 \times 50 \times 5 = 39000$ samples of the received signal in direction Θ is collected and a Hanning window function is applied before the discrete Fourier transform (DFT) is calculated. The signal is filtered in an ideal 30 kHz bandpass filter centered at ω_e and the power level in the unoccupied channel is measured (this is equivalent to the procedure described in the GSM 05.05 specification [158]). The intermodulation power is obtained using power spectrum analysis, define the Hanning windowed DFT of $z(n, \Theta)$ as [180]:

$$Z(p, \Theta) = \frac{1}{L_S} \sum_{n=0}^{L_S-1} z(n, \Theta) w_H(n) e^{j2\pi p(l/L)} \tag{6.50}$$

for $p = 0, \ldots, L_S - 1$ and $w_H(n)$ is the Hanning window function. The power in the measurement bandwidth in direction Θ can now be written as

$$P_r(\Theta) = 10 \cdot \log_{10} \left[\frac{1}{p_2 - p_1 + 1} \sum_{p=p_1}^{p_2} |Z(p, \Theta)|^2 \right] \tag{6.51}$$

and p_1, p_2 are chosen to obtained an ideal bandpass filter of 30 kHz bandwidth, centered at ω_q.

Cann's model, (6.20) was used to model the power amplifier with parameters $s = 8$, $D = 1$, $l = 1$ and the IBO was set to 10 dB, regardless of M_f, as described in Section 6.2. Figure 6.9 shows an example of the radiation pattern (6.51) in the empty frequency channel, centered at ω_q. With the number of active carriers M_f=8, a higher level of IM interference is measured, compared to when M_f=4, as expected. The number of third-order IMP terms can be calculated by (6.44) to $\nu_c^e = 20$ and 4 for the $M_f = 8$ and $M_f = 4$ case respectively[4]. When the variability υ in (6.46) is decreased, the IMP radiation pattern gets smoothed and

[4]Note that M_f=9 and M_f=5 is used in (6.44), but the center channel is left empty

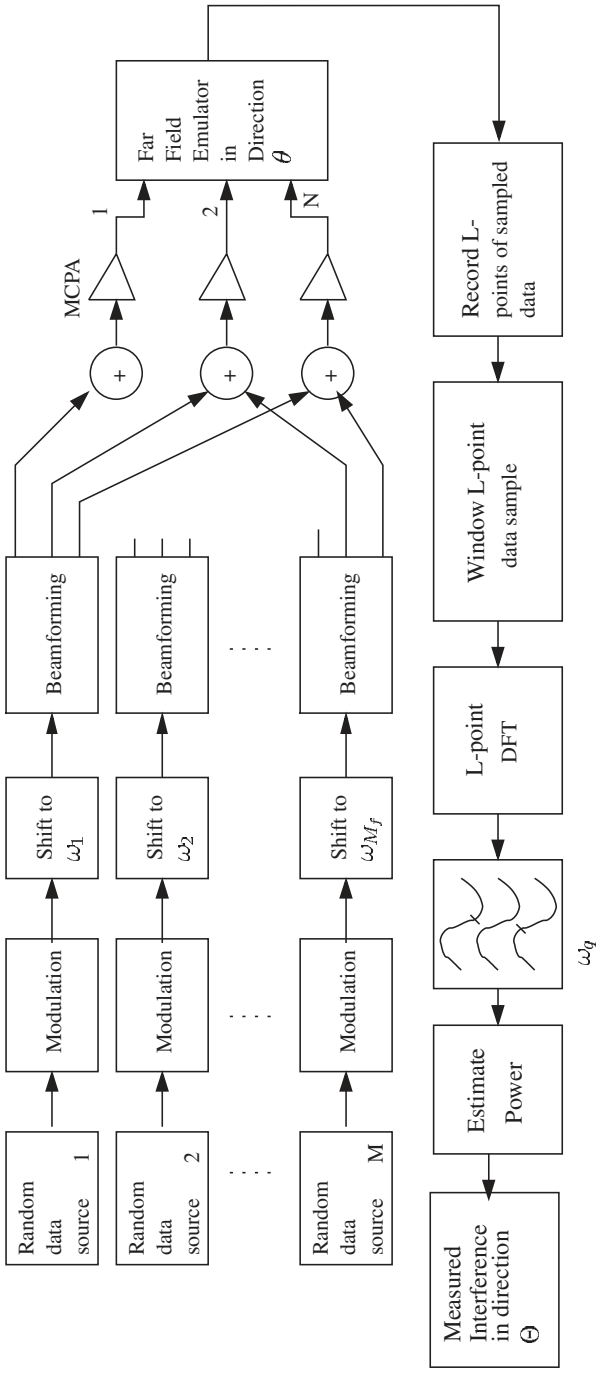


Figure 6.8: Set up for simulation to evaluate distortion in unoccupied channel in direction Θ

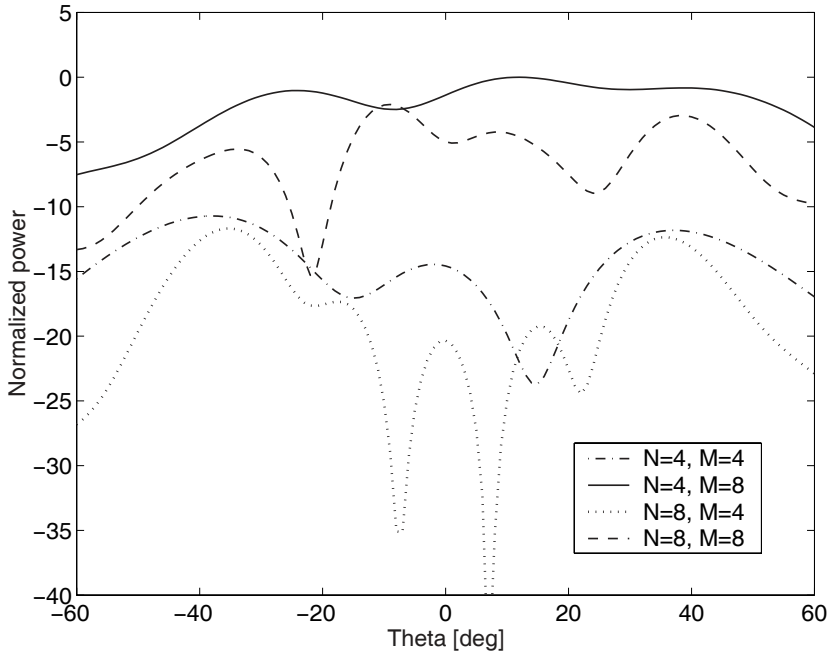


Figure 6.9: Simulated intermodulation radiation pattern using the power amplifier nonlinearity. When the number of beams n_t (denoted N in the figure) is decreased and the number of active frequency channels M_f (denoted M in the figure) is maintained constant, the radiation pattern approaches the single element antenna radiation pattern.

approaches the single element pattern. This is visible in the $n_t = 4$, $M_f = 8$ case in Figure 6.9.

If the received power at the test mobile when using $n_t = 1$ antennas at the basestation is used as a reference, then the pdf of the difference (in dB) between the $n_t = 1$ and $n_t > 1$ case for a particular set of mobile locations is estimated using 200 data bursts. Random frequency allocation is performed and no power control is used. See Figure 6.10 and 6.11 for the $M_f = 5$ and $M_f = 9$ case respectively (where the measurement is made in the empty center channel, as before). Note that the IBO is equal for the different antenna configurations, hence an increase in number of antennas, leads to a reduced output power of each MCPA, but the IBO is kept equal in the simulations, to allow an analysis of the spatial dispersion effect of the IMP only. Table 6.2 shows the mean of the pdfs in Figure 6.10 and Figure 6.11. According to the discussion in section 6.3.1, a reduction in the IM power of $10\log_{10}(n_t)$ dB is expected. For the M_f=9 case, the number of IMP is so large that approximately $1/n_t$ of them reach the mobile at maximum power, that is, in

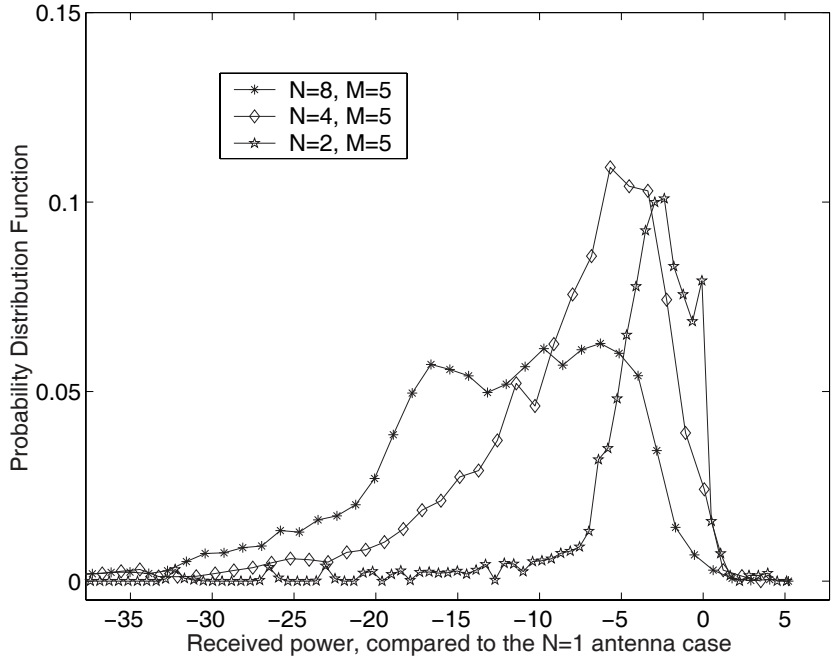


Figure 6.10: Probability distribution function of the difference in received power in the empty center channel for n_t (denoted N in the figure) antennas, for all angles in the sector. n_t=1 antenna case is set as a reference. Here, M_f=5 (denoted M in the figure) channels are used.

Antennas n_t	Channels M_f	Average [dB]
8	9	-9.2
8	5	-13.5
4	9	-6.1
4	5	-8.9
2	9	-2.9
2	5	-4.1

Table 6.2: Average of the received power in the empty measurement channel. Reference is the n_t=1 antenna case.

the main-lobe of the array antenna radiation pattern. However, the $M_f = 5$ case allows larger variations in the distribution of the IMPs over the n_t beams, i.e. a larger variability, and a deviation from the $10\log_{10}(n_t)$ expression is increased. Note that in the $n_t = 8, M_f = 5$ case, the probability that a mobile is not in

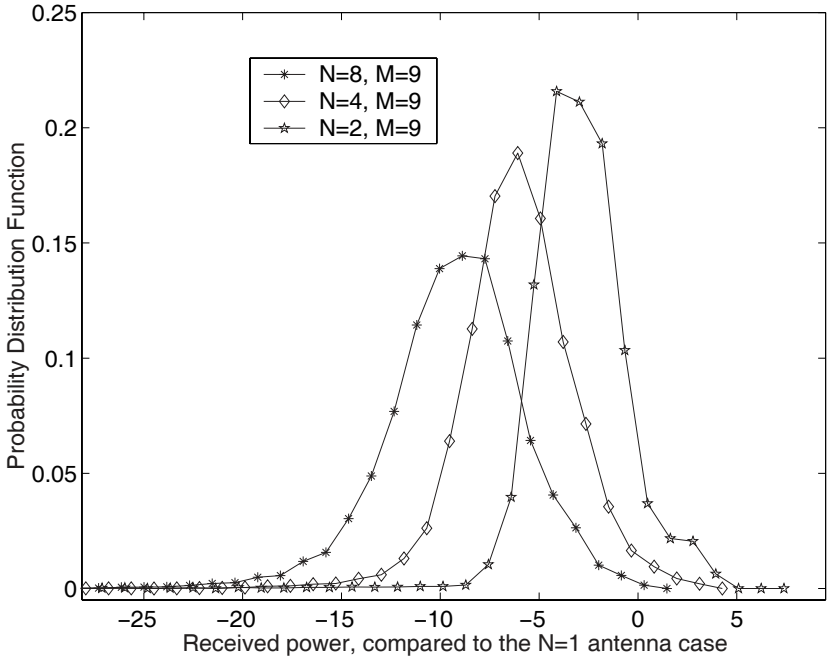


Figure 6.11: Probability distribution function of the difference in received power in the empty center channel for n_t (N) antennas, for all angles in the sector. n_t=1 antenna case is set as a reference. Here, M_f=9 (M) channels are used.

the direction of an IMP main beam in the empty center frequency channel is from Theorem 6.1, $P_\nu(0) = \left(\frac{8-1}{8}\right)^{\nu_c^e}$=0.59 where ν_c^e=4 is the total number of IMP in the center channel given by (6.44). This explains the relatively high probability for a very large reduction (5-20 dB) in IMP power for $n_t = 8$ in Figure 6.10, as compared to the $n_t = 2$ case in the same figure, where $P_\nu(0)$=0.0625.

The $M_f = 9$ case in Figure 6.11 shows the case of a large number of third order IMP. As the number of beams n_t is increased, the mean of the IMP is decreased but the width of the pdf increases (increased variability), due to larger variations in the number of IMP per beam. Furthermore, the probability that a mobile receives no main beam IMP $P_\nu(0)$ is in this case very small.

6.3.3 Conclusions

The effect of a nonlinear transmit amplifiers in a switched beam base-station antenna for a cellular system using FDMA has been analyzed. It has been found that the main beam direction of the IMP is in general different from the directions of the principal signal beam directions. It was shown how the increased number of

degrees of freedom when introducing an array antenna could be utilized to reduce the received IMD even further, for the mobile users. This is achieved by assigning new users to frequency channels that minimizes the IMD at other mobile users in the same sector, hence the IMP are "placed" in directions were no mobile user exists on that frequency. The proposed algorithm outperforms a random channel allocation method and the improvement is increasing with the traffic load and the number of antenna elements.

