MIMO OFDM for Wireless LANs

MIMO OFDM for Wireless LANs

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
Rector Magnificus, prof.dr. R.A. van Santen, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op woensdag 14 april 2004 om 16.00 uur

door

Albert van Zelst

Geboren te Waalwijk

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr.ir. G. Brussaard
en
prof.dr.ir. L.P. Ligthart

Copromotor:
dr.ir. P.F.M. Smulders

aan Jessica
aan onze ouders

Abstract

Broadband applications – such as high-speed computer networks, in-home delivery of multimedia services, or hospital data networks for telediagnostics involving digital imaging – and the demand for flexibility drive the need for broadband wireless communication systems. Since the available frequency spectrum is scarce, future systems should be characterised by significantly enhanced spectral efficiency in order to increase link throughput and network capacity. A very promising approach is to use multiple antennas at both the transmitter and the receiver (i.e., a Multiple-Input Multiple-Output (MIMO) system). With such a system the throughput can be increased by simultaneously transmitting different streams of data on the different transmit antennas but at the same carrier frequency. Although these parallel data streams are mixed up in the air, they can be recovered at the receiver by using spatial sampling (i.e., multiple receive antennas) and corresponding signal-processing algorithms, provided that the MIMO channel is well conditioned. This is in general the case in rich-scattering environments, e.g., indoor environments. Above technique is referred to as Space Division Multiplexing (SDM).

The combination of the throughput enhancement of SDM with the robustness of Orthogonal Frequency Division Multiplexing (OFDM) against frequency-selective fading caused by severe multipath scattering and narrowband interference is regarded as a very promising basis for future (indoor) high data-rate radio communication systems. SDM OFDM is the focus of this dissertation and its main contents and contributions, in the natural order from fundamental understanding, theoretical analysis to practical measurements, are as follows.

1. By means of a physical interpretation, a fundamental and intuitive explanation is given of the spectral efficiency and stability of a wireless MIMO system in rich-scattering environments such as indoor environments.

2. A generic wideband indoor MIMO channel model is proposed, including a Line-Of-Sight (LOS) component and spatial correlation, compacting the typically large number of channel parameters into a very few carefully selected ones. A major contribution is the

presented simple spatial correlation model with only one or two coefficient(s) that is proved to match the statistics of measured correlation matrices – typically consisting of a large number of parameters – in terms of capacity and error-rate performance.

3. Different narrowband SDM algorithms are described, including soft-decision demapping schemes for cases in which outer coding and decoding is applied. The error-rate performance and complexity of the algorithms are evaluated for different antenna configurations, for various constellation sizes, for different channel properties (some of which also include spatial correlation or a LOS component), with and without coding. It is shown that Maximum Likelihood Detection (MLD) outperforms the other schemes. Its complexity, however, is the highest and growing exponentially with the number of transmit antennas. Less complex alternatives are found that have only a slightly worse performance.

4. Through a unified view on (coded) MIMO techniques presented in this dissertation, we observed that the best performance is achieved by doing an exhaustive maximum likelihood search over the non-redundant lattice representing all possible space-time codewords. The complexity of such a search, however, grows exponentially with the number of lattice points. The turbo SDM scheme introduced in this dissertation allows for a significant complexity reduction, while performing very close to the overall exhaustive search. The complexity is reduced by splitting the temporal and spatial processing. The high performance at the receiver is achieved by iterating between the spatial demapping and temporal decoding. This approach stems from the turbo decoding principle.

5. Since OFDM already forms the basis of the current Wireless Local Area Network (WLAN) standards, IEEE 802.11a and g, the combination of SDM and OFDM is seen as an attractive solution for future high-speed indoor WLANs. In general, OFDM splits a wideband frequency-selective fading channel into a number of narrowband frequency-flat fading channels. As a result, all presented narrowband SDM algorithms can be readily applied to these subchannels. The combination SDM OFDM is evaluated in theory, with performance simulations, and with measurements. The theoretical evaluation is carried out by means of a general Space-Frequency performance analysis. It is shown that the maximum diversity gain equals the product of the number of transmit and receive antennas and the effective length of the channel impulse response. Coded SDM OFDM schemes are proposed that achieve a significant part of this diversity gain.

6. When a practical implementation is envisioned, the system has to deal with system impairments such as frequency offset, timing offset, phase noise, DC offset, etc. To tackle these impairments for SDM OFDM in the WLAN context, we propose training and synchronisation (tracking) algorithms which are extensions of those for IEEE 802.11a systems. In order to validate these algorithms and the general SDM OFDM concept, a test system with three transmit and three receive antennas, and based on IEEE 802.11a parameters, was built within Agere Systems, The Netherlands. Results from measurements with this test system in a typical office environment show successful transmissions up to 162 Mb/s, which is three times the data-rate of a regular IEEE 802.11a OFDM system.

Finally, it is concluded that SDM OFDM, although there is room for improvements, is an attractive and practical solution to enhance the throughput and/or robustness of wireless communication systems based on standards such as IEEE 802.11a considerably.

Contents

Acronyms

ADC	Analogue to Digital Converter
AoA	Angle of Arrival
AoD	Angle of Departure
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BLAST	Bell-Labs Layered Space-Time
BPSK	Binary Phase Shift Keying
cdf	cumulative distribution function
CDP	Correlation Delay Profile
CIR	Channel Impulse Response
CP	Cyclic Prefix
CPE	Common Phase Error
CSI	Channel State Information
CT	Coarse Timing
DAC	Digital to Analogue Converter
D-BLAST	Diagonal BLAST
DFE	Decision Feedback Equalisation
DFT	Discrete Fourier Transform
FD	Frame Detection
FFT	Fast Fourier Transform
FO	Frequency Offset
GI	Guard Interval
GSM	Global System for Mobile communications
ICI	Inter-Carrier Interference
IDFT	Inverse Discrete Fourier Transform
IEEE	Institute of Electrical and Electronics Engineers
IF	Intermediate Frequency
IFFT	Inverse Fast Fourier Transform
i.i.d.	independent, identically distributed
ISI	Inter-Symbol Interference

JC	Joint Coding
LGI	Long Guard Interval
LLR	Log-Likelihood Ratio
LNA	Low Noise Amplifier
LOS	Line Of Sight
LT	Long Training
MAC	Medium Access Control
MAP	Maximum A-Posteriori Probability
Mbps	Megabits per second
MIMO	Multiple-Input Multiple-Output
MLD	Maximum Likelihood Detection
MMSE	Minimum Mean Squared Error
MNC	Maximum Normalised Correlation
MSE	Mean Square Error
NLOS	Non Line Of Sight
pdf	probability density function
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PAC	Per-Antenna Coding
PC	Personal Computer
PDP	Power Delay Profile
PEP	Pairwise Error Probability
PER	Packet Error Rate
PHY	PHYsical layer
PLL	Phase Locked Loop
PN	Phase Noise
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
rms	root mean square
RX	Receiver
SDM	Space Division Multiplexing
SDMA	Space Division Multiple Access
SIC	Successive Interference Cancellation
SISO	Single-Input Single-Output
SNR	Signal-to-Noise Ratio
SOMLD	Soft-Output Maximum Likelihood Detection
SOMMSE	Soft-Output Minimum Mean Squared Error
SOZF	Soft-Output Zero Forcing
STBC	Space-Time Block Coding
ST	Short Training
STC	Space-Time Coding
STTC	Space-Time Trellis Coding
SVD	Singular Value Decomposition
TU/e	Eindhoven University of Technology
TX	Transmitter
V-BLAST	Vertical BLAST
WLAN	Wireless Local Area Network
ZF	Zero Forcing

Functions, Operators, and Transforms

x^* complex conjugate of x

$E[f(x)] = \int_{-\infty}^{\infty} f(x)p(x)dx$ expected value of $f(x)$

$p(x)$ probability density function (pdf) of random variable x

$p(x|y)$ probability density function (pdf) of random variable x conditioned on y

$\delta(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$ Discrete-time Dirac function

$\lceil x \rceil$ The ceil operator, providing the lowest integer value larger than or equal to x

x_i or $(\mathbf{x})_i$ i-th element of vector $\mathbf{x}$

$\mathbf{x}_i$ i-th column of matrix $\mathbf{X}$

$\mathbf{x}^i$ i-th row of matrix $\mathbf{X}$

x_{ij} or $(\mathbf{X})_{ij}$ element (i,j) of matrix $\mathbf{X}$

$\mathbf{X}^T$ transpose of matrix/vector $\mathbf{X}$

$\mathbf{X}^H$	complex conjugate transpose of matrix/vector $\mathbf{X}$ (Hermitian transpose)
$\mathbf{X}^*$	complex conjugate of matrix $\mathbf{X}$, i.e., component wise conjugation
$\mathbf{X}^{-1}$	inverse of matrix $\mathbf{X}$
$\mathbf{X}^\dagger$	pseudo-inverse of matrix $\mathbf{X}$, defined by $\left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product, of $M \times N$ matrix $\mathbf{A}$ and $X \times Y$ matrix $\mathbf{B}$, the result is an $MX \times NY$ dimensional matrix (see Appendix A.1.9)
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbf{0}_N$	$N \times N$ all zeros matrix
$\mathbf{0}_{N \times M}$	$N \times M$ all zeros matrix
$\mathbf{1}_N$	$N \times N$ all ones matrix
$\mathbf{1}_{N \times M}$	$N \times M$ all ones matrix
$\det(\mathbf{X})$	determinant of matrix $\mathbf{X}$
$\mathrm{tr}(\mathbf{X})$	trace of matrix $\mathbf{X}$, i.e., the sum of its diagonal elements
$\mathrm{diag}(\mathbf{x})$	diagonal matrix with the elements of vector $\mathbf{x}$ on its diagonal
$\mathrm{diag}(\mathbf{X})$	a column vector containing the diagonal elements of matrix $\mathbf{X}$

1

Introduction

1.1 Communication Trends

During the 19th and 20th century, the way of communication underwent revolutionary changes. While in the earlier ages the communication mainly took place from mouth to mouth or by sending letters, the introduction of the telegraph, the telephone, the fax machine, and the later transition to mobile phone services hugely improved the connectivity. Now, at the beginning of the 21st century, a transition that might turn out even more revolutionary is taking place as the Internet and other data communication applications move into the wireless domain. *Ubiquitous connectivity* (i.e., connectivity anytime and everywhere) to the Internet, to company's Intranets, or to other data services is creating room for applications that might not even be thought of today.

Regarding the latter transition, it is very interesting to observe the following two recent trends. Firstly, the amount of Internet (data) traffic in the United States is growing 300% per year and recently exceeded the amount of voice traffic, as is shown by L.G. Roberts in [95]. His findings are depicted in Figure 1-1 where the amount of traffic in bits per second is depicted versus the time in years. The historical data to 1995 are obtained from the United States National Science Foundation (NSF) and from the Advanced Research Projects Agency (ARPA) which founded ARPANET, the predecessor of Internet. Note that the annual growth of 300% cannot be maintained but will (most likely) saturate to a fixed percentage of the Gross Domestic Product (GDP). The main conclusion, however, is that the amount of data traffic surpassed the amount of voice traffic. When we would have the figures of the global traffic to our disposal, we would not be surprised to observe the same trend.

Secondly, the International Telecommunication Union (ITU) forecasted in its world telecommunication development report of 2002 that the number of mobile voice subscriptions would exceed the number of fixed voice subscriptions in and beyond 2002. The results of this study are given in Figure 1-2 in which the global number of

subscriptions (in millions) is depicted against the time in years. So, we can conclude that recently wireless/mobile voice surpassed wired/fixed voice in terms of number of subscriptions and, consequently, most likely also in terms of amount of traffic.
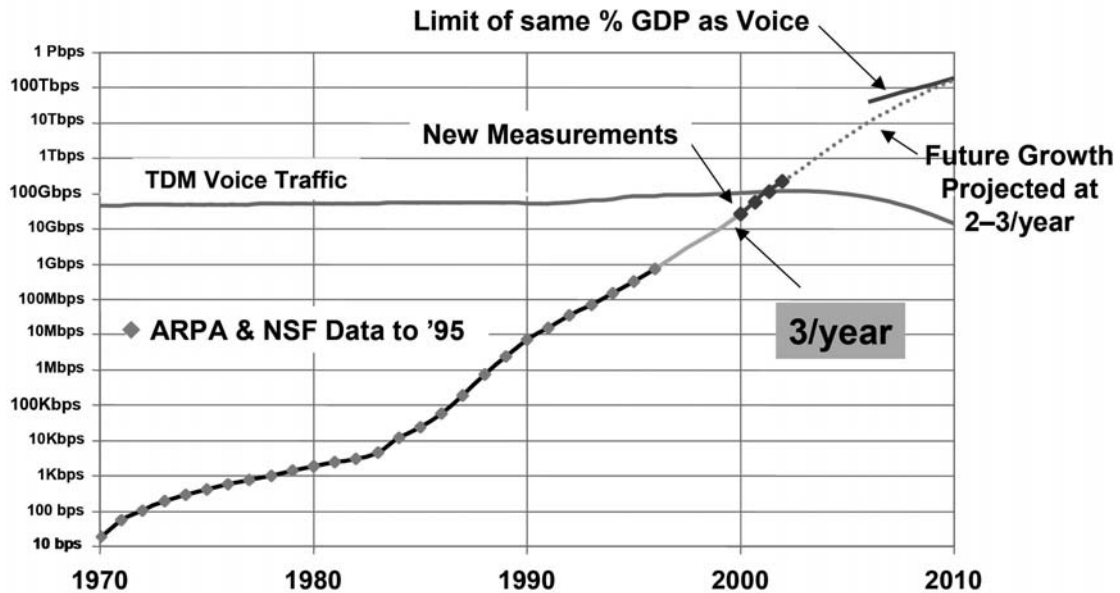
Figure 1-1: Historical and forecasted U.S. Internet traffic ([95]).
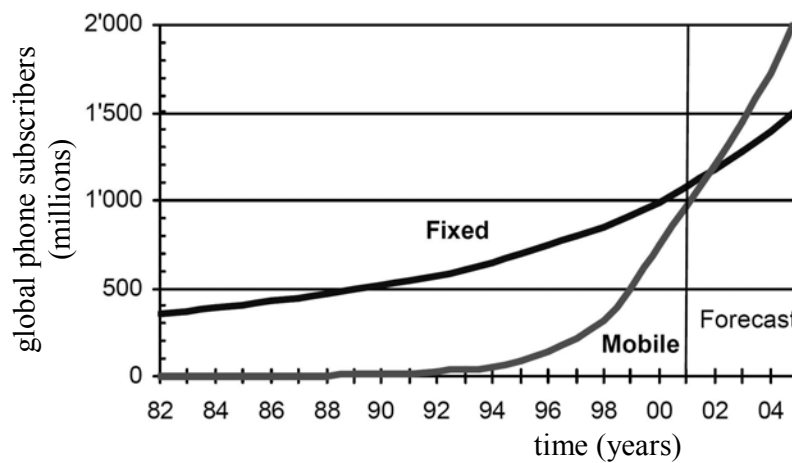
Figure 1-2: Mobile and fixed telephone subscribers worldwide, 1982 – 2005 ([59]).

When combining above two trends, a logic consequence must be that the amount of *wireless* data traffic will overtake the amount of *wired* data traffic (at least from an end-user perspective). This statement is being supported by the increasing demand for augmented capacity, data rates, and data services due to:

-   the tremendous momentum in wireless technology created both by the successful deployment of second generation mobile systems, e.g., GSM (including the quest for cheaper, smaller and more power-efficient handsets), and that of wireless data systems such as Wireless Local Area Networks (WLANs).

-   the increasing demand on wireless services, both for voice and data communications. In particular the demand for multimedia services such as video-on-demand, downloading music and movies, video conferencing, etc., is expected to diversify services and increase the volume of data traffic. As a result, emerging wireless/mobile networks are more and more networks that can integrate both voice and data services, opposed to the traditional voice-oriented networks. In general these emerging networks are operating on a packet basis (i.e., packet-switched) instead of setting up an end-to-end connection (i.e., circuit-switched).

-   the growing demand for flexibility and ubiquity. Communication services are expected to be available anytime and everywhere. Moreover, ubiquitous services aim to expand the objects of communication services, which have mostly been limited to humans so far, to everything and anything. In principle, everything or anything that moves is a potential object for mobile/wireless communication. For example, extremely small wireless chips may be attached to all products in a store to facilitate automatic billing of products that are taken to the cash desk (see, e.g., the efforts of the Radio Frequency Identification (RFID) association: [91]). Another example is the trend of the workplace becoming increasingly mobile. Ultimately, the worker should be able to log onto the company's Intranet, anytime and everywhere.

-   the continued scaling of Integrated Chips (IC) technology, allowing for more low-cost, power-efficient computing and resulting in increased integration and complex systems-on-a-chip.

Hence, it is obvious that the main goals in developing next generations of wireless communication systems (still) are increasing the link throughput (i.e., bit rate) and the network capacity. Since equipment cost (at least for the near future) and radio propagation conditions appear to limit the realm of wireless and mobile systems to the range around 1 GHz to 6 GHz, the available frequency spectrum is limited. So to fulfil above goals, future systems should be characterised by improved spectral efficiency.

Research in the information theory, performed in the early nineties, has revealed that important improvements in spectral efficiency can be achieved when multiple antennas are applied at both the transmitter and receiver side, especially in rich-scattering environments. This has been shown for wireless communication links in both narrowband channels ([36])[1] as well as wideband channels ([93]), and it initiated a lot of research activity to practical communication schemes that exploit this spectral-efficiency enhancement. The resulting multiple-transmit multiple-receive antenna, i.e., Multiple-Input Multiple-Output (MIMO), techniques can basically be split into two groups: Space-Time Coding (STC) ([5, 82, 116]) and Space Division Multiplexing (SDM) ([36, 93, 132]). STC increases the robustness/ performance of the wireless communication system by transmitting different representations of the same data stream (by means of coding) on the different transmitter branches, while SDM achieves a higher throughput by transmitting independent data streams on the different transmit branches simultaneously and at the same carrier frequency. In case of STC, advanced signal processing algorithms at the receiver combine the signals originated from the different transmitters to enhance the performance. In case of SDM, advanced signal processing algorithms at the receiver recover the parallel streams of

---

[1] Seeing the notation [36] as "abbreviation" of the exact text of the corresponding reference, when the exact text does not fit directly in the sentence, we will place it between brackets.

data that are mixed-up in the air. The latter technique usually requires multiple receive antennas, too, to ensure adequate performance.

The highest spectral efficiency gains are (on average) achieved when the individual channels from every transmit antenna to every receive antenna can be regarded to be independent. In practice this is the case in rich-scattering environments, with, preferably, no direct communication path (i.e., Line Of Sight (LOS) path) present between transmitter and receiver. So, especially for enhancement of the throughput of wireless applications in rich-scattering environments, MIMO techniques are appealing. In general, MIMO can be considered as an extension to any Single-Input Single-Output (SISO), Single-Input Multiple-Output (SIMO), i.e., receiver diversity, or Multiple-Input Single-Output (MISO), i.e., transmit diversity, system operating in these environments. To narrow the overall picture and to get a more concrete feeling of the application area, an overview of current and future wireless (data) standards as function of data rate and typical "cell" radius is provided in Figure 1-3. The typical "cell" radii of the Wireless Local Area Network (WLAN) standards IEEE 802.11b and 802.11a/g indicate that they are usually deployed in an indoor environment, while the probability of having no direct communication path between transmitter and receiver is high. So, we can conclude that the deployment conditions of WLAN systems are most favourable for applying MIMO.
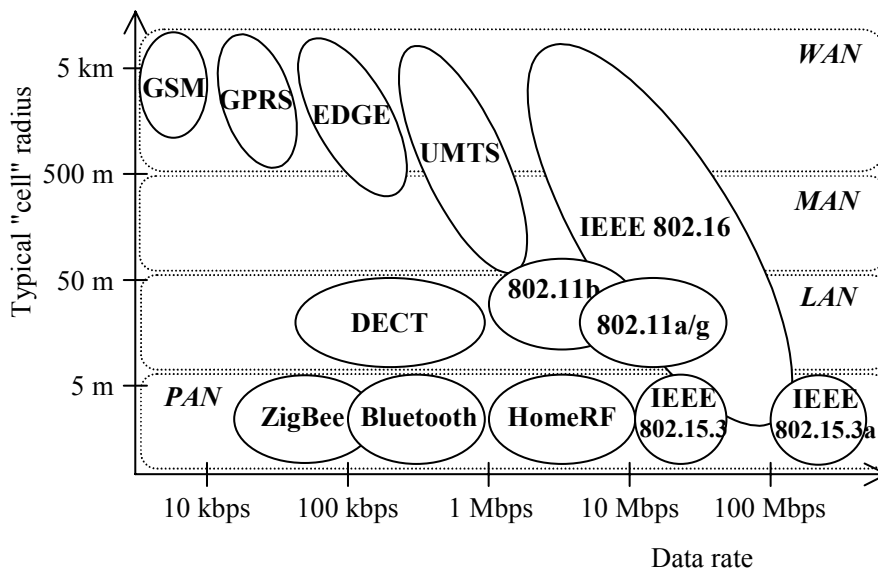
Figure 1-3: Overview of existing and future wireless data communication standards.

The standards of WLAN that currently gain the most momentum are IEEE 802.11a ([57]) and IEEE 802.11g ([56]). These two standards are based on Orthogonal Frequency Division Multiplexing (OFDM) ([23, 126, 141]). The main reason that OFDM was selected as basis for these standards is its capability to deal with the strong multipath propagation present in indoor propagation channels. In severe multipath, the multipath components add constructively and destructively and, as a result, the received signal can vary as a function of frequency, location and time. These variations are collectively referred to as *fading* and can lead to severe distortion of the received signal. OFDM, however, can mitigate this problem efficiently, since in OFDM, essentially, a wideband frequency-selective fading channel is split up into multiple orthogonal narrowband frequency-flat fading channels (i.e., subchannels or subcarriers) of which each can be

equalised in a trivial way (see Section 5.2 for a more thorough explanation of OFDM). Combined with coding, this principle also results in robustness against narrowband interference. Moreover, the ability to include a proper guard interval between subsequent OFDM symbols provides an effective mechanism to handle Inter-Symbol Interference (ISI) caused by severe multipath propagation.

The robustness of OFDM against frequency-selective fading and the favourable properties of indoor radio channels for SDM techniques ([137]) lead to the very promising combination of OFDM SDM as potential solution to satisfy the main goals in developing next generations of wireless communication systems. As such, OFDM SDM techniques are attractive candidates for high data rate extensions of the IEEE 802.11a and 802.11g standards. As example the IEEE 802.11 Task Group 'n' (TGn) can be mentioned which is planning to define high-data rate WLAN extensions up to 250 Megabits per second (Mbps) ([56]). The main focus of this dissertation is this promising combination of the data rate enhancement of SDM with the relatively high spectral efficiency of OFDM in the context of WLAN.

Since OFDM is also a potential candidate for the emerging standard IEEE 802.15.3a and, furthermore, forms the basis of one of the air interfaces defined in IEEE 802.16, we believe that our results can be extended to these systems as well, provided the underlying propagation channels exhibit sufficiently rich multipath and do not change too rapidly over time or frequency. Otherwise additional channel tracking methods are required.

1.2   The History of WLAN

Since the beginning of the nineties, WLANs for the 900 MHz, 2.4 GHz and 5 GHz license-free ISM (Industrial, Scientific and Medical) bands have been available, based on a range of proprietary techniques ([121, 126]). In June 1997 the Institute of Electrical and Electronics Engineers (IEEE) defined an international interoperability standard, called IEEE 802.11 ([56]). This standard specifies a number of Medium Access Control (MAC) protocols and three different Physical Layers (PHYs). Two of these PHYs are based on radio communication and use the 2.4 GHz band and the other PHY uses infrared light. All three PHYs support a data rate of 1 Mbps and optionally 2 Mbps.

User demand for higher bit rates and the international availability of the 2.4 GHz band has spurred the development of a higher speed extension to the 802.11 standard. In July 1998, a new standard was defined, named IEEE 802.11b, which describes a PHY providing a basic rate of 11 Mbps and a more robust rate, i.e., a fall-back rate, of 5.5 Mbps. Current widely-available products support both the 11 and 5.5 Mbps modes as well as the 1 and 2 Mbps modes (see, e.g., Figure 1-4). Meanwhile, in Europe, the European Telecommunication Standards Institute (ETSI) specified its own WLAN standard, called HIPERLAN/1 ([31]), which defines data rates ranging from 1 Mbps to 20 Mbps. In contrast to the IEEE 802.11b standard, no commercial products have been developed that support the HIPERLAN/1 standard.

Motivated by the demand for even higher data rates and by the opening of new unlicensed spectrum in the 5 GHz band for the use of a new category of equipment called Unlicensed National Information Infrastructure (UNII) devices ([34]), a new IEEE 802.11 working

group, called Task Group 'a' (TGa), started working on third generation (3G) WLANs. In July 1998, this group selected Orthogonal Frequency Division Multiplexing (OFDM) as transmission technique for the newly available spectrum in the 5 GHz band. In 2000, the standard was ratified and called IEEE 802.11a. It defines data rates between 6 and 54 Mbps ([57]). To make sure that these data rates are also available in the 2.4 GHz band, mid 2003 IEEE standardisation group finalised a similar standard for this band named IEEE 802.11g ([56]).

Figure 1-4: An IEEE 802.11b wireless LAN PC card.

Following the IEEE standardisation effort for the 5 GHz band, similar activities were started in Europe by a new ETSI working group named Broadband Radio Access Networks (BRAN) and in Japan by the MMAC group. BRAN was working on the next-generation HIPERLAN known as HIPERLAN/2 ([30]). The Multimedia Mobile Access Communication (MMAC) project is a cooperation of Japanese equipment manufacturers, service providers, and the Japanese Ministry of Post and Telecommunications ([76]). Following the selection of OFDM by the IEEE 802.11a standardisation group, both the ETSI BRAN and MMAC working groups adopted OFDM for their PHY. The three standardisation groups have worked in close co-operation since then to ensure that the standards are harmonised as much as possible thereby enabling equipment to be compatible worldwide. The main differences, however, are in the way the Medium Access Control (MAC) is defined. For instance, the HIPERLAN/2 MAC is based on a centralised protocol with Quality of Service (QoS) capabilities, including priority rules and provisions to ensure traffic-dependent maximum delays are not exceeded, whereas the IEEE 802.11 MAC is based on a random access protocol (i.e., decentralised). The QoS requirement of multimedia applications in particular urged the IEEE 802.11 body to also include QoS capabilities in their standard. This is currently pursued in IEEE 802.11 Task Group 'e' ([56]).

Based on the commercial availability of the higher data rate IEEE 802.11a and IEEE 802.11g products and the demand for high data rates, and building on the tremendous success of IEEE 802.11b products, it is expected that the former will soon surpass the latter in terms of sold volumes per month. In fact, in the second quarter (Q2) of 2003 9.9 million

Wireless-Fidelity (Wi-Fi)[1] certified WLAN units were shipped ([58]). In Q1 of 2003 6.9 million units entered the WLAN market. For 2002 overall, 18.7 million units were shipped, a 90% growth over 2001. From the 9.9 million units in Q2 of 2003, 1.7 million were IEEE 802.11g products (and almost all other 8.2 million units were 802.11b products), whereas the amount of IEEE 802.11g units shipped in 2002 was only marginal.

The growing success of the WLAN product family together with the demand for even higher bit rates confirms the need for research to high data-rate extensions for WLANs. Based on the argument that the success of new products also depends on its capability to be coexistent and interoperable with current standards, it appears to be logical to restrict this research to high data-rate extensions for OFDM. Together with the arguments of the previous section, this once more confirms the high potential of the combination of SDM and OFDM. The concept of SDM is explained in more detail in the next section.

1.3   Space Division Multiplexing

As already mentioned in Section 1.1, exploiting the spatial dimension by applying multiple antennas at both sides of the communication link is seen as a promising solution to significantly increase the bandwidth efficiency. Information theoretical research has namely revealed that the multipath wireless channel is capable of enormous capacities, provided that the multipath scattering is sufficiently rich ([35, 36, 93, 94, 98, 142]). The multipath scattering can be properly exploited through the use of an appropriate processing architecture. The diagonally-layered space-time architecture proposed in [35], known as Diagonal BLAST (Bell Laboratories Layered Space Time) or D-BLAST, is such an approach. See Figure 1-5a for a schematic representation of its transmission structure (where $s_p[n]$ denotes the n-th symbol originating from the p-th transmitter branch – before the cycling operation). The detector of this diagonal approach is, however, very complex and hard to implement. Therefore, a simplified version of BLAST, known as Vertical BLAST or V-BLAST was proposed in [144]. Note that "vertical" in V-BLAST does not denote the way the parallel data streams are encoded (in general, this is done "horizontally", see Figure 1-5b), but it refers to the way the detection at the receiving end is performed, namely, vertically, i.e., per time instant). In Bell Labs, a prototype with 12 transmit and 15 receive antennas and with V-BLAST detection was built by which it was demonstrated that bandwidth efficiencies up to 70 bits/s/Hz can be achieved in an indoor propagation environment at realistic SNRs and error rates ([88]).

The techniques based on multiplexing transmit signals over multiple antennas, i.e., over space, can be captured under the more general term Space Division Multiplexing (SDM) or Space Division Multiple Access (SDMA). SDM techniques exploit the spatial dimension by using multiple antennas to transmit. Basically, these techniques simultaneously transmit different signals on different transmit antennas, at the same carrier frequency. These parallel streams of data are mixed-up in the air, but can be recovered at the receiver by using advanced signal processing algorithms, which usually require multiple receive antennas, too, to ensure adequate error-rate performance. The difference between SDM and SDMA is that the latter allows different users to transmit simultaneously on a single

---

[1] The Wireless-Fidelity (Wi-Fi) Alliance is a nonprofit international organisation formed in 1999 to certify interoperability of WLAN products based on IEEE 802.11 specification (http://www.wi-fi.com/).

antenna each, whereas in SDM a single user transmits simultaneously on multiple antennas. Hybrid schemes can also be envisioned.

(a)

(b)
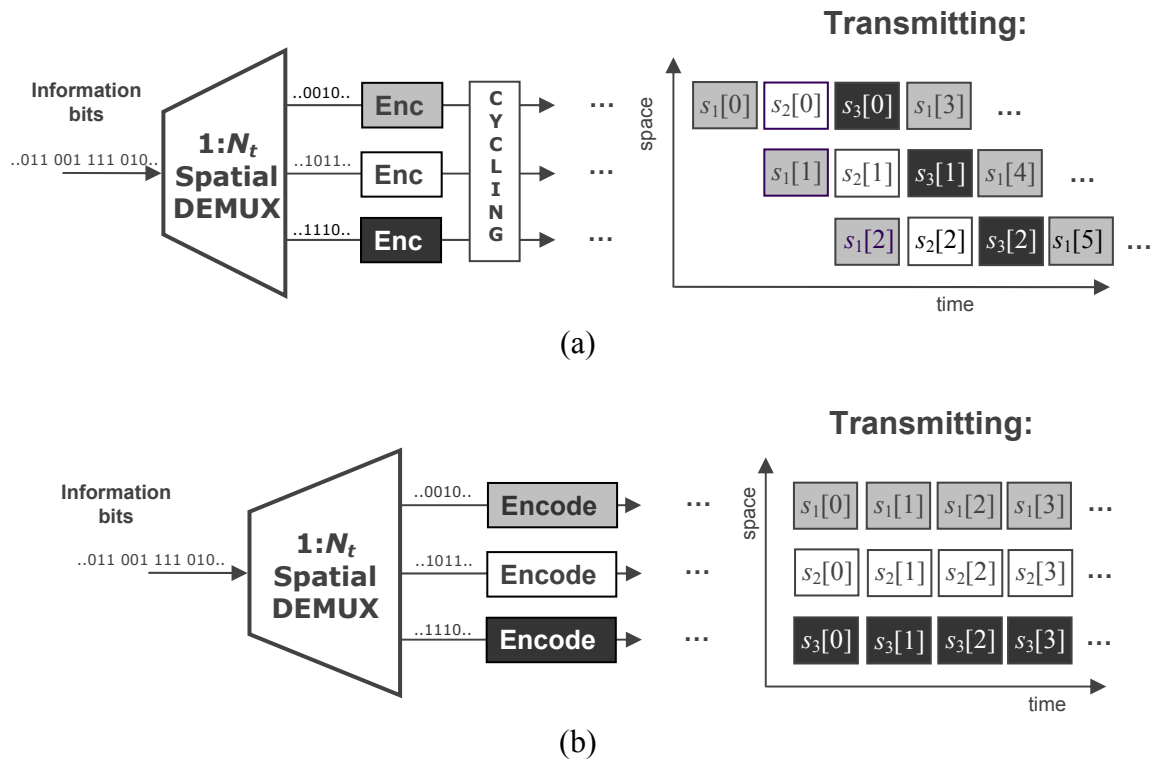
Figure 1-5: Transmission scheme of D-BLAST (a) and V-BLAST (b).

One can naturally ask in which way SDM(A) techniques differ from traditional multiple access techniques. Some of these differences are worth pointing out ([144]): First, unlike code-division or other spread-spectrum multiple access techniques, the total channel bandwidth utilised by an SDM(A) system is only slightly higher than the symbol rate, i.e., similar to the bandwidth required by a conventional single-carrier transmission technique like Amplitude Modulation (AM). Second, unlike Frequency Division Multiple Access (FDMA), each transmitted signal occupies the entire system bandwidth. Finally, unlike Time Division Multiple Access (TDMA), the entire system bandwidth is used simultaneously by all of the transmitters all of the time. These differences together are precisely what give SDM(A) the potential to realise higher bandwidth efficiencies than the other multiple-access techniques.

After the theoretical proof of the MIMO gains by Foschini in [36], various measurement systems and prototypes were built to verify its potential gains in practice as, among others, reported in [1, 68, 88, 136]. Moreover, recently the first successful implementations were announced in [3] and [37].

1.4 Framework and Goals

The research reported in this dissertation was conducted within the framework of a Dutch cooperative research project called B4 and funding was provided by Agere Systems, The

Netherlands. B4 ("BraBant BreedBand"; North-Brabant is a Dutch province and "breedband" is Dutch for broadband) is a research alliance initiated by KPN (a Dutch telecom provider), Lucent Technologies and the Eindhoven University of Technology (TU/e) in the area of broadband communication technologies ([12]). Main goals of the alliance are to further enhance the strong position of The Netherlands in this field in response to the explosive growth of the telecommunication technologies market, and to consistently create innovations for future products and applications. An essential element of the alliance is therefore joint pre-competitive research in broadband networks, including fibre and wireless technologies and services, from the conceptual phase up to and including verification and exploitation in pilot trials involving students using laptops. Its activities are organised in a number of task forces, in which specific partners support the three initiating organisations.

In one of these task forces named Broadband Radio@Hand, Agere Systems, TNO Telecom (formerly KPN Research), Philips, and the TU/e have joined their forces to investigate how future wireless (data) networks (based on UMTS and/or WLAN) can be realised that guarantee bandwidth and quality on demand, at the office and on the road. The primary objective of Broadband Radio@Hand is the development of a new state-of-the-art for (wireless and radio-over-fibre) networks with high capacity and corresponding services, to facilitate the development of new multimedia services within The Netherlands. The strength of the consortium is that the partners cover the complete field that is required to pursue this development; from system design, antenna knowledge, channel modelling, modulation and detection techniques, signal processing, RF-circuit and ASIC design, network technologies, to know-how on IC-production. More concretely, the main goals of the project are to substantially improve:

-   the practical achievable system capacity,
-   the transmission quality (i.e., QoS).

An important constraint is that above points should be realised with reasonable investment efforts and with low resulting operational cost.

Within this framework and based on the trends identified in Section 1.1, the focus of this dissertation is the development of high data rate WLANs for indoor scenarios, which, based the IEEE 802.11a/b/g parameters, may also be applicable in low-range low-mobility urban (outdoor) scenarios. Two fundamental problems complicate the design of high data rate indoor networks. Firstly, regulatory restrictions on bandwidth and transmit power exist in the frequency bands exploited by WLANs and inherently limit the capacity achievable with conventional SISO techniques. As mentioned before, SDM is a very promising technology to overcome this problem. Secondly, the indoor propagation channel exhibits strong multipath propagation. As we already explained, OFDM can effectively deal with severe multipath. Furthermore, due to the high cost and deployment in wireless communications, it is crucial that next-generation standards are a logical evolution of current standards and, as such, are coexistent and (preferably) backwards compatible. Based on above arguments, the combination of SDM and OFDM is regarded as a very promising candidate for next-generation WLANs. In this context, the main objectives of this dissertation are:

-   get a more fundamental understanding of MIMO,

- introduce a good and useful wideband MIMO channel model (for indoor environments),
- evaluate and find efficient SDM detection techniques in terms of performance and complexity,
- evaluate these techniques in combination with OFDM, e.g., by performing simulations and making use of the proposed wideband MIMO channel model,
- verify the SDM OFDM combination in real-life channels by means of a test system. This also requires tackling of radio imperfections encountered in practical systems.

The next section gives an overview of the work that is reported in this dissertation and that was performed based on above objectives.

1.5   Survey of this Dissertation and Contributions

In this section, a general preview is given of the different chapters of this dissertation. Furthermore, the main contributions are pointed out. The content of this dissertation follows the logical order from the fundamental understanding of MIMO, the theoretical analysis of MIMO and MIMO OFDM, to practical measurements with a MIMO OFDM test system based on WLAN parameters. Below, a short summary is presented per chapter and the main contributions are given by means of bulleted indices.

In Chapter 2, the effect of one of the simplest SDM algorithms called Zero Forcing (ZF) on the antenna array pattern is evaluated. It is shown that in free space, i.e., when the MIMO channel only consists of Line Of Sight (LOS) components, a weight generated to retrieve the signal from a particular TX antenna alters the antenna array pattern such that a null is placed in the *direction* of the unwanted TX antenna(s). When reflections are present, however, a weight vector generated for this latter case places a "null spot" at the *location* of an unwanted TX antenna. Moreover, in the latter case, the wanted TX antenna turns out to be positioned in a local maximum, resulting in a maximum separation. This provides an intuitive explanation of the robustness of MIMO in rich-scattering environments.

- Based on a physical interpretation, an intuitive and fundamental explanation of the MIMO principle was found. It shows why MIMO achieves a higher spectral efficiency and stability in rich multipath scattering. This work was published in [131].

Chapter 3 describes the properties of the richly scattered propagation channel. The indoor propagation channel is characterised by rich multipath scattering due to the reflection of the transmitted electromagnetic waves on walls and objects in the environment. Based on these geometrical considerations, a stochastic wideband MIMO channel model is proposed. It is based on the Non Line Of Sight (NLOS) tapped delay line model commonly used for wideband MIMO simulations and extended with two critical impairments for MIMO, namely a LOS component and spatial correlation. By this model, the typically large number of channel parameters are captured by a very few carefully selected ones, in order to take into account the crucial properties of the propagation channel that impose the main constraints on the design of a WLAN.

- A stochastic wideband MIMO channel model was developed based on a tapped delay line and including key parameters of the propagation channel like fading depth, root

mean square (rms) delay spread, propagation loss, a LOS component, and spatial correlation. A number of narrowband models were shown to be specific cases of the introduced generic wideband model.

- To model Additive White Gaussian Noise (AWGN) in the MIMO case, a specific constant-modulus orthogonal MIMO channel model was generated.

- Spatial fading correlation, representing the correlation between the various elements of a MIMO channel, is generally characterised by many parameters. For two commonly used performance measures, namely capacity and error-rate performance, we introduced a mathematical mapping of the many correlation parameters to one or two parameter(s) while maintaining the same performance. Based on this mapping a spatial correlation model was developed that allows for easy inclusion of spatial correlation in MIMO simulations. The strength of the model is that, by ranging the one or two parameter(s) from zero to one, all scenarios ranging from totally uncorrelated to fully correlated spatial fading can be considered. The mathematics, the model, and simulation results were submitted for publication in [133]. Earlier work was published in [134].

Chapter 4 discusses MIMO techniques suitable for channels with fading that is flat over frequency (i.e., narrowband techniques). First, an overview is given of the various MIMO algorithms presented in a vast amount of literature, including STC and SDM algorithms. It is shown that basically all these techniques can be mapped on a general structure including an encoder, a space-time mapper, and constellation mappers. This effort can be considered as a good starting point for a unified theory on MIMO. Second, a number of capacity definitions are given for MIMO systems with different properties. Based on these definitions, the outage Packet Error Rate (PER) is defined. Third, various SDM algorithms are described and their complexities are evaluated. For some algorithms, a theoretical error-rate analysis is presented to be able to verify the simulation results. Fourth, the described SDM algorithms are compared in terms of error-rate performance for systems without and with coding on top of the SDM scheme, and in terms of complexity. Fifth, the turbo principle is introduced to the SDM context and is called turbo SDM. In this scheme, the SDM mapping is seen as inner code and combined with some form of outer coding. This combination allows the receiver to iterate between the inner and outer code and, as a result, improve the error-rate performance substantially.

- The thorough study of different MIMO techniques resulted in the fact that we found a general structure on which basically all techniques can be mapped. This resulted in the presentation of a unified framework that can be envisioned as a good starting point for the development of a unified theory on MIMO.

- Different SDM algorithms were studied. Their complexity was calculated in terms of number of additions and number of multiplications. Early results were presented in [129, 132]. Next to the complexity analyses, for some algorithms theoretical error-rate analyses were carried out and reported, and for a larger set of techniques, soft decision output values were defined. Soft values can be used by the decoder of the outer code to achieve a better performance, since soft values do not only provide the estimated value of a bit but also a measure for the reliability of that estimate.

- The described algorithms were implemented in MATLAB and extensive simulations were performed in order to compare their performance. The simulations were done for different antenna configurations, for various constellation sizes, for different channel properties (some of which also include spatial correlation or a LOS component), without and with additional coding. Moreover, the simulations were verified with the findings of the theoretical error-rate analyses.

- Chapter 4 also reports the analysis that was performed to validate the proposed compact representation of spatial correlation (by means of the introduced simple model) with respect to the error-rate performance. The results were submitted for publication in [133].

- Independent from the work reported in [106, 123], the idea of adding turbo processing to coded SDM was born and published in [138]. This idea was further worked out and reported at the end of Chapter 4. Its performance was investigated by an evaluation tool developed for turbo codes named the EXIT chart method. Next to that, a turbo SDM scheme was programmed in C++ and a number of simulations were performed and compared with other MIMO techniques.

In Chapter 5, first the principle of OFDM is explained. Second, the combination of MIMO and OFDM is described. The core idea is that the wideband frequency-selective MIMO channel by means of the MIMO OFDM processing is transferred to a number of parallel flat-fading MIMO channels. Third, the wideband MIMO capacity is determined and the corresponding outage PER is defined. Fourth, a theoretical Space-Frequency analysis is presented based on the Pairwise Error Probability (PEP) to better understand the achievable performance and to deduce proper design criteria for MIMO OFDM systems. Fifth, two practical coding schemes are described; one is called Joint Coding (JC) and the other Per-Antenna Coding (PAC). For the latter, based on the results of Chapter 4, an efficient decoding scheme is proposed that has a much lower complexity than the optimal decoding scheme, but achieves comparable performances. This is shown by an extensive set of simulations based on IEEE 802.11a parameters.

- A concise tutorial on OFDM was developed describing its principle, how multipath distortion is handled, the main advantages of OFDM, and the general block diagram of an OFDM transceiver.

- The basis of a unified framework for MIMO introduced in Chapter 4 was extended with OFDM resulting in a general structure for a space-time-frequency scheme. Based on the main goal of this dissertation, enhancing the throughput of WLANs, we saw (and still see) the combination of SDM with OFDM as a very promising approach. The publications [129] and [137] are regarded to be among the first that presented the MIMO OFDM concept in the WLAN context.

- We introduced a MIMO OFDM signal model using a compact matrix notation. The strength of this matrix signal model is that it allows for mathematical derivations for MIMO OFDM systems, such as the in Section 5.5 introduced capacity definition for wideband channels and its corresponding outage PER. Also the space-frequency analysis performed in Section 5.6 and pointed out in the next bulleted index is done with this concise model. Furthermore, it can be used for impairment studies such as

timing offset and frequency offset analyses (see [102, 103, 135]), phase noise analysis, etc.

- A theoretical space-frequency error-rate analysis was performed in which the analysis of [20] was extended to include next to the spatial correlation of the receive side also the spatial correlation at the transmit side. This analysis can be used to design space-frequency codes/schemes that are not only performing well in idealised situations but are also robust in scenarios where spatial correlation is present. In order to simplify the space-frequency code design, we added the idea of subcarrier grouping described in [147] to our analysis. Moreover, we showed (based on the ideas of [4]) that under realistic conditions dedicated space-frequency code design rules can be overruled by the established Euclidean distance criterion.

- For transmission schemes in which the encoding is done per transmitter branch, we introduced an efficient decoding scheme with a low complexity and called it Per-Antenna-Coding Successive-Interference-Cancellation (PAC SIC). This scheme performs closely to the optimal performing scheme at the expense of a manageable latency. It was published in [130].

- The proposed (coded) SDM OFDM algorithms were programmed in MATLAB and an extensive set of simulations based on WLAN parameters was performed in order to evaluate their performance. The simulations were performed for various antenna configurations, rms delay spreads, constellation sizes, coding rates, and NLOS and LOS scenarios. Parts of the results were published in [130, 135, 137].

In the preceding chapters of Chapter 6, system impairments are assumed to be negligible. Practical implementations of digital communication systems, however, have to deal with impairments such as frequency offset, timing offset, phase noise, IQ imbalance, DC offset, etc. Therefore, in order to validate the implementability of MIMO OFDM algorithms, including the effect of impairments, a test system with three transmit and three receive antennas was built within Agere Systems, The Netherlands. Chapter 6 reports on the design choices of the preamble, how the time and frequency synchronisation is performed, how the propagation channel is estimated, and how the synchronisation is tracked. Furthermore, the test system is described in detail and the results of a set of performed measurements based on IEEE 802.11a parameters are presented.

- Early results of the work presented in this dissertation were the initiator for building a test system with three transmit and three receive antennas within Agere Systems, The Netherlands, that operates in the license free 5.x GHz band.

- A preamble design and corresponding processing for a MIMO OFDM WLAN application were developed. The processing provides extensions of the way impairments are handled in OFDM to include MIMO. The proposed preamble design has the IEEE 802.11a preamble as basis and supports backwards compatibility. The strength of the presented preamble design and processing for MIMO OFDM is that they are straightforward extensions of those for OFDM. More enhanced solutions are also possible as described in 5 (co-)authored patents that are pending.

- Measurements were performed with the test system within the office building of Agere Systems, Nieuwegein, The Netherlands. By these measurements we demonstrated the concept of SDM OFDM in practice and showed successful transmissions of data rates up to 162 Mbps. These results are seen as the industry first demonstration of 162 Mbps based on MIMO OFDM in the WLAN context and attracted a lot of attention in the world press (see, e.g., [24]). The proposed preamble design and processing, and the measurement results were accepted for publication in [135].

Finally, Chapter 7 describes the major conclusions of this work and indicates some promising directions for future research.

2

Physical Interpretation of MIMO Transmissions

2.1 Introduction

In the previous chapter, we already introduced the MIMO concept as a communication technique that exploits the spatial dimension by applying multiple antennas at both the transmitter and receiver side. This MIMO principle has been thoroughly studied by mathematical evaluations in literature, but to the author's knowledge, it has never been explained by a physical interpretation. In this chapter, such a physical interpretation is presented providing a fundamental understanding of the MIMO concept in radio communication. Moreover, it gives an intuitive explanation why the spectral efficiency and stability of MIMO are especially high in rich-scattering environments.

In Section 2.2, the MIMO communication principle is explained and a detection technique called Zero Forcing (ZF) is described. In Sections 2.3, 2.4, and 2.5, the effect of the environment on the antenna array pattern of the receiver (after ZF detection is applied) is evaluated by considering in each section a different number of reflecting planes. Section 2.6 describes the effect on the antenna array patterns of the receiver when the receiver does not perfectly know the communication channel, but only has a noisy estimate of the channel. Finally, in Section 2.7 conclusions are drawn.

2.2 Multiple-Input Multiple-Output Communication

Consider a wireless communication system with N_t transmit (TX) and N_r receive (RX) antennas. The idea is to transmit different streams of data on the different transmit antennas, but at the same carrier frequency. The stream on the p-th transmit antenna, as function of the time t, will be denoted by $s_p(t)$. When a transmission occurs, the transmitted signal from the p-th TX antenna might find different paths to arrive at the q-th RX antenna, namely, a direct path and indirect paths through a number of reflections. This principle is called *multipath*. Suppose that the bandwidth B of the system is chosen such that the time

delay between the first and last arriving path at the receiver is considerably smaller than $1/B$. In this case the system is called a *narrowband* system. For such a system, all the multipath components between the *p*-th TX and *q*-th RX antenna can be summed up to one term, say $h_{qp}(t)$. Since the signals from all transmit antennas are sent at the same frequency, the *q*-th receive antenna will not only receive signals from the *p*-th, but from all N_t transmitters. This can be denoted by the following equation (in this chapter, the additive noise at the receiver is omitted for clarity, but will be introduced in Section 3.4)

$$x_q(t) = \sum_{p=1}^{N_t} h_{qp}(t) s_p(t). \tag{2.1}$$

To capture all N_r received signals into one equation, the matrix notation can be used. With

$$\mathbf{s}(t) = \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_{N_t}(t) \end{pmatrix}, \quad \mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_{N_r}(t) \end{pmatrix} \text{ and } \mathbf{H}(t) = \begin{pmatrix} h_{11}(t) & h_{12}(t) & \cdots & h_{1N_t}(t) \\ h_{21}(t) & h_{22}(t) & \cdots & h_{2N_t}(t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_r1}(t) & h_{N_r2}(t) & \cdots & h_{N_rN_t}(t) \end{pmatrix}, \tag{2.2}$$

this results in

$$\mathbf{x}(t) = \mathbf{H}(t)\mathbf{s}(t). \tag{2.3}$$

A schematic representation of a MIMO communication scheme can be found in Figure 2-1.
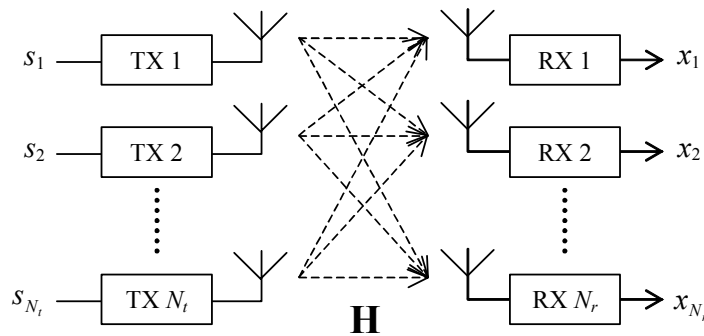
Figure 2-1: A schematic representation of a MIMO communication system.

Mathematically, a MIMO transmission can be seen as a set of equations (the recordings on each RX antenna) with a number of unknowns (the transmitted signals). If every equation represents a unique combination of the unknown variables and the number of equations is equal to the number of unknowns, then there exists a unique solution to the problem. If the number of equations is larger than the number of unknowns, a solution can be found by performing a projection using the *least squares* method ([113]), also known as the *Zero Forcing* (ZF) method (see Section 4.6). For the symmetric case (i.e., $N_t = N_r$), the ZF solution results in the unique solution.

Suppose the coefficients of the unknowns are gathered in the channel matrix $\mathbf{H}(t)$ and the number of parallel transmit signals (unknown variables) equals to the number of received

signals (equations), i.e., $N_t = N_r$, then the equations are solvable when $\mathbf{H}(t)$ is invertible. Under this condition, the solution of (2.3) can be found by multiplying both sides with the inverse of $\mathbf{H}(t)$:

$$\mathbf{H}^{-1}(t)\mathbf{x}(t) = \mathbf{H}^{-1}(t)\mathbf{H}(t)\mathbf{s}(t) = \mathbf{I}_{N_t}\mathbf{s}(t) = \mathbf{s}(t), \qquad (2.4)$$

where $\mathbf{I}_N$ is the $N \times N$ dimensional identity matrix. Thus, to estimate the transmitted signals at the receiver, the vector $\mathbf{x}(t)$ must be multiplied by the inverse of the channel matrix $\mathbf{H}(t)$. To that end, the channel matrix must be known at the receiver. This can be done by, e.g., sending a training sequence, that is known to the receiver, to train the channel.

In the next sections, a system with two transmit antennas ($N_t = 2$) and two receive antennas ($N_r = 2$), or shortly, a 2×2 system is considered. It will be assumed that the receiver perfectly knows the channel. With this assumption, we may write the two solutions $s_1(t)$ and $s_2(t)$ as

$$s_1(t) = \mathbf{w}^1(t)\mathbf{x}(t), \qquad (2.5)$$

$$s_2(t) = \mathbf{w}^2(t)\mathbf{x}(t), \qquad (2.6)$$

where $\mathbf{w}^i(t)$ denotes the weight vector that is applied at the receiver to estimate the i-th transmitted signal and can be shown to be equal to the i-th row of $\mathbf{H}^{-1}(t)$. In the next sections, for a specific antenna setup in different environments (with and without reflections), we will determine the channel coefficients and the weights, and show what the influence of these weights is on the RX antenna array pattern.

2.3 Free Space Aspects

A free-space scenario is considered where a 2×2 system is placed in an (artificial) environment where no reflections occur. Both the antenna set-up and the environment are assumed static and, therefore, the channel coefficients are constant over time. Hence, the time index can be omitted. Since no reflections take place, the channel coefficient between the p-th TX antenna and the q-th RX antenna, h_{qp}, only consists of the direct path between these antennas. Denote the length of this path by d_{qp} in metres, then both the power and phase of the channel coefficient are a function of d_{qp}. Since the system is operation in free space, the power at a distance d_{qp} from the p-th transmitter is given by the Friis free space equation ([92]):

$$P_r(d_{qp}) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d_{qp}^2 L_s} \text{ Watts}, \qquad (2.7)$$

where P_t is the transmitted power per TX antenna, G_t and G_r are, respectively, the transmitter and receiver antenna gains, L_s is the system loss factor not related to propagation and λ is the wavelength in metres. In the next analysis, we assume that there is no system loss ($L_s = 1$) and that unity gain antennas are used ($G_t = G_r = 1$). The phase at a distance d_{qp} equals $-2\pi d_{qp}/\lambda$ rad. This results in the following channel coefficient

$$h_{qp} = \sqrt{\frac{P_t \lambda^2}{(4\pi)^2 d_{qp}^2}} \exp\left(-j2\pi \frac{d_{qp}}{\lambda}\right). \tag{2.8}$$

Once the four elements of the channel matrix **H** are known, the weights for the Zero Forcing MIMO processing can be determined. The weight vectors $\mathbf{w}^1$ and $\mathbf{w}^2$ are obtained as described in Section 2.2. We want to see what the effect of these weights is. To that end, a dummy antenna is placed at a given two dimensional spot (x,y) and the received vector as function of (x,y) is determined: $\mathbf{x}(x,y)$. This vector is multiplied by the weights $\mathbf{w}^1$ and $\mathbf{w}^2$, respectively, and we now can, e.g., show what the power is of the resulting signals as function of (x,y). These plots can be seen as the RX antenna array patterns after applying the weights.

Here, this is worked out for an antenna set-up as depicted in Figure 2-2. Assume that the TX antennas and RX antennas are centred on the y-axis, with an antenna spacing of respectively $d_{TX} = 1\lambda$ and $d_{RX} = 1\lambda$, furthermore, assume that the distance between the transmitter array and receiver array equals $D = 100\lambda$, and that the power per TX antenna equals 0.035 Watts[1]. Then, the channel matrix can be shown to be

$$\mathbf{H} = \begin{pmatrix} \sqrt{\frac{0.035}{(4\pi)^2 \cdot 10000}} \exp(-j2\pi \cdot 100) & \sqrt{\frac{0.035}{(4\pi)^2 \cdot 10001}} \exp(-j2\pi \cdot \sqrt{10001}) \\ \sqrt{\frac{0.035}{(4\pi)^2 \cdot 10001}} \exp(-j2\pi \cdot \sqrt{10001}) & \sqrt{\frac{0.035}{(4\pi)^2 \cdot 10000}} \exp(-j2\pi \cdot 100) \end{pmatrix}, \tag{2.9}$$

from which the weight vectors can be determined by taking the rows of the inverse of **H**.
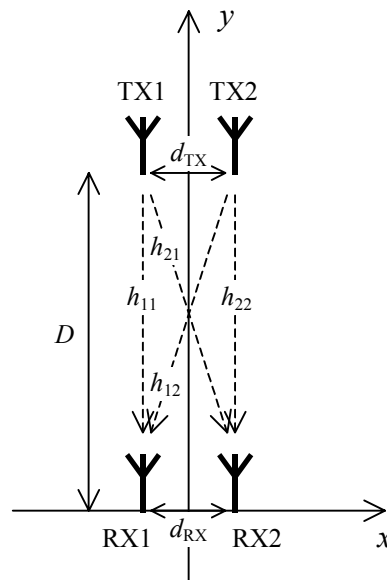
Figure 2-2: Antenna set-up.

Applying these weight vectors results in the RX antenna array patterns as given in Figure 2-3. The points in the plots are calculated using a grid in polar coordinates, with an angular

---

[1] This more or less equals 15 dBm which is the average TX power commonly used in WLAN products.

grid of ½·180/π·tan(½·d_{TX}/D) ≈ 0.143 degrees and a radius grid of 1λ. To smooth the plots, interpolation is applied. Note that the TX antenna positions are denoted by the white crosses and the RX antenna positions by the black ones. We clearly see that, when weight $\mathbf{w}^1$ is used, the signal from the second antenna (and all spots in line with that TX antenna and the receiver array) is suppressed, and vice versa when $\mathbf{w}^2$ is applied. Clearly, the undesired signal is forced to zero. Furthermore, it can be seen that the larger the distance between a given spot (x,y) and the receiver array, the weaker the signal that is received. This is the result of applying the free-space path loss model.

(a)   (b)
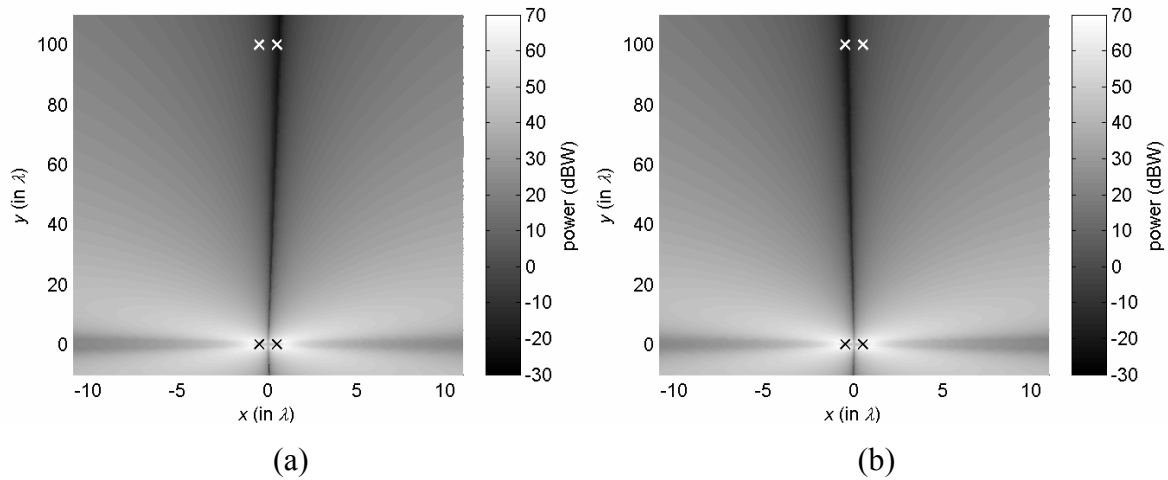
**Figure 2-3: RX antenna array patterns after applying the
first (a) and second (b) weight vector in free space.**

2.4 One Perfectly Reflecting Plane

Here, the scenario of the previous section is extended with one perfectly reflecting plane, parallel to the transmitter-receiver line. In addition to the direct paths of the free-space case, one indirect path per channel element has to be taken into account due to the reflection. At the receiver side, this indirect path can be seen as if it would be a direct path from the *image* of the transmitter, mirrored in the reflecting plane (see Figure 2-4). So, for the channel between the p-th TX and the q-th RX antenna this means that, besides the direct path, an extra path must be added, virtually being the direct path from the image of the p-th TX antenna to the q-th receiver (see Figure 2-4).

Using the same parameters as in Section 2.3 (d_{TX} = 1λ, d_{RX} = 1λ, D = 100λ and P_t = 0.035 Watts) and with the extra information that D_{refl} is chosen to be 8λ, the channel matrix and the weight vectors can be determined. The antenna patterns after applying both weights are given in Figure 2-5.

From these figures, we can see that the reflecting plane at x = 8λ causes an interference pattern. Again, we see that applying the right weight vector suppresses the signals from the antenna that is by this weight vector considered as interferer.

Figure 2-4: Antenna set-up with a perfectly reflecting plane. Only the extra paths that have to be taken into account in addition to the direct paths of Figure 2-2 are shown.

| (a) | (b) |

Figure 2-5: RX antenna array patterns after applying the first (a) and second (b) weight vector in a scenario with one perfectly reflecting plane at $x = 8\lambda$.

2.5   Two Perfectly Reflecting Planes

In the final scenario that is considered, another perfectly reflecting plane is added to the scenario of Section 2.4. Again, the following parameters are used: $d_{TX} = 1\lambda$, $d_{RX} = 1\lambda$, $D = 100\lambda$ and $P_t = 0.035$ Watts. Furthermore, we assume that the first reflecting plane is positioned at $x = 8\lambda$, whereas the other plane is positioned at $x = -6\lambda$. Since the two reflecting planes are parallel to each other, there will be paths that only arrive at the receiver after a multiple of bounces between the two planes. Here, we will only consider a maximum of one bounce and two bounces, respectively. The channel matrix and weight vectors can be determined for both cases. The RX antenna array patterns after application

of the weight vectors in case of a maximum of one and two bounces are shown, respectively, in Figure 2-6 and Figure 2-7.

From comparing these figures with Figure 2-5, it becomes clear that the more reflections occur, the more chaotic the interference patterns are. In Figure 2-6 and Figure 2-7, we can see that the undesired antenna is nulled with a spot, instead of with a beam (like in Figure 2-3), and that the desired antenna is (almost) located at a local maximum. This maximal separation between the wanted and unwanted antenna shows that the signals from both antennas can be treaded as uncorrelated (or independent). This observation speaks in favour of the robustness of MIMO systems in environments with many reflecting objects, i.e., *rich-scattering* environments.

(a)   (b)

Figure 2-6: RX antenna array patterns after applying the first (a) and second (b) weight vector in a scenario with two perfectly reflecting planes (at $x = -6\lambda$ and $x = 8\lambda$), where only paths with a maximum of one bounce are taken into account.

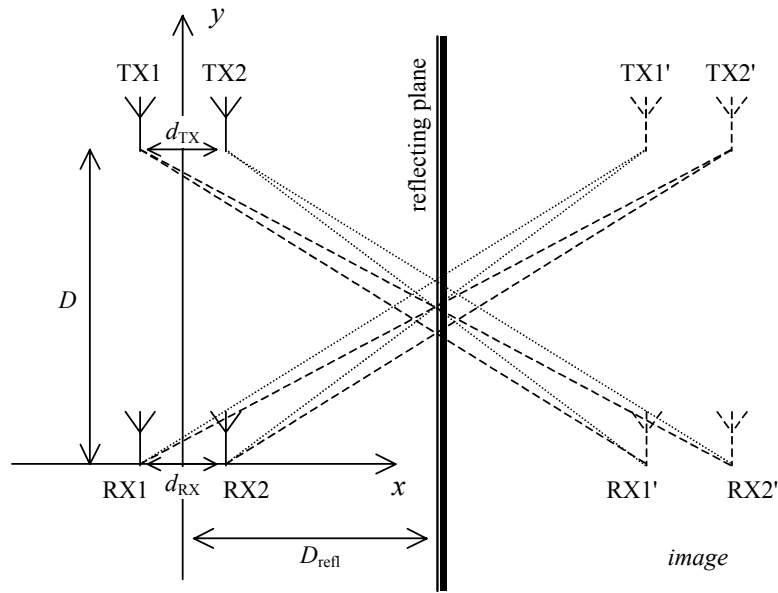(a)   (b)
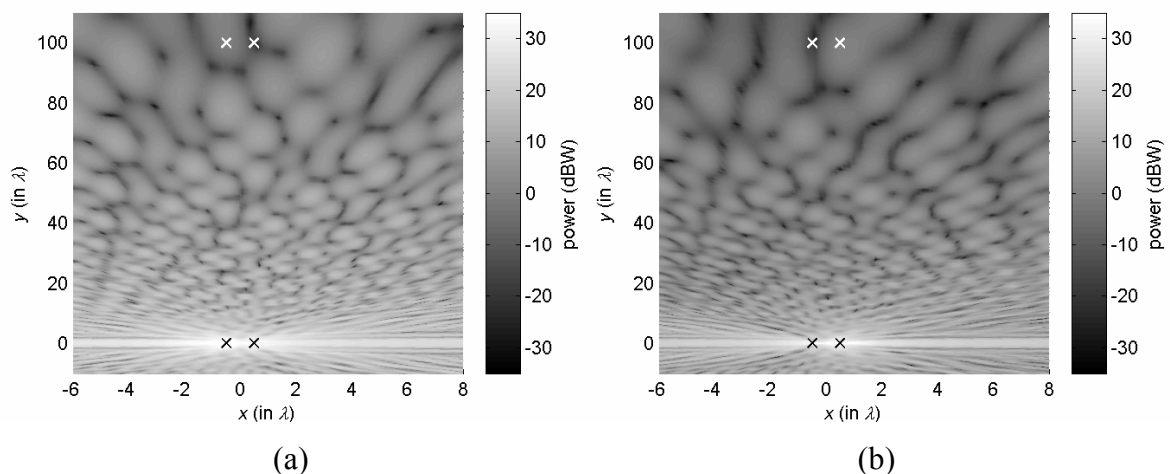
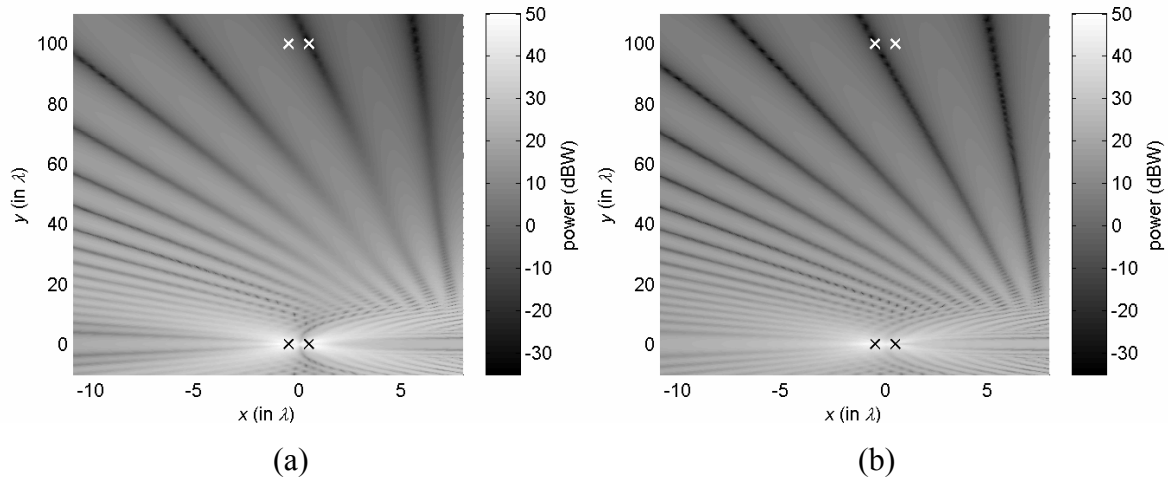**Figure 2-7: RX antenna array patterns after applying the first (a) and second (b) weight vector in a scenario with two perfectly reflecting planes (at $x = -6\lambda$ and $x = 8\lambda$), where only paths with a maximum of two bounces are taken into account.**

2.6   Channel Estimation Errors

The observation of the previous section that MIMO is more robust in rich-scattering environments can be confirmed by adding white Gaussian noise to the channel observation. This provides insight in the MIMO performance when the MIMO system experiences noise. More concrete, it illustrates how the antenna patterns are altered when the channel estimation is corrupted by noise. To include the influence of noise, we can add independent and identically distributed (i.i.d.) complex Gaussian noise to the four channel elements of the 2×2 cases of the previous sections. With an average noise power of 10% of the average channel element power (i.e., the Signal-to-Noise Ratio (SNR) = 10 dB), and the assumption that the average power per channel element is normalised to one, an example of the Additive White Gaussian Noise (AWGN) is given by

$$\begin{pmatrix} 0.0091 - j \cdot 0.0186 & -0.0826 - j \cdot 0.0933 \\ 0.0464 - j \cdot 0.0858 & -0.0326 + j \cdot 0.0658 \end{pmatrix}. \tag{2.10}$$

When adding this noise to the channel coefficients of the free space (pure LOS) case of Section 2.3 and applying correct scaling to maintain the SNR of 10 dB, the resulting RX antenna array patterns after applying the weight vectors are given in Figure 2-8. Adding the same noise to the case with two reflecting planes where up to two bounces are considered (see Section 2.5), results in the antenna patterns of Figure 2-9.

(a)   (b)
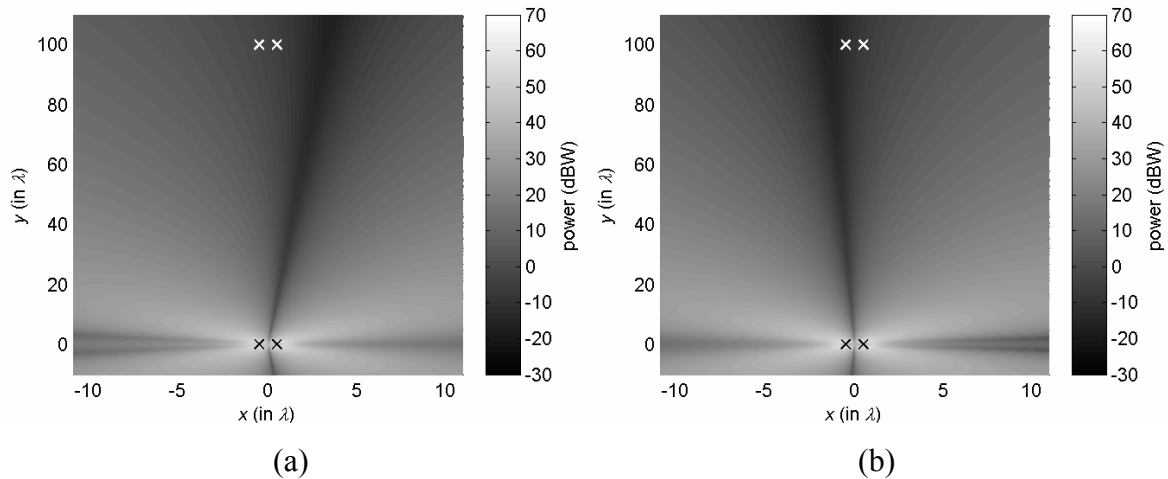
Figure 2-8: RX antenna array patterns after applying the first (a) and second (b) weight vector in free space with noise added to the channel observation.

When comparing the results of Figure 2-8 and Figure 2-9 with Figure 2-3 and Figure 2-7, respectively, we clearly see that the pure-LOS case strongly suffers from the additive noise. This can be explained by the fact that for this case the columns of the channel matrix have a strong resemblance (i.e., are highly correlated), see (2.9), resulting in a big error when noise is added. For the "richly-scattered" case, the channel matrix is highly orthogonal and this property is hardly changed when noise is added. As a result, the antenna patterns for the latter case are barely altered. Similar results are achieved when

other noise realisations are investigated, from which we can conclude that a MIMO system is indeed more robust in environments with many reflecting objects.

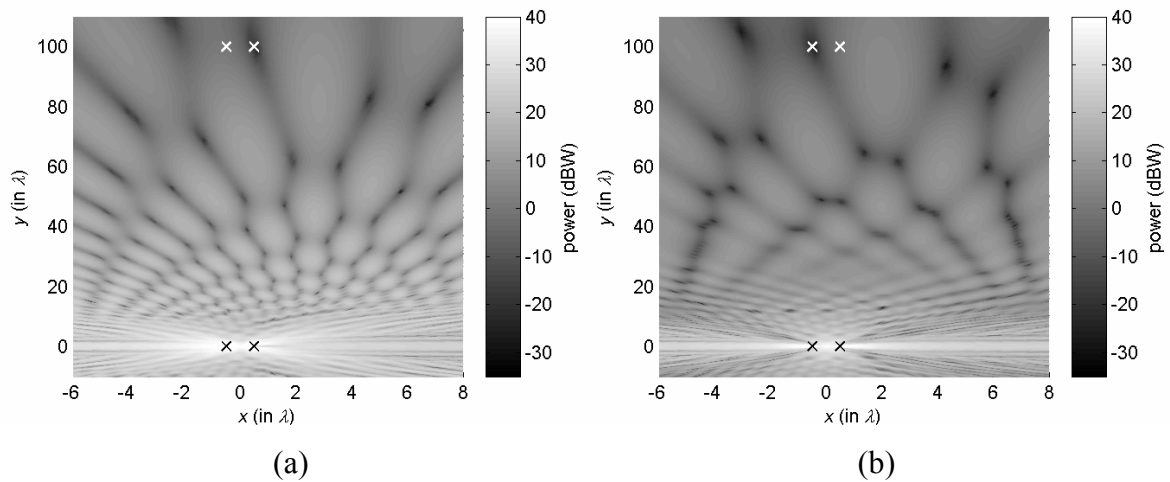<div align="center">(a)   (b)</div>
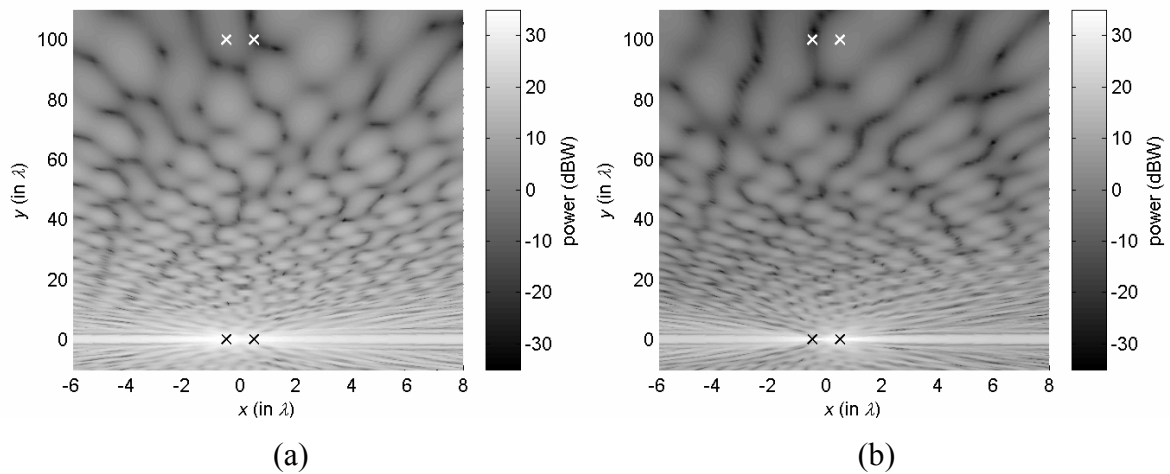
Figure 2-9: RX antenna array patterns after applying the first (a) and second (b) weight vector in a scenario with two perfectly reflecting planes (at $x = -6\lambda$ and $x = 8\lambda$), where only paths with up to two bounces are considered, and noise is added to the channel observation.

2.7 Conclusions

In this chapter, we showed that, for a communication system with multiple transmit and multiple receive antennas, the different signals from the different TX antennas (sent at the same frequency) can be separated at the receiver, under the assumption that the right weights can be found and applied. The ability of separating the different streams from the different transmit antennas, results in a linear growth in data rate with the number of TX antennas, by which the potential capacity enhancement of MIMO is intuitively explained. Furthermore, for cases with many reflecting paths, it is shown that the undesired antenna is nulled by a spot, whereas a local maximum is placed at the position of the desired antenna. This maximal separation between the two antennas speaks in favour of the robustness of MIMO systems in rich-scattering environments.

3

Multiple-Input Multiple-Output Channel Modelling

3.1 Introduction

In the ideal case the data rate of MIMO systems grows linearly with the number of TX antenna as we explained intuitively in Chapter 2. In general, however, the maximum transmission rate in a given bandwidth (i.e., the spectral efficiency) that can be exploited in MIMO systems depends on a number of parameters observed at the receiver, including the average received power of the desired signal, thermal and system-related noise, as well as co-channel interference. Moreover, the multidimensional statistical behaviour of the MIMO fading channel is of foremost significance to the system performance (e.g., influence of the spatial fading correlation). Therefore, it is important for the designer of a MIMO communication system to have an appropriate MIMO channel simulation model. Generally, important requirements for such a model are:

1. The representation of real-life MIMO channel statistics according to the targeted radio environment and system parameters (like antenna spacing, polarization, antenna element directionalities),
2. The possibility to easily cover a wide range of best-case to worst-case scenarios,
3. The ease of use and possibility to convey the relevant parameters between various groups of researchers to reliably compare results.