To study the expected value of the IMD reduction at the mobiles, a Monte Carlo simulation was used to estimate the pdf of the mobile CIR for an example GSM system. With increased number of antenna elements and a low number of active channels, the IMD reduction as compared to the n_t=1 reference antenna case has a larger dynamic range. In this case, it is possible, especially if a smart frequency allocation strategy is used, that the mobile user experiences no main beam IMD, hence a large IMD reduction compared to the reference case is gained. It was also shown that when the number of active users is larger than the number of available beams (i.e. the number of transmit antennas), the pdf is more sharply peaked and $10\log_{10}(n_t)$ dB reduction in IMD is the most likely improvement.

6.4 Intermodulation distortion in a SISO case

In this section IMD in a SISO channel with log-normal fading is investigated. In a wireless LAN (WLAN) systems for indoor office environments it is common to use a single BS and there are no other BS in the vicinity that interfere with the MS, so the only interference comes from IMD generated in the own BS. Due to the near-far ratio of two users connected to a WLAN access point[5] (AP), the IMD generated by the AP might block a near transmitter if the AP is simultaneously transmitting with high power to a MS located on the fringe of the coverage area. This blocking will manifests itself as a reduction in system capacity from the ideal case.

Often, in a WLAN environment, there is a shadow fading due to obstacles as office furniture and walls that fades the signals on a large scale. It is often modelled as a log-normal distribution around the mean given by a path loss equation. The integration over all possible MS locations has a closed form solution as is shown in this section. The problem was earlier addressed in [181] where a closed form expression was derived that characterize the relationship between the blocking probability and linearity requirements of the power amplifier in a simple SISO channel with an inverse power law attenuation of the signal power. Here the corresponding closed form expression for the blocking probability in the log-normal fading case is presented which in [181] was found only by using a Monte Carlo simulation. Hence, the novelty in this section is the closed form expression for this outage probability. Monte Carlo simulations are used to verify the analytical result.

Analysis

The assumptions are similar to the assumptions in [181] where an AP is the receiver (RX) and the desired transmitter TX_d and an interfering transmitter TX_i are situated on the distance r_d and r_i from the AP respectively, see Figure 6.12. The maximum range for the basestation coverage is R. Some more assumptions are made to make the problem analytically tractable: Assume that the transmitters are transmitting with the same power and without power control. The transmitter positions are independent and uniformly area distributed over a disc with radius R from the AP. When the desired signal power to IMD power ratio is below a certain threshold Γ, the desired transmitter is completely blocked.

If a simple power law equation with decay index p is assumed, then the expec-

[5] In WLAN systems, the basestation is commonly denoted the *access point*.

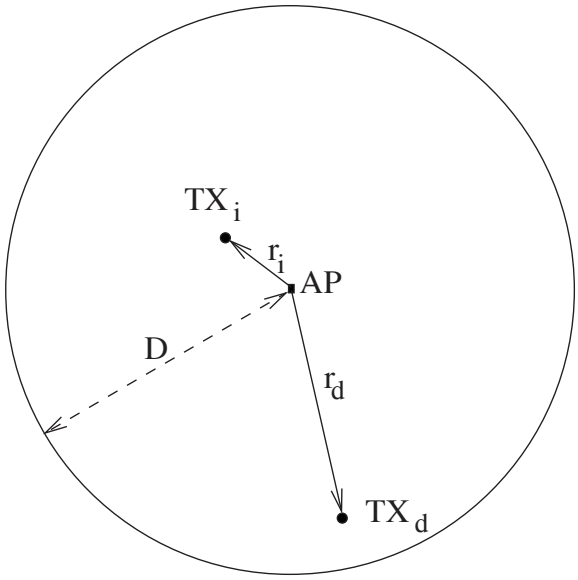


Figure 6.12: The near far scenario where IMD from TX_i is interfering the signal from TX_d at the AP. R is the maximum range of the AP.

tation of the received desired and IMD power at the AP is

$$\begin{aligned} m_d &= E\{p_d\} = 1/r_d^p \\ m_i &= E\{p_i\} = P_{IMD}/r_i^p \end{aligned} \tag{6.52}$$

where $10\log_{10}(P_{IMD})$ dBc is the IMD product level transmitted from TX_i and the transmitted power from TX_d has been normalized to $P_T = 1$ without loss of generality.

At the AP, the blocking probability is defined as

$$Pr(blocking) \stackrel{\triangle}{=} Pr\{p_d/p_i < \Gamma\} \tag{6.53}$$

where p_n and p_i are the received desired and IMD power respectively averaged over the eventual fast fading. The specified threshold Γ depends on the tolerance of the modulation to interference. The received power depends on the distances r_d and r_i, the path loss and on the probability distribution of the fading. Using Bayes theorem, rewrite the blocking probability (6.53) as

$$Pr\{p_d/p_i < \Gamma\} = (2\pi)^2 \int_0^R \int_0^R Pr\{p_d/p_i < \Gamma | r_d, r_i\}\, p(r_d)p(r_i)dr_d dr_i \tag{6.54}$$

where the factor $(2\pi)^2$ is due to the integrated uniform angle distribution and independence of the TX_d and TX_i positions are assumed. The marginal distribution of the radial position in an area uniform pdf is

$$p(r) = \frac{2r}{R^2} \ . \tag{6.55}$$

The blocking probability conditioned on the area mean power m_d, m_i for a log-normal shadowed channel with equal standard deviations for the desired and the IMD signal fading statistics, $\sigma_d = \sigma_i = \sigma$, can be written as [182]:

$$Pr\left\{p_d/p_i < \Gamma | m_d, m_i\right\} = Q\left(\frac{\ln\left(\frac{m_d}{\Gamma m_i}\right)}{\sqrt{2}\sigma}\right) \ . \tag{6.56}$$

The blocking probability is now obtained using (6.52),(6.55),(6.56) in (6.54) as

$$Pr\left\{p_d/p_i < \Gamma\right\} = \frac{16\pi^2}{R^4}\int_0^R\int_0^R Q\left(\frac{1}{\sqrt{2}\sigma}\ln\left\{\left(\frac{r_i}{r_d}\right)^p / P_{IMD}\Gamma\right\}\right) r_i r_d dr_i dr_d \tag{6.57}$$

Equation (6.57) reduces after integration to

$$\begin{aligned} &Pr\left\{p_d/p_i < \Gamma\right\} = \\ =&Q\left(c\right) + \frac{1}{2}e^{\frac{2\sigma^2}{p^2}}\left[(P_{IMD}\Gamma)^{2/p}Q\left(\frac{2\sigma}{p} - c\right) - \frac{1}{(P_{IMD}\Gamma)^{2/p}}Q\left(\frac{2\sigma}{p} + c\right)\right] \end{aligned} \tag{6.58}$$

where $c = -\ln(P_{IMD}\Gamma)/\sigma$. These integrations are carried out in Appendix 6.A. Note that this expression is independent of the maximum range R. In the limit $\sigma \to 0$, i.e. the non-fading case, the expression (6.58) approaches the blocking probability in [181] for the simple power law decay of the signals, if $P_{IMD} < 1/\Gamma$ which is true for practical IMD levels, is assumed :

$$\lim_{\sigma\to 0} Pr\left\{p_d/p_i < \Gamma\right\} = \frac{1}{2}(P_{IMD}\Gamma)^{2/p} \tag{6.59}$$

The desired to interference level tolerance of the modulation is set to Γ=15 dBc and plot the blocking probability as a function of the IMD level P_{IMD}, the family of curves in Figure 6.13 is obtained for different values of the log-normal pdf parameter σ. The path loss index was set to $p = 4$. To verify the expression (6.58), a Monte Carlo simulation was performed by randomly placing terminals around the AP and selecting a shadowing component from a log-normal distribution. Then equation (6.53) is evaluated to determine whether blocking has occurred. The correspondence between the theory and simulated results are satisfactory.

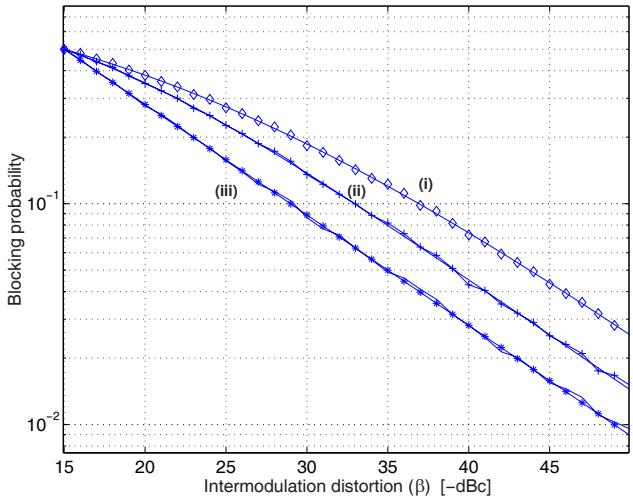


Figure 6.13: Theoretical (solid) and results from Monte Carlo simulation (marked) of channel blocking probability with respect to linearity of the transmit power amplifier. Different cases of log-normal fading (i) σ=9 dB, (ii) σ=6 dB, (iii) σ=$-\infty$ dB. (no fading)

In Figure 6.13 the curve for the non-fading case is displayed, and it shows the lowest probability of blocking for a given IMD level. Hence, the shadow fading increases the blocking probability. When the IMD level equals the IMD tolerance ($P_{IMD}\Gamma$=1)the blocking probability is 0.5, because blocking will the occur when TX_i is closer to the AP than TX_d and this occurs at equal probabilities $Pr(r_d < r_i) = Pr(r_i < r_d) = 0.5$.

It can be seen from Figure 6.13 that for a blocking probability of 10%, Γ is -37 dBc, -33 dBc with a shadow fading standard deviation of σ=9 dB and 6 dB respectively.

6.5 Conclusions

The aim of this chapter was to investigate the impact of using MCPA:s at the basestation for the downlink in a wireless system. The derivations was simplified by assuming a switched beam configuration, often implemented using a beamforming network, such as the Butler matrix. This technique was shown to have the interesting property that the IMD follows the same radiation patterns as the original desired signals patterns. This allowed for the concept of a beam-frequency scheme and a frequency channel allocation algorithm for reducing the harmful IMD in that particular cell was developed. A basestation with more antenna elements gave a reduction of intermodulation distortion which could be of the order of several dB.

The one-cell study concluded that in the many user case, the reduction in IMD is on average n_t-fold, where n_t is the number of transmit antennas. The distribution function of the power of the transmitted IMD was calculated using a simulation of a system resembling the GSM standard. It was seen that the variability decreased as the number of co-amplified signals increased.

A WLAN scenario was also investigated and it was found that the outage (or *blocking*) probability integral in the log-normal fading SISO case had a closed form analytical solution. This integral had been solved by previous authors [181] using Monte Carlo integration techniques but in this chapter a closed form expression is presented, for the first time. The result can be applied to IMD studies in WLAN system where multicarrier modulation such as OFDM is used. By varying the log-normal fading parameter, it could be seen that shadow fading increased the blocking probability. So an open space office has a lower blocking probability from intermodulation interference than an WLAN environment with blocked propagation paths.

Appendix 6.A The solution to the integral in (6.57)

To solve the integral

$$\int_0^R \int_0^R Q\left(\frac{1}{\sqrt{2}\sigma}\ln\left\{\left(\frac{r_i}{r_d}\right)^p / P_{IMD}\Gamma\right\}\right) r_i r_d dr_i dr_d \tag{6.60}$$

we start with the inner integral and define $A = p/(\sqrt{2}\sigma)$, $r_i = r$ and $c = -p\ln(r_d)/(\sqrt{2}\sigma) - \ln(P_{IMD}\Gamma)$ and rewrite it as

$$\int_0^R Q(A\ln(r) + c)dr \ . \tag{6.61}$$

Now, make the variable substitution $z = A\ln(r) + c$, which yields the equivalent integral

$$\frac{e^{-c/A}}{A}\int_{-\infty}^{A\ln(R)+c} Q(z)e^{z/A}dz \ . \tag{6.62}$$

This integral is solved by partial integration as

$$\int_{-\infty}^{A\ln(R)+c} Q(z)e^{z/A}dz = \left[AQ(z)e^{z/A}\right]_{-\infty}^{A\ln(R)+c} - \int_{-\infty}^{A\ln(R)+c} A\frac{dQ(z)}{dz}e^{z/A}dz \ . \tag{6.63}$$

Hence, the derivative of the Q-function is required, which is easily found from its definition (2.20)

$$\frac{dQ(z)}{dz} = -\frac{1}{\sqrt{2\pi}}e^{-z^2/2} \ . \tag{6.64}$$

So, (6.63) can be written as

$$\begin{aligned}&\int_{-\infty}^{A\ln(R)+c} Q(z)e^{z/A}dz \\ &= AQ(A\ln(R)+c)e^{\ln(R)+c/A} + \frac{A}{\sqrt{2\pi}}\int_{-\infty}^{A\ln(R)+c} Ae^{z/A-z^2/2}dz \ .\end{aligned} \tag{6.65}$$

Hence, we need to solve a new integral which solution is found in [80, Eq. 2.33]

$$\int_{-\infty}^{A\ln(R)+c} Ae^{z/A-z^2/2}dz = \sqrt{\frac{\pi}{2}}e^{1/2A^2}\left[\text{erf}\left(\frac{A\ln(R)+c}{\sqrt{2}} + \frac{1}{\sqrt{2}A}\right) + 1\right] \ . \tag{6.66}$$

By utilizing the relations $\text{erf}(x) + \text{erfx}(x) = 1$ and $Q(x) = \frac{1}{2}\text{erfc}(x/\sqrt{2})$ [14], (6.66) is written as

$$\int_{-\infty}^{A\ln(R)+c} Ae^{z/A-z^2/2}dz = \sqrt{\frac{\pi}{2}}e^{1/2A^2}\left\{2 - 2Q\left(A\ln(R) + c + \frac{1}{A}\right)\right\} \ . \tag{6.67}$$

So, the integral (6.61) can, if (6.67),(6.65) is used in (6.62) be written as

$$\begin{aligned}&\int_0^R Q(A\ln(r)+c)dr \\ &= RQ(A\ln(R)+c)+e^{-c/A-1/2A^2}\left(1-Q\left(A\ln(R)+c+\frac{1}{A}\right)\right) \quad .\end{aligned} \tag{6.68}$$

So we see that the inner integral in (6.60) consists of Q-functions as well. Now, make the substitutions $B=-p/\sqrt{2}\sigma$, $d_1=A\ln(R)-\ln(P_{IMD}\Gamma)$, $E=e^{-c/A-1/2A^2}$, $r_d=r$ and $d_2=d_1+A^{-1}$. Then (6.60) can be written as

$$\int_0^R RQ(B\ln(r)+d_1)+E(1-Q(B\ln(r)+d_2))dr \tag{6.69}$$

which is solved by applying the derived formula (6.68) once more. The calculations are straightforward but tedious and the result is finally given in (6.57).

Chapter 7

Switched Parasitic Antenna Applications for Wireless Systems

A novel techniqe to introduce pattern diversity is studied in this chapter. A switched parasitic antenna (SPA) is used to electronically steer the radiation pattern of the receiving or transmitting antenna. The correlation between different beams are shown to be sufficiently low to yield a diversity gain and the capacity is shown to be comparable with the antenna array in certain situations. Furthermore, a space time block code is used to evaluate the BER of a MIMO system with SPA. It is found that a SPA requires a 5 dB higher SNR than the array antenna to acheive a BER of 10^{-2} but the benefits are the lower costs for radio hardware since the SPA utilizes only one transceiver. The SPA is also shown to be useful in direction finding (DF) applications. The DF performance of a SPA is found by calculating the lower bound on the DF accuracy, i.e. the Cramer Rao lower Bound (CRB). It is found that the SPA offers a compact implementation with high resolution DF performance with only a single radio receiver and is thus an interesting low-cost alternative to traditional multiple antenna arrays.

7.1 Introduction

An unfortunate aspect of multiple antenna systems is the high cost of multiple radio transceivers at the access point and at the user terminal. There is always a need to reduce the size of the antenna array, to lower the fabrication costs but also the reduce the negative impact on the visual environment by improving the aestethics of the antenna. Furthermore, it is expensive to calibrate and maintain antenna arrays with many antenna elements. Recently, switched parasitic antennas

(SPAs) have been subject of an increased interest in the literature, for improving capacity in indoor LANs [183], as a diversity antenna [184, 185], for tracking of base-stations [186] and for high resolution direction of arrival estimation [109]. See also the recent book [187] for an extensive overview of SPA:s.

In a sense, SPA offers characteristics similar to an array antenna with several fixed beams, but is more compact in size, and might be more suitable on certain mobile equipments. The use of SPAs in MIMO systems will be examined in this chapter, where the SPA is primarily used at the mobile station (MS). It has been observed that when the fades of the MIMO receive channels are correlated, the channel capacity can be significantly smaller than when the fades are independent and a identically distributed (i.i.d.) [88]. Hence, the capacity of the SPA in this type of scenarios is studied and compared to a completely uncorrelated scenario. Furthermore, space time block coding (STBC) techniques are applied to examine how this correlation affects the BER of a SPA system.

7.2 Switched parasitic antennas

SPAs offering directional patterns dates back to the early work of Yagi and Uda in the 1930's. For mobile communications, Vaughan [185] gave some examples of this technology to provide angle diversity as two parasitics on a mobile phone handset. The concept is to use a single active antenna element, connected to a radio transceiver, in a structure with one or several *passive* antenna elements, operating near resonance. The passive elements are called parasitic elements (PE) and act together with the active element to form an array, as in the well known Yagi-Uda array [108]. To alter the radiation pattern, the termination impedances of the PEs are switchable, to change the current flowing in those elements. The PE become reflectors when shorted to the ground plane using *pin* diodes [188] and when not shorted, the PE have little effect on the antenna characteristics. The receiver is always connected to the center antenna element so there are no switches in the RF signal path.

The parasitic antennas can be designed using monopoles on a ground plane [185, 188] or as parasitic patch antennas [189]. In this chapter, the monopole on the ground plane is considered for its omnidirectional properties. Examples of parasitic antennas are shown in Figure 7.1 for a 4-direction symmetry and in Figure 7.2 for a 3-direction symmetric antenna. The antenna in Figure 7.2 have an additional circle of parasitic elements that always are shorted to ground. The effect of this arrangement is an increased directivity as their length are shorter than the corresponding resonant length ($\approx \lambda_c/4$) and will lead the induced emf [108].

The lengths and distances displayed in Figure 7.1 and 7.2 have not been opti-

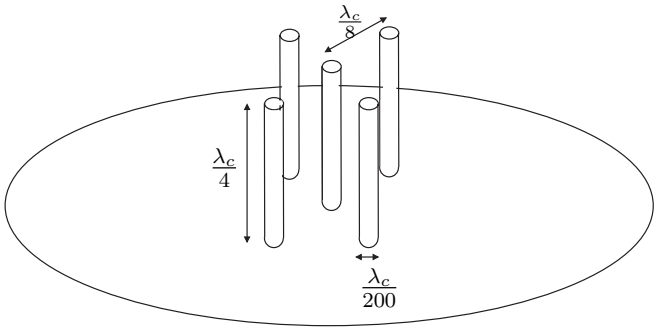


Figure 7.1: A five element monopole SPA. The center element is active and connected to the transceiver. The four passive antenna elements can be switched in or out of resonance using appropriately biased pin diodes.

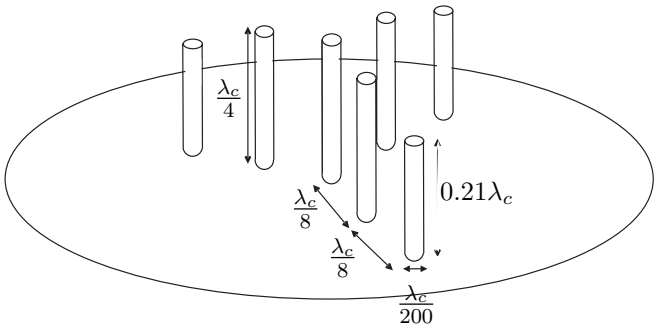


Figure 7.2: A seven element monopole SPA. The center element is active and connected to the transceiver. The three passive antenna elements closest to the active can be switched in or out of resonance using appropriately biased pin diodes. The three outermost monopoles are hardwired to ground.

mized. The lengths can be adjusted to give the antenna certain characteristics, such as directivity and/or dual band tuning, as demonstrated in [188], where a genetic algorithm approach was taken to optimize a six element switched beam antenna. If the parasitics are moved closer to the active element, then the mutual coupling increases and the change in the radiation pattern when switching is greater, however, the antenna impedance changes also more dramatically, which makes the antenna matching difficult. Hence, a too large mutual coupling renders in an inefficient antenna. The trade-off is thereby between compactness and high directivity on one hand and antenna efficiency on the other.

The antennas in Figure 7.1 and 7.2 was simulated using HFSS (High Frequency Structure Simulator) from Agilent Technologies Inc. which is a 3D simulator using the finite element method to solve for the electromagnetic field. The software was used to calculate the far-field radiation pattern of the antenna for different set-

tings of the switched parasitics. The monopole elements were cylindrical with an length to radius ratio $l/r = 100$ which have a first resonance at approximately $0.24\lambda_c$ [108]. The monopoles were assumed to be perfect conductors which forces the electrical field to be normal to the surface. Furthermore, to shorten the simulation time, the ground plane was assumed to be of infinite extent. The chosen frequency was 2.15 GHz, suitable for the downlink in UMTS FDD mode. Hence, the wavelength is λ_c=0.1395 meter.

The far-field power radiation pattern for three shorted parasitics and one open for the parasitic antenna in Figure 7.1 is shown in Figure 7.3 and the corresponding plot for Figure 7.2 is shown in Figure 7.4 for two shorted and one open parasitic monopole. The directivity of the two antennas are 9.9 dB and 10.0 dB respectively.

7.3 MIMO channel capacity and diversity gain

To achieve a high capacity in MIMO systems or a large diversity gain, the signals received by different settings of the parasitics, called the m_r modes, must have low correlation (ideally zero). Hence, define the correlation coefficient of the signal voltages received by two patterns as [190]

$$\rho_{12} = \int_0^{2\pi} \int_0^{\pi} S(\theta,\phi) E_1(\theta,\phi) \cdot E_2^*(\theta,\phi) \sin\theta d\theta d\phi \tag{7.1}$$

where the two far field patterns $E_1(\theta,\phi)$ and $E_2(\theta,\phi)$ are normalized as

$$\int_0^{2\pi} \int_0^{\pi} S(\theta,\phi) |E_i(\theta,\phi)|^2 \sin\theta d\theta d\phi = 1 \tag{7.2}$$

for $i = 1, 2$. Above, $S(\theta,\phi)$ is the pdf of the incident waves.

The correlation coefficients for the complex voltage patterns corresponding to the power radiation patterns in Figure 7.3 and 7.4 are calculated assuming the Clark scenario [191] with pdf $S(\theta,\phi) = \delta(\theta - \pi/2)/(2\pi \sin\theta)$ to model a ring of dense sources on a horizon about the receiving antenna. Assume that three (two) parasitics are always shorted, to get four (three) different directions with 90° (120°) separation. This gives the correlation coefficient for adjacent patterns, or modes, for the parasitic antenna in Figure 7.3 as $|\rho_{i,i+1}|^2 = 0.1157$ and for opposite patterns $|\rho_{i,i+2}|^2 = 0.0120$. For the parasitic antenna in Figure 7.4, $|\rho_{i,i+1}|^2 = 0.1002$ is obtained. The envelope correlation coefficients for the signals received from the modes are taken as $\rho_{eij} \approx |\rho_{ij}|^2$ [192]. A well known rule of thumb is that "uncorrelated" signals in diversity branches corresponds to an envelope correlation lower than 0.5. Hence, the SPAs presented here would achieve a diversity gain in Clarke's scenario.

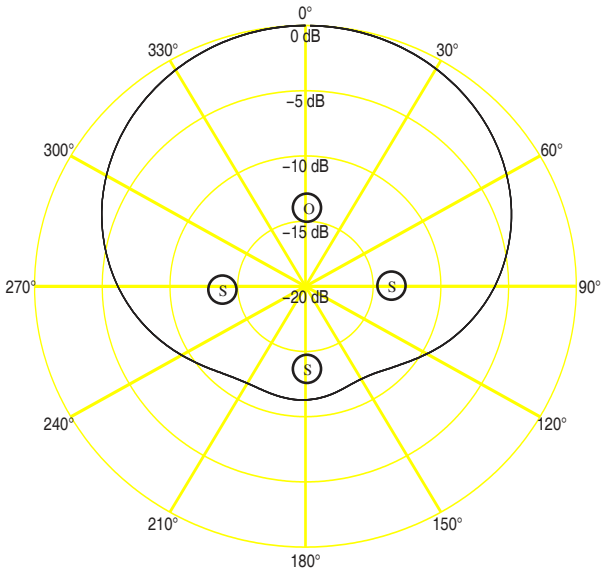


Figure 7.3: Power radiation pattern of the five element monopole antenna shown in Figure 7.1 with three parasitics shorted (S) to ground and one open (O).

7.3.1 Channel capacity

The MIMO channel capacity is calculated using the SPA at the base station (BS) and/or at the mobile station (MS). A single user, point to point communication over a flat-fading channel with m_t transmit modes (or array antenna elements) and m_r receive modes (or array antenna elements) and no co-channel interference is considered[1]. Assume that the block fading model, discussed in Section 1.1.2 is valid. Then the channel output corresponding to an input block spanning N symbol times is, from (1.22),

$$\mathbf{Y} = \mathbf{HC} + \mathbf{V} \ , \tag{7.3}$$

where the received signal $\mathbf{Y}$ is $m_r \times N$, the fading channel $\mathbf{H}$ is $m_r \times m_t$, the codeword matrix $\mathbf{C}$ is $m_t \times N$ and the receiver noise $\mathbf{V}$ is a $m_r \times N$ matrix. The entries of the noise matrix are i.i.d. complex Gaussian with zero mean. Recall that it is assumed that the channel is quasi-static, i.e. constant over the block of length N symbols.

With n_p parasitic elements, there are 2^{n_p} different modes, or settings of the switchable diodes. The (m_r or m_t) transmit and receive modes of the parasitic antennas are chosen among these to minimize the envelope correlation. Here, it is

[1]The number of antenna modes is denoted m_t as opposed to the number of antennas n_t in previous chapters since the definition of "antenna elements" does not hold on SPA:s.