In this chapter, a MIMO channel model is introduced, based on these requirements. In Section 3.2, a wideband MIMO channel model is introduced, based on a geometric interpretation of the communication link, as an extension to the narrowband geometric MIMO interpretation of Chapter 2. For various environments, the variations in the different paths between transmitter and receiver as function of time, location and frequency, generally called fading, can be represented by statistical distributions. For these cases, the geometrically based model is transferred to a stochastic channel model. A number of distributions, i.e., fading characteristics, are described in Section 3.3. In Section 3.4, a wideband MIMO signal model is introduced, which includes the fading characteristics as

well as additive receiver noise. Most MIMO algorithms, however, are not introduced as wideband, but as narrowband techniques. Therefore, Section 3.5 describes a narrowband signal model and its fading statistics. The narrowband model is shown to be a special case of the wideband model. An impairment that is specific for multi-antenna systems is spatial fading correlation. A simple model to cover this impairment is introduced in Section 3.6.

3.2   A Geometrically Based Stochastic MIMO Channel Model

3.2.1   Continuous-Time Channel Model

Consider a wireless MIMO system, with N_t transmit (TX) and N_r receive (RX) antennas, that is operating in an environment with reflecting objects (see Figure 3-1). In such a scattering environment, during a transmission, reflections will occur and a transmitted signal that is launched by a given TX antenna arrives at a given RX antenna along a number of distinct paths. This effect is referred to as *multipath*. Because of movement of the user and/or movement of objects, each of these paths has its own time-varying angle of departure, path delay (i.e., excess delay), angle of arrival, and power. Due to constructive and destructive interference of the multipath components, the received signal can vary as a function of frequency, location and time. These variations are referred to as *fading*.
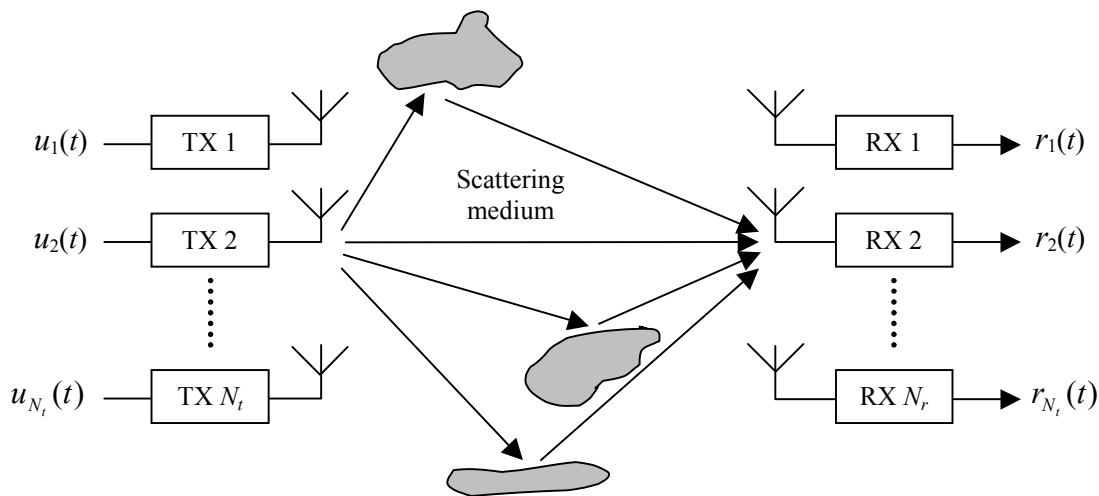
Figure 3-1: A MIMO communication system operating in a scattering environment.

To model the channel behaviour, we will extend the narrowband geometric MIMO interpretation of Chapter 2. The narrowband assumption is dropped, since, in general, the fading characteristics are not necessarily flat over frequency. To make a clear distinction with the narrowband case of Chapter 2, here, other symbols will be used.

In a MIMO system, all TX antennas transmit simultaneously and on the same carrier frequency and, therefore, the received signal on a given RX antenna q consists of a linear combination of contributions from the N_t transmitters. Furthermore, when considering the contribution of the p-th transmit antenna, due to the multipath the q-th RX antenna records a sum of scaled and phase-shifted copies of the original TX signal, where the i-th copy,

received at time t, experienced a time delay of, say, $\tau_{i,qp}(t)$. So, at time t, for a MIMO system operating at a carrier frequency f_c, the equivalent baseband transfer function from the p-th TX to the q-th RX antenna is in the time domain defined by[1] ([90, 92])

$$g_{qp}(t,\tau) = \sum_{i=0}^{N_m(t)-1} \gamma_{i,qp}(t) \exp\left(-j\phi_{i,qp}(t)\right) \delta\left(\tau - \tau_{i,qp}(t)\right), \tag{3.1}$$

where $N_m(t)$ is the number of observed multipath components at time t, and $\gamma_{i,qp}(t)$ and $\tau_{i,qp}(t)$ are the gain and delay, respectively, of a signal travelled through the i-th path from TX p to RX q and received at time t. Furthermore, $\delta(t)$ represents the Dirac function and $\phi_{i,qp}(t) = 2\pi f_c \tau_{i,qp}(t)$. Note that with c representing the speed of electromagnetic waves in air, and with $f_c = c/\lambda$ and $\tau_{i,qp} = d_{i,qp}/c$, we arrive at the same phase as in (2.8).

The channel transfer function $g_{qp}(t,\tau)$ can be physically interpreted as the channel response recorded on the q-th RX antenna at time t to an impulse signal from TX antenna p that is sent at time $t - \tau$, i.e., τ seconds in the past. Therefore, $g_{qp}(t,\tau)$ is called the equivalent baseband channel *impulse response* from TX p to RX q at time t. Note, for clarity, that $\tau_{i,qp}(t)$ does *not* represent the delay/travelling time an impulse transmitted at time t will encounter when travelling through the i-th path between TX p and RX q, but it denotes the delay/travelling time of the i-th path signal component having already travelled through the channel and *arriving* on the receive side at time t.

At this point, it is useful to introduce a number of parameters that are commonly used to characterise the time-domain channel impulse response:

- The *Power Delay Profile* (PDP) is defined as the power of the channel impulse response as function of τ. For a specific channel impulse response from the p-th TX to the q-th RX antenna, $g_{qp}(t,\tau)$, it is given by

$$P_{qp}(t,\tau) = \sum_{i=0}^{N_m(t)-1} \left|\gamma_{i,qp}(t)\right|^2 \delta\left(\tau - \tau_{i,qp}(t)\right). \tag{3.2}$$

- The *mean excess delay* is the first moment of the PDP and is, for a specific channel $g_{qp}(t,\tau)$, defined as ([92])

$$\bar{\tau}(t) = \frac{\displaystyle\int_{-\infty}^{\infty} P_{qp}(t,\tau)\tau\,d\tau}{\displaystyle\int_{-\infty}^{\infty} P_{qp}(t,\tau)d\tau} = \frac{\displaystyle\sum_{i=0}^{N_m(t)-1} \left|\gamma_{i,qp}(t)\right|^2 \tau_{i,qp}(t)}{\displaystyle\sum_{i=0}^{N_m(t)-1} \left|\gamma_{i,qp}(t)\right|^2}. \tag{3.3}$$

- The *root mean squared (rms) delay spread* is defined as the square root of the second central moment of the PDP:

---

[1] Although more than one scaled and phase-shifted replica of the original signal might arrive at a receiving antenna at the same time instant, i.e., through paths with the same length but with different angle of arrivals, it is assumed that these paths are summed and, hence, observed as a single path at the receiver.

$$\tau_d(t) = \sqrt{\frac{\int\limits_{-\infty}^{\infty} P_{qp}(t,\tau)(\tau - \bar{\tau}(t))^2 \, d\tau}{\int\limits_{-\infty}^{\infty} P_{qp}(t,\tau) d\tau}} = \sqrt{\frac{\sum\limits_{i=0}^{N_m(t)-1} |\gamma_{i,qp}(t)|^2 (\tau_{i,qp}(t) - \bar{\tau})^2}{\sum\limits_{i=0}^{N_m(t)-1} |\gamma_{i,qp}(t)|^2}} \; . \qquad (3.4)$$

Note that the mean excess delay and rms delay spread are measured relative to the first detectable signal arriving at the receiver (i.e., $\tau_{0,qp}$ is set to 0). In the remainder of this dissertation, we will assume that $\tau_{0,qp} = 0$ s. Furthermore, note that above parameters are often averaged over time and/or space. In this way, they don't specify a specific channel response as above, but characterise a particular propagation environment. These average values are commonly used for stochastic channel modelling as will be described in Subsection 3.2.2.

The architecture of a wireless communication system fundamentally depends on its envisioned application environments and, as such, the expected range of the rms delay spread has a big influence on the system design. If the average rms delay spread is sufficiently small compared to the symbol period of the transmitted signals, then all multipath components arrive within one signalling interval. Consequently, consecutive symbols do not interfere with one another and only amplitude and phase correction is required. On the other hand, if the symbol period is small compared to the delay spread, then inter-symbol interference (ISI) will result and equalisation is required ([121]).

For the introduction of some other parameters by which a channel can be characterised, it is convenient to determine the channel transfer function in the frequency domain. The frequency domain representation of the channel impulse response is called the *frequency response*. The frequency response is obtained by computing the Fourier transformation of the time domain impulse response:

$$g_{qp}^F(f,t) = F\{g_{qp}(f,t)\} = \int\limits_{-\infty}^{\infty} g_{qp}(t,\tau) \exp(-j2\pi f\tau) d\tau, \qquad (3.5)$$

where $F\{.\}$ denotes the Fourier transform of the corresponding function. In an environment with many reflections, a wireless communication channel will consist of many multipath components. From above equation, it becomes clear that the channel attenuation is then heavily dependent on the frequency. This can be explained by the fact that the phase of a multipath component, observed at the receiver, is frequency dependent. As a result, for some frequencies the multipath components add constructively while for other frequencies they add destructively. Therefore, the frequency response fluctuates over frequency and such a channel is generally called a *frequency-selective fading* channel.

A parameter to quantify the fluctuation of wireless channels over frequency is the *frequency correlation*. It defines the correlation between the channel response at different frequencies and is given by

$$R_{qp}(\Delta f, t) = \frac{1}{f_2 - f_1} \int\limits_{f_1}^{f_2} g_{qp}^F(f,t) \left(g_{qp}^F(f + \Delta f, t)\right)^* df, \qquad (3.6)$$

where f_1 and f_2 are two preferably widely-spaced frequencies in the frequency band of interest, Δf is the frequency separation for which the correlation is determined and * denotes the complex conjugate. This metric can be averaged over time and/or space in order to obtain average values. If the *coherence bandwidth B_c* is defined as the frequency separation for which the frequency correlation function is above 0.9, then the average coherence bandwidth is approximately ([90])

$$B_c \approx \frac{1}{50\tau_d},$$

(3.7)

where τ_d in seconds is the rms delay spread averaged over time. If the definition is relaxed so that the frequency correlation function is above 0.5, then B_c is about

$$B_c \approx \frac{1}{5\tau_d}.$$

(3.8)

So, the rms delay spread and coherence bandwidth are inversely proportional to one another. Note, however, that an exact relationship is hard to obtain and depends on the exact channel response and the required frequency correlation.

Based on the relationship between the coherence bandwidth and the rms delay spread, the impact of the rms delay spread on the architecture of a wireless communication system can also be explained from the coherence bandwidth point of view. If a transmission signal occupies a bandwidth smaller than B_c, the channel transfer function that is influencing the transmission is approximately flat over frequency and correspondingly only amplitude and phase correction is required and no equalisation. The system is said to experience *flat fading* and, consequently, called a *narrowband* system. When the bandwidth of the transceiver is considerably larger than B_c, however, the channel frequency response exhibits large fluctuations that need to be equalised in order to have a proper detection. In that case, the system is called a *wideband* system and, as explained above, the fading as function of frequency is referred to as *frequency-selective fading* and it is the counterpart of ISI in the time-domain.

3.2.2 Discrete-Time Channel Model

When the system bandwidth is limited, most likely the system is not able to distinguish every multipath component, but the receiver observes (weighted) summations of multipath components. This can be explained by the fact that the transmit and receive filters, which form part of any practical communication system, perform an integrating function and as a result rake together the multipath components.

A general flow graph of the communication link between the p-th TX and q-th RX antenna is depicted in Figure 3-2 ([121]). At the transmitter, the discrete-time input symbols $u_p(nT_s)$ are modulated using a modulation technique and shaping filter with impulse response $f_{TX}(\tau)$. The modulated signal is transmitted through the channel $g_{qp}(t,\tau)$. At the receiver, noise is added to the channel output (denoted by $v_q(t)$) and the result is filtered by the receiver filter with impulse response $f_{RX}(\tau)$. Subsequently, the receiver filter output is

sampled with period T_s. Regarding the sampling instants, we assume that the receiver is perfectly synchronised with the transmitter. Assuming that sampling results in time observations at time instants $t = nT_s$ and $\tau = lT_s$, where n is the sampling index and l is the impulse response tap number, the impulse response of the equivalent discrete-time baseband channel, $g_{qp}(nT_s, lT_s)$, is given by

$$g_{qp}(nT_s, lT_s) = \int_0^{\tau_{N_m(nT_s)-1,qp}} f(lT_s - \tau) g_{qp}(nT_s, \tau) d\tau$$

$$= \sum_{i=0}^{N_m(nT_s)-1} \gamma_{i,qp}(nT_s) \exp(-j\phi_{i,qp}(nT_s)) f(lT_s - \tau_{i,qp}(nT_s)),$$

(3.9)

where $f(\tau) = f_{TX}(\tau) * f_{RX}(\tau)$ is the convolution of the transmit and receive filter impulse responses, i.e., the combined transmit/receive filter impulse response. Note that, due to the filtering operations, the number of nonzero channel taps $g_{qp}(nT_s, lT_s)$ can be much higher than the number of nonzero propagation paths in the corresponding physical channel, resulting in a non-causal equivalent discrete-time baseband channel which has an infinite number of taps. In practical systems, however, the transmit and receive filters are causal and consist of Finite Impulse Response (FIR) filters which in turn results in a causal discrete channel impulse response that can be modelled by a (time-dependent) FIR filter with a limited number of taps.
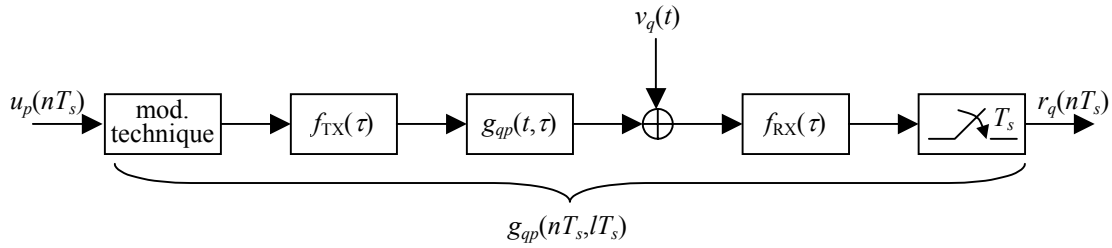
Figure 3-2: General flow graph of the communication link between the *p*-th TX and *q*-th RX antenna.

As example for the non-causal case assume that both the transmit and receive filter are ideal low-pass filters with

$$f_{TX}^F(f) = f_{RX}^F(f) = F\{f_{TX}(\tau)\} = F\{f_{RX}(\tau)\} = \begin{cases} \sqrt{T_s}, & \text{for } \dfrac{-1}{2T_s} < f \leq \dfrac{1}{2T_s}, \\ 0, & \text{elsewhere}. \end{cases}$$

(3.10)

The impulse response of the equivalent discrete-time baseband channel can be shown to be ([121])

$$g_{qp}(nT_s, lT_s) = \sum_{i=0}^{N_m(nT_s)-1} \gamma_{i,qp}(nT_s) \exp(-j\phi_{i,qp}(nT_s)) \text{sinc}\left(\pi\left(l - \frac{\tau_{i,qp}(nT_s)}{T_s}\right)\right).$$

(3.11)

Clearly, the channel impulse response is not causal, but from this example, it becomes clear that the filters perform an integrating function and as a result rake together the multipath components.

In the remainder of this dissertation, we will assume that the channel transfer function is causal. For convenience and without loss of generality, we assume that the multipath components within one sampling interval are linearly combined. Then, the coefficients of the FIR channel can be written as

$$g_{qp}(nT_s, lT_s) = \sum_{i=N_{\min}(nT_s)}^{N_{\max}(nT_s)} \gamma_{i,qp}(nT_s) \exp(-j\phi_{i,qp}(nT_s)), \tag{3.12}$$

where $N_{\min}(nT_s)$ and $N_{\max}(nT_s)$ are chosen such that the paths observed at time nT_s with indices i, with $N_{\min}(nT_s) \leq i \leq N_{\max}(nT_s)$ and with an encountered delay $\tau_{i,qp}(nT_s)$, stem from the l-th sampling interval of the time-variant channel impulse response, i.e.,

$$lT_s \leq \tau_{i,qp}(nT_s) < (l+1)T_s. \tag{3.13}$$

Note that above channel model can be represented by a *tapped delay line* as shown in Figure 3-3, and that l denotes the channel tap index, with $l = 0, \ldots, L_{qp}(nT_s) - 1$. $L_{qp}(nT_s)$ is the channel length and equals the number of channel taps that are required to also include the last path with index $N_m(nT_s) - 1$, i.e.,

$$L_{qp}(nT_s) = \left\lceil \frac{\tau_{N_m(nT_s)-1,qp}(nT_s)}{T_s} \right\rceil, \tag{3.14}$$

where $\lceil x \rceil$ provides the lowest integer value larger than or equal to x. Let the transmitted discrete-time complex baseband signal on the p-th TX antenna as a function of the sampling index n be $u_p(nT_s)$, with a bandwidth $B \leq 1/T_s$. Since in a MIMO system, all TX antennas transmit simultaneously and on the same carrier frequency, the received signal on a given RX antenna q consists of a linear combination of contributions from the N_t transmitters. Furthermore, under the assumption that the TX antennas are not spaced too far apart and that this also holds for the RX antennas, it is reasonable to assume that $L_{qp}(nT_s)$ is equal for all p and q, with $p = 1, \ldots, N_t$ and $q = 1, \ldots, N_r$, and set to $L(nT_s)$. Then, when omitting the additive noise at the receiver, the baseband signal $r_q(nT_s)$ recorded at the q-th receiver is given by

$$r_q(nT_s) = \sum_{p=1}^{N_t} \sum_{l=0}^{L(nT_s)-1} g_{qp}(nT_s, lT_s) u_p(nT_s - lT_s). \tag{3.15}$$

Moreover, to capture all N_r received signals into one equation, the matrix notation can be used:

$$\mathbf{r}(nT_s) = \sum_{l=0}^{L(nT_s)-1} \mathbf{G}(nT_s, lT_s)\mathbf{u}(nT_s - lT_s), \tag{3.16}$$

where,

$$\mathbf{u}(t)=\begin{pmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_{N_t}(t) \end{pmatrix}, \mathbf{r}(t)=\begin{pmatrix} r_1(t) \\ r_2(t) \\ \vdots \\ r_{N_r}(t) \end{pmatrix}, \mathbf{G}(t,\tau)=\begin{pmatrix} g_{11}(t,\tau) & g_{12}(t,\tau) & \cdots & g_{1N_t}(t,\tau) \\ g_{21}(t,\tau) & g_{22}(t,\tau) & \cdots & g_{2N_t}(t,\tau) \\ \vdots & \vdots & \ddots & \vdots \\ g_{N_r 1}(t,\tau) & g_{N_r 2}(t,\tau) & \cdots & g_{N_r N_t}(t,\tau) \end{pmatrix}. \quad (3.17)$$
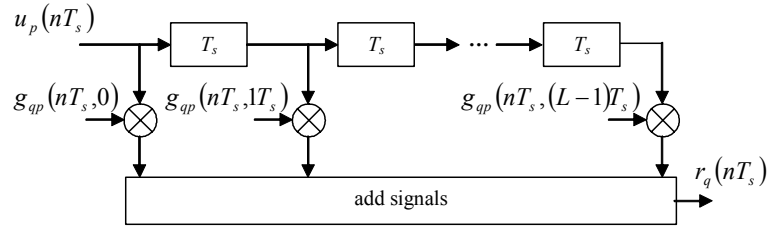
Figure 3-3: Tapped delay line channel model for time-variant channels.

From this point onwards, only the discrete-time domain is used unless mentioned otherwise. To simplify the notations and without loss of generality T_s is dropped and the discrete time-domain MIMO channel model is defined by

$$\mathbf{r}(n)=\sum_{l=0}^{L-1}\mathbf{G}(n,l)\mathbf{u}(n-l). \quad (3.18)$$

3.2.3   Quasi-Static Discrete-Time Channel Model

When the user and/or objects hardly move during a packet transmission, the MIMO radio channel can be considered constant for the respective time interval. Such a channel is called a *quasi-static* channel. In other words, the packet duration is assumed less than the *coherence time*, i.e., the time in which the communication channel can be considered static. According to [92] a conservative measure, i.e., a lower bound, for the coherence time T_c is given by

$$T_c \approx \frac{9}{16\pi \cdot f_m}, \quad (3.19)$$

where f_m is the maximum Doppler shift defined by $f_m = v/\lambda$. Taking indoor wireless LAN communication at 5 GHz as an example, we could say that the velocity v of the user will not exceed 10 km/h or roughly 3 m/s. In this case, the maximum Doppler shift is

$$f_m = \frac{v}{\lambda} = \frac{v}{c/f} \approx \frac{3}{3\cdot 10^8 / 5\cdot 10^9} \approx 50\,\text{Hz}. \quad (3.20)$$

This Doppler shift leads to a coherence time of $T_c \approx 3.6$ ms. So, for this example, the assumption that the channel is quasi-static is justified as long as the packet duration is

significantly less than 3.6 ms. The lowest communication rate of the IEEE 802.11a standard is 6 Mbps ([57]). The maximum IP (Internet Protocol) packetlength equals 1500 bytes. A transmission of a 1500 byte packet at 6 Mbps would require about 2 ms, so the quasi-static assumption is valid for this example.

Under the quasi-static assumption, $\mathbf{G}(n,l)$ is independent of n during a packet transmission. So, let $\mathbf{G}(l)$ denote the quasi-static MIMO channel transfer characteristic of a wireless medium observed by a communication system with bandwidth B, then the relation between the vectors $\mathbf{u}(n)$ and $\mathbf{r}(n)$ is defined by the quasi-static wideband MIMO relation:

$$\mathbf{r}(n) = \sum_{l=0}^{L-1} \mathbf{G}(l)\mathbf{u}(n-l). \tag{3.21}$$

For indoor environments, measurement campaigns have indicated that the amplitudes of the elements of $\mathbf{G}(l)$, $|g_{qp}(l)|$, are approximately Rayleigh distributed ([48, 62, 71, 77]). This can be understood intuitively from the reasonable assumption that an indoor environment contains a large number of scatterers and, as a result, a channel tap consists of a large sum of multipath components (see (3.12)). When the multipath components have similar gains and the phase is uniformly distributed between 0 and 2π, then, according to the Central Limit Theorem ([90]), the resulting complex channel coefficients are complex Gaussian distributed (and their amplitude is Rayleigh distributed). So, the geometric-based model can be transformed to a stochastic model. As an extension to this model, in Sections 3.3 and 3.5 more stochastic fading characteristics are defined. The assumptions made to arrive at a stochastic model allow for direct modelling of the components of $\mathbf{G}(l)$ instead of fully describing the physical geometric propagation paths, thus, resulting in more convenient and faster channel modelling.

Although the channel is assumed static for a packet transmission, it is commonly known that the multipath characteristics change over time, due to movement of the user, movement of objects in the environment, etc. This is modelled by changing $\mathbf{G}(l)$ on a packet by packet basis according to given fading statistics (see Sections 3.3 and 3.5). Note that above channel model can be represented by a tapped delay line that describes the linear transformation between $\mathbf{u}(n)$ and $\mathbf{r}(n)$ as shown in Figure 3-4.
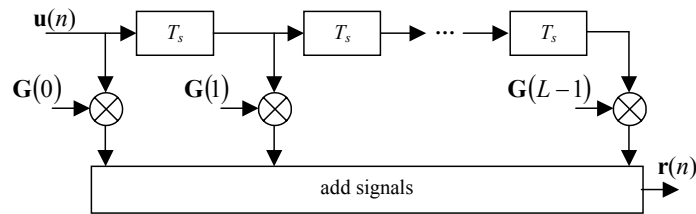
Figure 3-4: MIMO signal model represented by a tapped delay line.

To simplify the channel modelling, it is furthermore assumed that the average power of all transmission coefficients for a given tap, i.e., of all elements of $\mathbf{G}(l)$ is identical for a given tap, so

$$E\left[\left|g_{qp}(l)\right|^2\right] \equiv P(l), \text{ for all } p = 1, 2, \cdots, N_t \text{ and } q = 1, 2, \cdots, N_r. \tag{3.22}$$

$P(l)$, with $l = 0, \ldots, L-1$, is the discrete-time equivalent of the continuous-time PDP given in Subsection 3.2.1. The total power of the discrete-time PDP, i.e., the large-scale channel gain encompassing distance-dependent decay, is given by

$$P_c = \sum_{l=0}^{L-1} P(l).$$
(3.23)

As explained for the continuous-time case, the mean excess delay is the first moment of the PDP. When performing the calculation in the time domain instead of the "sampling domain", the mean excess delay is defined as,

$$E[\tau] = \frac{\sum_{l=0}^{L-1} P(lT_s) \cdot lT_s}{\sum_{l=0}^{L-1} P(lT_s)} = \frac{\sum_{l=0}^{L-1} P(lT_s) \cdot lT_s}{P_c}.$$
(3.24)

In the discrete-time case, the rms delay spread equals:

$$\tau_d = \sqrt{E[\tau^2] - E[\tau]^2},$$
(3.25)

where

$$E[\tau^2] = \frac{\sum_{l=0}^{L-1} P(lT_s) \cdot (lT_s)^2}{\sum_{l=0}^{L-1} P(lT_s)} = \frac{\sum_{l=0}^{L-1} P(lT_s) \cdot (lT_s)^2}{P_c}.$$
(3.26)

Note that these delays are measured relative to the first detectable signal arriving at the receiver[1] (i.e., l starts at 0).

3.3   Fading Characteristics of Indoor-like Environments

3.3.1 Motivation

From fundamental work published in [48, 80, 97] and many other measurements reported in literature, it can be deduced that the average received multipath power of a transmitted impulse in indoor-like environments tends to fall off exponentially over time. Furthermore, in general, the amplitudes of the channel tap coefficients are Rayleigh distributed. In case the communication channel also has a Line-of-Sight (LOS) or specular component, one can imagine that, generally speaking, the first path is Ricean distributed. The physical mechanisms driving the indoor-like characteristics are mainly based on:

---

[1] The rms delay spread does not depend on the start of l, but the mean excess delay does.

- slow movement of objects and/or transceivers, resulting in a negligible Doppler shift (see Section 3.2);
- (in case of a LOS or specular component) a component that illuminates the arrays entirely and is thus deterministic from antenna to antenna on a packet by packet basis;
- the large number of reflectors within a typical indoor-like environment (resulting in Rayleigh fading);
- the propagation loss characteristic, where longer propagation paths arriving linearly later in time have logarithmically weaker energy (see Figure 3-5);
- analogue-to-digital conversion, which results in the fact that in the digital domain the channel is monitored at the sampling rate, leading to equidistant time observations of, e.g., the channel impulse response.

The above points have led to simplified channel models as used in [21, 33, 43, 46]. These models assume a fixed number of channel taps with equidistant delays and an exponentially decaying PDP, with or without a specular component added to the first tap. To explicitly include the PDP, from this point forward we will normalise $\mathbf{G}(l)$, such that every element of $\mathbf{G}(l)$ has an average power of one. This results in the discrete sampling-domain wideband MIMO channel model:

$$\mathbf{r}(n) = \sum_{l=0}^{L-1} \sqrt{P(l)}\,\mathbf{G}(l)\mathbf{u}(n-l).$$

(3.27)

The wideband MIMO channel models, as described in the following subsections, result from a combining of the simplified models given in [21, 33, 43, 46].
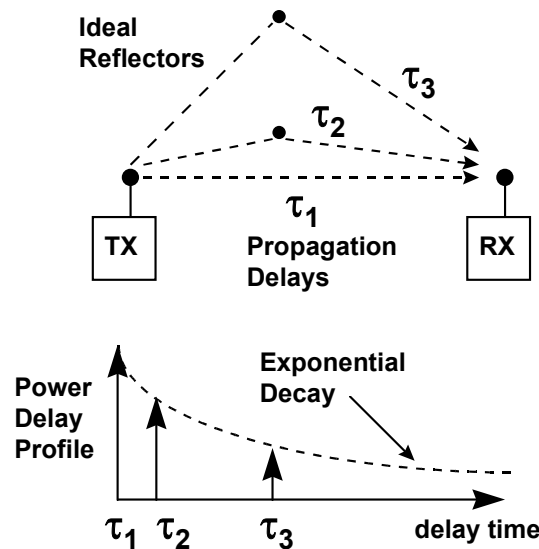
Figure 3-5: Exponentially decaying Power Delay Profile

3.3.2 Wideband Rayleigh Fading Model

When no strong LOS or specular path is present, the large number of reflectors within a typical indoor-like environment results in Rayleigh fading ([92]). For a MIMO system

operating in such a rich-scattering environment, when the antenna spacing is chosen equal to or larger than half the carrier wavelength, the channel coefficients can be assumed i.i.d. ([36]). To model this, $\mathbf{G}(l)$ in (3.27) is defined as $\mathbf{G}_{\text{iid}}(l)$. The elements of the $N_r \times N_t$ dimensional channel transfer matrix $\mathbf{G}_{\text{iid}}(l)$ are i.i.d. *circularly-symmetric complex Gaussian* variables with zero mean and unit variance, with an independent realisation for all l, $l = 0, 1, \ldots, L-1$. The definition of a circularly-symmetric complex Gaussian random variable, say z, with zero mean and variance σ^2 is given by $z = x + jy$ with x and y being i.i.d. zero mean real Gaussian variables with variance $\sigma^2/2$. Equivalently, each entry of $\mathbf{G}_{\text{iid}}(l)$ has uniformly distributed phase and Rayleigh distributed magnitude. Furthermore, to consider the quasi-static property of the channel, an independent realisation of $\mathbf{G}_{\text{iid}}(l)$ is generated on a packet by packet basis.

To model the effect that longer propagation paths arriving linearly later in time have logarithmically-weaker energy (see Figure 3-5), the PDP is characterised by the one-sided exponentially decaying function with equidistant delays

$$P(l) = \begin{cases} \dfrac{P_c \exp(-l/(\tau_d f_s))}{\displaystyle\sum_{l=0}^{L-1} \exp(-l/(\tau_d f_s))}, & \text{for } \tau_d > 0 \text{ and } l = 0,1,\ldots,L-1, \\[4mm] P_c, & \text{for } \tau_d = 0 \text{ and } L \equiv 1. \end{cases} \tag{3.28}$$

where τ_d represents the rms delay spread defined by (3.25), and $f_s = 1/T_s$ denotes the sampling frequency.

3.3.3 Wideband Ricean Fading Model

An unobstructed direct path or a very strong path between transmitter and receiver is called the LOS or specular component, respectively. To model such a component, it is reasonable to assume that the corresponding path travels the shortest possible distance between transmitter and receiver and, thus, arrives before any multipath components. Therefore, the LOS or specular path will be modelled being part of the first tap of the MIMO channel impulse response. The components due to scattering can still be modelled following the NLOS model of Subsection 3.3.2. Note that in this case the distribution of the first tap is modified to a Ricean distribution.

Let the $N_r \times N_t$ matrix $\mathbf{G}_{\text{spec}}$ denote the normalised LOS or specular part of the MIMO communication channel, then we can define $\mathbf{G}(l)$ of (3.27) as

$$\mathbf{G}(l) = \begin{cases} \dfrac{1}{\sqrt{P(l)}} \sqrt{\dfrac{K}{K+1}} \sqrt{\displaystyle\sum_{l=0}^{L-1} P(l)} \, \mathbf{G}_{\text{spec}} + \sqrt{\dfrac{1}{K+1}} \mathbf{G}_{\text{iid}}(l), & \text{for } l = 0 \\[4mm] \sqrt{\dfrac{1}{K+1}} \mathbf{G}_{\text{iid}}(l), & \text{for } l = 1,\ldots,L-1. \end{cases} \tag{3.29}$$

From (3.27) it follows that for the above model, the total channel gain is given by

$$P_c = P_{\text{spec}} + P_{\text{scatt}} = \frac{K}{K+1}\sum_{l=0}^{L-1}P(l) + \sum_{l=0}^{L-1}\frac{1}{K+1}P(l) = \sum_{l=0}^{L-1}P(l). \qquad (3.30)$$

In these equations the parameter K, known from single-antenna Ricean fading ([92]), denotes the ratio between specular-to-scattered energy according to $K = P_{\text{spec}}/P_{\text{scatt}}$. Note that in case K is set to zero, this model reduces to the wideband Rayleigh fading channel model described in the previous subsection.

Two cases can be distinguished for the definition of a LOS component. Firstly, the case where the TX-RX distance is significantly larger than the antenna spacing, referred to as the *far-field* case, and, secondly, the situation where TX-RX distance is only up to tens of antenna spacing. The latter is called the *near-field* case. In this dissertation, we will only consider far-field communication, but for completeness, we will include some notes about the near field.

As defined in [33], when the TX and RX are well in the *far field* of one another, the LOS or specular component is given by the "product" of the steering vectors at TX and RX:

$$\mathbf{G}_{\text{spec}} = \mathbf{a}_{\text{RX}}(\theta_{\text{RX}})\mathbf{a}_{\text{TX}}^T(\theta_{\text{TX}}), \qquad (3.31)$$

where $\mathbf{a}_{\text{TX}}(\theta_{\text{TX}})$ and $\mathbf{a}_{\text{RX}}(\theta_{\text{RX}})$ are the steering vectors at TX and RX of which the entries denote the phase per antenna element that is applied or observed for a given Angle of Departure (AoD) θ_{TX} or Angle of Arrival (AoA) θ_{RX}, respectively. $(.)^T$ denotes the transpose of the corresponding vector (or matrix). As example, the steering vector of an N_x-element linear array is given by

$$\left(1 \quad \exp(j2\pi d\cos(\theta)) \quad \cdots \quad \exp(j2\pi d(N_x-1)\cos(\theta))\right)^T, \qquad (3.32)$$

where θ is the AoA or AoD and d is the antenna spacing in wavelengths. When the AoD and AoA are unknown, we could assume, for simplicity, that they are random i.i.d. variables taken uniformly from $[0,2\pi\rangle$ (and changing on a packet by packet basis).

If the antenna configuration is also unknown, in order to still model the specular component, we could simply assume that for each channel realization, the steering vectors are randomised according to the following expression:

$$\mathbf{a} = \left(\exp(j\phi_1) \quad \exp(j\phi_2) \quad ... \quad \exp(j\phi_{N_x})\right)^T, \qquad (3.33)$$

where the ϕ_x are i.i.d. variables taken uniformly from $[0,2\pi\rangle$ and N_x equals the number of transmit or receive antennas for the TX or RX steering vector, respectively. This model is appropriate to describe a Line-of-Sight scenario in which, again, the RX and TX antenna structures are well in the far field of one another but in which we do not want to become concrete about antenna spacing, polarization, orientation etc.

Note that a "product" of two vectors results in a rank-one matrix (Appendix A.1.5). From a MIMO perspective, such a channel is ill conditioned and does not have a high capacity (see

Section 4.4). So when $K \to \infty$, the capacity for the far field case goes to that of a rank-1 channel.

When the TX and RX are in each others *near field*, the LOS component might have a higher rank than one. Since the TX-RX distance is relatively small, two LOS paths originating from different TX antennas to the same RX antenna cannot be assumed to have the same AoD. Therefore, to model this, we must consider every single specular channel element as, e.g., is done in Chapter 2 and not simply take the "product" of the steering vectors. As a result, the matrix of the LOS component for the near-field case might be better conditioned and its capacity might be higher than that of a rank-1 channel.

Finally, in Figure 3-6, as an example, an average and instantaneous PDP of the wideband stochastic Ricean model, as described in this section, is shown for $K = 10$.
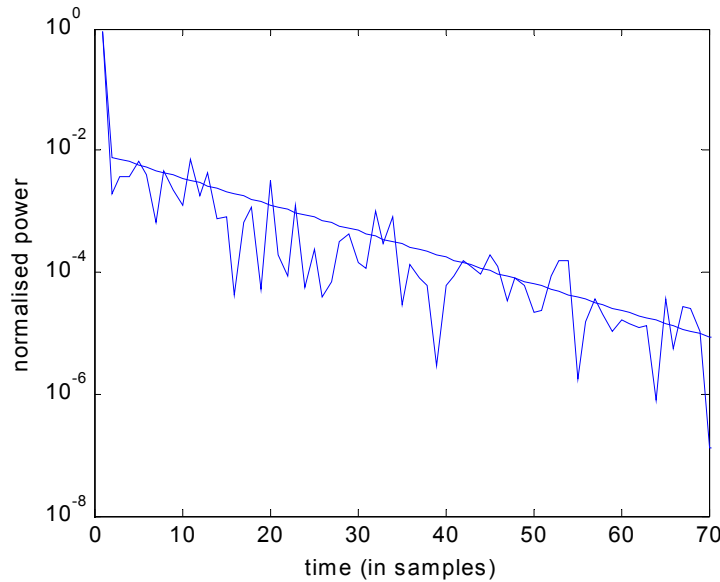
Figure 3-6: Example of generated average and instantaneous Power Delay Profiles for an rms delay spread of 10 sampling intervals and $K = 10$.

3.3.4 Uniformly Distributed PDP Model

For theoretical analysis, it is useful to introduce a hypothetical channel model that is referred to as the uniformly distributed PDP model. The fading properties of this model are equivalent to the wideband Rayleigh fading model, whereas the PDP is a block function, such that

$$P(l) = \begin{cases} \dfrac{P_c}{L}, & \text{for } l = 0,1,...,L-1, \\ 0, & \text{for } l < 0 \text{ and } l > (L-1). \end{cases} \tag{3.34}$$

Note that for this PDP, following (3.25), the rms delay spread (in samples) can be shown to be

$$\tau_d = \sqrt{\frac{1}{L}\sum_{l=0}^{L-1} l^2 - \left(\frac{1}{L}\sum_{l=0}^{L-1} l\right)^2}$$

$$= \sqrt{\frac{(L-1)(2L-1)}{6} - \left(\frac{L-1}{2}\right)^2} = \sqrt{\frac{L^2-1}{12}} \text{ samples}. \tag{3.35}$$

3.4 Wideband MIMO Signal Model

As explained in Section 3.1, besides the propagation characteristics the receiver also observes thermal and system related noise, and sometimes co-channel interference. Thermal noise is treated in this section, system related impairments are tackled in Chapter 6, but Co-channel interference is outside the scope of this dissertation.

The thermal noise is usually modelled as Additive White Gaussian Noise (AWGN) ([99]). We will assume that we also can model the baseband equivalent noise as AWGN. With this assumption, the following wideband MIMO signal model is introduced:

$$\mathbf{r}(n) = \sum_{l=0}^{L-1} \sqrt{P(l)}\,\mathbf{G}(l)\mathbf{u}(n-l) + \mathbf{v}(n), \tag{3.36}$$

where $\mathbf{v}(n)$ represents AWGN at the n-th sample with N_r i.i.d. zero-mean, circularly-symmetric complex Gaussian elements with variance σ_v^2. The elements of $\mathbf{u}(n)$ are assumed to be zero mean, uncorrelated random variables with variance σ_u^2. When we furthermore assume that the average channel gain P_c is 1, then, the expected SNR per receive antenna can be shown to be $\rho = N_t\sigma_u^2/\sigma_v^2$. Note that, for a fair comparison, we want to keep the total TX power the same as in the Single-Input Single-Output (SISO) case. Therefore, the power per TX antenna is scaled down by N_t.

3.5 Stochastic Narrowband MIMO Channel Models

3.5.1 Motivation

Since most MIMO algorithms are proposed as narrowband techniques, it is convenient to specify narrowband MIMO channel models. These narrowband models can be seen as special cases of the wideband models: the channel is assumed to be flat over frequency, i.e., the system bandwidth B is assume to be much smaller than the coherence bandwidth of the channel B_c. Hence, only the first tap of (3.27), $\mathbf{G}(0)$, is assumed to have a significant contribution, i.e., the channel length L is set to one. This results in a PDP

$$P(l) = \begin{cases} P_c, & \text{for } l = 0, \\ 0, & \text{for } l \neq 0. \end{cases} \tag{3.37}$$

To distinguish from the wideband model, we will introduce a narrowband signal model following the notation of Chapter 2. When omitting the time index and assuming that the average path loss is normalised, such that $P_c = 1$, the result is:

$$\mathbf{x} = \mathbf{Hs} + \mathbf{n}, \tag{3.38}$$

where $\mathbf{H}$ equals $\mathbf{G}(0)$, the narrowband MIMO transmit vector $\mathbf{s}$ is equal to $\mathbf{u}(n)$ under the assumption that $B \ll B_c$, the receive vector $\mathbf{x}$ equals $\mathbf{r}(n)$ with $B \ll B_c$, and $\mathbf{n}$ is AWGN, equivalent to $\mathbf{v}(n)$.

3.5.2 Flat Rayleigh Fading Model

To specify a scenario of frequency-flat Rayleigh fading, a model is introduced called flat Rayleigh fading model. This is a reference model, which can be envisaged as a simplification of dense scattering occurring locally only. The flat Rayleigh fading MIMO model can be obtained from its wideband counterpart, introduced in Subsection 3.3.2, by setting the channel length L to 1. To model uncorrelated frequency-flat Rayleigh fading, $\mathbf{H}$ is defined as an $N_r \times N_t$ dimensional matrix, denoted by $\mathbf{H}_{\text{iid}}$ and having i.i.d. zero-mean, circularly-symmetric complex Gaussian elements with a variance of one.

3.5.3 Flat Ricean Fading Model

The frequency-flat Rayleigh fading model of Subsection 3.5.2 can be extended with a LOS or specular component. This results in a model that is equivalent to the first tap of the wideband Ricean model of Subsection 3.3.3:

$$\mathbf{H} = \sqrt{\frac{K}{K+1}} \mathbf{H}_{\text{spec}} + \sqrt{\frac{1}{K+1}} \mathbf{H}_{\text{iid}}, \tag{3.39}$$

where $\mathbf{H}_{\text{iid}}$ and $\mathbf{H}_{\text{spec}}$ are the complex Gaussian distributed matrix and specular component equivalent to $\mathbf{G}_{\text{iid}}$ and $\mathbf{G}_{\text{spec}}$ defined in Subsections 3.3.2 and 3.3.3, respectively, and K denotes the Ricean factor.

3.5.4 Pure-LOS versus AWGN MIMO Channel Model

For a single antenna (i.e., SISO) system operating in a Ricean fading environment and with additive white Gaussian noise, when the Ricean factor K goes to infinity, the result is an AWGN channel. This can be easily seen by filling in the SISO variant of the Ricean distribution of Subsection 3.5.3 into the SISO variant of the signal model of (3.38) and assuming $K \to \infty$.

For the MIMO models introduced in Subsections 3.3.3 and 3.5.3, however, the result would be that the channel is completely defined by a LOS or specular component, since for $K \to \infty$, $\mathbf{H} = \mathbf{H}_{\text{spec}}$ (see Subsection 3.5.3). In case the TX and RX are well in the far field of one another, recalling (3.31), this yields

$$\mathbf{H} = \mathbf{H}_{\text{spec}} = \mathbf{a}_{\text{RX}}\left(\theta_{\text{RX}}\right)\mathbf{a}_{\text{TX}}^{T}\left(\theta_{\text{TX}}\right). \tag{3.40}$$

As mentioned in Subsection 3.3.3, from a MIMO perspective, such a channel is ill-conditioned and does not have a high capacity.

The highest MIMO capacity is achieved by an orthogonal MIMO channel. To examine AWGN properties for a MIMO system, it is convenient to introduce a channel model based on such an orthogonal channel. Keeping the property in mind that the average power of all elements of $\mathbf{H}$ should be equal to one, we propose the following AWGN MIMO channel model (under the restriction that $N_r \geq N_t$):

$$\mathbf{H}_{\text{orthogonal}} = \mathbf{F}_{N_r}\begin{pmatrix} \mathbf{I}_{N_t} \\ \mathbf{0}_{(N_r - N_t) \times N_t} \end{pmatrix}, \tag{3.41}$$

where $\mathbf{F}_N$ equals the $N \times N$ Fourier matrix, defined by ([113])

$$\mathbf{F}_N = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W^1 & W^2 & \cdots & W^{(N-1)} \\ 1 & W^2 & W^4 & \cdots & W^{2(N-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & W^{(N-1)} & W^{(N-1)2} & \cdots & W^{(N-1)^2} \end{pmatrix}, \tag{3.42}$$

with $W = \exp(-j2\pi/N)$, and $\mathbf{0}_{M \times N}$ is the $M \times N$ all zeros matrix.

A physical interpretation of this model can be found as follows. Assume that all N_t TX antennas can be considered as independent omni-directional sources, located in the far field and that their distance to RX antenna 1 is equivalent and equal to an integer number of wavelengths. Furthermore, assume a linear RX array. In free space, at the receiver, the incoming wavefront from every TX antenna will be observed (approximately) as a planar wavefront. For a linear array with N_r antennas, such a wavefront coming in from direction θ, as depicted in Figure 3-7, results in a steering vector given by

$$\begin{pmatrix} 1 & \exp\left(j2\pi d_{\text{RX}}\cos(\theta)\right) & \cdots & \exp\left(j2\pi d_{\text{RX}}\left(N_r - 1\right)\cos(\theta)\right) \end{pmatrix}^T, \tag{3.43}$$

where d_{RX} denotes the RX antenna spacing in wavelengths.
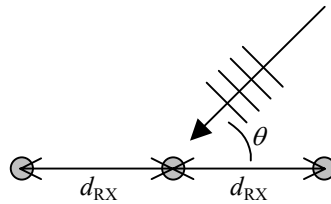
Figure 3-7: An incoming planar wavefront from direction θ.

For above sketched situation, the p-th column of $\mathbf{H}_{\text{orthogonal}}$ defined by (3.41) must equal the steering vector of a planar wavefront coming from direction θ_p, i.e., the p-th TX antenna is located at an angle θ_p with respect to the RX array. Thus, the (q,p)-th element of $\mathbf{H}_{\text{orthogonal}}$ must be equal to the q-th element of the steering vector with angle θ_p:

$$\exp\left(-j2\pi/N_r\right)^{(p-1)(q-1)} = \exp\left(j2\pi d_{\text{RX}}(q-1)\cos\left(\theta_p\right)\right). \tag{3.44}$$

This results in the relation

$$\cos\left(\theta_p\right) = -\frac{(p-1)}{N_r d_{\text{RX}}}. \tag{3.45}$$

When d_{RX} is equal to one wavelength, this results for, e.g., a 3×3 channel in the directions of the three transmit antennas as given in Figure 3-8. More generally, θ_p can be found geometrically as the angle of the line that goes through the origin and through the crossing point between the circle with a radius of one wavelength and the line $x = -(p-1)/N_r$ wavelengths.
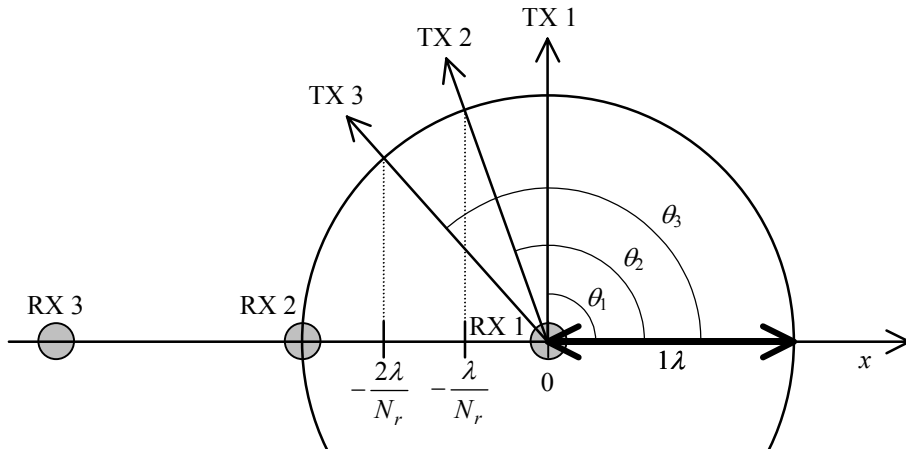
Figure 3-8: An example of the angles of arrivals for a 3×3 orthogonal MIMO channel. The three RX antennas are denoted by the grey dots. Furthermore, the direction of the TX antennas is shown. They are located in the far field.

3.6   *Spatial Correlation*

In general, the MIMO system performance deteriorates when spatial correlation is present. Therefore, if system designers have enough spatial freedom, they should design the system such that spatial correlation is reduced to a minimum. For instance, when, for indoor-like environments, the antenna spacing is chosen to be equal or larger than half the carrier wavelength, the spatial correlation can be assumed negligible ([36]). If space is the limiting design factor, then spatial correlation should be taken into account in the channel model.

In [25], the spatial (or fading) correlation for a narrowband flat-fading MIMO channel $\mathbf{H}$, where the elements of $\mathbf{H}$ are zero mean and have an average power of one, is defined as

$$\mathbf{R}_H = E\left[\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^H\right] ,\tag{3.46}$$

where $\mathrm{vec}(\mathbf{H})$ denotes the $N_r N_t \times 1$ dimensional vector composed by stacking the columns of $\mathbf{H}$, and ^{H} denotes the Hermitian transpose, i.e., conjugate transpose of a matrix. Thus, the correlation between the fading path from TX a to RX b, h_{ba}, and the fading path from TX p to RX q, h_{qp}, which equals

$$\rho_{yx,qp} = E\left[h_{ba}h_{qp}^*\right] ,\tag{3.47}$$

can be found on position ($(a{-}1)N_r{+}b$, $(p{-}1)N_r{+}q$) of $\mathbf{R}_H$. Note that when the elements of $\mathbf{H}$ are not zero mean and/or not having a unit average power, a correlation measure as defined in [65] must be used (here extended for the complex case):

$$\rho_{yx,qp} = \frac{E\left[h_{ba}h_{qp}^*\right] - E\left[h_{ba}\right]E\left[h_{qp}^*\right]}{\sqrt{\left(E\left[|\,h_{ba}\,|^2\right] - \left|E\left[h_{ba}\right]\right|^2\right)\left(E\left[|\,h_{qp}\,|^2\right] - \left|E\left[h_{qp}\right]\right|^2\right)}} .\tag{3.48}$$

In [15, 22, 25, 86, 134], the correlation between different MIMO channel elements is modelled with the assumption that the correlation among receive antennas is independent from the correlation between transmit antennas (and vice versa). The underlying justification for this approach is to assume that only immediate surroundings of the antenna array impose the correlation between array elements and have no impact on correlations observed between the elements of the array at the other end of the link, which is a reasonable assumption for indoor environments ([64]). Based on this assumption, the following model is suggested:

$$\mathbf{R}_H = \mathbf{R}_{\mathrm{TX}}^T \otimes \mathbf{R}_{\mathrm{RX}} ,\tag{3.49}$$

where $\otimes$ represents the Kronecker product. The intuitive explanation of the equation is that the correlation between two elements of the channel matrix is the product of the correlation seen from the transmitter side, $\mathbf{R}_{\mathrm{TX}}$ ($N_t \times N_t$ dimensional), and the correlation seen from the receiver side, $\mathbf{R}_{\mathrm{RX}}$ ($N_r \times N_r$ dimensional). Under the assumption that the elements of $\mathbf{H}$ are zero mean and have a variance of one, the correlation matrices can be found as follows:

$$\mathbf{R}_{\mathrm{TX}} = E\left[\left(\mathbf{h}^q\right)^H \mathbf{h}^q\right] , \quad \text{for all } q, \quad q = 1,...,N_r ,\tag{3.50}$$

$$\mathbf{R}_{\mathrm{RX}} = E\left[\mathbf{h}_p \mathbf{h}_p^H\right] , \quad \text{for all } p, \quad p = 1,...,N_t ,\tag{3.51}$$

where $\mathbf{h}^q$ is the q-th row of $\mathbf{H}$ and $\mathbf{h}_p$ is the p-th column of $\mathbf{H}$.

To generate independent narrowband flat-fading MIMO channel realisations with spatial correlation, one can use ([86])

$$\mathbf{H} = \mathrm{unvec}\left(\mathbf{R}_H^{\frac{1}{2}}\mathbf{h}_{\mathrm{iid}}\right),\tag{3.52}$$

where $\mathbf{h}_{\text{iid}}$ is an $N_t N_r \times 1$ stochastic vector with i.i.d. zero-mean unit variance complex Gaussian elements and unvec(.) is the reverse of the vec(.) operation. The question is how to determine the "square-root" of $\mathbf{R}_H$. To answer this question, we need to state some observations:

1. $\mathbf{R}_H$ is Hermitian:

$$\begin{aligned}
\mathbf{R}_H^H &= \left(E\left[\text{vec}(\mathbf{H}) \text{vec}(\mathbf{H})^H \right] \right)^H = E\left[\left(\text{vec}(\mathbf{H}) \text{vec}(\mathbf{H})^H \right)^H \right] \\
&= E\left[\left(\text{vec}(\mathbf{H})^H \right)^H \text{vec}(\mathbf{H})^H \right] = \mathbf{R}_H .
\end{aligned} \tag{3.53}$$

2. A Hermitian matrix can be diagonalised by a unitary matrix ([113]), so, $\mathbf{R}_H$ can be diagonalised as

$$\mathbf{U}\mathbf{R}_H\mathbf{U}^H = \mathbf{\Lambda}_H , \tag{3.54}$$

where $\mathbf{\Lambda}_H$ is a diagonal matrix with the eigenvalues of $\mathbf{R}_H$ on its diagonal (i.e., the eigenvalue matrix) and the columns of $\mathbf{U}$ equal the corresponding orthonormal eigenvectors, thus $\mathbf{U}^H\mathbf{U} = \mathbf{I}$, which is the definition of a unitary matrix.

3. $\mathbf{R}_H$ is nonnegative definite (see Appendix A.1.7), because for all nonzero complex vectors $\mathbf{z}$ it follows that

$$\begin{aligned}
\mathbf{z}^H \mathbf{R}_H \mathbf{z} &= \mathbf{z}^H E\left[\text{vec}(\mathbf{H}) \text{vec}(\mathbf{H})^H \right] \mathbf{z} \\
&= E\left[\mathbf{z}^H \text{vec}(\mathbf{H}) \text{vec}(\mathbf{H})^H \mathbf{z} \right] = E\left[\left| \text{vec}(\mathbf{H})^H \mathbf{z} \right|^2 \right] \geq 0 ,
\end{aligned} \tag{3.55}$$

thus, the eigenvalues of $\mathbf{R}_H$ are nonnegative real numbers.

Based on these three observations, we may write

$$\mathbf{R}_H = \mathbf{U}^H \mathbf{\Lambda}_H \mathbf{U} = \mathbf{U}^H \mathbf{\Lambda}_H^{\frac{1}{2}} \mathbf{\Lambda}_H^{\frac{1}{2}} \mathbf{U} = \left(\mathbf{\Lambda}_H^{\frac{1}{2}} \mathbf{U} \right)^H \mathbf{\Lambda}_H^{\frac{1}{2}} \mathbf{U} = \left(\mathbf{R}_H^{\frac{1}{2}} \right)^H \mathbf{R}_H^{\frac{1}{2}} , \tag{3.56}$$

from which we obtain the "square-root" of $\mathbf{R}_H$. Above observations also hold for $\mathbf{R}_{\text{TX}}$ and $\mathbf{R}_{\text{RX}}$, so from (3.49) it follows that

$$\begin{aligned}
\mathbf{R}_H &= \mathbf{R}_{\text{TX}}^T \otimes \mathbf{R}_{\text{RX}} \\
&= \left(\left(\mathbf{R}_{\text{TX}}^{\frac{1}{2}} \right)^H \mathbf{R}_{\text{TX}}^{\frac{1}{2}} \right)^T \otimes \left(\left(\mathbf{R}_{\text{RX}}^{\frac{1}{2}} \right)^H \mathbf{R}_{\text{RX}}^{\frac{1}{2}} \right) \\
&= \left(\left(\left(\mathbf{R}_{\text{TX}}^{\frac{1}{2}} \right)^H \right)^T \otimes \mathbf{R}_{\text{RX}}^{\frac{1}{2}} \right)^H \left(\left(\left(\mathbf{R}_{\text{TX}}^{\frac{1}{2}} \right)^H \right)^T \otimes \mathbf{R}_{\text{RX}}^{\frac{1}{2}} \right) \\
&= \left(\mathbf{R}_H^{\frac{1}{2}} \right)^H \mathbf{R}_H^{\frac{1}{2}} ,
\end{aligned} \tag{3.57}$$

where the property is used that, for any matrix $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ with proper dimensions, $(\mathbf{AB}) \otimes (\mathbf{CD}) = (\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D})$. Based on the Kronecker product identity that for any (complex) $M \times N$ matrix $\mathbf{A}$, $N \times P$ matrix $\mathbf{B}$, and $P \times Q$ matrix $\mathbf{C}$ (see Appendix A.1.10),

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B}), \tag{3.58}$$

(3.52) can be rewritten as

$$
\begin{aligned}
\mathbf{H} &= \text{unvec}\left(\mathbf{R}_H^{\frac{1}{2}}\mathbf{h}_{\text{iid}}\right) \\
&= \text{unvec}\left(\left(\left(\left(\mathbf{R}_{\text{TX}}^{\frac{1}{2}}\right)^H\right)^T \otimes \mathbf{R}_{\text{RX}}^{\frac{1}{2}}\right)\text{vec}(\mathbf{H}_{\text{iid}})\right) \\
&= \mathbf{R}_{\text{RX}}^{\frac{1}{2}}\mathbf{H}_{\text{iid}}\left(\mathbf{R}_{\text{TX}}^{\frac{1}{2}}\right)^H,
\end{aligned}
\tag{3.59}
$$

where $\mathbf{H}_{\text{iid}} = \text{unvec}(\mathbf{h}_{\text{iid}})$ is a stochastic $N_r \times N_t$ matrix with i.i.d. circularly-symmetric complex Gaussian zero-mean unit variance elements. This result is equivalent to the spatial correlation model introduced in [22]. Note that, according to (3.56), the "square-root" of $\mathbf{R}_H$ is not Hermitian. Likewise, the "square-roots" of $\mathbf{R}_{\text{TX}}$ and $\mathbf{R}_{\text{TX}}$ are not Hermitian.

When system simulations need to be carried out, one way to proceed is to explicitly state specific correlation matrices $\mathbf{R}_{\text{TX}}$ and $\mathbf{R}_{\text{RX}}$ covering various propagation scenarios. To obtain these specific correlation matrices, either ray tracing or correlation measurements have to be performed representing different scenarios. This approach is cumbersome and has as major disadvantage that the essential MIMO properties (relating to the achievable MIMO capacity) are concealed in a large number of parameters, namely, the various correlation matrix entries and, therefore, it is hard to cover a wide range of best-case to worst-case scenarios. This leads to the question how to reduce this amount of parameters.

We start with the observation that the capacity and BER performance are frequently used measures to evaluate MIMO systems. Below, we introduce a compact representation of the spatial correlation that nevertheless results in an equivalent capacity and BER performance.

The capacity of an $N_r \times N_t$ narrowband MIMO channel $\mathbf{H}$ is given by ([36], Section 4.4)

$$C = \log_2 \det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t}\mathbf{HH}^H\right) \text{ bits/s/Hz}, \tag{3.60}$$

where ρ represents the SNR per receive antenna and $\mathbf{I}_N$ is the $N \times N$ dimensional identity matrix. Furthermore, the channel matrix $\mathbf{H}$ is normalised, meaning that the elements of $\mathbf{H}$ have an average power of one.

When spatial correlation is present, the capacity equals

$$C = \log_2 \det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t}\mathbf{R}_{\text{RX}}^{\frac{1}{2}}\mathbf{H}_{\text{iid}}\left(\mathbf{R}_{\text{TX}}^{\frac{1}{2}}\right)^H \mathbf{R}_{\text{TX}}^{\frac{1}{2}}\mathbf{H}_{\text{iid}}^H\left(\mathbf{R}_{\text{RX}}^{\frac{1}{2}}\right)^H\right) \text{ bits/s/Hz}. \tag{3.61}$$

With the equality $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$, this can be rewritten to

$$C = \log_2 \det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t} \mathbf{H}_{\text{iid}} \mathbf{R}_{\text{TX}} \mathbf{H}_{\text{iid}}^H \mathbf{R}_{\text{RX}} \right) \text{ bits/s/Hz.} \qquad (3.62)$$

For high SNRs, we get

$$\begin{aligned} C &\approx \log_2 \det\left(\frac{\rho}{N_t} \mathbf{H}_{\text{iid}} \mathbf{R}_{\text{TX}} \mathbf{H}_{\text{iid}}^H \mathbf{R}_{\text{RX}} \right) \\ &= \log_2 \det\left(\frac{\rho}{N_t} \mathbf{I}_{N_r} \right) \det(\mathbf{H}_{\text{iid}}) \det(\mathbf{R}_{\text{TX}}) \det(\mathbf{H}_{\text{iid}}^H) \det(\mathbf{R}_{\text{RX}}), \end{aligned} \qquad (3.63)$$

since the determinant of a product is the product of the determinants. So, apparently, the capacity distributions of two different situations will be the same when the determinant of the $\mathbf{R}_{\text{TX}}$'s and $\mathbf{R}_{\text{RX}}$'s are equal. Or, it must be possible to introduce a model that is, in a capacity sense, a mapping of measured correlation matrices. To that end, we require

$$\det(\mathbf{R}_{\text{TX,mod}}) = \det(\mathbf{R}_{\text{TX,meas}}), \qquad (3.64)$$

$$\det(\mathbf{R}_{\text{RX,mod}}) = \det(\mathbf{R}_{\text{RX,meas}}). \qquad (3.65)$$

Note that in case the correlation matrices of the model would be set equal on both sides of the communication link, i.e., $\mathbf{R}_{\text{mod}} = \mathbf{R}_{\text{TX,mod}} = \mathbf{R}_{\text{RX,mod}}$, we would get the criterion

$$\det(\mathbf{R}_{\text{mod}}) = \sqrt{\det(\mathbf{R}_{\text{TX,meas}}) \ \det(\mathbf{R}_{\text{RX,meas}})}. \qquad (3.66)$$

Now the question is if there exists a unique solution for the requirements (3.64) and (3.65). To answer that question, note that, like $\mathbf{R}_H$, both $\mathbf{R}_{\text{TX}}$ and $\mathbf{R}_{\text{RX}}$ are nonnegative definite. According to Hadamard's inequality for an $N \times N$ nonnegative definite matrix $\mathbf{A}$ ([54]),

$$\det(\mathbf{A}) \le \prod_{i=1}^{N} a_{ii}, \qquad (3.67)$$

where a_{ii} represents the i-th diagonal element of $\mathbf{A}$. For the correlation matrices $\mathbf{R}_{\text{TX}}$ and $\mathbf{R}_{\text{RX}}$ this means that $\det(\mathbf{R}_{\text{TX}}) \le 1$ and $\det(\mathbf{R}_{\text{RX}}) \le 1$. Furthermore, since the determinant of a matrix is the product of the eigenvalues of that matrix ([113]) and we have shown at the beginning of this section that the eigenvalues of a nonnegative definite matrix are real and nonnegative, this yields $\det(\mathbf{R}_{\text{TX}}) \ge 0$ and $\det(\mathbf{R}_{\text{RX}}) \ge 0$. So, the determinant of the *measured* correlation matrices will always be real, larger than or equal to zero and less than or equal to one.

Next, it is shown that a unique match can be found with respect to capacity using the following simple and generic definitions for the transmitter and receiver correlation:

$$\mathbf{R}_{\mathrm{TX,mod}} = \begin{pmatrix} 1 & r_{\mathrm{TX}} & r_{\mathrm{TX}}^2 & \cdots & r_{\mathrm{TX}}^{N_t-1} \\ r_{\mathrm{TX}} & 1 & r_{\mathrm{TX}} & \ddots & \vdots \\ r_{\mathrm{TX}}^2 & r_{\mathrm{TX}} & 1 & \ddots & r_{\mathrm{TX}}^2 \\ \vdots & \ddots & \ddots & \ddots & r_{\mathrm{TX}} \\ r_{\mathrm{TX}}^{N_t-1} & \cdots & r_{\mathrm{TX}}^2 & r_{\mathrm{TX}} & 1 \end{pmatrix}, \tag{3.68}$$

$$\mathbf{R}_{\mathrm{RX,mod}} = \begin{pmatrix} 1 & r_{\mathrm{RX}} & r_{\mathrm{RX}}^2 & \cdots & r_{\mathrm{RX}}^{N_r-1} \\ r_{\mathrm{RX}} & 1 & r_{\mathrm{RX}} & \ddots & \vdots \\ r_{\mathrm{RX}}^2 & r_{\mathrm{RX}} & 1 & \ddots & r_{\mathrm{RX}}^2 \\ \vdots & \ddots & \ddots & \ddots & r_{\mathrm{RX}} \\ r_{\mathrm{RX}}^{N_r-1} & \cdots & r_{\mathrm{RX}}^2 & r_{\mathrm{RX}} & 1 \end{pmatrix}, \tag{3.69}$$

where r_{TX} and r_{RX} represent (real-valued) correlation coefficients. Note that a similar model has been introduced in [74] with the difference that in [74] the correlation is defined as $\mathbf{R} = E[\mathbf{HH}^H]$. The most powerful property of this model is that, when ranging the coefficients between 0.0 and 1.0, we can go from fully uncorrelated scenarios (all off-diagonal elements of both matrices equal to 0.0) to fully correlated scenarios (all entries equal to 1.0). Another property is the simple form of the determinants of the matrices. The determinant of, e.g., $\mathbf{R}_{\mathrm{TX,mod}}$ can be shown to be

$$\det\left(\mathbf{R}_{\mathrm{TX,mod}}\right) = \left(1 - r_{\mathrm{TX}}^2\right)^{N_t-1}. \tag{3.70}$$
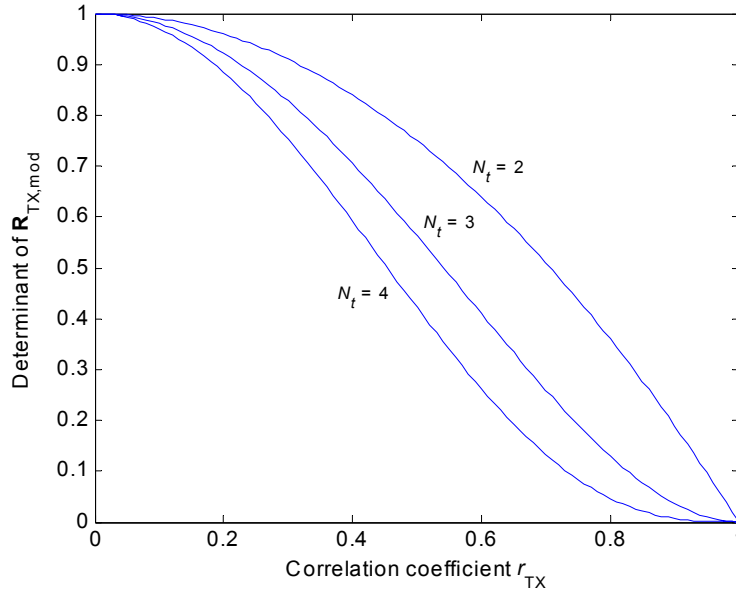
Figure 3-9: The determinant of the correlation model matrix $\mathbf{R}_{\mathrm{TX,mod}}$ versus the correlation coefficient r_{TX} for a various number of TX antennas.

Finally, it can be shown that the determinant for the *modelled* matrices is monotonically decreasing in the range of interest, e.g., $\mathbf{R}_{\mathrm{TX,mod}}$ as function of r_{TX} is monotonically

decreasing for $0 \leq r_{TX} \leq 1$ (see Figure 3-9 for N_t is 2, 3 and 4). Based on these observations, it can be concluded that there will always be a unique solution that satisfies the criteria (3.64) and (3.65).

Since a mathematical link is found to match the MIMO capacity of measured correlation matrices with that of the model, we can suffice with one example. The result is presented for complex correlation matrices measured in a picocell environment ([68]) and given by

$$
\mathbf{R}_{TX,meas} = \begin{pmatrix} 1 & -0.45+0.53i & 0.37-0.22i & 0.19+0.21i \\ -0.45-0.53i & 1 & -0.35-0.02i & 0.02-0.27i \\ 0.37+0.22i & -0.35+0.02i & 1 & -0.10+0.54i \\ 0.19-0.21i & 0.02+0.27i & -0.10-0.54i & 1 \end{pmatrix}, \quad (3.71)
$$

$$
\mathbf{R}_{RX,meas} = \begin{pmatrix} 1 & -0.13-0.62i & -0.49+0.23i & 0.15+0.28i \\ -0.13+0.62i & 1 & -0.13-0.52i & -0.38+0.12i \\ -0.49-0.23i & -0.13+0.52i & 1 & 0.02-0.61i \\ 0.15-0.28i & -0.38-0.12i & 0.02+0.61i & 1 \end{pmatrix}. \quad (3.72)
$$

For these measured matrices, it can be shown that $\det(\mathbf{R}_{TX,meas}) = 0.2372$ and $\det(\mathbf{R}_{RX,meas}) = 0.2796$, respectively. From the criteria (3.64), (3.65), $0 \leq r_{TX} \leq 1$, and $0 \leq r_{RX} \leq 1$, we obtain $r_{TX} = 0.6172$ and $r_{RX} = 0.5883$. With these results, the capacity of the measured correlation matrices can be compared with that of the model. Note that for every realisation of $\mathbf{H}_{iid}$, (3.61) produces a different instantaneous capacity value. The average of these capacity values, i.e., the ergodic capacity, as function of the SNR per receive antenna is shown in Figure 3-10 for the measured and modelled correlation matrices. From these curves, we indeed see that the match is perfect, even for low SNR values.
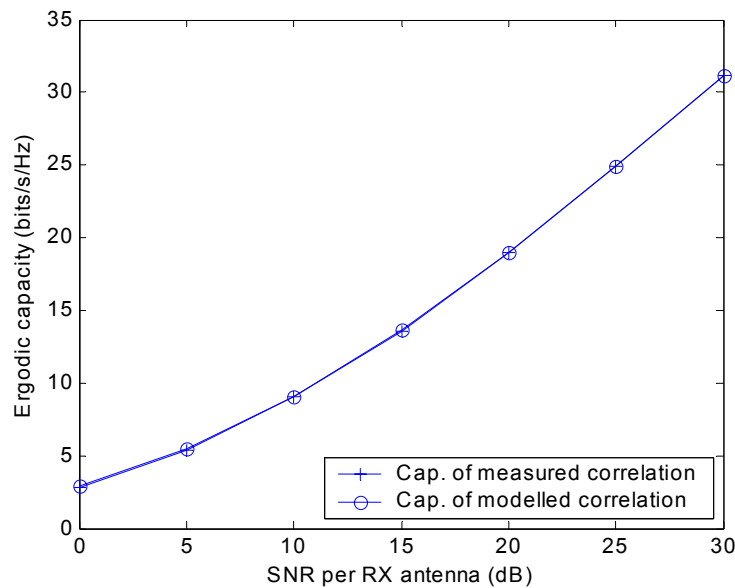
Figure 3-10: Ergodic capacity versus SNR per RX antenna for measured and modelled spatial correlation for a 4×4 system.

Obviously, the introduced spatial correlation model may not be an accurate model for some real-world scenarios, but it is a simple double-coefficient model that allows one to study the effect of correlation on the MIMO capacity in an explicit way. Moreover, with the criteria (3.64) and (3.65), we can mathematically link it to measured correlations. A mathematical match in BER performance can also be obtained, but the matching criteria are somewhat different. To determine these criteria, a BER analysis needs to be performed. Such an analysis depends on the chosen MIMO algorithm and, therefore, this analysis is performed at the end of Chapter 4, in Section 4.12.

In order to simplify the model further, for small r_{TX} and r_{RX} (much smaller than 1) we can discard the higher order terms in (3.68) and (3.69), resulting in tridiagonal matrices. As an example, for a 4×4 system this leads to

$$\mathbf{R}_{TX} = \begin{pmatrix} 1 & r_{TX} & 0 & 0 \\ r_{TX} & 1 & r_{TX} & 0 \\ 0 & r_{TX} & 1 & r_{TX} \\ 0 & 0 & r_{TX} & 1 \end{pmatrix} \text{ and } \mathbf{R}_{RX} = \begin{pmatrix} 1 & r_{RX} & 0 & 0 \\ r_{RX} & 1 & r_{RX} & 0 \\ 0 & r_{RX} & 1 & r_{RX} \\ 0 & 0 & r_{RX} & 1 \end{pmatrix}. \tag{3.73}$$

This can be viewed as approximating a scenario in which any pair of adjacent antenna elements is correlated, whereas any other pair exhibits independent fading.

For further simplification, we can set $r_{TX} = r_{RX} = r$, which leads to a single-parameter MIMO correlation model. Now we can depict the capacity as function of r for the introduced "exponential" spatial correlation model and its derivative, the tridiagonal model. The average capacity, i.e., the ergodic capacity, for a 4×4 system and an SNR per RX antenna of 20 dB is given in Figure 3-11 for the exponential and tridiagonal correlation models together with the capacity for uncorrelated (full rank) and fully correlated (rank-1) fading scenarios. Clearly, the tridiagonal correlation model is only valid for small r, whereas the exponential model ranges from the fully uncorrelated to the totally correlated scenario by simply changing r from 0 to 1.

The ergodic full-rank capacity in Figure 3-11 is found by determining

$$C = E \left[\log_2 \det \left(\mathbf{I} + \frac{\rho}{N_t} \mathbf{H}_{iid} \mathbf{H}_{iid}^H \right) \right], \tag{3.74}$$

where $E[.]$ denotes the expectation. The ergodic rank-1 capacity is found by setting $\mathbf{H}$ equal to

$$\mathbf{H} = \begin{pmatrix} h & \cdots & h \\ \vdots & \ddots & \vdots \\ h & \cdots & h \end{pmatrix} = h\mathbf{1}_{N_r \times N_t}, \tag{3.75}$$

where $\mathbf{1}_{N \times M}$ is the $N \times M$ all ones matrix and h is a circularly-symmetric complex Gaussian random variable with unit variance. Since in the latter case all channel elements are fully correlated, it is obvious that such a channel has only one eigenmode, with a gain equal to

the nonzero eigenvalue of $\mathbf{HH}^H$, i.e., $N_t N_r |h|^2$. So, the ergodic rank-1 capacity can be shown to be

$$C = E\left[\log_2\left(1 + N_r \rho |h|^2\right)\right] .$$

(3.76)
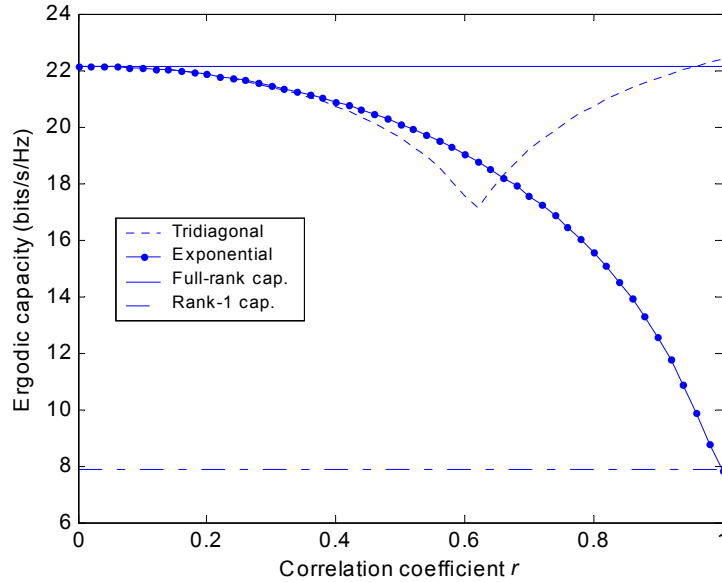
Figure 3-11: Ergodic capacity versus correlation coefficient *r* for different correlation models, for a 4 × 4 MIMO system and an SNR of 20 dB per RX antenna.