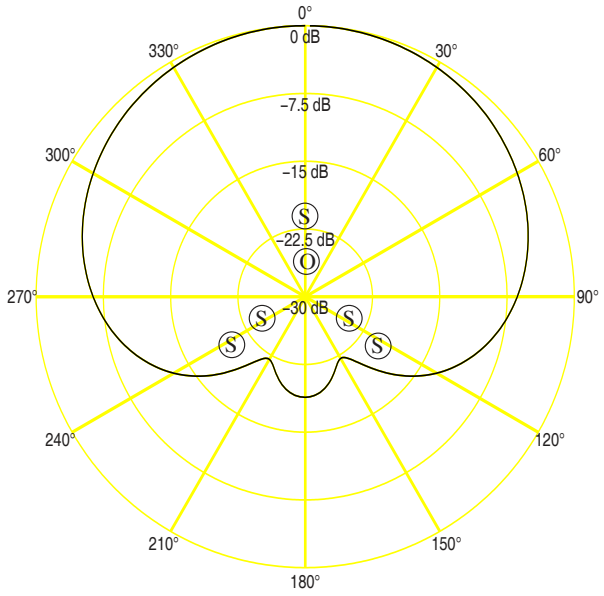


Figure 7.4: Power radiation pattern of the seven element monopole antenna shown in Figure 7.2 with two parasitics shorted (S) to ground and one open (O) plus three hardwired to ground (S).

assumed that the receiver switches through and samples the chosen modes during one symbol interval. This is possible in many systems, since the switching time for a pin diode is on the order of a few nanoseconds. The technique of oversampling the received signal is common in many communication systems, but here the oversampling is performed in both space and time, i.e. spatio-temporal oversampling. If the increased sampling rate (or bandwidth) poses a problem, then a bandpass sampling strategy could also be employed. In this chapter, the potential in using the different radiation patterns of a SPA for diversity and capacity enhancement is examined. It is important to note that further work is needed on the practical aspects of the antenna design as well as sampling strategies.

As the underlying concept here is angle diversity, a simple, yet detailed channel model is used that includes the spatial dimension. It was defined in Section 3.3.1 as the scattering disc model. Different channel scenarios is easily be obtained by varying the radius of the scattering disc as well as the location of the BS relative the disc [104]. The channel capacity are investigated in the following cases:

- BS with antenna array and MS with antenna array
- BS with antenna array and MS with SPA

- BS with SPA and MS with SPA.

When the array is used, omnidirectional antenna elements are assumed with $\delta_{BS} = 2\lambda_c$ spacing at the BS and $\delta_{MS} = 0.5\lambda_c$ spacing at the MS.

Shannon's capacity formula for a m_t input, m_r output MIMO channel, assuming equal power radiated from each transmitting antenna/mode and CSI unknown at the transmitter was given by (2.58) and is repeated here

$$C = E_{\mathbf{H}} \left[\log_2 \det \left\{ \mathbf{I}_m + \frac{P_T}{\sigma_n^2 m_t} \mathbf{H}\mathbf{H}^* \right\} \right] \tag{7.4}$$

where P_T/σ_n^2 is the signal to noise ratio (SNR) for each receive antenna/mode. The channel matrix $\mathbf{H}$ in (7.4) is normalized in the sense that $\|\mathbf{H}\|_F = \sqrt{\min(m_r, m_t)}$, where F denotes the Frobenius norm.

Since the channel matrix $\mathbf{H}$ is a function of the random position and reflection coefficients of the scatterers, the channel $\mathbf{H}$ is a stochastic matrix and thus the capacity C in (7.4) is obtained as the expectation value over the mutual information. With the SNR set to 4 dB, a Monte Carlo simulation with 4000 trials was performed to estimate the Complementary Cumulative Distribution Function (CCDF) of the capacity C for a scenario with $L = 20$ scatterers. The calculations are performed for an equal number of modes at the transmitter and the receiver (m_t=m_r=4). The distance between the BS and MS is $D = 50\lambda_c$. The CCDF of the capacity is shown in Figure 7.5 for the scattering disc radius $R = 50\lambda_c$. The curves are compared to the i.i.d. channel matrix $\mathbf{H}_w$ according to Definition 3.2 on page 81. Using the array antenna at the MS results in a slightly higher capacity than the array-parasitic configuration, however at the expense of more hardware due to the use of four transceivers instead of one. If the SPA is used at the BS, then the capacity is further decreased, as the signals from the modes becomes correlated due to the small angular spread as seen at the BS.

The outage capacity ${C_{P_o}}^2$ at $P_o = 10\%$ outage probability is presented in Figure 7.6 and 7.7 for the two types of antenna configurations respectively. A large disc correspond to an indoor scenario, where both BS and MS are surrounded by scatterers. The other extreme, with a small scattering disc centered at the MS, as in an outdoor to indoor channel, results in a smaller capacity, due to the reduced angular spread, and hence, lower diversity advantage. This can especially be seen in the case where the BS and the MS both are equipped with parasitic antennas, the capacity increases when the radius of the scattering disc exceeds the BS to MS distance, where full angular diversity also is possible at the BS.

At high bit rates, it might not be possible to switch through several modes during a symbol interval. Therefore a system with only two modes was investigated.

[2] See the definition of outage capacity in Definition 2.2 on page 36.

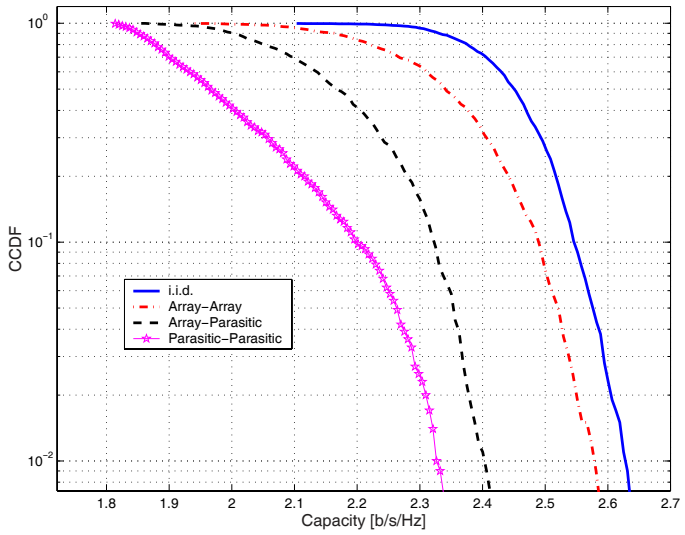


Figure 7.5: The complementary cumulative distribution function of the MIMO channel capacity for the m_t=m_r=4 case. The SNR is 4 dB and the scattering disc radius is 50λ_c. The parasitic antenna is shown in Figure 7.1.

The parasitic antenna in Figure 7.1 was used, but only two modes with lowest signal correlation (opposite in direction) was utilized. The outage capacity at 10% outage probability is shown in Figure 7.8. Comparing Figure 7.6 and 7.8, it is observed that the overall capacity is lower, but the difference between the array and the SPA is slightly smaller at scattering disc radius around 10λ_c.

7.4 Evaluating BER using STBC

In this section, an orthogonal space time block coding (STBC) is used, see Section 2.3.2, to exploit the available channel capacity discussed in the previous section. Assume that the transmitter have no knowledge of the CSI. For complex symbol constellations, rate R=1 orthogonal STBC only exists for $N = 2$ [29], however for real constellations there exists orthogonal, delay optimal rate one codes for $m_t = 2, 4, 8$, as for example the $m_t = 4$ case [29]

$$\mathbf{C} = \sqrt{\frac{P_T}{4}} \begin{pmatrix} s(n) & -s(n+1) & -s(n+2) & -s(n+3) \\ s(n+1) & s(n) & s(n+3) & -s(n+2) \\ s(n+2) & -s(n+3) & s(n) & s(n+1) \\ s(n+3) & s(n+2) & -s(n+1) & s(n) \end{pmatrix} . \quad (7.5)$$

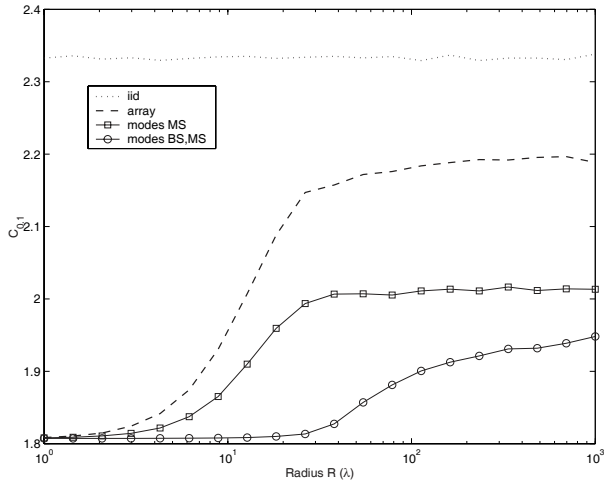


Figure 7.6: The outage capacity $C_{0.1}$ at 10% outage probability versus the radius of the scattering disc for the m_t=m_r=4 case. The parasitic antenna is shown in Figure 7.1. SNR=4 dB.

To study the performance in terms of BER using the SPA, the systems are simulated employing the STBCs (2.70) and (7.5), using the 4-direction SPA in Figure 7.3 and compare the results with the array antenna solution. When the Alamouti's STBC scheme is used, two opposite (180° separation direction) modes are utilized as the two receiving/transmitting modes for the SPA. The BER for these configurations is compared in Figure 7.9 where BPSK modulation is used in both cases, hence the bit rate 1 bit/s/Hz. It is observed that at the bit error rate of 10^{-2}, the 2-mode SPA gives 8 dB gain over a SISO and the 4-mode gives about 16 dB gain in SNR. The large and small scattering radius gives a difference in SNR gain less than 1 dB. The array-array configuration is about 5 dB better than the array-SPA configuration.

7.5 High-Resolution Direction Finding Using a SPA

Direction finding (DF) is of great importance in a variety of applications, such as radar, sonar, communications, and recently also personal locating services. In the last two decades, direction finding and sensor array processing has attracted considerable interest in the signal processing community. The focus of this work has been on high resolution, i.e. a resolution higher than the width of the main lobe, DOA estimation algorithms [193]. These algorithms exploit the fact that an electromagnetic wave that is received by an array of antenna elements reaches each element at different time instants. Although the performance of these systems is

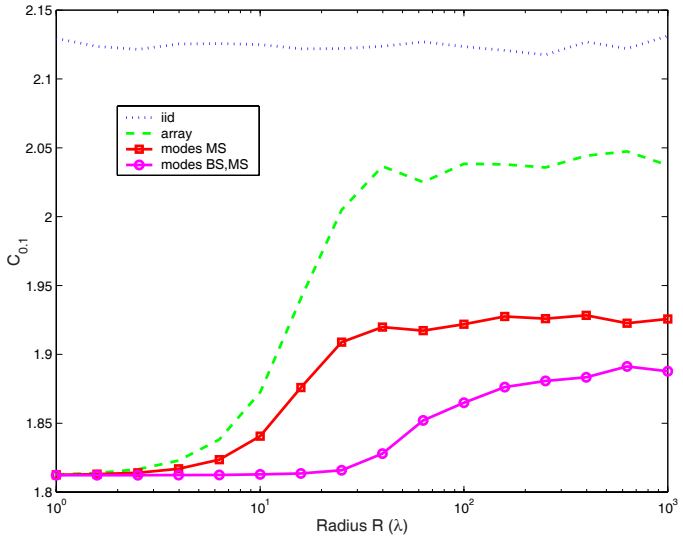


Figure 7.7: The outage capacity $C_{0.1}$ at 10% outage probability versus the radius of the scattering disc for the m_r=m_t=3 case. The parasitic antenna is shown in Figure 7.2.

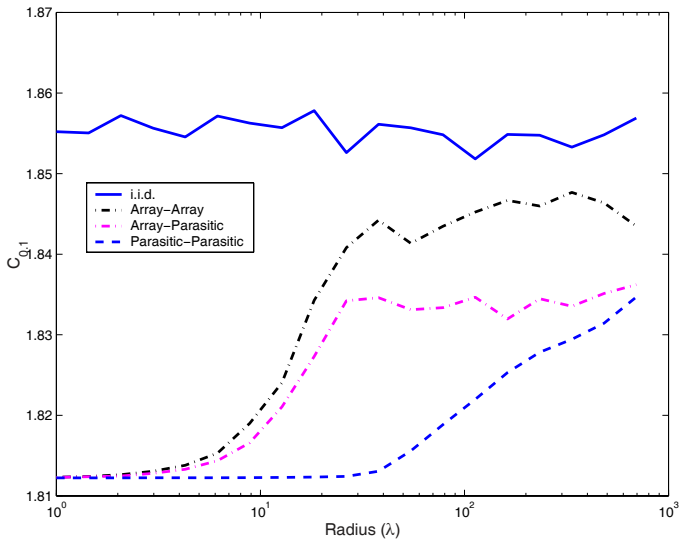


Figure 7.8: The outage capacity $C_{0.1}$ at 10% outage probability versus the radius of the scattering disc. Only m_r=m_t=2 of the 4 modes in the parasitic antenna in Figure 7.2 are used. SNR=4 dB.

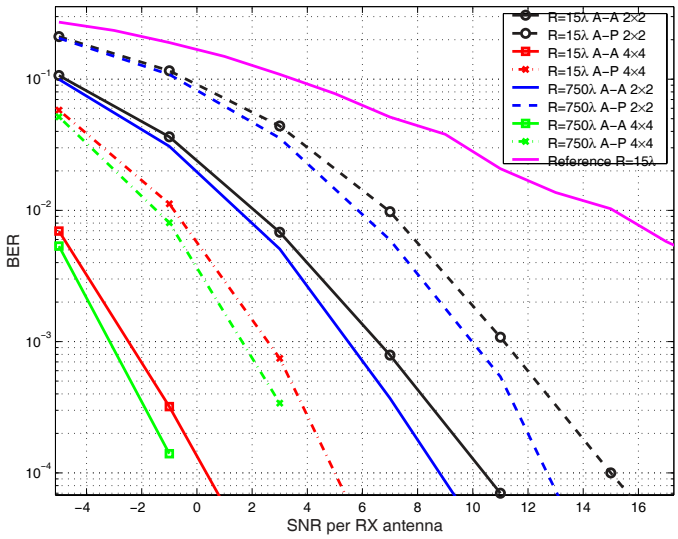


Figure 7.9: BER using STBC with $m_t = m_r = 2$ and $m_t = m_r = 4$ modes. The scattering disc radius is $750\lambda_c$ and $15\lambda_c$. The BER curves are compared to a SISO system (labelled reference).

excellent, an unfortunate aspect is the high costs of employing a radio receiver for each antenna element. Furthermore, it is expensive to calibrate and maintain antenna arrays with many antenna elements.

Recently, it was proposed to employ an SPA for direction finding [186, 189] that only uses a single active radio receiver, thereby significantly reducing the cost. The possibilities of exploiting the SPA for high-resolution DOA estimation will be examined in this section, since no attempt to employ high-resolution DOA methods was undertaken in [186, 189].

7.5.1 The SPA in a DF application

The far field radation patterns in the azimuth plane $F(\phi)$ is shown in Figure 7.3 and 7.4 for the 4-mode and 3-mode SPA respectively. Once these far-field radiation properties are found, it is straightforward to derive a model for the received voltages [194]. If p waves are incident upon an antenna with m_r symmetry directions, then the received voltages can be written in matrix form as

$$\mathbf{x}(t) = \mathbf{A}(\phi)\mathbf{s}(t) + \mathbf{v}(t), \tag{7.6}$$

where the vector of measured voltages $\mathbf{x}(t)$ is $m_r \times 1$. The matrix $\mathbf{A}(\phi)$ $(m_r \times p)$ corresponds to the response of the different symmetry directions and has elements

$[\mathbf{A}(\phi)]_{qp} = F(\phi_p + 2q\pi/m_r)$. This matrix is typically called the steering matrix in the sensor array processing literature. The signal vector $\mathbf{s}(t)$ is $p \times 1$ and contains the strength of the received fields. Finally, the noise vector $\mathbf{v}(t)$ is $n_r \times 1$.

In order for the analysis in the following sections to be valid, some additional assumptions are needed:

- The steering matrix has full rank, i.e., $\text{rk}(\mathbf{A}) = p$.
- $\mathbf{v}(t)$ is temporally white and circularly Gaussian distributed with zero mean and covariance matrix $\text{E}\{\mathbf{v}(t)\mathbf{v}(s)^*\} = \delta_{ts}\sigma_n^2\mathbf{I}$.
- $\mathbf{s}(t)$ is also temporally white and circularly Gaussian distributed with zero mean and covariance matrix $\mathbf{R_{ss}}$.

The noise is both spatially and temporally white, while the signal is only assumed to be temporally white. Furthermore, the signal is assumed to be uncorrelated with the noise.

7.5.2 Direction Finding Performance

The data model (7.6) is identical to the usual data model used in sensor array processing [193], except for a new steering matrix. This will of course change the direction finding properties. Before the properties of a specific DOA estimation scheme is studied, a lower bound, the Cramer-Rao Bound, on the variance of the DOA estimates will be analyzed. Note that it is possible to asymptotically achieve this bound with many methods in the literature [193].

Expressions for the CRB was derived for an array of antenna elements in [195]; and can also be applied to the parasitic antenna by changing the steering matrix. For any unbiased estimator $\hat{\phi}$, the following holds

$$E\{(\hat{\phi} - \phi_0)(\hat{\phi} - \phi_0)^T\} \geq \mathbf{B} \tag{7.7}$$

where $\mathbf{B}^{-1}$ is the Fisher information matrix which in the white noise case can be shown to be [195]

$$\mathbf{B} = \frac{\sigma_n^2}{2N_S}\left[Re\{(\mathbf{D}^*\mathbf{P}_A^{\perp}\mathbf{D}) \odot (\mathbf{R_{ss}}\mathbf{A}^*\mathbf{R}^{-1}\mathbf{A}\mathbf{R_{ss}})^T\}\right]^{-1}, \tag{7.8}$$

where the elements of $\mathbf{D}_{qp} = \frac{\partial F(\phi + 2q\pi/m_r)}{\partial \phi}\Big|_{\phi=\phi_p}$. Furthermore, $\odot$ denotes the Hadamard (or Schur) product, i.e., element-wise multiplication and $\mathbf{P}_A^{\perp} = \mathbf{I} - \mathbf{P}_A = \mathbf{I} - \mathbf{A}\mathbf{A}^{\dagger}$ [3] is the orthogonal projector onto the null space of $\mathbf{A}^*$. The matrix

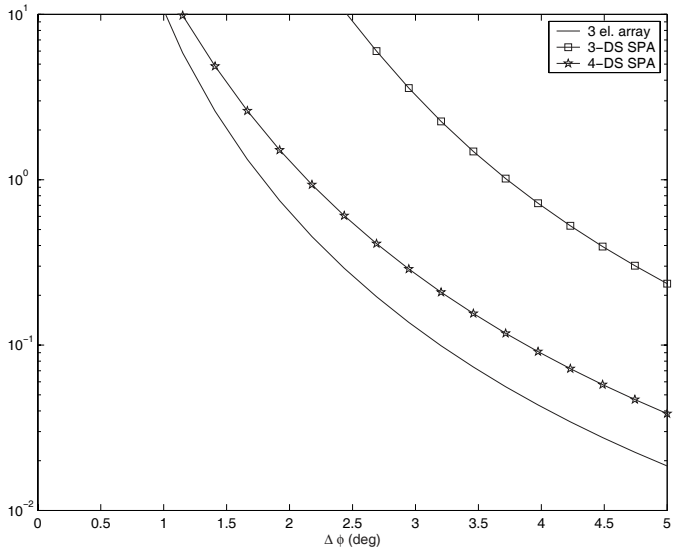


Figure 7.10: The square root of the CRB for the configurations in Figure 7.1 and 7.2 when two waves are incident from $(30^\circ, 30^\circ + \Delta)$ *with SNR=10dB and 1000 samples.*

$\mathbf{R} = \mathbf{A}\mathbf{R}_{\mathbf{ss}}\mathbf{A}^* + \sigma_n^2\mathbf{I}$ is the covariance matrix of the measured voltages $\mathbf{x}(t)$ and N_S denotes the number of time samples.

The square root of the CRB, i.e. the standard deviation, is shown in Figure 7.10 for the antenna configurations in Figure 7.1 and 7.2 as two waves are incident from $(30^\circ, 30^\circ + \Delta)$. Only the CRB for the first DOA, i.e. the wave arriving from 30°, is shown since the CRB for the second DOA will behave similarly. The standard deviation for a uniform linear array of three elements spaced $\lambda_c/2$ apart is compared to the 4-mode and 3-mode SPA:s. As expected, the performance is better when using four rather than the three symmetry directions. Also, note that the three element array performance slightly better the 4-mode SPA. However, these results depend on the incidence angles, since the array will work best for broadside and worst for end-fire incidence.

In Figure 7.11, the standard deviation is shown for the same antenna configurations as in Figure 7.10 when two waves are incident from $(\phi_0, \phi_0 + 5^\circ)$. The parasitic antenna, due to its symmetrical properties, offers similar direction finding performance properties for all incidence angles. The linear array performs worse than the parasitic antenna at end-fire incidence, while performing much better at

[3] $\mathbf{A}^\dagger$ is the Moore-Penrose pseudo inverse of $\mathbf{A}$.

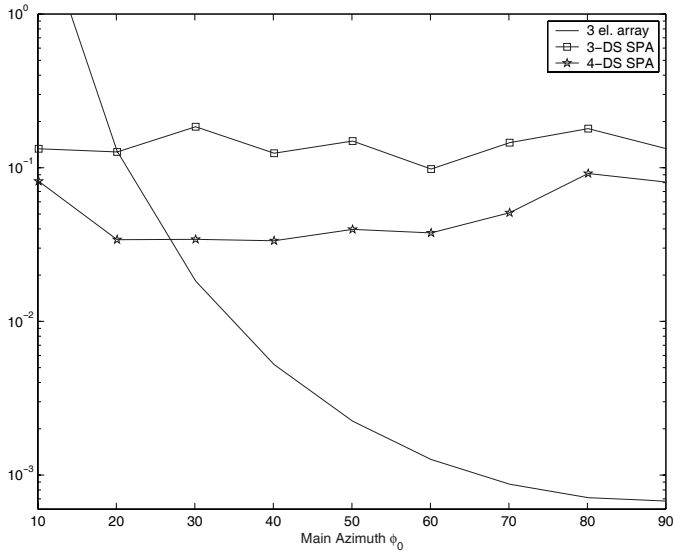


Figure 7.11: The square root of the CRB for the configurations in Figure 7.1 and 7.2 when two waves are incident from $(\phi_0, \phi_0 + 5^\circ)$ with SNR=10dB and 1000 samples.

broad-side incidence. However, for many direction finding applications, the direction finding performance of the parasitic antenna is sufficient and the cost reduction of using only a single radio receiver outweighs the loss in performance for broad-side angles. It should also be stressed that the antenna designs in Figure 7.1 and 7.2 are by no means optimal and better DOA properties may be obtained by a proper optimization.

7.5.3 Estimation Methods

The analysis in the previous section was based on the CRB on the estimation error. In this section, algorithms that approximately achieve this lower bound will be discussed. In principle, all DOA estimation schemes derived for a general antenna array can also be applied to a parasitic antenna by inserting a new steering matrix. For an overview of DOA estimation methods, see [193].

In [196], a popular high resolution DOA estimation method, MUSIC, was introduced where the DOA estimates are taken as those ϕ that maximizes the MUSIC criterion function

$$\hat{\phi} = \arg\max_{\phi} \frac{\mathbf{a}^*(\phi)\mathbf{a}(\phi)}{\mathbf{a}^*(\phi)\hat{\mathbf{E}}_n\hat{\mathbf{E}}_n^*\mathbf{a}(\phi)}, \tag{7.9}$$

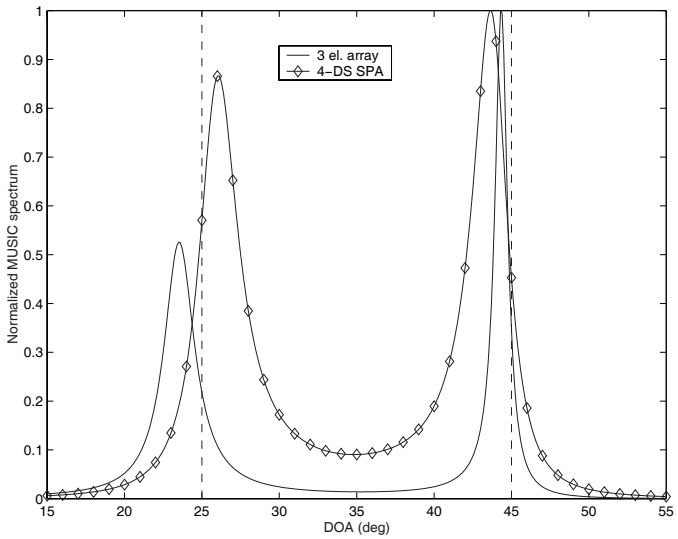


Figure 7.12: The normalized MUSIC spectrum when two waves are incident from 25° and 45° upon a 4-mode parasitic antenna and a three element array with SNR=10dB and 1000 samples.

where the steering vector $\mathbf{a}_q(\phi) = F(\phi + 2q\pi/m_r)$. Usually this is formulated as finding the p largest peaks in the "MUSIC spectrum". Here, $\hat{\mathbf{E}}_n$ denotes the $m_r - p$ eigenvectors corresponding to the $m_r - p$ smallest eigenvalues of the estimated covariance matrix $\hat{\mathbf{R}}$. A typical example of a MUSIC spectrum is shown in Figure 7.12, where two waves are incident from 25° and 45° upon a 4-mode SPA and a three element array with SNR=10dB and $N_S = 1000$ samples. This figure indicates that the SPA, in this case, offers a high-resolution direction finding performance similar to that of an antenna array without the cost of multiple radio receivers. Most other DOA estimation schemes [193] can also be applied to SPAs with similar results. For instance, the SML algorithm [195] for this type of antenna was implemented. The root mean square error of the maximum likelihood estimator achieved the CRB bound from Section 7.5.2, as expected.

7.6 Conclusions

A MIMO system using a switched parasitic antenna has been analyzed, in terms of capacity and BER, assuming orthogonal space time block codes. The channel capacity was simulated using a spatial channel model. It was found that the SPA offered capacities close to the capacities offered by an array antenna, in realistic

MIMO scattering environments. Since the SPA only requires one receiver, it could be an attractive low cost solution to future user terminals using space time coding to increase bit rates.

Furthermore, the potential use of an SPA for high-resolution direction finding was investigated. The SPA generates directional radiation patterns that can be used successfully to estimate DOAs. The direction finding performance was examined by calculating the CRB and the MUSIC estimator. It was found that the SPA offers a compact implementation with high-resolution direction finding performance using only a single radio receiver. Thus, exploiting SPAs for direction finding is an interesting alternative that offers several advantages over traditional arrays. Another advantage of the SPA is that a very compact implementation of the antenna is possible.

The examples of SPA:s presented here are only for demonstration purpose, the design of the SPA depends on the dimensions of the user equipment. Further work is needed on the practical aspects of the antenna design as well as sampling strategies.

Appendix A

The Butler Matrix

In this appendix, the theory of the Butler matrix is explained more thoroughly. The Butler matrix was developed as a multiple beam feed network for a phased array antenna. It is attributed to Butler [20], but it was independently discovered by Shelton [78]. It uses 3-dB directional couplers, or hybrid junctions along with fixed phase shifters to form n_r beams from an n_r element array where n_r is an integer expressed as some power of 2, that is $n_r = 2^p$. The 3-dB directional couplers is a four port junction that divides the power equally on two ports with +90° phase shift and no power is present at the fourth port. Similarly, a signal entering port four will be equal-power divided between the two ports with -90° phase difference and no signal at port one.