The fact that the tridiagonal model is only valid for small *r* can be explained by verifying if the tridiagonal structure satisfies the observation made above that a correlation matrix is nonnegative definite. For a nonnegative definite matrix, the eigenvalues are real and nonnegative. The eigenvalues of

$$\mathbf{R} = \mathbf{R}_{H,\mathrm{TX}} = \mathbf{R}_{H,\mathrm{RX}} = \begin{pmatrix} 1 & r & 0 & 0 \\ r & 1 & r & 0 \\ 0 & r & 1 & r \\ 0 & 0 & r & 1 \end{pmatrix},$$

(3.77)

can be found using Appendix A.1.2, and are given by

$$
\begin{aligned}
\lambda_1 &= 1 + \tfrac{1}{2}\left(1 + \sqrt{5}\right)\cdot r, \\
\lambda_2 &= 1 - \tfrac{1}{2}\left(1 - \sqrt{5}\right)\cdot r, \\
\lambda_3 &= 1 + \tfrac{1}{2}\left(1 - \sqrt{5}\right)\cdot r, \\
\lambda_4 &= 1 - \tfrac{1}{2}\left(1 + \sqrt{5}\right)\cdot r.
\end{aligned}
$$

(3.78)

In the range $0 \le r \le 1$, λ_4 becomes negative for $r > 2/(1+\sqrt{5}) \approx 0.618$, resulting in a negative definite matrix and, thus, leading to an incorrect model. As shown in Figure 3-11,

this result indeed corresponds to the region where the capacity of the tridiagonal model goes up again, when the correlation r exceeds 0.618. When setting r to $2/(1+\sqrt{5})$ and checking the amount of nonzero eigenvalues of $\mathbf{HH}^H$, it can be observed that $\mathbf{HH}^H$ still has 3 nonzero eigenvalues, leading to a 3×3 ergodic capacity of approximately 16.7 bits/s/Hz. This explains the capacity level of the tridiagonal model in Figure 3-11 at $r \approx 0.618$.

For the cases where the number of TX and RX antennas are set equal and are ranging from 3 to 25, the r value above which the tridiagonal model becomes negative definite is given in Figure 3-12. From this figure, it can be concluded that the spatial model based on the tridiagonal structure is valid for values of r that are smaller than 0.5.
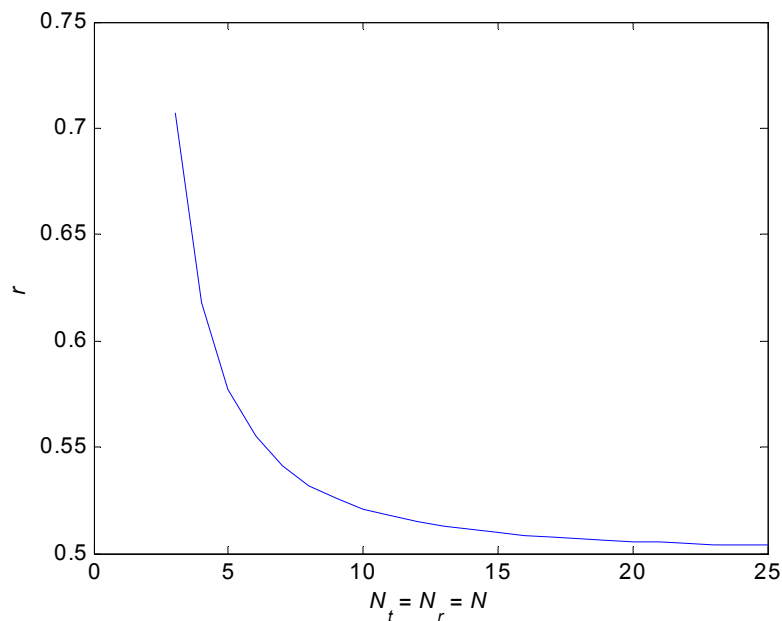
Figure 3-12: The region above which the matrices of the tridiagonal correlation model are not longer valid.

Note that the spatial correlation model described in this section is a narrowband model. To take spatial fading correlation into account in a wideband channel model, a Correlation Delay Profile (CDP) must be introduced. This can be explained by the intuition that in real-world environments, the first channel taps are mainly determined by a few strong paths, e.g., the LOS path and some dominant reflections, whereas towards the last taps of the channel impulse response, the spread of the AoD and AoA, i.e., the angular spread, is omni-directional with many (equally strong) contributing paths. This results in a high correlation coefficient at the beginning of the CDP and low values at the end. Two examples of a CDP are given in Figure 3-13, namely, the linearly and exponentially decaying CDP. For convenience, we used the same horizontal scale as for Figure 3-6.

Finally, note that in LOS scenarios, the specular component inherently includes spatial correlation and, thus, the correlation matrices as described in this section are not needed for the specular component. Hence, the extension of the above to include a LOS component is straightforward.
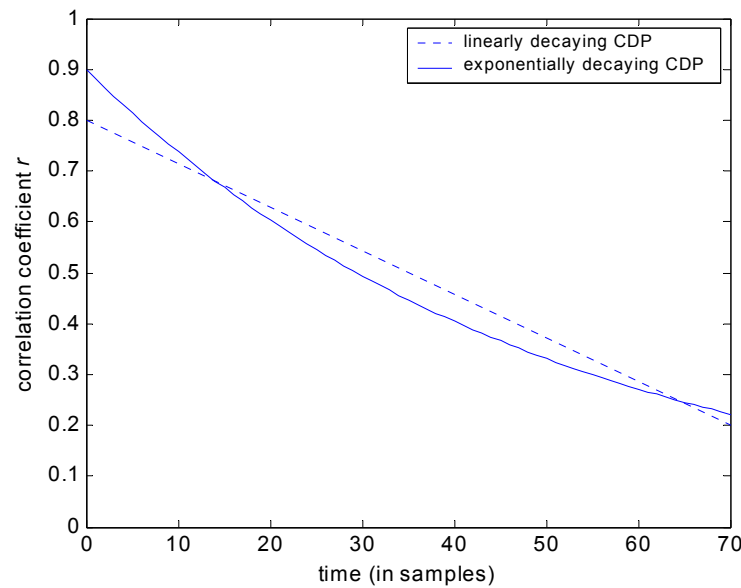
Figure 3-13: Two examples of a Correlation Delay Profile for an rms delay spread of 10 sampling intervals; linearly and exponentially decaying CDP.

3.7   Conclusions

In this chapter, a stochastical wideband MIMO channel model is introduced based on a geometric contemplation. The result is a channel model that models the typically large number of channel parameters by a very few carefully selected ones. Since the model consists of only a few parameters, the ease of use is high. Two of the selected parameters are introduced to model the two most critical impairments for MIMO, namely a LOS component and spatial correlation.

With respect to spatial correlation, we introduced a simple representation of spatial correlation in MIMO radio channels. For the frequently used evaluation measures of a MIMO system, namely capacity and BER performance (see Section 4.12), the amount of parameters representing the spatial correlation can be reduced to at most two. With a proper choice of these coefficients, the correlation can be varied controllably from the totally uncorrelated scenario to the fully correlated scenario. This simplified correlation model allows one to perform simulations with spatial correlation, while it is not required to explicitly specify the hardware (e.g. antenna) setup and wave propagation environment to include the spatial correlation. Altogether, this makes the model powerful, yet simple to use.

4

Flat-fading MIMO Techniques

4.1 Introduction

Besides the channel conditions (see the previous chapter), also the structure of the transmit signal of a MIMO system has a strong impact on the achievable capacity and performance. In addition, the signal design directly influences the complexity of the transmitter and, particularly, the receiver. These observations have led to numerous research activities to proper MIMO techniques. Basically, the proposed schemes can be split in two groups: Space Time Coding (STC) ([116]) and Space Division Multiplexing (SDM) ([36, 93, 132]). STC increases the robustness/performance of the communication system by coding over the different transmitter branches, while SDM achieves a higher data rate by transmitting independent data streams on the different transmitter branches simultaneously and at the same carrier frequency. These basic concepts have been the basis for various flavours of transmission approaches, which resulted in a multiplicity of candidate transmission schemes. Combined with corresponding receiver techniques, these schemes offer a variety of trade-offs between capacity-attainment capability, frame error-rate performance, computational complexity/simplicity and sensitivity to channel/interference estimation mismatch. In Section 4.2, an overview of some standard MIMO transmit techniques is given. We will restrict ourselves to narrowband techniques, and thus the flat-fading channel models of Chapter 3 are used. A method to extend these techniques to wideband applications is treated in Chapter 5.

Based on the different techniques that are highlighted, a conceptual unified view on MIMO is introduced in Section 4.2. Since the main goal of this dissertation is the search for data rate enhancement, SDM is the most interesting technique. Moreover, recent findings in literature have shown that, under certain conditions, SDM together with one-dimensional outer coding can outperform space-time codes ([18, 63, 148]).

After the introduction of the MIMO signal model in Section 4.3, the capacity of MIMO channels is determined and a Packet Error Rate (PER) lowerbound is formulated based on

the capacity in Section 4.4. In Section 4.5, the relation between the SNR per receive antenna and the energy per bit versus noise density (Eb/N0) for MIMO is defined.

Section 4.6 through Section 4.10 describe various SDM algorithms, namely, Zero Forcing (ZF), Minimum Mean Squared Error (MMSE), ZF with Successive Interference Cancellation (SIC), MMSE with SIC, and Maximum Likelihood Detection (MLD). For ZF, MMSE and MLD, soft-decision output algorithms are defined. Furthermore, an error-rate performance analysis is performed for ZF and MLD. The error-rate performance of the algorithms with and without outer coding is compared in the next section.

Another method that is used to compare the different SDM techniques is by determining the complexity of the different algorithms in terms of number of additions and multiplications. The resulting complexity numbers are a good measure for the final complexity of the (receiver) hardware.

In Section 4.12, a match is derived between the spatial correlation model introduced in Section 3.6 and measured correlation matrices with respect to the error-rate performance.

When outer coding is used and the MIMO mapping is seen as an inner code, one can imagine that exchanging information between the MIMO demapper and the decoder of the outer code could potentially improve the performance. This exchange of information is commonly organised in a number of iterations and this principle is generally referred to as turbo processing. Section 4.13 describes the application of this principle to SDM, named Turbo SDM.

Finally, conclusions are drawn in Section 4.14.

4.2   A Unified Framework of MIMO Techniques

4.2.1   General Structure

In this section, an overview of (most of) the existing MIMO techniques is provided and a framework is introduced in which a general TX and RX structure is proposed. This framework could form the basis of a unified theory on MIMO techniques.

Regarding the TX structure, in general, a TX signal for a MIMO system with N_t transmit antennas is generated by performing the following tasks on the incoming bit stream:

- channel encoding,
- mapping of the encoded bits on the spatial and/or temporal dimensions,
- mapping the (coded) bits onto a constellation diagram (and, if necessary, weighting of the resulting streams).

On the receive side, generally speaking, detection is performed jointly over the spatial and temporal dimension. The complexity strongly depends on the TX signal design. When nothing is undertaken to reduce the complexity, the number of codewords can grow exponentially with the size of the spatial and temporal dimension. Proper design of the TX

signal, however, allows for less complex receivers achieving (near) optimal performance. Examples of such MIMO schemes are discussed later on in this section. The general structure of a MIMO system is given in Figure 4-1.
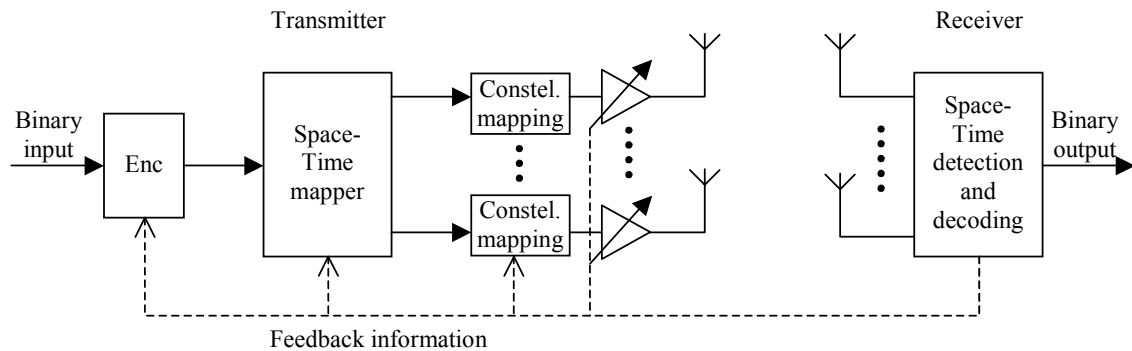
Figure 4-1: General structure of a MIMO system.

Although, in our opinion, the introduced general MIMO scheme can cover (most of) the MIMO algorithms reported in literature, generally, a number of distinctions are made to classify the different algorithms. The commonly used classifications are:

- *Open-loop* versus *closed-loop* techniques. The distinction is made between systems that do not rely on knowledge of the channel responses at the transmitter, i.e., open-loop schemes, and systems that do assume partial or full availability of the channel information at the TX through some form of feedback mechanism, i.e., closed-loop schemes. In general, the feedback loop is designed to provide information for selection of the coding rate, constellation size, type of space-time mapping, and/or TX power per antenna (see Figure 4-1).

- *Transmit diversity* versus *spatial multiplexing* algorithms. If the wireless communication channel is richly scattered, a distinction can be made depending on to what extent the algorithms exploit the transmit diversity provided by the channel. On the one hand, transmit diversity schemes fully use the spatial dimension for adding more redundancy, thus, keeping the data rate equivalent to a single antenna system, with the goal to increase robustness. When the redundancy is generated through coding over the spatial and temporal dimension, the principle is called Space-Time Coding. On the other hand, spatial multiplexing algorithms exploit the spatial dimension by transmitting multiple data streams in parallel on different antennas, with the goal to achieve high data rates ([35, 36, 132, 144]). These algorithms are referred to as Space Division Multiplexing (SDM) algorithms. Note that the distinction is not very strict, since hybrid schemes can be envisioned that combine transmit diversity and spatial multiplexing and partly benefit from both robustness and data rate enhancement. In [73], the possible trade-off between what is called *diversity gain* and *spatial multiplexing gain* is theoretically analysed, and an example is given in [19]. Furthermore, it should be mentioned that a MIMO channel could be ill conditioned in a MIMO capacity sense (e.g., see the LOS cases in Chapter 2). These cases call for beamforming type of approaches, both at the transmitter and receiver side, aiming at maximising the signal-to-noise ratio by forming proper beams. When the antennas are only partially correlated, hybrid schemes of transmit diversity/beamforming may be

applicable ([44, 87]). So, beamforming can be included in the context of the unified framework.

-   *Joint Coding (JC)* versus *Per-Antenna Coding (PAC)*. When the original bit stream is first encoded and then demultiplexed into coded substreams of which each is modulated and mapped onto the corresponding transmit antenna, it is called Joint Coding (or vertical encoding, [146]). With Per-Antenna Coding (or horizontal encoding, [146]), the original bit stream is first demultiplexed into a number of uncoded bit substreams which are then individually encoded, modulated and mapped onto the transmit antennas ([130]). The advantage of the former is that the coding is performed over the space and time dimension, which could result in a better performance than the latter. The advantage of the latter, however, is that its receiver architecture might be less complex, since the encoding over the time and spatial dimension are separated. Note that from the description of PAC it seems that the general diagram of Figure 4-1 does not cover the latter case and that the order of encoding and Space-Time mapping should be changed. When complexity is not taken into account, however, it is not difficult to imagine that the encoder internally already can perform the demultiplexing and encoding per substream and, to make the picture complete, multiplexes the result again, such that the "official" demultiplexing still is done by the Space-Time mapper. In this way, the general structure also holds for PAC schemes.

These different flavours of transmission approaches result in a multiplicity of candidate transmission schemes. Combined with corresponding receiver techniques, these schemes offer a variety of trade-offs between capacity performance, frame error-rate performance, computational complexity/simplicity and sensitivity to channel and/or interference estimation mismatch. An overview of some standard MIMO TX techniques is given in the next subsections.

4.2.2 Space-Time Coding (STC)

As explained before, in Space-Time Coding techniques the coding is performed over the spatial and temporal dimension. In this way, the spectral efficiency of MIMO is exploited by adding extra redundancy to improve the performance/robustness.

The question emerges how to efficiently code over the two dimensions, space and time. As an answer to this question, some basic code design criteria were defined in [39] and [116] by means of an upperbound analysis on the pairwise error probability. Given two codewords $\mathbf{C}$ and $\mathbf{E}$, it is found that the following two quantities should be maximised:

-   the *diversity order*, which determines the exponential decay of the error-rate with the SNR (on a linear scale). For independent fading per channel element, the diversity order is shown to be $r \cdot N_r$, where r equals the minimum rank of the codeword difference matrix $\mathbf{C} - \mathbf{E}$ over all possible codeword pairs. Thus, to maximise the diversity order, r must be maximised. Therefore, this criterion is called the *rank criterion*.

-   the *coding gain*, which defines the SNR gain compared to an uncoded scheme with the same diversity order. The coding gain is given by the minimum product of the nonzero

eigenvalues of $(\mathbf{C} - \mathbf{E})(\mathbf{C} - \mathbf{E})^H$ over all distinct pairs of codewords. Since the product of eigenvalues equals the determinant, this criterion is often called the *determinant criterion*.

These criteria are difficult to relate to traditional code designs ([90]) and, thus, many space-time codes are simply handcrafted using computer search methods. More general design rules are presented in [45], albeit only for small PSK constellations.

Recently, however, it has been shown that under certain properties the traditional code design criterion of maximising the minimum Euclidean distance ($\|\mathbf{C} - \mathbf{E}\|$) between any pair of codewords is more appropriate:

- In [18, 63] it is shown analytically that this design rule is applicable when $N_r \to \infty$, or when $N_t \to \infty$ and $N_r \to \infty$ and the ratio between N_t and N_r is finite, while computer simulations show that this result holds even for a number of antennas as low as $N_t = N_r = 4$.

- In [148] it is shown that this criterion applies when the diversity order $r \cdot N_r \geq 4$, or, in other words, when the product of the number of transmit and receive antennas is relatively large.

This can be explained by the fact that, when a reasonably large diversity gain is provided by transmit and/or receive diversity, a MIMO fading channel converges to a Gaussian channel, provided proper encoding is applied across the diversity dimensions. As a result, under above properties standard SISO codes together with some form of spatial multiplexing may outperform handcrafted Space-Time Codes.

The rank criterion, the determinant criterion, and, later, the Euclidean distance criterion, have stimulated various design activities, which have resulted in a number of different space-time codes, namely, Space-Time Block Codes (STBCs), Space-Time Trellis Codes (STTCs), Space-Time Turbo Codes and Linear Dispersive Codes.

Space-Time Block Codes

In Space-Time Block Codes, the input signal is assumed to be a stream of modulated symbols from a real or complex constellation. The STBC encoder maps these symbols onto codewords that span a block in both the spatial and temporal dimension. The encoder operates on a block of, say, Y input symbols producing an $N_t \times N_s$ codeword $\mathbf{C}$ whose rows correspond to transmit antennas and columns correspond to symbol time instants ([75]):

$$\mathbf{C} = \begin{pmatrix} c_{11} & \cdots & c_{1N_s} \\ \vdots & \cdots & \vdots \\ c_{N_t 1} & \cdots & c_{N_t N_s} \end{pmatrix}, \tag{4.1}$$

The rate of STBCs is defined as $R = Y/N_s$. This definition is chosen such that when N_t and, correspondingly, the number of input symbols Y grows, R improves related to the fact that number of transmitted symbols per unity of time increases. From an efficiency point of view, the codes should be designed such that they achieve a rate as high as possible. On the

other hand, redundancy is added to obtain a robust communication link. Both goals cannot always be achieved at the same time. For instance, *Orthogonal STBCs* seek to fully exploit the available transmit diversity, however, sometimes at the expense of data rate/coding rate. Namely, in [42] it is shown that for *linear complex* orthogonal designs with more than two antennas, the rate cannot be larger than 4/5. "Linear" in this case means that the encoder consists of a linear operation by which the input bits are transformed to ST codewords. Examples of orthogonal STBCs for more than two TX antennas (achieving part of this rate) can be found in [115] and [140]. When *non-linear* processing is allowed, however, rate-1 full diversity schemes can be found for $N_t = 3$ and $N_t = 4$ with Quadrature Phase Shift Keying (QPSK) modulation, see [72]. For the design of *real* orthogonal STBCs, full diversity rate-1 codes exist for all number of transmit antennas ([32]).

Regarding the efficiency it should be noted that orthogonal STBCs do not always fully exploit the available MIMO channel capacity. In [101] it is shown that a space-time block code is optimal with respect to capacity when it is rate one and it is used over a channel of rank one. In other words, only rate-1 STBCs, used over any channel with one receive antenna, are optimal with respect to capacity.

Since the rate of complex orthogonal STBCs with more that two transmit antennas drops below one, the result is a capacity penalty. For this reason *non-orthogonal STBCs* have been designed, which are able to achieve rate one, but at the expense of performance. Examples can be found in [60, 61, 83, 122].

When we have a close look at STBC encoders, we can observe that they, in a sense, act like space-time mappers or interleavers and thus always require the assistance of a one-dimensional outer code.

In [50], linear space-time block codes are introduced that are designed to optimise the mutual information between transmit and receive signals. These codes are called *Linear Dispersive Codes*. They subsume many STBCs as special cases and, like those, necessitate the use of outer one-dimensional coding. Furthermore, like STBCs, they constrain the channel capacity to some extent. When $N_t \leq N_r$, they revert to Space Division Multiplexing (SDM) architectures. Therefore, their domain of interest is scenarios with $N_t > N_r$.

One of the most popular orthogonal space-time block codes is the *Alamouti scheme* for two transmit antennas ([5]). In this scheme, the data is transmitted as is shown in Figure 4-2. At a given symbol period, two symbols s_1 and s_2 are transmitted simultaneously on TX 1 and TX 2, respectively. During the next symbol period, $-s_2^*$ and s_1^* are transmitted on TX 1 and TX 2.

Under the assumption that the channel is a flat-fading channel and constant during the two symbol periods, using the notation of Subsection 3.5.1, the signal at the q-th RX antenna at symbol time 1 and 2, $x_q(1)$ and $x_q(2)$, respectively, can be expressed as a function of an orthogonal code matrix:

$$\begin{pmatrix} x_q(1) \\ x_q(2) \end{pmatrix} = \begin{pmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{pmatrix} \begin{pmatrix} h_{q1} \\ h_{q2} \end{pmatrix} + \begin{pmatrix} n_q(1) \\ n_q(2) \end{pmatrix}, \tag{4.2}$$

where $n_q(t)$ represents the additive noise at symbol time t and h_{qp} denotes the channel between TX p and RX q. Alternatively, the signal at RX q can be given as function of an equivalent orthogonal channel matrix:

$$\begin{pmatrix} x_q(1) \\ x_q^*(2) \end{pmatrix} = \begin{pmatrix} h_{q1} & h_{q2} \\ h_{q2}^* & -h_{q1}^* \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \begin{pmatrix} n_q(1) \\ n_q^*(2) \end{pmatrix}. \tag{4.3}$$

When this equation is left multiplied by the Hermitian transpose of the orthogonal channel, scaled estimates of s_1 and s_2 are retrieved:

$$\begin{pmatrix} h_{q1}^* & h_{q2} \\ h_{q2}^* & -h_{q1} \end{pmatrix} \begin{pmatrix} x_q(1) \\ x_q^*(2) \end{pmatrix} = \left(|h_{q1}|^2 + |h_{q2}|^2 \right) \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \begin{pmatrix} h_{q1}^* n_q(1) + h_{q2} n_q^*(2) \\ h_{q2}^* n_q(1) - h_{q1} n_q^*(2) \end{pmatrix}. \tag{4.4}$$

When the receiver consists of only a single RX antenna, $s_{1,\text{est}}$ and $s_{2,\text{est}}$ can be obtained through straightforward slicing, keeping the proper scaling in mind. Otherwise, the signals related to the multiple antennas can be combined using the Maximum Ratio Combining (MRC) principle ([90]), after which the estimates of s_1 and s_2 can be found to be

$$\sum_{q=1}^{N_r} \begin{pmatrix} h_{q1}^* & h_{q2} \\ h_{q2}^* & -h_{q1} \end{pmatrix} \begin{pmatrix} x_q(1) \\ x_q^*(2) \end{pmatrix} = \left(\sum_{q=1}^{N_r} \sum_{p=1}^{2} |h_{qp}|^2 \right) \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \sum_{q=1}^{N_r} \begin{pmatrix} h_{q1}^* n_q(1) + h_{q2} n_q^*(2) \\ h_{q2}^* n_q(1) - h_{q1} n_q^*(2) \end{pmatrix}. \tag{4.5}$$

Through this processing, the Alamouti scheme can achieve a diversity order of $2N_r$, depending on the channel properties.
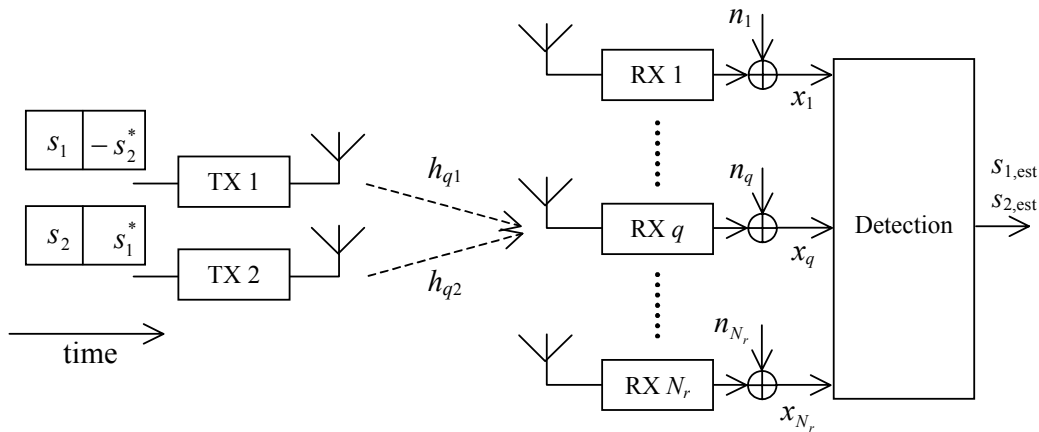
Figure 4-2: The Alamouti STBC scheme.

Recalling that in [101] it is shown that a space-time block code is optimal with respect to capacity when it is rate one and it is used over a channel of rank one, it is easy to verify that the Alamouti scheme is optimal for the case that $N_r = 1$. Moreover, in [73] it is proven that the 2×1 Alamouti scheme is providing an optimal trade-off between diversity gain and multiplexing gain.

When we compare the transmitter part of Figure 4-2 with that of Figure 4-1, it becomes clear that the Alamouti scheme does not directly fit onto the general block diagram since

the Alamouti scheme requires modulated symbols as input to the encoder/space-time mapper. It is possible, however, to redesign the encoder such that it also works on bit level. As a result, the modulation can be performed after the encoding. Without loss of generality, this is shown for QPSK modulation[1] in Figure 4-3. Assume that two subsequent input bits of the constellation mapping are mapped to QPSK constellation points using the mapping of Table 4-1.

Table 4-1: QPSK constellation mapping.

Input bits, e.g., $b_0 b_1$:	00	01	10	11
Output symbol, e.g., s_1:	$\frac{1}{\sqrt{2}}(-1-j)$	$\frac{1}{\sqrt{2}}(-1+j)$	$\frac{1}{\sqrt{2}}(1-j)$	$\frac{1}{\sqrt{2}}(1+j)$

Furthermore, let $!b$ represent the outcome of the binary NOT operation on the corresponding bit b. Then, if the encoder is designed such that its output to an input sequence $b_0, b_1, b_2, b_3, \dots$ equals $b_0, b_2, b_1, b_3, !b_2, b_0, b_3, !b_1, \dots$ and that the demultiplexer alternates these bits among the two TX branches, the input sequence to the first constellation mapping block is $b_0, b_1, !b_2, b_3, \dots$ and its corresponding output is $s_1, -s_2^*, \dots$ Likewise, we can show that the output of the second constellation mapping block is given by $s_2, s_1^*, \dots$, which is in agreement with Figure 4-2.
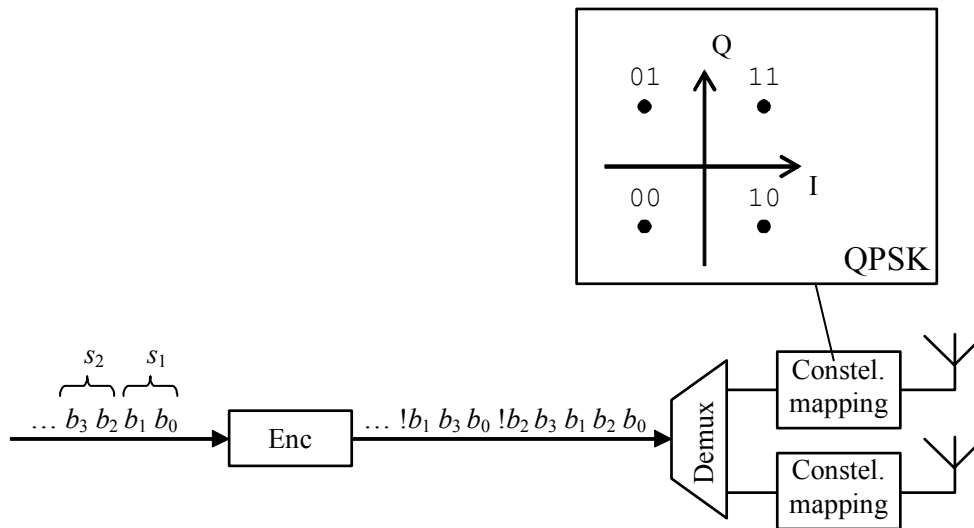
Figure 4-3: Alamouti scheme fitted onto the general structure (see Figure 4-1), for QPSK modulation.

From (4.5), we observe that the receiver for the Alamouti scheme (and in general for all orthogonal designs) can be much simpler than full MLD over the spatial and temporal dimension, since first the channel is decoupled into the scalar channel $\Sigma_q \Sigma_p |h_{qp}|^2$, before the final detection is performed.

Equivalent to the fitting of the Alamouti code to the general structure, other STBCs can be matched to it.

---

[1] Unless mentioned otherwise, we will use Gray mapping. In Gray mapping, the corresponding binary representation of adjacent constellation points only differ in one position.

Space-Time Trellis Codes

Space-Time Trellis Codes (STTCs) were first introduced in [116]. These codes combine the space-time mapping principle of STBCs with proper channel coding and thus they provide significant coding gain, in addition to diversity gain. A disadvantage is that their complexity depends on the number of states, which on its turn grows exponentially with the number of transmit antennas. Therefore, the many contributions in literature following [116] are mostly constrained to two, three or four transmit antennas.

In space-time trellis coding, in general, the incoming bits are first encoded before they are mapped into the space-time format, and modulated. This order perfectly matches that of the general structure of the transmitter of Figure 4-1. Hence, STTCs fit within the unified view on MIMO systems of this section.

As example, we will show the match for the 8-state QPSK code proposed in [116]. The trellis diagram together with the outputs of the encoder is given in Figure 4-4a. It is assumed that the input to the encoder consists of two bits at a time. The encoder state is represented binary by $D_2D_1D_0$ and the two input bits result in four possible transitions per state. The four transitions result in four possible encoder outputs, which are listed per input state in Figure 4-4a. The outputs are denoted as a combination of two quaternary numbers that represent the QPSK symbols to be transmitted on TX 1 and TX 2, respectively. A possible implementation scheme of this code can be found in Figure 4-4b. When comparing this figure with Figure 4-1, the match with the general block diagram is obvious. The best decoder is the full MLD operating over the spatial and temporal dimension, which can be implemented efficiently and without much loss by the Viterbi decoder ([116]). Note, however, that its complexity grows exponentially with the number of states.
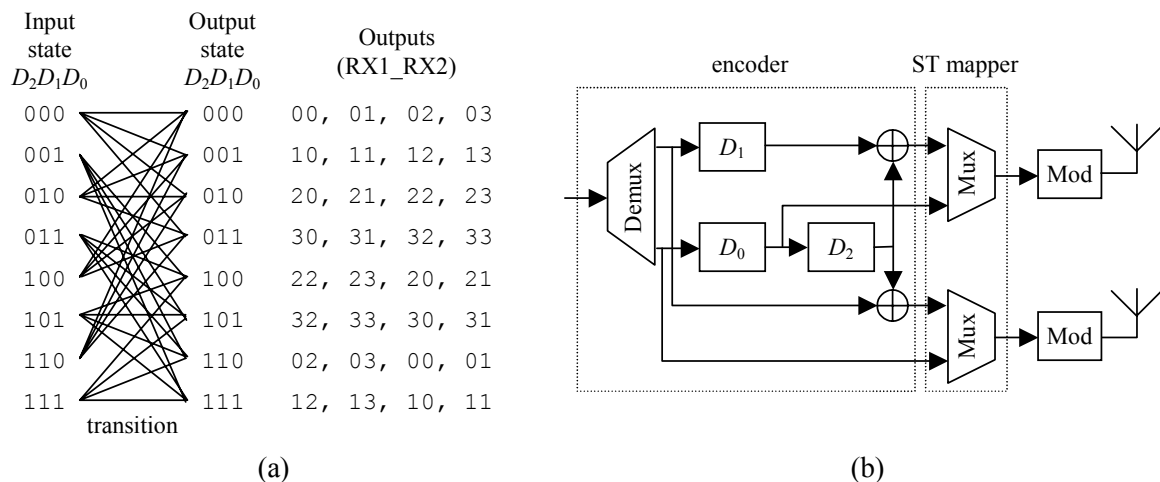
Figure 4-4: (a) The trellis of the 8-state QPSK STTC proposed in [116] and (b) its mapping to the general block diagram.

Space-Time Turbo Codes

It is possible to replace the encoder of a space-time code by a turbo encoder, to apply the turbo principle that generated much attention for the SISO channel ([16, 17]) to the space-

time domain. The multiple outputs of a turbo code can be launched from multiple TX antennas ([28, 124]), or, alternatively, the MIMO mapping itself can be regarded as inner code that can be concatenated with an outer channel code ([55, 111]). At the receiver, the turbo-decoding principle is applied by performing iterative decoding between the concatenated codes. This could be an attractive alternative to the full space-time MLD of STTCs, which may achieve almost the same performance at lower complexity, but at the expense of latency.

Equivalent to the results for STBCs and STTCs, it is possible to find a mapping of the space-time turbo codes onto the general framework.

4.2.3 Space Division Multiplexing (SDM)

Instead of exploiting the spatial dimension by introducing more redundancy in order to enhance the robustness like in Space-Time Coding, the multiple antennas can also be used to increase the data rate. The latter can be achieved by simultaneously transmitting different streams of data on the different transmit antennas (at the same carrier frequency). Although these parallel data streams are mixed up in the air, when the MIMO channel is well conditioned they can be recovered at the receiver by using spatial sampling (i.e., multiple receive antennas) and corresponding signal-processing algorithms. This technique is referred to as *Space Division Multiplexing (SDM)*. The main advantage of SDM is that it directly exploits the MIMO channel capacity to improve the data rate. The main disadvantage is that no redundancy is added and, thus, it might suffer from a poor link reliability. To overcome this problem additional channel coding can be introduced. This, however, reduces its data rate advantage.

Recall that at the beginning of Subsection 4.2.2 it has been cited that, when the number of antennas and diversity potential of the channel are large enough, the probability of error appears to depend only on the Euclidean distance of the code. This would indicate that a one-dimensional code designed for AWGN channels, of which its codewords are properly interleaved across the space and time, may be as effective as a space-time code designed according to the rank and determinant criterion. In [148] it is shown that this would already be the case when the diversity order is equal to or larger than four, which would restrict the domain of interest of space-time coding to architectures with only 2–3 antennas.

Besides the data rate advantage of SDM, also its decoding complexity might not be as complex as for space-time codes, especially when the number of transmit antennas is relatively large. This can be explained by the fact that, in general, space-time codes require joint detection over the spatial and temporal dimension of the signals radiated by the various transmit antennas, resulting in a complexity that explodes when the number of TX antennas increases. With SDM it is, however, generally possible to separate the spatial and temporal processing. This will (most likely) result in a performance degradation. A possible way to overcome this, at the expense of latency, is to apply the turbo-decoding principle and iterate between the spatial mapping/code (i.e., the inner code) and the temporal code (i.e., the outer code), see [53, 106, 123, 138].

When encoding is applied to SDM, the coding of the input data can be done before or after the demultiplexing, which is referred to as Joint Coding and Per-Antenna Coding,

respectively (see Subsection 4.2.1). In the concept of PAC, the TX antennas can be either co-located or not. Especially the latter option can be seen as multiple access scheme and is generally called Space Division Multiple Access (SDMA) ([121]). From a reception point of view, the two options are not very different and the detection at the receiver can be performed by multi-user detectors operating in the spatial domain. Since for SDMA with dislocated single-branch transmitters, however, each transmitter has its own local oscillator(s), time and frequency synchronisation in case of SDMA is cumbersome ([121]). Therefore, we will focus on SDM.

The concept of PAC (also called layered architecture) was introduced in [35, 84]. One of the main contributions of [35] was to recognize that, because each transmit antenna encounters a different propagation channel, PAC incurs a capacity penalty. Hence, a diagonally layered architecture was proposed (Diagonal Bell-Labs Layered Space-Time (D-BLAST)) in which successive symbols of a given encoded data stream are sent on a different TX antenna, by cyclically selecting another TX antenna per symbol period (see Figure 1-5). In this way, each data stream is exposed to the distinct propagation channels within the MIMO channel. In fact, this eliminates the capacity penalty compared to cases in which no cycling is used ([8]).

Despite the losslessness of D-BLAST in terms of capacity, the cyclic association between the encoded data streams and transmit antennas makes the decoding of D-BLAST complicated. When this "diagonal" spatial mapping is dropped, the decoding is significantly simplified and can be performed "vertically", i.e., per spatial vector, which, however, most likely will result in a capacity penalty. As mentioned before, the receivers of this PAC principle adopt the form of multi-user detectors operating in the spatial domain. Both linear and non-linear processing is possible. As examples of linear detection, the Zero-Forcing (ZF) and Minimum Mean Squared Error (MMSE) algorithms can be mentioned ([132]). The V-BLAST scheme proposed in [144] and Maximum Likelihood Detection ([132]) are examples of non-linear detectors.

The matching of PAC and joint encoding to the unified structure of Figure 4-1 is already performed in Subsection 4.2.1. Based on this unified view, it is interesting to note that:

- Any PAC scheme can be regarded as a space-time code wherein each codeword $\mathbf{C}$ is made up of N_t independent one-dimensional codewords.

- Conversely, any space-time code can be regarded a PAC scheme wherein the N_t transmit antennas radiate portions of a single codeword $\mathbf{C}$.

Furthermore, note that several activities have been undertaken to combine SDM with STC ([51, 73, 112, 114, 139]). The basic idea behind these hybrid schemes is to form space-time codewords $\mathbf{C}$ using n one-dimensional codewords, with $1 \leq n \leq N_t$. This idea may be used as a starting point to further work out the presented conceptual unified view.

4.2.4 Discussion

In the previous subsections we have seen that we can map the various MIMO TX techniques onto a general TX structure. A logical question is how this reflects onto the

receiver. It is commonly known that at the receiver side the best performance can be achieved when a full maximum likelihood search is performed over the complete dimensions that are spanned by the space-time encoding process. It is obvious that the complexity of such a receiver grows exponentially with the size of the spatial and temporal dimension. Especially when the space-time codeword sizes result in unmanageable complexity, a search for less complex RX architectures is required. Sometimes the TX signal structure allows for less complex decoding while still achieving full performance, like for some STBCs (see Subsection 4.2.2), but in general less complex RX schemes result in a performance loss. The goal is to design a receiver with a manageable complexity that performs closely to the maximum likelihood bound.

Following the general TX structure given in Figure 4-1, an RX structure as shown in Figure 4-5 can be envisioned. When the encoding is seen as encoding over time, the receiver processing is split in spatial and temporal processing. Because the processing over the two dimensions is separated, a performance loss is inevitable. But, as we will see in Subsection 4.11.2, for certain conditions the loss is manageable. Moreover, the performance can be improved by applying the principle of turbo-decoding processing ([16, 17]). In our case, the turbo-decoding principle can be applied by iterating over the spatial and temporal detection/decoding. With this principle, in general, the performance gets very close to that of the full maximum likelihood search, at the cost of latency.
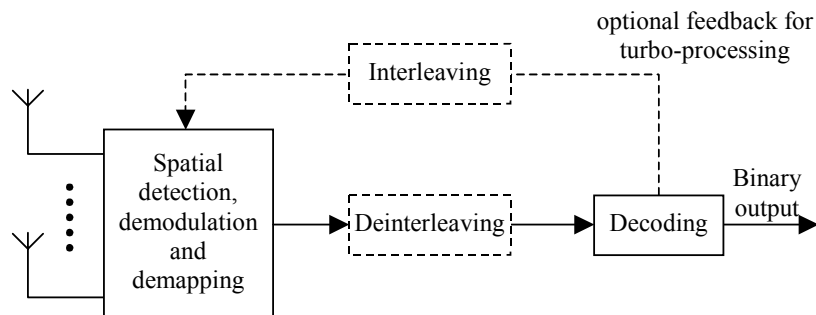
Figure 4-5: General RX architecture.

Besides the above proposed low-complex receiver architecture, a number of other observations made in the previous subsections speak in favour of separately treating the spatial and temporal processing and justify research to SDM:

-   the data rate enhancement of SDM (one of the main goals of this dissertation is searching for schemes that enhance the data rate). To achieve a certain robustness, some temporal outer code may be required.

-   the flexibility of spatial multiplexing with temporal encoding in terms of data rate. With the general structure of Figure 4-1 it is easy to adapt the rate of the encoder and/or the size of the constellation scheme, resulting in variable performance and data rate.

-   the Euclidean distance design rule. When the diversity order is large enough, the Euclidean distance criterion holds as a design rule and simply demultiplexing a code optimised for AWGN channels over the spatial dimension might outperform handcrafted STCs (designed following the rank and determinant criterions). In

[148] it is shown that this would already be the case when the diversity order is equal to or larger than four, which would restrict the domain of interest of space-time coding to architectures with only 2–3 antennas. Note that the spatial processing of such a scheme can be seen as SDM and above RX structure can be used for detection.

Overall, we can conclude that, based on the unified view presented in this section, the inner processing of most MIMO schemes is equivalent to SDM and the outer processing is some form of spatial mapping and temporal encoding as outer processing. As a result, (coded) SDM may be also applicable in the context of STC.

4.3 The Single-Carrier MIMO Signal Model

In this section, a baseband equivalent signal model for a Multiple-Input Multiple-Output (MIMO) communication system is stated in which the communication channel bandwidth is assumed to be so narrow that the channel can be treated as flat with frequency (i.e., flat fading).

A communication system comprising N_t transmit (TX) and N_r receive (RX) antennas is considered. It is assumed to operate in a flat-fading environment and exploits the spatial dimension by the MIMO technology (see Figure 4-6). At discrete times, the transmitter sends an N_t-dimensional (complex) signal vector $\mathbf{s}$. Note that the elements of $\mathbf{s}$ are sent at the same carrier frequency. The receiver records an N_r-dimensional complex vector $\mathbf{x}$. Then, recalling from subsection 3.5.1, the following baseband equivalent signal model describes the relation between $\mathbf{s}$ and $\mathbf{x}$:

$$\mathbf{x} = \mathbf{Hs} + \mathbf{n} ,\tag{4.6}$$

where $\mathbf{H}$ is an $N_r \times N_t$ complex propagation matrix that is assumed constant for the length of a packet transmission (i.e., a quasi-static channel) and assumed known at the receiver (e.g., via transmission of training sequences). It is assumed that the statistics of the channel transfer matrix $\mathbf{H}$ can be described by the fading statistics introduced in Section 3.5, namely, Rayleigh fading, Ricean fading or AWGN. Furthermore, it is assumed that the elements of $\mathbf{H}$ have a variance of one, or, in other words, the average channel gain P_c is normalised to one.
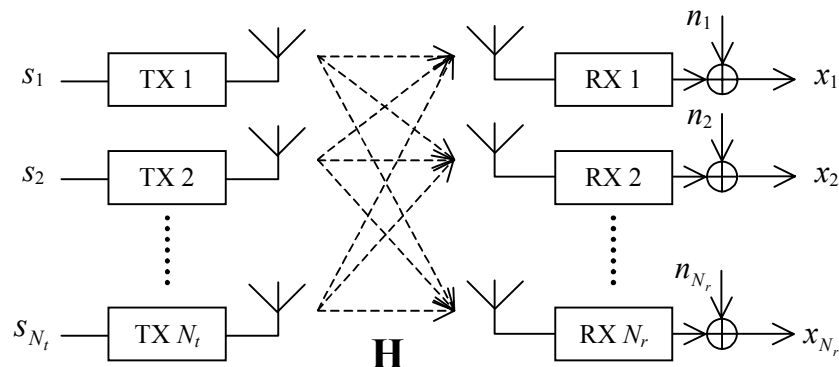
Figure 4-6: The physical model of a MIMO system.

The N_r-dimensional vector $\mathbf{n}$ represents zero mean ($\boldsymbol{\mu}_n = \mathbf{0}$), i.i.d. complex Additive White Gaussian Noise (AWGN) with a variance σ_n^2 per element. This means that $\mathbf{n}$ follows a complex multivariate normal (or Gaussian) distribution (see Appendix A.2) and its probability density function (pdf) equals

$$p(\mathbf{n}) = \det(\pi \mathbf{Q}_n)^{-1} \exp\left(-(\mathbf{n} - \boldsymbol{\mu}_n)^H \mathbf{Q}_n^{-1}(\mathbf{n} - \boldsymbol{\mu}_n)\right). \tag{4.7}$$

Its covariance matrix equals

$$\mathbf{Q}_n = E\left[(\mathbf{n} - \boldsymbol{\mu}_n)(\mathbf{n} - \boldsymbol{\mu}_n)^H\right] = E\left[\mathbf{n}\mathbf{n}^H\right] = \sigma_n^2 \mathbf{I}_{N_r}. \tag{4.8}$$

With the focus on spatial multiplexing in mind, unless mentioned otherwise, the total transmit power P_t is assumed to be uniformly distributed among the different transmit antennas. More precisely, the vector $\mathbf{s}$ is assumed to have zero-mean, uncorrelated variables with equal variance, σ_s^2, and the total transmit power is $E[\mathbf{s}^H\mathbf{s}] = N_t\sigma_s^2 = P_t$. The covariance matrix of $\mathbf{s}$ equals

$$\mathbf{Q}_s = E\left[\mathbf{s}\mathbf{s}^H\right] = \sigma_s^2 \mathbf{I}_{N_t} = \frac{P_t}{N_t} \mathbf{I}_{N_t}. \tag{4.9}$$

Furthermore, the vectors $\mathbf{s}$ and $\mathbf{n}$ are assumed to be independent ($E[\mathbf{s}\mathbf{n}^H] = \mathbf{0}$). With this information, the covariance matrix of $\mathbf{x}$ for a given $\mathbf{H}$ can be found, using the fact that, if both $\mathbf{s}$ and $\mathbf{n}$ are zero mean, $\mathbf{x}$ is zero mean as well:

$$\begin{aligned}
\mathbf{Q}_x &= E\left[\mathbf{x}\mathbf{x}^H\right] = E\left[(\mathbf{H}\mathbf{s} + \mathbf{n})(\mathbf{H}\mathbf{s} + \mathbf{n})^H\right] \\
&= \mathbf{H}\mathbf{Q}_s\mathbf{H}^H + \mathbf{Q}_n = \sigma_s^2 \mathbf{H}\mathbf{H}^H + \sigma_n^2 \mathbf{I}_{N_r}.
\end{aligned} \tag{4.10}$$

Furthermore, the expected Signal-to-Noise Ratio (SNR) over the ensemble of all possible realisations for the q-th receive antenna, i.e., the average SNR for the q-th component of $\mathbf{x}$, can be found and, with $P_c = 1$, equals

$$\rho_q = \frac{E_s}{N_0} = \frac{\sigma_s^2 \left(E\left[\mathbf{H}^H\mathbf{H}\right]\right)_{qq}}{(\mathbf{Q}_n)_{qq}} = \frac{N_t\sigma_s^2 P_c}{\sigma_n^2} = \frac{P_t}{\sigma_n^2}, \tag{4.11}$$

where E_s represents the average signal power per receive antenna, N_0 denotes the average noise power per receive antenna, and $(.)_{qq}$ represents the (q,q)-th element of the corresponding matrix. The expected SNR per RX antenna, ρ_q, is assumed identical for all N_r receive antennas and will be written as ρ.

As mentioned above, it is assumed that (an estimate of) the channel transfer matrix $\mathbf{H}$ is available at the receiver, i.e., the receiver knows the Channel State Information (CSI). A common way to obtain the CSI at the receiver is by sending a preamble containing known training sequences in front of the payload, i.e., the data packet, and using these sequences to estimate the channel coefficients. Since the channel is assumed quasi-static, these channel coefficients can be used throughout the payload to retrieve the transmitted data.

Besides training sequences the preamble usually also consists of synchronisation symbols. Synchronisation is an essential task for any digital communication system and required for reliable reception of the transmitted data, but in this chapter, synchronisation is said to be perfect, in order to focus on its main topic, that is, performance comparison of MIMO algorithms. Furthermore, unless mentioned otherwise, it is assumed that the receiver perfectly knows the CSI.

4.4 Capacity

4.4.1 Definition of Capacity

One way to express the gain of a MIMO system over a SISO system is by means of the capacity. In general, the capacity is defined by information theory as an upperbound on the information rate for error-free communication. The question is what the capacity gain of a MIMO system is under certain conditions. The capacity of a MIMO communication link depends not only on the fading statistics, as for a SISO link, but also on the spatial correlation of the channel. This results in a random capacity whose instant value depends on the corresponding instantaneous $\mathbf{H}$ matrix. An ensemble of $\mathbf{H}$ matrices results for a given average SNR per RX antenna in a cumulative distribution function (cdf) of the capacity. In general, however, the complementary cdf is used because then, e.g., the 99% point denotes that for 99% of the instants of channel use the corresponding capacity can be achieved. So, note that when the transmitter does not have channel knowledge, for any information or bit rate it chooses there is a nonzero probability that an instantaneous $\mathbf{H}$ is incapable of supporting it even when ideal channel coding for that chosen rate is employed. When the instant capacity is less than the chosen rate, a channel outage is said to occur. This leads to an *outage probability*. For a required outage probability, say 1%, the corresponding capacity value corresponds to the 99% point of the complementary cdf of the capacity. Furthermore, the *ergodic capacity* is defined as the average capacity over the distribution of $\mathbf{H}$.

In general, the challenge for both SISO and MIMO systems is to design efficient coding/decoding algorithms that can approach the information theory bound on bit rate (i.e., capacity) as close as possible, preferable with a complexity that makes implementation feasible. To quantify this bound, in the next subsections, capacity expressions and the corresponding outage performance are introduced.

4.4.2 Physical Interpretation

A nice and intuitive way to visualise the physical interpretation of a given channel transfer matrix and its impact on the channel capacity is given in [117] and is based on the Singular Value Decomposition (SVD) (Appendix A.1.4) of a given channel matrix $\mathbf{H}$. The SVD of $\mathbf{H}$ equals

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^H, \tag{4.12}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices with dimensions $N_r \times N_r$ and $N_t \times N_t$, respectively. $\mathbf{D}$ is an $N_r \times N_t$ dimensional diagonal matrix having the singular values of $\mathbf{H}$ on its diagonal. These singular values are the square roots of the nonzero eigenvalues λ_k of $\mathbf{HH}^H$ or $\mathbf{H}^H\mathbf{H}$ ([113]), with $k = 1, \ldots, N_k$, where $N_k = \text{rank}(\mathbf{HH}^H) \leq \min(N_t, N_r)$ denotes the *rank* of the matrix $\mathbf{HH}^H$. Based on the SVD, we thus can rewrite (4.6) as

$$\mathbf{x} = \mathbf{UDV}^H\mathbf{s} + \mathbf{n} . \qquad (4.13)$$

Assume we transmit $\mathbf{s}' = \mathbf{Vs}$ instead of $\mathbf{s}$, and at the receiver, the received vector $\mathbf{x}$ is multiplied by $\mathbf{U}^H$. Then this results in

$$\mathbf{y} = \mathbf{U}^H\mathbf{x} = \mathbf{U}^H\mathbf{UDV}^H\mathbf{Vs} + \mathbf{U}^H\mathbf{n} = \mathbf{Ds} + \mathbf{n}' . \qquad (4.14)$$

Note that a matrix or vector that is multiplied by a unitary matrix results in a matrix or vector that is transformed from one set of basis vectors that span a space to another set of basis vectors. So, multiplying by a unitary matrix can be seen as a rotation, and therefore, the channel capacity does not change by above operations. Based on the same arguments, the multiplication of the noise vector $\mathbf{n}$ by a unitary matrix does not affect its distribution.

A component-wise notation of (4.14) results in

$$y_k = \lambda_k^{\frac{1}{2}} s_k + n'_k . \qquad (4.15)$$

In this case, the equivalence of the physical model of Figure 4-6 is given in Figure 4-7, where it is shown that the equivalence of $\mathbf{H}$ consists of N_k parallel spatial subchannels with the k-th eigenvalue, λ_k, as gain for the k-th subchannel ([6]).
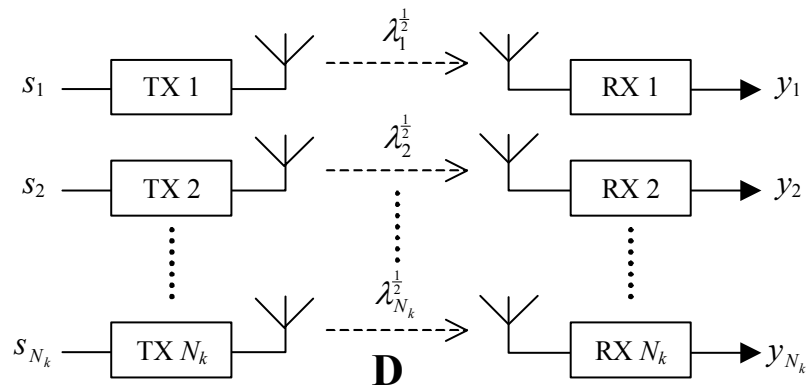
Figure 4-7: Equivalence of the physical model of a MIMO system.

Since the rank equals the number of nonzero eigenvalues, it represents the number of available spatial subchannels. The number of spatial subchannels (or eigenmodes) indicates the number of parallel symbol streams that can be transmitted through the MIMO channel, using the same frequency bandwidth, and is hence a measure of the capacity of the MIMO channel. To find this capacity, denote the average SNR at the k-th RX antenna by ρ_k and the transfer function of the k-th subchannel by h_k. Then, by using Shannon's capacity formula ([107]), we can find the total capacity per unit of bandwidth to be

$$C = \sum_{k=1}^{N_k} \log_2\left(1 + \rho_k |h_k|^2\right) = \sum_{k=1}^{N_k} \log_2\left(1 + \lambda_k \frac{\sigma_k^2}{\sigma_n^2}\right) \text{ bits/s/Hz,} \qquad (4.16)$$

where σ_k^2 is the power allocated to the k-th subchannel. Clearly, the capacity depends on the power allocation per subchannel as will be explained in more detail in Subsections 4.4.4 and 4.4.5. Furthermore, when the gains λ_k would be equal and the power allocation uniform, we can observe that the capacity grows linearly with N_k.

With spatial multiplexing, the virtual subchannels of a MIMO channel are exploited by sending independent data streams on multiple transmit antennas to improve data rates. The spread of the eigenvalues λ_k, with $k = 1, \dots, N_k$, is a measure of the orthogonality of the MIMO channel. A large eigenvalue spread means that the channel matrix is highly non-orthogonal and vice-versa. Symmetric orthogonal channels are desirable since they do not have null modes, i.e., eigenvalues equal to zero, and hence do not lose transmitted information. Moreover, these channels can be inverted in the receiver without noise amplification, leading to a good system performance.

Two concrete measures for the orthogonality of a MIMO channel are the condition number and Effective Degrees Of Freedom (EDOF). The *condition number* of a matrix is defined by the ratio of its largest and smallest nonzero eigenvalues ([113]). We will define the condition number of a MIMO channel by the condition number of $\mathbf{H}^H\mathbf{H}$ and assume that its N_k nonzero eigenvalues are sorted in decreasing order of magnitude, then the condition number equals

$$\kappa\left(\mathbf{H}^H\mathbf{H}\right) = \frac{\lambda_1}{\lambda_{N_k}} . \qquad (4.17)$$

A condition number of unity means that the channel matrix $\mathbf{H}$ is orthogonal. A large condition number, however, implies that the channel is highly non-orthogonal or *ill conditioned*, resulting in a poor channel capacity.

The *Effective Degrees Of Freedom* (EDOF) represents the number of subchannels actively participating in conveying information over a given wireless MIMO link. For a SISO channel h, the capacity is given by $C = \log_2(1 + \rho|h|^2)$, and it is obvious that at a high SNR, ρ, an a-fold increase in SNR results in a capacity improvement of approximately $\log_2(a)$ bit/s/Hz. For a MIMO system, if the capacity increases by EDOF·$\log_2(a)$ bit/s/Hz for a-fold raise of the SNR, then, the EDOF is defined as the number of parallel SISO channels that is required for an equivalent capacity increase. So, when α is the number of factor-2 increases in SNR, we define the EDOF at a given average SNR per receive antenna ρ as the derivative of C versus α at $\alpha = 0$ ([25]):

$$\text{EDOF} \equiv \frac{d}{d\alpha} C\left(2^\alpha \rho\right)\Big|_{\alpha=0} . \qquad (4.18)$$

Note that the EDOF is a real number between one and N_k and that it is determined by the spatial correlation of $\mathbf{H}$. Furthermore, for low SNR values, it is SNR dependent and its value increases when the SNR is increased. For high SNR, it will saturate.

4.4.3 The Capacity Expression

Although the physical interpretation of the previous subsection provides an intuitive capacity expression, we are interested in a more general capacity expression based on information theory. To obtain the theoretical expression of the *information capacity* of a MIMO system for a given channel matrix $\mathbf{H}$, the mutual information between transmitted vector $\mathbf{s}$ and received vector $\mathbf{x}$ should be determined. This mutual information is given by:

$$I(\mathbf{s};\mathbf{x}) = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{s}) = H(\mathbf{x}) - H(\mathbf{n}), \qquad (4.19)$$

where $H(\mathbf{y})$ denotes the entropy of a multivariate distribution $\mathbf{y}$. The capacity equals the maximum mutual information. Maximising $I(\mathbf{s};\mathbf{x})$ is equivalent to maximising $H(\mathbf{x})$. In [117] it is shown that the entropy is largest when $\mathbf{x}$ is complex multivariate normal (or Gaussian) distributed. Therefore, we will assume that $\mathbf{x}$ follows a complex multivariate normal distribution with covariance matrix $\mathbf{Q}_n + \mathbf{HQ}_s\mathbf{H}^H$ and mean $\boldsymbol{\mu}_x = \mathbf{0}$. The pdf of a complex multivariate normal distribution $\mathbf{z}$ with covariance matrix $\mathbf{Q}_z$ and mean $\boldsymbol{\mu}_z$ is given by (see Appendix A.2):

$$p(\mathbf{z}) = \det(\pi\mathbf{Q}_z)^{-1} \exp\left(-(\mathbf{z}-\boldsymbol{\mu}_z)^H \mathbf{Q}_z^{-1}(\mathbf{z}-\boldsymbol{\mu}_z)\right). \qquad (4.20)$$

The entropy of this complex multivariate normal distribution is given by:

$$
\begin{aligned}
H(\mathbf{z}) &= -\int p(\mathbf{z})\log_2 p(\mathbf{z})d\mathbf{z} \\
&= -\int p(\mathbf{z})\log_2\left(\det(\pi\mathbf{Q}_z)^{-1} e^{-(\mathbf{z}-\boldsymbol{\mu}_z)^H \mathbf{Q}_z^{-1}(\mathbf{z}-\boldsymbol{\mu}_z)}\right)d\mathbf{z} \\
&= -\int p(\mathbf{z})\log_2\left(\det(\pi\mathbf{Q}_z)^{-1}\right)d\mathbf{z} - \int p(\mathbf{z})\log_2\left(e^{-(\mathbf{z}-\boldsymbol{\mu}_z)^H \mathbf{Q}_z^{-1}(\mathbf{z}-\boldsymbol{\mu}_z)}\right)d\mathbf{z} \\
&= \log_2\det(\pi\mathbf{Q}_z) + (\log_2 e)\int p(\mathbf{z})(\mathbf{z}-\boldsymbol{\mu}_z)^H \mathbf{Q}_z^{-1}(\mathbf{z}-\boldsymbol{\mu}_z)d\mathbf{z} \\
&= \log_2\det(\pi\mathbf{Q}_z) + (\log_2 e)E\left[(\mathbf{z}-\boldsymbol{\mu}_z)^H \mathbf{Q}_z^{-1}(\mathbf{z}-\boldsymbol{\mu}_z)\right] \\
&= \log_2\det(\pi\mathbf{Q}_z) + (\log_2 e)\mathrm{tr}\left(E\left[(\mathbf{z}-\boldsymbol{\mu}_z)(\mathbf{z}-\boldsymbol{\mu}_z)^H \mathbf{Q}_z^{-1}\right]\right) \\
&= \log_2\det(\pi\mathbf{Q}_z) + (\log_2 e)\mathrm{tr}(\mathbf{I}) \\
&= \log_2\det(\pi e\mathbf{Q}_z),
\end{aligned}
\qquad (4.21)
$$

where the fact is used that $\mathbf{a}^H\mathbf{b} = \mathrm{tr}(\mathbf{ba}^H)$, where $\mathrm{tr}(.)$ stands for the *trace* of a matrix, i.e., the sum of the diagonal elements of a matrix.

Above analysis leads with $\mathbf{Q}_n = \sigma_n^2\mathbf{I}_{N_r}$ to the following capacity expression:

$$
\begin{aligned}
C &= \max(I(\mathbf{s};\mathbf{x})) = \max(H(\mathbf{x})) - H(\mathbf{n}) \\
&= \log_2\det(\pi e\mathbf{Q}_x) - \log_2\det(\pi e\mathbf{Q}_n) \\
&= \log_2\det\left(\pi e\left(\mathbf{Q}_n + \mathbf{HQ}_s\mathbf{H}^H\right)\right) + \log_2\det(\pi e\mathbf{Q}_n)^{-1} \\
&= \log_2\det\left(\mathbf{I}_{N_r} + \frac{1}{\sigma_n^2}\mathbf{HQ}_s\mathbf{H}^H\right) = \log_2\det\left(\mathbf{I}_{N_t} + \frac{1}{\sigma_n^2}\mathbf{Q}_s\mathbf{H}^H\mathbf{H}\right) \text{ bits/s/Hz},
\end{aligned}
\qquad (4.22)
$$

where the last equality follows from the determinant identity $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$. Furthermore, we used the properties that $\det(\mathbf{A})^{-1} = \det(\mathbf{A}^{-1})$ and $\det(\mathbf{A})\det(\mathbf{B}) = \det(\mathbf{AB})$ ([113]). Note that G.J. Foschini arrives at the same result in [36] using a similar analysis. Furthermore, note that the capacity strongly depends on the distribution of $\mathbf{H}$. Therefore, usually the capacity distribution or the ergodic capacity is given. The *ergodic capacity* is simply the average capacity over the distribution of $\mathbf{H}$. In the next subsections, above result is used to determine the capacity of both a closed-loop and open-loop MIMO system.

4.4.4 Closed-loop Capacity

To determine the capacity of a closed-loop scheme in which the receiver feeds back the CSI to the transmitter, optimal feedback with zero delay is assumed. When the transmitter has perfect knowledge of the channel matrix, the total available power P_t can be optimally distributed among the N_t transmit antennas. This solution is usually referred to as *water filling* ([69, 117]).

To obtain the capacity for this case, we must choose a $\mathbf{Q}_s$ in (4.22) that maximises this quantity subject to a total transmit power constraint $\mathrm{tr}(\mathbf{Q}_s) \leq P_t$ ([117]). Based on the SVD of $\mathbf{H}$, we can rewrite (4.22) which leads to the task to find a $\mathbf{Q}_s$ that maximises

$$\log_2\det\left(\mathbf{I}_{N_r} + \frac{1}{\sigma_n^2}\mathbf{HQ}_s\mathbf{H}^H\right) = \log_2\det\left(\mathbf{UU}^H + \frac{1}{\sigma_n^2}\mathbf{UDV}^H\mathbf{Q}_s\mathbf{VD}^H\mathbf{U}^H\right)$$

$$= \log_2\det\left(\mathbf{I}_{N_r} + \frac{1}{\sigma_n^2}\mathbf{DV}^H\mathbf{Q}_s\mathbf{VD}^H\right) \qquad (4.23)$$

$$= \log_2\det(\mathbf{B})\,\text{bits/s/Hz} .$$

According to Hadamard's inequality for an $N \times N$ nonnegative definite matrix $\mathbf{A}$ ([54]),

$$\det(\mathbf{A}) \leq \prod_{i=1}^{N} a_{ii} , \qquad (4.24)$$

with equality when $\mathbf{A}$ is diagonal. In this equation, a_{ii} represents the i-th diagonal element of $\mathbf{A}$. So, in order to maximise $\det(\mathbf{B})$, it must be diagonal. So, choose $\mathbf{Q}_s = \mathbf{VQV}^H$, where $\mathbf{Q}$ is a diagonal matrix with elements q_{kk}. Note that the q_{kk}'s are the eigenvalues of matrix $\mathbf{Q}_s$. The optimal diagonal entries can be found via "water filling" to be ([69, 117])

$$q_{kk} = \left(\mu - \frac{\sigma_n^2}{\lambda_k}\right)^+ , \quad k = 1, \ldots, N_k , \qquad (4.25)$$

where μ is chosen to satisfy $\sum_k q_{kk} = P_t$, and where

$$(x)^+ = \begin{cases} x, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \qquad (4.26)$$

An illustration of the water filling principle is depicted in Figure 4-8. The corresponding maximum capacity is given by

$$C = \sum_{k=1}^{N_k} \log_2\left(1 + \frac{1}{\sigma_n^2} q_{kk} \lambda_k\right).$$ (4.27)
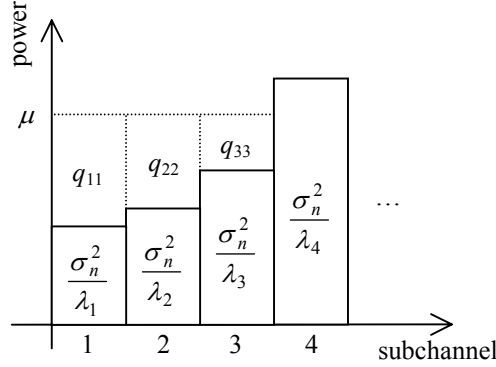
Figure 4-8: The water filling principle.

4.4.5 Open-loop Capacity

If the CSI is not known at the transmitter, it has been shown in [36, 117] that the optimal transmit strategy is to distribute the available power uniformly over the transmit antennas, i.e.,

$$\mathbf{Q}_s = \frac{P_t}{N_t} \mathbf{I}_{N_t}.$$ (4.28)

This leads to the open-loop capacity given by

$$C = \log_2 \det\left(\mathbf{I}_{N_r} + \frac{P_t}{N_t \sigma_n^2} \mathbf{H}\mathbf{H}^H\right) = \log_2 \det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t} \mathbf{H}\mathbf{H}^H\right) \text{bits/s/Hz}.$$ (4.29)

For the open-loop capacity, the following two scenarios can be distinguished:

1. The channel matrix **H** is *orthogonal*. It can be shown that orthogonal MIMO channels have $N_k = \min(N_t, N_r)$ spatial subchannels with equal gain. In other words, the condition number of the channel is one. Therefore, a uniform distribution of the total available transmit power on all subchannels is the optimal transmit strategy and, hence, the closed-loop and open-loop capacity are the same and given by

$$C = \log_2 \det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t} \mathbf{H}\mathbf{H}^H\right) = \sum_{k=1}^{N_k} \log_2\left(1 + \frac{\rho}{N_t} \lambda_k\right) \text{bits/s/Hz}.$$ (4.30)

Since the average power of the elements of **H** is assumed to be one and the channel is orthogonal, it can be shown that for this case the nonzero eigenvalues of $\mathbf{H}\mathbf{H}^H$ and $\mathbf{H}^H\mathbf{H}$ are all equal to $\lambda_k = \max(N_t, N_r)$, $k = 1, \ldots, N_k$. Thus, in this case, the capacity grows linearly with the minimum of the transmit and receive antennas, $N_k = \min(N_t, N_r)$.

2. When the MIMO channel is not perfectly orthogonal, the closed-loop capacity will be higher than the open-loop capacity. The reason is that for non-orthogonal channels, the eigenmodes have unequal gains. To achieve the closed-loop capacity, the optimal transmitter having perfect CSI will use more power on the stronger subchannel and enhance the capacity, whereas the transmitter that applies equal power on all subchannels wastes energy in the null modes of the channel and, thus, loses in capacity. This is illustrated for different antenna configurations in Figure 4-9. In this figure, the average capacity, i.e., ergodic capacity, is determined for 10,000 realisations of **H**, where the elements of **H** are assumed to be i.i.d. circularly-symmetric complex Gaussian distributed (see Subsection 3.5.2). Note that the number of nonzero eigenvalues of $\mathbf{H}\mathbf{H}^H$ or $\mathbf{H}^H\mathbf{H}$ for this type of channels equals $N_k = \min(N_t, N_r)$. From Figure 4-9, we observe that there is a big difference in closed-loop vs. open-loop capacity for the 4×2 case. Such a difference is not observed for the 2×4 case, which we did not plot for clarity. This observation is supported by [69] and the references therein, where it is noted that for $N_r \geq N_t$, the increase in capacity achieved via water filling is only of interest when small number of antennas is employed, and when the receiver observes a low SNR. Significant gains when applying water filling, however, are achieved when $N_t > N_r$. Also for scenarios when high spatial correlation occurs, the water-filling solution might be interesting, as is shown in Figure 4-10 for a 4×4 system and the correlation model of Section 3.6, with $r = r_{TX} = r_{RX}$, where $r = \{0, 0.6, 0.8, 1\}$. We have to conclude, however, that a closed-loop scheme is only providing significant capacity enhancements over an open-loop scheme for scenarios in which spatial multiplexing is most likely not performing properly anyway. Since the focus of this dissertation is on spatial multiplexing, we will focus on open-loop MIMO algorithms.
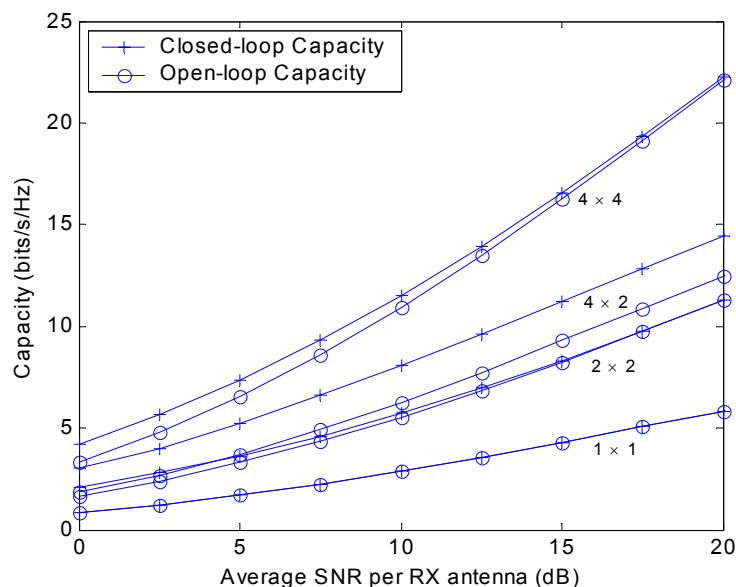
Figure 4-9: Ergodic channel capacity for different MIMO configurations.
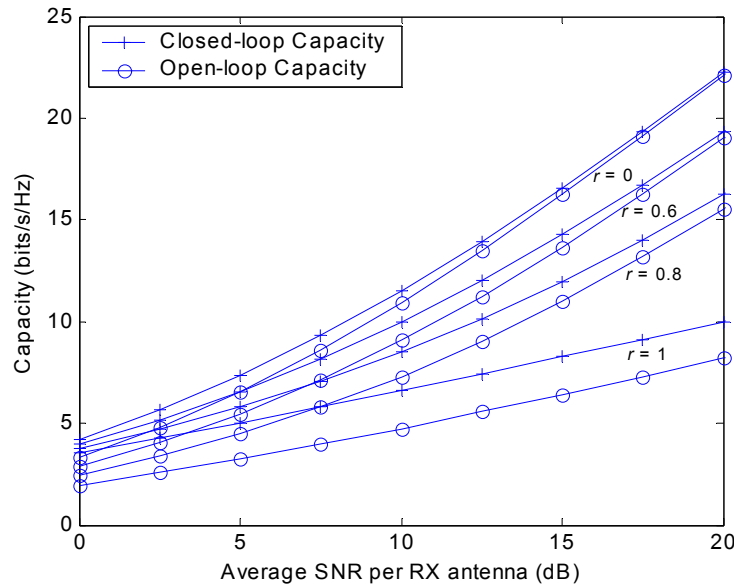
Figure 4-10: Closed-loop and open-loop capacity versus SNR per RX antenna for a 4 × 4 system and different correlation coefficients.

4.4.6 Outage Packet Error Rate Performance

As mentioned already in Subsection 4.4.1, the MIMO channel capacity is typically a random variable that is varying according to the fading statistics and spatial fading correlation of the channel. If we know the channel at the transmitter and therefore its capacity, we can select a signalling rate equal to or less than the channel capacity and guarantee error-free transmission as long as we use optimal encoding with a large enough block size. If the channel, however, is unknown to the transmitter, there will be a finite probability that an instantaneous channel does not support the bit rate it chooses. This leads to an outage probability and, therefore, despite optimal coding and block sizes, packet errors will occur whenever the bit rate exceeds the actual but unknown channel capacity. In other words, for optimal coding and large enough block sizes, the MIMO Packet Error Rate (PER) equals the probability that the chosen bit rate is larger than the channel capacity (see also [85]).

Consider a 2 × 2 MIMO system as an example. The complementary cdf of the capacity for such a system operating in a flat Rayleigh fading channel is shown in Figure 4-11 for $\rho =$ 15 dB. From this figure, it can be seen that for a bit rate of, e.g., 6 bits/s/Hz, approximately 90% of the channels supports this bit rate, that is 90% of the packet transmissions could ideally be error free. In other words, the outage probability (i.e., lowerbound on the PER) equals 0.1. When applying this principle for a fixed bit rate at different SNR values, it is possible to obtain a PER versus SNR curve. For a fixed rate of 6 bits/s/Hz, this results in the optimal MIMO outage curve of Figure 4-12. Note that this curve implies that the PER will always be greater than zero and depends on the average SNR much like the PER in an uncoded (or sub-optimal coded) Rayleigh fading channel. Furthermore, if we define the *diversity order* as the number of decades the PER improves when the SNR is increased by 10 dB, the diversity order of the optimal MIMO PER for this 2 × 2 case can be shown to be

4 by looking at the slope of the curve. In general, it can be shown that the diversity order of the optimal MIMO PER equals N_tN_r. This indicates that even at the maximum bit rate, a diversity order of N_tN_r is achievable.
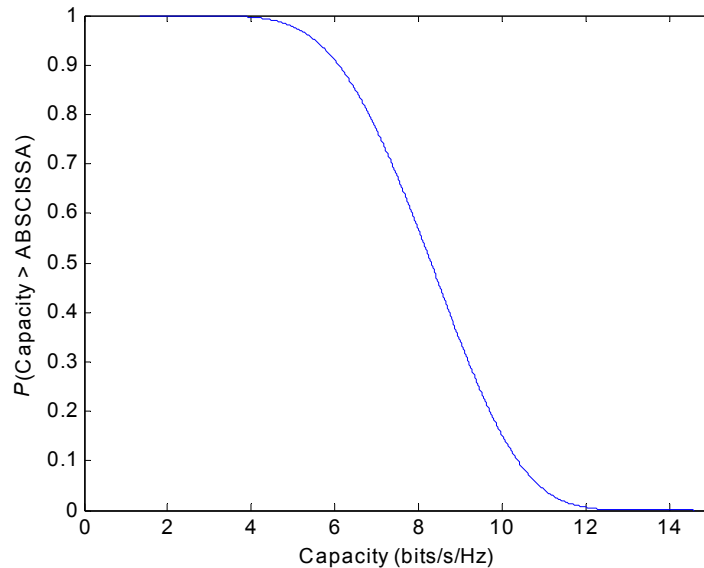
Figure 4-11: Complementary probability distribution function of the capacity of a 2×2 MIMO system operating in flat Rayleigh fading at an average SNR per receive antenna of 15 dB.
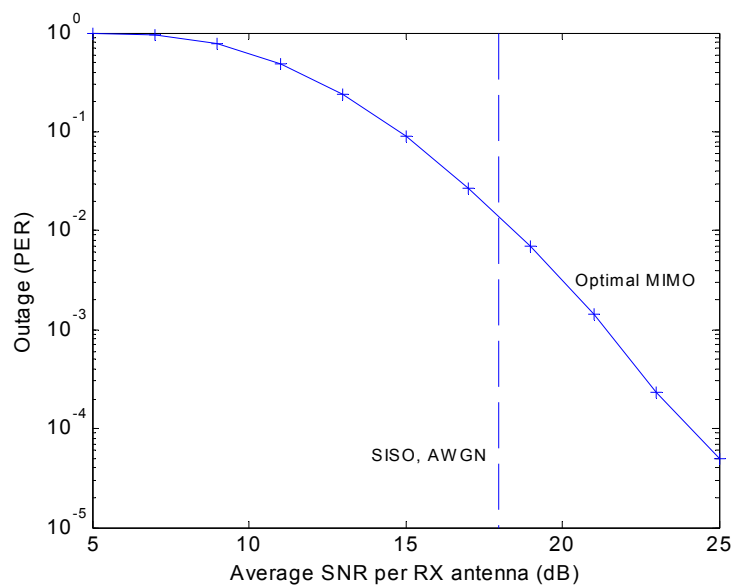
Figure 4-12: Outage probability (or optimal PER) versus the average SNR per RX antenna for a 2×2 MIMO system operating in flat Rayleigh fading and a SISO system experiencing AWGN, for a bit rate of 6 bits/s/Hz.

For comparison, also the outage performance curve for a SISO AWGN channel with a bit rate of 6 bits/s/Hz is depicted in Figure 4-12. It appears to be a step function at $\rho \approx 18$ dB.

This can be explained by the fact that the capacity for a SISO AWGN channel equals $C = \log_2(1 + \rho)$. By setting C equal to 6 bits/s/Hz we can find the corresponding SNR. As a result, in this case, an error is always made if we attempt to transmit at 6 bits/s/Hz on the SISO AWGN channel when $\rho < 18$ dB and conversely we can transmit error-free if the SNR is above 18 dB.

4.5   SNR versus Bit Energy-to-Noise Density Ratio

To make a fair performance comparison of systems with, e.g., different coding schemes or modulation schemes in terms of received energy per bit, usually in literature, the error-rate performance is given as function of the bit energy-to-noise density ratio E_b/N_0. Whereas for wireless communication system simulations, in general, the SNR at the receive antenna is used as input parameter. Let E_s/N_0 denote the SNR per sample at the input of the RX baseband processing. Then, clearly, there is a relation between E_b/N_0 and E_s/N_0. The relation that is used for the MIMO simulations of this chapter is given in this section.

The baseband processing of a MIMO transmission system consists of a number of subsequent blocks that have an influence on the relation between E_b/N_0 and E_s/N_0. These blocks are (see Figure 4-1):

- the encoder with coding rate R ($R < 1$),
- the spatial mapper that maps N_t symbols on N_t transmit antennas,
- and the modulation block that maps m bits onto a 2^m-ary modulation scheme.

Now, assume that the communication between transmitter and receiver is scaled such that the variance of the propagation attenuation is $\sigma_c^2 = 1$. When $R_b = 1/T_b$ denotes the bit rate and T_s the symbol duration, in general, the relation between E_b/N_0 and E_s/N_0 is given by

$$\frac{E_b}{N_0} = \frac{E_s}{N_0}\frac{T_b}{T_s}.$$ (4.31)

Given this equation, the relation between E_b/N_0 and E_s/N_0 can be determined per block described above.

For an encoder with coding rate R, it can be shown that $T_s = RT_b$, thus, substituting this into (4.31) leads to

$$\frac{E_s}{N_0} = R\frac{E_b}{N_0}.$$ (4.32)

Sending N_t bits parallel on N_t transmit antennas reflects on the relation between T_s and T_b as follows: $T_s = N_tT_b$. Therefore the symbol-energy to noise-density ratio per receive antenna equals

$$\frac{E_s}{N_0} = N_t\frac{E_b}{N_0}.$$ (4.33)

The mapping of m bits on an M-ary modulation scheme, with $M = 2^m$, leads to the relation $T_s = mT_b$, so

$$\frac{E_s}{N_0} = m \frac{E_b}{N_0}.$$ (4.34)

When all the above blocks are combined in a serial way, this leads to the following relation:

$$\frac{E_s}{N_0} = RmN_t \frac{E_b}{N_0} = \eta_{\text{eff}} \frac{E_b}{N_0},$$ (4.35)

where η_{eff} denotes the spectral efficiency in bits/s/Hz, i.e., the ratio between the bit rate and the system bandwidth $B = 1/T_s$:

$$\frac{E_s}{N_0} = \frac{T_s}{T_b} \frac{E_b}{N_0} = \frac{R_b}{B} \frac{E_b}{N_0} = \eta_{\text{eff}} \frac{E_b}{N_0}.$$ (4.36)

In the next sections, first some SDM algorithms are described before simulation results are shown for which above relations are applicable.

4.6   Zero Forcing (ZF)

4.6.1   Algorithm Description

The principle of Zero Forcing is already introduced in Chapter 2. Here, the algorithm is described in more detail, and its performance and complexity are evaluated in the next subsections. Zero Forcing is a linear MIMO technique. In Section 2.2, it is shown already that the processing takes place at the receiver where, under the assumption that the channel transfer matrix $\mathbf{H}$ is invertible, $\mathbf{H}$ is inverted and the transmitted MIMO vector $\mathbf{s}$ is estimated by

$$\mathbf{s}_{\text{est}} = \mathbf{H}^{-1}\mathbf{x}.$$ (4.37)

This principle is based on a conventional adaptive antenna array (AAA) technique, namely, linear combinatorial nulling ([144]). In this technique, each substream in turn is considered to be the desired signal, and the remaining data streams are considered as "interferers". Nulling of the interferers is performed by linearly weighting the received signals such that all interfering terms are cancelled. For Zero Forcing, nulling of the "interferers" can be performed by choosing $1 \times N_r$ dimensional weight vectors $\mathbf{w}^i$ (with $i = 1, 2, \ldots, N_t$), referred to as nulling vectors ([144]), such that

$$\mathbf{w}^i \mathbf{h}_p = \begin{cases} 0, & p \neq i, \\ 1, & p = i, \end{cases}$$ (4.38)

where $\mathbf{h}_p$ denotes the p-th column of the channel matrix $\mathbf{H}$. Let $\mathbf{w}^i$ be the i-th row of a matrix $\mathbf{W}$, then it follows that

$$\mathbf{WH} = \mathbf{I}_{N_t},\tag{4.39}$$

where $\mathbf{W}$ is a matrix that represents the linear processing in the receiver. So, by forcing the "interferers" to zero, each desired element of $\mathbf{s}$ can be estimated.

If $\mathbf{H}$ is not square, $\mathbf{W}$ equals the *pseudo-inverse* of $\mathbf{H}$ (denoted by $\mathbf{H}^\dagger$):

$$\mathbf{W} = \mathbf{H}^\dagger = \left(\mathbf{H}^H \mathbf{H}\right)^{-1}\mathbf{H}^H.\tag{4.40}$$

If the elements of $\mathbf{H}$ are assumed to be i.i.d., the pseudo-inverse exists when N_t is less than or equal to N_r. For N_t larger than N_r, $\mathbf{H}^H\mathbf{H}$ is singular and its inverse does not exist ([113]). When the pseudo-inverse exists, the estimates of $\mathbf{s}$ (given by $\mathbf{s}_{est}$) can be found by

$$\begin{aligned}\mathbf{s}_{est} &= \mathbf{Wx} = \mathbf{H}^\dagger\mathbf{x} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{x}\\&= \mathbf{s} + \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{n}\end{aligned}.\tag{4.41}$$

Denote the i-th component of $\mathbf{s}_{est}$ by $(\mathbf{s}_{est})_i$, then, as a final step, $(\mathbf{s}_{est})_i$ must be sliced to the nearest constellation point. In this way, all N_t elements of $\mathbf{s}$ can be decoded at the receiver.

A big disadvantage of Zero Forcing is that it suffers from noise enhancement, especially for channels with a high condition number $\kappa(\mathbf{H}^H\mathbf{H})$. This can be readily observed from (4.41).

4.6.2   Performance Analysis

The ZF algorithm described in the previous subsection can be simulated and its BER and PER performance can be obtained (see Section 4.11). To verify the simulations it is useful to deduce a theoretical representation of the BER performance of the ZF algorithm. We start by recalling the relation between $\mathbf{s}$ and $\mathbf{x}$ from (4.41):

$$\begin{aligned}\mathbf{s}_{est} &= \mathbf{H}^\dagger\mathbf{x} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{x}\\&= \mathbf{s} + \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{n}\end{aligned}.\tag{4.42}$$

This leads to an estimation error

$$\boldsymbol{\varepsilon} = \mathbf{s}_{est} - \mathbf{s} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{n}.\tag{4.43}$$

The covariance matrix of the estimation error equals

$$\mathbf{Q} = E[\boldsymbol{\varepsilon}\cdot\boldsymbol{\varepsilon}^H] = \sigma_n^2\left(\mathbf{H}^H\mathbf{H}\right)^{-1}.\tag{4.44}$$

Since $\mathbf{n}$ is multivariate complex Gaussian distributed (see Section 4.3) and $\boldsymbol{\varepsilon}$ is a linear transformation of $\mathbf{n}$, $\boldsymbol{\varepsilon}$ will also follow a multivariate complex normal distribution. For a given $\mathbf{H}$, this results in the conditional probability density function

$$p(\mathbf{s}_{est}|\mathbf{s},\mathbf{H}) = \det(\pi\mathbf{Q})^{-1}\exp\left(-(\mathbf{s}_{est}-\mathbf{s})^H\mathbf{Q}^{-1}(\mathbf{s}_{est}-\mathbf{s})\right). \tag{4.45}$$

For the ZF solution, it is assumed that the elements of the estimated vector $\mathbf{s}_{est}$ are independent and, thus, $\mathbf{Q}$ is assumed diagonal. As a result, it can be said that, without loss of generality, the conditional probability density function of the p-th element of $\mathbf{s}_{est}$ equals

$$p\left((\mathbf{s}_{est})_p|s_p,\mathbf{H}\right) = \frac{\alpha}{\pi\sigma_n^2}\exp\left(\frac{-\alpha\left|(\mathbf{s}_{est})_p-s_p\right|^2}{\sigma_n^2}\right), \tag{4.46}$$

where $\alpha = 1/\left(\mathbf{H}^H\mathbf{H}\right)_{pp}^{-1}$ and $\left(\mathbf{H}^H\mathbf{H}\right)_{pp}^{-1}$ corresponds to the element (p,p) of matrix $\left(\mathbf{H}^H\mathbf{H}\right)^{-1}$.

A commonly used method to find an upperbound on the performance is by means of the Pairwise Error Probability (PEP) analysis ([20, 90, 116]). Assume two different space vectors of size $N_t \times 1$, $\mathbf{s}_i$ and $\mathbf{s}_k$, of which the elements are taken from an M-ary constellation. Denote the p-th element of the first and second vector by $(\mathbf{s}_i)_p$ and $(\mathbf{s}_k)_p$, respectively, with $i, k \in \{1, \dots, M\}$. Then, based on above conditional pdf and exploiting the equivalency with BPSK as shown in Figure 4-13 (cf. [90]), the probability that the receiver decides erroneously in favour of $(\mathbf{s}_k)_p$ while $(\mathbf{s}_i)_p$ ($i \neq k$) has been sent is given by

$$\Pr\left((\mathbf{s}_i)_p \to (\mathbf{s}_k)_p|\mathbf{H}\right) = Q\left(\sqrt{\frac{d_{ik}^2}{2\sigma_{BSPK}^2}}\right) = Q\left(\sqrt{\frac{\alpha\left|(\mathbf{s}_i)_p-(\mathbf{s}_k)_p\right|^2}{2\sigma_n^2}}\right). \tag{4.47}$$

In this equation, d_{ik} is the Euclidean distance between $(\mathbf{s}_i)_p$ and $(\mathbf{s}_k)_p$, σ_{BPSK} is the standard deviation of the equivalent complex BPSK noise (see Figure 4-13), and Q is the function that defines the area under the tail of the Gaussian pdf and equals

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp\left(-t^2/2\right)dt \text{ with } x \geq 0. \tag{4.48}$$
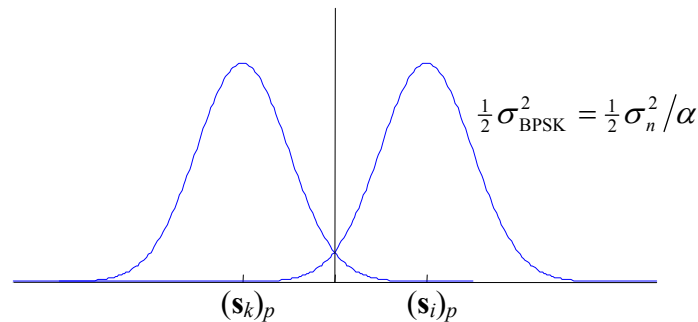
Figure 4-13: Pdfs of received symbols. Note the equivalence with BPSK.

Using the Chernoff bound $Q(x) \leq \exp(-x^2/2)$ yields

$$\Pr\left(\left(\mathbf{s}_i\right)_p \to \left(\mathbf{s}_k\right)_p \middle| \mathbf{H}\right) \leq \exp\left(-\frac{\alpha\left|\left(\mathbf{s}_i\right)_p - \left(\mathbf{s}_k\right)_p\right|^2}{4\sigma_n^2}\right). \tag{4.49}$$

In [142] it is shown that when $\mathbf{H}$ is Rayleigh distributed (see Subsection 3.5.2) α is chi-square distributed with $2(N_r - N_t + 1)$ degrees of freedom. As a result, the pdf of α can be shown to be

$$p(\alpha) = \frac{1}{(N_r - N_t)!\sigma^{2(N_r - N_t + 1)}} \alpha^{N_r - N_t} \exp\left(-\alpha/\sigma^2\right) \text{ for } \alpha \geq 0. \tag{4.50}$$

Note that usually the chi-square distribution with n degrees of freedom is defined as the distribution of the sum of n squared i.i.d. zero mean *real* Gaussian variables with a variance σ^2. Here, however, a slightly modified form of this definition is used where the underlying $n/2$ *complex* Gaussian variables are said to have a variance of σ^2 so that the n real Gaussian variables from which the complex ones are composed have a variance of $\sigma^2/2$.

The average PEP over all channel realisations is now given by

$$\begin{aligned}
\Pr\left(\left(\mathbf{s}_i\right)_p \to \left(\mathbf{s}_k\right)_p\right) &\leq \int_0^\infty \exp\left(-\frac{\alpha d_{ik}^2}{4\sigma_n^2}\right) p(\alpha) d\alpha \\
&= \int_0^\infty \exp\left(-\frac{\alpha d_{ik}^2}{4\sigma_n^2}\right) \exp\left(-\frac{\alpha}{\sigma^2}\right) \frac{\alpha^{N_r - N_t}}{(N_r - N_t)!\sigma^{2(N_r - N_t + 1)}} d\alpha \\
&= \int_0^\infty \exp\left(-\frac{\alpha'}{\sigma_{tot}^2}\right) \frac{\alpha'^{N_r - N_t}}{(N_r - N_t)!\sigma_{tot}^{2(N_r - N_t + 1)}} \frac{\sigma_{tot}^{2(N_r - N_t + 1)}}{\sigma^{2(N_r - N_t + 1)}} d\alpha' \\
&= \frac{\sigma_{tot}^{2(N_r - N_t + 1)}}{\sigma^{2(N_r - N_t + 1)}},
\end{aligned} \tag{4.51}$$

where in the third equality α is redefined as a chi-square distribution α' with $2(N_r - N_t + 1)$ degrees of freedom. The variance of the $2(N_r - N_t + 1)$ underlying real Gaussian variables of this chi-square distribution equals

$$\sigma_{tot}^2 = \left(\frac{\left|\left(\mathbf{s}_i\right)_p - \left(\mathbf{s}_k\right)_p\right|^2}{4\sigma_n^2} + \frac{1}{\sigma^2}\right)^{-1}. \tag{4.52}$$

From the signal model of Section 4.3 it can be easily deduced that $\sigma^2 = 1$ and, thus, this finally results in a PEP equal to

$$\Pr\!\left((\mathbf{s}_i)_p \to (\mathbf{s}_k)_p\right) \le \left(1 + \frac{\left|(\mathbf{s}_i)_p - (\mathbf{s}_k)_p\right|^2}{4\sigma_n^2}\right)^{-(N_r - N_t + 1)}. \tag{4.53}$$

When we normalise $(\mathbf{s}_i)_p$ and $(\mathbf{s}_k)_p$ such that $(\mathbf{s}_i)_p = \sigma_s(\mathbf{s'}_i)_p$ and $(\mathbf{s}_k)_p = \sigma_s(\mathbf{s'}_k)_p$, respectively, we may write

$$\begin{aligned}
\Pr\!\left((\mathbf{s}_i)_p \to (\mathbf{s}_k)_p\right) &\le \left(1 + \frac{\sigma_s^2\left|(\mathbf{s'}_i)_p - (\mathbf{s'}_k)_p\right|^2}{4\sigma_n^2}\right)^{-(N_r - N_t + 1)} \\
&= \left(1 + \frac{\rho\left|(\mathbf{s'}_i)_p - (\mathbf{s'}_k)_p\right|^2}{4N_t}\right)^{-(N_r - N_t + 1)}.
\end{aligned} \tag{4.54}$$

Note that the Symbol Error Rate (SER) performance is a (weighted) sum of the PEPs. This, e.g., follows from the *union bound* which defines that the symbol error probability when $(\mathbf{s}_i)_p$ is sent equals ([92])

$$\Pr_s\!\left((\mathbf{s}_i)_p\right) \le \sum_{\substack{k=1 \\ k \ne i}}^{M} \Pr\!\left((\mathbf{s}_i)_p \to (\mathbf{s}_k)_p\right). \tag{4.55}$$

Or, when averaging over the M possible TX symbols on the p-th antenna, the SER of this TX stream reads

$$\Pr_s \le \frac{1}{M} \sum_{i=1}^{M} \sum_{\substack{k=1 \\ k \ne i}}^{M} \Pr\!\left((\mathbf{s}_i)_p \to (\mathbf{s}_k)_p\right). \tag{4.56}$$

On average, the total SER can be shown to be equivalent to this result. Moreover, from this bound, an approximation on the bit error probability can be obtained via

$$\Pr_b \approx \frac{1}{m}\Pr_s, \tag{4.57}$$

where $m = \log_2 M$ denotes the amount of bits per constellation point. By filling in (4.54), it can be observed that the BER as function of the SNR per RX antenna ρ tends to fall off in an exponential way with the exponent $N_r - N_t + 1$. In other words, the *diversity order* is equal to $N_r - N_t + 1$, which means that when the SNR is increased by 10 dB the BER drops by $N_r - N_t + 1$ decades.

Based on above results, it is easily verified that the performance of an $N_t \times N_r$ system with ZF processing equals that of a system with Maximal Ratio Combining (MRC), one TX antenna (sending with the same TX power as a TX antenna of the ZF MIMO system) and $N_r - N_t + 1$ RX antennas, as also is observed in [142]. Closed-form solutions of the MRC performance can be found in [90]. For instance, for BPSK, when s_i and s_k are two different

TX symbols for the single TX antenna, the closed-form solution for the PEP of the MRC system averaged over all channel realisations is given by

$$\Pr(s_i \rightarrow s_k) = \left(\tfrac{1}{2}(1-\mu)\right)^{N_r - N_t + 1} \sum_{l=0}^{N_r - N_t} \frac{(N_r - N_t + l)!}{(N_r - N_t)!l!} \left(\tfrac{1}{2}(1+\mu)\right)^l, \tag{4.58}$$

where

$$\mu = \sqrt{\frac{\rho/N_t}{1 + \rho/N_t}}. \tag{4.59}$$

Discarding the actual transmit power, we can say that BPSK only transmits symbols from the set $\{-1, 1\}$ with an equal probability. Therefore, the total bit error probability can be found to be $\Pr_b = \Pr(s_1 \rightarrow s_2)/2 + \Pr(s_2 \rightarrow s_1)/2 = \Pr(s_1 \rightarrow s_2)$. For a large SNR per RX antenna ρ, the term $(1 + \mu)/2 \approx 1$ and the term $(1 - \mu)/2 \approx N_t/(4\rho)$. Furthermore,

$$\sum_{l=0}^{N_r - N_t} \frac{(N_r - N_t + l)!}{(N_r - N_t)!l!} = \frac{(2N_r - 2N_t + 1)!}{(N_r - N_t)!(N_r - N_t + 1)!}. \tag{4.60}$$

Therefore, when γ_c is sufficiently large (greater than 10 dB), the probability of error in (4.58) can be approximated by

$$\Pr_b \approx \frac{(2N_r - 2N_t + 1)!}{(N_r - N_t)!(N_r - N_t + 1)!} \left(\frac{N_t}{4\rho}\right)^{N_r - N_t + 1}. \tag{4.61}$$

From this equation, we again observe that the error rate decreases with the $(N_r - N_t + 1)$-th power of the SNR.

The theoretical result, given in (4.58), will be used in Section 4.11 to verify the simulated BER performance of the ZF algorithm.

4.6.3   Complexity

As described in Subsection 4.6.1, the Zero Forcing technique is based on calculation of the pseudo-inverse of the channel transfer matrix **H**. Because it is assumed that the MIMO system is operating in a quasi-static environment, i.e., **H** is constant during a packet transmission, the pseudo-inverse of **H** needs to be calculated only once per packet. The pseudo-inverse can be calculated after the channel training in the preamble processing. During the payload processing, the pseudo-inverse is used for the estimation of every transmitted MIMO vector **s** of the corresponding packet.

A detailed analysis of the complexity of the ZF algorithm is presented in Appendix C.2. From this analysis, it follows that the complexity of the preamble processing equals $4N_t^3 + N_t^2(8N_r - 2) - 2N_tN_r$ real additions (R_ADDs) and $4N_t^3 + 8N_t^2N_r$ real multiplications (R_MULs).

The complexity of ZF in the payload processing can be found in the same appendix and is per transmitted vector **s** equal to $2N_tN_r + 2N_t(N_r - 1) + N_t \cdot \log_2(M)$ R_ADDs and $4N_tN_r$ R_MULs. When N_s vectors are transmitted within a packet, these numbers must be multiplied by N_s to obtain the payload complexity of ZF per packet.

4.6.4 Soft-decision Output ZF

When coding is applied, it is commonly known that the performance improves when the decoder also has knowledge about the reliability of its input bits next to their sliced values (see [40]). The former values are generally referred to as soft-decision values, whereas the latter are called hard-decision values. For ZF, the generation of soft-decision output values is based on the transformation of the received vector **x** from the "**x**-space" into the "**s**-space" ([110]) so that the search for the possible transmitted vector can be done directly in the "**s**-space". This transformation is already performed in Subsection 4.6.2 and we start by recalling the conditional probability density function of the estimated **s** vector for a given channel **H** from (4.45):

$$p(\mathbf{s}_{\text{est}}|\mathbf{s}, \mathbf{H}) = \det(\pi\mathbf{Q})^{-1} \exp\left(-(\mathbf{s}_{\text{est}} - \mathbf{s})^H \mathbf{Q}^{-1}(\mathbf{s}_{\text{est}} - \mathbf{s})\right). \tag{4.62}$$

Note that, generally speaking, the covariance matrix **Q** cannot be assumed diagonal, i.e., the estimation errors for different antennas are correlated.

The generation of soft-decision output values for ZF is based on the maximum a posteriori probability (MAP) principle. Suppose that, at a given time instant, $N_b = N_t \cdot m$ bits are sent by a MIMO vector/symbol, where $m = \log_2 M$ denotes the amount of bits used per M-ary constellation point. Then, if b_k is the k-th bit, k = 1, ..., N_b, of the transmitted vector $\mathbf{s}_i$ to estimate, the binary symbol-by-symbol MAP decoder at the receiver decides, conditioned on the received vector **x**, that this bit was "1" if

$$\Pr(b_k = 1|\mathbf{x}) > \Pr(b_k = 0|\mathbf{x}), \tag{4.63}$$

and "0" otherwise. This rule can be written in a more compact form by means of the definition of the Log-Likelihood Ratio (LLR) as follows

$$L(b_k) = \ln \frac{\Pr(b_k = 1|\mathbf{x})}{\Pr(b_k = 0|\mathbf{x})}, \tag{4.64}$$

or, equivalently, when transferring the problem to the "**s**-space"

$$L(b_k) = \ln \frac{\Pr(b_k = 1|\mathbf{s}_{\text{est}})}{\Pr(b_k = 0|\mathbf{s}_{\text{est}})} = \ln \frac{\sum_{\mathbf{s}_i|b_k=1} \Pr(\mathbf{s}_i|\mathbf{s}_{\text{est}})}{\sum_{\mathbf{s}_i|b_k=0} \Pr(\mathbf{s}_i|\mathbf{s}_{\text{est}})}. \tag{4.65}$$

where the ensemble $\mathbf{s}_i$ ($1 \le i \le I$) denotes all possible transmitted MIMO vectors, thus,

$$I = M^{N_t} . \tag{4.66}$$

Now, above decision rule can easily be implemented as $b_k = (\text{sgn}(L(b_k))+1)/2$. When we apply Bayes' rule, $\Pr(A|B) = \Pr(B|A) \cdot \Pr(A)/\Pr(B)$, the LLR becomes

$$L(b_k) = \ln \frac{\sum\limits_{\mathbf{s}_i | b_k = 1} p(\mathbf{s}_{\text{est}} | \mathbf{s}_i) \Pr(\mathbf{s}_i)}{\sum\limits_{\mathbf{s}_i | b_k = 0} p(\mathbf{s}_{\text{est}} | \mathbf{s}_i) \Pr(\mathbf{s}_i)} . \tag{4.67}$$

Because the vectors $\mathbf{s}_i$ are equally likely to be transmitted, $\Pr(\mathbf{s}_i)$ is equal for all vectors $\mathbf{s}_i$. Using the distribution of $p(\mathbf{s}_{\text{est}} | \mathbf{s}_i)$ that is provided in (4.45), the LLR can be shown to be

$$L(b_k) = \ln \frac{\sum\limits_{\mathbf{s}_i | b_k = 1} \exp\left(-(\mathbf{s}_{\text{est}} - \mathbf{s}_i)^H \mathbf{Q}^{-1}(\mathbf{s}_{\text{est}} - \mathbf{s}_i)\right)}{\sum\limits_{\mathbf{s}_i | b_k = 0} \exp\left(-(\mathbf{s}_{\text{est}} - \mathbf{s}_i)^H \mathbf{Q}^{-1}(\mathbf{s}_{\text{est}} - \mathbf{s}_i)\right)} . \tag{4.68}$$

Note that, up to this point, we did not perform any form of approximation, thus, as far as performance is concerned, the achievable result is identical to the exhaustive search method MLD (see Subsection 4.10.4), since the transformation of $\mathbf{x}$ to the "$\mathbf{s}$-space" is free of information loss.

Applying the max-log approximation to the above equation yields

$$L(b_k) \approx \ln \frac{\exp\left(\max\limits_{\mathbf{s}_i | b_k = 1}\left(-(\mathbf{s}_{\text{est}} - \mathbf{s}_i)^H \mathbf{Q}^{-1}(\mathbf{s}_{\text{est}} - \mathbf{s}_i)\right)\right)}{\exp\left(\max\limits_{\mathbf{s}_i | b_k = 0}\left(-(\mathbf{s}_{\text{est}} - \mathbf{s}_i)^H \mathbf{Q}^{-1}(\mathbf{s}_{\text{est}} - \mathbf{s}_i)\right)\right)}$$
$$= \min\limits_{\mathbf{s}_i | b_k = 0}(\mathbf{s}_{\text{est}} - \mathbf{s}_i)^H \mathbf{Q}^{-1}(\mathbf{s}_{\text{est}} - \mathbf{s}_i) - \min\limits_{\mathbf{s}_i | b_k = 1}(\mathbf{s}_{\text{est}} - \mathbf{s}_i)^H \mathbf{Q}^{-1}(\mathbf{s}_{\text{est}} - \mathbf{s}_i) . \tag{4.69}$$

According to [110], the performance degradation of this approximation is negligible. Unfortunately, for above LLR definitions an exhaustive search is required over all I transmitted vectors $\mathbf{s}_i$, and since I grows exponentially with N_t, the complexity also grows exponentially with N_t. In order to obtain an algorithm with a complexity that grows linearly with the number of TX antennas, the following simplification is made: the elements of the estimation error vector $\boldsymbol{\varepsilon}$ are assumed to be uncorrelated and, thus, $\mathbf{Q}$ is assumed diagonal. Then, the LLR becomes equal to

$$L(b_k) \approx \ln \frac{\sum\limits_{\mathbf{s}_i | b_k = 1} \exp\left(-\frac{|(\mathbf{s}_{\text{est}})_1 - (\mathbf{s}_i)_1|^2}{q_{11}}\right) \cdot \ldots \cdot \exp\left(-\frac{|(\mathbf{s}_{\text{est}})_{N_t} - (\mathbf{s}_i)_{N_t}|^2}{q_{N_t N_t}}\right)}{\sum\limits_{\mathbf{s}_i | b_k = 0} \exp\left(-\frac{|(\mathbf{s}_{\text{est}})_1 - (\mathbf{s}_i)_1|^2}{q_{11}}\right) \cdot \ldots \cdot \exp\left(-\frac{|(\mathbf{s}_{\text{est}})_{N_t} - (\mathbf{s}_i)_{N_t}|^2}{q_{N_t N_t}}\right)} . \tag{4.70}$$

where q_{ab} denotes element (a,b) of $\mathbf{Q}$, and $(\mathbf{s}_{est})_p$ and $(\mathbf{s}_i)_p$ correspond to the p-th element of $\mathbf{s}_{est}$ and $\mathbf{s}_i$, respectively. Now, when applying the max-log approximation on this result and using the fact that one of the terms in the summation is the maximum and equals $\exp(\max(\ldots))\cdot\ldots\cdot\exp(\max(\ldots))$, the soft value of a certain bit b_k, which is assumed to be mapped on the k'-th element of $\mathbf{s}$, can be shown to be

$$L(b_k) \approx \ln \frac{\exp\left(\max\limits_{(\mathbf{s}_i)_1 \mid b_k=1} -d_1\right) \cdot \ldots \cdot \exp\left(\max\limits_{(\mathbf{s}_i)_{k'} \mid b_k=1} -d_{k'}\right) \cdot \ldots \cdot \exp\left(\max\limits_{(\mathbf{s}_i)_{N_t} \mid b_k=1} -d_{N_t}\right)}{\exp\left(\max\limits_{(\mathbf{s}_i)_1 \mid b_k=0} -d_1\right) \cdot \ldots \cdot \exp\left(\max\limits_{(\mathbf{s}_i)_{k'} \mid b_k=0} -d_{k'}\right) \cdot \ldots \cdot \exp\left(\max\limits_{(\mathbf{s}_i)_{N_t} \mid b_k=0} -d_{N_t}\right)}$$

$$= \ln \frac{\exp\left(-\min\limits_{(\mathbf{s}_i)_1} d_1\right) \cdot \ldots \cdot \exp\left(-\min\limits_{(\mathbf{s}_i)_{k'} \mid b_k=1} d_{k'}\right) \cdot \ldots \cdot \exp\left(-\min\limits_{(\mathbf{s}_i)_{N_t}} d_{N_t}\right)}{\exp\left(-\min\limits_{(\mathbf{s}_i)_1} d_1\right) \cdot \ldots \cdot \exp\left(-\min\limits_{(\mathbf{s}_i)_{k'} \mid b_k=0} d_{k'}\right) \cdot \ldots \cdot \exp\left(-\min\limits_{(\mathbf{s}_i)_{N_t}} d_{N_t}\right)} \tag{4.71}$$

$$= \frac{1}{q_{k'k'}}\left(\min\limits_{(\mathbf{s}_i)_{k'} \mid b_k=0} \left|(\mathbf{s}_{est})_{k'} - (\mathbf{s}_i)_{k'}\right|^2 - \min\limits_{(\mathbf{s}_i)_{k'} \mid b_k=1} \left|(\mathbf{s}_{est})_{k'} - (\mathbf{s}_i)_{k'}\right|^2\right),$$

where

$$d_p = \frac{\left|(\mathbf{s}_{est})_p - (\mathbf{s}_i)_p\right|^2}{q_{pp}}, \tag{4.72}$$

and

$$k' = \left\lceil \frac{k}{\log_2 M} \right\rceil, \tag{4.73}$$

with the operator $\lceil a \rceil$ giving the lowest integer value that is larger than or equal to a. From (4.44) and (4.71), it can be deduced that for (very) high SNRs, the LLR could become (very) large. A practical implementation, however, generally uses a fixed-point approach in which the signals are quantised. Suppose that a linear N_q-bit quantiser is used. To make sure that the LLRs are within the quantisation range, we propose to scale (4.71).

In order to find the proper scaling factor, we first need to determine the expectation of the covariance matrix of the estimation error for a large number of channel realisations. If the communication system is assumed to operate in an environment that can be modelled by the flat Rayleigh fading model of Subsection 3.5.2, the expectation can be shown to be

$$E[\mathbf{Q}] = E\left[(\boldsymbol{\varepsilon} - E[\boldsymbol{\varepsilon}])(\boldsymbol{\varepsilon} - E[\boldsymbol{\varepsilon}])^H\right]$$
$$= \sigma_n^2 E\left[(\mathbf{H}^H \mathbf{H})^{-1}\right] = \sigma_n^2 (N_r \mathbf{I}_{N_t})^{-1} = \frac{\sigma_n^2}{N_r} \mathbf{I}_{N_t}. \tag{4.74}$$

To analyse the soft value of the k-th bit, some assumptions need to be made. Firstly, it is assumed that $b_k = 1$ has been transmitted. Secondly, it is assumed that the correct $(\mathbf{s}_i)_{k'}|b_k = 1$ is found, such that

$$E\left[\min_{(s_i)_{k'}|b_k=1}\left|(\mathbf{s}_{\text{est}})_{k'} - (\mathbf{s}_i)_{k'}\right|^2\right] = \frac{\sigma_n^2}{N_r}.$$ (4.75)

Finally, it is assumed that $(\mathbf{s}_i)_{k'}|b_k = 0$ is chosen completely wrong, thus, the distance between $(\mathbf{s}_{\text{est}})_{k'}$ and $(\mathbf{s}_i)_{k'}|b_k = 0$ is assumed to be the maximum Euclidean distance between two arbitrary constellation points. For BPSK and QPSK, this distance equals 2. Combining these results and using (4.71), the expectation of the soft-value of the k-th bit equals

$$E[L(b_k)] \approx E\left[\frac{1}{q_{k'k'}}\left(\min_{(\mathbf{s}_i)_{k'}|b_k=0}\left|(\mathbf{s}_{\text{est}})_{k'} - (\mathbf{s}_i)_{k'}\right|^2 - \min_{(\mathbf{s}_i)_{k'}|b_k=1}\left|(\mathbf{s}_{\text{est}})_{k'} - (\mathbf{s}_i)_{k'}\right|^2\right)\right]$$

$$= \frac{4 - \dfrac{\sigma_n^2}{N_r}}{\dfrac{\sigma_n^2}{N_r}} = 4\frac{N_r}{\sigma_n^2} - 1.$$ (4.76)

Note that, if it was assumed that $b_k = 0$ was transmitted, the following result would have been obtained:

$$E[L(b_k)] = 1 - 4\frac{N_r}{\sigma_n^2}.$$ (4.77)

With N_q quantisation bits, the quantisation range is given by

$$[-2^{N_q-1}, 2^{N_q-1} - 1].$$ (4.78)

It is reasonable to assume that when we set the "worst case" average of (4.76) to 80% of the upperbound of the quantisation range, it provides a sufficient margin to avoid clipping as much as possible. Then, with $4N_r \gg \sigma_n^2$, the quantisation scaling factor C_q can be found as follows:

$$C_q E[L(b_k)] \approx 4C_q \frac{N_r}{\sigma_n^2} \equiv \tfrac{4}{5}\left(2^{N_q-1}\right),$$

$$C_q = \tfrac{1}{5}\left(2^{N_q-1}\right)\frac{\sigma_n^2}{N_r}.$$ (4.79)

Note that, when applying this quantisation constant, the resulting LLRs are independent of the SNR, so no estimation of the SNR is required. For the SNR region of interest, applying this quantisation constant hardly influences the performance as is shown for SOMLD (see Subsection 4.10.4) in Figure 5-16.

4.7   Minimum Mean Squared Error (MMSE) Solution

4.7.1   Algorithm Description

Another approach in estimation theory to the problem of estimating a random vector $\mathbf{s}$ on the basis of observations $\mathbf{x}$ is to choose a function $\mathbf{f}(\mathbf{x})$ that minimises the Mean Square Error (MSE) ([92]):

$$\varepsilon^2 = E\left[(\mathbf{s} - \mathbf{s}_{\text{est}})^H (\mathbf{s} - \mathbf{s}_{\text{est}})\right] = \left[(\mathbf{s} - \mathbf{f}(\mathbf{x}))^H (\mathbf{s} - \mathbf{f}(\mathbf{x}))\right] . \tag{4.80}$$

An exact function $\mathbf{f}(\mathbf{x})$ is usually hard to obtain, however, if we restrict this function to be a linear function of the observations, an exact solution can be achieved. Using linear processing, the estimates of $\mathbf{s}$ can be found by

$$\mathbf{s}_{\text{est}} = \mathbf{W}\mathbf{x} . \tag{4.81}$$

Now, to obtain the linear Minimum Mean Squared Error (MMSE) solution, $\mathbf{W}$ must be chosen such that the Mean Square Error ε^2 is minimised:

$$\varepsilon^2 = E\left[(\mathbf{s} - \mathbf{s}_{\text{est}})^H (\mathbf{s} - \mathbf{s}_{\text{est}})\right] = E\left[(\mathbf{s} - \mathbf{W}\mathbf{x})^H (\mathbf{s} - \mathbf{W}\mathbf{x})\right] . \tag{4.82}$$

Since $\mathbf{a}^H \mathbf{a} = \text{tr}(\mathbf{a}\mathbf{a}^H)$, where $\text{tr}(\cdot)$ stands for the *trace* of a matrix, the MSE can be written as

$$\begin{aligned} E\left[(\mathbf{s} - \mathbf{W}\mathbf{x})^H (\mathbf{s} - \mathbf{W}\mathbf{x})\right] &= \text{tr}\left(E\left[(\mathbf{s} - \mathbf{W}\mathbf{x})(\mathbf{s} - \mathbf{W}\mathbf{x})^H\right]\right) \\ &= \text{tr}\left(\mathbf{Q}_s - \mathbf{Q}_{sx}\mathbf{W}^H - \mathbf{W}\mathbf{Q}_{xs} + \mathbf{W}\mathbf{Q}_x\mathbf{W}^H\right), \end{aligned} \tag{4.83}$$

where, $\mathbf{Q}_s = E[\mathbf{s}\mathbf{s}^H]$, $\mathbf{Q}_{sx} = E[\mathbf{s}\mathbf{x}^H]$, $\mathbf{Q}_{xs} = E[\mathbf{x}\mathbf{s}^H]$ and $\mathbf{Q}_x = E[\mathbf{x}\mathbf{x}^H]$. Based on the same observations as for the spatial correlation matrices in Section 3.6, we may write $\mathbf{Q}_x$ (because $\mathbf{Q}_x$ is Hermitian and nonnegative definite) as follows:

$$\mathbf{Q}_x = \mathbf{U}^H \boldsymbol{\Lambda}_x \mathbf{U} = \mathbf{U}^H \boldsymbol{\Lambda}_x^{1/2} \boldsymbol{\Lambda}_x^{1/2} \mathbf{U} = \left(\boldsymbol{\Lambda}_x^{1/2} \mathbf{U}\right)^H \boldsymbol{\Lambda}_x^{1/2} \mathbf{U} = \mathbf{A}^H \mathbf{A} , \tag{4.84}$$

where $\boldsymbol{\Lambda}_x$ is a diagonal matrix with the eigenvalues of $\mathbf{Q}_x$ on its diagonal and $\mathbf{A}$ is said to be the "square-root" of $\mathbf{Q}_x$. Now, we can rewrite the MMSE problem to a form from which a solution for $\mathbf{W}$ can be obtained that minimises the Mean Square Error:

$$\begin{aligned} \varepsilon^2 &= \text{tr}\left(\mathbf{Q}_s - \mathbf{Q}_{sx}\mathbf{W}^H - \mathbf{W}\mathbf{Q}_{xs} + \mathbf{W}\mathbf{Q}_x\mathbf{W}^H\right) \\ &= \text{tr}\left(\mathbf{Q}_s - \mathbf{Q}_{sx}\mathbf{A}^{-1}\mathbf{A}\mathbf{W}^H - \mathbf{W}\mathbf{A}^H\left(\mathbf{A}^H\right)^{-1}\mathbf{Q}_{xs} + \mathbf{W}\mathbf{A}^H\mathbf{A}\mathbf{W}^H\right) \\ &= \text{tr}(\mathbf{Q}_s - \mathbf{Q}_{sx}\mathbf{A}^{-1}\left(\mathbf{A}^H\right)^{-1}\mathbf{Q}_{xs} + \mathbf{Q}_{sx}\mathbf{A}^{-1}\left(\mathbf{A}^H\right)^{-1}\mathbf{Q}_{xs} \\ &\quad - \mathbf{Q}_{sx}\mathbf{A}^{-1}\mathbf{A}\mathbf{W}^H - \mathbf{W}\mathbf{A}^H\left(\mathbf{A}^H\right)^{-1}\mathbf{Q}_{xs} + \mathbf{W}\mathbf{A}^H\mathbf{A}\mathbf{W}^H) \\ &= \text{tr}\left(\mathbf{Q}_s - \mathbf{Q}_{sx}\mathbf{A}^{-1}\left(\mathbf{A}^H\right)^{-1}\mathbf{Q}_{xs} + \left(\mathbf{W}\mathbf{A}^H - \mathbf{Q}_{sx}\mathbf{A}^{-1}\right)\left(\mathbf{W}\mathbf{A}^H - \mathbf{Q}_{sx}\mathbf{A}^{-1}\right)^H\right). \end{aligned} \tag{4.85}$$

Since the first and the second term of this result do not depend on $\mathbf{W}$, the result is minimised when

$$\mathbf{W}\mathbf{A}^H - \mathbf{Q}_{sx}\mathbf{A}^{-1} = \mathbf{0} \tag{4.86}$$

When this is worked out, finally, $\mathbf{W}$ is obtained:

$$\mathbf{W} = \mathbf{Q}_{sx}\mathbf{A}^{-1}\left(\mathbf{A}^H\right)^{-1} = \mathbf{Q}_{sx}\left(\mathbf{A}^H\mathbf{A}\right)^{-1} = \mathbf{Q}_{sx}\mathbf{Q}_x^{-1}. \tag{4.87}$$

When we go back to our channel model ($\mathbf{x} = \mathbf{Hs} + \mathbf{n}$) and assume that $\mathbf{Q}_s = E[\mathbf{ss}^H] = \sigma_s^2\mathbf{I}$ and $\mathbf{Q}_n = E[\mathbf{nn}^H] = \sigma_n^2\mathbf{I}$ and that $\mathbf{Q}_x = E[\mathbf{xx}^H] = \mathbf{HQ}_s\mathbf{H}^H + \mathbf{Q}_n$ is invertible, then, with $\mathbf{Q}_{sx} = E[\mathbf{sx}^H] = E[\mathbf{s}(s^H\mathbf{H}^H + \mathbf{n}^H)] = \mathbf{Q}_s\mathbf{H}^H$ ($\mathbf{s}$ and $\mathbf{n}$ are assumed independent), $\mathbf{W}$ becomes

$$
\begin{aligned}
\mathbf{W} &= \mathbf{Q}_s\mathbf{H}^H\left(\mathbf{Q}_n + \mathbf{HQ}_s\mathbf{H}^H\right)^{-1} \\
&= \sigma_s^2\mathbf{H}^H\left(\sigma_n^2\mathbf{I}_{N_r} + \sigma_s^2\mathbf{HH}^H\right)^{-1} = \mathbf{H}^H\left(\alpha\mathbf{I}_{N_r} + \mathbf{HH}^H\right)^{-1},
\end{aligned}
\tag{4.88}
$$

where α equals σ_n^2/σ_s^2. When we have an inverse of a matrix with the form $\mathbf{A} + \mathbf{BCD}$, it is often useful to see what happens when the Matrix Inversion Lemma is applied (see Appendix A.1.8), which is given by

$$\left(\mathbf{A} + \mathbf{BCD}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}, \tag{4.89}$$

where, $\mathbf{A}$ and $\mathbf{C}$ are square invertible matrices. When applying the Matrix Inversion Lemma, $\mathbf{W}$ can be rewritten as follows:

$$
\begin{aligned}
\mathbf{W} &= \mathbf{Q}_s\mathbf{H}^H\left(\mathbf{Q}_n + \mathbf{HQ}_s\mathbf{H}^H\right)^{-1} \\
&= \mathbf{Q}_s\mathbf{H}^H\left(\mathbf{Q}_n^{-1} - \mathbf{Q}_n^{-1}\mathbf{H}\left(\mathbf{Q}_s^{-1} + \mathbf{H}^H\mathbf{Q}_n^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{Q}_n^{-1}\right) \\
&= \mathbf{Q}_s\mathbf{H}^H\mathbf{Q}_n^{-1} - \mathbf{Q}_s\mathbf{H}^H\mathbf{Q}_n^{-1}\mathbf{H}\left(\mathbf{Q}_s^{-1} + \mathbf{H}^H\mathbf{Q}_n^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{Q}_n^{-1} \\
&= \left(\mathbf{Q}_s\left(\mathbf{Q}_s^{-1} + \mathbf{H}^H\mathbf{Q}_n^{-1}\mathbf{H}\right) - \mathbf{Q}_s\mathbf{H}^H\mathbf{Q}_n^{-1}\mathbf{H}\right)\left(\mathbf{Q}_s^{-1} + \mathbf{H}^H\mathbf{Q}_n^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{Q}_n^{-1} \\
&= \left(\mathbf{Q}_s^{-1} + \mathbf{H}^H\mathbf{Q}_n^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{Q}_n^{-1} \\
&= \left(\alpha\mathbf{I}_{N_t} + \mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H.
\end{aligned}
\tag{4.90}
$$

So, summarising, it can be said that in order to minimise the Mean Square Error (over $\mathbf{W}$), the processing at the receiver must be equal to

$$\mathbf{W} = \left(\alpha\mathbf{I}_{N_t} + \mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H, \tag{4.91}$$

with $\alpha = N_t/\rho$. Based on this result, we can say intuitively that the MMSE solution trades off signalling separation quality for noise enhancement reduction. Furthermore, from above equation, it becomes clear that the ZF solution corresponds to an MMSE solution with $\alpha = 0$. On the other hand, the MMSE solution can be shown to be a ZF solution with the substitutions

$$\mathbf{H} \rightarrow \begin{pmatrix} \mathbf{H} \\ \sqrt{\alpha}\mathbf{I}_{N_t} \end{pmatrix} \text{ and } \mathbf{x} \rightarrow \begin{pmatrix} \mathbf{x} \\ \mathbf{0} \end{pmatrix}. \tag{4.92}$$

Note that in practice it is hard to measure the SNR, so, sometimes α is set to a fixed value that does not depend on the Signal-to-Noise Ratio.

4.7.2 Complexity

The complexity of the MMSE algorithm is almost equal to that of the ZF and is derived in Appendix C.3. The complexity of the MMSE processing in the preamble phase is found to be $4N_t^3 + N_t^2(8N_r - 2) - 2N_tN_r + N_t$ R_ADDs and $4N_t^3 + 8N_t^2N_r$ R_MULs.

The complexity of the MMSE in the payload processing is $2N_tN_r + 2N_t(N_r-1) + N_t \cdot \log_2(M)$ R_ADDs and $4N_tN_r$ R_MULs for every transmitted vector $\mathbf{s}$ to decode.

4.7.3 Soft-decision Output MMSE

Except for the fact that the covariance matrix of the estimation error $\mathbf{Q}$ for MMSE is somewhat different from that of ZF, the soft-decision values of MMSE are similar and can be obtained by substituting the $\mathbf{Q}$ of MMSE in the equations of subsection 4.6.4. The covariance matrix of the estimation error $\mathbf{s} - \mathbf{s}_{est}$ for MMSE can be shown to be

$$\begin{aligned}
\mathbf{Q} &= E\left[(\mathbf{s} - \mathbf{s}_{est})(\mathbf{s} - \mathbf{s}_{est})^H\right] = E\left[(\mathbf{s} - \mathbf{W}(\mathbf{Hs} + \mathbf{n}))(\mathbf{s} - \mathbf{W}(\mathbf{Hs} + \mathbf{n}))^H\right] \\
&= E\left[((\mathbf{I} - \mathbf{WH})\mathbf{s} - \mathbf{Wn})((\mathbf{I} - \mathbf{WH})\mathbf{s} - \mathbf{Wn})^H\right] \\
&= \sigma_s^2\left(\mathbf{I} - \mathbf{WH} - \mathbf{H}^H\mathbf{W}^H + \mathbf{WHH}^H\mathbf{W}^H\right) + \sigma_n^2\mathbf{WW}^H \\
&= [\left(\sigma_s^2\mathbf{I} - \sigma_s^2(\alpha\mathbf{I} + \mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{H}\right)(\alpha\mathbf{I} + \mathbf{H}^H\mathbf{H}) - \sigma_s^2\mathbf{H}^H\mathbf{H} \\
&\quad + \sigma_s^2(\alpha\mathbf{I} + \mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{HH}^H\mathbf{H} \\
&\quad + \sigma_n^2(\alpha\mathbf{I} + \mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{H}](\alpha\mathbf{I} + \mathbf{H}^H\mathbf{H})^{-1} \\
&= \sigma_n^2(\alpha\mathbf{I} + \mathbf{H}^H\mathbf{H})^{-1}.
\end{aligned} \tag{4.93}$$

4.8   ZF with SIC

4.8.1   Algorithm Description

The linear approaches of Section 4.6 and 4.7 are viable, but as will become clear from the results in Section 4.11, superior performance is obtained if non-linear techniques are used. One can imagine that if somehow first the most reliable element of the transmitted vector $\mathbf{s}$ could be decoded and used to improve the decoding of the other elements of $\mathbf{s}$, a better performance can be achieved. This is called *Successive Interference Cancellation* (SIC)

and it exploits the timing synchronism inherent in the system model (the assumption of co-located transmitters makes this reasonable). Furthermore, linear nulling (i.e., ZF) or MMSE is used to perform the detection. In other words, SIC is based on the subtraction of interference of already detected elements of **s** from the receiver signal vector **x**. This results in a modified receiver vector in which effectively fewer interferers are present.

When SIC is applied, the order in which the components of **s** are detected is important to the overall performance of the system. To determine a good detection order, the covariance matrix of the estimation error $\mathbf{s} - \mathbf{s}_{\text{est}}$ is used. Recalling from (4.44), the covariance matrix is given by

$$\mathbf{Q} = E\left[(\mathbf{s} - \mathbf{s}_{\text{est}})(\mathbf{s} - \mathbf{s}_{\text{est}})^H\right] = \sigma_n^2\left(\mathbf{H}^H\mathbf{H}\right)^{-1} \equiv \sigma_n^2\mathbf{P} , \tag{4.94}$$

or, using the pseudo-inverse,

$$\mathbf{P} = \mathbf{H}^\dagger\left(\mathbf{H}^\dagger\right)^H . \tag{4.95}$$

Let $(\mathbf{s}_{\text{est}})_p$ be the p-th entry of $\mathbf{s}_{\text{est}}$, then, the "best" estimate is the one for which $\mathbf{P}_{pp}$ (i.e., the p-th diagonal element of **P**) is the smallest, because this is the estimate with the smallest error variance. From (4.95) it becomes clear that $\mathbf{P}_{pp}$ is equal to the squared length of row p of $\mathbf{H}^\dagger$. Hence, finding the minimum squared length row of $\mathbf{H}^\dagger$ is equivalent.

Summarising, the decoding algorithm consists of three parts:

1. Ordering: determine the TX stream with the lowest error variance.
2. Interference nulling: estimate the strongest TX signal by nulling out all the weaker TX signals.
3. Interference cancellation: remodulate the data bits, subtract their contribution from the received signal vector and return to the ordering step.

Note that this principle is analogous to ZF Decision Feedback Equalisation (DFE) in the spatial domain ([38, 145]), where interference nulling corresponds to the feedforward filter and interference cancellation corresponds to the feedback filter. This algorithm was first introduced to the context of SDM in [144] where the algorithm was called Vertical-BLAST (V-BLAST). Furthermore, note that the principle is similar to successive interference cancellation in Multi-User Detection schemes.

Below, a more detailed description of the three recursive steps above is worked out:

1. Compute $\mathbf{H}^\dagger$, find the minimum squared length row of $\mathbf{H}^\dagger$, say it is the p-th, and permute it to be the last row. Permute the columns of **H** accordingly.

2. Form the estimate of the corresponding element of **s**. In case of ZF:

$$\left(\mathbf{s}_{\text{est}}\right)_p = \mathbf{w}^{N_t}\mathbf{x} , \tag{4.96}$$

where the weight vector $\mathbf{w}^{N_t}$ equals row N_t of the permuted $\mathbf{H}^\dagger$. Slice $(\mathbf{s}_\text{est})_p$ to the nearest constellation point $(\mathbf{s}_\text{est,sliced})_p$.

3. While $N_t - 1 > 0$ go back to step 1, but now with:

$$\mathbf{H} \to \mathbf{H}^{(N_t-1)} = \left(\mathbf{h}_1 \quad \cdots \quad \mathbf{h}_{N_t-1}\right), \ \mathbf{x} \to \mathbf{x} - \mathbf{h}_{N_t}\left(\mathbf{s}_\text{est,sliced}\right)_p \ \text{and} \ N_t \to N_t - 1. \quad (4.97)$$

Further simplification is possible when the QR decomposition ([113]) is used. Assume that the recursive process is in its $(k + 1)$-th run, then the dimensions of $\mathbf{H}$ are $N_r \times (N_t - k)$, determined with the original N_t. Based on the QR decomposition, we may write $\mathbf{H} = \mathbf{Q}_\text{QR}\mathbf{R}$, where $\mathbf{Q}_\text{QR}$ is a unitary matrix representing the feedforward filtering, and the upper triangular structure of $\mathbf{R}$ implies that in the $(k + 1)$-th run

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}_{(N_t-k)\times(N_t-k)} \\ \mathbf{0}_{(N_r-N_t+k)\times(N_t-k)} \end{pmatrix} \mathbf{R}', \quad (4.98)$$

where $\mathbf{R}'$ (the top $N_t - k$ rows of $\mathbf{R}$) is invertible. Then, based on the original N_t, the weight vector in (4.96) becomes

$$\mathbf{w}^{N_t-k} = \frac{1}{r_{(N_t-k)(N_t-k)}} \mathbf{q}_{N_t-k}^H, \quad (4.99)$$

where r_{yy} denotes element (y, y) of $\mathbf{R}$ and $\mathbf{q}_y$ the y-th column of $\mathbf{Q}_\text{QR}$.

4.8.2 Complexity

With respect to ZF, the ZF with SIC algorithm introduces extra complexity in the preamble phase as well as in the payload phase. The exact complexity analysis is given in Appendix C.4. From this analysis it is concluded that the complexity in the preamble-processing phase equals $N_t(6N_t^3 + 8N_t^2(1 + 2N_r) + 3N_t(1 + 6N_r) + 2N_r - 5)/6$ real additions and a number of real multiplications equal to $N_t^2(N_t + 1)^2 + 8N_rN_t(N_t + 1)(2N_t + 1)/6$. The complexity in the payload processing is shown to be $2N_t(4N_r - 1) + N_t\cdot\log_2(M)$ R_ADDs and $8N_tN_r$ R_MULs per transmitted vector $\mathbf{s}$. Note that an efficient low-complexity implementation of the SIC principle can be found in [49].

4.9 MMSE with SIC

4.9.1 Algorithm Description

In order to perform Successive Interference Cancellation with Minimum Mean Squared Error estimation, the SIC algorithm of Subsection 4.8.1 has to be adapted somewhat. Again, the covariance matrix of the estimation error $\mathbf{s} - \mathbf{s}_\text{est}$ will be used to determine a

good ordering for detection. Recalling from subsection 4.7.3, this covariance matrix can be shown to be

$$E\left[(\mathbf{s} - \mathbf{s}_{\text{est}})(\mathbf{s} - \mathbf{s}_{\text{est}})^H\right] = \sigma_n^2 \left(\alpha\mathbf{I} + \mathbf{H}^H\mathbf{H}\right)^{-1} \equiv \sigma_n^2 \mathbf{P}. \tag{4.100}$$

Note that $\mathbf{P}$ is somewhat different from the case where ZF is used as estimation technique. In order to do SIC based on the MMSE algorithm, the SIC algorithm is adapted resulting in

1. Compute $\mathbf{W}$ ($\mathbf{P}$ is obtained while determining $\mathbf{W}$). Find the smallest diagonal entry of $\mathbf{P}$ and suppose this is the p-th entry. Permute the p-th column of $\mathbf{H}$ to be the last column and permute the rows of $\mathbf{W}$ accordingly.

2. Form the estimate of the corresponding element of $\mathbf{s}$. In case of MMSE:

$$\left(\mathbf{s}_{\text{est}}\right)_p = \mathbf{w}^{N_t}\mathbf{x}, \tag{4.101}$$

where the weight vector $\mathbf{w}^{N_t}$ equals row N_t of the permuted $\mathbf{W}$. Slice $(\mathbf{s}_{\text{est}})_p$ to the nearest constellation point $(\mathbf{s}_{\text{est,sliced}})_p$.

3. While $N_t - 1 > 0$ go back to step 1, but now with:

$$\mathbf{H} \rightarrow \mathbf{H}^{(N_t-1)} = \left(\mathbf{h}_1 \quad \cdots \quad \mathbf{h}_{N_t-1}\right), \ \mathbf{x} \rightarrow \mathbf{x} - \mathbf{h}_{N_t}\left(\mathbf{s}_{\text{est,sliced}}\right)_p \ \text{and} \ N_t \rightarrow N_t - 1. \tag{4.102}$$

The BER performance of ZF with SIC and MMSE with SIC will be compared in Section 4.11.

4.9.2 Complexity

The complexity of MMSE with SIC is almost the same as that of Zero Forcing with SIC (see Subsection 4.8.2). Compared to ZF with SIC, there is only a slight difference in the preamble processing, namely, the addition of $\alpha\mathbf{I}$ in the calculation of $\mathbf{W}$. As is explained in Appendix C.5, this difference results in a complexity of MMSE with SIC in the preamble phase of $N_t(3N_t^3 + 4N_t^2(1 + 2N_r) + 3N_t(1 + 3N_r) + N_r - 1)/3$ real additions and a number of real multiplications equal to $N_t^2(N_t + 1)^2 + 8N_rN_t(N_t + 1)(2N_t + 1)/6$. The complexity in the payload processing does not change compared to ZF with SIC and can be shown to be $2N_t(4N_r - 1) + N_t \cdot \log_2(M)$ R_ADDs and $8N_tN_r$ R_MULs per transmitted vector $\mathbf{s}$.

4.10 Maximum Likelihood Detection (MLD)

4.10.1 Algorithm Description

Maximum Likelihood Detection (MLD) is a method that performs a maximum likelihood search over all possible transmitted vectors $\mathbf{s}$. The most likely transmitted vector is found as follows:

$$\mathbf{s}_{\mathrm{ml}} = \arg \min_{\mathbf{s}_i \in \{\mathbf{s}_1, \ldots, \mathbf{s}_I\}} \left\| \mathbf{x} - \mathbf{H}\mathbf{s}_i \right\|^2, \qquad (4.103)$$

where a search is performed over all vectors $\mathbf{s}_i$ that are part of the ensemble $\{\mathbf{s}_1, \ldots, \mathbf{s}_I\}$ formed by all possible transmitted vectors. Their number equals

$$I = M^{N_t}, \qquad (4.104)$$

where M denotes the number of constellation points. Note that for MLD it is not required that $N_t \leq N_r$.

A way to arrive at the most likely transmitted vector is by stating that we want to find the vector $\mathbf{s}_i$ from the ensemble $\{\mathbf{s}_1, \ldots, \mathbf{s}_I\}$ for which the probability $\Pr(\mathbf{s} = \mathbf{s}_i | \mathbf{x})$, or $\Pr(\mathbf{s}_i | \mathbf{x})$ in short, is maximal. This is called the *Maximum A posteriori Probability* (MAP). Note that according to [90], finding such a vector leads to the minimisation of the probability of error. When applying Bayes' rule, $\Pr(A|B) = \Pr(B|A) \cdot \Pr(A)/\Pr(B)$, the probability $\Pr(\mathbf{s}_i | \mathbf{x})$ may be expressed as

$$\Pr(\mathbf{s}_i | \mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{s}_i)\Pr(\mathbf{s}_i)}{p(\mathbf{x})}, \qquad (4.105)$$

where $p(\mathbf{x}|\mathbf{s}_i)$ is the conditional probability density function of the observed vector given that $\mathbf{s}_i$ has been sent and $\Pr(\mathbf{s}_i)$ is the probability of the i-th vector being transmitted. When no *a priori* knowledge is available on the probability that a certain vector is sent, it is best to assume that the I vectors are equally probable to be transmitted, hence $\Pr(\mathbf{s}_i) = 1/I$. When this assumption is made, the resulting detection method is not longer the MAP method but is generally called Maximum Likelihood Detection ([90]). In this case, using the fact that the denominator in (4.105) does not depend on $\mathbf{s}_i$, the decision rule based on finding the signal that maximises $\Pr(\mathbf{s}_i | \mathbf{x})$ is equivalent to finding the $\mathbf{s}_i$ that maximises $p(\mathbf{x}|\mathbf{s}_i)$.

Based on the signal model of Section 4.3 it can be shown that the conditional probability density function $p(\mathbf{x}|\mathbf{s}_i)$ is a multivariate complex normal distribution (see Appendix A.2). Under the assumption that $\mathbf{s}_i$ is sent, for a specific channel $\mathbf{H}$, the mean of $\mathbf{x}$ equals $\mathbf{H}\mathbf{s}_i$. This leads to the following probability density function:

$$p(\mathbf{x}|\mathbf{H},\mathbf{s}_i) = \det(\pi\mathbf{Q})^{-1} \exp\left(-(\mathbf{x} - \mathbf{H}\mathbf{s}_i)^H \mathbf{Q}^{-1}(\mathbf{x} - \mathbf{H}\mathbf{s}_i)\right), \qquad (4.106)$$

where $\mathbf{Q}$ is the covariance matrix and equals

$$\mathbf{Q} = E\left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^H\right] = E\left[(\mathbf{x} - \mathbf{H}\mathbf{s}_i)(\mathbf{x} - \mathbf{H}\mathbf{s}_i)^H\right] = E\left[\mathbf{n}\mathbf{n}^H\right] = \sigma_n^2 \mathbf{I}_{N_r}. \qquad (4.107)$$

The resulting conditional probability density function is

$$p(\mathbf{x}|\mathbf{H},\mathbf{s}_i) = \frac{1}{\left(\pi\sigma_n^2\right)^{N_r}} \exp\left(-\frac{1}{\sigma_n^2}(\mathbf{x} - \mathbf{H}\mathbf{s}_i)^H(\mathbf{x} - \mathbf{H}\mathbf{s}_i)\right). \qquad (4.108)$$

Consequently, finding the maximum of the conditional probability $\Pr(\mathbf{s}_i|\mathbf{x})$ leads to:

$$\arg \max_{\mathbf{s}_i \in \{\mathbf{s}_1, \ldots, \mathbf{s}_I\}} p(\mathbf{x}|\mathbf{H}, \mathbf{s}_i) = \arg \min_{\mathbf{s}_i \in \{\mathbf{s}_1, \ldots, \mathbf{s}_I\}} \|\mathbf{x} - \mathbf{H}\mathbf{s}_i\|^2 = \mathbf{s}_{\mathrm{ml}}, \qquad (4.109)$$

which equals (4.103) and because finding the maximum of the conditional probability $\Pr(\mathbf{s}_i|\mathbf{x})$ leads to the minimisation of the error probability ([90]), MLD is optimal in terms of BER performance (under the assumption that no a priori knowledge is available at the receiver on the probability that a given vector is transmitted, i.e., $\Pr(\mathbf{s}_i) = 1/I$). A major disadvantage of MLD, however, is that its complexity grows exponentially with N_t, because the complexity is proportional to I (see Subsection 4.10.3 and Appendix C.6).

4.10.2 Performance Analysis

In order to verify the simulation results of uncoded MLD, again the PEP approach is used to obtain an analytical upperbound on the performance. Let $\mathbf{s}_i$ and $\mathbf{s}_k$ be two possible spatial TX vectors with dimensions $N_t \times 1$ and assume that $\mathbf{s}_i$ is transmitted, with $i, k \in \{1, \ldots, I\}$. For a given MIMO channel $\mathbf{H}$, the probability that the MLD receiver decides erroneously in favour of $\mathbf{s}_k$ ($k \neq i$) equals (equivalent to Subsection 4.6.2)

$$\Pr(\mathbf{s}_i \to \mathbf{s}_k|\mathbf{H}) = Q\left(\sqrt{\frac{\|\mathbf{H}(\mathbf{s}_i - \mathbf{s}_k)\|^2}{2\sigma_n^2}}\right). \qquad (4.110)$$

Define $\mathbf{y} = \mathbf{H}(\mathbf{s}'_i - \mathbf{s}'_k)$, where $\mathbf{s}'_i$ and $\mathbf{s}'_k$ are the normalised versions of $\mathbf{s}_i$ and $\mathbf{s}_k$, respectively, such that $\mathbf{s}'_i = \mathbf{s}_i/\sigma_s$ and $\mathbf{s}'_k = \mathbf{s}_k/\sigma_s$. Then, using the Chernoff bound, the conditional PEP can be upperbounded by

$$\Pr(\mathbf{s}_i \to \mathbf{s}_k|\mathbf{H}) \leq \exp\left(-\frac{\|\mathbf{H}(\mathbf{s}_i - \mathbf{s}_k)\|^2}{4\sigma_n^2}\right) = \exp\left(-\frac{\sigma_s^2}{4\sigma_n^2}\mathbf{y}^H\mathbf{y}\right). \qquad (4.111)$$

Under the assumption of flat Rayleigh fading, the elements of $\mathbf{H}$ are assumed to be i.i.d. complex Gaussian distributed. As a result, $\mathbf{y}$ is a multivariate complex Gaussian distribution (see Appendix A.2). So, averaging over all channel realisations equals averaging over all $\mathbf{y}$. This leads to the following PEP

$$\Pr(\mathbf{s}_i \to \mathbf{s}_k) \leq \int_{\mathbf{y}} \exp\left(-\frac{\sigma_s^2}{4\sigma_n^2}\mathbf{y}^H\mathbf{y}\right) p(\mathbf{y})d\mathbf{y}$$

$$= \int_{\mathbf{y}} \exp\left(-\frac{\sigma_s^2}{4\sigma_n^2}\mathbf{y}^H\mathbf{y}\right) \frac{1}{\pi^{N_r} \det(\mathbf{Q}_y)} \exp\left(-\mathbf{y}^H\mathbf{Q}_y^{-1}\mathbf{y}\right)d\mathbf{y} \qquad (4.112)$$

$$= \frac{\det(\mathbf{B})}{\det(\mathbf{Q}_y)} \int_{\mathbf{y}'} \frac{1}{\pi^{N_r} \det(\mathbf{B})} \exp\left(-\mathbf{y}'^H \mathbf{B}^{-1}\mathbf{y}'\right)d\mathbf{y}' = \frac{\det(\mathbf{B})}{\det(\mathbf{Q}_y)},$$

where in the last line $\mathbf{y}$ is redefined as a multivariate complex Gaussian distribution $\mathbf{y}'$ with the covariance matrix

$$\mathbf{B} = \left(\frac{\sigma_s^2}{4\sigma_n^2} \mathbf{I}_{N_r} + \mathbf{Q}_y^{-1} \right)^{-1} . \tag{4.113}$$

Since $\det(\mathbf{B}) = 1/\det(\mathbf{B}^{-1})$, now the PEP upperbound can be written as

$$\Pr(\mathbf{s}_i \rightarrow \mathbf{s}_k) \leq \frac{\det(\mathbf{B})}{\det(\mathbf{Q}_y)} = \det\left(\mathbf{I}_{N_r} + \frac{\sigma_s^2}{4\sigma_n^2} \mathbf{Q}_y \right)^{-1} . \tag{4.114}$$

The covariance matrix of $\mathbf{y}$ for a given $\mathbf{s}_i$ and $\mathbf{s}_k$ is defined as

$$\mathbf{Q}_y = E\left[\mathbf{y}\mathbf{y}^H \right] = E\left[\mathbf{H}(\mathbf{s}'_i - \mathbf{s}'_k)(\mathbf{s}'_i - \mathbf{s}'_k)^H \mathbf{H}^H \right] . \tag{4.115}$$

Since the matrix $(\mathbf{s}'_i - \mathbf{s}'_k)(\mathbf{s}'_i - \mathbf{s}'_k)^H$ is Hermitian, we already know from Section 3.6 that their exists a unitary matrix $\mathbf{U}$ and a real diagonal matrix $\mathbf{\Lambda}$ containing the eigenvalues on its diagonal such that $(\mathbf{s}'_i - \mathbf{s}'_k)(\mathbf{s}'_i - \mathbf{s}'_k)^H = \mathbf{U}^H \mathbf{\Lambda} \mathbf{U}$. But it is also clear that $(\mathbf{s}'_i - \mathbf{s}'_k)(\mathbf{s}'_i - \mathbf{s}'_k)^H$ is rank one and consequently has only one eigenvalue, namely $\lambda_{ik} = (\mathbf{s}'_i - \mathbf{s}'_k)^H(\mathbf{s}'_i - \mathbf{s}'_k)$. Assume without loss of generality that this eigenvalue is positioned on the p-th diagonal entry of $\mathbf{\Lambda}$. Then we may write

$$\mathbf{Q}_y = E\left[\mathbf{H} \begin{pmatrix} \mathbf{0}_{(p-1)\times N_r} \\ \lambda_{ik}\mathbf{h}_p^H \\ \mathbf{0}_{(N_t - p)\times N_r} \end{pmatrix} \right], \tag{4.116}$$

where $\mathbf{h}_p^H$ denotes the Hermitian transpose of the p-th column of $\mathbf{H}$, or, equivalently, the p-th row of $\mathbf{H}^H$. When we assume, according to the flat Rayleigh fading model of Subsection 3.5.2, that the elements of $\mathbf{H}$ are i.i.d. zero mean complex Gaussian distributed with a variance of one, it can be shown that

$$\mathbf{Q}_y = \lambda_{ik}\mathbf{I}_{N_r} . \tag{4.117}$$

This finally results in the PEP upperbound

$$\Pr(\mathbf{s}_i \rightarrow \mathbf{s}_k) \leq \det\left(\mathbf{I}_{N_r} + \frac{\sigma_s^2}{4\sigma_n^2} \lambda_{ik}\mathbf{I}_{N_r} \right)^{-1} = \left(1 + \lambda_{ik}\frac{\rho}{4N_t} \right)^{-N_r} . \tag{4.118}$$

Since it is already shown in Subsection 4.6.2 that the BER is proportional to a (weighted) sum of PEPs, it can be concluded that MLD achieves a diversity order of N_r (which was also shown in [125], but with a much looser upperbound).

Using the union bound (see (4.55)), we can show that the average MIMO vector (or symbol) probability equals

$$\mathrm{Pr}_s \leq \frac{1}{M^{N_t}} \sum_{i=1}^{M^{N_t}} \sum_{\substack{k=1 \\ k \neq i}}^{M^{N_t}} \mathrm{Pr}\!\left(\mathbf{s}_i \to \mathbf{s}_k\right), \qquad (4.119)$$

and, an approximation on the bit error probability can be found with $\mathrm{Pr}_b \approx \mathrm{Pr}_s/(N_t M)$.

4.10.3 Complexity

As can be deduced from Appendix C.6, the complexity of MLD strongly depends on the available memory size. To determine the complexity figures, in Appendix C.6 two cases are distinguished: a *minimum amount of memory* and a *maximum amount of memory* available. This led to the following complexity numbers.

For the minimum amount of memory, the complexity during the preamble processing is $2MN_rN_t$ R_ADDs and $4MN_rN_t$ R_MULs, and the complexity during the payload phase per transmitted vector equals $2N_rI$ R_MULs and

$$2N_r M \frac{I-1}{M-1} + 2N_r I - 1 \ \ \mathrm{R\_ADDs.} \qquad (4.120)$$

In the case enough memory is available (i.e., the "maximum amount of memory" case), the complexity figures for the preamble phase are, respectively, $4MN_rN_t$ R_MULs and

$$2M^2 N_r \frac{M^{N_t-1}-1}{M-1} + 2MN_rN_t \ \ \mathrm{R\_ADDs.} \qquad (4.121)$$

While for the payload processing, the complexity count equals in this case $4N_rI - 1$ R_ADDs and $2N_rI$ R_MULs per transmitted vector.

To obtain the complexity for the entire packet, these complexity numbers have to be multiplied by the number of spatial vectors within a packet N_s. Note that these complexity figures increase linearly with the number of receiving antennas and exponentially with the number of transmit antennas.

It is possible to reduce this complexity considerably without too much loss in performance. From Appendix C.6, it is clear that one of the significant terms in the complexity calculation for MLD is the determination of the norm of $\mathbf{x} - \mathbf{Hs}_i$ (commonly known as the squared l_2 norm ([54]): $\|\mathbf{x} - \mathbf{Hs}_i\|^2$). In order to reduce the complexity, an approximation of the l_1 norm can be used:

$$\begin{aligned}
\|\mathbf{x} - \mathbf{Hs}_i\| &\approx \\
&\left|\mathrm{Re}\!\left(x_1 - \mathbf{h}^1\mathbf{s}_i\right)\right| + \left|\mathrm{Im}\!\left(x_1 - \mathbf{h}^1\mathbf{s}_i\right)\right| + \ldots + \left|\mathrm{Re}\!\left(x_{N_r} - \mathbf{h}^{N_r}\mathbf{s}_i\right)\right| + \left|\mathrm{Im}\!\left(x_{N_r} - \mathbf{h}^{N_r}\mathbf{s}_i\right)\right|,
\end{aligned} \qquad (4.122)$$

where $\mathbf{h}^q$ stands for the q-th row of $\mathbf{H}$. This approximation consists only of real additions (and no multiplications) making the MLD algorithm less complex. The drawback of MLD

with the approximated norm is that the BER performance deteriorates by approximately 0.5 dB (see, e.g., Figure 4-18 and [129]).

The new norm definition does not influence the complexity of the preamble processing. The complexity of the payload phase, however, is significantly reduced since the calculation of the approximated l_1 norm results in a complexity of $(2N_r - 1)I$ real additions and no multiplications. As a result, the overall complexity per TX vector for this reduced complexity case equals $4N_rI - 1$ R_ADDs. Clearly, this is a significant reduction in complexity compared to the previous numbers, but the complexity still increases exponentially with N_t. More ideas for complexity reduction are published, e.g., in [11].