For an n_r-element antenna array, $(n_r/2)\log_2(n_r)$ directional couplers and $(n_r/2)(\log_2(n_r) - 1)$ phase shifters are required to create the Butler matrix. The Butler matrix is theoretically lossless, but there will always be a finite insertion loss, due to the inherent losses in directional couplers, phase shifters and transmission lines. There is no theoretical limit of the bandwidth, but due to the hardware, the bandwidth is usually limited, to about 30%.

The normalized input-output relationship can be written as

$$F_r = \frac{1}{\sqrt{n_r}} \sum_{l=0}^{n_r-1} f_l \exp\left[-j\frac{2\pi lr}{n_r}\right] \tag{A.1}$$

and

$$f_l = \frac{1}{\sqrt{n_r}} \sum_{r=0}^{n_r-1} F_r \exp\left[j\frac{2\pi lr}{n_r}\right] \tag{A.2}$$

where f_l is the signal at the lth input port and F_r is the signal at the rth output port, connected to antenna r. Note the similarities with the fast Fourier transform (FFT).

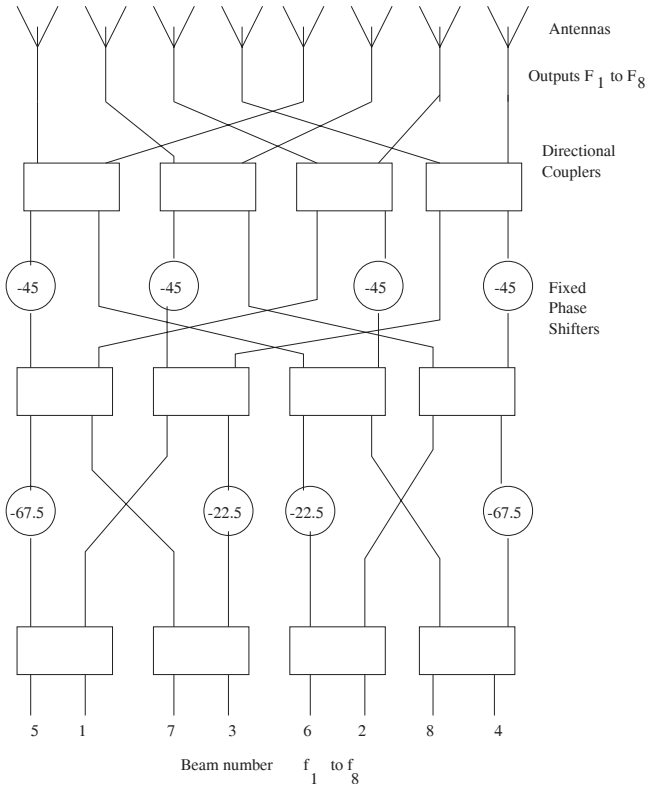


Figure A.1: Eight-element Butler matrix for beamforming

The fact is that the Butler matrix had existed a few years before the FFT appeared in 1965.

If the Butler matrix is connected to an antenna array, then the matrix will act so that the array will have a uniform amplitude distribution and constant phase difference between neighboring elements. This will then generate the beams depending on which input port is used. The direction of the beam depends on the wavelength of the transmitted signal, and will thus squint with frequency. If port one is used, then all other ports are grounded, the output of antenna r is

$$F_r = \frac{1}{\sqrt{n_r}} f_1 \exp\left[-j\frac{2\pi r}{n_r}\right] \tag{A.3}$$

and the radiation pattern in direction θ is

$$U(\theta) = \sum_{r=1}^{n_r} F_r(t) e^{-jr\frac{2\pi\delta}{\lambda_c}\sin(\theta) - j\frac{2\pi r}{n_r}} \tag{A.4}$$

where we assumed iso-tropical radiation patterns. This expression will have a maxima when the exponent is unity, given by

$$\frac{2\pi\delta}{\lambda_c}\sin(\theta_{max}) + \frac{2\pi r}{n_r} = 0 \tag{A.5}$$

Here δ is the inter-element spacing (linear uniformly spaced array is assumed) and λ_c is the wavelength. To maintain the same direction of the maximums when the frequency is changed, the distance δ must be changed to maintain the quotient δ/λ_c constant.

Bibliography

[1] P.K. Bondyopadhyay, "The first application of array antenna," in *Proceedings of IEEE International Conference on Phased Array Systems and Technology.*, Dana Point , USA, 2000, pp. 29–32.

[2] S. Andersson, B. Carlqvist, B. Hagerman, and R. Lagerholm, "Enhancing cellular network capacity with adaptive antennas," *Ericsson Review*, vol. 76, pp. 138–141, 1999.

[3] H. Dam, M. Berg, S. Andersson, R. Bormann, M. Frerich, and T. Henβ, "Performance evaluation of adaptive antenna base stations in a commercial GSM network," in *Proceedings of Vehicular Technology Conference (VTC)*, Piscataway, USA, 1999, pp. 47–51.

[4] K. Sheikh, D. Gesbert, D. Gore, and A.J. Paulraj, "Smart antennas for boradband wireless access networks," *IEEE Signal Processing Magazine*, vol. 37, no. 11, pp. 100–105, Nov. 1999.

[5] R.T. Derryberry, S.D. Gray, D.M. Ionescu, G.Mandyam, and B. Raghothaman, "Transmit diversity in 3G CDMA systems," *IEEE Communications Magazine*, vol. 40, no. 4, pp. 68–75, April 2002.

[6] S. Andersson and B. Hagerman, "Adaptive antennas in wireless systems-basic background and field-trial results," in *Proceedings of Radiovetenskaplig Konferens*, Karlskrona, Sweden, 1999, pp. 249–253.

[7] S. Andersson, U.Forssén, J.Karlsson, T.Witzschel, P. Fischer, and A. Krug, "Ericsson/Mannesmann GSM field trials with adaptive antennas," in *Proceedings of Vehicular Technology Conference (VTC)*, Phoenix,USA, May 1997, pp. 1587–1591.

[8] J. Strandell, M. Wennström, A. Rydberg, T. Öberg, O. Gladh, L. Rexberg, E. Sandberg, B.V. Andersson, and M. Appelgren, "Experimental evaluation of an adaptive antenna for a TDMA telephony system," in *IEEE Personal Indoor and Mobile Radio Communications Conference (PIMRC)*, Helsinki, Finland, Sep. 1997, pp. 79–84.

[9] J. Strandell, M. Wennström, T. Öberg, and A. Rydberg, "Design and evaluation of a fully adaptive antenna for telecommunication systems," in *Proceedings Antenn97 conference*, Gothenburg, Sweden, 1997, pp. 357–366.

[10] S.M. Simmonds and M. Beach, "Downlink calibration requirements for the TSUNAMI(II) adaptive antenna testbed," in *Proceedings of the Ninth International Symposium on Personal, Indoor and Mobile Radio Communications*, Boston, USA, 8-11 September 1998.

[11] P.E. Mogensen, K.I. Pedersen, P. Leth-Espensen, B. Fleury, F. Fredriksen, and K.Olesen, "Preliminary results from an adaptive antenna array testbed for GSM/UMTS," in *Proceedings of Vehicular Technology Conference (VTC)*, Phoenix,USA, May 1997, pp. 1592–1596.

[12] G. Tsoulos, M. Beach, and J. McGeehan, "Space division multiple acess (SDMA) field trials. Part 2: Calibration and linearity issues," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 145, no. 1, pp. 79–84, Feb. 1998.

[13] S.M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on selected areas in communications*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[14] T. Öberg, *Modulation, Detection and Coding*. Chichester: John Wiley and Sons, 2001.

[15] J.G. Proakis, *Digital Communications*. Singapore: McGraw-Hill, 1989.

[16] E. Lindskog. *Space-Time Processing and Equalization for Wireless Communications*, PhD thesis, Signals and Systems Group, Uppsala University, Uppsala, Sweden, 1999.

[17] C. Tidestav. *The Multivariable Decision Feedback Equalizer- Multiuser Detection and Interference Rejection*, PhD thesis, Signals and Systems Group, Uppsala University, Uppsala, Sweden, 1999.

[18] J.P. Kermoal, L. Schumacher, F. Fredriksen, and P.E. Mogensen, "Polarization diversity an MIMO radio channels: Experimental validation of a

stochastic model and performance assessment," in *Proceedings of Vehicular Technology Conference, VTC Fall*, Atlantic City, USA, Oct. 2001, pp. 22–26.

[19] R.U. Nabar, H. Bölcskei, and A.J. Paulraj, "Transmit optimization for spatial multiplexing in the presence of spatial fading correlation," in *IEEE Global Telecommunications Conference*, San Antonio, USA, Nov. 2001, pp. 131–135.

[20] J.L. Butler, *Digital matrix and intermediate frequency scanning, Scanning Antennas*, vol. 3: Academic Press, 1966.

[21] T. Svantesson, "An antenna solution for MIMO channels: The multimode antenna," in *Conference Record of the Thirty-Fourth Asilomar Conference*, Pacific Grove, USA, Oct. 2000.

[22] M. Wennström and T. Svantesson, "An antenna solution for MIMO channels: the switched parasitic antenna," in *IEEE Personal Indoor and Mobile Radio Communications Conference (PIMRC)*, San Diego, USA, Oct. 2001, pp. 159–163.

[23] C.C. Martin, J.H. Winters, and N.R. Sollenberger, "MIMO radio channel measurements: Performance comparison of antenna configurations," in *Proceedings of Vehicular Technology Conference, VTC Fall*, Atlantic City, USA, Oct. 2001, pp. 1225–1229.

[24] A.S. Macedo and E.S. Sousa, "Coded OFDM in broadband indoor wireless systems," in *Proceedings of IEEE Int. conf. on Communications (ICC)*, Montreal, Canada, June 1997, pp. 934–938.

[25] H. Bölcskei, D. Gesbert, and A.J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Transactions on Communications*, vol. 50, no. 2, pp. 225–234, Feb. 2002.

[26] T. Ottosson, A. Ahlen, A. Brunström, M. Sternad, and A. Svensson, "Towards 4G IP-based wireless systems," in *Future Telecommunications Conference*, Beijing,China, 2001.

[27] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.

[28] N. Seshadri and J.H. Winters, "Two signalling schemes to improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitter antenna diversity," in *Proceedings of Vehicular Technology Conference (VTC)*, Secaucus, USA, May 1993, pp. 508–511.

[29] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[30] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antenna," *Bell Labs technical Journal*, pp. 41–59, 1996.

[31] B. Hassibi and B. Hochwald, "High-rate linear space time codes," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, Salt Lake City, USA, May 2001, pp. 2461–2464.

[32] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining transmit beamforming and orthogonal space-time block codes by utilizing side information," in *Proceedings IEEE Sensor Array and Multichannel Signal Processing Workshop*, Cambridge, USA, March 2000, pp. 153–157.

[33] M.T. Ivrlac, T.P. Kurpjuhn, C. Brunner, and W. Utschick, "Efficient use of fading correlations in MIMO systems," in *Proceedings of Vehicular Technology Conference, VTC Fall*, Atlantic City, USA, Oct. 2001, pp. 2763–2767.

[34] M.T. Ivrlac and J.A. Nossek, "Correlated fading in MIMO systems - blessing or curse ?," in *Proceedings of the Allerton Conference on communications*, Monticello, USA, 2001.

[35] N. Sharma and E. Geraniotis, "Analyzing the performance of the space time block codes with partial channel state feedback," in *IEEE Wireless Communications and Networking Conference*, Chicago, USA, Sep. 2000, pp. 1362–1366.

[36] G. Ganesan and P. Stoica, "Differential detection based on space-time block codes," *Wireless Personal Communications*, vol. 21, pp. 163–180, 2002.

[37] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, March 1998.

[38] E. N. Onggosanusi, A. Gatherer, A. G. Dabak, and S. Hosur, "Performance analysis of closed-loop transmit diversity in the presence of feedback delay," *IEEE Transactions on Communications*, vol. 49, pp. 1618–1630, 2001.

[39] A.T. James, "Distribution of matrix variate and latent roots derived from normal samples," *Ann. Math. Stat*, vol. 35, pp. 475–501, 1964.

[40] S.M. Kay, *Fundamentals of Statistical Signal Processing, vol.II, Detection theory*. New Jersey: Prentice Hall, 1998.

[41] R.W. Heath.Jr. and A. Paulraj, "Switching between multiplexing and diversity based on constellation distance," in *Proceedings of the Allerton Conference on communications*, Monticello, USA, Oct. 2000.

[42] H. Lutkepohl, *Handbook of matrices*. Chichester: John Wiley and Sons, 1999.

[43] R.J. Muirhead, *Aspects of multivariate statistical theory*. Ann Arbour, USA: John Wiley and Sons, 1982.

[44] G. Taricco and E. Biglieri, "Exact pairwise error probability of space-time codes," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 510–513, Feb. 2002.

[45] R. Knopp and P.A. Humblet, "On coding for block-faded channels," *IEEE Transactions on Information Theory*, vol. 46, no. 1, pp. 189–205, Jan. 2000.

[46] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels:information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[47] I.E. Telatar, "Capacity of multi antenna Gaussian channels," *European transactions on telecommunication*, pp. 585–595, Nov. 1999, Originally published as an AT&T Bell Labs Technical Memo, June 1995.

[48] G.J. Foschini and M.J. Gans, "On limits of wireless communications in a fading environment using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[49] R.G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, 1968.

[50] A.J. Goldsmith and P.P. Varaiya, "Capacity of fading channels with side-information," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1896–1992, Nov. 1997.