4.10.4 Soft-decision Output MLD

The generation of soft-decision output values for MLD is based on the maximum a posteriori probability (MAP) principle. Again, the Log-Likelihood Ratio is used as an indication for the reliability of a bit. Based on the analysis of Subsection 4.6.4 and assuming that b_k is the k-th bit of the transmitted vector to estimate, the LLR value of the estimated bit can be shown to be

$$L(b_k) = \ln \frac{\Pr(b_k = 1 | \mathbf{x})}{\Pr(b_k = 0 | \mathbf{x})} = \ln \frac{\sum_{\mathbf{s}_i | b_k = 1} \Pr(\mathbf{s}_i | \mathbf{x})}{\sum_{\mathbf{s}_i | b_k = 0} \Pr(\mathbf{s}_i | \mathbf{x})} = \ln \frac{\sum_{\mathbf{s}_i | b_k = 1} p(\mathbf{x} | \mathbf{s}_i) \Pr(\mathbf{s}_i)}{\sum_{\mathbf{s}_i | b_k = 0} p(\mathbf{x} | \mathbf{s}_i) \Pr(\mathbf{s}_i)}. \tag{4.123}$$

Because the vectors $\mathbf{s}_i$ are equally likely to be transmitted, $\Pr(\mathbf{s}_i)$ is equal for all vectors $\mathbf{s}_i$. From Subsection 4.10.1, we already know that

$$p(\mathbf{x} | \mathbf{H}, \mathbf{s}_i) = \det(\pi \mathbf{Q})^{-1} \exp\left(-(\mathbf{x} - \mathbf{H}\mathbf{s}_i)^H \mathbf{Q}^{-1} (\mathbf{x} - \mathbf{H}\mathbf{s}_i)\right), \tag{4.124}$$

with

$$\begin{aligned} \mathbf{Q} &= E\left[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^H\right] \\ &= E\left[(\mathbf{x} - \mathbf{H}\mathbf{s}_i)(\mathbf{x} - \mathbf{H}\mathbf{s}_i)^H\right] = E\left[\mathbf{n}\mathbf{n}^H\right] = \sigma_n^2 \mathbf{I}_{N_r} \end{aligned} \tag{4.125}$$

This leads to the LLR of b_k:

$$L(b_k) = \ln \frac{\sum_{\mathbf{s}_i | b_k = 1} \exp\left(-\dfrac{\|\mathbf{x} - \mathbf{H}\mathbf{s}_i\|^2}{\sigma_n^2}\right)}{\sum_{\mathbf{s}_i | b_k = 0} \exp\left(-\dfrac{\|\mathbf{x} - \mathbf{H}\mathbf{s}_i\|^2}{\sigma_n^2}\right)}. \tag{4.126}$$

When applying the max-log approximation this results in

$$L(b_k) \approx \ln \frac{\exp\left(\max_{\mathbf{s}_i|b_k=1} -\frac{\|\mathbf{x}-\mathbf{Hs}_i\|^2}{\sigma_n^2}\right)}{\exp\left(\max_{\mathbf{s}_i|b_k=0} -\frac{\|\mathbf{x}-\mathbf{Hs}_i\|^2}{\sigma_n^2}\right)} \tag{4.127}$$

$$= \frac{1}{\sigma_n^2}\left(\min_{\mathbf{s}_i|b_k=0}\|\mathbf{x}-\mathbf{Hs}_i\|^2 - \min_{\mathbf{s}_i|b_k=1}\|\mathbf{x}-\mathbf{Hs}_i\|^2\right).$$

From this result, it can be observed that Soft-decision Output MLD (SOMLD) is hardly more complex than hard-decision output MLD. Namely, at the same time that the norm of $\mathbf{x} - \mathbf{Hs}_i$ is determined, the search for the minima can be performed "on the fly". For SOMLD, we need to keep track of $2N_t m$ minima, e.g., in case of a 2×2 system with BPSK we need to keep track of the minima $b_{\{0,1\}} = 0$ and $b_{\{0,1\}} = 1$, and finally calculate the LLR according to (4.127). A possible architecture for this example is shown in Figure 4-14.
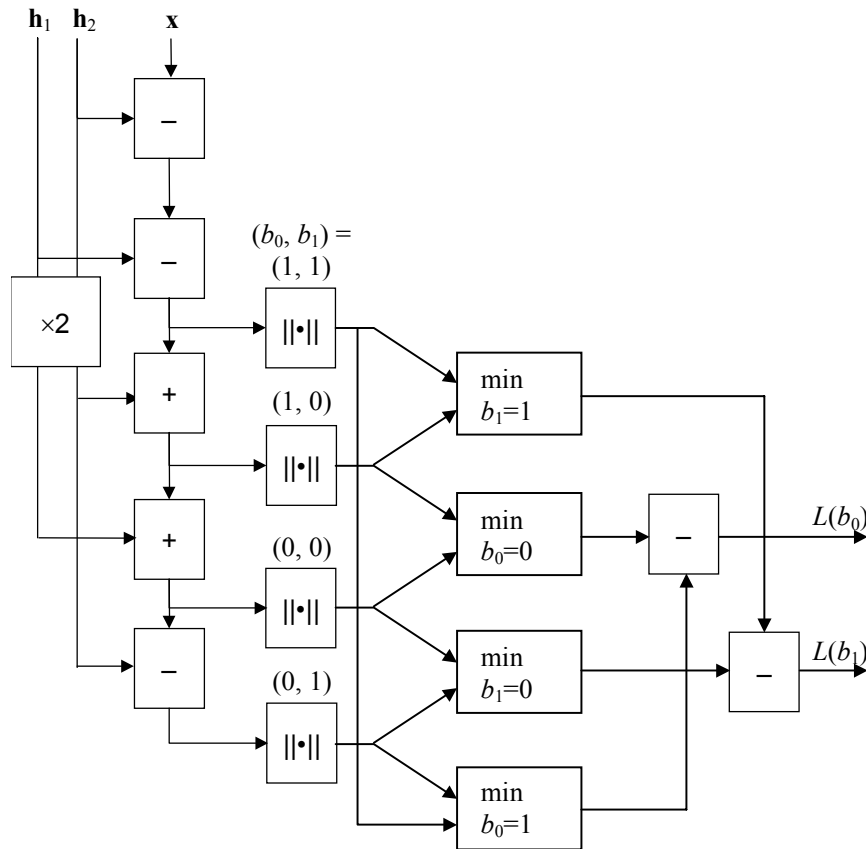
**Figure 4-14: A simplified 2 × 2 Soft-decision Output
Maximum Likelihood Detector for BPSK.**

From (4.127), it can furthermore be observed that for (very) high SNRs the LLR can become (very) large. Following the analysis of Subsection 4.6.4, it can be shown that, for a practical implementation, the same quantisation constant as given by (4.79) can be used to limit the soft-decision values of SOMLD.

4.11 Performance Comparison

4.11.1 Simulations without Coding

The SDM techniques, which are described in this chapter, are programmed in MATLAB and some simulations are performed to obtain BER characteristics. These characteristics are used to compare the performance of the different SDM algorithms. In all simulations, it is assumed that the channel is perfectly known to the receiver and, unless mentioned otherwise, the BER performance is obtained by averaging over 100,000 channel realisations and packets with a length of 64 bytes (counted at the input of the TX processing).

In Figure 4-15, the BERs for different SDM techniques are depicted against the average SNR per receive antenna for a 2×2 system that operates in a flat Rayleigh fading environment (see Subsection 3.5.2). A BPSK modulation scheme is used and there is no coding applied. As a reference, the BER of a 1×1 system is included. Note that for all SDM algorithms, the 1×1 performance is equivalent. Furthermore, the upperbounds of ZF and MLD are depicted, which are given by (4.58) and (4.119), respectively. Clearly, the closed-form expression for the BER of ZF, (4.58), provides an exact match proving the correctness of the simulation, whereas the union bound of MLD is loose. The slopes of MLD and its upperbound, however, are equivalent.
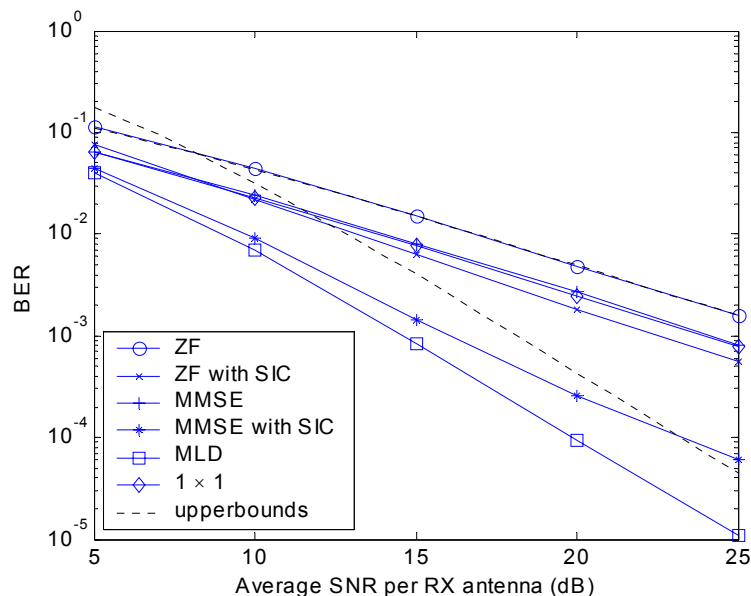
Figure 4-15: BER versus the average SNR per RX antenna for a 2×2 system in flat Rayleigh fading, BPSK, no coding, and different SDM algorithms. As a reference, the 1×1 case is provided as well as the ZF and MLD upperbounds.

From Figure 4-15, it can be concluded that MLD has the best performance. An intuitive explanation for this is that MLD performs detection and uses a priori knowledge about the possible constellation points that are sent, whereas the other techniques just performs

estimation. Furthermore, the slope of the MLD curve shows a diversity order of 2, whereas the slopes of the other curves tend towards diversity order 1. This conclusion is supported by the PEP analysis of Subsections 4.6.2 and 4.10.2, where it is shown that the diversity order of ZF is $N_r - N_t + 1 = 1$ and that of MLD equals N_r. Furthermore, note that the performance of the 1×1 system is 3 dB better than that of 2×2 ZF. This can also be concluded from the PEP analysis of ZF. Finally, note that especially the successive interference cancellation in combination with MMSE detection significantly outperforms regular MMSE.

In Figure 4-16, the results of similar simulations are presented, but now for a system with 2 TX and 4 RX antennas. In this case, the diversity order of ZF is 3 and that of MLD is 4. Clearly, the performance of MMSE with SIC in this case is very close to the performance of MLD.
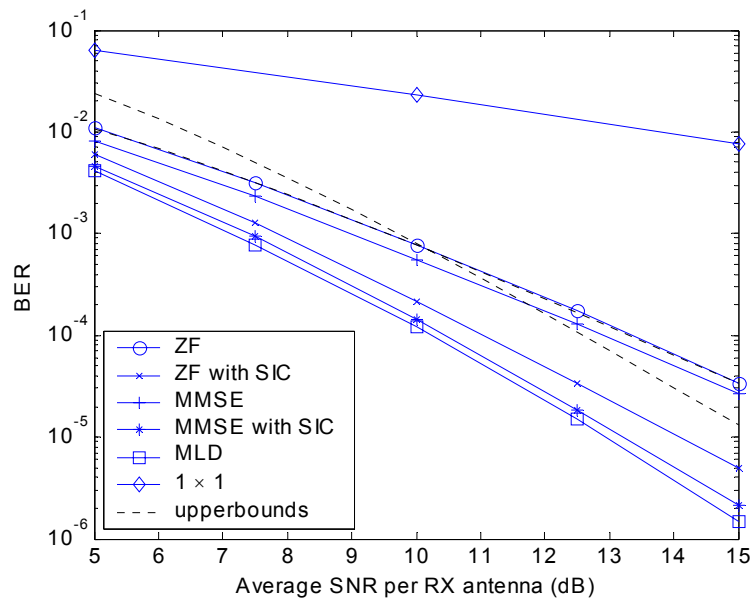

Figure 4-16: BER versus the average SNR per RX antenna for a 2 × 4 system in flat Rayleigh fading, BPSK, no coding, and different SDM algorithms. As a reference, the 1 × 1 case is provided as well as the ZF and MLD upperbounds.

In line with the previous simulations, Figure 4-17 shows the results for a 4×4 system. From these results it can be seen that MMSE with SIC still follows the performance of MLD quite well, whereas MLD by far outperforms the other schemes due to its diversity order of 4. Furthermore, the 4×4 ZF results are 6 dB worse than the performance of the SISO case, which again also follows from the PEP analysis of ZF (see (4.54)).

When we, however, would go to a higher constellation order, say 16-QAM, the performance of MMSE with SIC is dropped considerable compared to MLD as can be seen in Figure 4-18. So, SIC does not work that well for large constellation sizes. This can be explained by the fact that higher constellation orders are more vulnerable for noise, resulting in more errors in the SIC information and, correspondingly, resulting in a performance loss. Moreover, the performance of ZF and MMSE almost coincide.
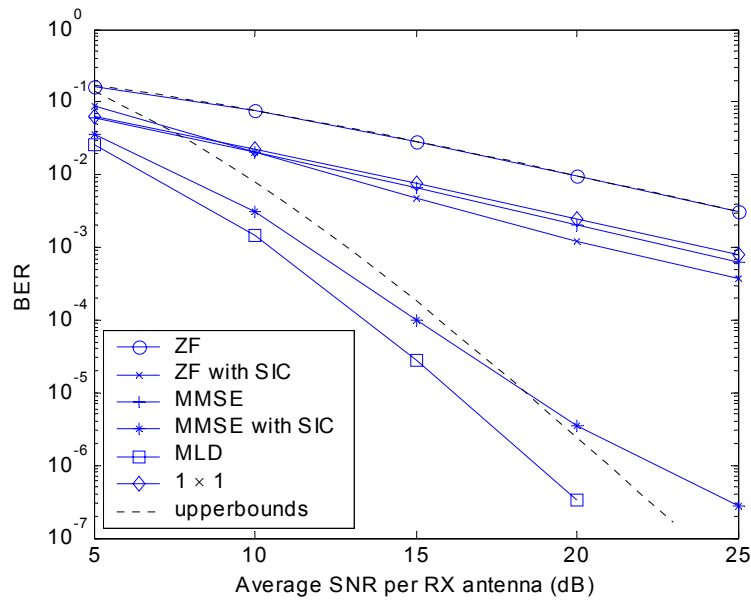
Figure 4-17: BER versus the average SNR per RX antenna for a 4 × 4 system in flat Rayleigh fading, BPSK, no coding, and different SDM algorithms. As a reference, the 1 × 1 case is provided as well as the ZF and MLD upperbounds.
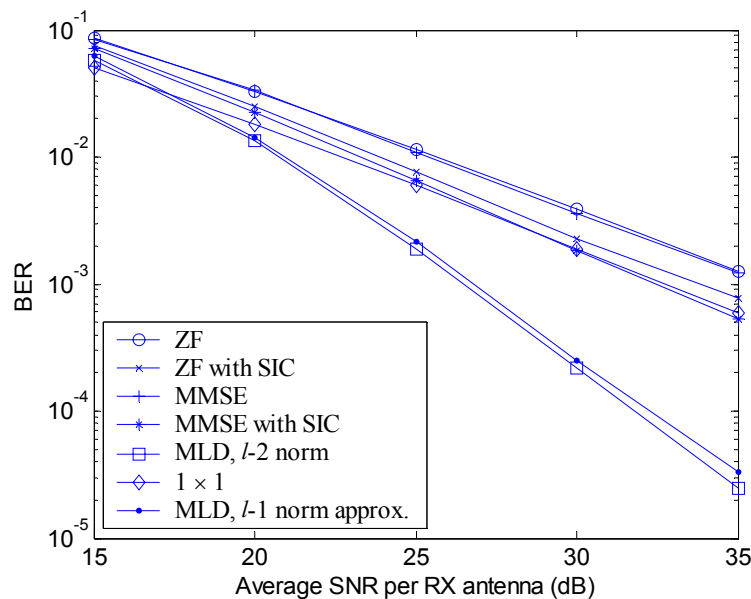
Figure 4-18: BER versus the average SNR per RX antenna for a 2 × 2 system in flat Rayleigh fading, 16-QAM, no coding, and different SDM algorithms (including a comparison between MLD with the l_1-norm approximation and MLD based on the l_2 norm). As a reference, the 1 × 1 case is provided.

In Figure 4-19, simulation results are depicted for a 2 × 2 system operating in an orthogonal channel that is only influenced by AWGN (see Subsection 3.5.4). By writing out the signal model for this type of channels, it can be easily verified that the performance of an $N \times N$ system operating in an orthogonal AWGN channel equals that of a SISO

system under AWGN. Consequently, the curves of Figure 4-19 are equal to those of a SISO system presented in [90] and do not depend on the SDM processing that is used. The SNR difference between the curves of about 10 dB can be explained as follows. From Table 3.1 in [126] we learn that in E_b/N_0 the theoretical shift between BPSK and 16-QAM is 3.98 dB. Furthermore, based on (4.34), an extra 6 dB shift can be expected in the SNR when going from BPSK to 16-QAM modulation. In total, this results in the 10 dB shift observed in Figure 4-19. When comparing the performance between BPSK and 16-QAM for a flat Rayleigh fading channel, the shift strongly depends on the applied SDM technique. The shift for MLD between Figure 4-17 and Figure 4-18 at a BER of 10^{-3} is equal to $22 - 11 = 11$ dB. For MMSE with SIC, this is about $29.5 - 12 = 17.5$ dB and for ZF at a BER of 10^{-2} it is about $29 - 20 = 9$ dB. These differences most likely can be explained by the different non-linear effects that occur in the MIMO algorithms when operating in flat Rayleigh fading.
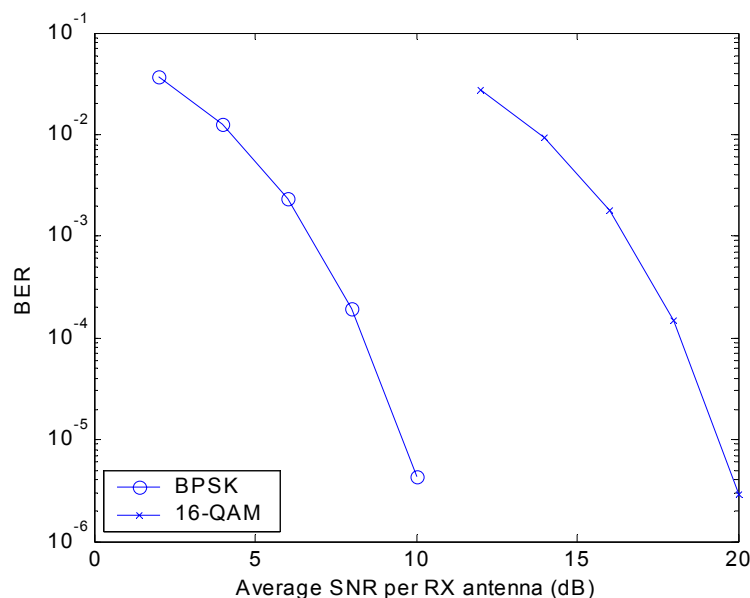
**Figure 4-19: BER versus the average SNR per RX antenna for a
2×2 system in an AWGN channel, BPSK and 16-QAM, no coding.**

So far, we have looked to simulations without correlation. The question is what the effect of spatial correlation is. To that end, simulations are performed with the correlation matrices introduced in (3.68) and (3.69). Assume that the correlation coefficients r_{TX} and r_{RX} are equal: $r = r_{TX} = r_{RX}$. For a 4×4 system with a spatial correlation of $r = 0.6$, the BER characteristics are shown in Figure 4-20 for BPSK. Compared to the results of Figure 4-17, a performance penalty is observed (for MLD, also the curve without correlation is depicted in Figure 4-20). The performance penalty strongly depends on the detection algorithm. For MLD, we see a degradation of about 2.5 dB, whereas for ZF the performance drops by more than 5 dB. Moreover, the performance penalty strongly depends on the SNR and on r. To make this clear, in Figure 4-21 the BER versus r is shown for the average SNR per RX values 10, 20 and 30 dB, both for ZF and for MLD. Note that when the BER is similar, the degradation as function of r for ZF and MLD is comparable, which can be concluded from comparing the result of ZF with an average SNR per RX of 30 dB with that of MLD with an average SNR per RX of 10 dB. Note that

the degradation we talk about here is the BER degradation for a given SNR (i.e., a vertical shift of a performance point at a given SNR). Since the MLD curve has a steeper slope due to its diversity advantage, the corresponding horizontal shift of MLD is less than that of ZF, which is observed in Figure 4-20.
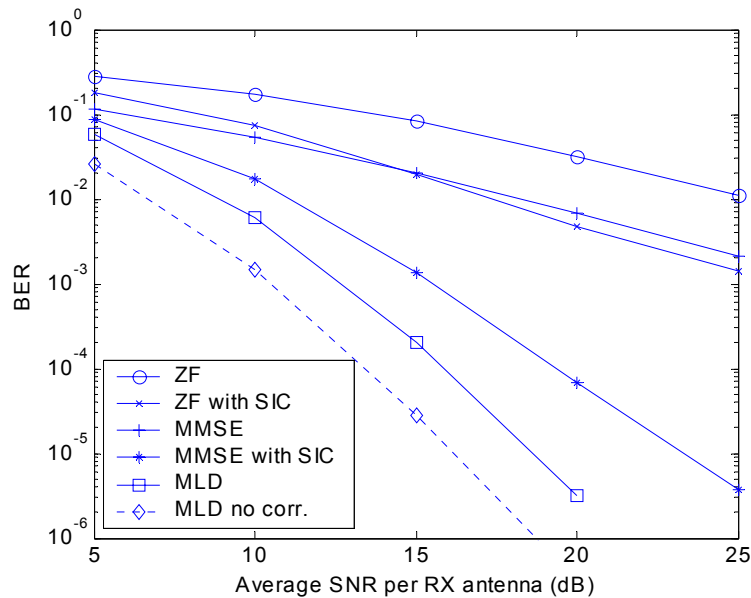
Figure 4-20: BER versus the average SNR per RX antenna for a 4 × 4 system in flat Rayleigh fading with correlation following (3.68) and (3.69) with $r = r_{TX} = r_{RX} = 0.6$, BPSK, no coding, and different SDM algorithms. As a reference, MLD without correlation is provided.
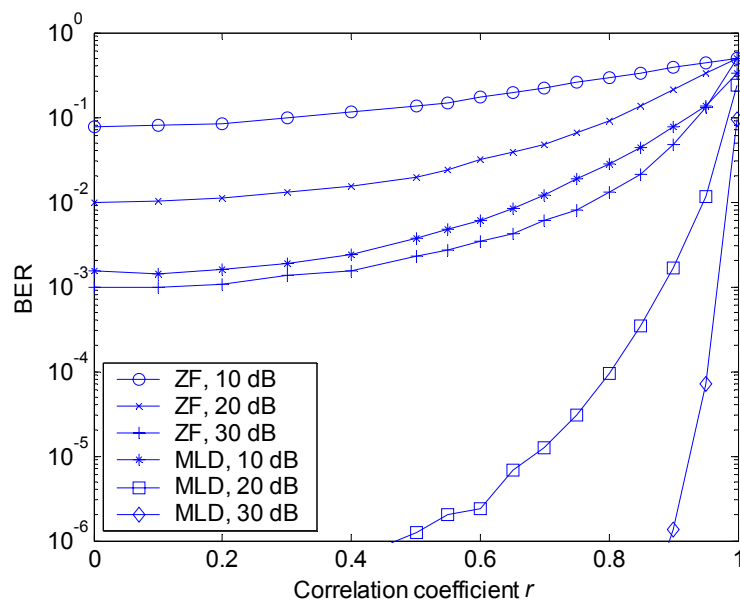
Figure 4-21: BER versus the correlation coefficient $r = r_{TX} = r_{RX}$ for a 4 × 4 system in flat Rayleigh fading with spatial correlation following (3.68) and (3.69), BPSK, no coding, and ZF and MLD for different average SNRs per RX antenna.

When the channel conditions change from a rich-scattering environment, modelled by flat Rayleigh fading, to a channel with more LOS, the performance also deteriorates. The amount of deterioration can be read from Figure 4-22 for a 4×4 system with ZF or MLD as detection technique.
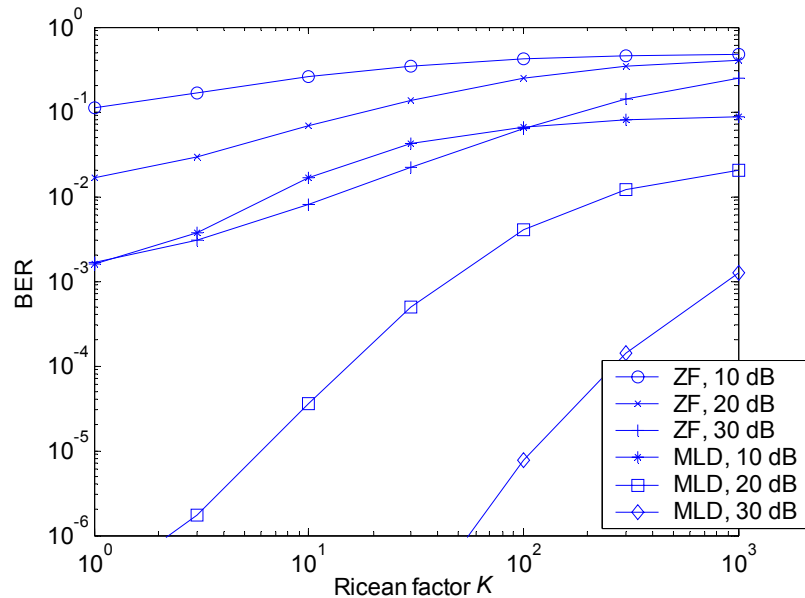
Figure 4-22: BER versus the Ricean factor K for a 4×4 system in flat Ricean fading without spatial correlation, with BPSK modulation, no coding, and ZF and MLD for different average SNRs per RX antenna.

4.11.2 Simulations with Coding

In this subsection, simulation results are shown of MIMO combined with coding. The performance evaluations of this subsection all focus on a $2 \times N_r$ MIMO system with coding rate $r = \frac{1}{2}$ and QPSK modulation. This results in an overall spectral efficiency of 2 bits/s/Hz. The codes that are compared are the Alamouti scheme (see Subsection 4.2.2), an 8-state STTC (see Figure 4-4), and SDM with joint encoding (see Figure 4-23) based on an 8-state convolutional code with (15, 17) as generator polynomial denoted in octal notation. For the coded SDM, the modulation is performed after the spatial mapping, which in this case simply is a demultiplexer.
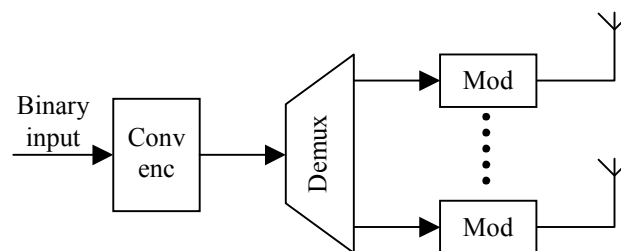
Figure 4-23: Joint encoding for SDM.

As receiver algorithm for the Alamouti scheme, we used the algorithm described by (4.5). For the 8-state STTC, the Viterbi algorithm is used to compute the path with the lowest accumulated metric and the branch metric for a given transition, denoted by, say, $s_i(t)$ is given by $\|x(t) - Hs_i(t)\|^2$ ([116]). For SDM with joint encoding SOMLD is chosen as receiving algorithm, since in the previous section it is shown that MLD has the best performance. After the spatial processing, the convolutional encoded bits are decoded by a Viterbi algorithm that accepts the soft-decision output values of SOMLD.

To obtain the BER and PER performance, the average of the bit and packet errors is taken over 100,000 channel realisations, with one 64-byte packet transmission per channel realisation. Note that the packetlength is defined as the length of the binary input bit sequence.

Figure 4-24 shows the BER and PER results of the 2×1 and 2×2 simulations. For the 8-state STTC scheme, only the PER results are shown. For the 2×1 scenario, the Alamouti scheme and the 8-state STTC perform similarly, whereas SOMLD with an 8-state convolutional outer code performs considerably worse. For the 2×2 case, however, it can be observed that for low SNRs, the PER performance of this coded SDM system is better than the Alamouti scheme. Interestingly, this is not the case for the BER performance. Apparently, the errors of SOMLD with coding occur more in bursts, resulting in a lower PER.
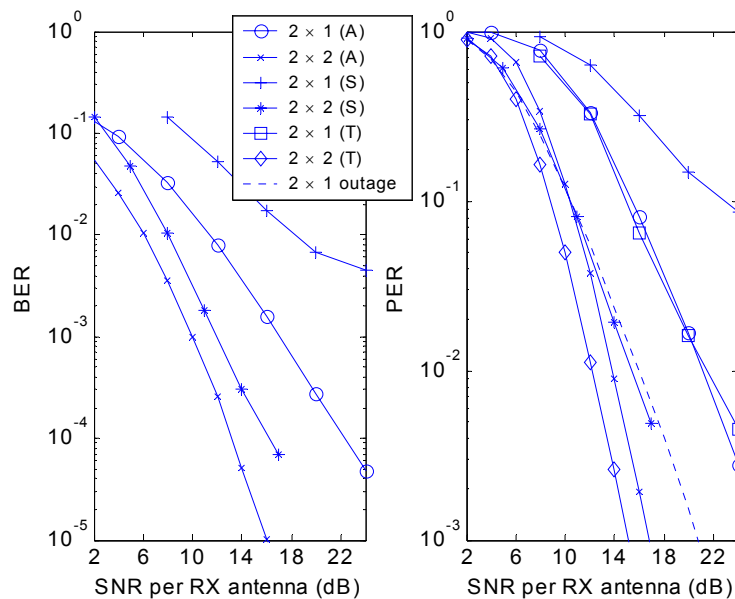
Figure 4-24: BER and PER versus the average SNR per RX antenna for a 2×1 and 2×2 system in flat Rayleigh fading with an efficiency of 2 bits/s/Hz and as coding scheme: Alamouti (A), 8-state STTC (T), and SOMLD with an 8-state convolutional outer code (S). Also the 2×1 PER lowerbound (i.e., outage) is shown.

In Figure 4-24 also the 2×1 outage PER is shown as it is defined in Subsection 4.4.6. This curve represents the PER lowerbound for a spectral efficiency of 2 bit/s/Hz. In Section 4.2 we observed that in the 2×1 case, the Alamouti scheme is optimal, in the sense that it is capable of attaining the system's Shannon capacity. In a performance sense, however, we

see that the performance is still about 4.5 dB worse than the PER lowerbound. This performance gap can be explained by the fact that according to Shannon the PER lowerbound is only achievable with progressively strong overlay encoding, and infinite packet sizes. Since we used a limited packetlength and no outer code on top of the Alamouti scheme, the observed performance penalty occurs.

For the 2×4 and 2×8 case, the simulations results are depicted in Figure 4-25. We clearly see that both the 8-state STTC and SOMLD with an 8-state convolutional outer code outperform the Alamouti scheme. Furthermore, we see that for the 2×8 scenario, for a PER higher than 0.005, SOMLD with coding performs better than the 8-state STTC, and also for the 2×4 case in the region of interest the PERs of both schemes are very close. This provides a strong indication that the Euclidean distance criterion (see Subsection 4.2.2) for higher diversity orders is a better design rule that the rank and determinant criteria. Note, by the way, that the curves of the STCs (Alamouti and 8-state STTC) are steeper than SOMLD with coding due to their TX diversity advantage. In Figure 4-25 also the 2×4 outage PER is shown as it is defined in Subsection 4.4.6. Apparently the 2×4 schemes as still more the 4 dB worse than this theoretical PER lowerbound. Maybe applying the turbo principle ([138]) might result in a performance closer to this bound (see Section 4.13).
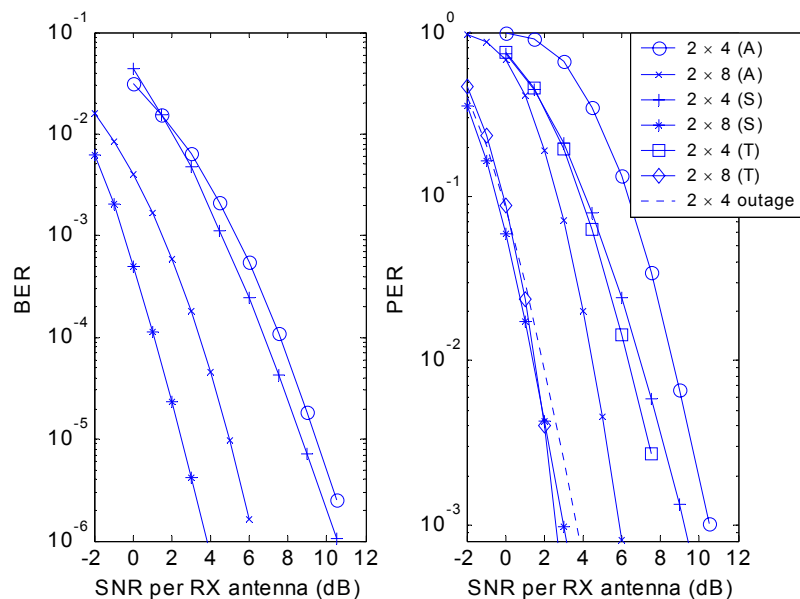
Figure 4-25: BER and PER versus the average SNR per RX antenna for a 2×4 and 2×8 system in flat Rayleigh fading with an efficiency of 2 bits/s/Hz and as coding scheme: Alamouti (A), 8-state STTC (T), and SOMLD with an 8-state convolutional outer code (S). Also the 2×4 PER lowerbound (i.e., outage) is shown.

4.11.3 Complexity Comparison

Based on the complexity figures of the previous sections, it is possible to compare the complexity of the described SDM algorithms. In this comparison, we will assume that the algorithms will be implemented with 8-bit operations as basis. Based on this assumption

we will further assume that the complexity of an 8-bits multiplier is ten times higher than that of an 8-bits adder. This is based on the fact that a multiplication by a power of 2 is in octal notation equivalent to a shift operation, so multiplying two 8 bits numbers in general requires 8 shifts and the summation of the results (i.e., 7 additions). It is reasonable to assume that the total complexity of 8 shifts and 7 additions is roughly ten times that of a single addition, thus, it is reasonable to assume that the complexity of an 8-bits multiplier is ten times higher than that of an 8-bits adder.

Based on above consideration, we can express the complexity as a single number in terms of equivalent ADDs[1]. Furthermore, the complexity is determined for 64 byte packets and it is assumed that the symbol time of one MIMO vector is 4 μs. Note that the number of MIMO vectors within one packet (N_s) is inversely proportional to the number of TX antennas (N_t). Moreover, in the complexity measure, we assume that the complexity of the packet detection, time and frequency synchronisation, and channel estimation does not depend on the chosen MIMO technique and, therefore, we do not take it into account.

Based on above parameters, the overall complexity, i.e., the sum of the preamble and payload processing complexity, of symmetric systems, i.e., $N_t = N_r$, for BPSK and various SDM algorithms is given in Figure 4-26. For MLD, only the complexity of the processing with the maximum amount of memory is displayed. Furthermore, L1 and L2 in the legend stand for the l_1 and l_2 norm, respectively.
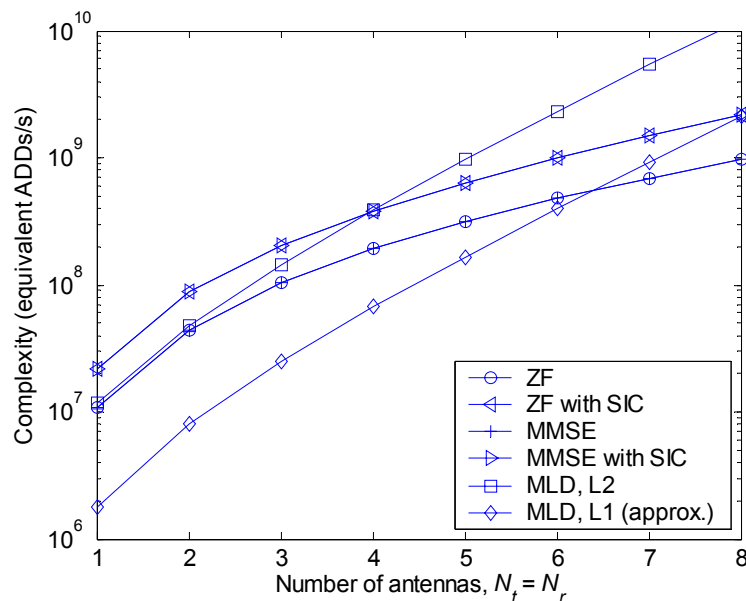
Figure 4-26: Complexity in equivalent additions per second versus $N_t = N_r$ for BPSK, a 64 byte packet, and different SDM algorithms.

From Figure 4-26 it can be seen that the complexity of ZF and MMSE is similar, as well as that of ZF with SIC and MMSE with SIC. It is obvious that the performance improvement of the schemes with SIC is paid by a complexity increase compared to the schemes without SIC. Note, however, that this complexity increase is about a factor of two and therefore in

---

[1] Although we know that the amount of equivalent ADDs/s is an uncommon measure, it can be used to compare the relative complexity of different MIMO systems.

general manageable. Surprisingly, it can be noted that the complexity of MLD, although it grows exponentially with N_t, is reasonable for BPSK and a small number of antennas, while performance wise MLD is also outperforming the other schemes (see Figure 4-15, for instance). For MLD based on the l_2 norm, the complexity is less than or equal to that of the algorithms with SIC up to $N_t = N_r = 4$. For the l_1 norm approximation, this is even the case up to $N_t = N_r = 8$. For higher constellation sizes, however, the complexity of MLD quickly diverges from the complexity of the other techniques (see Figure 4-27).
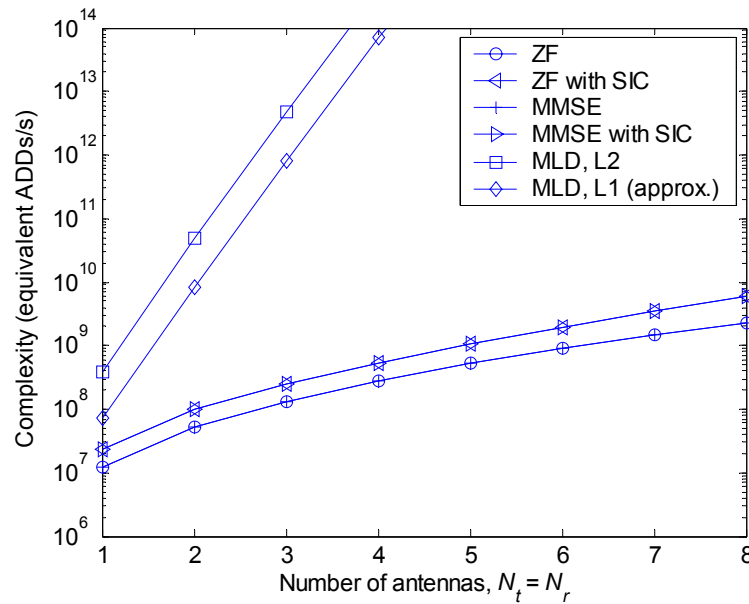
Figure 4-27: Complexity in equivalent additions per second versus $N_t = N_r$ for 64-QAM, a 64 byte packet, and different SDM algorithms.

4.12 Spatial Correlation

In Section 3.6 we have introduced a compact representation of the spatial fading correlation when the capacity is used as performance measure. We also promised in that section to show a similar match for the BER performance. Such a link depends on the chosen MIMO detection method and since we introduced a number of methods in this chapter, we can work out the compact representation of the spatial correlation in terms of BER performance in this section. We will assume the correlation models introduced in (3.68) and (3.69) as starting point. Moreover, we assume that MLD is chosen as detection method.

Since the PEP forms a good basis for the BER we start with that. Recalling the PEP of MLD from (4.114),

$$\Pr(\mathbf{s}_i \to \mathbf{s}_k) \le \det\left(\mathbf{I}_{N_r} + \frac{\sigma_s^2}{4\sigma_n^2}\mathbf{Q}_y\right)^{-1},\qquad(4.128)$$

it can be readily deduced that for a high SNR the PEP can be approximated by

$$\Pr\!\left(\mathbf{s}_i \to \mathbf{s}_k\right) \le \det\!\left(\frac{\rho}{4N_t}\mathbf{Q}_y\right)^{-1}. \tag{4.129}$$

Hence, in the asymptotic case the PEP (and thus the BER performance) depends inversely on the determinant of $\mathbf{Q}_y$. Now the question arises what $\mathbf{Q}_y$ is in scenarios with spatial correlation. To find the answer, we start by rewriting $\mathbf{y}$:

$$\mathbf{y} = \mathbf{H}\!\left(\mathbf{s}'_i - \mathbf{s}'_k\right) = \left(\left(\mathbf{s}'_i - \mathbf{s}'_k\right)^T \otimes \mathbf{I}_{N_r}\right)\mathrm{vec}(\mathbf{H}). \tag{4.130}$$

Now it can be shown that, averaging over $\mathbf{H}$, the covariance matrix of $\mathbf{y}$ equals

$$
\begin{aligned}
\mathbf{Q}_y &= E\!\left[\left(\left(\mathbf{s}'_i - \mathbf{s}'_k\right)^T \otimes \mathbf{I}_{N_r}\right)\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^H\left(\left(\mathbf{s}'_i - \mathbf{s}'_k\right)^T \otimes \mathbf{I}_{N_r}\right)^H\right] \\
&= \left(\left(\mathbf{s}'_i - \mathbf{s}'_k\right)^T \otimes \mathbf{I}_{N_r}\right)\mathbf{R}_H\left(\left(\mathbf{s}'_i - \mathbf{s}'_k\right)^* \otimes \mathbf{I}_{N_r}\right) \\
&= \left(\left(\mathbf{s}'_i - \mathbf{s}'_k\right)^T \otimes \mathbf{I}_{N_r}\right)\left(\mathbf{R}_{TX}^T \otimes \mathbf{R}_{RX}\right)\left(\left(\mathbf{s}'_i - \mathbf{s}'_k\right)^* \otimes \mathbf{I}_{N_r}\right) \\
&= \left(\mathbf{s}'_i - \mathbf{s}'_k\right)^T \mathbf{R}_{TX}^T\left(\mathbf{s}'_i - \mathbf{s}'_k\right)^* \otimes \mathbf{R}_{RX} = \beta\,\mathbf{R}_{RX}.
\end{aligned}
\tag{4.131}
$$

From this result, we can observe that, in order to have the same PEP for the modelled and measured $\mathbf{R}_{RX}$, the determinants of both matrices must be the same. In other words, $\det(\mathbf{R}_{RX,mod}) = \det(\mathbf{R}_{RX,meas})$. And by using (3.69) we can deduce an r_{RX} that achieves an equivalent MLD performance compared to the performance with the measured spatial receiver correlation $\mathbf{R}_{RX,meas}$.

Regarding the spatial correlation at the transmitter, it is obvious that β strongly depends on $(\mathbf{s}_i - \mathbf{s}_k)$. Therefore, to find a link between $\mathbf{R}_{TX,mod}$ and $\mathbf{R}_{TX,meas}$, one has to average over all possible difference vectors $(\mathbf{s}_i - \mathbf{s}_k)$, which is equivalent to using the overall error-rate performance. This overall error-rate performance can be found by averaging over all PEPs by means of, e.g., the union bound. Since the $\mathbf{s}_i$'s are taken from a discrete set that depends on the constellation size, the easiest and most effective way to find a link is through numerical evaluation.

Since we found a (numerical) mathematical link between the MLD error-rate performance for measured and modelled spatial correlation matrices, one example showing the match is sufficient. To that end, the same parameters as in Section 3.6 are used, where the match in terms of the ergodic capacity is already shown. The measured spatial correlation matrices are given by (3.71) and (3.72). Clearly, the matching criterion of the measured and modelled spatial correlation at the receiver side for the MLD error-rate performance is equivalent to (3.65). So to link (3.69) with (3.72), r_{RX} must be set to 0.6172. Furthermore, from numerical evaluation we found that, to link the MLD BER performances, r_{TX} must be set to 0.38. Finally, the match is shown graphically in Figure 4-28 in which a perfect match of the upperbounds can be observed. The slight mismatch between the simulation curves at high SNR can be explained largely by their limited accuracy. The curves are namely obtained by averaging over 100,000 channel realisations, with one 64-byte packet transmission per channel realisation. Hence, the BER is derived by averaging over

51,200,000 bits and since, for a good BER measure, we want to have at least 500 bit errors, the accuracy for BERs low than 10^{-5} is marginal.

Note that in [70] a tighter upperbound is found on MLD with spatial correlation, but from this bound the same results can be deduced, namely, that for a performance match the spatial correlation at the receiving end should obey $\det(\mathbf{R}_{RX,meas}) = \det(\mathbf{R}_{RX,mod})$ and for the transmitter side a numerical evaluation must be performed.
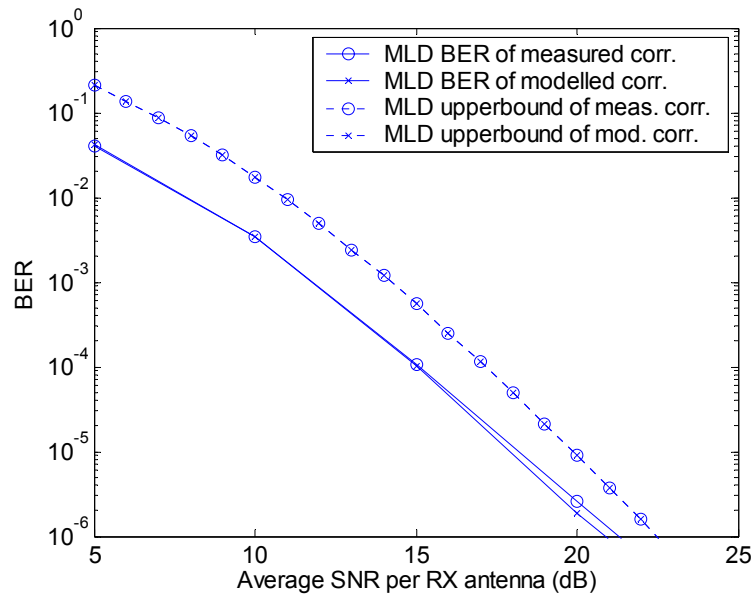
Figure 4-28: MLD BER performance and upperbound versus average SNR per RX antenna for measured and modelled spatial correlation for a 4 × 4 system.

4.13 Turbo SDM

4.13.1 Introduction

In this section, the concept of turbo processing is introduced to coded SDM in order to investigate the potential gains of iterative processing in the context of MIMO. It has been shown in [119, 120] that the performance of demapping a multilevel modulated signal (e.g., like QPSK or 16-QAM) can be improved by using anti-Gray mapping, and iterative demapping and decoding, based on the turbo-decoding principle. In this section, we extend the iterative demapping idea from a single-transmit single-receive wireless communication system to the MIMO case. Since this idea is based on the turbo-coding principle ([16]), it will be called Turbo SDM (also known as Turbo-BLAST, see [138]).

Following [120], the proposed method can be regarded as a serially concatenated iterative decoding scheme in which the inner decoder is replaced by an SDM-demapping device accepting *a priori* information. This leads to almost the same system configuration as described in [120], except that the "regular" (de)mapper is replaced by an SDM (de)mapper as shown in Figure 4-29.

Next, the system configuration of Figure 4-29 will be described in more detail. Starting at the transmitter, the bits from the binary source are encoded using a convolutional code and passed through a pseudo-random interleaver (Π). Then they are mapped onto a MIMO vector by the SDM mapper. In this SDM mapper the interleaved coded bits are demultiplexed, mapped (either using Gray mapping or anti-Gray mapping) and sent in parallel on the N_t TX antennas (see Figure 4-30). This TX vector will be represented by the N_t-dimensional complex vector $\mathbf{s}$.
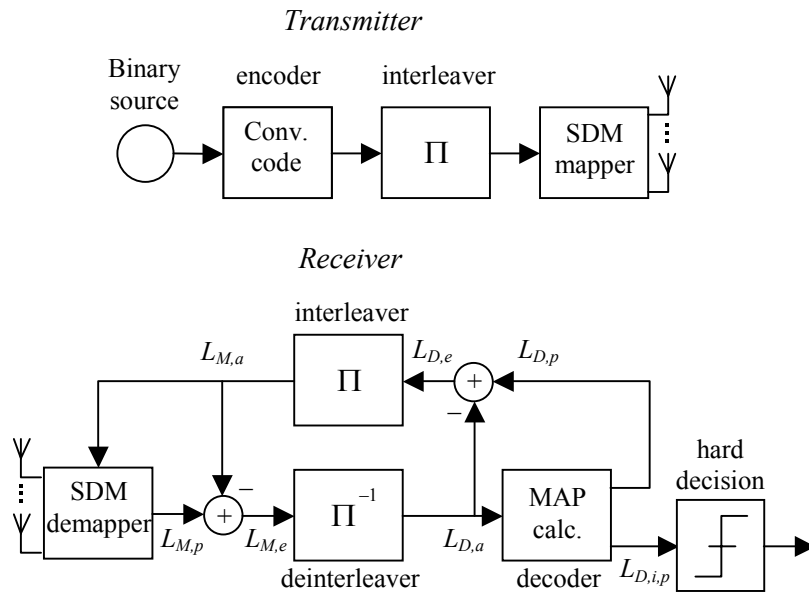
Figure 4-29: Turbo SDM system configuration.
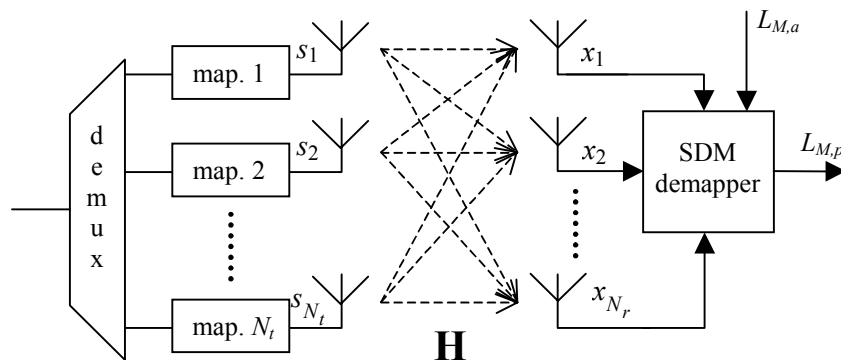
Figure 4-30: The block diagram of an SDM mapper and demapper.

At the receiver, the N_r RX antennas record an N_r-dimensional complex vector $\mathbf{x}$. In the SDM demapper, this vector $\mathbf{x}$ is demapped and soft-decision values are determined by a log-likelihood ratio calculation for all of the transmitted coded bits. After deinterleaving (Π^{-1}) and soft-decision input/soft-decision output decoding with a *Maximum A posteriori Probability* (MAP) decoder ([96]) that contains an implementation of the BCJR algorithm ([13]), the estimates on the transmitted information bits are available at the output of the hard decision block. In the iterative demapping/decoding path, extrinsic information $L_{D,e}$

from the decoder is interleaved and fed back as *a priori* values $L_{M,a}$ to the soft-input soft-output SDM demapper. The extrinsic information at the decoder is the difference of the soft-input and the soft-output LLRs (see [41]) on the coded bits: $L_{D,e} = L_{D,p} - L_{D,a}$. The demapper utilises the extrinsic information from the decoder and calculates improved *a posteriori* values $L_{M,p}$, which are passed as $L_{M,e} = L_{M,p} - L_{M,a}$ to the decoder for further, iterative decoding steps. $L_{M,e}$ is the difference between *a priori* and *a posteriori* LLR values at the demapper and consists of channel information and extrinsic information.

In Subsection 4.13.2, an SDM demapper that accepts *a priori* LLR values and generates *a posteriori* LLR values is derived based on MAP decoding. In Subsection 4.13.3 we evaluate this SDM demapper using the Extrinsic Information Transfer (EXIT) chart methodology which visualises the relation between mutual information at the input of the demapper and the mutual information at its output. Finally, some simulation results are presented in Subsection 4.13.4.

4.13.2 MAP SDM demapper

The SDM demapper that is used has a strong resemblance with SOMLD as described in Subsection 4.10.4, except for the fact that it accepts soft-decision input values. This principle is commonly called Maximum A posteriori Probability (MAP) decoding ([90]). Recalling (4.123), the LLR of the k-th bit of a given transmitted vector to estimate, conditioned on the corresponding received vector $\mathbf{x}$, equals

$$L(b_k|\mathbf{x}) = \ln \frac{P(b_k = 1|\mathbf{x})}{P(b_k = 0|\mathbf{x})} = \ln \frac{\displaystyle\sum_{\mathbf{s}_i|b_k=1} P(\mathbf{s}_i|\mathbf{x})}{\displaystyle\sum_{\mathbf{s}_i|b_k=0} P(\mathbf{s}_i|\mathbf{x})}. \tag{4.132}$$

For example, for a 2×2 SDM system and QPSK modulation, the LLR of e.g. bit b_1 conditioned on the received vector, is given by

$$L(b_1|\mathbf{x}) = \ln \frac{P(b_1 = 1|\mathbf{x})}{P(b_1 = 0|\mathbf{x})}$$

$$= \ln \left(\frac{P(b_1 = 1, b_2 = 0, b_3 = 0, b_4 = 0|\mathbf{x}) +}{P(b_1 = 0, b_2 = 0, b_3 = 0, b_4 = 0|\mathbf{x}) +} \right. \tag{4.133}$$

$$\left. \frac{+ P(b_1 = 1, b_2 = 0, b_3 = 0, b_4 = 1|\mathbf{x}) + \ldots + P(b_1 = 1, b_2 = 1, b_3 = 1, b_4 = 1|\mathbf{x})}{+ P(b_1 = 0, b_2 = 0, b_3 = 0, b_4 = 1|\mathbf{x}) + \ldots + P(b_1 = 0, b_2 = 1, b_3 = 1, b_4 = 1|\mathbf{x})} \right).$$

Due to the bit interleaver that is placed between the encoder and the SDM mapper, the bits that are transmitted can be assumed independent. So, when applying Bayes' rule, the previous equation can be expressed as

$$L(b_1|\mathbf{x}) = \ln \left(\frac{\begin{array}{l} p(\mathbf{x}|b_1=1,b_2=0,b_3=0,b_4=0)\dfrac{P(b_1=1)P(b_2=0)P(b_3=0)P(b_4=0)}{P(\mathbf{x})} + \\[4pt] p(\mathbf{x}|b_1=0,b_2=0,b_3=0,b_4=0)\dfrac{P(b_1=0)P(b_2=0)P(b_3=0)P(b_4=0)}{P(\mathbf{x})} + \\[4pt] +\ldots+ p(\mathbf{x}|b_1=1,b_2=1,b_3=1,b_4=1)\dfrac{P(b_1=1)P(b_2=1)P(b_3=1)P(b_4=1)}{P(\mathbf{x})} \end{array}}{\begin{array}{l} +\ldots+ p(\mathbf{x}|b_1=0,b_2=1,b_3=1,b_4=1)\dfrac{P(b_1=0)P(b_2=1)P(b_3=1)P(b_4=1)}{P(\mathbf{x})} \end{array}} \right) \quad (4.134)$$

$$= L_a(b_1) + \ln \left(\frac{\begin{array}{l} p(\mathbf{x}|b_1=1,b_2=0,b_3=0,b_4=0) + \\ p(\mathbf{x}|b_1=0,b_2=0,b_3=0,b_4=0) + \\ +\ldots+ p(\mathbf{x}|b_1=1,b_2=1,b_3=1,b_4=1)\exp(L_a(b_2)+L_a(b_3)+L_a(b_4)) \end{array}}{+\ldots+ p(\mathbf{x}|b_1=0,b_2=1,b_3=1,b_4=1)\exp(L_a(b_2)+L_a(b_3)+L_a(b_4))} \right).$$

So, by assuming that the *a priori* soft values,

$$L_a(b_k) = \ln \frac{P(b_k=1)}{P(b_k=0)}, \qquad (4.135)$$

are available as input, an SDM demapper accepting *a priori* values has been found. To remove statistically dependent information for further, iterative decoding steps, the additive term $L_a(b_1)$ in   (4.134) is ignored, in order to gain the 'extrinsic' plus channel information ([41]) of the demapping device.

More general, for N_b coded bits, with $N_b = N_t \cdot \log_2 M$, the soft value of the k-th bit can be obtained as follows:

$$L(b_k|\mathbf{x}) = L_a(b_k) + \ln \frac{\displaystyle\sum_{\beta=0}^{2^{N_b-1}-1} p\left(\mathbf{x}|b_k=1,(b_1 \quad \ldots \quad b_{k-1} \quad b_{k+1} \quad \ldots \quad b_{N_b})=\mathbf{b}_\beta\right)\exp(\mathbf{b}_\beta \mathbf{L}_a)}{\displaystyle\sum_{\beta=0}^{2^{N_b-1}-1} p\left(\mathbf{x}|b_k=0,(b_1 \quad \ldots \quad b_{k-1} \quad b_{k+1} \quad \ldots \quad b_{N_b})=\mathbf{b}_\beta\right)\exp(\mathbf{b}_\beta \mathbf{L}_a)}, \quad (4.136)$$

where,

$$\mathbf{b}_\beta \equiv \mathrm{bin}(\beta) \text{ and } \mathbf{L}_a = \begin{pmatrix} L_a(b_1) \\ \vdots \\ L_a(b_{k-1}) \\ L_a(b_{k+1}) \\ \vdots \\ L_a(b_{N_b}) \end{pmatrix}. \qquad (4.137)$$

where bin(β) is a row vector having the values 0 and 1 according to the binary decomposition of β. Finally, following Subsection 4.10.1, for a given channel matrix $\mathbf{H}$, the conditional probability density function can be shown to be

$$p\big(\mathbf{x}|\mathbf{H},\mathbf{s}_i\big)=\det(\pi\mathbf{Q})^{-1}\exp\big(-(\mathbf{x}-\mathbf{H}\mathbf{s}_i)^H\mathbf{Q}^{-1}(\mathbf{x}-\mathbf{H}\mathbf{s}_i)\big). \tag{4.138}$$

With covariance matrix $\mathbf{Q}$ equal to

$$\begin{aligned}\mathbf{Q}&=E\Big[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^H\Big]\\&=E\Big[(\mathbf{x}-\mathbf{H}\mathbf{s}_i)(\mathbf{x}-\mathbf{H}\mathbf{s}_i)^H\Big]=E\big[\mathbf{n}\mathbf{n}^H\big]=\sigma_n^2\mathbf{I}_{N_r}\,,\end{aligned} \tag{4.139}$$

this leads to the soft-decision outputs

$$\begin{aligned}L\big(b_k|\mathbf{x}\big)=&L_a\big(b_k\big)\\&+\ln\frac{\displaystyle\sum_{\beta=0}^{2^{N_b-1}-1}\exp\left(-\frac{\left\|\mathbf{x}-\mathbf{H}\cdot\text{map}\big(\big((\mathbf{b}_\beta)_{1:k-1}\quad 1\quad(\mathbf{b}_\beta)_{k+1:N_b}\big)\big)\right\|^2}{\sigma_n^2}\right)\exp\big(\mathbf{b}_\beta\mathbf{L}_a\big)}{\displaystyle\sum_{\beta=0}^{2^{N_b-1}-1}\exp\left(-\frac{\left\|\mathbf{x}-\mathbf{H}\cdot\text{map}\big(\big((\mathbf{b}_\beta)_{1:k-1}\quad 0\quad(\mathbf{b}_\beta)_{k+1:N_b}\big)\big)\right\|^2}{\sigma_n^2}\right)\exp\big(\mathbf{b}_\beta\mathbf{L}_a\big)},\end{aligned} \tag{4.140}$$

where map(.) denotes the SDM mapping of the corresponding bit vector and results in $\mathbf{s}_i|b_k = 0$ or $\mathbf{s}_i|b_k = 1$, corresponding to the value to which b_k is set. Furthermore, $(\mathbf{b}_\beta)_{p:q}$ denotes the p-th up to and including the q-th element of $(\mathbf{b}_\beta)$.

When applying the max-log approximation, the result equals

$$\begin{aligned}L\big(b_k|\mathbf{x}\big)\approx&L_a\big(b_k\big)\\&+\max_\beta\left(-\frac{\left\|\mathbf{x}-\mathbf{H}\cdot\text{map}\big(\big((\mathbf{b}_\beta)_{1:k-1}\quad 1\quad(\mathbf{b}_\beta)_{k+1:N_b}\big)\big)\right\|^2}{\sigma_n^2}+\mathbf{b}_\beta\mathbf{L}_a\right)\\&-\max_\beta\left(-\frac{\left\|\mathbf{x}-\mathbf{H}\cdot\text{map}\big(\big((\mathbf{b}_\beta)_{1:k-1}\quad 0\quad(\mathbf{b}_\beta)_{k+1:N_b}\big)\big)\right\|^2}{\sigma_n^2}+\mathbf{b}_\beta\mathbf{L}_a\right).\end{aligned} \tag{4.141}$$

4.13.3 EXIT Characteristics of SDM Demapper

To evaluate the performance of Turbo SDM we will use the Extrinsic Information Transfer (EXIT) chart method as described in [119]. In this method, the idea is to predict the iterative decoding behaviour by solely looking at the input/output relations of the

demapper and decoder in terms of *bitwise mutual information*. The bitwise mutual information is defined as the mutual information between the coded bits at the transmitter and the soft-decision values at some stage in the loop of the iterative demapping at the receiver.

The *a posteriori* bitwise mutual information of the SDM demapper $I_{M,e}$ is a function of the *a priori* bitwise mutual information $I_{M,a}$ and the SNR per RX antenna, ρ,

$$I_{M,e} = f\left(I_{M,a}, \rho\right) \tag{4.142}$$

Since the coded bits at the transmitter are discrete, while the soft-decision values in the iterative demapping are continuous, we have to find the bitwise mutual information between a discrete and continuous stochastic variable. The bitwise mutual information between a discrete stochastic variable X with N "states" and a continuous stochastic variable Y can be found in [90] and is given by

$$I(X;Y) = \sum_{i=1}^{N} \int_{-\infty}^{\infty} p\left(y|x_i\right) \Pr(x_i) \log_2 \frac{p\left(y|x_i\right)}{p(y)} dy . \tag{4.143}$$

With equiprobable binary input symbols B to the SDM mapper, the *a priori* bitwise mutual information between the input of the demapper and the coded bits at the transmitter equals ([119])

$$I_{M,a} = \tfrac{1}{2} \sum_{b=0}^{1} \int_{-\infty}^{\infty} p_{M,a}\left(\xi|B=b\right) \log_2 \frac{2 p_{M,a}\left(\xi|B=b\right)}{p_{M,a}\left(\xi|B=0\right) + p_{M,a}\left(\xi|B=1\right)} d\xi , \tag{4.144}$$

where $p_{M,a}(\xi|B = b)$ is the conditional pdf of the LLRs at the input of the SDM demapper. Furthermore, the property $p_{M,a}(\xi) = \tfrac{1}{2} \cdot (p_{M,a}(\xi|B = 0) + p_{M,a}(\xi|B = 1))$ is used, with the assumption that the input bits are equiprobable, i.e., $\Pr(B = b) = \tfrac{1}{2}$. For $N_b = N_t \cdot \log_2 M$ bits per mapped MIMO vector, the *a posteriori* bitwise mutual information is

$$I_{M,e} = \frac{1}{N_b} \sum_{k=0}^{N_b-1} I_{M,e,k} , \tag{4.145}$$

with

$$I_{M,e,k} = \tfrac{1}{2} \sum_{b=0}^{1} \int_{-\infty}^{\infty} p_{M,e,k}\left(\xi|B=b\right) \log_2 \frac{2 p_{M,e,k}\left(\xi|B=b\right)}{p_{M,e,k}\left(\xi|B=0\right) + p_{M,e,k}\left(\xi|B=1\right)} d\xi . \tag{4.146}$$

The conditional pdf $p_{M,a}$ of $L_{M,a}$ is modelled based on the following. In [119] it is stated that the EXIT characteristics prove to be very robust due to the robustness of the entropy measure, because when different distributions $p_{M,a}$ were used to calculate the demapper EXIT characteristics hardly any changes were noticed in the shape of the characteristics. This justifies the idea that the conditioned *a priori* input $L_{M,a}$ can be modelled as an independent Gaussian random variable with a mean of $\mu_{M,a}s$ and a variance of $\sigma_{M,a}^2$, or,

when $n_{M,a}$ is an independent Gaussian random variable with zero mean and variance $\sigma_{M,a}^2$, this yields

$$L_{M,a} = \mu_{M,a}s + n_{M,a} \qquad (4.147)$$

where $s \in \{-1, 1\}$ represents the BPSK modulation of the corresponding transmitted coded bit. Since $L_{M,a}$ is assumed to be an LLR based on a Gaussian distribution, the following must hold:

$$
\begin{aligned}
L_{M,a} &= \ln \frac{p\left(L_{M,a}\middle|s=+1\right)}{p\left(L_{M,a}\middle|s=-1\right)} \\
&= \ln \frac{\dfrac{1}{\sqrt{2\pi}\sigma_{M,a}}\exp\left(-\dfrac{\left(L_{M,a}-\mu_{M,a}\right)^2}{2\sigma_{M,a}^2}\right)}{\dfrac{1}{\sqrt{2\pi}\sigma_{M,a}}\exp\left(-\dfrac{\left(L_{M,a}+\mu_{M,a}\right)^2}{2\sigma_{M,a}^2}\right)} \\
&= -\frac{\left(L_{M,a}-\mu_{M,a}\right)^2}{2\sigma_{M,a}^2} + \frac{\left(L_{M,a}+\mu_{M,a}\right)^2}{2\sigma_{M,a}^2} \\
&= \frac{4L_{M,a}\mu_{M,a}}{2\sigma_{M,a}^2} ,
\end{aligned}
\qquad (4.148)
$$

by which it is shown that $\mu_{M,a}$ must equal $\sigma_{M,a}^2/2$. With the Gaussian approximation of $L_{M,a}$, $I_{M,a}$ can be determined using (4.144). Let $\xi = L_{M,a}$, then it can be shown that ([118])

$$I_{M,a}\left(\sigma_{M,a}\right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_{M,a}} \exp\left(-\frac{\left(\xi - \sigma_{M,a}^2/2\right)^2}{2\sigma_{M,a}^2}\right)\cdot\left(1 - \ln\left(1 + \exp\left(-\xi\right)\right)\right)d\xi. \quad (4.149)$$

$I_{M,a}(\sigma_{M,a})$ goes to zero when $\sigma_{M,a}$ goes to zero and $I_{M,a}(\sigma_{M,a})$ goes to one when $\sigma_{M,a}$ goes to infinity. $I_{M,a}(\sigma_{M,a})$ cannot be expressed in closes-form, but it can be shown to be monotonically increasing ([118]) and, thus, it is invertible. Hence, for a given $I_{M,a}$ we can find a corresponding $\sigma_{M,a}$ and use this to model the *a priori* soft-decision inputs for the demapper. Now, the conditional pdf $p_{M,e,k}$ of $L_{M,e,k}$ can be obtained through simulations based on a large set of channel realisations, noise realisations, and $L_{M,a}$'s. Consequently, with (4.145) we can find $I_{M,e}$ as a function of $I_{M,a}$ or $\sigma_{M,a}^2$. This relation between the bitwise mutual information of the input values and that of the output values of the demapper results in the EXIT characteristics of the SDM demapper.

Iterative demapping and decoding can be applied to any multilevel/multiphase modulation scheme such as *M*-PAM, *M*-PSK or *M*-QAM ([120]). In this section, we will only focus on QPSK, either with Gray mapping or anti-Gray mapping.

The EXIT characteristics of the SDM demapper for different mappings, SNRs per RX antenna, and antenna configurations for AWGN and flat-fading channels are depicted in

Figure 4-31 and Figure 4-32, respectively. Note that for $N_t \times N_r = 2 \times 2$ and AWGN, we used the channel matrix (see Subsection 3.5.4)

$$\mathbf{H} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{4.150}$$
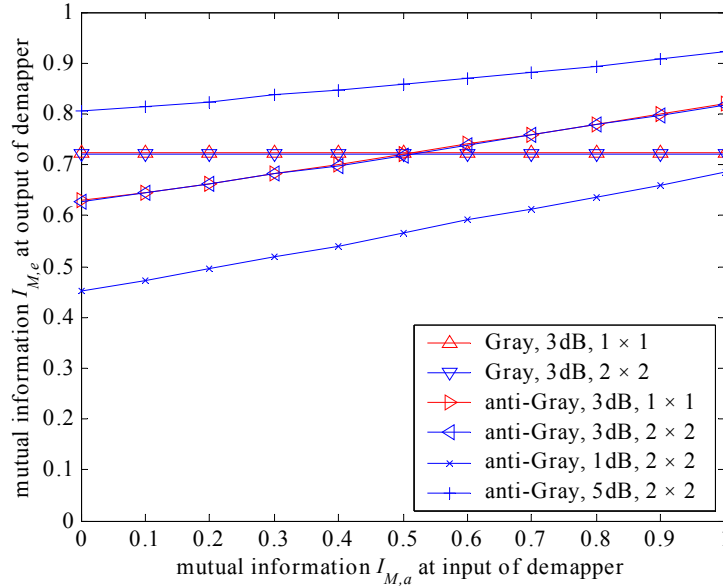
Figure 4-31: EXIT characteristics of QPSK SDM demapper operating in AWGN for different mappings, SNRs per RX antenna, and antenna configurations ($N_t \times N_r$).
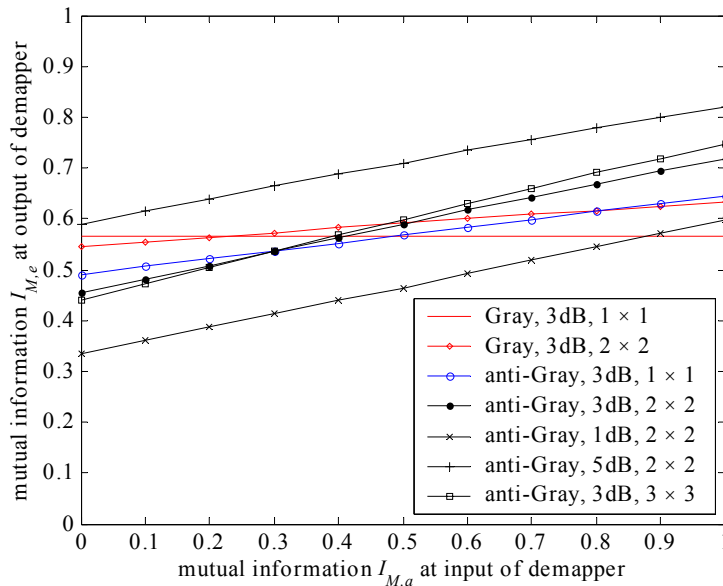
Figure 4-32: EXIT characteristics of QPSK SDM demapper in flat Rayleigh fading with a different realisation per MIMO vector, for different mappings, SNRs per RX antenna, and antenna configurations ($N_t \times N_r$).

A number of things can be observed from the demapper transfer characteristics. First, note that the characteristics are almost straight lines. Second, keeping the mapping and antenna configuration $N_t \times N_r$ fixed, different SNR values just shift the curve up and down (note that for SNR values of 5 dB and larger, also the slope is affected; see the 5 dB curve in Figure 4-31). Third, keeping the SNR fixed, different mappings and/or antenna configurations result in lines of different slope. Fourth, for AWGN, for $N_t \times N_r$ is 1×1 and 2×2, the results are the same, which seems logical when choosing the **H** of (4.150) and dividing the power equally among the two TX antennas in the latter case. Fifth, even for Gray mapping, MIMO in the Rayleigh flat-fading case results in a non-horizontal curve, which can be explained by the fact that the instantaneous MIMO channels are not completely orthogonal but provide some spatial "memory". Finally, the curves in Figure 4-31 and Figure 4-32 with non-horizontal slopes unveil the big potential performance improvements in an iterative demapping and decoding scheme compared to the configurations resulting in horizontal transfer characteristics. Namely, for the horizontal lines, the *a priori* information is of no means, whereas for the curves with a slope larger than zero, the better the *a priori* information, the better the resulting *a posteriori* information. The latter property shows that applying turbo processing is beneficial.

4.13.4 Simulation Results

In the first set of simulations, we will use a half rate, memory 4, non-systematic convolutional code with generator polynomial $(G_1, G_2) = (23, 35)$ (octal notation) as outer code, to be able to compare our results with those of [120]. Furthermore, like in [120], the channel is assumed to change for every successive MIMO vector, i.e., "fast time fading". Following the analysis of (4.144)-(4.146), the decoder transfer characteristic on the coded bits is defined as $I_{D,e} = f(I_{D,a})$. To obtain the transfer characteristic of the decoder, it is assumed that the pdf of soft-decision inputs, $L_{D,a}$, can be modelled by a Gaussian distribution, like for the SDM demapper (cf. (4.147)).

Since the demapper and decoder are only connected by the interleavers, the extrinsic output of the demapper becomes the *a priori* input of the decoder $I_{D,a} = I_{M,e}$ and vice versa $I_{M,a} = I_{D,e}$. This exchange of extrinsic information can be visualised in a graph called EXIT-chart. When the pseudo-random interleaver size is set to 1000 coded bits, for Rayleigh flat-fading, QPSK, anti-Gray mapping, an average SNR per RX antenna of $\rho = 3$ dB and $N_t \times N_r = 2 \times 2$, this results in the EXIT-chart shown in Figure 4-33.

The BER contour plots of Figure 4-33 can be obtained using the same approach as [118]. Assume that not only the input of the MAP decoder, but also the output of the MAP decoder, $L_{D,p}$ can be modelled by a Gaussian distribution with variance $\sigma_{D,p}^2$ and mean $\mu_{D,p} = \sigma_{D,p}^2/2$. Furthermore, assume that if there still is an error in the soft-decision outputs of the coded bits, there will also be an error in the corresponding information bits, i.e., the sliced values of $L_{D,i,p}$ (see Figure 4-29). Then, the bit error probability is approximated by

$$P_b \approx Q\left(\frac{\mu_{D,p}}{\sigma_{D,p}}\right) = Q\left(\frac{\sigma_{D,p}}{2}\right). \tag{4.151}$$

Since $L_{D,p} = L_{D,a} + L_{D,e}$ and assuming independency between the *a priori* information and the extrinsic information, we may write

$$\sigma_{D,p}^2 = \sigma_{D,a}^2 + \sigma_{D,e}^2. \tag{4.152}$$

Now, when we fill in a certain P_b and use the inverse of the Q function, we can find the corresponding $\sigma_{D,p}$. With above equation, using the inverse relations of $I_{D,a}(\sigma_{D,a})$, $\sigma_{D,e}$ is obtained for a given P_b and $I_{D,a}$, and based on the analysis of Subsection 4.13.3, $I_{D,e}$ can be found as function of $\sigma_{D,e}$, $I_{D,e}(\sigma_{D,e})$. This results in the contour plots of Figure 4-33.
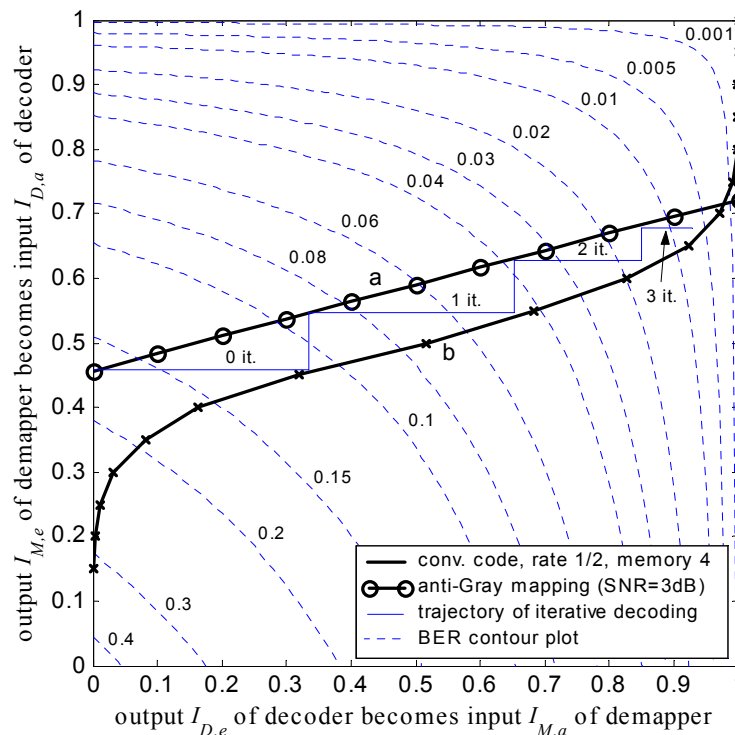
Figure 4-33: Simulated EXIT characteristics and real trajectory of iterative decoding for QPSK, anti-Gray mapping, SNR per RX antenna of 3dB, rate ½ memory 4 code, and $N_t \times N_r = 2 \times 2$.

From Figure 4-33 a number of things can be seen. First, following the trajectory for an SNR per RX antenna of $\rho = 3$dB, the Turbo SDM system appears to converge after 4 or 5 iterations. Second, the Bit Error Rate (BER) floor at this SNR is approximately equal to 0.001 as can be found at the intersection of curve 'a' and 'b', which can be read off from the BER contour plots. Third, when the SNR goes down, curve 'a' is also shifting downwards. This narrows the tunnel between curve 'a' and 'b'. At a certain SNR, the tunnel is blocked. At this SNR, the turbo cliff starts in the Bit Error Rate (BER) curve. For the example of Figure 4-33, this is around 1.5 dB (which can be verified with Figure 4-34). Fourth, the real trajectory after 3 iterations ends in a BER of around 0.01, which also can be verified with Figure 4-34. Fifth, note that, because of its horizontal EXIT chart, it does not make sense to perform any iterations for the 1×1 case with Gray mapping. Sixth, the MIMO configurations perform better than their SISO counterpart, while achieving a two times higher data rate. Seventh, for a high SNR and one or more iterations, the anti-Gray

mappings outperform the Gray mappings. Only for zero iterations anti-Gray mapping performs worse than Gray mapping which can be explained by the fact that the anti-Gray constellation contains less mutual information when no *a priori* knowledge is available (see Figure 4-31 and Figure 4-32). Finally, according to Figure 4-32 the slope of curve 'a' changes when changing $N_t \times N_r$. From this observation some design guidelines become apparent, e.g., the steeper the slope of the SDM demapper curve, the later the turbo cliff in the BER versus SNR chart, but the lower the BER floor. Moreover, in [119] it is shown that the EXIT chart of the MAP decoder for convolutional codes with larger memories becomes less steep, from which the same reasoning follows. Note that the BER results of [120] are shown versus the E_b/N_0 of the coded bits. According to (4.34), for QPSK this results in a 3 dB shift of the performance.
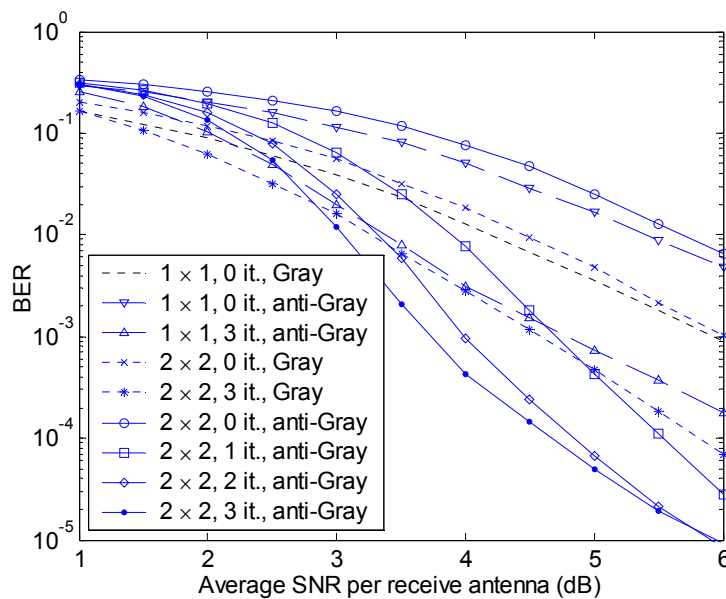
Figure 4-34: BER performance of iterative decoding in a flat Rayleigh fading channel with a different realisation per MIMO vector, for QPSK, rate ½ memory 4 code, different mappings and antenna configurations.

In the next set of simulations, we will compare the performance of turbo SDM with the results of Subsection 4.11.2. For the turbo scheme, the same encoder as for SDM with joint encoding is used, namely, a rate ½, memory 3 convolutional encoder with generator polynomial (15, 17). Furthermore, the channel is assumed *quasi static*, i.e., constant throughout the duration of a packet, and the packet size is set to 64 bytes.

When random interleaving is applied with an interleaver depth of 64 bytes, the results for a 2×2 and a 2×4 system are shown in Figure 4-35 and Figure 4-36, respectively. Clearly, turbo decoding with 3 iterations and anti-Gray mapping at the transmitter gives the best PER performance. For the 2×4 case, it is even only 2.5 dB away from the outage PER. Note, however, that its BER performance is not always the best, especially not for the 2×2 case. Again, the most logical explanation is that for anti-Gray mapping the bit errors are more bursty, resulting in a higher BER, but lower PER than its counterpart with Gray mapping. Furthermore, Gray-mapped Turbo SDM without iterations shows a better

performance than SOMLD with the same convolutional outer code. This can be explained by the fact that a MAP decoder in general performs better than a Viterbi decoder ([96]).
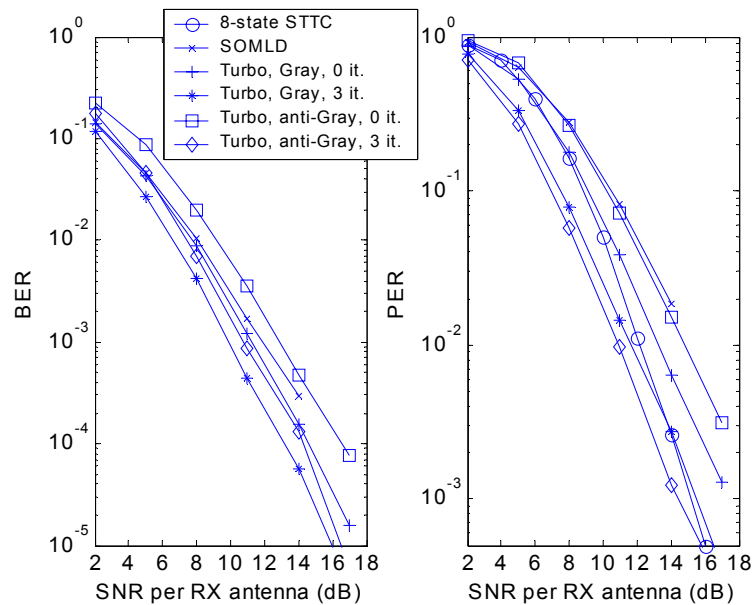
Figure 4-35: BER and PER versus the average SNR per RX antenna for a 2 × 2 system in quasi-static flat Rayleigh fading with an efficiency of 2 bits/s/Hz and as coding scheme: 8-state STTC, SOMLD with an 8-state convolutional outer code, and Turbo SDM with different mappings and a different number of iterations.
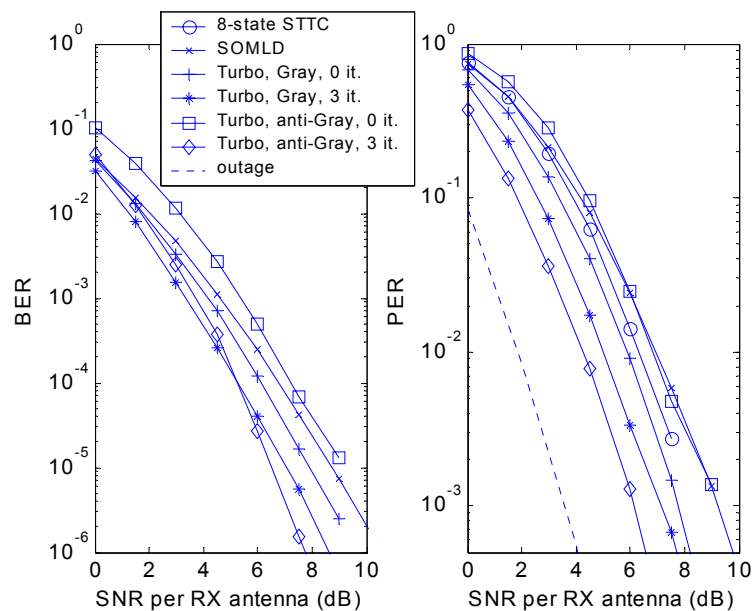
Figure 4-36: BER and PER versus the average SNR per RX antenna for a 2 × 4 system in flat Rayleigh fading with an efficiency of 2 bits/s/Hz and as coding scheme: 8-state STTC, SOMLD with an 8-state convolutional outer code, and Turbo SDM with different mappings and a different number of iterations.

4.14 Conclusions

In this chapter, we introduced a unified view on MIMO techniques like space-time block coding, space-time trellis coding, and Space Division Multiplexing (SDM). The general transmitter consists of an encoder, a space-time mapper, and per transmit antenna a modulation block. When the space-time mapper is simply a demultiplexer, we get the powerful concept of SDM with outer coding. In literature it was namely shown that, when the number of antennas and diversity potential of the channel are large enough, the probability of error appears to depend only on the Euclidean distance of the code. This would indicate that a one-dimensional code designed for AWGN channels, of which its codewords are properly interleaved across the space and time, may be as effective as a space-time code that follows the rank and determinant criteria introduced in [116]. In [148] it was shown that this would already be the case when the diversity order is equal to or larger than four, which would restrict the domain of interest of space-time coding to architectures with only 2–3 antennas. In Subsection 4.11.2, we also showed that SDM with a certain outer code outperforms Space-Time Coding (STC) with comparable coding strength, for a high enough diversity. The performance can be improved even further by iterating over the spatial processing and outer code of coded SDM, i.e., by Turbo SDM (see Section 4.13), at the cost of latency.

Moreover, the concept of SDM with outer coding is more flexible than STC, since a transmit antenna can easily be added or dropped, and puncturing can be applied easily. These are good properties for defining sub-schemes that are able to fallback in rate or enhance robustness. These definitions are generally captured in modes referred to as *fallback modes*. For STC, however, a different encoder and decoder have to be implemented for every rate.

More concretely, the following SDM algorithms were described in this chapter: Zero Forcing (ZF), Minimum Mean Squared Error (MMSE), ZF with Successive Interference Cancellation (SIC), MMSE with SIC, and Maximum Likelihood Detection (MLD). Their performance and complexity were compared and it turned out that MLD is the best performing algorithm. Its complexity, however, grows exponentially with the number of transmit antennas (N_t). The complexity of the least complex scheme, ZF, grows polynomial with N_t. For ZF, the required processing during the payload is the highest and proportional to N_t^3. It should be noted, though, that (hardly) no attempt is made to optimise the algorithms (for special hardware architectures), so, the results of the complexity comparison presented in this chapter are not binding. What we already can conclude from the presented results, however, is that the implementation of MIMO schemes (with a low number of antennas) is feasible.

All above points provide a strong indication of the potential of SDM with outer coding and, therefore, we will apply this concept in the next chapter to OFDM.

5

MIMO OFDM

5.1 Introduction

In general, the MIMO techniques described in the previous chapter provide an interesting basis for next generations of wireless communication systems. One of the potential application areas is that of Wireless Local Area Networks (WLANs). The current WLAN standards IEEE 802.11a ([57]) and IEEE 802.11g ([56]) are based on Orthogonal Frequency Division Multiplexing (OFDM) ([127]). A high-data-rate extension of these standards could be based on SDM ([137]). This leads to the promising combination of the data rate enhancement of SDM with the relatively high spectral efficiency and the robustness against frequency-selective fading and narrowband interference of OFDM. An advantage of wireless LAN systems is that they are mainly deployed in indoor environments. These environments are typically characterised by richly scattered multipath. As explained in [36], this is a good condition for having a high MIMO capacity.

In this chapter, first the principle of OFDM is explained (Section 5.2). Then, the combination of MIMO and OFDM is described in Section 5.3. Section 5.4 introduces a signal model for MIMO OFDM which shows that the MIMO OFDM processing transfers the wideband frequency-selective MIMO channel into a number of parallel flat-fading MIMO subchannels. In the next section the frequency-selective MIMO capacity is determined and the corresponding outage PER is defined.

In Section 5.6, a theoretical Space-Frequency analysis is performed based on the Pairwise Error Probability (PEP) to better understand the achievable performance and required design rules. Then, Section 5.7 introduces practical coding schemes in which the SDM algorithms of the previous chapter can be combined with OFDM. Two basic transmitter architectures are described, namely Joint Coding (JC) and Per-Antenna-Coding (PAC). These schemes are programmed in MATLAB and simulation results for different SDM algorithms, different antenna configurations, and a number of rms delay spreads are shown

in Section 5.8. Based on the analyses and simulation results of this chapter, Section 5.9 presents conclusions and recommendations.

5.2   Orthogonal Frequency Division Multiplexing (OFDM)

5.2.1   Background

In classical data systems in which more data rate was sought by exploiting the frequency domain, parallel transmissions were achieved by dividing the total signal frequency band into N_c *non-overlapping* frequency subchannels. This technique is referred to as Frequency Division Multiplexing (FDM). In this technique, each subchannel or subcarrier is modulated with a separate symbol and then the N_c subchannels are frequency multiplexed. Spectral overlap is avoided by putting enough guard space between adjacent subchannels. In this way, Inter Carrier Interference (ICI) is eliminated. This method, however, leads to a very inefficient use of the available spectrum. A more efficient use of bandwidth can be obtained with parallel transmissions if the spectra of the individual subchannels are permitted to partly overlap. This requires that specific orthogonality constraints are imposed to facilitate separation of the subchannels at the receiver.

Figure 5-1 shows the general structure of a multi-carrier system ([109]). The data stream $s(i)$ is converted to parallel data streams, which are modulated onto separate subchannels. The resulting signals are summed and transmitted. At the receiver, the different subchannels are down converted to parallel baseband signals, demodulated, and then concatenated to a serial data stream.
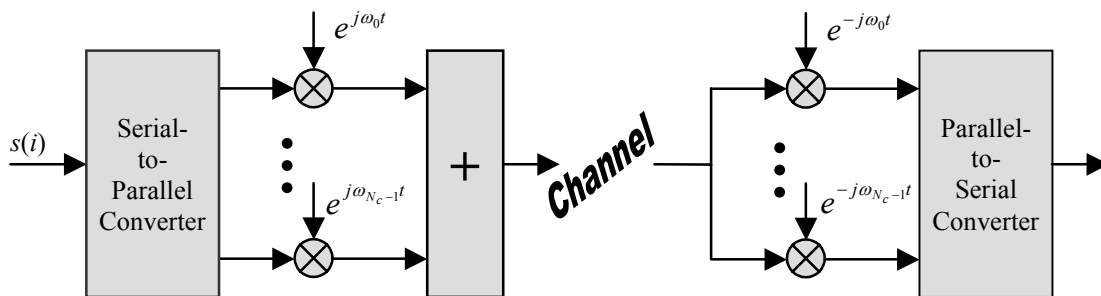
Figure 5-1: Basic structure of a multi-carrier system.

Orthogonal Frequency Division Multiplexing (OFDM) is an example of a multi-carrier technique that operates with specific orthogonality constraints between the subcarriers. Due to these constraints, it achieves a very high spectral efficiency. For adequate transmission quality, however, it is important to preserve the subcarrier orthogonality that is inherent to the OFDM concept. Although the OFDM principle has been around for many years [141], only the current technology level makes satisfactory implementation feasible. As a result, more and more systems that operate in the Gigahertz bands are based on OFDM, such as wireless LANs ([30] and [57]), Digital Video Broadcasting (DVB) ([29]), and Digital Audio Broadcasting (DAB) ([26]). In the next subsection the basics of OFDM are explained.

5.2.2 Principle

In OFDM, the subcarrier pulse used for transmission is chosen to be rectangular. This has the advantage that the task of pulse forming and modulation can be performed by a simple Inverse Discrete Fourier Transform (IDFT) which results in a remarkable reduction in equipment complexity (filters, modulators, etc.), as is explained in [23] and [141]. Note that when the number of subcarriers is a power of two, the (Inverse) Discrete Fourier Transform ((I)DFT) can be implemented very efficiently by means of the (Inverse) Fast Fourier Transform ((I)FFT). A schematic representation of the equivalent complex-valued baseband core of an OFDM system is depicted in Figure 5-2. Most of the symbols in this figure have already been introduced in Subsection 3.2.2. The others are explained in this subsection. The purpose of the Cyclic Prefix (CP) is explained in Subsection 5.2.3.

Let $s(i)$ be a sequence of discrete-time, QAM-modulated symbols. Furthermore, consider an OFDM system that utilizes N_c subcarriers. In communications, in general, these N_c subcarriers are distributed equally around the 0th subcarrier. When only N_u out of the N_c subcarriers are used, commonly the OFDM symbol is padded with zeros in such a way that the zeros result in extra guard space between adjacent OFDM bands. Furthermore, to avoid the influence of DC offset on the DC subcarrier (i.e., the 0th subcarrier), it is often set to zero. Note that sampling in the time domain results in periodicity in the frequency domain. Hence, the negative subcarriers might as well be seen as positive subcarriers by increasing their index with N_c. In this case, the $N_c \times 1$ OFDM symbol, $\mathbf{s'}$, has the following form (when omitting the OFDM symbol index)

$$\mathbf{s'} = \begin{pmatrix} 0 & s(0) & \cdots & s(\lceil \tfrac{1}{2} N_u \rceil - 1) & 0 & \cdots & 0 & s(\lceil \tfrac{1}{2} N_u \rceil) & \cdots & s(N_u - 1) \end{pmatrix}^T, \quad (5.1)$$

where $\lceil x \rceil$ denotes the nearest integer value that is equal to or higher than x. When N_u out of N_c subcarriers are used, the IDFT output is given by

$$z(n) = \frac{1}{\sqrt{N_u}} \sum_{i=0}^{N_c - 1} s'(i) \exp\left(j2\pi \frac{in}{N_c} \right), \text{ with } i = 0, \ldots, N_c - 1, \text{ and } n = 0, \ldots, N_c - 1, \quad (5.2)$$

where $s'(i)$ denotes the i-th element of $\mathbf{s'}$. Note that the uncommon scaling factor $1/\sqrt{N_u}$ is chosen such that the average power of the $z(n)$'s equals that of the $s(i)$'s, i.e., $\sigma_z^2 = \sigma_s^2$:

$$
\begin{aligned}
\sigma_z^2 &= E\left[|z(n)|^2 \right] \\
&= E\left[\left| \frac{1}{\sqrt{N_u}} \sum_{i=0}^{N_c-1} s'(i) \exp\left(j2\pi \frac{in}{N_c} \right) \right|^2 \right] \\
&= \frac{1}{N_u} \sum_{i=0}^{N_c-1} E\left[\left| s'(i) \exp\left(j2\pi \frac{in}{N_c} \right) \right|^2 \right] = \sigma_s^2 ,
\end{aligned}
\quad (5.3)
$$

where the assumption is made that the elements of $\mathbf{s'}$ are independent.
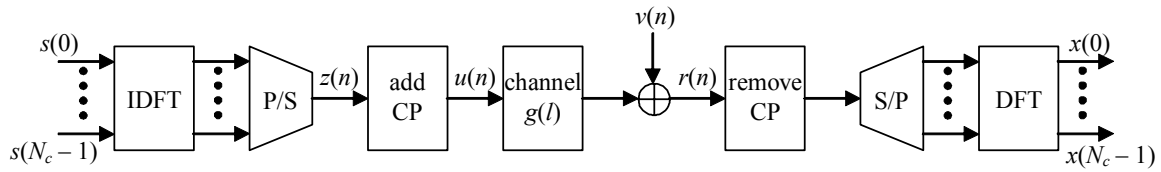
Figure 5-2: Schematic representation of an equivalent baseband OFDM system.

To show the orthogonality principle of OFDM, we next examine the frequency spectrum of a continuous-time equivalent of $z(n)$. Based on the fact that sampling in the time domain results in periodicity in the frequency domain, it is convenient to write the continuous-time equivalent of $z(n)$ for the signalling interval $0 \le t < N_c T_s$ as

$$z(t) = \frac{1}{\sqrt{N_u}} \sum_{i=-\left\lceil \frac{1}{2} N_c \right\rceil}^{\left\lceil \frac{1}{2} N_c \right\rceil - 1} s'(i \bmod N_c) \exp(j 2\pi f_i t) \mathrm{rect}\left(\frac{t}{N_c T_s} - \frac{1}{2} \right), \qquad (5.4)$$

where

$$\mathrm{rect}(x - \tfrac{1}{2}) = \begin{cases} 1, & \text{for } 0 \le x < 1, \\ 0, & \text{for } x < 0 \text{ and } x \ge 1, \end{cases} \qquad (5.5)$$

$f_i = i\Delta f$ denotes the frequency of the i-th subcarrier, $\Delta f = 1/(N_c T_s) = f_s/N_c$ is the subcarrier spacing, T_s is the sampling time, $f_s = 1/T_s$ is the sampling rate, and $x \bmod N$ denotes the modulo reduction of x by N. Note the equivalence of $z(t)$ with the general multi-carrier signals that can be deduced from Figure 5-1. The frequency spectrum of $z(t)$ is obtained by performing the Fourier transformation on $z(t)$:

$$
\begin{aligned}
z^F(f) = F\{z(t)\} &= \int_{-\infty}^{\infty} z(t) \exp(-j 2\pi f t) dt \\
&= \int_0^{N_c T_s} \frac{1}{\sqrt{N_u}} \sum_{i=-\left\lceil \frac{1}{2} N_c \right\rceil}^{\left\lceil \frac{1}{2} N_c \right\rceil - 1} s'(i \bmod N_c) \exp(-j 2\pi (f - f_i) t) dt \qquad (5.6) \\
&= C_z \sum_{i=-\left\lceil \frac{1}{2} N_c \right\rceil}^{\left\lceil \frac{1}{2} N_c \right\rceil - 1} s'(i \bmod N_c) e^{-j\pi (f - f_i) N_c T_s} \mathrm{sinc}\left(\pi (f - f_i) N_c T_s \right).
\end{aligned}
$$

In this result, C_z is some power related constant. Clearly, the act of truncating the continuous-time signal to the interval $[0, N_c T_s\rangle$ in the time domain imposes a $\mathrm{sinc}(x)$ function on each subchannel in the frequency domain with zeros at multiples of $1/(N_c T_s)$, which leads to the orthogonality principle of OFDM. As example, suppose the number of subcarriers is 16 ($N_c = 16$) of which only 10 subcarriers are used ($N_u = 10$) and suppose that for a given OFDM symbol all used subcarriers are modulated by the same QAM symbol, then the normalised squared absolute frequency responses of the individual subcarriers of this OFDM symbol are given by the dashed lines of Figure 5-3. The total squared absolute frequency response is proportional to the solid line in Figure 5-3.

At the receiver, the transmitted signals are recovered using the Discrete Fourier Transform. Based on the resulting signals, it can be observed that the orthogonality principle leads to a number of disadvantages of OFDM. OFDM is, e.g., vulnerable for carrier frequency offset and phase noise. When the local oscillators at the transmitter and receiver are not aligned, a frequency offset occurs and as a result, the frequency spectrum is not "sampled" at the peaks of the sinc-functions, but to the left or right of the peaks. At these points, the other subcarriers are not zero, and Inter-Carrier Interference (ICI) takes place. Likewise, the OFDM signal is vulnerable for non-linearities, like non-linear amplifiers, IQ imbalance, DC offset, etc. Moreover, due to multipath distortion, recovering the transmitted signals brings some extra problems along as will be explained in the next subsection.
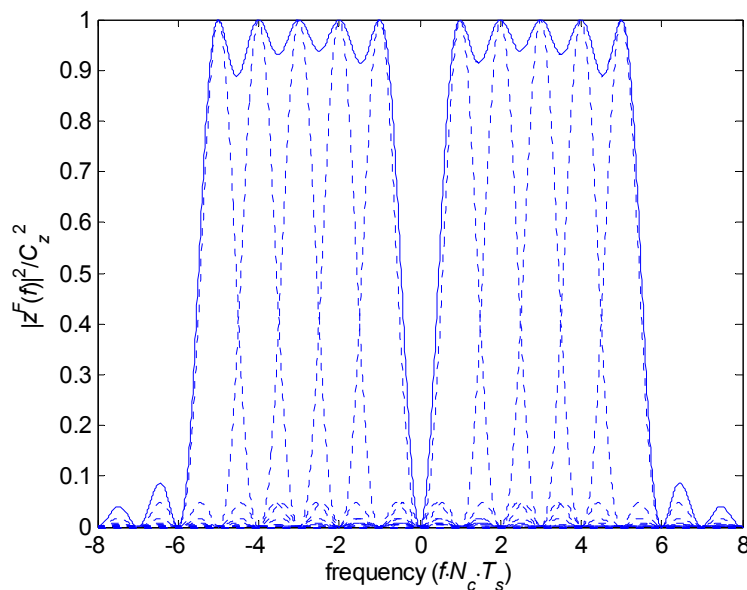
Figure 5-3: Normalised squared absolute version of the frequency response of an OFDM symbol with $N_c = 16$ and $N_u = 10$.

5.2.3 Multipath Distortion

The reason that the information transmitted on the subcarriers can be separated at the receiver is the orthogonality relation giving OFDM its name. By using an IDFT for modulation, the spacing of the subcarriers is implicitly chosen in such a way that, at the frequencies where the received signals are evaluated (at the peaks of the sinc-functions in Figure 5-3), all other signals are zero. In order for this orthogonality to be preserved, the following must be true ([109]):

1. The receiver and the transmitter must be accurately synchronised. This means they both must have exactly the same modulation frequency and the same time-scale for transmission.
2. The analogue components, part of transmitter and receiver, must be of very high quality.
3. There should be no multipath channel.

Unfortunately, multipath distortion is (almost) unavoidable in radio communication systems and, thus, the received signal is affected. It is shown in [141] that, although the truncated subchannel sinusoids are delayed by different amounts (i.e. channel delays), the distortion is mainly concentrated at the on-off transmissions of these waveforms. Hence, a guard space (in frequency), and a guard time by means of a Cyclic Prefix (CP) or cyclic extension (see Figure 5-5), chosen longer than the maximal delay spread, will eliminate most interference among channels (i.e., Inter Carrier Interference (ICI)) and between adjacent transmission blocks (i.e., Inter Symbol Interference (ISI)). The CP, in general, is chosen equal to the last part of the OFDM symbol and, therefore, often is referred to as cyclic extension.

To avoid out of band radiation, the on-off transitions must be smoothened. This can be implemented by, e.g., windowing each OFDM symbol by a raised cosine window ([126]). Figure 5-4 depicts schematically the implementation of the windowing in an OFDM symbol.

$$T_{\text{tot}} = T_G + T_{FFT}$$

Time $\longrightarrow$
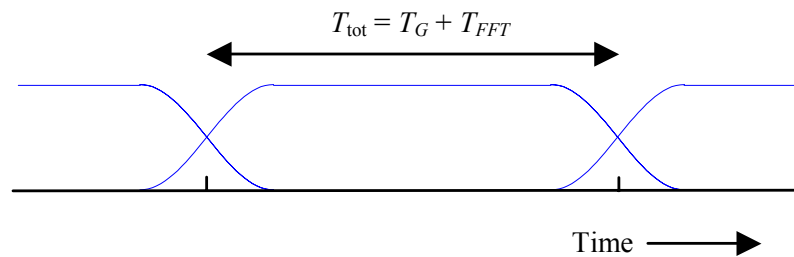
Figure 5-4: The principle of windowing.

The validity of windowing each OFDM symbol by a raised cosine window can be explained by looking to the shapes of the subcarrier sinusoids after the multipath-fading channel. An OFDM receiver uses only a part of this signal to calculate the FFT. This part should be chosen such that in this FFT interval with a length of T_{FFT} seconds, which at the Nyquist rate equals N_cT_s seconds, every subcarrier has an integer number of cycles, which ensures orthogonality. In the multipath-fading channel, the receiver-input signal will be a sum of delayed and scaled replicas of the transmitted subcarriers ([127], Figure 5-5). Note that a sum of scaled and delayed sinusoids is again a sinusoid. So, as long as the Guard Interval (GI) time T_G minus half the windowing ramp-up and role-off time is larger than the maximal channel delay we can choose the window for applying the FFT such that there will still be an integer number of cycles within this FFT interval for each multipath component keeping the reflections of previous symbols out and preserving the orthogonality.

So, thanks to the guard interval and windowing, the wideband multipath fading is experienced in OFDM as a set of narrowband fading subcarriers without ICI and ISI. The only remaining effect of multipath is a random phase and amplitude per subcarrier. This effect can be minimised by correcting the subcarriers of the received signal with reference amplitudes and phases, i.e., channel estimates, per subcarrier, which can be obtained during a training phase. The subcarriers in deep fades still are a problem but in order to deal with these weak subcarriers that have a large probability to be detected erroneously, forward error correction across the subcarriers can be applied.

5.2.4 Main Advantages of OFDM

One significant advantage to use OFDM is its low complexity compared to an equivalent single carrier system designed to work with the same amount of delay spread. This is due to the fact that single-carrier systems need complex equalizers, whereas OFDM allows for efficient (I)FFT processing. In, e.g., [126], it is shown that the complexity of a 64-point FFT compared to single-carrier equalisation of a channel with an rms delay spread of 250 ns is 10 times less.

The other advantage of OFDM over single-carrier systems with equalizers is that for the latter systems, the performance degrades abruptly if the delay spread exceeds the value for which the equalizers are designed. Because of error propagation, the raw bit error probability increases so quickly that introducing lower rate coding or a lower constellation size does not significantly improve the delay-spread robustness. For OFDM, however, such non-linear effects as error propagation do not occur, and coding and lower constellation sizes can be employed to provide fall-back rates that are significantly more robust against delay spread. This is an important consideration, as it enhances the coverage area and avoids the situation that users in bad spots cannot get any connection at all.
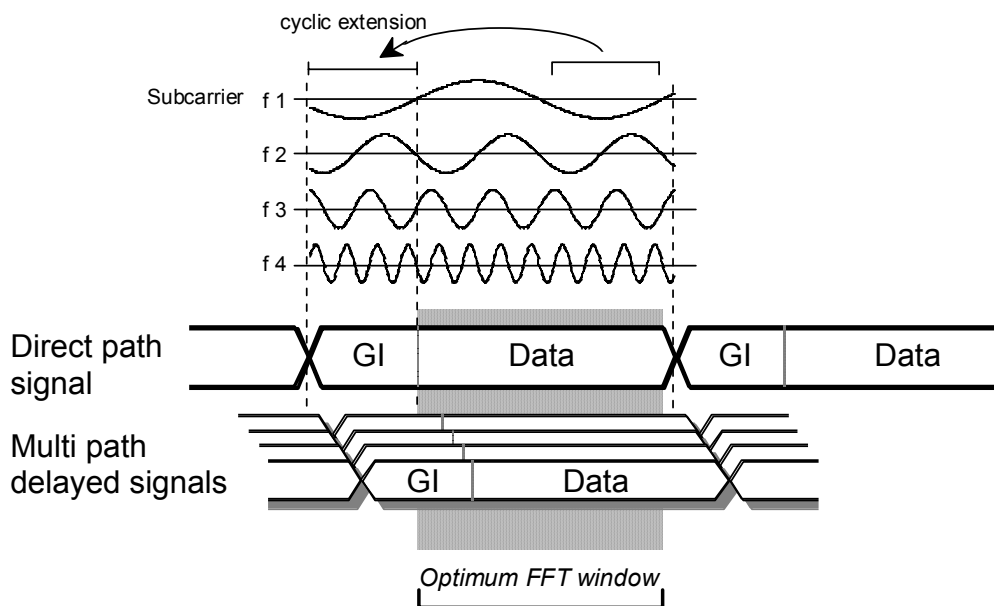
Figure 5-5: OFDM symbol with cyclic prefix.

5.2.5 OFDM Transceiver

A general block diagram of the common baseband processing of an OFDM transceiver is shown in Figure 5-6 ([127]). On the binary input data, the transmitter performs encoding, interleaving, QAM mapping, N_c-point IFFT, and adds a cyclic extension before the final TX signal is windowed, transferred to the analogue domain by the Digital-to-Analogue Converter (DAC) and converted up to the Radio Frequency (RF) and transmitted. Note that, in order to get an output spectrum with a relative low out-of-band radiation, the size

of the IFFT can be chosen larger than the number of subcarriers that is actually used for transmission. In addition, windowing provides (in general) an extra fall-off of the output spectrum.

For reliable detection, it is, in general, necessary that the receiver knows the wireless communication channel and keeps track of phase and amplitude drifts. To enable estimation of the wireless communication channel, the transmitter occasionally sends known training symbols. In Wireless Local Area Networks (WLANs) a preamble, which includes channel training sequences, is added to every packet. Moreover, to track the phase drift, pilot symbols are inserted to every OFDM data symbol on predefined subcarriers.
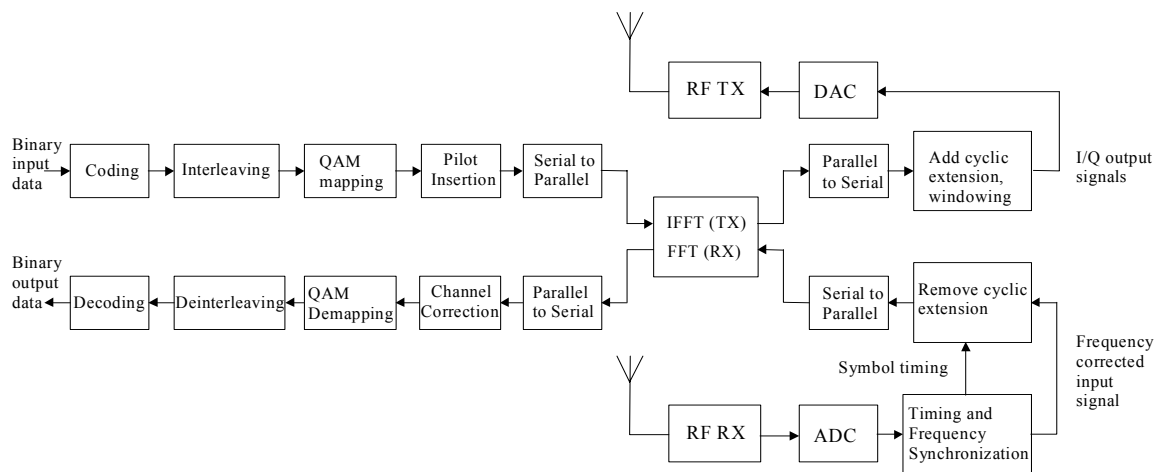
Figure 5-6: Block diagram of an OFDM transceiver.

The OFDM receiver, basically, performs the reverse operation of the transmitter, together with additional training tasks. The received RF signal is converted down to baseband and subsequently converted from the analogue to the digital domain by the Analogue-to-Digital Converter (ADC). Then, the receiver must estimate and correct for the frequency offset and the symbol timing, e.g., by using the training sequence in the preamble. Next, the CP is removed and the N_c-point FFT is performed. The resulting signal is corrected for the channel influences, and finally demapping, deinterleaving and decoding are performed to obtain the binary output data.

5.3   MIMO OFDM

The MIMO algorithms as described in Chapter 4 are narrowband algorithms. In order to deal with the frequency-selective nature of wideband wireless channels, MIMO can be combined with OFDM. Effectively, OFDM transforms a frequency-selective channel into parallel flat-fading subchannels, i.e., the signals on the subcarriers undergo narrowband fading. Hence, by performing MIMO transmission and detection per subcarrier, MIMO algorithms can be applied in broadband communication ([2, 93, 137]).

Consider a MIMO OFDM system with N_t transmit (TX) and N_r receive (RX) antennas. In addition to the spatial and temporal dimension of MIMO, OFDM adds one extra dimension

to exploit, namely, the frequency dimension. In line with the unified view on MIMO (see Section 4.2), a MIMO OFDM transmitter can be envisioned as presented in Figure 5-7. In general, the incoming bit stream is first encoded by a one-dimensional encoder after which the encoded bits are mapped onto the three available dimensions by the Space-Time-Frequency (STF) mapper. After the STF mapper, each TX branch consists of almost an entire OFDM transmitter (see Subsection 5.2.5). Note that the symbols of Figure 5-7 are explained in Section 5.4.
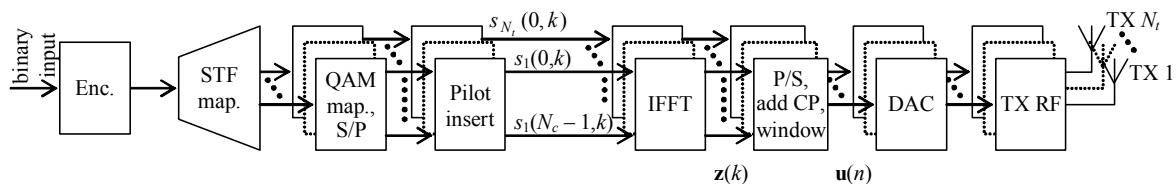
Figure 5-7: Schematic representation of a MIMO OFDM transmitter.

A schematic representation of a MIMO OFDM receiver is given in Figure 5-8. After a digital representation of the N_r received signals is obtained by the ADCs, the receiver first must estimate and correct for the frequency offset and retrieve the symbol timing, e.g., by making use of training sequences. Note that it is convenient for the remaining processing to have all receiver branches jointly synchronised and, therefore, the synchronisation task should not be performed in parallel per branch, but jointly. Furthermore, for proper frequency synchronisation of the multiple branches it is beneficial to have all branches at one end of the communication link connected to the same local oscillator in a homodyne structure, or to the same local oscillators providing the multiple frequency levels in a heterodyne structure. This speaks in favour of Space Division Multiplexing (SDM) as opposed to Space Division Multiple Access (SDMA), since for SDMA with dislocated single-branch transmitters, each having their own local oscillator(s), frequency synchronisation is cumbersome ([121]).
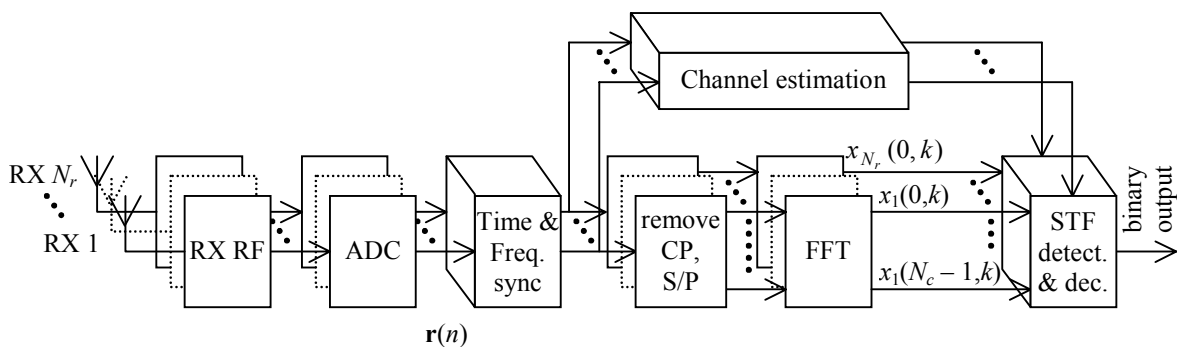
Figure 5-8: Schematic representation of a MIMO OFDM receiver.

After synchronisation, the CP is removed and the N_c-point FFT is done per receiver branch. In the context of the unified view, at this point, overall STF detection and decoding must be performed to recover the binary data stream. In general, however, because the MIMO algorithms are single carrier algorithms, MIMO detection is performed per OFDM subcarrier. To that end, the received signals of subcarrier i are routed to the i-th MIMO detector to recover the N_t QAM symbols transmitted on that subcarrier. Next, the symbols per TX stream are combined and, finally, STF demapping/deinterleaving and decoding are

performed on these N_t parallel streams and the resulting data are combined to obtain the binary output data. For reliable detection, it is typically necessary that the receiver knows the wireless communication channel and keeps track of phase and amplitude drifts. To enable estimation of the wireless communication channel, the transmitter occasionally sends known training symbols. In WLANs a preamble, which includes channel-training sequences, is added to every data packet. Moreover, to track the phase drift, pilot symbols are inserted into every MIMO OFDM data symbol on predefined subcarriers.

Finally, note that OFDM has as advantage that it introduces a certain amount of parallelism by means of its N_c subcarriers. This fact can be exploited by MIMO OFDM. Namely, if MIMO detection is performed per subcarrier, then a given detector is allowed to work N_c times slower than the MIMO detector of an equivalent single carrier system with comparable data rate. Although in the case of MIMO OFDM N_c of such detectors are required, they can work in parallel, which might ease the implementation.

5.4   *The Multi-Carrier MIMO Signal Model*

In this section, a signal model is introduced for a MIMO OFDM system in which the relation between the transmitted and received MIMO OFDM symbols is captured in matrix form. With this concise matrix notation we mathematically show that the signal model per subcarrier equals the narrowband signal model introduced in Section 4.3. The strength of this matrix signal model is that it allows for mathematical evaluations of MIMO OFDM systems, including the outage performance and space-frequency analysis of this chapter. Furthermore, it can be used for impairment studies such as timing offset and frequency offset analyses (see [102, 103, 135]), phase noise analysis ([104]), etc.

Consider a communication system with N_t transmit (TX) and N_r receive (RX) branches where, at a sampling interval n, the transmitter sends an N_t-dimensional complex vector $\mathbf{u}(n)$ and the receiver records an N_r-dimensional complex vector $\mathbf{r}(n)$. Furthermore, assume that the system is operating in a frequency-selective Rayleigh fading environment and that the channel remains constant during a packet transmission, i.e., quasi-static fading. Suppose that the channel impulse response can be recorded with L consecutive samples. Then, following Section 3.3, the fading channel between the p-th TX and q-th RX antenna can be modelled by a discrete-time baseband equivalent $(L-1)$-th order finite impulse response (FIR) filter with filter taps $g_{qp}(l)$, with $l = 0, \ldots, L-1$. Suppose $g_{qp}(l)$ represents the (q,p)-th element of $\mathbf{G}(l)$, with $p = 1, \ldots, N_t$ and $q = 1, \ldots, N_r$, then, recalling (3.36), the discrete-time MIMO baseband signal model is given by

$$\mathbf{r}(n) = \sum_{l=0}^{L-1} \sqrt{P(l)} \, \mathbf{G}(l) \mathbf{u}(n-l) + \mathbf{v}(n), \qquad (5.7)$$

where $\mathbf{v}(n)$ represents Additive White Gaussian Noise (AWGN) at the n-th sample with N_r independent and identically distributed (i.i.d.) zero-mean, complex Gaussian elements with variance σ_v^2. The elements of $\mathbf{u}(n)$ are assumed zero mean, uncorrelated random variables with variance σ_u^2. So, the expected SNR per receive antenna, say the q-th receive antenna, can be shown to be

$$\rho = \frac{E\left[\left(\sum_{l=0}^{L-1}\sqrt{P(l)}\,\mathbf{u}^H(n-l)\mathbf{g}^{q^H}(l)\right)\left(\sum_{l=0}^{L-1}\sqrt{P(l)}\,\mathbf{g}^{q^H}(l)\mathbf{u}^H(n-l)\right)\right]}{E\left[v_q^*(n)v_q(n)\right]}$$

$$= \frac{1}{\sigma_v^2}E\left[\sum_{l=0}^{L-1}\sum_{p=1}^{N_t}P(l)\left|g_{qp}(l)\right|^2\left|u_p(n-l)\right|^2\right] = \frac{N_t\sigma_u^2}{\sigma_v^2}, \qquad (5.8)$$

where it is assumed that $\mathbf{g}^q(l)$ and $\mathbf{g}^q(l')$ for $l \neq l'$ are independent, and u_p is the p-th element of $\mathbf{u}$, v_q is the q-th element of $\mathbf{v}$, and $\mathbf{g}^q(l)$ represents the q-th row of $\mathbf{G}(l)$. Clearly, the SNR per receive antenna equals the total transmitted power (normalised by the path loss) divided by the noise power per receive antenna.

To deal with the frequency selectivity of the channel, we apply OFDM utilizing a maximum of N_c subcarriers per antenna transmission. To combat Inter Symbol Interference (ISI), a guard interval of N_g samples is added per OFDM symbol. When sampled at the Nyquist rate, an OFDM symbol, including the guard interval, consists of $N_{tot} = N_c + N_g$ complex samples. Based on this, n can be written as $n = k \cdot N_{tot} + n'$, where k represents the k-th OFDM symbol and n', with $n' = 0, \ldots, N_{tot} - 1$, the sample number within that symbol.

OFDM can be introduced to the signal model of (5.7) as follows. Denote the collection of QAM symbols to be sent on the k-th MIMO OFDM symbol by

$$\mathbf{s}'(k) = \begin{pmatrix} \mathbf{s}(0,k) \\ \vdots \\ \mathbf{s}(N_c-1,k) \end{pmatrix}, \qquad (5.9)$$

where $\mathbf{s}(i,k)$ denotes the $N_t \times 1$ MIMO vector that is transmitted on the i-th subcarrier of the k-th MIMO OFDM symbol. The dimensions of $\mathbf{s}'(k)$ are $N_c N_t \times 1$.

First, the IDFT is applied at the transmitter (see Figure 5-7). This transforms the frequency domain vector $\mathbf{s}'(k)$ into the time domain. Without loss of generality, we will assume for now that all subcarriers are used. Using (5.2), with $N_u = N_c$, and denoting the result by $\mathbf{z}'(k)$, yields

$$\mathbf{z}'(k) = \sqrt{N_c}\left(\mathbf{F}^{-1} \otimes \mathbf{I}_{N_t}\right)\mathbf{s}'(k), \qquad (5.10)$$

where $\otimes$ represents the Kronecker product and $\mathbf{F}^{-1}$ equals the $N_c \times N_c$ inverse Fourier matrix, defined by ([113])

$$\mathbf{F}^{-1} = \frac{1}{N_c}\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W^{-1} & W^{-2} & \cdots & W^{-(N_c-1)} \\ 1 & W^{-2} & W^{-4} & \cdots & W^{-2(N_c-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & W^{-(N_c-1)} & W^{-(N_c-1)2} & \cdots & W^{-(N_c-1)^2} \end{pmatrix}, \qquad (5.11)$$

with $W = \exp(-j2\pi/N_c)$.

Second, the cyclic prefix is added. This is done by taking the last $N_g N_t$ elements of $\mathbf{z}'(k)$ and stacking them on top of $\mathbf{z}'(k)$ to produce the vector $\mathbf{u}'(k)$. In matrix notation this can be written as

$$\mathbf{u}'(k) = \mathbf{A}_1 \mathbf{z}'(k) = \left(\begin{pmatrix} \mathbf{0} & \mathbf{I}_{N_g} \\ \mathbf{I}_{N_c} & \end{pmatrix} \otimes \mathbf{I}_{N_t} \right) \mathbf{z}'(k), \qquad (5.12)$$

where, in this case, $\mathbf{0}$ is an $N_g \times (N_c - N_g)$ dimensional all-zeros matrix.

Finally, the signal is transmitted over the frequency-selective communication channel as defined in (5.7). In case ISI occurs, we assume that the only significant interference comes from the preceding MIMO OFDM symbol, $\mathbf{u}'(k - 1)$. With this assumption, the convolution in (5.7) can be represented as

$$\mathbf{r}'(k) = \mathbf{G}'\mathbf{u}'(k) + \mathbf{G}_{\text{ISI}}\mathbf{u}'(k-1) + \mathbf{v}'(k), \qquad (5.13)$$

where the size of both $\mathbf{G}'$ and $\mathbf{G}_{\text{ISI}}$ equals $N_r(N_c + N_g) \times N_t(N_c + N_g)$, and $\mathbf{G}'$ is defined by

$$\mathbf{G}' = \begin{pmatrix} \mathbf{G}_0 & & & & & \\ \vdots & \ddots & & & \mathbf{0} & \\ \mathbf{G}_{L-1} & \cdots & \mathbf{G}_0 & & & \\ & \ddots & & \ddots & & \\ & & \ddots & & \ddots & \\ \mathbf{0} & & & \mathbf{G}_{L-1} & \cdots & \mathbf{G}_0 \end{pmatrix}, \qquad (5.14)$$

where $\mathbf{G}_l$ represents

$$\mathbf{G}_l = \sqrt{P(l)}\mathbf{G}(l) , \qquad (5.15)$$

while $\mathbf{G}_{\text{ISI}}$ is a matrix with all elements equal to zero, except for the $N_r(L{-}1) \times N_t(L{-}1)$ elements of the upper-right corner, they are defined by the upper-triangular block matrix

$$\begin{pmatrix} \sqrt{P(L-1)}\mathbf{G}(L-1) & \cdots & \sqrt{P(1)}\mathbf{G}(1) \\ & \ddots & \vdots \\ \mathbf{0} & & \sqrt{P(L-1)}\mathbf{G}(L-1) \end{pmatrix}. \qquad (5.16)$$

Furthermore, $\mathbf{u}'(k)$ and $\mathbf{v}'(k)$ are, respectively, defined by

$$\mathbf{u}'(k) = \begin{pmatrix} \mathbf{u}(kT_{\text{tot}} + 0) \\ \vdots \\ \mathbf{u}(kT_{\text{tot}} + T_{\text{tot}} - 1) \end{pmatrix} \text{ and } \mathbf{v}'(k) = \begin{pmatrix} \mathbf{v}(kT_{\text{tot}} + 0) \\ \vdots \\ \mathbf{v}(kT_{\text{tot}} + T_{\text{tot}} - 1) \end{pmatrix}. \qquad (5.17)$$

Note that, when ISI from more than one MIMO OFDM symbol has to be taken into account, matrix notation becomes impractical and one can better use the discrete convolution of (5.7).

At the receiver, first, the cyclic prefix is removed. This is done by discarding the first $N_g N_r$ samples of $\mathbf{r'}(k)$, and, second, the DFT is performed. Together, this results in

$$\mathbf{x'}(k) = \frac{1}{\sqrt{N_c}}\left(\mathbf{F} \otimes \mathbf{I}_{N_r}\right)\mathbf{A}_2\mathbf{r'}(k) = \frac{1}{\sqrt{N_c}}\left(\mathbf{F} \otimes \mathbf{I}_{N_r}\right)\left(\left(\mathbf{0}_{N_c \times N_g} \quad \mathbf{I}_{N_c}\right) \otimes \mathbf{I}_{N_r}\right)\mathbf{r'}(k), \quad (5.18)$$

where $\mathbf{F}$ denotes the $N_c \times N_c$ Fourier matrix of which element (a,b) equals $W^{(a-1)(b-1)}$ ([113]), and $\mathbf{x'}(k)$ is defined as

$$\mathbf{x'}(k) = \begin{pmatrix} \mathbf{x}(0,k) \\ \vdots \\ \mathbf{x}(N_c-1,k) \end{pmatrix}, \quad (5.19)$$

where $\mathbf{x}(i,k)$ denotes the $N_r \times 1$ received MIMO vector on the i-th subcarrier of the k-th MIMO OFDM symbol.

Combining all above steps and assuming that no ISI occurs on a MIMO OFDM symbol basis (i.e., $L \leq N_g + 1$), this leads to the following relation between $\mathbf{s'}(k)$ and $\mathbf{x'}(k)$:

$$\mathbf{x'}(k) = \left(\mathbf{F} \otimes \mathbf{I}_{N_r}\right)\mathbf{A}_2\mathbf{G'}\mathbf{A}_1\left(\mathbf{F}^{-1} \otimes \mathbf{I}_{N_t}\right)\mathbf{s'}(k) + \frac{1}{\sqrt{N_c}}\left(\mathbf{F} \otimes \mathbf{I}_{N_r}\right)\mathbf{A}_2\mathbf{v'}(k). \quad (5.20)$$

This result may be rewritten as

$$\mathbf{x'}(k) = \left(\mathbf{F} \otimes \mathbf{I}_{N_r}\right)\mathbf{C}\left(\mathbf{F}^{-1} \otimes \mathbf{I}_{N_t}\right)\mathbf{s'}(k) + \mathbf{n'}(k), \quad (5.21)$$

where $\mathbf{C}$ is an $N_r N_c \times N_t N_c$ block circulant matrix given by

$$\mathbf{C} = \begin{pmatrix} \mathbf{G}_0 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{G}_{L-1} & \cdots & \mathbf{G}_1 \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & & \ddots & \mathbf{G}_{L-1} \\ \mathbf{G}_{L-1} & & & \ddots & \ddots & & \mathbf{0} \\ & \ddots & & & \ddots & \ddots & \vdots \\ & & \ddots & & & \ddots & \mathbf{0} \\ \mathbf{0} & & & \mathbf{G}_{L-1} & \cdots & \cdots & \mathbf{G}_0 \end{pmatrix}, \quad (5.22)$$

and $\mathbf{n'}(k)$ represents the frequency domain noise, equal to

$$\mathbf{n'}(k) = \frac{1}{\sqrt{N_c}}\left(\mathbf{F} \otimes \mathbf{I}_{N_r}\right)\mathbf{A}_2\mathbf{v'}(k). \quad (5.23)$$

C is a special kind of Toeplitz block matrix where each block of columns is obtained by doing a block wrap-around downshift of the previous "block vector". For the remainder of this dissertation, we will define the first block of columns by

$$
\mathbf{K} = \begin{pmatrix} \sqrt{P(0)}\mathbf{G}(0) \\ \vdots \\ \sqrt{P(L-1)}\mathbf{G}(L-1) \\ \mathbf{0}_{N_r(N_c-L)\times N_t} \end{pmatrix}.
\tag{5.24}
$$

As shown in Appendix A.1.11, the block circulant matrix **C** can be block-diagonalised by pre-multiplying it by $(\mathbf{F} \otimes \mathbf{I}_A)$ and post-multiplying it by $(\mathbf{F}^{-1} \otimes \mathbf{I}_B)$ under the assumption that the dimensions of the blocks in **C** are $A \times B$. The resulting block-diagonal matrix contains the Fourier transform of **K** on its block diagonal. So, finally, (5.21) can be written as

$$
\mathbf{x}'(k) = \begin{pmatrix} \mathbf{H}(0) & & 0 \\ & \ddots & \\ 0 & & \mathbf{H}(N_c-1) \end{pmatrix} \mathbf{s}'(k) + \mathbf{n}'(k) = \mathbf{H}'\mathbf{s}'(k) + \mathbf{n}'(k),
\tag{5.25}
$$

where (see Appendix A.1.11)

$$
\mathbf{H}(i) = \sum_{l=0}^{L-1} \sqrt{P(l)}\mathbf{G}(l)e^{-j2\pi\frac{il}{N_c}}.
\tag{5.26}
$$

So, for the i-th subcarrier we may write

$$
\mathbf{x}(i,k) = \mathbf{H}(i)\mathbf{s}(i,k) + \mathbf{n}(i,k),
\tag{5.27}
$$

which results in a flat-fading signal model (as defined in Section 4.3) per subcarrier. Assume that the elements of $\mathbf{s}(i,k)$ are zero mean, uncorrelated random variables with variance σ_s^2 and that $\mathbf{n}(i,k)$ represents Additive White Gaussian Noise (AWGN) with N_r independent and identically distributed (i.i.d.) zero-mean, complex Gaussian elements with variance σ_n^2. Then, the expected SNR per receive antenna, say the q-th receive antenna, for this equation is given by

$$
\begin{aligned}
\rho &= \frac{E\left[\mathbf{s}^H(i,k)\mathbf{h}^{q^H}(i)\mathbf{h}^q(i)\mathbf{s}(i,k)\right]}{E\left[n_q^*(i,k)n_q(i,k)\right]} \\
&= \frac{1}{\sigma_n^2}E\left[\mathbf{s}^H(i,k)\left(\sum_{l=0}^{L-1}\sqrt{P(l)}\mathbf{g}^{q^H}(l)e^{j2\pi\frac{il}{N_c}}\right)\left(\sum_{l=0}^{L-1}\sqrt{P(l)}\mathbf{g}^q(l)e^{-j2\pi\frac{il}{N_c}}\right)\mathbf{s}(i,k)\right] \\
&= \frac{1}{\sigma_n^2}E\left[\sum_{l=0}^{L-1}\sum_{p=1}^{N_t}P(l)\left|g_{qp}(l)\right|^2\left|s_p(i,k)\right|^2\right] = \frac{N_t\sigma_s^2}{\sigma_n^2},
\end{aligned}
\tag{5.28}
$$

where s_p is the p-th element of $\mathbf{s}$, n_q is the q-th element of $\mathbf{n}$, and $\mathbf{h}^q$ represents the q-th row of $\mathbf{H}$. Since we also know that $\mathbf{x}'(k)$ equals

$$\mathbf{x}'(k) = \frac{1}{\sqrt{N_c}}\left(\mathbf{F} \otimes \mathbf{I}_{N_r}\right)\mathbf{A}_2\mathbf{G}'\mathbf{u}'(k) + \frac{1}{\sqrt{N_c}}\left(\mathbf{F} \otimes \mathbf{I}_{N_r}\right)\mathbf{A}_2\mathbf{v}'(k),\qquad(5.29)$$

it is easily verified that the SNR definition of (5.28) equals $N_t\sigma_u^2/\sigma_v^2$, which is equivalent to the SNR definition of (5.8). Note that, when only N_u out of N_c subcarriers are used for transmission, the SNR per receive antenna per signal-bearing subcarrier (frequency-domain) is higher than SNR per receive antenna per sample (time-domain). This is because the signal power in the frequency domain is concentrated on the N_u subcarriers, whereas the noise power is equally spread among all N_c tones (see Subsection 5.8.2). When we average over all N_c subcarriers (i.e., signal bearing and zero subcarriers), this again would (of course) give the same SNR in the frequency domain as in the time domain.

5.5 Capacity

5.5.1 Definition of the Capacity of Wideband Channels

In this section we define the capacity of a MIMO OFDM system, using the signal model introduced in the previous section. The open-loop capacity of a frequency-selective wideband channel, defined in bits/s/Hz, can be obtained by dividing the frequency band in a number of narrowband channels and averaging over the capacity of these narrowband channels. The open-loop capacity of a narrowband channel is defined in Section 4.4.5. For an infinite amount of narrowband subchannels this results in

$$C = \frac{1}{B}\int_B \log_2 \det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t}\mathbf{H}(f)\mathbf{H}^H(f)\right)df \;\text{ bits/s/Hz },\qquad(5.30)$$

where $\mathbf{H}(f)$ represents the frequency response of the MIMO channel and B the system bandwidth. When the frequency band is divided in a discrete number of frequency-flat subchannels, say N_c, then the open-loop capacity is given by

$$C = \frac{1}{N_c}\sum_{i=0}^{N_c-1}\log_2 \det\left(\mathbf{I}_{N_r} + \frac{\rho}{N_t}\mathbf{H}(i)\mathbf{H}^H(i)\right) \;\text{ bits/s/Hz },\qquad(5.31)$$

where $\mathbf{H}(i)$ denotes the channel of the i-th subchannel.

5.5.2 Outage Packet Error Rate Performance

In line with the narrowband definition of outage PER performance provided in Subsection 4.4.6, we can also define the outage PER for a wideband channel. In an open-loop system, there will always be a finite probability that a given frequency-selective channel does not

support the bit rate that is chosen by the transmitter, simply because the theoretical capacity of that channel is lower. This results in a packet error. By averaging over different channel realisations, the outage PER for a given bit rate is obtained.

5.6 *Space-Frequency Analysis*

In this section, a Space-Frequency (SF) analysis is performed to provide theoretical guidelines for proper SF code design. In the analysis, we omit the MIMO OFDM symbol index k for brevity. Let $\mathbf{c}'$ and $\mathbf{e}'$ be two possible space-frequency codewords, where $\mathbf{c}' = (\mathbf{c}^T(0) \ldots \mathbf{c}^T(N_c - 1))^T$ and $\mathbf{e}' = (\mathbf{e}^T(0) \ldots \mathbf{e}^T(N_c - 1))^T$, and $\mathbf{c}(i)$ and $\mathbf{e}(i)$ are the $N_t \times 1$ parts of the two possible codewords for the i-th subcarrier, and let $\mathbf{H}'$ be the representation of the SF channel as introduced in Section 5.4. Then the PEP between any pair of codewords, for a given SF channel $\mathbf{H}'$, can be shown to be (by applying the principle of Figure 4-13)

$$\Pr(\mathbf{c}' \to \mathbf{e}' | \mathbf{H}') = Q\left(\sqrt{ \frac{\sigma_s^2}{2\sigma_n^2} \sum_{i=0}^{N_c-1} \| \mathbf{H}(i)(\mathbf{c}(i) - \mathbf{e}(i)) \|^2 } \right), \tag{5.32}$$

or, using the Chernoff upperbound,

$$\Pr(\mathbf{c}' \to \mathbf{e}' | \mathbf{H}') \le \exp\left(-\frac{\sigma_s^2}{4\sigma_n^2} \sum_{i=0}^{N_c-1} \| \mathbf{H}(i)(\mathbf{c}(i) - \mathbf{e}(i)) \|^2 \right). \tag{5.33}$$

Define $\mathbf{y}$ as

$$\mathbf{y} = \mathbf{H}'(\mathbf{c}' - \mathbf{e}') = \begin{pmatrix} \mathbf{H}(0) & & 0 \\ & \ddots & \\ 0 & & \mathbf{H}(N_c - 1) \end{pmatrix} \begin{pmatrix} \mathbf{c}(0) - \mathbf{e}(0) \\ \vdots \\ \mathbf{c}(N_c - 1) - \mathbf{e}(N_c - 1) \end{pmatrix}. \tag{5.34}$$

Following the analysis of Subsection 4.10.2 and averaging over all channel realisations leads to the PEP upperbound:

$$\Pr(\mathbf{c}' \to \mathbf{e}') \le \det\left(\mathbf{I}_{N_r N_c} + \frac{\sigma_s^2}{4\sigma_n^2} \mathbf{Q}_y \right)^{-1}, \tag{5.35}$$

where $\mathbf{Q}_y$ denotes the covariance matrix of the above defined $\mathbf{y}$. With the channel definition of (5.26), with $\mathbf{C} = (\mathbf{c}(0) \ldots \mathbf{c}(N_c - 1))$ and $\mathbf{E} = (\mathbf{e}(0) \ldots \mathbf{e}(N_c - 1))$, and with the Kronecker identity (A.12) of Appendix A.1.10, it can be shown that $\mathbf{y}$ equals

$$\mathbf{y} = \left(\sum_{l=0}^{L-1} \mathbf{D}^l \otimes \sqrt{P(l)} \mathbf{G}(l) \right) \mathrm{vec}(\mathbf{C} - \mathbf{E}) = \mathrm{vec}\left(\sum_{l=0}^{L-1} \sqrt{P(l)} \mathbf{G}(l)(\mathbf{C} - \mathbf{E})\mathbf{D}^l \right)$$

$$= \mathrm{vec}\left(\mathbf{I}_{N_r} \sum_{l=0}^{L-1} \sqrt{P(l)} \mathbf{G}(l)(\mathbf{C} - \mathbf{E})\mathbf{D}^l \right) = \sum_{l=0}^{L-1} \left(\left(\mathbf{D}^l (\mathbf{C} - \mathbf{E})^T \right) \otimes \mathbf{I}_{N_r} \right) \mathrm{vec}\left(\sqrt{P(l)} \mathbf{G}(l) \right), \tag{5.36}$$

where $\mathbf{D}$ represents the $N_c \times N_c$ diagonal matrix

$$\mathbf{D} = \begin{pmatrix} \exp(-j2\pi 0 / N_c) & & 0 \\ & \ddots & \\ 0 & & \exp(-j2\pi(N_c - 1)/N_c) \end{pmatrix}. \tag{5.37}$$

Since we assume that there is no correlation between the taps of the channel impulse response, it follows that $E[\mathrm{vec}(G(l))\mathrm{vec}(G(l'))^H] = \mathbf{0}$ for all $l \neq l'$. Furthermore note that $\mathbf{y}$ has a mean of zero (because $\mathbf{c}'$ and $\mathbf{e}'$ are also zero mean). As a result, the covariance matrix $\mathbf{Q}_y$ can be shown to be equal to

$$\begin{aligned} \mathbf{Q}_y &= E\left[\mathbf{y}\mathbf{y}^H \right] \\ &= E\left[\sum_{l=0}^{L-1} P(l)\left(\left(\mathbf{D}^l(\mathbf{C}-\mathbf{E})^T\right)\otimes \mathbf{I}_{N_r} \right)\mathrm{vec}(\mathbf{G}(l))\mathrm{vec}(\mathbf{G}(l))^H \left(\left(\mathbf{D}^l(\mathbf{C}-\mathbf{E})^T\right)\otimes \mathbf{I}_{N_r}\right)^H \right]. \end{aligned} \tag{5.38}$$

To include spatial correlation, an equivalent spatial correlation definition as presented in Section 3.6 in (3.49) is used per channel delay tap. This results in

$$\begin{aligned} \mathbf{Q}_y &= \sum_{l=0}^{L-1} P(l)\left(\left(\mathbf{D}^l(\mathbf{C}-\mathbf{E})^T\right)\otimes \mathbf{I}_{N_r} \right)\left(\mathbf{R}_{\mathrm{TX},l}^T \otimes \mathbf{R}_{\mathrm{RX},l}\right)\left(\left(\mathbf{D}^l(\mathbf{C}-\mathbf{E})^T\right)\otimes \mathbf{I}_{N_r}\right)^H \\ &= \sum_{l=0}^{L-1} \left(P(l)\mathbf{D}^l(\mathbf{C}-\mathbf{E})^T \mathbf{R}_{\mathrm{TX},l}^T (\mathbf{C}-\mathbf{E})^*\left(\mathbf{D}^l\right)^H \right)\otimes \mathbf{R}_{\mathrm{RX},l}, \end{aligned} \tag{5.39}$$

which is similar to the solution found in [20], except that we not only include the spatial correlation of the receiver side, but also that of the transmitter side. Factorising $\mathbf{R}_{\mathrm{TX},l}$ and $\mathbf{R}_{\mathrm{RX},l}$ into their square roots (according to (3.56)) yields

$$\begin{aligned} \mathbf{Q}_y &= \sum_{l=0}^{L-1} \left(P(l)\mathbf{D}^l(\mathbf{C}-\mathbf{E})^T \left(\mathbf{R}_{\mathrm{TX},l}^{\frac{1}{2}}\right)^* \left(\mathbf{R}_{\mathrm{TX},l}^{\frac{1}{2}}\right)^T (\mathbf{C}-\mathbf{E})^*\left(\mathbf{D}^l\right)^H \right)\otimes \left(\mathbf{R}_{\mathrm{RX},l}^{\frac{1}{2}}\left(\mathbf{R}_{\mathrm{RX},l}^{\frac{1}{2}}\right)^H \right) \\ &= \sum_{l=0}^{L-1} P(l)\left(\left(\mathbf{D}^l(\mathbf{C}-\mathbf{E})^T \left(\mathbf{R}_{\mathrm{TX},l}^{\frac{1}{2}}\right)^* \right)\otimes \mathbf{R}_{\mathrm{RX},l}^{\frac{1}{2}} \right)\left(\left(\left(\mathbf{R}_{\mathrm{TX},l}^{\frac{1}{2}}\right)^T (\mathbf{C}-\mathbf{E})^*\left(\mathbf{D}^l\right)^H \right)\otimes \left(\mathbf{R}_{\mathrm{RX},l}^{\frac{1}{2}}\right)^H \right) \\ &= \mathbf{B}(\mathbf{C},\mathbf{E})\mathbf{B}^H(\mathbf{C},\mathbf{E}), \end{aligned} \tag{5.40}$$

where $\mathbf{B}(\mathbf{C},\mathbf{E})$ is the $N_c N_r \times L N_t N_r$ block matrix

$$\mathbf{B}(\mathbf{C},\mathbf{E}) = \left(\mathbf{B}(\mathbf{C},\mathbf{E},0) \quad \cdots \quad \mathbf{B}(\mathbf{C},\mathbf{E},L-1) \right), \tag{5.41}$$

with the l-th block equal to

$$\mathbf{B}(\mathbf{C},\mathbf{E},l) = \sqrt{P(l)}\left(\left(\mathbf{D}^l(\mathbf{C}-\mathbf{E})^T \left(\mathbf{R}_{\mathrm{TX},l}^{\frac{1}{2}}\right)^* \right)\otimes \mathbf{R}_{\mathrm{RX},l}^{\frac{1}{2}} \right). \tag{5.42}$$

When we assume that $N_c > LN_t$, clearly, the maximum rank of $\mathbf{B(C,E)}$ (and $\mathbf{Q}_y$) is LN_tN_r. This means that the maximum achievable diversity order equals LN_tN_r. Based on above analysis, [20] introduces a number of design criteria for SF codes. When the available diversity is high enough, however, in line with the arguments of Subsection 4.2.2 and the simulation results of Subsection 4.11.2, the Euclidean distance criterion might be more appropriate.

To show this, assume that the channel elements are independent, i.e., there is no spatial fading correlation. Under this assumption, $\mathbf{R}_{\mathrm{TX},l}$'s and $\mathbf{R}_{\mathrm{RX},l}$'s are identity matrices. Hence, $\mathbf{Q}_y$ equals

$$\mathbf{Q}_y = \sum_{l=0}^{L-1}\left(P(l)\mathbf{D}^l\left(\mathbf{C}-\mathbf{E}\right)^T\left(\mathbf{C}-\mathbf{E}\right)^*\left(\mathbf{D}^l\right)^H\right)\otimes\mathbf{I}_{N_r} \equiv \mathbf{Q'}_y\otimes\mathbf{I}_{N_r}. \qquad (5.43)$$

Moreover, we will use the property that a determinant of the form $\det(\mathbf{I}_N + \alpha\mathbf{A})$ can be factorised as

$$\det\left(\mathbf{I}_N + \alpha\mathbf{A}\right) = 1 + \alpha\sum_{x=1}^{N}a_{xx} + \alpha^2\sum_{x=1}^{N-1}\sum_{y=x+1}^{N}\det\begin{pmatrix} a_{xx} & a_{xy} \\ a_{yx} & a_{yy} \end{pmatrix}$$

$$+ \alpha^3\sum_{x=1}^{N-2}\sum_{y=x+1}^{N-1}\sum_{z=y+1}^{N}\det\begin{pmatrix} a_{xx} & a_{xy} & a_{xz} \\ a_{yx} & a_{yy} & a_{yz} \\ a_{zx} & a_{zy} & a_{zz} \end{pmatrix} + ... + \alpha^N\det(\mathbf{A}). \qquad (5.44)$$

Let λ_i be the i-th nonzero eigenvalue of $\mathbf{Q'}_y$. Since we already found that $\mathbf{Q}_y$ has a maximum rank of LN_tN_r, $\mathbf{Q'}_y$ will have a maximum rank of LN_t. When N_i denotes the actual rank of $\mathbf{Q'}_y$, based on above factorisation and using the fact that the sum of the diagonal entries of a matrix equals the sum of its eigenvalues and that the determinant of a matrix equals the product of its eigenvalues, we can upperbound the PEP as follows:

$$\Pr\left(\mathbf{c'}\rightarrow\mathbf{e'}\right) \le \det\left(\mathbf{I}_{N_rN_c} + \frac{\sigma_s^2}{4\sigma_n^2}\mathbf{Q}_y\right)^{-1} = \det\left(\mathbf{I}_{N_rN_c} + \frac{\sigma_s^2}{4\sigma_n^2}\left(\mathbf{Q'}_y\otimes\mathbf{I}_{N_r}\right)\right)^{-1}$$

$$\le \left(1 + \frac{\sigma_s^2}{4\sigma_n^2}N_r\sum_{i=1}^{N_i}\lambda_i + \left(\frac{\sigma_s^2}{4\sigma_n^2}\right)^{N_iN_r}\left(\prod_{i=1}^{N_i}\lambda_i\right)^{N_r}\right)^{-1}. \qquad (5.45)$$

Now, using the same argumentation as in [4], one can see that this upperbound captures the effect of the frequency and receiver diversity on the code design, by means of N_i and N_r, respectively. For example, for a large frequency diversity and/or a large number of receive antennas, the packet error rate of interest (e.g., 10^{-2}) is typically achieved at small SNRs. Note that the SNR is proportional to σ_s^2/σ_n^2. For such small SNRs, it is clear that the term consisting of the summation of the eigenvalues will dominate the bound. For asymptotically high SNRs, however, the term containing the product of the eigenvalues will be dominant, resulting in the diversity and coding gain criteria introduced in [116]. The rule to optimise the former term is usually called the *trace criterion* because the sum

of the eigenvalues of a matrix equals the sum of its diagonal entries, i.e. the trace of the matrix. It can be shown that the trace of $\mathbf{Q}'_y$ equals

$$
\begin{aligned}
\operatorname{tr}\!\left(\mathbf{Q}'_{y}\right) &= \sum_{l=0}^{L-1}\operatorname{tr}\!\left(P(l)\mathbf{D}^{l}(\mathbf{C}-\mathbf{E})^{T}(\mathbf{C}-\mathbf{E})^{*}\left(\mathbf{D}^{l}\right)^{H}\right) \\
&= \sum_{l=0}^{L-1}\operatorname{tr}\!\left(P(l)\mathbf{D}^{l}\begin{pmatrix} \left\|\mathbf{c}(0)-\mathbf{e}(0)\right\|^{2} & & \times \\ & \ddots & \\ \times & & \left\|\mathbf{c}(N_{c}-1)-\mathbf{e}(N_{c}-1)\right\|^{2} \end{pmatrix}\left(\mathbf{D}^{l}\right)^{H}\right) \\
&= \sum_{l=0}^{L-1}P(l)\left\|\mathbf{C}-\mathbf{E}\right\|^{2}=\left\|\mathbf{C}-\mathbf{E}\right\|^{2},
\end{aligned}
\tag{5.46}
$$

where we assumed that the total power of the PDP equals one. Since in this case the trace equals the Euclidean distance between the SF codewords, the trace criterion is also called the Euclidean distance criterion. As a result, for high diversity orders and/or low SNR values, maximising the Euclidean distance between any two codewords is a better design rule than those based on the diversity and coding gain criteria. In [63] it is shown by simulations that the trace criterion is already applicable for values as low as $L = 2$ and $N_r = 2$.

To further simplify the code search, a technique called *subcarrier grouping* can be applied. From the channel model of (5.26), it follows that subcarriers are independent when they are separated by a multiplicity of N_c/L subcarriers. Following [147], we can group the subcarriers that have independent fades into N_{cg} groups of L subcarriers. Without loss of generality we will assume for this analysis that N_{cg} is a positive integer such that $N_{cg}L = N_c$. When defining $\mathbf{c}(\eta, g) = \mathbf{c}(\eta \cdot N_{cg} + g)$ and $\mathbf{e}(\eta, g) = \mathbf{e}(\eta \cdot N_{cg} + g)$, for $g = 0, \ldots, N_{cg} - 1$ and $\eta = 0, \ldots, L - 1$, this results in the conditional PEP

$$
\Pr\!\left(\mathbf{c}'\rightarrow\mathbf{e}'\big|\mathbf{H}'\right)\le\exp\!\left(-\frac{\sigma_{s}^{2}}{4\sigma_{n}^{2}}\sum_{g=0}^{N_{cg}-1}\sum_{\eta=0}^{L-1}\left\|\sum_{l=0}^{L-1}\sqrt{P(l)}\,\mathbf{G}(l)e^{-\frac{j2\pi\eta l}{L}}e^{-\frac{j2\pi gl}{N_{c}}}\left(\mathbf{c}(\eta,g)-\mathbf{e}(\eta,g)\right)\right\|^{2}\right).
\tag{5.47}
$$

The principle of subcarrier grouping is explained in Figure 5-9, in which it is assumed without loss of generality that a certain SISO channel undergoes block fading over frequency. The different grey shades represent independent fading snapshots.
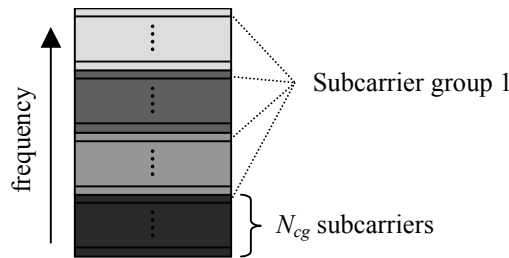
Figure 5-9: The principle of subcarrier grouping.

Substituting

$$\mathbf{y}(g)=\begin{pmatrix} \sum_{l=0}^{L-1}\sqrt{P(l)}\mathbf{G}(l)e^{-\frac{j2\pi l\cdot 0}{L}}e^{-\frac{j2\pi gl}{N_c}}\big(\mathbf{c}(0,g)-\mathbf{e}(0,g)\big) \\ \vdots \\ \sum_{l=0}^{L-1}\sqrt{P(l)}\mathbf{G}(l)e^{-\frac{j2\pi l(L-1)}{L}}e^{-\frac{j2\pi gl}{N_c}}\big(\mathbf{c}(L-1,g)-\mathbf{e}(L-1,g)\big) \end{pmatrix},\qquad(5.48)$$

yields

$$\Pr\big(\mathbf{c}'\rightarrow\mathbf{e}'\big|\mathbf{H}'\big)\le\exp\left(-\frac{\sigma_s^2}{4\sigma_n^2}\sum_{g=0}^{N_{cg}-1}\|\mathbf{y}(g)\|^2\right).\qquad(5.49)$$

Averaging over all channel realisations, in line with the analysis of Subsection 4.10.2, results in the following PEP upperbound:

$$\Pr\big(\mathbf{c}'\rightarrow\mathbf{e}'\big)\le\det\left(\mathbf{I}_{N_rL}+\frac{\sigma_s^2}{4\sigma_n^2}\sum_{g=0}^{N_{cg}-1}\mathbf{Q}_y(g)\right)^{-1},\qquad(5.50)$$

where $\mathbf{Q}_y(g)$ is the covariance matrix of $\mathbf{y}(g)$. Now, we can follow the same steps as in the general analysis at the beginning of this section to analyse the PEP in case of subcarrier grouping. In accordance with those steps, we define $\mathbf{C}(g) = (\mathbf{c}(0, g) \ldots \mathbf{c}(L - 1, g))$ and $\mathbf{E}(g) = (\mathbf{c}(0, g) \ldots \mathbf{c}(L - 1, g))$ as two possible $N_t \times L$ codeword matrices of the g-th subcarrier group. Then by applying (A.12) we can show that

$$\mathbf{y}(g)=\sum_{l=0}^{L-1}\big(\mathbf{D}^l(g)(\mathbf{C}(g)-\mathbf{E}(g))^T\otimes\mathbf{I}_{N_r}\big)\mathrm{vec}\big(\sqrt{P(l)}\,\mathbf{G}(l)\big),\qquad(5.51)$$

where $\mathbf{D}(g)$ equals the $L \times L$ diagonal matrix

$$\mathbf{D}(g)=\mathrm{diag}\left\{\exp\left(-\frac{j2\pi\eta}{L}\right)\exp\left(-\frac{j2\pi g}{L}\right)\right\}_{\eta=0,\ldots,L-1}.\qquad(5.52)$$

Finally, we can show that

$$\mathbf{Q}_y(g)=\mathbf{B}_g\big(\mathbf{C}(g),\mathbf{E}(g)\big)\mathbf{B}_g^H\big(\mathbf{C}(g),\mathbf{E}(g)\big),\qquad(5.53)$$

where $\mathbf{B}_g(\mathbf{C}(g),\mathbf{E}(g))$ is the $LN_r \times LN_tN_r$ block matrix

$$\mathbf{B}_g\big(\mathbf{C}(g),\mathbf{E}(g)\big)=\big(\mathbf{B}_g\big(\mathbf{C}(g),\mathbf{E}(g),0\big)\quad\cdots\quad\mathbf{B}_g\big(\mathbf{C}(g),\mathbf{E}(g),L-1\big)\big),\qquad(5.54)$$

of which the l-th block equals

$$\mathbf{B}_g\left(\mathbf{C}(g),\mathbf{E}(g),l\right) = \sqrt{P(l)}\left(\left(\mathbf{D}^l(g)(\mathbf{C}(g)-\mathbf{E}(g))^T \left(\mathbf{R}_{\mathrm{TX},l}^{\frac{1}{2}}\right)^*\right) \otimes \mathbf{R}_{\mathrm{RX},l}^{\frac{1}{2}}\right). \qquad (5.55)$$

From above analysis, we clearly see that the maximum rank of $\mathbf{B}_g(\mathbf{C}(g),\mathbf{E}(g))$ and, thus, $\mathbf{Q}_y(g)$ equals LN_r. Apparently, the transmit diversity is lost due to the application of subcarrier grouping. This can be explained by the fact that every subcarrier group is loaded with independent data, hence, achieving a "multiplexing gain". If we require the full diversity gain LN_tN_r, we can add the time dimension to the analysis, which leads to the STF results of [147]. When taking these concepts together, we can trade off multiplexing gain and diversity gain (see also [73]).