[51] D.W. Bliss, K.W. Forsythe, A.O. Hero, and A.L. Swindlehurst, "MIMO environmental capacity sensitivity," in *Conference Record of the Thirty-Fourth Asilomar Conference*, Pacific Grove, USA, Nov. 2000, pp. 764–768.

[52] B.A. Bjerke and J.G. Proakis, "Multiple antenna diversity techniques for transmitting over fading channels," in *IEEE Wireless Communication and Networking Conference*, New Orleans, USA, Sep. 1999, pp. 1038–1042.

[53] B. Hassibi and H. Vikalo, "On the expected complexity of sphere decoding," in *Conference Record of the Thirty-Fourth Asilomar Conference*, Pacific Grove, USA, Nov. 2001, pp. 1051–1055.

[54] S. Sandhu, R.W. Heath Jr., and A. Paulraj, "Space time block codes versus space-time trellis codes," in *Proceedings of IEEE Int. conf. on Communications (ICC)*, vol. 4, Helsinki, Finland, June 2001, pp. 1132–1136.

[55] P. Stoica and G. Ganesan, "Space-time block codes: Trained, semi-blind and blind detection," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Orlando, USA, May 2002, pp. 1609–1612.

[56] G. Ganesan and P. Stoica, "Space-time diversity using orthogonal and amicable orthogonal designs," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Istanbul, Turkey, June 2000, pp. 2561–2564.

[57] P. Stoica and G. Ganesan, "Maximum-SNR space-time designs for MIMO channels," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, USA, May 2001, pp. 2425–2428.

[58] G.B. Giannakis, editor, *Signal Processing Advances in Wireless and Mobile Communications: Trends in Single- and Multi-user systems*, vol. 2, chapter 3, Prentice Hall, Upper Saddle River, USA, 2001, Space-Time Diversity Chapter, authored by G. Ganesan and P. Stoica.

[59] S. Sandhu and A. Paulraj, "Union bound on error probability of linear space-time block codes," in *IEEE Transactions on Acoustics, Speech and Signal Processing*, Salt Lake City, USA, May 2001, pp. 2473–2476.

[60] S. Sandhu and A. Paulraj, "Space time block codes: A capacity perspective," *IEEE Communications Letters*, vol. 4, no. 12, pp. 384–386, Dec. 2000.

[61] A.J. Goldsmith and S-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[62] S.T. Chung and A.J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1561–1571, Sept. 2001.

[63] D.W. Lin, "On optimal bit loading for multitone ADSL," in *IEEE International Symposium on Circuits and Systems*, Geneva, Switzerland, May 2000, pp. 597–600.

[64] B.S. Krongold, K. Ramchandran, and D.L. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *IEEE Transactions on Communications*, vol. 48, no. 1, pp. 23–27, Jan. 2000.

[65] T.H. Liew and L. Hanzo, "Space-time block coded adaptive modulation aided OFDM," in *IEEE Global Telecommunications Conference*, San Antonio, USA, Nov. 2001, pp. 136–140.

[66] M.-S. Alouini and A.J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity combining techniques," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1165–1181, 1999.

[67] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Transactions on Communications*, vol. 49, pp. 2198–2206, 2001.

[68] J. Campello, "Optimal discrete bit loading for multicarrier modulation," in *IEEE International Symposium on Information Theory*, Cambridge, USA, Aug. 1998, p. 193.

[69] M. Wennström, "Smart antenna implementation issues for wireless communications,", Technical report, Signals and Systems Group, Uppsala University, Uppsala, Sweden, 1999, Technical Licentiate Thesis in Signal Processing, http://www.signal.uu.se/Publications/abstracts/l991.html.

[70] B. Noble and J.W. Daniel, *Applied Linear Algebra*. Englewood Cliffs, USA: Prentice-Hall, 1988.

[71] J. Litva and T. Lo, *Digital beamforming in wireless communications*. Boston: Artech-House Publishers, 1996.

[72] D. Shim and S. Choi, "Should the smart antenna be a tracking beam array or switching beam array?," in *Proceedings of Vehicular Technology Conference (VTC)*, Ottawa, Canada, May 1998, pp. 494–498.

[73] S.S. Swales, M.A. Beach, D.J. Edwards, and J.P. McGeehan, "The performance enhancement of multibeam adaptive base-station antennas for cellular land mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 39, no. 1, pp. 56–67, Feb. 1990.

[74] C.R. Ward, D.N. Adams, F.M. Wilson, K.S. Wilson, and A.K. Bush, "The live-air trial of a multi-beam cellular base station antenna system," in *IEE National Conference on Antennas and Propagation*, London, UK, 1999, pp. 169–172.

[75] M.J. Ho, G.L. Stuber, and M.D. Austin, "Performance of switched-beam smart antenna systems," *IEEE Transactions on Vehicular Technology*, vol. 47, pp. 10–19, 1998.

[76] P. Petrus, R.B. Ertel, and J.H. Reed, "Capacity enhancement using adaptive arrays in an AMPS system," *IEEE Transactions on Vehicular Technology*, vol. 47, pp. 717–727, 1998.

[77] K.I. Pedersen, P.E. Mogensen, and F. Fredriksen, "Joint directional properties of uplink and downlink channel in mobile communication," *Electronic Letters*, vol. 35, no. 16, pp. 1311–1312, Aug. 1999.

[78] J.P Shelton and K.S. Kelleher, "Multiple beams from linear arrays," *IRE Transactions on antennas and propagation*, vol. 9, pp. 154–161, 1961.

[79] M-S. Alouini and A.J. Goldsmith, "Comparison of fading channel capacity under different CSI assumptions," in *Proceedings of Vehicular Technology Conference, VTC Fall*, Boston, USA, Sep. 2000, pp. 1844–1849.

[80] I.S. Gradshteyn, I.M. Ryzhik, and A. Jeffrey, *Table of integrals, series, and products*. San Diego, USA: Academic Press, 6 edition, 2000.

[81] J.W. Wallace and M.A. Jensen, "Modeling the indoor MIMO wireless channel," *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 5, pp. 591–599, May 2002.

[82] M. Nakagami, "The m-distribution - a general formula of intensity distribution of rapid fading,", in *Statistical Methods in Radio Wave Propagation*, Pergamon, Oxford, England, 1960.

[83] J. Luo, J.R. Zeidler, and S. McLaughlin, "Performance analysis of compact antenna arrays with MRC in correlated Nakagami fading channels," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 1, pp. 267–277, Jan. 2001.

[84] S. Rice, "Mathematical analysis of random noise," *Bell System Technical Journal*, vol. 23, 1944.

[85] C.N. Chuah, D. Tse, J.M. Kahn, and R.A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 637–650, March 2002.

[86] H. Bölcskei and A. Paulraj, "Performance of space-time codes in the presence of spatial fading correlation," in *Conference Record of the Thirty-Fourth Asilomar Conference*, Pacific Grove, USA, October 2000.

[87] J. Wallace and M.A. Jensen, "Mutual coupling in MIMO wireless systems: A rigorous network theory analysis,", 2002, Submitted to IEEE Journal on Selected Areas in Communication.

[88] D-S. Shiu, G.J Foschini, M.J. Gans, and J.M Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502–513, March 2000.

[89] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M.Beach, "Second order statistics of NLOS indoor MIMO channels based on 5.2 GHz measurements," in *IEEE Global Telecommunications Conference*, San Antonio, USA, Nov. 2001, pp. 156–160.

[90] A. Grant, "Rayleigh fading multiple-antenna channnels," *EURASIP Journal on applied signal processing, Special issue on space time coding (Part 1)*, vol. 2002, no. 3, pp. 316–329, March 2002.

[91] Z.D. Bai and Y.Q. Yin, "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," *Ann. Probab.*, vol. 21, pp. 1275–1294, 1993.

[92] Z.D. Bai, J.W. Silverstein, and Y.Q. Yin, "A note on the largest eigenvalue of a large dimensional sample covariance matrix," *Journal on Multivariate Analysis*, vol. 26, pp. 166–168, 1988.

[93] K. Johansson, "Shape fluctuations and random matrices," *Communications in mathematical physics*, vol. 209, pp. 437–476, 2000.

[94] A.A.M. Saleh and R.A. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE Journal on Selected Areas in Communications*, vol. 5, pp. 128–132, 1987.

[95] Q.H. Spencer, B.D. Jeffs, M.A. Jensen, and A.L. Swindlehurst, "Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 347–360, March 2000.

[96] P. Petrus, J.H. Reed, and T.S. Rappaport, "Geometrical-based statistical macrochannel model for mobile environments," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 495–502, March 2002.

[97] W.C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.

[98] A. Ábdi and M. Kaveh, "Space-time correlation modeling of multielement antennas systems in mobile fading channels," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, USA, 2001, pp. 2505–2508.

[99] D.P. McNamara, M.A. Beach, and P.N. Fletcher, "Experimental investigation of the temporal variation of mimo channels," in *Proceedings of Vehicular Technology Conference, VTC Fall*, Proceedings of Vehicular Technology Conference, VTC Fall, Oct. 2001, pp. 1063–1067.

[100] J. Cheng and N.C. Beaulieu, "Maximum-likelihood based estimation of the Nakagami-m parameter.," *IEEE Communications letters*, vol. 5, no. 3, pp. 101–103, March 2001.

[101] L.J. Greenstein, D.G.Michelson, and V.Erceg, "Moment-method estimation of the ricean K-factor," *IEEE Communications Letters*, vol. 3, no. 6, pp. 175–176, June 1999.

[102] J. Salz and J.H. Winters, "Effect of fading correlation on adaptive arrays in digital mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 4, pp. 1049–1057, Nov. 1994.

[103] H. Bölcsei and A.J. Paulraj, "Efficient use of fading correlations in MIMO systems," in *Conference Record of the Thirty-Fourth Asilomar Conference*, Pacific Grove, USA, 2000, pp. 687–693.

[104] T. Svantesson. *Antennas and Propagation from a Signal Processing Perspective*, PhD thesis, Chalmers University of Technology, Gothenburg, Sweden, 2001.

[105] T. Svantesson and J. Wallace, "On signal strength and multipath richness in muliple-input multiple-output systems," in *Proceedings of Radiovetenskaplig Konferens*, Kista, Sweden, 2002.

[106] S. Loyka and G. Tsoulos, "Estimating MIMO system performance using the correlation matrix approach," *IEEE Communication Letters*, vol. 6, pp. 19–21, 2002.

[107] Q.T. Zhang, "Maximum-ratio combining over Nakagami fading channels with an arbitrary branch covariance matrix," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1141–1150, July 1999.

[108] C.A. Balanis, *Antenna Theory, Analysis and Design*. New York, USA: John Wiley and Sons, 1997.

[109] T. Svantesson and A. Ranheim, "Mutual coupling effects on the capacity of multielement antenna systems," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, USA, May 2001, pp. 2485–2488.

[110] "Url:http://functions.wolfram.com,".

[111] I.J. Gupta and A.K. Ksienski, "Effect of mutual coupling on the performance of adaptive arrays," *IEEE Transactions on Antennas and Propagation*, vol. 31, pp. 785–789, 1983.

[112] D.A. Gore, R.U. Nabar, and A. Paulraj, "Selecting an optimal set of transmit antennas for a low rank matrix channel," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Istanbul, Turkey, June 2000, pp. 2785–2788.

[113] A.F. Molisch, M.Z. Win, and J.H. Winters, "Reduced-complexity transmit/receive-diversity systems," in *Proceedings of Vehicular Technology Conference, VTC Spring*, Rhodes, Greece, May 2001, pp. 1996–2000.

[114] Q.H. Spencer and A.L. Swindlehurst, "Some results on channel capacity when using multiple antennas," in *Proceedings of Vehicular Technology Conference, VTC Fall*, Boston, USA, Sep. 2000, pp. 681–688.

[115] M. Hayajneh and A. Scaglione, "Space-time codes for high bit rate wireless communications: asymptotic performance of space-time random codes," in *IEEE International symposium on circuits and systems*, 2002.

[116] J.K. Han, D.K. Kim, and H.K. Park, "Performance comparison of the systems employing multiple antennas over spatially correlated rayleigh fading channel," in *IEEE Global Telecommunications Conference*, San Antonio, USA, Nov. 2001, pp. 3217–3221.

[117] G. Jöngren and B. Ottersten, "Combining transmit antenna weights and orthogonal space-time block codes by utilizing side information," in *Conference Record of the Thirty-Third Asilomar Conference on Signals, Systems, and Computers, 1999.*, Pacific Grove, USA, Oct. 1999, pp. 1562–1566.

[118] A. Hottinen and R. Wichman, "Transmit diversity using filtered feedback weights in the FDD/WCDMA system," in *International Zürich seminar on Broadband Communications, 2000. Proceedings.*, Zürich, Switzerland, Feb. 2000, pp. 15–21.

[119] S. Hu and A. Duel-Hallen, "Combined adaptive modulation and transmitter diversity using long range prediction for flat fading mobile radio channels," in *IEEE Global Telecommunications conference*, San Antonio, USA, 2001, pp. 1256–1261.

[120] J.N. Pierce and S. Stein, "Multiple diversity with nonindependent fading," in *Proceedings of IRE*, vol. 48, 1960.

[121] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. Singapore: McGraw-Hill, 1984.

[122] S-S. Jeon, Y. Wang, Y. Qian, and T. Itoh, "A novel smart antenna system implementation for broad-band wireless communications," *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 5, pp. 600–606, 2002.

[123] H. Andersson and M. Landing, "Adaptive antenna for mobile telephone systems,", Technical report, Signals and Systems, Uppsala University, Uppsala, Sweden, 1997, Technical Licentiate Thesis in Electronics.

[124] A.O. Boukalov and S-G. Häggman, "System aspects of smart-antenna technology in cellular wireless communications-an overview," *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, pp. 919–929, 2000.