Also for the subcarrier grouping result, the PEP upperbound can be factorised. Under the assumption that there is no spatial correlation, this yields

$$\Pr\left(\mathbf{c'}\rightarrow\mathbf{e'}\right) \le \det\left(\mathbf{I}_{N_rL} + \frac{\sigma_s^2}{4\sigma_n^2}\sum_{g=0}^{N_{cg}-1}\mathbf{Q}_y(g)\right)^{-1}$$

$$= \det\left(\mathbf{I}_{N_rL} + \frac{\sigma_s^2}{4\sigma_n^2}\sum_{g=0}^{N_{cg}-1}\left(\mathbf{Q'}_y(g)\otimes\mathbf{I}_{N_r}\right)\right)^{-1} \qquad (5.56)$$

$$\le \left(1 + \frac{\sigma_s^2}{4\sigma_n^2}N_r\mathrm{tr}\left(\sum_{g=0}^{N_{cg}-1}\mathbf{Q'}_y(g)\right) + \left(\frac{\sigma_s^2}{4\sigma_n^2}\right)^{N_iN_r}\det\left(\sum_{g=0}^{N_{cg}-1}\mathbf{Q'}_y(g)\right)^{N_r}\right)^{-1},$$

where N_i is the rank of the summation over all $\mathbf{Q'}_y(g)$'s. Using (5.46), it is easily verified that

$$\mathrm{tr}\left(\sum_{g=0}^{N_{cg}-1}\mathbf{Q'}_y(g)\right) = \sum_{g=0}^{N_{cg}-1}\left\|\mathbf{C}(g)-\mathbf{E}(g)\right\|^2 = \left\|\mathbf{C}-\mathbf{E}\right\|^2. \qquad (5.57)$$

As a result, due to subcarrier grouping, the trace criterion, i.e., maximising the Euclidean distance between any pair of Space-Frequency codewords, is simplified to maximising the Euclidean distance between any two codewords of a subcarrier group.

## 5.7   Coded Space Division Multiplexing OFDM

5.7.1   Introduction

In the previous section it was shown that, when the potential diversity gain is high enough and the SNR of interest is low enough, the traditional code design criterion of maximising the minimum Euclidean distance between any pair of codewords ($\|\mathbf{C}-\mathbf{E}\|$) is more appropriate than specific Space-Frequency code design rules, i.e., the diversity and coding gain criteria as defined in, e.g., [20]. This can be explained by the fact that, when a reasonably large diversity gain is achievable through transmit, receive, and/or frequency diversity, a frequency-selective MIMO fading channel converges to a Gaussian channel

(based on the Central Limit Theorem ([90])) under the condition that proper encoding is applied across the diversity dimensions. As a result, under above conditions, standard SISO codes together with some form of space and frequency multiplexing may outperform handcrafted Space-(Time-)Frequency codes. Based on this argument, the concatenation of coding with the straightforward multiplexing over space and frequency of Space Division Multiplexing (SDM) OFDM is a promising starting point. Moreover, such a coded SDM OFDM scheme offers the flexibility of easily adapting the constellation order and/or coding rate. Basically, there are two options to add coding to SDM OFDM, namely Joint Coding (JC) and Per-Antenna-Coding (PAC), which are explained in the next subsections.
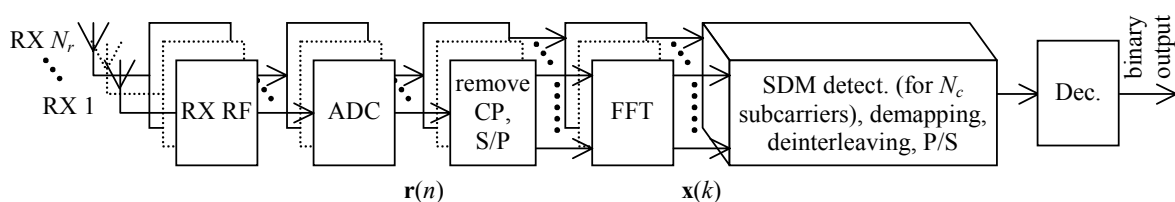
5.7.2   Joint Coding

In Joint Coding (JC), also referred to as vertical coding ([146]), the information bit stream is first encoded and then converted into N_t parallel substreams of which each is modulated and mapped onto the corresponding transmit antenna. A transmitter scheme in which JC is applied to SDM OFDM is shown in Figure 5-10, where S/P denotes the serial-to-parallel conversion. After the S/P block, each branch in parallel performs interleaving (Π), QAM mapping, pilot insertion, N_c-point IFFT, and adds a Cyclic Prefix before the final TX signal is shaped, converted up to the Radio Frequency (RF), and transmitted.
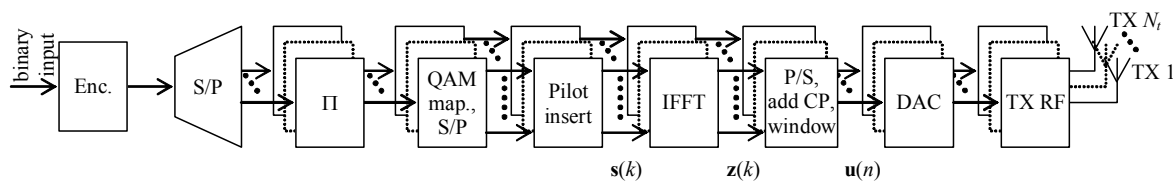
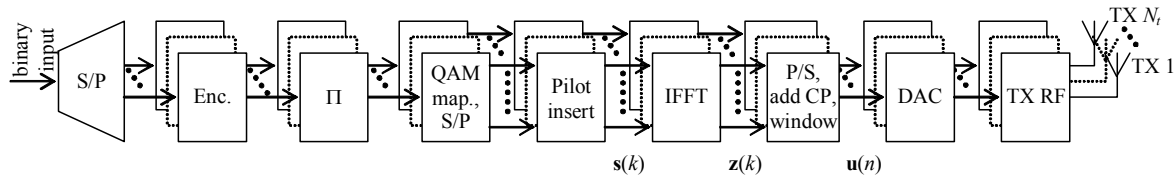**Figure 5-10: Schematic representation of a Joint-Coded SDM OFDM transmitter.**

A schematic representation of an SDM OFDM receiver for JC transmissions is given in Figure 5-11. Disregarding the blocks for synchronisation tasks, the receiver first generates a digital representation of the N_r down-converted received signals by means of the ADCs. Second, the OFDM processing is performed per branch up to the FFT outputs. At this point, SDM detection is performed per subcarrier. To that end, the received signals of subcarrier i are routed to the i-th SDM detector to recover the N_t transmitted symbols on the corresponding subcarrier. After that, symbols per transmitter stream are combined, and demapping and deinterleaving is performed for these N_t parallel streams. Finally, the resulting streams are converted to a serial stream that, in turn, is decoded to generate the binary output sequence. Note that for the SDM detection, we can apply all SDM schemes described in Chapter 4.

Figure 5-11: Schematic representation of a SDM OFDM receiver for a Joint-Coding architecture at the transmitter.
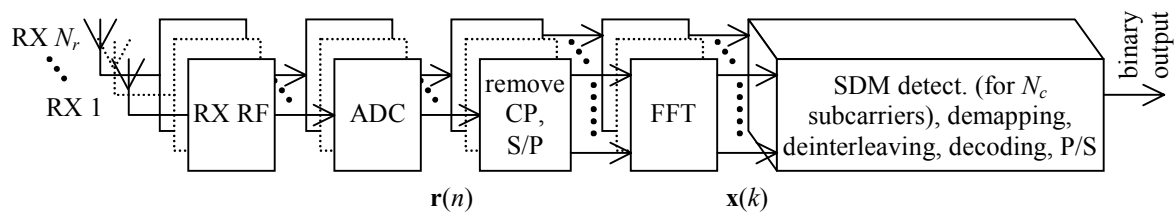
5.7.3 Per-Antenna-Coding

In Per-Antenna-Coding (PAC) schemes, the incoming bit stream is first transformed to N_t parallel substreams and then encoding is performed per substream. So, basically, the transmitter consists of N_t OFDM transmitters among which the information bits are multiplexed ([130]), as shown in Figure 5-12.

Figure 5-12: Schematic representation of a Per-Antenna-Coded SDM OFDM transmitter.

The receiver for a PAC transmitter is exactly the same as that for a JC transmitter up to and including the interleavers. The difference is that after interleaving the N_t detected substreams are first decoded per stream before they are converted into a serial stream. A schematic representation of such a receiver is given in Figure 5-13.

Figure 5-13: Schematic representation of an SDM OFDM receiver for a Per-Antenna-Coding architecture at the transmitter.

For the SDM detection in the PAC receiver all SDM schemes of Chapter 4 can be applied. Moreover, for Successive-Interference-Cancellation schemes there exists an extra option, namely, feeding the SIC signals first through the decoding stage before actually performing the cancellation. In this way, Forward-Error-Correcting coding is performed on the SIC information. How this can be applied to MIMO OFDM is schematically represented in Figure 5-14. We introduced this concept in [130] where we called it PAC V-BLAST.

To improve the performance even more, the SIC scheme (either using ZF or MMSE) should produce soft-decision outputs. In order to obtain them, the estimated values of each Nulling Unit should not be sliced to their respective QAM points as done in [144], but they should be used to generate soft values according to (4.71), with the proper **Q** as defined in Subsections 4.6.4 and 4.7.3 for SOZF and SOMMSE, respectively. Note that, during the SIC process, **H** changes (see Subsections 4.8.1 and 4.9.1), and since **Q** (required to obtain the $q_{k'k'}$ of (4.71)) is a function of **H**, it must change accordingly. It turned out that when we used these soft value definitions for SOZF with SIC or SOMMSE with SIC in the JC case that the performance did not improve, but deteriorated. The explanation is that apparently the scaling of (4.71) due to the changing **Q** during the SIC process is incorrect. For PAC,

however, the inherent structure of a PAC receiver results in a decoding per TX stream and, as a result, the inputs per convolutional decoder all have the same (but incorrect) scaling factor $q_{k'k'}$, still leading to a proper decoding because the soft values still have a correct relative value. This last argument does not hold for the decoding of a JC scheme. A proper definition of soft values for SIC schemes is a topic for further research.

Note that since the SIC information is passed through a decoder and encoder stage, a disadvantage of this scheme could be its latency. But when the interleaver size is small, and convolutional encoding and Viterbi decoding are used, the encoder in the SIC feedback loop can start its operation as soon as the Viterbi decoder produces outputs. Then, for a limited number of transmit antennas, the latency is manageable.

As example, consider an implementation based on the WLAN IEEE 802.11a standard (more details on the parameters can be found in the next section). The interleaver and deinterleaver in the SIC loop both operate on an OFDM symbol basis, so their respective latency is at most 3.2 µs (4 µs minus Guard Interval). The delay of the Viterbi decoder is directly related to its trace-back depth. A common rule of thumb is to take the trace-back depth equal to six times the constraint length of the encoder. The constraint length of the IEEE 802.11a encoder is seven. Consequently, the delay of the Viterbi decoder equals 42 samples or 42/20 MHz = 2.1 µs. Finally, the delay of the encoder is equal to its memory size times the sample period. For the IEEE 802.11a encoder this is 300 ns. So, the total latency per loop is at most 2*3.2 + 2.1 + 0.3 = 8.8 µs.

Another disadvantage may be that the optimal ordering ([144]), for the sequence in which the SIC is performed (e.g., it is best to perform detection in the order of signal strength) cannot be applied per subcarrier but only per substream, inherent to the decoding structure. Hence, this per-substream ordering does not take into account the different frequency selectivity of the various substreams, but simulations in this chapter show that its performance is nevertheless close to SOMLD.
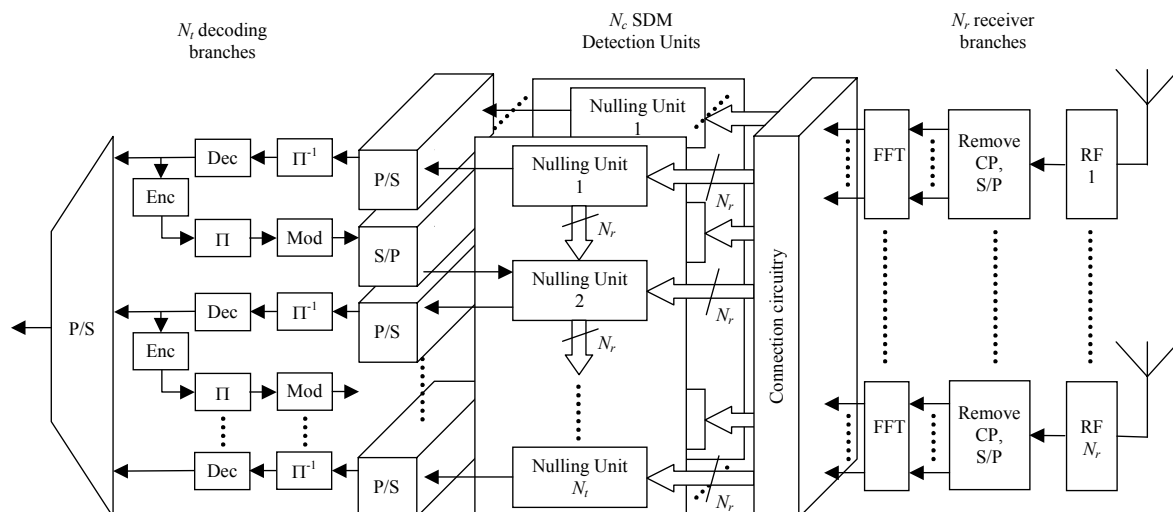
Figure 5-14: Multi-antenna receiver with Per-Antenna-Coded SIC and OFDM.

5.8 Simulations

5.8.1 Simulation parameters

The JC and PAC systems proposed in the previous section have been programmed in MATLAB to be able to evaluate their performance. The main parameters that are used for the simulations are based on the IEEE 802.11a WLAN standard ([57]). An overview of these parameters is listed in Table 5-1.

Table 5-1: Main parameters based on the IEEE 802.11a OFDM standard.

Constellation diagrams	BPSK, QPSK, 16-QAM, 64-QAM
Coding rate (R)	1/2, 2/3, 3/4
Number of subcarriers (N_c)	64
Number of subcarriers used for data (N_u)	48
Number of pilot tones (N_p)	4
OFDM symbol duration	4 μs
Guard interval	800 ns
Subcarrier spacing	312.5 kHz
–3 dB bandwidth	16.56 MHz
Channel spacing	20 MHz

A key parameter, which largely determines the choice of the other parameters, is the guard interval of 800 ns ([127]). This guard interval provides robustness to rms delay spreads up to several hundreds of nanoseconds, depending on the chosen coding rate and constellation diagram. In practice, this means that an OFDM system with a guard time of 800 ns is robust enough to be used in any indoor environment, including large factory buildings ([126]). It can also be used in outdoor environments, although directional antennas may be needed in this case to reduce the delay spread to an acceptable amount and to increase the range.

In order to limit the relative amount of energy spent on the guard time, the OFDM symbol duration was chosen to be 4 μs. This also determined the subcarrier spacing to be 312.5 kHz, which is the inverse of the symbol duration minus the guard time. By using 48 data subcarriers, uncoded data rates of 12 to 72 Mbps can be achieved by using variable constellation sizes from BPSK to 64-QAM. In addition to the 48 data subcarriers, each OFDM symbol contains four pilot subcarriers, which can be used to track the residual carrier frequency offset that remains after an initial frequency offset correction during the training phase of the packet. Moreover, the low frequency components of the phase noise generated by non-ideal oscillators can be estimated and corrected for. Without correction, these two impairments would cause a common phase drift on all subcarriers. Note that, in the IEEE 802.11a standard, the total of 52 subcarriers used are distributed equally around the 0th subcarrier. This 0th subcarrier and the others are set to zero. So, denoting only the used subcarriers out of the 64 with an index, this results in the following subcarrier layout: {0, 0, 0, 0, 0, 0, –26, …, –1, 0, 1, …, 26, 0, 0, 0, 0, 0}.

In order to correct for subcarriers in deep fades, forward error correction across the subcarriers is used with variable coding rates, giving coded data rates from 6 up to 54 Mbps. Convolutional coding is used with the industry standard rate 1/2, constraint length 7 code with generator polynomials (133,171). Higher coding rates of 2/3 and 3/4 are obtained by puncturing the rate 1/2 code.

For the simulations in this chapter we will assume that the receiver has perfect knowledge about the channel (i.e., perfect CSI) data rate, and packet size. Furthermore, we will assume that the time and frequency synchronisation is optimal. Hence, we can ignore any training and/or preamble processing and focus on the performance of the different MIMO OFDM schemes.

5.8.2 SNR versus Bit Energy-to-Noise Density Ratio

All simulation results in the next subsection are provided as function of the average SNR per RX antenna, but if one wants to make a fair performance comparison of systems with different coding schemes or constellation schemes in terms of energy spent per bit, commonly, the error-rate performance as function of the bit energy-to-noise density ratio E_b/N_0 is used. To be able to translate the results of the next subsection to average E_b/N_0 per RX antenna, this subsection provides the relation between E_b/N_0 and the average SNR per RX antenna, here defined as the average SNR at the input of the RX baseband processing and denoted by E_s/N_0.

In the baseband processing of a MIMO OFDM transmission system, there are a number of subsequent blocks that have an influence on the relation between E_b/N_0 and E_s/N_0. These blocks are:

-    the encoder with coding rate R ($R < 1$),
-    the modulation block that maps m bits on a 2^m-ary modulation scheme,
-    the spatial mapper that maps N_t symbols on N_t transmit antennas,
-    the N_c-point IFFT that maps N_u symbols on N_u subcarriers ($N_u < N_c$) ,
-    and a block that performs the guard time extension (cyclic prefix).

Now, assume that the communication between transmitter and receiver is scaled such that the variance of the propagation attenuation equals $\sigma_c^2 = 1$. Then, recalling Equation (4.31),

$$\frac{E_b}{N_0} = \frac{E_s}{N_0} \frac{T_b}{T_s} , \tag{5.58}$$

the relation between E_b/N_0 and E_s/N_0 can be determined per above described block, assuming that the TX processing only comprises that specific block.

The first three blocks give a result equivalent to the single carrier MIMO case described in Section 4.5. For this case the relation is

$$\frac{E_s}{N_0} = Rm N_t \frac{E_b}{N_0} . \tag{5.59}$$

Putting N_u bits on N_u subcarriers using an N_c-point IFFT results in $T_s = N_u T_b / N_c$. By substituting this into Equation (5.58), the symbol-energy to noise-density ratio per receive antenna equals

$$\frac{E_s}{N_0} = \frac{N_u}{N_c} \frac{E_b}{N_0}.$$

(5.60)

Note that when we want to take the overhead of pilot tones into account, we should not use the total amount of subcarriers used, but N_u minus the number of pilots in above equation.

Adding a cyclic prefix (i.e., guard time) of time T_G to an OFDM symbol of length T_{FFT} leads to the relation $T_s = T_{FFT} \cdot T_b / T_{tot}$, where $T_{tot} = T_G + T_{FFT}$. Thus

$$\frac{E_s}{N_0} = \frac{T_{FFT}}{T_{tot}} \frac{E_b}{N_0}.$$

(5.61)

For a MIMO OFDM system all the above blocks are combined in a serial way, so finally, for this type of systems, this leads to the following relation (with $T_{FFT} = N_c \cdot T_s$) :

$$\frac{E_s}{N_0} = N_t R m \frac{N_u T_{FFT}}{N_c T_{tot}} \frac{E_b}{N_0} = N_t R m \frac{N_u T_s}{T_{tot}} \frac{E_b}{N_0} = N_t R m \frac{N_u}{B T_{tot}} \frac{E_b}{N_0} = \eta_{eff} \frac{E_b}{N_0}.$$

(5.62)

Note that the spectra efficiency η_{eff} is bits/s/Hz equals the ratio of the bit rate, $N_t R m N_u / T_{tot}$, and bandwidth, B.

5.8.3 Simulations results

The main goal of this subsection is to compare the various proposed coded SDM OFDM schemes for different antenna configurations and for different rms delay spreads. Since, in general, the main application of WLANs are indoor environments, we will apply the uncorrelated NLOS wideband Rayleigh fading model described in Subsection 3.3.2 as channel model, unless mentioned otherwise. For the decoding of the convolutional code a Viterbi decoder is used. From [40], it is known that a Viterbi decoder performs better when soft decisions are applied as input. Therefore, for the detection of JC transmissions, SOZF, SOMMSE and SOMLD are used, and for the detection of PAC transmissions, MMSE with SIC is applied with soft-decision output values generated as described in Subsection 5.7.3. In the remainder of this section the last scheme is referred to as PAC SIC. As a performance measure, we will use the average SNR per RX antenna that is required to achieve a PER of 10^{-2}, or 1%. The packet size is chosen to be 64 bytes[1] and the BER and PER performances are obtained by averaging over 10,000 packets undergoing independent channel realisations. Furthermore, for each plot a 1×1 curve using the same coding rate and constellation size as each of the multiple antennas is given as benchmark and, when the outage PER as defined in Subsection 5.5.2 does not fall below the SNR range of the

---

[1] Note that, compared to 64 byte packets, for 1000 byte packets, on average we saw for comparable PERs an SNR degradation of about 1.5 dB, except for high delay spreads where we observed degradations over 5 dB.

simulation results, it is also shown. Note that for the 1×1 case all of above mentioned schemes overlap, since for this case there is no difference in the processing.

The first set of simulations is done for a wideband Rayleigh fading channel with an rms delay spread of 50 ns, for QPSK and a coding rate of ½. The BER and PER performance versus the average SNR per receive antenna (in dB) for a 2×2 system, 3×3 system, 4×4 system, and 2×3 system are given in Figure 5-15, Figure 5-16, Figure 5-17, and Figure 5-18, respectively. Since the constellation size and coding rate is kept the same, the 1×1 curves of these figures are the same. The total data rate is given by N_t times the coding rate times the number of bits per constellation symbol times the number of data subcarriers divided by the OFDM symbol length (in time). This results for the 1×1, 2×2, 3×3, and 4×4 case in a respective data rate of 12 Mbps, 24 Mbps, 36 Mbps, and 48 Mbps. Accordingly, the outage PER is shown in Figure 5-15, Figure 5-16, and Figure 5-17 for capacities of $24/20 = 1.2$ bits/s/Hz, $36/20 = 1.8$ bits/s/Hz, and $48/20 = 2.4$ bits/s/Hz, respectively. We can make a number of observations.

First, the diversity order increases for SOMMSE, PAC SIC, and SOMLD when going to antenna configurations with higher numbers of antennas. Apparently, the schemes benefit from higher receiver diversity. This leads for the region of interest to a better performance than the 1×1 case, even while the total data rate is increased. SOZF, however, tends to follow the 1×1 curves and slightly loses in performance when going to an antenna configuration with more antennas.

Second, for a PER of 1%, for the 2×2, 3×3, and 4×4 simulations, the results are far off from the outage PER, namely, about 7 dB. Part of this big difference can be explained by the inherent overhead of the MIMO OFDM symbol structure: only 48 out of 64 subcarriers are used for data and only 3.2 μs of the 4 μs packetlength is non-redundant due to the guard interval. As a result, the performance loss/energy loss compared to the outage PER is already $10 \cdot \log_{10}(4/3.2 \cdot 64/48) \approx 2.2$ dB. The remaining gap is 4.8 dB, so there is still room for improvement and it would be desirable to find schemes that perform closer to the outage PER.

Third, when an extra receive antenna is added to the 2×2 system (compare Figure 5-15 and Figure 5-18), the extra diversity gain makes that all coded SDM OFDM schemes at a PER of 1% perform in a small range of 2.5 dB from each other. Moreover, the performance is 5-7 dB better than the 1×1 case, while achieving a twice as high data rate.

Since in the previous simulation results the PAC SIC and JC SOMLD detection performed the best, these schemes are used to verify the performance for different rms delay spreads. The results for a 3×3 system with constellation scheme QPSK and a coding rate of ½ is given in Figure 5-19, for rms delay spreads of 10 ns, 100 ns, and 250 ns. The 50 ns curves can be found in the previous figures. Moreover, it turned out that the curves of the coded SDM OFDM schemes for an rms delay spread of 100 ns and 250 ns were very similar. Therefore, the former results are not shown. It can be concluded that the MIMO OFDM schemes, especially for low rms delay spreads, achieve a better performance than the SISO case due to the spatial diversity gain. For higher rms delay spreads the additional spatial diversity on top of the frequency diversity does not provide much gain compared to the 1×1 performance. This can be explained by the fact that, when an extra degree of

diversity is added, the highest performance improvements are achieved for the first diversity improvements.
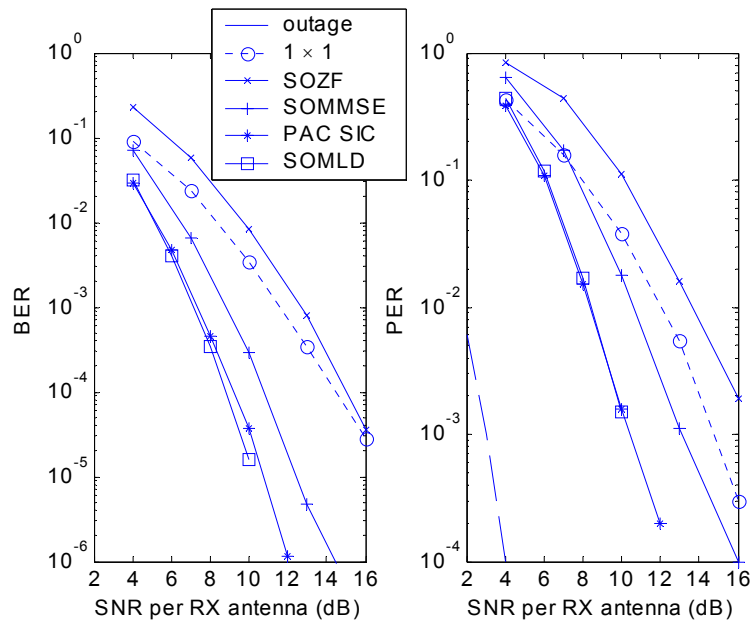
Figure 5-15: BER and PER versus mean SNR per receive antenna for a 2 × 2 system in a Rayleigh fading channel with an exponentially decaying PDP with 50 ns rms delay spread for QPSK, rate ½ convolutional coding (24 Mbps), 64 byte packets, and different SDM detection schemes. Also the 1 × 1 curves are given as reference.
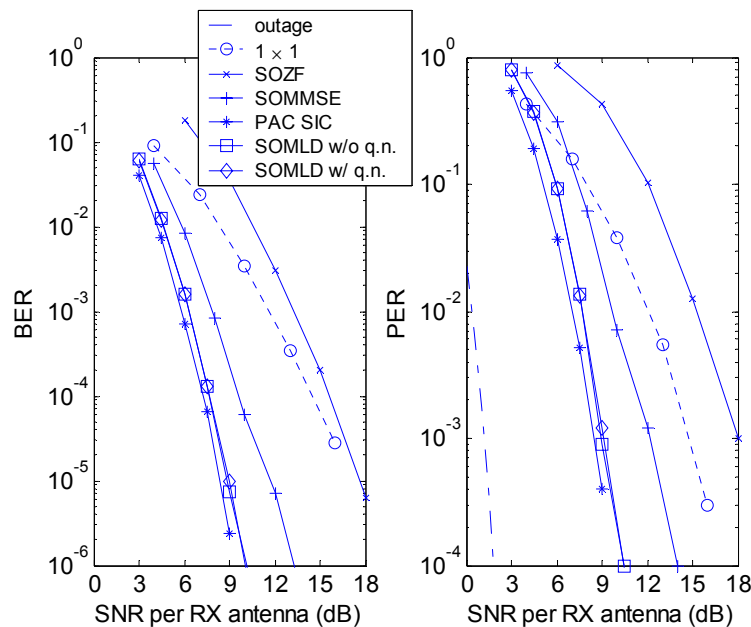
Figure 5-16: BER and PER versus mean SNR per receive antenna for a 3 × 3 system in a Rayleigh fading channel with an exponentially decaying PDP with 50 ns rms delay spread for QPSK, rate ½ convolutional coding (36 Mbps), 64 byte packets, and different SDM detection schemes. For SOMLD, both the curves with quantisation normalisation (q.n.) defined in (4.79) with $N_q = 8$ and without q.n. are presented. Also the 1 × 1 curves are given as reference.
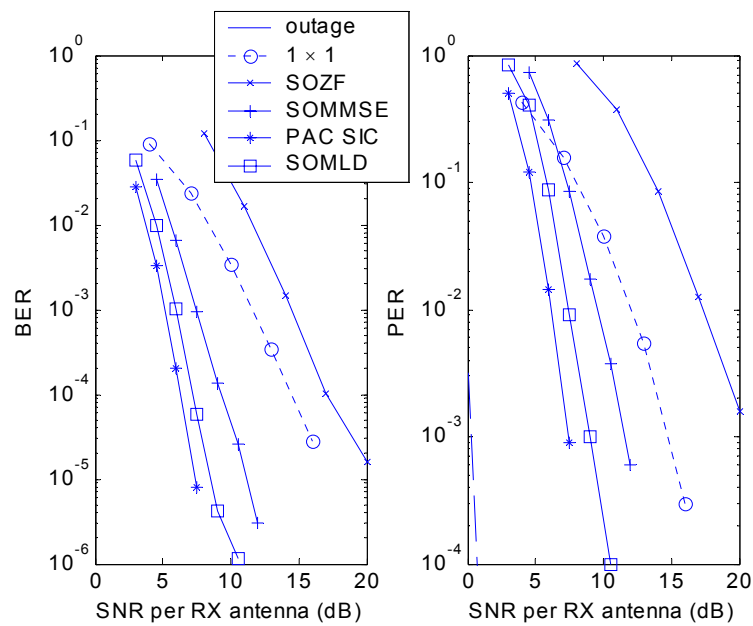
Figure 5-17: BER and PER versus mean SNR per receive antenna for a 4 × 4 system in a Rayleigh fading channel with an exponentially decaying PDP with 50 ns rms delay spread for QPSK, rate ½ convolutional coding (48 Mbps), 64 byte packets, and different SDM detection schemes. Also the 1 × 1 curves are given as reference.
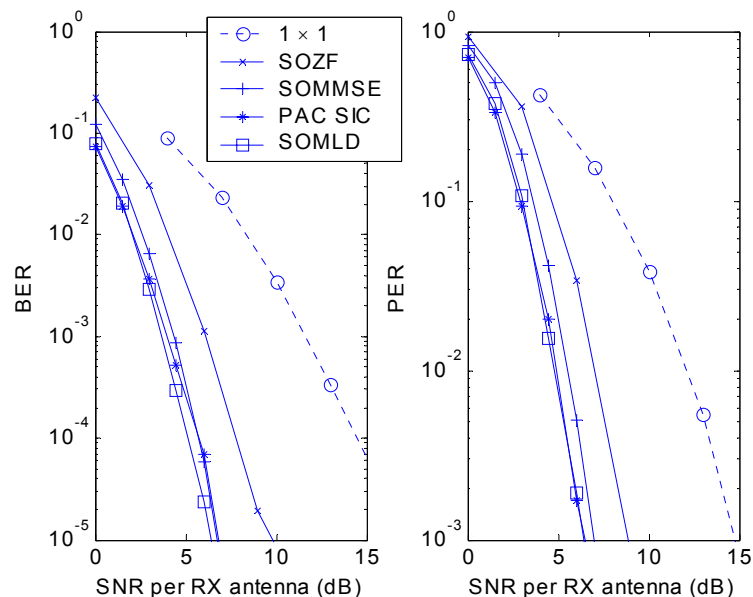

Figure 5-18: BER and PER versus mean SNR per receive antenna for a 2 × 3 system in a Rayleigh fading channel with an exponentially decaying PDP with 50 ns rms delay spread for QPSK, rate ½ convolutional coding (24 Mbps), 64 byte packets, and different SDM detection schemes. Also the 1 × 1 curves are given as reference.
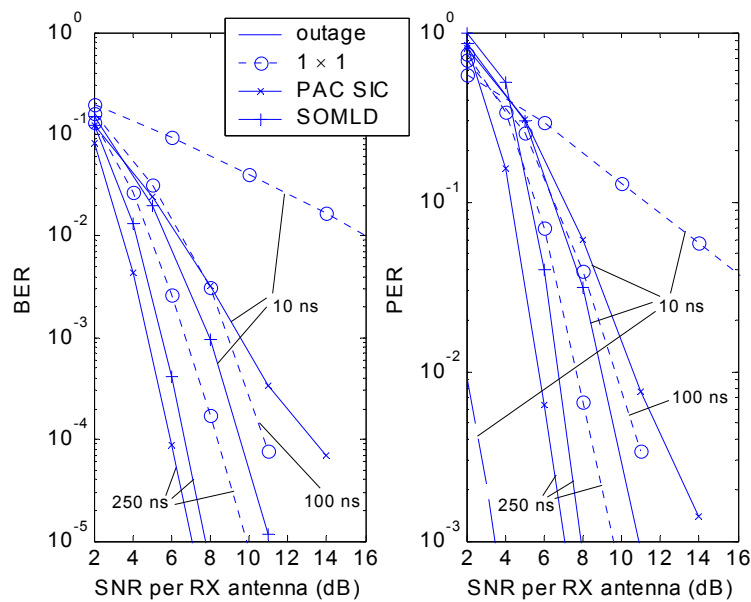
Figure 5-19: BER and PER versus mean SNR per receive antenna for a 3 × 3 system in a Rayleigh fading channel with an exponentially decaying PDP with various rms delay spreads for QPSK, rate ½ convolutional coding (36 Mbps), 64 byte packets, and different SDM detection schemes. Also the 1 × 1 curves are given as reference.

When comparing Figure 5-15 through Figure 5-20, it can be seen that PAC SIC performs better than SOMLD for high rms delay spreads (with an exponentially decaying PDP) and low constellation sizes. For low rms delay spreads SOMLD performs better, which can be explained by the fact that the spatial diversity order of MMSE with SIC for high SNRs tends towards $N_r - N_t + 1$, whereas that of SOMLD is equal to N_r (see Chapter 4). On the other hand, for high rms delay spreads, the frequency diversity plays a more important role which leads to advantages for PAC. Namely, in PAC, per TX stream a convolutional encoder with a free distance of 10 is used, by which up to a 10th-order frequency diversity can be exploited, whereas in JC the gain of this free distance is divided among both the frequency and spatial diversity.

With respect to higher constellation orders, the likelihood reduces that the feedback information used for SIC is correct, which results in a performance penalty for PAC SIC over SOMLD. The cases where PAC SIC performs better than SOMLD can be explained, next to above described extra frequency diversity gain, by the way the soft-decision output values are generated. Due to non-orthogonal MIMO channels, received MIMO vectors will have dependent elements, which for JC SOMLD will result in dependent soft values. It is well known that the Viterbi decoder only performs optimally if the input values are independent. A solution would be to calculate joint LLRs ([55]) and, accordingly, change the decoder to handle these joint soft-values. Since the PAC SIC results are based on MMSE with SIC, it first orthogonalises the data streams and then determines the soft values and, hence, PAC SIC does not have the above mentioned problem, which explains its better performance.

To further investigate the effect of larger constellation diagrams, Figure 5-20, Figure 5-21, and Figure 5-22 show the performance results for a 3 × 3 system with coding rate ¾, 16-

QAM (i.e., 108 Mbps) and 64-QAM (i.e., 162 Mbps), and rms delay spreads of 50 ns and 250 ns.
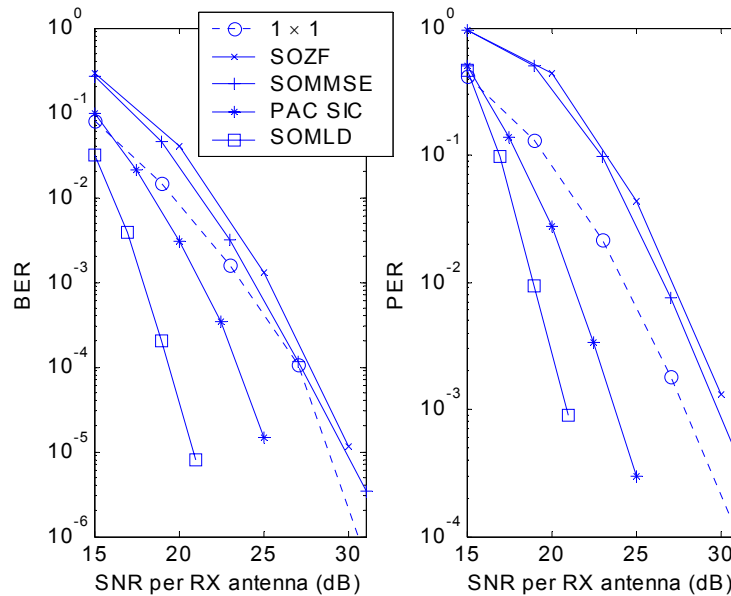

Figure 5-20: BER and PER versus mean SNR per receive antenna for a 3 × 3 system in a Rayleigh fading channel with an exponentially decaying PDP with 50 ns rms delay spread for 16-QAM, rate ¾ convolutional coding (108 Mbps), 64 byte packets, and different SDM detection schemes. Also the 1 × 1 curves are given as reference.
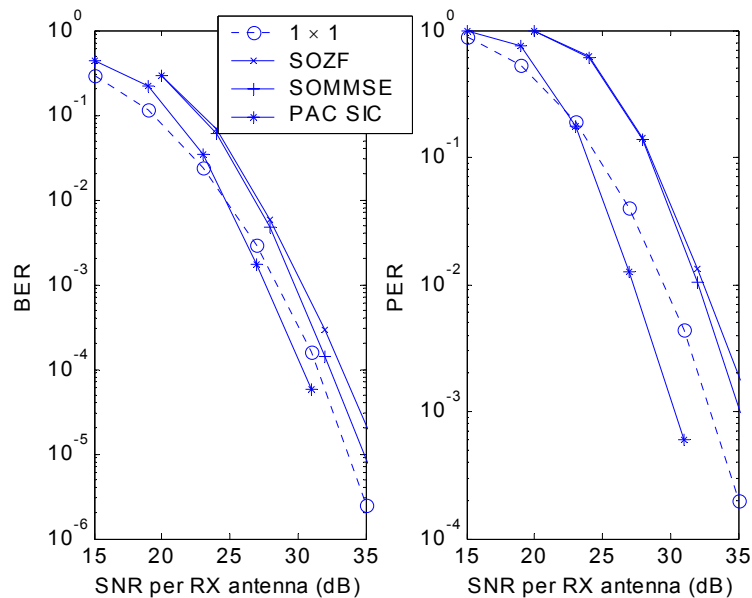
Figure 5-21: BER and PER versus mean SNR per receive antenna for a 3 × 3 system in a Rayleigh fading channel with an exponentially decaying PDP with 50 ns rms delay spread for 64-QAM, rate ¾ convolutional coding (162 Mbps), 64 byte packets, and different SDM detection schemes. Also the 1 × 1 curves are given as reference.
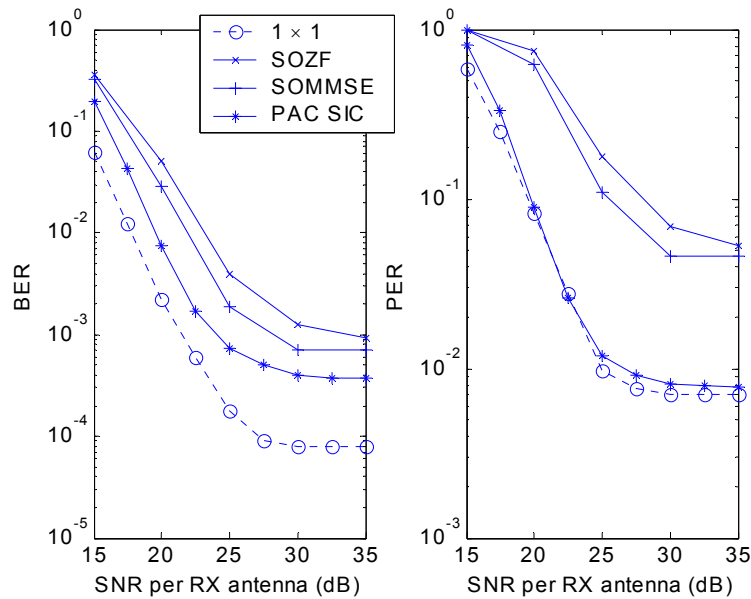
Figure 5-22: BER and PER versus mean SNR per receive antenna for a 3 × 3 system in a Rayleigh fading channel with an exponentially decaying PDP with 250 ns rms delay spread for 16-QAM, rate ¾ convolutional coding (108 Mbps), 64 byte packets, and different SDM detection schemes. Also the 1 × 1 curves are given as reference.

We can observe the following. Firstly, for high constellation sizes, SOZF and SOMMSE are performing very similar in the region of interest. Apparently, the required SNR is so high that the pseudo inverse of ZF and the weight matrix of MMSE are approximately the same resulting in approximately equivalent performance.

Secondly, for symmetric antenna configurations with high constellation orders, the coded SDM OFDM schemes perform worse than the 1 × 1 case. Since the lattice of all possible points to decode becomes more and more complex when going to a higher constellation size and to more transmit antennas, the performance particularly suffers from non-orthogonal MIMO channels. As a result, the performance penalty for MIMO systems when going to a higher constellation order is worse than that for SISO systems.

Thirdly, in line with above observation, for high constellation sizes coded SDM OFDM systems suffer more from high rms delay spreads than SISO coded OFDM systems; compare Figure 5-22 and Figure 5-19, where we see that in Figure 5-22, for 250 ns, the ISI already causes an irreducible error-rate floor.

As we can already conclude from Figure 5-21, it appears that for some scenarios average SNRs per receive antenna of over 30 dB are required for a PER of 1%. For a lower rms delay spread the performance, due to less frequency diversity gain, will even be worse. For example, see Figure 5-23 where the BER and PER performance of a 2 × 2 system is shown for 16-QAM, coding rate ¾, and an rms delay spread of 10 ns. Also in this case it can be observed that for SOZF and SOMMSE SNRs of over 30 dB are required for a PER of 10^{-2}. Compared to 16-QAM, simulations have shown that 64-QAM would even require an SNR of approximately 34 dB which, when also considering, say, 5 dB extra for implementation losses, would lead to unrealistic requirements for low-cost implementation. One way to

reduce these required average SNRs per RX antenna and, thus, to relax the transceiver design, is by adding an extra receiver branch. For 16-QAM, coding rate ¾, and an rms delay spread of 10 ns, this results in the performance curves of Figure 5-24, where we observe a substantial diversity gain. Also from 2 × 3 simulations with 64-QAM we observed a considerable diversity gains and an SNR of only 26 dB is required for a PER of 1%. Furthermore, note that higher rms delay spreads will result in better performances, due to extra frequency diversity gain (as long as ISI does not occur).

Figure 5-23: BER and PER versus mean SNR per receive antenna for a 2 × 2 system in a Rayleigh fading channel with an exponentially decaying PDP with 10 ns rms delay spread for 16-QAM, rate ¾ convolutional coding (72 Mbps), 64 byte packets, and different SDM detection schemes. Also the 1 × 1 curves are given as reference.
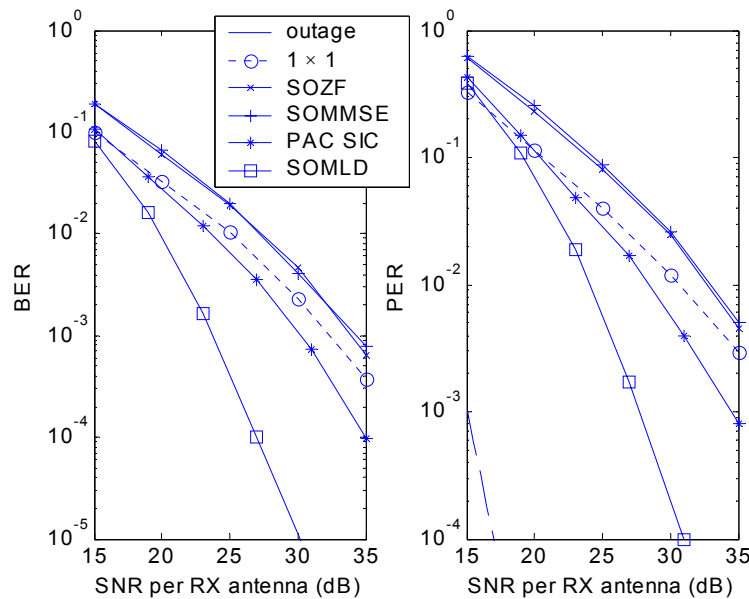
As already concluded at the end of Subsection 4.11.1, the SDM performance degrades for a given SNR when the channel is not Rayleigh faded but Ricean faded. Figure 5-25 and Figure 5-26 show the respective losses compared to Figure 5-23 and Figure 5-24 when systems with the same configuration undergo Ricean fading with a *K*-factor of 10. For these simulations, the wideband Ricean fading model described in Subsection 3.3.3 is used where the specular component is obtained by (3.31) with steering vectors with random phases as described by (3.33). When comparing Figure 5-23 and Figure 5-25, we clearly see that the SDM algorithms based on estimation techniques, i.e., SOZF, SOMMSE, and PAC SIC, require a 10 dB higher SNR to achieve a PER of 1%, whereas the algorithm based on detection, SOMLD, only loses 4 dB. For the 2 × 3 case, these losses are respectively 8 and 5 dB. So, apparently, SOMLD suffers the least from going to a scenario with a strong specular path. Finally, we can conclude from the 2 × 3 case that MIMO communication at an appreciable rate of 72 Mbps is still feasible in a Ricean fading environment with a *K*-factor as high as 10, even for SOZF and SOMMSE.
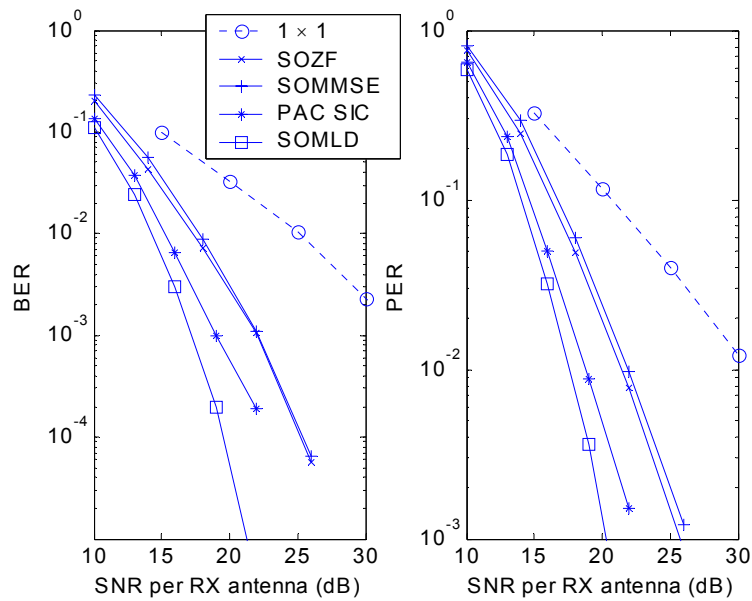
Figure 5-24: BER and PER versus mean SNR per receive antenna for a 2 × 3 system in a Rayleigh fading channel with an exponentially decaying PDP with 10 ns rms delay spread for 16-QAM, rate ¾ convolutional coding (72 Mbps), 64 byte packets, and different SDM detection schemes. Also the 1 × 1 curves are given as reference.

Figure 5-25: BER and PER versus mean SNR per receive antenna for a 2 × 2 system in a Ricean fading channel with $K = 10$ and an exponentially decaying PDP with 10 ns rms delay spread for 16-QAM, rate ¾ convolutional coding (72 Mbps), 64 byte packets, and different SDM detection schemes.
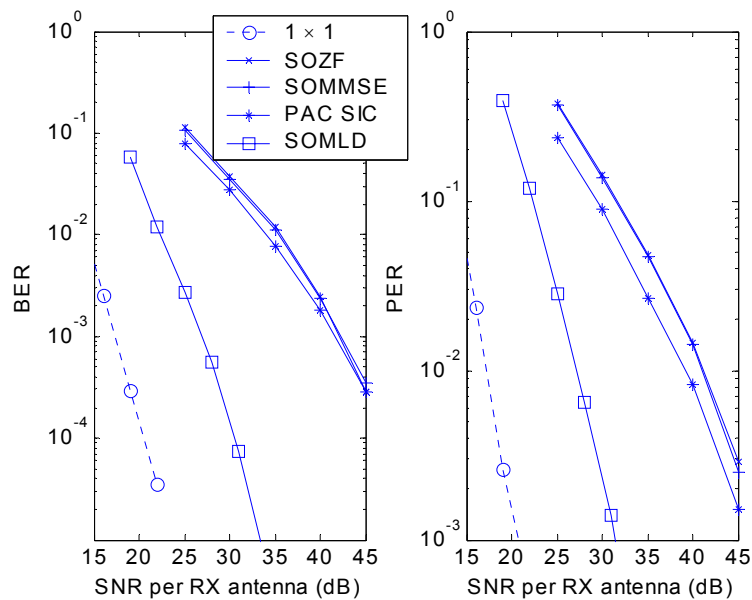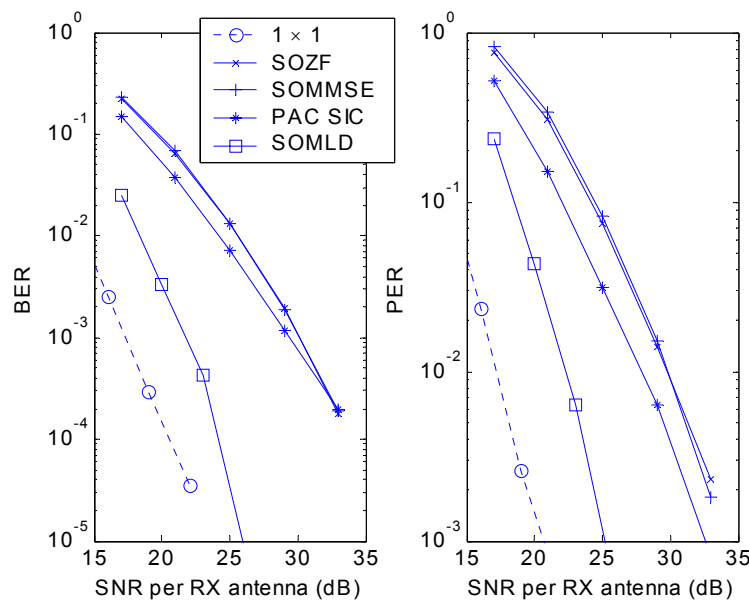
**Figure 5-26: BER and PER versus mean SNR per receive antenna for a 2 × 3 system
in a Ricean fading channel with $K = 10$ and an exponentially decaying PDP with 10 ns
rms delay spread for 16-QAM, rate ¾ convolutional coding (72 Mbps), 64 byte
packets, and different SDM detection schemes.**

5.9   Conclusions and Recommendations

The combination of SDM and OFDM is discussed in this chapter. First the general
principle of MIMO OFDM is explained and its signal model is introduced after which a
Space-Frequency analysis is performed. From this analysis, it is concluded that, when the
diversity order is high enough (four or more) and the SNR is low enough, maximising the
minimum Euclidean distance between any pair of codewords is a more appropriate design
rule than the rank and determinant criteria. As a result, under above conditions, standard
SISO codes together with proper form of space and frequency multiplexing outperforms
handcrafted Space-(Time-)Frequency codes.

The straightforward coded SDM OFDM schemes introduced in this chapter, based on
simply demultiplexing a one-dimensional code over the spatial and frequency dimension,
still perform at least 7 dB worse than the outage performance. 2.2 dB of the 7 dB, however,
is imposed by the OFDM overhead, i.e., the guard subcarriers and guard time. So,
apparently there is still enough room available for improvements, challenging us to find
more efficient ways of implementing the Euclidean distance criterion. When latency is not
the issue, a promising solution would be to introduce the turbo processing, as described for
narrowband MIMO transmissions in Section 4.13, to the MIMO OFDM context.

Simulations with different SDM detection algorithms for the proposed coded SDM OFDM
schemes showed that the introduced detection scheme for a PAC transmission, based on
SOMMSE and SIC in which the SIC information is fed through a Forward Error
Correcting loop (referred to as PAC SIC), achieves equivalent PER performances as the
complex SOMLD, but the complexity of PAC SIC only grows polynomial with the

number of TX antennas. The only disadvantage is that PAC SIC introduces extra latency, but for small interleaver and coding depths, this latency is manageable.

Furthermore, it is observed that, when an extra RX antenna is added (at least for the 2×2 case), the different schemes perform in a close range of each other. Moreover, the average performance is improved due to the extra antenna and diversity gain. For some symmetric MIMO schemes with a high constellation order this is shown to be a potential solution that shifts the required SNR to practical values and, as a result, relaxes the transceiver design. Other ways to improve the performance and to get closer to the outage PER may be found in more advanced code design that also benefits from the transmit diversity.

Based on above arguments, we can conclude that coded SDM OFDM is a promising scheme that combines the data rate enhancement and flexibility of SDM with the relatively high spectral efficiency and the robustness against frequency-selective fading and narrowband interference of OFDM. Within this concept, however, there is still room for improvement, namely, at a PER of 1% we observed a loss of 4.8 dB compared to the outage PER.

6

Implementation of a MIMO OFDM WLAN System

6.1 Introduction

In the previous chapters, in the analyses and evaluations we performed on MIMO (OFDM), we assumed perfect time and frequency synchronisation between the transmitter and receiver, and perfect knowledge of the channel, the packetlength, and data rate at the receiver. This is justified when the goal is to, e.g., compare and evaluate MIMO algorithms in controllable and idealised scenarios. When working towards practical implementations, we cannot make above assumptions anymore, but we have to deal with all sorts of impairments.

In this chapter, we will describe how these impairments can be tackled in practical applications of MIMO OFDM in the WLAN context. In order to restrict the solution space, we note that one of the key criteria for success of next generations of existing wireless communication products is coexistence and, maybe even more important, backwards compatibility with the current products. Hence, in order to introduce MIMO as an extension to a SISO product or standard, the definition of fallback mechanisms is essential. This compliance with existing and upcoming products or standards puts extra constraints on the TX signal format.

One of the current high data rate standards for WLAN is IEEE 802.11a. Since it is based on OFDM, it seems a perfect test case to apply the techniques described in the previous chapter. Keeping backwards compatibility in mind, we choose to follow the IEEE 802.11a TX signal format as closely as possible.

The basic Medium Access Control (MAC) protocol defined in the IEEE 802.11 standard (which also applies to IEEE 802.11a) is based on the unscheduled packet-switched principle. This means that there is no dedicated communication channel set up between the transmitter and receiver (i.e., circuit-switched), but the transmitter sends data on a packet-by-packet basis. In other words, WLAN is a packet-switched system with a random access

protocol. Consequently, the receiver does not have any a priori knowledge about packet-arrival times. The random nature of the arrival times and the high data rates require the synchronisation to be completed shortly after the start of the reception of a packet. To facilitate "quick" synchronisation, the data packet is preceded with a known sequence (i.e., the preamble). The preamble is carefully designed to provide enough information for a good start-of-packet detection, Automatic Gain Control (AGC), time synchronisation, frequency synchronisation, and channel estimation. Once these tasks have been performed, the payload can be processed.

In practical implementations, these synchronisation tasks are essential since they tackle system impairments, such as frequency offset and timing offset. Other system impairments that have influence on the performance are, e.g., phase noise, IQ imbalance, and DC offset ([143]). In the pervious chapter, the performance of various MIMO OFDM algorithms was compared (based on IEEE 802.11a parameters), however, impairments were not taken into account. Therefore, to validate MIMO OFDM algorithms in practice, a 3×3 test system was built within Agere Systems, The Netherlands.

In this chapter, data rate measurements are performed with this 3×3 test system. To tackle the impairments, frequency and time synchronisation, channel estimation, and phase tracking are implemented. Moreover, a specific receiver design based on the sampled-IF principle reduces the influence of IQ imbalance and DC offset. Successful transmissions up to 162 Mbps in the license-free 5.x GHz band are demonstrated in a typical office environment.

The organisation of this chapter is as follows. In Section 6.2, a description is given of the processing that is required for synchronisation of a MIMO OFDM system. The proposed preamble is an extension of the IEEE 802.11a preamble. Section 6.3 gives a description of the test system that has been built to evaluate the performance of different MIMO algorithms and synchronisation algorithms in real-life environments. The results of measurements that are performed are provided in Section 6.4. The chapter ends with conclusions in Section 6.5.

6.2   Implementation Description

6.2.1 Motivation

As already mentioned in the introductory section, synchronisation is an essential task for any digital communication system and required for reliable reception of the transmitted data. From the perspective of physical layer design, proper synchronisation algorithms are crucial to build a successful product. Therefore, for a potential application of MIMO OFDM in the WLAN context, it is important to examine the necessary changes to (e.g.) the IEEE 802.11a preamble, time and frequency synchronisation, channel estimation, synchronisation tracking and detection. The next subsections will give an overview of these necessary changes. A short description of the IEEE 802.11a preamble is provided in the next subsection.

6.2.2 The IEEE 802.11a Preamble

Since the IEEE 802.11a standard is based on the unscheduled packet-switched principle, the receiver needs to perform a start-of-packet detection (where the start of a packet is defined by the start of the preamble), Automatic Gain Control (AGC), time synchronisation, frequency synchronisation, and channel estimation before the actual data bits can be decoded. To that end, the standard defines a preamble by which above tasks can be performed. The preamble structure is shown in Figure 6-1.
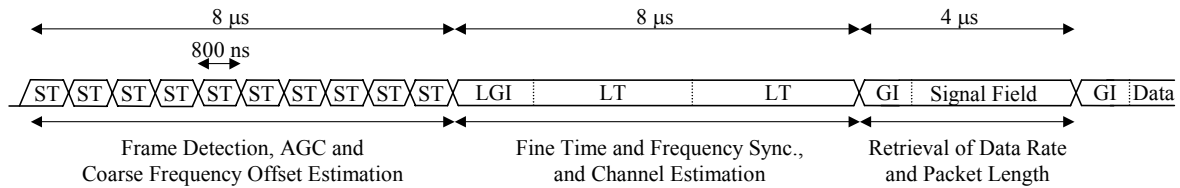
Figure 6-1: IEEE 802.11a preamble.

The first part of the preamble consists of 10 identical Short Training (ST) symbols with a duration of 800 ns each. These ST symbols are produced as follows. Only subcarriers with an index that is a multiple of 4 are bearing a predefined nonzero value ([57]). Hence, of all possible subcarrier numbers from –26 to 26, only the subset {–24, –20, –16, –12, –8, –4, 4, 8, 12, 16, 20, 24} is used ([126]). Taking the 64-point IFFT of the result leads to 64 samples containing 4 ST symbols. The ST symbols are used to perform Frame Detection (FD), AGC, and coarse frequency-offset estimation.

Firstly, the receiver is set up in listing mode, which means that it scans for the start of a packet using a Frame Detection (FD) / Coarse Timing (CT) algorithm. For IEEE 802.11a, the receiver can perform a running correlation between two subsequent intervals of 800 ns, which at a sampling rate of 20 MHz consist of 16 samples each. When a packet is received, the outcome of the correlator will go up and stay high for the duration of the ST symbols. In [102], three different criteria to determine the start of a packet (and thus the detection of a frame) based on the correlator output are compared:

1. The Maximum-Correlation criterion, proposed in [67]. In this criterion, the start of a frame is, based on a fixed time offset, directly related to the maximum of the magnitude of the correlation between two successive ST symbols. Usually, this criterion is combined with a threshold that triggers the search for the maximum and a search window that limits the search for the maximum to the length of the window.

2. The Schmidl criterion, proposed in [105]. To overcome numerical-range issues of the Maximum-Correlation criterion, Schmidl proposed a criterion in which the squared magnitude of the correlation between two successive ST symbols is normalised by the squared power of the received sequence windowed by a sliding window equivalent to the length of one ST symbol. The maximum of the result directly relates to the start of the frame.

3. The Maximum-Normalised-Correlation criterion, proposed in [81]. This criterion is a modification of the Schmidl criterion. The squared magnitude of the correlation

between two successive ST symbols is multiplied by four and normalised by the squared power of the received sequence windowed by a slicing window equivalent to two times the length of a ST symbol. And, again, its maximum indicates the start of a frame (after correction with a fixed time offset).

In [102], it was concluded that for high SNRs, the Schmidl criterion and Maximum-Normalised-Correlation (MNC) criterion perform the best. Depending on the performance measure, the MNC criterion in general was shown to be the better of the two.

Their relatively large number of repetitions allows multiple power measurements, obtained by correlating two successive ST symbols, to stepwise converge to the right AGC setting.

Furthermore, the short symbol period makes it possible to do coarse frequency offset estimation with a large unambiguous range. The repetition of ST symbols, namely, allows for estimation of the phase rotation between time-delayed versions of the same symbol, which is a measure for the frequency offset ([79]). For a repetitive signal with duration T, the range of the tolerable frequency offset, which is also referred to as lock-in range, is equal to $[-1/(2T), 1/(2T)]$, as frequency offsets outside this range result in a phase change exceeding $[-\pi,\pi]$ form one symbol to another, leading to ambiguity. Hence, with the ST symbols, absolute frequency offsets up to 625 kHz can be estimated, which is two times the subcarrier spacing (see Table 5-1).

The middle part of the preamble consists of two Long Training (LT) symbols. A single LT symbol is constructed by modulating the 52 subcarriers with the same subcarrier index as in a normal data symbol by known QPSK symbols and taking the 64-point IFFT. The two LT symbols are preceded by a Long Guard Interval (LGI) of 1.6 μs containing a copy of the last 1.6 μs of the LT symbol (see Figure 6-1). The repetition is used to perform fine frequency offset estimation ([102]), with a limited absolute frequency range up to 156.25 kHz. The advantage over the coarse frequency offset estimation is that in the fine frequency offset estimation the repetition period consists of more samples resulting in a more accurate (i.e., fine) estimate of the frequency offset. Furthermore, the two LT symbols are used for fine time synchronisation and channel estimation in order to obtain a reference amplitude and phase per subcarrier for doing coherent demodulation.

In order to deal with phase noise, the data in the payload of the packet is extended with 4 known pilot symbols per OFDM symbol. Based on the estimated phase offset of these pilot tones, the receiver can track the reference phase. The pilot symbols are scrambled by a length 127 pseudo-noise sequence ([57]) to avoid spectral lines exceeding the average power density of the OFDM spectrum ([126]).

A schematic representation of above described preamble and pilot tones is shown by the time-frequency structure of an OFDM packet in Figure 6-2 ([126]), where all known training values are marked grey. It clearly illustrates how the packet starts with 10 ST symbols, using only 12 subcarriers, followed by the LT sequence and OFDM data symbols, with each OFDM data symbol containing four known pilot subcarriers.

The last part of the preamble is the signal field, which contains information about the data rate and packetlength. This information is encoded in 16 binary bits. There is also a reserved bit and a parity check bit. These 18 bits are padded with 6 zeros (to allow the

Viterbi decoder to return to the zero state) and encoded with the convolutional encoder with coding rate ½. The resulting sequence is then interleaved and used to modulate the 48 data subcarriers using BPSK. Hence, the signal field is always sent at the lowest possible rate, namely 6 Mbps. After the 4 pilot symbols are added, the IFFT is taken of the result to obtain the time domain representation of the signal field. Finally, of course, a cyclic prefix is added as guard interval.

Further information on constellation sizes, coding schemes, etc., can be found in Subsection 5.8.1.
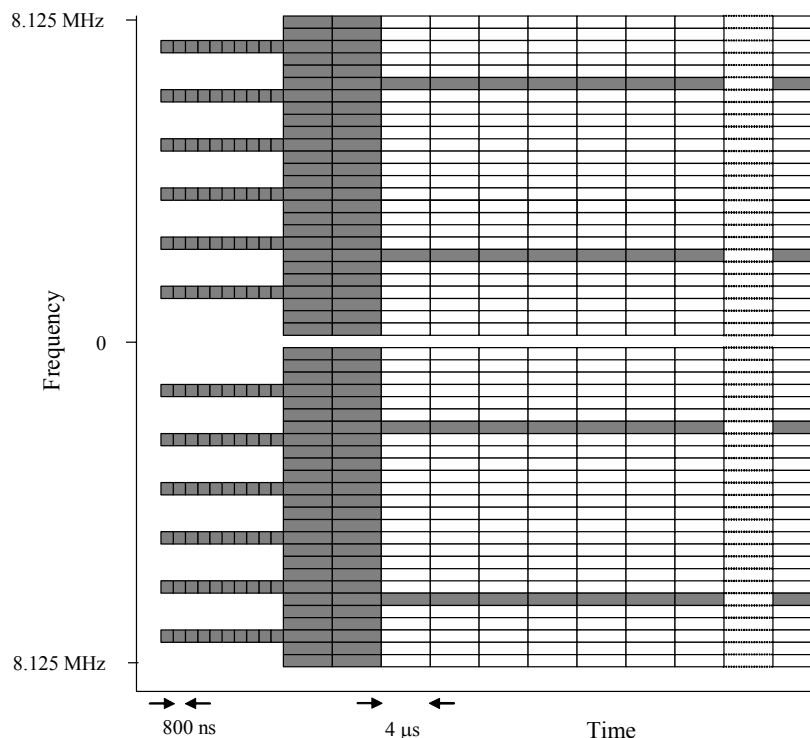
Figure 6-2: Time-frequency structure of an OFDM packet. The grey subcarriers contain known training values.

6.2.3 MIMO OFDM Preamble

In this subsection, the IEEE 802.11a preamble design as described in the previous subsection is extended to be able to perform synchronisation tasks in the MIMO OFDM case, while keeping the possibility to fall back to IEEE 802.11a rates (see also [135]). We will assume that the receiver gains (per branch) are already set by the AGC. Since the AGC for a MIMO system does not differ significantly from the SISO case, it will not be treated here.

The main difference between a SISO preamble and a MIMO preamble is that the latter should be designed such that it allows for estimation of all MIMO channel elements. In [14], it was shown that, in order to achieve the same accuracy per MIMO channel element as for the SISO case, the length of the training sequence must grow linearly with the

number of TX antennas. Once the MIMO channel is estimated, the MIMO processing can separate the signal components originating from the different transmit antennas.

To estimate the MIMO channel, it is important that the subchannels from the different TX antennas to every RX antenna can be uniquely identified. To achieve that it is shown in [14] that the preambles on the different TX antennas should be orthogonal and time-shift orthogonal, for at least the channel impulse response length. We have chosen for time orthogonality as shown in Figure 6-3.
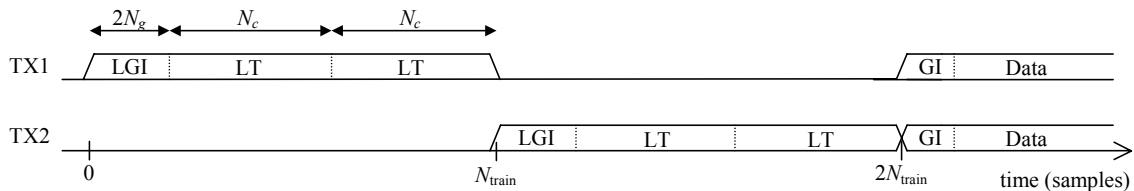
Figure 6-3: Concept of a time orthogonal preamble for a MIMO configuration with two transmit antennas.

Note that we ignored the Signal Field, since (for now) we will assume that both the transmitter and receiver know the data rate and packetlength. To include the Signal Field and maintain backwards compatibility, one could choose to place it between the LT symbols of TX1 and TX2, and use the reserved bit to indicate that the current transmission is a MIMO transmission.

To perform frequency offset estimation, a periodicity in the preamble is desired as explained in the previous subsection. Therefore, the proposed preamble consists of a concatenation of two identical training sequences per TX antenna.

Furthermore, to make the channel estimation less vulnerable to ISI, a LGI by means of a CP of length $2N_g$ is added. Altogether, this results in a preamble as depicted in Figure 6-3 for a 2×2 system. Following the IEEE 802.11a standard, the training sequence is chosen to be the standard's long training symbol, with $N_g = 16$ and $N_c = 64$ ([57]), resulting in a non-zero part per TX antenna of $N_{train} = 2N_g + 2N_c$ samples. In the remainder of this chapter, we will denote the signal on the i-th subcarrier of the long training symbol by $c(i)$.

Note that since the total length of the proposed preamble grows linearly with N_t, it is not highly efficient. More efficient channel training sequences for MIMO OFDM are proposed in [14]. It is claimed that training sequences for the different transmit antennas are optimal if they are orthogonal, as well as shift orthogonal over the sampling interval $\{-L + 1, \ldots, L - 1\}$. Since the impulse response length L is not known in advance, the simplest implementation would be to choose a fixed orthogonality interval. If the shift orthogonality is not satisfied, however, training following the design rules of [14] becomes unstable, leading to undesirable performance degradation. Therefore, stable but efficient channel estimation algorithms remain a topic for further research.

Finally, note that the preamble of Figure 6-3 does not contain ST symbols. In general, we could say that when it is assumed that the TX antennas are co-located and that the RX antennas are co-located, then the FD and coarse frequency offset estimation can be done in the same way as for the SISO case. The best thing to do is to send the SISO ST symbols on

all TX antennas simultaneously. Note that in an environment that results in high spatial correlation this might result in a beamforming kind of transmission. Therefore, either the TX antennas should be designed such that their antenna patterns are as orthogonal as possible, or some form of transmit diversity must be applied. Since we, however, assume that the AGC is already set, in the next subsections the ST symbols are discarded and the Frame Detection is performed on the LT symbols. When one wants to include the ST symbols to the preamble, the adaptation of the proposed algorithms of the next subsections is straightforward.

6.2.4 Time Synchronisation

Frame Detection / Coarse Timing

The task of FD is to identify the preamble in order to detect a packet arrival. This preamble detection algorithm can also be used as a CT algorithm, since it inherently provides a rough estimate of the starting point of the packet. In literature different data-aided FD algorithms have been proposed for OFDM ([67, 81, 105]). A simple MIMO extension of Schmidl's algorithm ([105]) was proposed in [78]. All these algorithms are based on the correlation between the repeated symbols constituting the preamble.

Based on the assumption at the end of Subsection 6.2.3 that we will discard the ST symbols, the next equations are derived from the preamble design of Figure 6-3. We define the complex correlation Λ between two subsequently received frames of N_c samples on the N_r receive antennas as

$$\Lambda(\tau) = \sum_{n=\tau-(N_c-1)}^{\tau} \mathbf{r}^H(n-N_c)\mathbf{r}(n) = \sum_{n=\tau-(N_c-1)}^{\tau} \sum_{q=1}^{N_r} r_q^*(n-N_c)r_q(n). \qquad (6.1)$$

where $r_q(n)$ is the n-th sample of the received signal on antenna q. The sum of the power of N_c subsequently received samples on the N_r receive antennas together, denoted by P, is defined as

$$P(\tau) = \sum_{n=\tau-(N_c-1)}^{\tau} \mathbf{r}^H(n)\mathbf{r}(n) = \sum_{n=\tau-(N_c-1)}^{\tau} \sum_{q=1}^{N_r} r_q^*(n)r_q(n). \qquad (6.2)$$

Here we propose to use the Maximum-Normalised-Correlation (MNC) criterion for FD, which was proposed in [81] for a SISO OFDM system and shown to perform better than techniques proposed in [67] and [105] (see also Subsection 6.2.2). In the MIMO extension of the MNC frame detection algorithm the estimated start of the data packet, i.e. the end of the preamble, is given by

$$\tau_{\mathrm{FD}} = \arg\max_{\tau} \left(\frac{\sum_{p=1}^{N_t} 4|\Lambda(\tau_p)|^2}{\sum_{p=1}^{N_t} P(\tau_p - N_c)^2 + P(\tau_p)^2} \right), \qquad (6.3)$$

where $\tau_p = \tau - (N_t - p)N_{\text{train}}$ to consider the offset of N_{train} per TX branch (see Figure 6-3). To avoid a false detection, i.e. a detection of a packet when none is present, a threshold should be set which triggers the above algorithm.

Symbol Timing

The symbol timing in an OFDM system decides where to place the start of the FFT window within the OFDM symbol. Although an OFDM system exhibits a Guard Interval (GI), making it somewhat robust against timing offsets, non-optimal symbol timing will cause more ISI and ICI in delay spread environments. This will result in performance degradation.

The symbol timing proposed here is designed to minimise the amount of ISI and ICI that is generated in the system and is an extension of the technique proposed for SISO OFDM in [126, pp. 88-92]. Since it relies on the knowledge of the Channel Impulse Responses (CIRs), their powers are estimated by correlating the received signals with the known training sequence (in time domain, thus after applying the IFFT on $c(i)$)

$$\eta_q(\tau) = \left\| \sum_{k=0}^{N_c-1} r_q(\tau - k) \left(\frac{1}{\sqrt{N_c}} \sum_{i=0}^{N_c-1} c^*(i) \exp\left(-j2\pi \frac{i(N_c - 1 - k)}{N_c} \right) \right) \right\|^2 . \tag{6.4}$$

Subsequently the powers of the $N_t N_r$ impulse response estimates are summed. Note that the estimates of the powers of the CIRs corresponding to the N_t transmitters are spaced N_{train} samples apart in η_q, and that η_q holds two estimates per TX spaced N_c samples apart, caused by the repetition in the preamble. The sum of the powers is window integrated over the length of the Guard Interval of the OFDM symbol N_g. The joint symbol timing for the whole MIMO receiver τ_{est}, i.e., a measure for the start of the payload, is then found by searching the maximum of the window integral:

$$\tau_{\text{est}} = \arg\max_\tau \sum_{l=0}^{N_g-1} \sum_{p=1}^{N_t} \sum_{q=1}^{N_r} \sum_{n=0}^{1} \eta_q(\tau_p - l - nN_c). \tag{6.5}$$

Note that, to reduce the amount of computations for the symbol timing, a search window can be defined based on the coarse timing. This search window should be centred around $\tau = \tau_{\text{FD}}$, as determined by (6.3).

6.2.5   Frequency Synchronisation

The frequency synchronisation has to correct for the Frequency Offset (FO) caused by the difference in oscillator frequencies at the transmitter and the receiver. We estimate this frequency offset and compensate the received signals for it. The frequency offset can be estimated using the phase of the complex correlation Λ between the two consecutive received training symbols (defined in (6.1)), as was shown for a SISO OFDM system by Moose in [79] and further worked out in [105]. A simple MIMO extension of Moose's algorithm was proposed in [78]. Here it is assumed that all transmit/receive branches of one MIMO transmitter/receiver use the same oscillator, which is a valid assumption if the

different transmit/receive branches are co-located. The best instant to estimate this frequency offset is at timing instant $\tau = \tau_{FD}$. The estimated frequency offset Δf_{est} is then given by

$$\Delta f_{est}(\tau) = \frac{\theta_{est}(\tau)}{2\pi T_c} = \frac{f_s \cdot \angle\left(\sum_{p=1}^{N_t} \Lambda(\tau_p)\right)}{2\pi N_c} , \qquad (6.6)$$

where θ_{est} denotes the phase of the summation of the complex correlations Λ of the training symbols originating from the different transmitters, f_s equals the sample frequency, $T_c = N_c/f_s$ stands for the training symbol duration, and $\angle(.)$ denotes the angle (in rad) of the corresponding argument. The maximum estimated frequency offset is limited, since the angle θ_{est} that can be estimated without phase ambiguity is limited to $\theta_{max} = \pm\pi$. This relates to a maximum frequency offset of $|\Delta f_{max}| = |\theta_{max}|f_s/(2\pi N_c) = f_s/(2N_c)$, which equals half the subcarrier spacing. A larger range can be achieved by first performing a coarse frequency offset estimation using shorter sequences, e.g., the ST symbols as defined in the IEEE 802.11a standard ([57], Subsection 6.2.2).