[125] ETSI / GSM Recommendation 05.05, "Radio transmission and reception,", 1991.

[126] R.T. Compton Jr., *Adaptive Antennas*. New Jersey: Prentice-Hall, 1988.

[127] R. Monzingo and T. Miller, *Introduction to adaptive arrays*. New York: John Wiley and Sons, 1980.

[128] L.C. Godara, "Applications of antenna arrays to mobile communications, part I:performance improvement, feasibility, and system considerations," *Proceedings of the IEEE*, vol. 85, no. 7, pp. 1031–1060, July 1997.

[129] L.C. Godara, "Applications of antenna arrays to mobile communications, part II:beam-forming and direction of arrival considerations," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1195–1245, Aug. 1997.

[130] J. Hudson, *Adaptive array principles*. New York: Peter Peregrinus Ltd, 1992.

[131] H. Lin, "Spatial correlation in adaptive arrays," *IEEE Transactions on Antennas and Propagation*, vol. 30, pp. 212–223, 1982.

[132] M. Wax and Y. Anu, "Performance analysis of the minimum variance beamformer," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 928–937, April 1996.

[133] TSUNAMI partners, "Adaptive antenna performance in mobile systems,", Technical report, ERA, 1998.

[134] M.W. Gantz, R.L. Moses, and S.L. Wilson, "Convergence of the SMI algorithms with weak interference," *IEEE Transactions on Antennas and Propagation*, vol. 38, no. 3, pp. 394–399, 1990.

[135] H. Steyskal, "Array error effects in adaptive beamforming," *Microwave journal*, vol. 34, pp. 101–112, 1991.

[136] J. Hudson, "The effects of signal and weight coefficient quantisation in adaptive array processors," in *Proceedings of the NATO Advanced Study Institute of Signal Processing*, Reidel, Dordrecht, Netherlands, 1977, pp. 423–428.

[137] T. Takahashi, I. Chiba, and Y. Karasawa, "The required resolution of A/D for null beamforming in a DBF antenna," in *IEEE Antennas and Propagation Society International Symposium Digest*, Seattle,USA, 1994, pp. 128–131.

[138] R. Nitzberg, "Computational precision requirements for optimal weights in adaptive processing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 16, pp. 418–425, 1980.

[139] I.S. Reed, J.D. Mallet, and L.E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Transactions on Antennas and Propagation*, vol. 10, no. 6, pp. 853, 1974.

[140] R. Davis and P. Sher, "Quantisation noise in adaptive weighting networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 20, pp. 547–559, 1984.

[141] R. Nitzberg, "Effect of errors in adaptive weights," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 12, pp. 369–373, 1976.

[142] Voloshina, Gabrielyan, and Shatskii, "The discretization effect of phase control on the noise-to-signal ratio in adaptive arrays," *Radioelectronics and Communication Systems*, vol. 38, pp. 44–45, 1995, Translated from: Izvestiya Vysshikh Uchebnykh Zavedenii Radioelektronika vol.38, no.7; 1995; p.67-69.

[143] L.C. Godara, "The effect of phase-shifter errors on the performance of an adaptive antenna-array beamformer," *IEEE Journal of Oceanic Engineering*, vol. 10, pp. 278–284, 1985.

[144] M. Beach, P. Brown, M. Dowds, J. Fonollosa, D. Greenaway, P. Kenington, S. Mayrargue, J.J. Monot, C. Passmann, C. Simmonds, J. Thibault, and G. Villino, "Requirements for flexible multi-standard adaptive antenna basestations,", Technical Report AC347/FTC/A2.1/DS/P/002/b1, SUNBEAM Project, 1998.

[145] G. Tsoulos and M. Beach, "Calibration and linearity issues for an adaptive antenna system," in *Proceedings of Vehicular Technology Conference (VTC)*, Phoenix, USA, 1997, pp. 1597–1600.

[146] B.D. Carlson, "Covariance estimation and diagonal loading in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 397–401, 1988.

[147] E. Lindskog, "Making SMI-beamforming insensitive to the sampling timing for gsm signals," in *Proceedings of the Sixth International Symposium on Personal, Indoor and Mobile Radio Communications*, Toronto, Canada, 1995, pp. 664–668.

[148] M. Dell'Anna and A. Hamid Aghvami, "Performance of optimum and suboptimum combining at the antenna array of a w-cdma system," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 2123–2137, 1999.

[149] A.F. Naguib, A. Paulraj, and T. Kailath, "Capacity improvement with base-station antenn arrays in cellular cdma," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 691–698, 1994.

[150] J. Mitola, "Guest editorial on software radios," *IEEE Communications Magazine*, vol. 33, no. 5, pp. 24–25, 1995.

[151] K.J. Maalouf, R.C. Gaus Jr., S. Sowelam, and A. Jacomb-Hood, "Error rate estimation in a multi-channel active phased array," in *Proceedings of IEEE Int. conf. on Communications (ICC)*, Atlanta, USA, 1998, pp. 402–406.

[152] P.A. Chiavacci, "Analysis of noise power ratio effects on satellite communication system performance," *Microwave Journal*, vol. 40, no. 6, pp. 24–34, 1997.

[153] R. De Gaudenzi, "Payload nonlinearity impact on the globalstar forward link multiplex part I: Physical layer analysis," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 3, pp. 960–976, May 1999.

[154] P.B. Kenington and D.W. Bennett, "Linear distortion correction using a feedforward system," *IEEE Transactions on vehicular technology*, vol. 45, no. 1, pp. 74–81, 1996.

[155] A.N. D'Andrea, V. Lottici, and R. Reggiannini, "RF power amplifier linearization through amplitude and phase predistortion," *IEEE Transactions on communications*, vol. 44, no. 11, pp. 1477–1484, 1996.

[156] P.B. Kenington, "Linearized transmitters: An enabling technology for software defined radio," *IEEE Communications Magazine*, vol. 40, pp. 156–162, 2002.

[157] P.B. Kenington, D.W. Bennett, and R.J. Wilkinson, "Determing the power rating of a multichannel power amplifier," *IEE Proceedings on Communications*, vol. 142, no. 4, pp. 274–280, 1995.

[158] R.J. Wilkinson, D.W. Bennett, and P.B. Kenington, "Distortion effects of multicarrier envelope limiting," *IEE Proceedings on Communications*, vol. 144, no. 5, pp. 349–356, 1997.

[159] W.A. Sandrin, "Spatial distribution of intermodulation products in active phased array antennas," *IEEE Transactions on Antennas and Propagation*, vol. 22, pp. 864–868, 1973.

[160] S.L. Loyka, "The influence of electromagnetic environment on operation of active array antennas: analysis and simulation techniques," *Antennas and Propagation Magazine*, vol. 41, pp. 23–39, 1999.

[161] C. Hemmi, "Pattern characteristics of harmonic and intermodulation products in broad-band active transmit arrays," *IEEE Transactions on Antennas and Propagation*, vol. 50, no. 6, pp. 858–865, June 2002.

[162] M. Wennström, "Considering downlink intermodulation distortion in switched multibeam antennas for cellular radio systems," in *Proceedings of Vehicular Technology Conference, VTC Fall*, Boston, USA, 2000, pp. 1858–1865.

[163] S. Loyka, "Characteristics of recieving intermodulation channel of active array antennas," *International Journal of Electronics*, vol. 80, no. 4, pp. 595–602, 1996.

[164] H. Xue, R. Davies, M. Beach, and J. McGeehan, "Linearity considerations in adaptive antenna array applications," in *IEEE Personal Indoor and Mobile Radio Communications Conference (PIMRC)*, Toronto, Canada, Sep. 1995, pp. 682–686.

[165] M.C. Jeruchim, *Simulation of communication systems*. New York: Plenum Press, 1992.

[166] K.W. Schneider and W.H. Tranter, "Efficient simulation of multicarrier digital communication systems in nonlinear channel environments," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 3, pp. 328–339, 1993.

[167] O. Shimbo, "Effects of intermodulation, AM-PM conversion, and additive noise in multicarrier TWT systems," *Proceedings of the IEEE*, vol. 59, no. 2, pp. 230–239, 1971.

[168] Merril I. Skolnik, *Introduction to Radar Systems*. Auckland: McGraw Hill, 1980.

[169] D.D. Siljak, *Nonlinear systems*. New York: John Wiley and Sons, 1969.

[170] A.J. Cann, "Nonlinearity model with variable knee sharpness," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 16, pp. 874–878, 1980.

[171] S. Loyka, "On the use of Cann's model for nonlinear behavioral-level simulation," *IEEE Transactions on Vehicular Technology*, vol. 5, pp. 1982–1985, 2000.

[172] O. Shimbo, L.N. Nguyen, J.P. De Almeida, and E. Alberquerque, "Modulation-transfer noise effects among FM and digital signals in memoryless nonlinear devices," *Proceedings of the IEEE*, vol. 74, no. 4, pp. 580–597, 1986.

[173] A. Kavak, W. Yang, G. Xu, and W.J. Vogel, "Characteristics of vector propagation channels in dynamic mobile scenarios," *IEEE Transactions on Antennas and Propagation*, vol. 49, no. 12, pp. 1695–1702, 2001.

[174] 3rd Generation Partnership Project; Technical Specification Group Radio Access Networks, "TS 25.141 base station conformance testing (FDD) (Release 1999),", 2000.

[175] J. Shi and K. Arvidsson, "Performance of multi-carrier power amplifiers in hierarchical cell structure," in *Proceedings of Vehicular Technology Conference (VTC)*, Ottawa, Canada, 1998, pp. 1622–1625.

[176] K. Hamied and G. Labedz, "AMPS cell transmitter interference to CDMA mobile receiver," in *Proceedings of Vehicular Technology Conference (VTC)*, New York, USA, 1996, pp. 1467–1471.

[177] H. Jung and O.K. Tonguz, "Random spacing channel assignment to reduce the nonlinear intermodulation distortion in cellular mobile communications," *IEEE Transactions on Vehicular Technology*, vol. 48, pp. 1666–1675, 1999.

[178] B. Hwang and O.K. Tonguz, "A generalized suboptimum unequally spaced channel allocation technique-part I:In IM/DD WDM systems," *IEEE Transactions on Communications*, vol. 46, pp. 1027–1037, 1998.

[179] M.D. Yacoub, *Foundations of Mobile Radio Engineering*. Boca Raton, Florida: CRC Press, 1993.

[180] M. Hayes, *Statistical digital signal processing and modeling*. New York: John Wiley and Sons, 1996.

[181] M. Li, T.A. Wilkinson, M. Beach, S.K. Barton, H. Xue, I.R. Johnson, and A. Nix, "Analysis of intermodulation distortion specification for radio LANs using multicarrier schemes," *Electronic Letters*, vol. 29, pp. 1229–1231, 1993.

[182] R. Prasad and A. Kegel, "Improved assessment of interference limits in cellular radio performance," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 2, pp. 412–419, May 1991.

[183] A.A. Almhdie, V. Kezys, and T.D. Todd, "Improved capacity in TDMA/SDMA using switched parasitic antennas," in *IEEE Personal Indoor and Mobile Radio Communications Conference (PIMRC)*, London,UK, 2000, pp. 363–367.

[184] N.L. Scott, O. L-Taylor, and R.G. Vaughan, "Diversity gain from a single-port adaptive antenna using switched parasitic elements illustrated with a wire and monolopole prototype," *IEEE Transactions on Antennas and Propagation*, vol. 47, no. 6, pp. 1066–1070, June 1999.

[185] R. Vaughan, "Switched parasitic elements for antenna diversity," *IEEE Transactions on Antennas and Propagation*, vol. 47, no. 2, pp. 399–405, Feb. 1999.

[186] S.L. Preston, D.V. Thiel, T.A. Smith, S.G. O'Keefe, and J.W. Liu, "Base-station tracking in mobile communications using a switched parasitic antenna array," *IEEE Transactions on Antennas and Propagation*, vol. 46, no. 6, pp. 841–844, June 1998.

[187] D.V. Thiel and S. Smith, *Switched Parasitic Antennas for Cellular Communications*. Boston: Artech House, 2002.

[188] R. Schlub, D.V. Thiel, J.W. Lu, and S.G. O'Keefe, "Dual-band six-element switched parasitic array for smart antenna cellular communications systems," *Electronic Letters*, vol. 36, no. 16, pp. 1342–1343, Aug. 2000.

[189] S.L. Preston, D.V. Thiel, J.W. Lu, S.G. O'Keefe, and T.S. Bird, "Electronic beam steering using switched parasitic patch elements," *Electronic letters*, vol. 33, no. 1, pp. 7–8, Jan. 1997.

[190] R. Vaughan and J.B. Andersen, "Antenna diversity in mobile communications," *IEEE Transactions on Vehicular Technology*, vol. 36, pp. 149–172, 1987.

[191] R.H. Clarke, "A statistical theory of mobile radio reception,", Technical report, Bell System Technical Journal, 1969.

[192] J.N. Pierce and S. Stein, ""Multiple Diversity with Nonindependent Fading"," *Proc. of the IRE*, vol. 48, pp. 89–104, Jan. 1960.

[193] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, July 1996.

[194] P-S. Kildal, "Equivalent circuits of receive antennas in signal processing arrays," *Microwave and Optical Technology Letters*, vol. 21, pp. 244–246, 1999.

[195] B. Ottersten, B. Wahlberg, M. Viberg, and T. Kailath, "Stochastic maximum likelihood estimation in sensor arrays by weighted subspace fitting," in *Conference Record of the Thirty-Third Asilomar Conference*, Pacific Grove, USA, 1989, pp. 599–603.

[196] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectrum Estimation Workshop*, Rome, NY, USA, 1979, pp. 243–258.