As an accuracy measure of the estimator we consider the normalised variance of the estimator in an AWGN channel as was derived in [79] for the SISO estimator. The normalised variance of the MIMO estimator is derived in [103] and given by

$$\text{var}\left(\frac{\Delta f_{est} - \Delta f}{\Delta F}\right) = E\left(\frac{\Delta f_{est} - \Delta f}{\Delta F}\right)^2 = \frac{1}{(2\pi)^2 N_r N_c \rho} , \qquad (6.7)$$

where ρ denotes the SNR per receive antenna, Δf the actual frequency offset, and $\Delta F = f_s/N_c$ the subcarrier spacing. It can be shown that the variance of the estimator equals the Mean Square Error (MSE), since the estimator is unbiased. It is clear from (6.7) that the accuracy increases linearly with the number of receive antennas N_r. Note that the Cramér-Rao lower bound is equal to this theoretical value of the variance, as was already concluded for the SISO version in [105].

In Figure 6-4 the MSE of the FO estimation is depicted as function of the average SNR per receive antenna. The theoretical value from (6.7) is shown together with simulation results averaged over 100,000 channel realisations for both AWGN channels and Rayleigh faded channels with an exponentially decaying PDP having different values of rms delay spread. The figure shows the results for the SISO and 3×3 configuration. It is clear that the theoretical value is a good estimate of the MSE for high SNR values, but underestimates the MSE compared to simulation results for low SNR. In case of AWGN, we see an improvement of 4.8 dB in performance going from a 1×1 to a 3×3 configuration, as expected from (6.7).

The simulations with multipath channels show a degradation in performance with respect to the theoretical and the simulated AWGN case. The degradation is the worst for the flat fading case and decreases when the rms delay spread increases. This is explained by the frequency diversity introduced by the delay-spread channel. The degradation, however, is much smaller for the 3×3 case than the SISO case. This shows that the frequency-offset

estimation in MIMO systems is more robust under different fading conditions, which can be explained by the space diversity introduced by the multiple antennas.
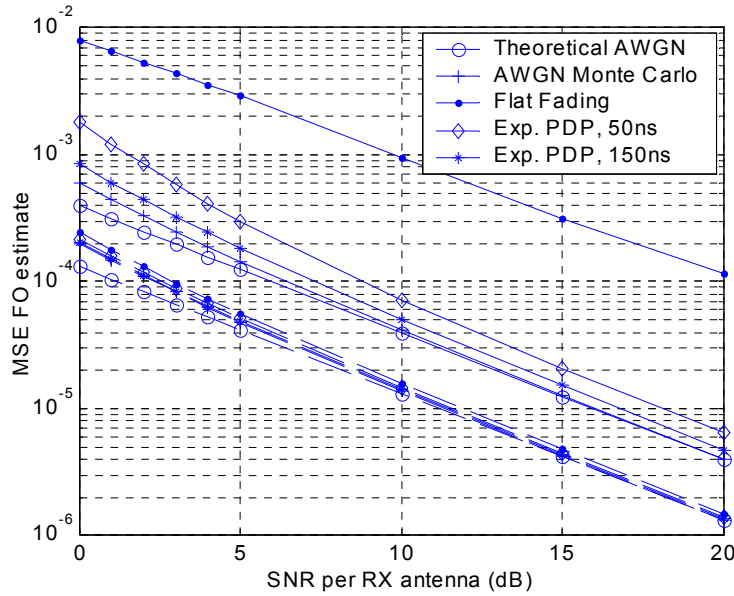
Figure 6-4: Mean Square Error of the frequency offset estimate for a 1 × 1 (solid lines) and a 3 × 3 (dashed lines) system from theory and simulations with AWGN and multipath channels.

6.2.6 Channel Estimation

When time synchronisation is performed at the receiver and after the received signals are corrected for the frequency offset, the channel can be estimated using the known training symbols within the preamble. When the timing is performed correctly, we know which received samples correspond to the training part. More precisely, we know exactly which part of the received preamble is sent by transmit antenna p (see Figure 6-3). Therefore, in the analysis of this subsection, we will omit the MIMO OFDM symbol index for brevity. Let us denote the Fourier transform of the symbol received on antenna q that corresponds to the training symbol from transmitter p by

$$\mathbf{x'}_q^p = \begin{pmatrix} x_q^p(0) \\ \vdots \\ x_q^p(N_c - 1) \end{pmatrix}. \tag{6.8}$$

Recall that $c(i)$ denotes the training signal on the i-th subcarrier. Then, rewriting the signal model of Section 5.4, it can be shown that

$$\mathbf{x'}_q^p = \mathbf{D} \begin{pmatrix} h_{qp}(0) \\ \vdots \\ h_{qp}(N_c - 1) \end{pmatrix} + \mathbf{n'}_q^p, \tag{6.9}$$

where $\mathbf{D} = \text{diag}(\mathbf{c})$, with $\mathbf{c} = (c(0), \ldots, c(N_c - 1))^T$. Furthermore, $\mathbf{n'}_q^p$ represents the frequency domain additive noise at the q-th receiver branch.

Now the estimates of the channel coefficients can be obtained by

$$\mathbf{h'}_{qp,\text{est}} = \begin{pmatrix} h_{qp,\text{est}}(0) \\ \vdots \\ h_{qp,\text{est}}(N_c - 1) \end{pmatrix} = \mathbf{D}^\dagger \mathbf{x'}_q^p . \tag{6.10}$$

Since $\mathbf{D}$ is diagonal and the IEEE 802.11a training symbol has a constant amplitude of 1.0 over the (used) subcarriers ([57]), the processing of (6.10) can be simplified to

$$\mathbf{h'}_{qp,\text{est}} = \mathbf{D}^H \mathbf{x'}_q^p . \tag{6.11}$$

Note that, because the proposed preamble contains a repetition of two identical training symbols per TX, we can improve the accuracy of the channel estimate by averaging over the two corresponding received symbols before the above described processing is performed. Furthermore, note that performing the channel estimation in the time domain has a potential higher performance, since in the time domain only an impulse response with a length not much higher than N_g has to be estimated and, because $N_g < N_c$, the available observations are effectively used to estimate a smaller number of values, i.e., the channel impulse response ([52]). On the other hand, we could also apply an interpolation filter on the result of (6.10) to average the estimation noise over frequency domain estimates and obtain a higher accuracy. Either of these techniques, however, is more sensitive to high timing offsets and high delay spreads due to extra constraints that are placed on, e.g., the observation length. Therefore, the more robust implementation of (6.10) has been chosen.

6.2.7 Synchronisation Tracking using Pilot Subcarriers

As described in the above sections, the processing of the preamble takes care of the initial synchronisation of the MIMO OFDM receiver. It is, however, likely that the frequency offset will vary during the reception of the packet due to, e.g., Phase Noise (PN), making solely initial frequency synchronisation insufficient. The observed PN at baseband is the residual error of the phase tracking of the RF oscillator in combination with a Phase Locked Loop (PLL). From [9] it is clear that PN causes a common phase turn for all subcarriers, called Common Phase Error (CPE), and a Gaussian like ICI term. Since the CPE is equal for all subcarriers, it can be estimated and corrected for. An initial estimation is done inherently by the channel estimation. The CPE, however, changes on a symbol-by-symbol basis, making this initial correction insufficient.

It is, therefore, necessary to estimate and correct the rotation of the received constellation points caused by the FO and CPE on a symbol-by-symbol basis. A convenient method is to use pilot subcarriers, i.e. subcarriers containing known data, which are inserted into the

data symbols. The rotation of these known pilot symbols, observed at the receiver, is a good measure for the CPE. An estimate of the rotation for the k-th symbol is given by

$$\vartheta(k) = \angle \left(\sum_{i \in \wp} \sum_{p=1}^{N_t} s_{p,\mathrm{est}}(i,k) s_p^*(i,k) \right), \tag{6.12}$$

where $\wp$ denotes the set of pilot subcarrier numbers of the N_p pilots per TX antenna and $s_{p,\mathrm{est}}(i,k)$ is the estimate, before slicing, of the transmitted signal $s_p(i,k)$. This rotation can now be corrected for at the receiver by multiplying the complex values of all subcarriers of the k-th symbol by $\exp(-j\vartheta(k))$.

6.3 The TRIO Test System

6.3.1 Introduction

In order to validate the principle of MIMO (OFDM) in practice, a MIMO test system was developed and built within Agere Systems, The Netherlands ([27, 128, 136, 135]). More concretely, this test system was developed with the following main goals: 1) to verify the theoretical MIMO propagation studies and 2) to serve as a platform for MIMO algorithm development. The system exploits the spatial dimension through three transmit branches and three receive branches and, hence, forms a 3 × 3 system. It is called the TRIO test system, which stands for TRiple-Input-Output test system.

With the test system, transmissions over (wideband) MIMO channels can be performed, e.g., for algorithm selection and development. Additionally, measurements can be used to verify propagation studies. As a platform for algorithm development, different approaches for MIMO decoding and synchronisation can be explored quickly by keeping the processing off-line to avoid implementation problems. Once a promising algorithm is found, efforts could be made to implement it in the test system to demonstrate the algorithm in real-time.

In the next subsections, a general description of the test system is given as well as more detailed descriptions of the specific transmitter and receiver hardware.

6.3.2 Configuration

To have enough flexibility in achieving the goals brought up above, the test system was composed with in-house developed components. To access this dedicated hardware, two Personal Computers (PCs) are used as TX and RX platform. Each PC contains three boards, where every single board represents a transmitter or receiver branch, resulting in a 3 × 3 MIMO system as schematically depicted in Figure 6-5. Every board consists of a baseband part, an Intermediate Frequency (IF) part and a Radio Frequency (RF) front-end based on a 5.x GHz GaAs radio chip. The test system operates in the 5-GHz band, and is capable of transmitting broadband signals with a bandwidth up to 20 MHz, which gives a time resolution of 50 ns. To a large extent the system design corresponds with the IEEE

802.11a standard ([57]). Taking into account the guard subcarriers, i.e. the zero subcarriers on the edges of the spectrum, the −3 dB bandwidth is 16.56 MHz.

For the moment, the TX and RX platforms are frame synchronised with a synchronisation cable, meaning the RX starts receiving when the TX triggers it. When the test system is upgraded in the future synchronisation processing will be implemented at the receiver and synchronisation is performed by transmitting a known preamble or training sequence (see Section 6.2). This makes the synchronisation cable redundant.
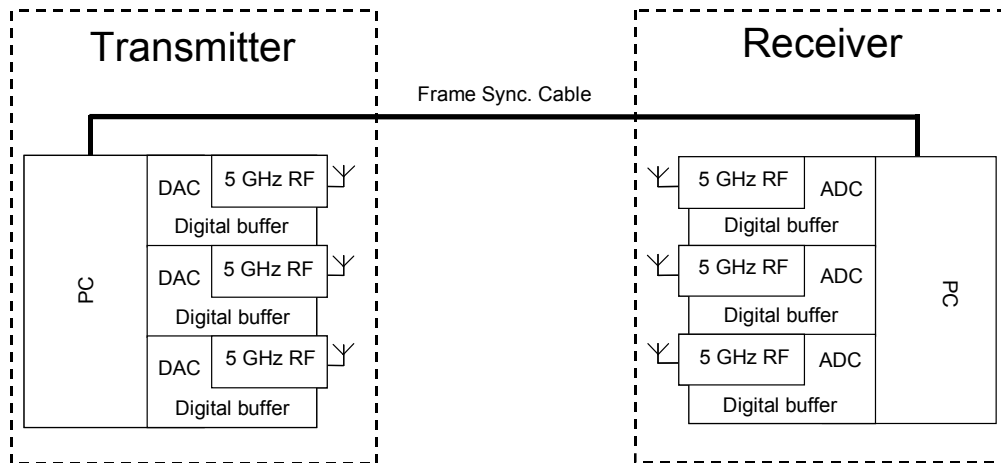
Figure 6-5: Diagram of the TRIO test system.

Both PCs are equipped with wireless LAN cards so that they can be accessed remotely. Additionally, for both PCs a program with a DCOM interface is developed and installed. This enables the user to operate the TRIO platform remotely via a Local Area Network. All the remote software used to compose transmit sequences, to operate the test system, and to perform various analyses/measurements is designed in MATLAB. For instance, by transmitting multiple packets, BER/PER tests can be performed in order to evaluate different MIMO algorithms in real-life communication channels.

Both the transmitter and receiver hardware platform are stalled on trolleys and connected to battery packs (UPS's). This enables the system to be moved freely and to operate at every location without the continuous need for external power supplies. Pictures of the receiver equipment are given in Figure 6-6.

6.3.3 Matching Transmitter and Receiver Hardware

The baseband processing is built around two Field Programmable Gate Arrays (FPGAs) per board with a possible extension of a third FPGA. In Figure 6-7 a schematic representation of a baseband board is presented. The FPGAs that are used are Xilinx Virtex 800's. The FPGAs can be reprogrammed and/or accessed via the ISA bus of the PCs very easily, providing a flexible baseband solution. The baseband boards are capable to both transmit and receive, while (currently) the IF and RF part are only set up to transmit or receive, respectively, for the TX and RX platform. Hence, at the moment, the communication stream is one way only.
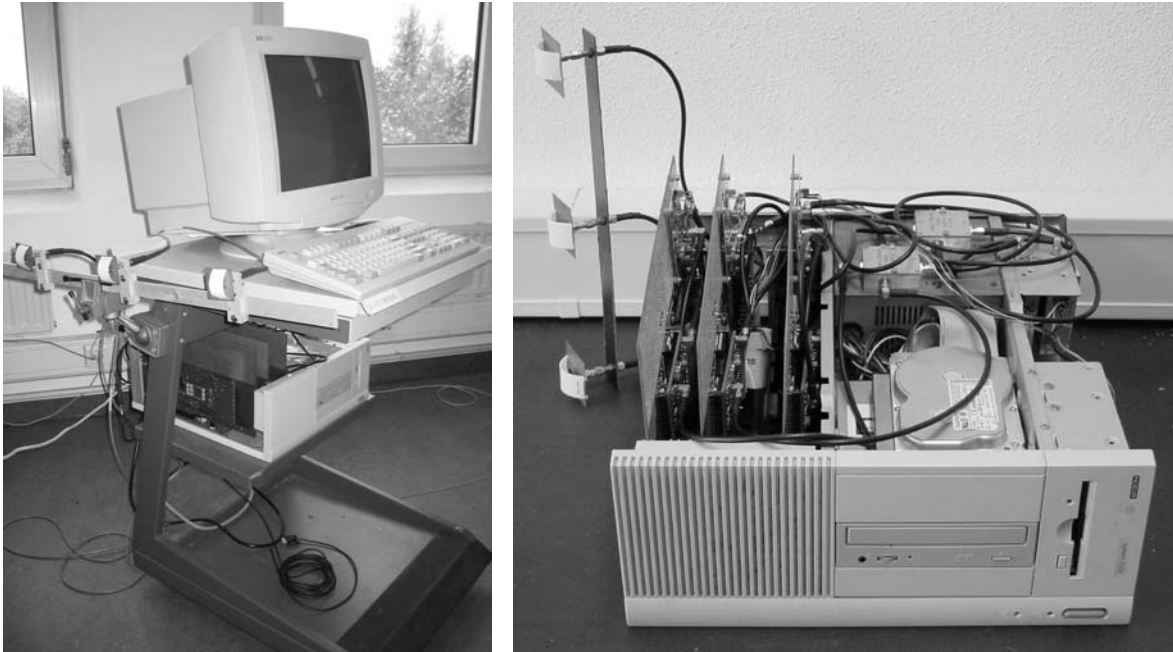
Figure 6-6: Receiver equipment, built around a PC with three receiver boards (i.e., the three receiver branches).
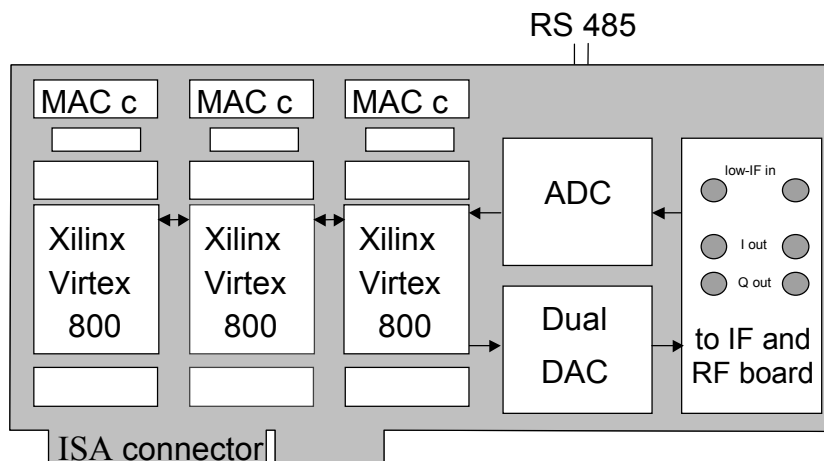
Figure 6-7: Schematic representation of a TRIO baseband board.

As mentioned before, the FPGAs provide a flexible baseband solution. It is, for instance, possible to load the FPGAs with a real-time implementation of the baseband processing. In this way, the boards can run stand-alone and real-time. Moreover, if they are connected to a Medium Access Control (MAC) board using the MAC connectors (denoted by "MAC c" in Figure 6-7), a complete protocol can be tested.

To be able to "quickly" compare and evaluate different algorithms, however, we make use of the possibility to program the FPGAs as memory banks and perform the necessary processing off-line. At the transmitter, waveforms can be loaded into and sent from the memory banks and they can be recorded at the receiver. These recorded data can then be processed off-line by software (so not in real-time) to not only compare different MIMO

algorithms, but also calibrate the system, perform channel measurements, perform capacity analysis, etc.

The RS 485 connector that is shown in Figure 6-7 is used (for the moment) as an interface between the TX and RX to do the frame synchronisation that is done per cable for the moment.

The antennas used in the system are ASCOM MBA-5, 5 GHz Miniature Broadband antennas. They radiate (in azimuth) a dual-half-plane wideband pattern and are preliminary designed for wireless LAN applications in the 5 GHz band. Information regarding the bandwidth and radiation pattern can be found in Appendix B.1. The antennas are mounted on a linear wooden rail that allows for gradual altering of the antenna spacing.

6.3.4 Transmitter Specific Hardware

The transmitter sends signals at zero-IF, meaning that it sends baseband signals centred around 0 Hz. To transmit such a signal, an in-phase (I) and quadrature (Q) part must be available, i.e., the baseband signal is complex and its real and imaginary parts are needed for further processing. The I and Q signals are converted to the analogue domain by 10-bit Digital-to-Analogue Converters (DACs) and after filtering sent to the IF and RF stages. At these stages the synthesisers (steered by local oscillators (LOs)) up-convert the signals to IF, 1.489 GHz, and subsequently to the RF carrier frequency, 5.150 GHz, and then the signals are fed to the TX antennas by low-loss coaxial cables. A block diagram of the components of baseband to IF stage is presented in Figure B-4 of Appendix B.2. The RF signal is amplified by power amplifiers (PAs), which have a maximum gain of 22 dB, to extend the range of the system. The output power of a transmitter is set to 15 dBm. This corresponds to about 32 mW. The cables that connect the transmitters with the antennas are 1 metre long and have a loss of 2 dB/m. Assuming no antenna loss, the antennas thus transmit with 13 dBm, or about 20 mW.

Per branch, the FPGAs are programmed as two (I and Q) 4096 10-bit word memory banks. Furthermore, the sampling rate at TX is set to 40 MHz. With respect to the baseband sampling rate as defined by the IEEE 802.11a standard, namely 20 MHz ([57]), this means oversampling by a factor of 2. So using the principle of oversampling, effectively, the transmitter can send 2048 words at 20 MHz simultaneously per branch. Since an IEEE 802.11a OFDM symbol consists of 80 words, each buffer can contain a maximum of 25 OFDM symbols.

6.3.5 Receiver Specific Hardware

At the receiver, the received signals are down-converted to low-IF signals, meaning that the baseband board gets a signal that is centred around a low Intermediate Frequency. For the TRIO system it is chosen to be 15 MHz. The RF signal is first down-converted from 5.150 GHz to 1.489 GHz and subsequently a 1.474 GHz LO converts it down to 15 MHz. The signals are then sampled and converted to the digital domain by 10-bit Analogue-to-Digital converters. The down-conversion to baseband is done off-line in the digital domain.

The components for the conversion of IF to low-IF are sketched in the block diagram in Figure B-5 of Appendix B.2.

The advantage of this generally called sampled-IF principle is that the DC-component can be easily filtered out and that hardly any IQ imbalance is introduced. The baseband transmitter, however, is sending at zero-IF, so IQ mismatches might occur, but since this then only happens at one side of the link, namely at TX, it can easily be calibrated out. In the memory banks of a receiver branches 8192 10-bit words can be stored. The sample rate of the ADC is 60 MHz. After the digital down conversion, the equivalence of the memory size is 2730 words at 20 MHz sampling rate.

Each receiver branch is also equipped with a Low Noise Amplifier (LNA) and an Automatic Gain Control (AGC). Depending on the field strength on the receive antennas the LNAs can be switched on or off. They have a gain of 23 dB and a noise figure of 1.5 dB. The AGCs have an attenuation range of 1 to 31 dB. An LNA (that can be switched on and off) in combination with an AGC allows the system to exploit the full resolution of the ADC for a wide range of field strengths at the receive antennas, being the dynamic range of the system.

6.4 Measurements

Measurements with the above described test system were done in a wing on the third floor of the Agere Systems building in Nieuwegein, The Netherlands, which can be regarded as a typical office environment. The wing is 42 m × 12.7 m and its floor plan is shown in Figure 6-8. The inner walls of the wing are modular walls, while the rest of the walls and floors consists of concrete. Besides the walls, the desks and metal cupboards are depicted. Furthermore, the RX and 9 different TX positions and orientations that were used for the measurements are shown. The antenna spacing was fixed at two wavelengths. The average SNR per RX antenna and the rms delay spread that we measured per location are given in Table 6-1. These values were measured at baseband level and, thus, include system influences. This means the estimated rms delay spread includes the influence of filters at both transmitter and receiver. Since the preamble for 3 TX antennas, as logic extension of the preamble shown in Figure 6-3, consumes 6 OFDM symbol lengths, the remaining memory of the transmitter has a storage capacity of 19 MIMO OFDM data symbols which we always fully exploited, leading to different packetlengths for different data rates. The tested data rates and corresponding packetlengths, i.e. the number of information bits, are given in Table 6-2. To obtain the average performance, 1000 packets were transmitted per rate. We applied per-antenna coding at the transmitter and the well-performing PAC MMSE with SIC algorithm introduced in Subsection 5.7.3 as receiving scheme. The BER and PER performance for different locations and different rates are depicted in Figure 6-9, Figure 6-10, and Figure 6-11. Note that for certain rates no performance values are given; this corresponds to the fact that at these rates all 1000 packets were received correctly. Furthermore, note that the 1 × 1 measurements were only performed for locations 5, 6 and 7 and were obtained with the same test system by switching off two of the three boards at both TX and RX.
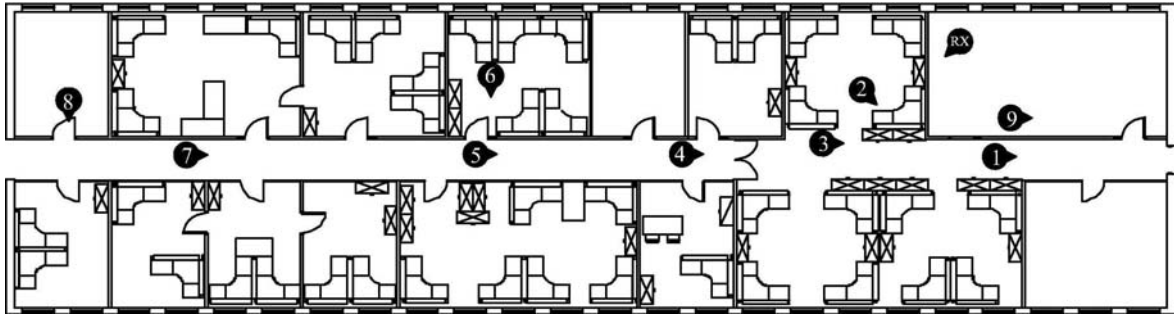
Figure 6-8: Floor plan of the wing where the measurements were performed, including the RX and TX locations and orientations.

Table 6-1: Average SNR per RX antenna and rms delay spread versus position.

Position	Average SNR (dB)	rms delay spread (ns)
1	24	98
2	24	97
3	24	100
4	23	104
5	25	91
6	24	100
7	22	109
8	14	131
9	26	100

Table 6-2: Data rates and packetlengths

Data rate (Mbps)	Antenna setup	Modulation	Coding rate	Packetlength (Bytes)
6	1×1	BPSK	1/2	57
9	1×1	BPSK	3/4	85.5
12	1×1	QPSK	1/2	114
18	1×1	QPSK	3/4	171
24	1×1	16-QAM	1/2	228
36	1×1	16-QAM	3/4	342
48	1×1	64-QAM	2/3	456
54	1×1	64-QAM	3/4	513
18	3×3	BPSK	1/2	171
27	3×3	BPSK	3/4	256.5
36	3×3	QPSK	1/2	342
54	3×3	QPSK	3/4	513
72	3×3	16-QAM	1/2	684
108	3×3	16-QAM	3/4	1026
144	3×3	64-QAM	2/3	1368
162	3×3	64-QAM	3/4	1539
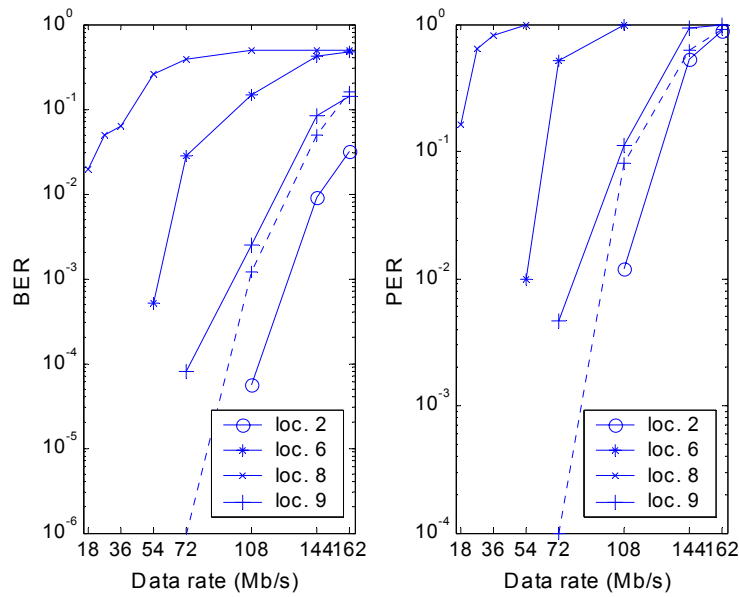
Figure 6-9: Measurement results of a 3 × 3 system with PAC SIC detection for the TX locations 2, 6, 8 and 9. For location 9, the measurements are compared with simulations (dotted line), with an average SNR per receive antenna of 26 dB and an rms delay spread of 100 ns.
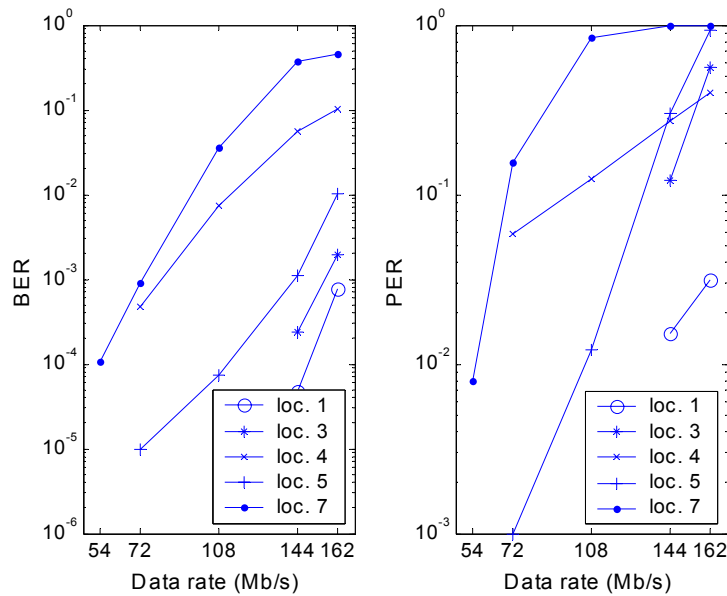
Figure 6-10: Measurement results of a 3 × 3 system with PAC SIC detection for the TX locations 1, 3, 4, 5 and 7.
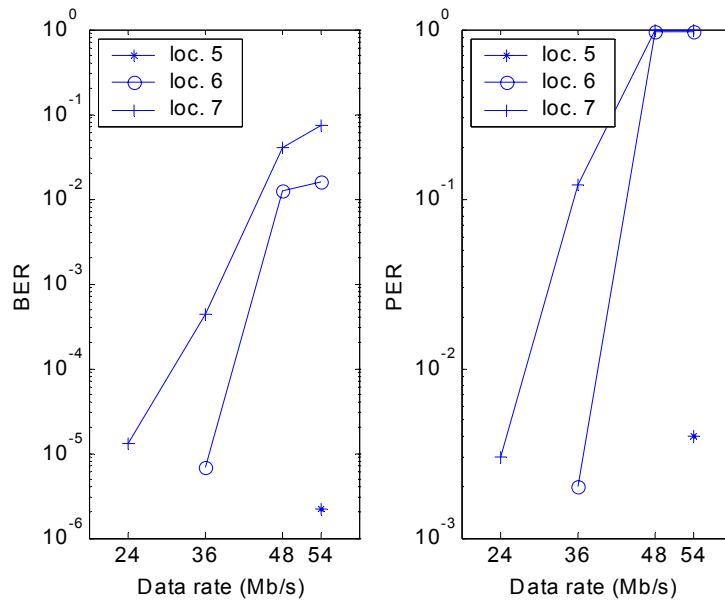
Figure 6-11: Measurement results of a 1 × 1 system with PAC SIC detection for the TX locations 5, 6 and 7.

In Figure 6-9, Figure 6-10, and Figure 6-11, it can be observed that the performance of the 3 × 3 measurements is worse than the performance of its 1 × 1 counterpart with 1/3 of the data rate. To have a PER of 1%, it can be shown that, in the 1 × 1 case, a rate of 54, 36, and 24 Mbps can be achieved for locations 5, 6, and 7, respectively. For the 3 × 3 case, these rates are respectively 108, 54 and 54 Mbps, resulting in an average data rate enhancement of 1.92. For well-conditioned MIMO channels, i.e. channels with i.i.d. channel elements, the data rate could theoretically be improved by a factor of 3. Clearly, the measured MIMO channels do not provide this improvement. The following arguments can explain this fact: firstly, the environment does not provide enough scattering, leading to ill-conditioned MIMO channels. Secondly, mutual coupling between the branches at the transmitter and the receiver leads to performance degradation. Since our system is not shielded very well, as can be seen from Figure 6-6, mutual coupling most likely has a high contribution to the observed performance/throughput loss in our case. Finally, system impairments such as RF non-linearities and quantisation noise might be an issue, though, in Subsections 6.2.4 and 6.2.5, and in [103] we observed that MIMO correction algorithms for impairments in general perform better than their SISO counterparts.

Next to measurements, simulations were performed to evaluate the Packet Error Rate (PER) performance for the 3 × 3 rates 72, 108, 144, and 162 Mbps (see Table 6-2) and to compare the results with the measurements results. Location 9 was used for this comparison (see Figure 6-8). At this position we observed an rms delay spread of about 100 ns and an average SNR per RX antenna of 26 dB. These parameters were used in the simulations. Furthermore, perfect synchronisation and i.i.d. channel elements were assumed and the PER was obtained by averaging over 10,000 packets. The results are shown in Figure 6-12. The solid curves are obtained by performing simulations including channel estimation based on the preamble design of Figure 6-3. The dashed curve is generated assuming perfect CSI at the receiver.
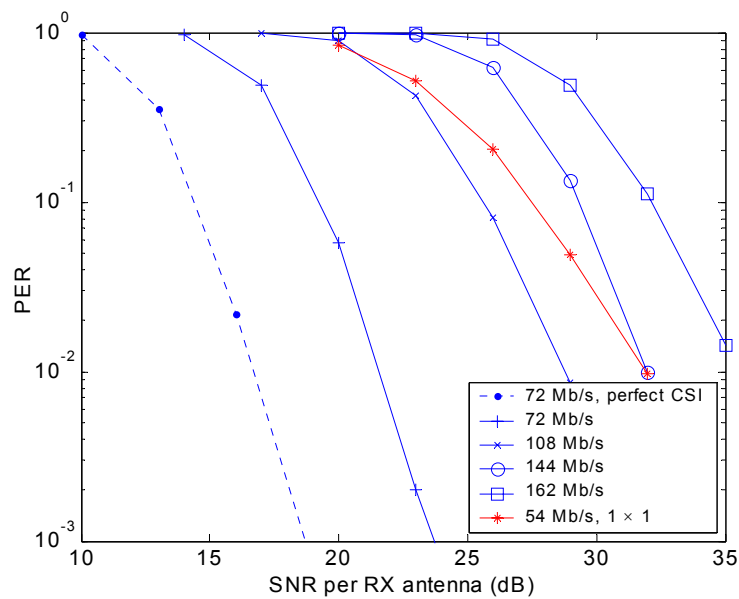
Figure 6-12: PER simulation results of PAC SIC for a 1 × 1 and 3 × 3 configuration, with perfect CSI at the receiver (dotted line) and with channel estimation (solid lines), different data rates, and frequency-selective Rayleigh fading with an exponential decaying PDP with 100 ns rms delay spread.

From these simulation results we see that the performance deteriorates going to higher data rates and we observe that applying channel estimation results in a loss of more than 4 dB. A more advanced channel estimation algorithm might reduce this loss (see Subsections 6.2.3 and 6.2.6). Furthermore, we note that the 3 × 3 curves fall off faster than the 1 × 1 curve (of 54 Mbps, 64-QAM, rate 3/4 and a packetlength of 513 bytes), such that at high SNRs, most of the higher MIMO rates outperform the SISO 54 Mbps rate, even having larger packetlengths. This can be explained by the fact that the MIMO configurations benefit not only from the frequency diversity, but also from the spatial diversity.

Taking the simulation results at an SNR of 26 dB, we can evaluate the performance of the measurements with the test system for location 9. For different data rates and an rms delay spread of 100 ns, the results are depicted in Figure 6-9 by the dotted lines. We clearly see that the performance of the test system in a real environment is worse than the performance of the idealized simulations. This can most likely be explained by system degradations that are not taken into account in the simulations, such as mutual coupling, residual IQ imbalance, errors in the frequency offset estimation, phase noise, and/or quantisation. Another reason could be the assumption of i.i.d. channel elements in the simulations.

6.5   Conclusions

In this chapter, an overview was given of the (changes in) signal processing, required to extend the physical layer of an OFDM system to Multiple-Input Multiple-Output (MIMO) OFDM. The signal processing of frame detection, time synchronisation, frequency synchronisation, channel estimation, synchronisation tracking using pilot subcarriers and MIMO detection algorithms was discussed. As a test case the MIMO extension of the

OFDM standard IEEE 802.11a was considered, but the results can be applied more generally. The necessary processing was not only evaluated through simulations, but also by measurements using a 3×3 MIMO test system that was set up in a typical office environment.

These error-rate measurements performed with the test setup (with a partly off-line implementation of the complete signal processing) showed a slightly worse performance than the idealised simulation results. The explanation is that in the simulations, system degradations such as phase noise and quantisation as well as propagation effects such as ill-conditioned MIMO channels are not taken into account.

Finally, measurements showed that an implementation of a 3×3 MIMO OFDM system achieves about a two times higher data rate than its 1×1 counterpart at a given range. Two reasons can be found for not reaching the theoretical tripling of the data rate: firstly, mutual coupling between the branches at the transmitter and receiver side. Secondly, the maximum data rate enhancement by a factor of 3 can only be achieved in well-conditioned MIMO channels, i.e. having i.i.d. channel elements. Thirdly, system impairments such as RF non-linearities and quantisation noise could be an issue, though, in Subsections 6.2.4 and 6.2.5, and in [103] we observed that MIMO correction algorithms for impairments in general perform better than their SISO counterparts.

Although some work needs to be done to improve the average performance, in general, it can be concluded that the MIMO technology is suitable to extend wireless communication standards like IEEE 802.11a to higher data rates.

7

Conclusions and Recommendations

7.1 Conclusions

7.1.1 Main Conclusion and Summary of Objectives

As answer to the demand for ever increasing data rates and augmented mobility, SDM OFDM provides an attractive and practical solution for future high-speed indoor wireless data communication networks. It combines the data-rate and spectral-efficiency enhancements of SDM with the relatively high spectral efficiency and the robustness against frequency-selective fading and narrowband interference of OFDM.

Specifically, the proposed coded-SDM detection scheme, based on a non-linear variant of the Minimum Mean Squared Error (MMSE) algorithm and named Per-Antenna-Coded Successive Interference Cancellation (PAC SIC) (see Subsection 5.7.3), was shown to perform close to the optimal performing Maximum Likelihood Detection with soft-decision output values (SOMLD). PAC SIC, however, has a much lower complexity than SOMLD as can be deduced from the complexity comparison of Subsection 4.11.3. This high performance and low complexity are achieved at the expense of a manageable latency. Through simulations and practical measurements it is shown that by SDM OFDM the data rate can be increased by a factor equal to the number of transmit (TX) antennas, while for many scenarios also the performance is increased with respect to standard OFDM. This performance gain is explained by the fact that, besides the exploitation of frequency diversity, also the spatial diversity is exploited.

Moreover, the inherent capability of coded SDM to fallback in coding rate, constellation scheme, and/or number of transmit antennas provides a means to deliver a good performance in a variety of scenarios. Altogether, these are strong advantages making the SDM OFDM combination an attractive solution for, e.g., next generation Wireless Local Area Networks (WLANs).

More detailed conclusions and results of this dissertation are presented in the next subsections. Instead of presenting a chronological summary of the results, which can be found anyhow in Section 1.5 and the concluding sections of the other chapters, we will discuss the conclusions and results following the objectives stated in the introductory chapter. Recalling from Section 1.4, the objectives were:

- get a more fundamental understanding of MIMO,
- introduce a good and useful wideband MIMO channel model (for indoor environments),
- evaluate and find efficient SDM detection techniques in terms of performance and complexity,
- evaluate these techniques in combination with OFDM, e.g., by performing simulations and making use of the proposed wideband MIMO channel model,
- verify the SDM OFDM combination in real-life channels by means of a test system. This also requires tackling of the non-idealities encountered in practical circumstances.

The next subsections highlight what was achieved in this dissertation with respect to each of these objectives.

7.1.2 A Fundamental Understanding

A more fundamental understanding of MIMO was obtained by means of an intuitive physical interpretation, a unified framework, and a theoretical performance analysis of MIMO (in combination with OFDM).

The intuitive physical interpretation of MIMO was presented in Chapter 2. In this chapter, we showed how one of the simplest SDM techniques called Zero Forcing (ZF) alters the RX antenna array patterns. In free space, i.e., when only the Line-Of-Sight (LOS) components are considered, a null is placed in the direction of the unwanted antenna to null the "interference", hoping that the wanted antenna is not nulled but properly received. When, on the other hand, reflections occur in the environment, the unwanted antenna is nulled by a spot instead of a beam. Moreover, the wanted antenna is coincidentally positioned in a local maximum, resulting in a maximum separation between the wanted and unwanted antenna. This provides an intuitive explanation of the robustness of MIMO techniques in rich-scattering environments. Moreover, the ability to effectively separate the TX antenna signals in these environments leads to the intuitive explanation why the data rate with MIMO can be increased with a factor equal to the number of transmit antennas.

In Section 4.2, we presented a unified framework for MIMO techniques, which we extended with OFDM in Section 5.3. By the presentation of an overview of various MIMO algorithms we came to the definition of a unified framework. We observed that basically all space-time processing techniques (like STC and SDM) can be mapped onto a general structure consisting of an encoder, a space-time mapper, and constellation mappers. This concept can be extended to the frequency dimension as well. This work provides a more fundamental understanding on how to design space-time(-frequency) schemes. Namely, based on the unified framework we intuitively think that the best codes should be designed such that a representation of an information bit (i.e., by means of the redundancy introduced by the encoder) is mapped on as many different fading instants (in time and

space (and frequency)) as possible, allowing for detection of this information bit at the receiver with a high reliability. For open-loop systems an overall MLD, that searches over all possible codewords included in the lattice spanning the entire dimensions, would be the best performing scheme. Unfortunately, the complexity of such an exhaustive search grows exponentially with the size of the code dimensions. So, the goal is to find attractive and effective space-time(-frequency) transmit schemes that allow for simple, low complex, high-performing detection schemes at the receiver.

The general structure of a space-time(-frequency) transmitter that follows from the unified framework forms a perfect basis for general theoretical performance analysis. We performed a theoretical space-frequency error-rate analysis of MIMO OFDM in which we also included the effect of spatial fading correlation (see Section 5.6). The main observation from this analysis was that the maximum achievable diversity order equals the product of the frequency diversity order, the number of TX antennas, and the number of RX antennas. The benefit of this analysis is that it can be used to design space-frequency coding schemes that perform well for a variety of spatial correlation scenarios. Furthermore, we showed that, when the diversity order is high enough (four or more) and the SNR is low enough, dedicated space-frequency code-design rules are overruled by the established Euclidean distance criterion. As a result, standard Single-Input Single-Output (SISO) codes together with some form of spatial and frequency multiplexing may outperform Space(-Time)-Frequency codes designed according to dedicated rules (i.e., the rank and determinant criteria introduced in [116]).

7.1.3 A Good and Useful Channel Model

The design choices for any communication system are largely constrained by the communication channel it experiences. The propagation channel of WLANs is characterised by severe multipath because of their primarily indoor exploitation resulting in the many reflections of the radio waves on walls and objectives in the environment. Due to constructive and destructive combination of these waves the received signal can strongly very as function of time, frequency, and/or location. The frequency-selectivity of the indoor channel can efficiently be mitigated by OFDM, which essentially splits a wideband frequency-selective fading channel into multiple orthogonal narrowband frequency-flat fading channels. These narrowband channels can be equalised in a trivial way. Together with the robustness and data-rate enhancement of MIMO in rich-scattering environments, MIMO OFDM is seen as a very promising solution for next generation WLAN systems.

In order to evaluate the various MIMO OFDM detection algorithms, performance simulations are an effective means. Since the characteristics of the MIMO fading channel have a strong influence on the system performance, it is of foremost significance to have an accurate but useful channel simulation model that takes the key channel characteristics into account. To that end, we introduced a wideband MIMO channel model that captures the typically large number of channel characteristics into a very few carefully selected ones such as multipath fading, propagation loss, rms delay spread, a LOS component, and spatial fading correlation (see Chapter 3).

Spatial fading correlation is identified as one of the key channel impairments specific for MIMO communications. The MIMO performance substantially degrades for correlation

values of over 0.4. Therefore, it is important to also evaluate the performance of MIMO algorithms in scenarios with spatial correlation. In general, however, the correlation is defined between any pair of MIMO channel elements. In the worst case, this leads for, e.g., a 4 × 4 MIMO system to 256 correlation parameters. One can imagine that in this case it is cumbersome to include spatial correlation in performance simulations. For cases in which the performance is expressed in capacity or error-rate, however, we presented a compact representation of the spatial fading correlation, having at most two parameters (see Section 3.6). This introduced spatial correlation model allows for easy inclusion of spatial correlation in MIMO simulations. The strength of the model is that, by ranging the one or two parameter(s) from zero to one, all scenarios ranging from totally uncorrelated to fully correlated spatial fading can be considered.

7.1.4 Performance and Complexity Evaluation of SDM Algorithms

Of the two basic MIMO techniques STC and SDM, the latter is especially designed to exploit the data-rate enhancement capabilities of the MIMO principle. Basically, in SDM independent data streams are transmitted on different transmit antennas simultaneously and at the same carrier frequency. Although these parallel streams of data are mixed-up in the air, when the MIMO channel is well conditioned they can be recovered at the receiver by using advanced signal processing algorithms, which usually require multiple receive antennas, too, to ensure adequate Bit Error Rate (BER) performance.

A number of such signal processing algorithms were described and evaluated in Chapter 4, namely, ZF, MMSE, ZF with SIC, MMSE with SIC, and MLD. Here, they are listed by increasing error-rate performance. We programmed these algorithms in MATLAB and did a number of performance simulations for different antenna configurations, for various constellation sizes, for different channel properties (including spatial correlation and/or a LOS component), without and with additional coding. Evaluating the results led to the following conclusions.

1. Provided enough diversity in the channel, the diversity order of the SDM algorithms based on the linear techniques ZF and MMSE tends towards $N_r - N_t + 1$, with $N_t \leq N_r$, where N_t and N_r denote the number of TX and RX antennas, respectively (see Section 4.6). The diversity order of MLD, however, equals N_r (see Section 4.10).

2. Linked to the previous conclusion, when an extra RX antenna is added to a symmetric MIMO configuration, the SDM performance improves significantly, relaxing the practical implementation of such systems considerably (see Section 4.11).

3. SDM with outer coding is a powerful concept. In literature it was namely shown that, when the number of antennas and diversity potential of the channel are large enough and the targeted SNR is low enough, the Euclidean distance criterion may be more appropriate than the rank and determinant criteria (see Section 4.2 and the simulation results of Subsection 4.11.2). This can be explained by the fact that, when a reasonably large diversity gain is provided by frequency, transmit and/or receive diversity, a wideband MIMO fading channel converges to a Gaussian channel, provided proper encoding is applied across the diversity dimensions. This would indicate that a one-dimensional code designed for AWGN channels, of which its codewords are properly

interleaved across the space and time, may be as effective as a space-time code that follows the rank and determinant criteria introduced in [116]. In [148] it was shown that this would already be the case when the diversity order is equal to or larger than four, which would restrict the domain of interest of space-time coding to architectures with only 2–3 antennas. In Subsection 4.11.2, we showed through simulations that SDM with a certain outer code outperforms Space-Time Coding (STC) with comparable coding strength, for a high enough diversity.

4. The concept of SDM with outer coding is much more flexible than STC, since it has the capability to easily fallback in coding rate, constellation scheme, and/or number of transmit antennas. Hence, coded SDM is an attractive technique able to deliver a good performance in a variety of scenarios. For STC, however, in general a different encoder and decoder have to be implemented for every rate.

5. The separation between the encoding process and spatial mapping (or multiplexing) in coded SDM allows the receiver to iterate between the spatial demapping and the temporal decoding. This processing stems from the turbo decoding principle and is in this context called Turbo SDM. Since its performance is very close to the overall exhaustive maximum likelihood search, Turbo SDM is an excellent way to get close to the outage performance bound (see Section 4.13).

We also evaluated the complexity of the presented SDM algorithms in terms of number of additions and number of multiplications. It turned out that the complexity of MLD grows exponentially with the number of transmit antennas (N_t). The complexity of the least complex scheme, ZF, grows only polynomial with N_t, namely, the complexity of the ZF processing during the payload phase is proportional to N_t^3. We can conclude from the presented results that the implementation of MIMO schemes (with a low number of antennas) is feasible.

7.1.5 SDM OFDM Algorithm Evaluation

As mentioned before, the combination of SDM and OFDM is seen as an attractive and practical solution for future high-speed indoor WLANs. This is based on the following arguments:

- OFDM (in combination with coding) effectively deals with the frequency-selectivity of the indoor channel, it provides a relatively high spectral efficiency, and it offers the ability to include a proper guard interval between subsequent OFDM symbols leading to sufficient delay spread robustness. Furthermore, when it is combined with coding it provides robustness against narrowband interference,
- the current high-data rate standards of IEEE, namely IEEE 802.11a and g, are based on OFDM, so SDM OFDM is a logical extension and offers simple means to also consider coexistence and interoperability,
- the data-rate and spectral-efficiency enhancements of SDM are the highest in rich-scattered multipath by which the indoor environments are typically characterised,
- the flexibility of coded SDM to fallback in coding rate, constellation scheme, and/or number of transmit antennas.

In order to evaluate the performance of (coded) SDM OFDM schemes, we extended the narrowband SDM simulation software we used in Chapter 4 with OFDM based on the IEEE 802.11a parameters and the in Chapter 3 introduced wideband MIMO channel model. Based on the performance simulations we presented in Chapter 5 for different antenna configurations, for various constellation sizes, for different rms delay spreads, without and with a LOS component, we concluded the following:

1. The diversity order of the SDM algorithms is multiplied by the frequency diversity order (see Section 5.6). As a result, SDM OFDM algorithms have a higher robustness, since the probability of having a low spatial *and* frequency diversity is particularly small in indoor environments.

2. Based on the additional spatial diversity of SDM, in general, the coded SDM OFDM schemes outperform their SISO counterparts, except for the symmetric antenna configurations with high constellation orders in scenarios with a delay spread of 250 ns. Since the lattice containing all possible TX symbols becomes more and more complex when going to a higher constellation size and to more transmit antennas, the performance particularly suffers from non-orthogonal MIMO channels. As a result, the performance penalty for MIMO systems when going to a higher constellation order is worse than that for SISO systems.

3. The introduced coded-SDM-OFDM detection scheme based on a non-linear variant of the Minimum Mean Squared Error (MMSE) algorithm and named Per-Antenna-Coded Successive Interference Cancellation (PAC SIC), see Subsection 5.7.3, achieves equivalent PER performances as the complex SOMLD, but the complexity of PAC SIC grows only polynomial with the number of TX antennas. The potential disadvantage of PAC SIC is that it introduces extra latency, but for small interleaver and coding depths, this latency is manageable.

4. Detection schemes, like PAC SIC, designed for the straightforward coded-SDM-OFDM transmission schemes, based on multiplexing a one-dimensional code over the spatial and frequency dimension, still perform at least 7 dB worse than the outage performance at the PER of interest: 1%. Note that 2.2 dB of the 7 dB is imposed by the OFDM overhead, i.e., the guard subcarriers and guard time. But apparently there is still enough room left for improvements, challenging us to find more efficient ways of implementing the Euclidean distance criterion. A promising solution may be to introduce the turbo processing as described for narrowband MIMO transmissions in Section 4.13 to the MIMO OFDM context. Other ways to improve the performance and get closer to the outage PER may be found in more advanced code design that also benefits from the transmit diversity.

5. When an extra RX antenna is added to a 2×2 system, the different SDM OFDM detection schemes perform very similarly; their required SNRs to achieve a PER of 1% do not differ more than 2.5 dB. An additional effect is that the average performance is improved due to the extra diversity gain. For some symmetric MIMO schemes with a high constellation order this is shown to be a potential solution to shift the required SNR to practical values and, as a result, relaxes the transceiver design.

7.1.6 Implementation of a MIMO OFDM WLAN system

For the simulations of Chapter 4 and 5 we idealised the system under investigation, in order to focus on the main goal of these chapters, namely, algorithm comparison and evaluation. We assumed that system impairments such as frequency offset, timing offset, phase noise, IQ imbalance, DC offset, and quantisation noise, were negligible. When practical implementation is envisioned, however, the system has to deal with these non-idealities. Commonly, this is done by performing training and synchronisation. In packet-switched systems, a preamble (i.e., a piece of known data exceeding the packet) can be used to initiate these tasks. Since we were searching for extensions to the IEEE 802.11a standard, a number of necessary changes to the 802.11a preamble, together with the required processing changes, were proposed in Chapter 6, such that it allows for time and frequency synchronisation, channel estimation, and synchronisation tracking in case of MIMO OFDM.

To validate the MIMO OFDM concept in practice, including the effect of system impairments, a 3×3 test system was built within Agere Systems, The Netherlands. With this test system, as set of measurements based on IEEE 802.11a parameters was performed in a typical office environment. The measurement results led to the following conclusions.

First, the proposed preamble and channel estimation approach result in a performance degradation of more than 4 dB compared to the hypothetical case where perfect knowledge of the channel was available at the receiver (see Figure 6-12).

Second, the error-rate measurements performed with the test set up (with a – partly off-line – implementation of the complete signal processing) showed a slightly worse performance than the idealised simulation results (including channel estimation). The explanation is that in the simulations, system degradations such as phase noise and quantisation as well as propagation effects such as ill-conditioned MIMO channels are not taken into account.

Finally, measurements showed that an implementation of a 3×3 MIMO OFDM system achieves about a two times higher data rate than its 1×1 counterpart at a given range. Two reasons can be found for not reaching the theoretical tripling of the data rate: 1) mutual coupling between the branches at the transmitter and receiver side and 2) the maximum data rate enhancement by a factor of 3 can only be achieved in well-conditioned MIMO channels, i.e. having i.i.d. channel elements.

Although there is some room for improvement in the SDM OFDM concept, we can overall conclude that SDM OFDM is an attractive and practical solution to extend wireless communication systems based on standards such as IEEE 802.11a to higher data rates.

7.2   Recommendations and Open Issues

At the end of this dissertation, we would like to present a number of recommendations and open issues to stimulate further research. Since the recommendations are very diverse, we will present them by means of bulleted indices:

- The overview of different MIMO techniques presented in Section 4.2 raises the question what the effect of each of these techniques will be on the RX antenna array patterns (see Chapter 2). What is, e.g., the effect of a closed-loop technique based on the singular value decomposition on the antenna patterns? Note that the approach of Chapter 2 limits the answering of the first question to linear techniques. Is it possible to define other visualisation methods to also include non-linear techniques?

- Although the wideband MIMO channel model described in Chapter 3 is an extension of an indoor SISO channel model that is supported by numerous measurements presented in literature, it would be beneficial to verify the extensions (such as the spatial correlation and LOS component) by performing MIMO channel measurements. If necessary, the model should be adapted accordingly.

- The unified framework on MIMO (OFDM) introduced in Sections 4.2 and 5.3 can serve as a good starting point to define a unified theory on the various MIMO (OFDM) techniques. Based on the unified framework one could, e.g., search for a unified signal model, which again can be used to derive a general theoretical error-rate performance analysis. From this analysis, general space-time-frequency code design rules may be deduced. Perhaps the unified theory can be extended even further and beyond the boarders of the speciality of radio communication, because in areas like equalisation, remote sensing, and image processing, techniques are used with a lot of commonalities with SDM algorithms.

- In Subsections 4.6.2 and 4.10.2, we derived upperbounds for the error-rate performance of ZF and MLD. It would be useful to find tight upperbounds on the error-rate performance of MMSE, ZF with SIC, and MMSE with SIC as well. Especially when the upperbounds are tight this would ease the evaluation of the performance of these algorithms for different MIMO configurations, since then long-lasting simulations are only needed when exact performance figures are required.

- A more thorough study under which conditions the MIMO OFDM schemes based on the Euclidean distance criterion outperform dedicated Space-Frequency codes (see Subsection 5.7.1) would be beneficial.

- Proper codes for MIMO OFDM, that are robust for a wide range of channel conditions including different correlation scenarios, can be designed, based on the space-frequency error-rate analysis of Section 5.6 or extensions of it. (According to Subsection 5.8.3 there is still room for improvement with respect to the proposed coded SDM OFDM schemes.)

- A possible solution to get closer to the outage performance might be the application of the turbo-SDM concept described in Section 4.13 to coded SDM OFDM. In order to reduce the complexity, it would be beneficial to find less complex algorithms that can replace the complex MAP MIMO detector, such as, e.g., a form of ZF that accepts a priori inputs.

- Other solutions might be found by evaluating the performance of coded SDM OFDM with different types of coding schemes, e.g., Low-Density Parity Check (LDPC) coding.

- It would be desirable to find more flexible coded MIMO schemes, that like STC also benefit from the available transmit diversity.

- It was observed in Subsection 5.7.3 that the defined soft values for PAC SIC are in general not correct, but coincidentally work properly for the detection of PAC transmissions. The definition of appropriate soft-decision output values for ZF with SIC and MMSE with SIC is still an open issue.

- With respect to the overall throughput, it would be good to find efficient Medium Access Control (MAC) protocols to embed the principle of MIMO OFDM. Note that the training overhead increases linearly with the number of transmit antennas (N_t), moreover the packet size decreases by N_t, so the overhead grows quadratic with N_t. Therefore, more efficient MAC protocols must be found to deal with this problem, for instance by making the packet size larger (e.g., through frame aggregation), or reducing the overhead, or both.

- The big performance loss observed in Figure 6-12, the remarks made in Subsections 6.2.3 and 6.2.6, and the previous bulleted index call for the definition of stable and efficient channel training algorithms for MIMO OFDM.

- Define good fallback mechanisms. The main question is when to fallback in coding rate/constellation size, and when to fallback in number of TX antennas.

- In order to obtain a better understanding of the effect of system impairments such as frequency offset, timing offset, phase noise, IQ imbalance, DC offset, and non-linear amplifiers on MIMO, it would be useful to perform more fundamental analyses on their effect on, e.g., the error-rate performance. A good basis for these analyses forms the concise matrix MIMO OFDM signal model introduced in Section 5.4. When necessary, the estimation and synchronisation algorithms of Chapter 6 must be updated based on new findings. Maybe it is possible to find a general error model that includes all impairments.

- Search for efficient mathematical algorithms that allow for low-cost hardware implementations of SDM OFDM.

- In this dissertation we have only looked at open-loop MIMO (OFDM) schemes. For particular environments closed-loop schemes might be feasible and beneficial. Hence, it might be good to list and define a number of closed-loop MIMO (OFDM) schemes and evaluate their performances and compare the performances with those of their open-loop counterparts.

- When a cellular system is the target application, interference issues might cause problems. For such an application, it would be useful to perform MIMO OFDM simulations with different interference sources, such as co-channel interference, adjacent-channel interference, etc. and evaluate the results. If possible, appropriate interference cancellation schemes must be found. In this context, adding an extra RX antenna opens new possibilities such as using this extra degree of freedom for interference cancellation.

- Extend the application of MIMO OFDM to networks beyond WLANs. Investigate how the MIMO OFDM concept can be embedded in a 4G framework.

- The high potential of MIMO OFDM brings along an interesting observation, namely, that MIMO OFDM exploits two dimensions: the time/frequency dimension and the spatial dimension. This raises a fundamental question for research to future wireless communication systems: what is the next dimension to exploit? Or do we, by this concept, reach the limit and can we only achieve higher data rates by exploiting these two dimensions more and more efficiently?

Appendix A

Mathematical Appendix

A.1 Matrix Theory

A.1.1 References

Most of the matrix theory that is described in the next subsections can be found in [54] and [113], and explicit references to these books will not be given. When other literature is used, however, the references will be included.

A.1.2 Eigenvalues and Eigenvectors

Let $\mathbf{A}$ be a matrix with complex entries and $\mathbf{z}$ be a complex vector, consider the equation

$$\mathbf{A}\mathbf{z} = \lambda\mathbf{z}, \qquad (A.1)$$

where λ is a scalar. If a scalar λ and a nonzero vector $\mathbf{z}$ happen to satisfy this equation, then λ is called an *eigenvalue* of $\mathbf{A}$ and $\mathbf{z}$ is called an *eigenvector* of $\mathbf{A}$ associated with λ. Note that the two occur inextricably as a pair, and that an eigenvector cannot be the zero vector.

Or in other words, the number λ is an eigenvalue of matrix $\mathbf{A}$ if and only if

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0, \qquad (A.2)$$

This is the *characteristic equation*, and for a given solution λ_i, the corresponding eigenvector $\mathbf{z}_i$ can be found by solving

$$\left(\mathbf{A} - \lambda_i \mathbf{I}\right)\mathbf{z}_i = 0 \,. \tag{A.3}$$

A.1.3 Hermitian Matrix

A matrix is said to be Hermitian if $\mathbf{A} = \mathbf{A}^H$.

A.1.4 The Singular Value Decomposition

The Singular Value Decomposition (SVD) is closely associated with the eigenvalue-eigenvector factorisation of a Hermitian matrix: $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, where the eigenvalues are in the diagonal matrix $\boldsymbol{\Lambda}$, and the eigenvector matrix is *unitary*: $\mathbf{U}^H\mathbf{U} = \mathbf{I}$, because the eigenvectors of a Hermitian matrix are orthonormal. The requirement that a matrix is Hermitian makes above factorisation very strict. The SVD, however, can be applied to any matrix: any $M \times N$ matrix $\mathbf{A}$ can be factored into:

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^H \,. \tag{A.4}$$

The columns of the $M \times M$ matrix $\mathbf{U}$ are eigenvectors of $\mathbf{A}\mathbf{A}^H$, and the columns of matrix $\mathbf{V}$ ($N \times N$ dimensional) are eigenvectors of $\mathbf{A}^H\mathbf{A}$. The singular values σ_i are found on the main diagonal of the $M \times N$ matrix $\mathbf{D}$ and they are the square roots of the nonzero eigenvalues of both $\mathbf{A}\mathbf{A}^H$ and $\mathbf{A}^H\mathbf{A}$:

$$\mathbf{A}\mathbf{A}^H = \left(\mathbf{U}\mathbf{D}\mathbf{V}^H\right)\left(\mathbf{U}\mathbf{D}\mathbf{V}^H\right)^H = \mathbf{U}\mathbf{D}\mathbf{D}^H\mathbf{U}^H \,, \text{ and similarly } \mathbf{A}^H\mathbf{A} = \mathbf{V}\mathbf{D}^H\mathbf{D}\mathbf{V}^H \,. \tag{A.5}$$

A.1.5 Rank and Condition Number

The *rank* of a matrix $\mathbf{A}$ equals the number of nonzero eigenvalues of that matrix $\mathbf{A}$.

The *condition number* of a matrix $\mathbf{A}$ is defined as the ratio between its maximum and minimum singular value ([100]):

$$\kappa(\mathbf{A}) = \frac{\sigma_{\max}}{\sigma_{\min}} \,. \tag{A.6}$$

A condition number of one implies that the matrix $\mathbf{A}$ is orthogonal. A large condition number means that $\mathbf{A}$ is highly non-orthogonal or *ill-conditioned*.

A.1.6 (Non-)singular

The following are equivalent for an $N \times N$ matrix $\mathbf{A}$:

1. $\mathbf{A}$ is non-singular;
2. $\mathbf{A}^{-1}$ exists;

3. rank($\mathbf{A}$) = N;
4. the rows of $\mathbf{A}$ are linearly independent;
5. the columns of $\mathbf{A}$ are linearly independent;
6. det($\mathbf{A}$) $\neq$ 0;
7. the only solution to $\mathbf{Az} = \mathbf{0}$ is $\mathbf{z} = \mathbf{0}$;
8. 0 is not an eigenvalue of $\mathbf{A}$.

The proof of 8) goes as follows: the matrix $\mathbf{A}$ is singular if and only if $\mathbf{Az} = \mathbf{0}$ for some $\mathbf{z} \neq \mathbf{0}$. This happens if and only if $\mathbf{Az} = 0 \cdot \mathbf{z}$, that is, if and only if $\lambda = 0$ is an eigenvalue.

A.1.7 Nonnegative or Positive Semidefinite

An $N \times N$ Hermitian matrix $\mathbf{A}$ is said to be *nonnegative definite* or *positive semidefinite* if

$$\mathbf{z}^H \mathbf{Az} \geq 0, \tag{A.7}$$

for all nonzero complex vectors $\mathbf{z}$. Note that if $\mathbf{A} = \mathbf{A}^H$, then for all complex vectors $\mathbf{z}$, the number $\mathbf{z}^H \mathbf{Az}$ is real. For the proof, compute $(\mathbf{z}^H \mathbf{Az})^H$. One would expect the conjugate of the 1×1 matrix $\mathbf{z}^H \mathbf{Az}$, but actually the same number is got back again, so the number must be real:

$$\left(\mathbf{z}^H \mathbf{Az}\right)^H = \mathbf{z}^H \mathbf{A}^H \left(\mathbf{z}^H\right)^H = \mathbf{z}^H \mathbf{Az}. \tag{A.8}$$

Furthermore, note that each eigenvalue of a nonnegative definite matrix is a nonnegative real number. Let $\mathbf{A}$ be nonnegative definite, let λ be an eigenvalue of $\mathbf{A}$, let $\mathbf{z}$ be an eigenvector of $\mathbf{A}$ associated with λ, and calculate $\mathbf{z}^H \mathbf{Az} = \mathbf{z}^H \lambda \mathbf{z} = \lambda \|\mathbf{z}\|^2$. Then $\lambda = \mathbf{z}^H \mathbf{Az}/\|\mathbf{z}\|^2$ is real and nonnegative, since it is a ratio of a real nonnegative and a real positive number.

A.1.8 Matrix Inversion Properties

For any invertible square matrices $\mathbf{A}$ and $\mathbf{B}$, with the same dimensions,

$$\begin{aligned} \mathbf{A}^{-1} - \mathbf{B}^{-1} &= \mathbf{B}^{-1}\mathbf{B}\left(\mathbf{A}^{-1} - \mathbf{B}^{-1}\right) = \mathbf{B}^{-1}\left(\mathbf{BA}^{-1} - \mathbf{I}\right) \\ &= \mathbf{B}^{-1}\left(\mathbf{BA}^{-1} - \mathbf{AA}^{-1}\right) = \mathbf{B}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{A}^{-1} \end{aligned} \tag{A.9}$$

When we have a matrix of the form $\mathbf{A} + \mathbf{BCD}$, where $\mathbf{A}$ and $\mathbf{C}$ are square invertible matrices, the *Matrix Inversion Lemma* is defined as ([66])

$$\left(\mathbf{A} + \mathbf{BCD}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}. \tag{A.10}$$

A.1.9 The Kronecker Product

The Kronecker product of any $M \times N$ matrix $\mathbf{A}$ and $X \times Y$ matrix $\mathbf{B}$, is defined by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1N}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2N}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}\mathbf{B} & a_{M2}\mathbf{B} & \cdots & a_{MN}\mathbf{B} \end{bmatrix}. \tag{A.11}$$

The result on the right-hand side is an $MX \times NY$ matrix.

A.1.10 Kronecker Product Identities

In this subsection, a proof (as found in [47]) will be given for the theorem that for any $M \times N$ matrix $\mathbf{A}$, $N \times P$ matrix $\mathbf{B}$, and $P \times Q$ matrix $\mathbf{C}$,

$$\text{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right)\text{vec}(\mathbf{B}), \tag{A.12}$$

where $\otimes$ denotes the Kronecker product and vec(.) stands for the vector operation, in which a vector is formed from the respective matrix by stacking its columns under each other. To prove Equation (A.12), we need a number of statements. Firstly note that, under the assumption that $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ have the right dimensions,

$$(\mathbf{AB}) \otimes (\mathbf{CD}) = (\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D}), \tag{A.13}$$

which is simple but tedious to show. Secondly, vec($\mathbf{A}$+$\mathbf{B}$) = vec($\mathbf{A}$) + vec($\mathbf{B}$) and finally, if $\mathbf{a}$ and $\mathbf{b}$ are vectors, then vec($\mathbf{ba}^T$) = $\mathbf{a} \otimes \mathbf{b}$, both of which are obvious.

Now, Equation (A.12) is proved as follows. $\mathbf{B}$ can be expressed as

$$\mathbf{B} = \sum_{p=1}^{P} \mathbf{b}_p \mathbf{u}_p^T, \tag{A.14}$$

where $\mathbf{b}_p$ is column p of $\mathbf{B}$ and $\mathbf{u}_p$ is column p of the (non-square) identity matrix. Using this property and after changing the order of the 'vec'-operation and the sum, we can rewrite (A.12) as

$$\text{vec}(\mathbf{ABC}) = \text{vec}\left\{\mathbf{A}\left(\sum_p \mathbf{b}_p \mathbf{u}_p^T\right)\mathbf{C}\right\} = \sum_p \text{vec}\left(\mathbf{A}\mathbf{b}_p \mathbf{u}_p^T \mathbf{C}\right). \tag{A.15}$$

Because both $\mathbf{A}\mathbf{b}_p$ and $\mathbf{C}^T\mathbf{u}_p$ are vectors,

$$\sum_p \text{vec}\left(\mathbf{A}\mathbf{b}_p \mathbf{u}_p^T \mathbf{C}\right) = \sum_p \left(\mathbf{C}^T\mathbf{u}_p\right) \otimes \left(\mathbf{A}\mathbf{b}_p\right)$$
$$= \sum_p \left(\mathbf{C}^T \otimes \mathbf{A}\right)\left(\mathbf{u}_p \otimes \mathbf{b}_p\right). \tag{A.16}$$

To finish the proof, just reverse the previous steps, like:

$$\sum_p \left(\mathbf{C}^T \otimes \mathbf{A} \right)\left(\mathbf{u}_p \otimes \mathbf{b}_p \right) = \left(\mathbf{C}^T \otimes \mathbf{A} \right)\mathrm{vec}\left(\sum_p \mathbf{b}_p \mathbf{u}_p^T \right)$$
$$= \left(\mathbf{C}^T \otimes \mathbf{A} \right)\mathrm{vec}(\mathbf{B}). \tag{A.17}$$

Note that (A.12) also holds for the complex case. This can be shown by writing a complex matrix $\mathbf{X}$ as $\mathbf{X}_R + j\mathbf{X}_I$. By doing so, (A.12) becomes

$$\begin{aligned}
\mathrm{vec}(\mathbf{ABC}) &= \mathrm{vec}\left((\mathbf{A}_R + j\mathbf{A}_I)(\mathbf{B}_R + j\mathbf{B}_I)(\mathbf{C}_R + j\mathbf{C}_I) \right) \\
&= \mathrm{vec}(\mathbf{A}_R \mathbf{B}_R \mathbf{C}_R) - \mathrm{vec}(\mathbf{A}_R \mathbf{B}_I \mathbf{C}_I) - \mathrm{vec}(\mathbf{A}_I \mathbf{B}_I \mathbf{C}_R) \\
&\quad - \mathrm{vec}(\mathbf{A}_I \mathbf{B}_R \mathbf{C}_I) + j\mathrm{vec}(\mathbf{A}_I \mathbf{B}_R \mathbf{C}_R) - j\mathrm{vec}(\mathbf{A}_I \mathbf{B}_I \mathbf{C}_I) \\
&\quad + j\mathrm{vec}(\mathbf{A}_R \mathbf{B}_I \mathbf{C}_R) + j\mathrm{vec}(\mathbf{A}_R \mathbf{B}_R \mathbf{C}_I) \\
&= \left((\mathbf{C}_R^T \otimes \mathbf{A}_R) - (\mathbf{C}_I^T \otimes \mathbf{A}_I) + j(\mathbf{C}_I^T \otimes \mathbf{A}_R) + j(\mathbf{C}_R^T \otimes \mathbf{A}_I) \right) \\
&\quad \cdot \mathrm{vec}(\mathbf{B}_R + j\mathbf{B}_I) \\
&= \left((\mathbf{C}_R^T + j\mathbf{C}_I^T) \otimes (\mathbf{A}_R + j\mathbf{A}_I) \right)\mathrm{vec}(\mathbf{B}_R + j\mathbf{B}_I) \\
&= \left(\mathbf{C}^T \otimes \mathbf{A} \right)\mathrm{vec}(\mathbf{B}).
\end{aligned} \tag{A.18}$$

A.1.11 Block Circulant

In [108] it is shown that a circulant matrix can be diagonalised by a Fourier matrix. Here, this will be extended to the block circulant case. An $NA \times NB$ is called block circulant if it has the form

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_0 & \mathbf{C}_{N-1} & \mathbf{C}_{N-2} & \cdots & \mathbf{C}_1 \\ \mathbf{C}_1 & \mathbf{C}_0 & \mathbf{C}_{N-1} & \cdots & \mathbf{C}_2 \\ \mathbf{C}_2 & \mathbf{C}_1 & \mathbf{C}_0 & \cdots & \mathbf{C}_3 \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{C}_{N-1} & \mathbf{C}_{N-2} & \mathbf{C}_{N-3} & \cdots & \mathbf{C}_0 \end{pmatrix}, \tag{A.19}$$

where $\mathbf{C}_i$ is an $A \times B$ matrix. $\mathbf{C}$ is a special kind of Toeplitz matrix where each block of columns is obtained by doing a block wrap-around downshift of the previous "block vector". When denoting $\mathbf{K}$ as its first block of columns,

$$\mathbf{K} = \begin{pmatrix} \mathbf{C}_0 & \mathbf{C}_1 & \mathbf{C}_2 & \cdots & \mathbf{C}_{N-1} \end{pmatrix}^T, \tag{A.20}$$

$\mathbf{C}$ can be specified as

$$\mathbf{C} = \begin{pmatrix} \mathbf{K} & \mathbf{RK} & \mathbf{R}^2\mathbf{K} & \cdots & \mathbf{R}^{N-1}\mathbf{K} \end{pmatrix}, \tag{A.21}$$

where

$$\mathbf{R} = \mathbf{E} \otimes \mathbf{I}_A \text{ with } \mathbf{E} = \begin{pmatrix} \mathbf{e}_2 & \mathbf{e}_3 & \cdots & \mathbf{e}_N & \mathbf{e}_1 \end{pmatrix}, \tag{A.22}$$

and $\mathbf{e}_i$ is the i-th column of the $N \times N$ identity matrix. Now, the following theorem can be stated:

Theorem: if $\mathbf{C}$ is a block circulant matrix, then it can be block diagonalised like

$$\boldsymbol{\Delta} = \begin{pmatrix} \mathbf{F} \otimes \mathbf{I}_A \end{pmatrix} \mathbf{C} \begin{pmatrix} \mathbf{F}^{-1} \otimes \mathbf{I}_B \end{pmatrix}, \tag{A.23}$$

where $\mathbf{F}$ represents the $N \times N$ Fourier matrix and $\boldsymbol{\Delta}$ is a block diagonal matrix defined by

$$\boldsymbol{\Delta} = \begin{pmatrix} \begin{pmatrix} \mathbf{f}^1 \otimes \mathbf{I}_A \end{pmatrix} \mathbf{K} & & 0 \\ & \ddots & \\ 0 & & \begin{pmatrix} \mathbf{f}^N \otimes \mathbf{I}_A \end{pmatrix} \mathbf{K} \end{pmatrix}, \tag{A.24}$$

with $\mathbf{f}^i$ denoting the i-th row of the Fourier matrix and

$$\begin{pmatrix} \mathbf{f}^i \otimes \mathbf{I}_A \end{pmatrix} \mathbf{K} = \sum_{n=0}^{N-1} \mathbf{C}_n e^{-j2\pi \frac{n \cdot (i-1)}{N}}. \tag{A.25}$$

Thus, the blocks on the block diagonal of $\boldsymbol{\Delta}$ are obtained by performing the DFT on the blocks of "block vector" $\mathbf{K}$.

Note that the Fourier matrix is defined as

$$\mathbf{F} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W^1 & W^2 & \cdots & W^{(N-1)} \\ 1 & W^2 & W^4 & \cdots & W^{2(N-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & W^{(N-1)} & W^{2(N-1)} & \cdots & W^{(N-1)^2} \end{pmatrix}, \tag{A.26}$$

with $W = \exp(-j2\pi/N)$.

To prove this theorem, we start with verifying that a block circulant matrix is a polynomial in the downshift operator $\mathbf{E}$. The polynomial is defined by

$$\mathbf{P} = \begin{pmatrix} \mathbf{I}_N \otimes \mathbf{C}_0 \end{pmatrix} + \begin{pmatrix} \mathbf{E} \otimes \mathbf{C}_1 \end{pmatrix} + \begin{pmatrix} \mathbf{E}^2 \otimes \mathbf{C}_2 \end{pmatrix} + \dots + \begin{pmatrix} \mathbf{E}^{N-1} \otimes \mathbf{C}_{N-1} \end{pmatrix}. \tag{A.27}$$

Comparing the i-th block of columns of above polynomial and using $\mathbf{E}^k \mathbf{e}_i = \mathbf{e}_{(i+k) \bmod N}$ yields

$$
\begin{aligned}
\mathbf{P}\left(\mathbf{e}_i \otimes \mathbf{I}_B\right) &= \left(\mathbf{I}_N \otimes \mathbf{C}_0\right)\left(\mathbf{e}_i \otimes \mathbf{I}_B\right)+\ldots+\left(\mathbf{E}^{N-1} \otimes \mathbf{C}_{N-1}\right)\left(\mathbf{e}_i \otimes \mathbf{I}_B\right) \\
&= \left(\mathbf{e}_i \otimes \mathbf{C}_0\right)+\left(\mathbf{E}\mathbf{e}_i \otimes \mathbf{C}_1\right)+\ldots+\left(\mathbf{E}^{N-1}\mathbf{e}_i \otimes \mathbf{C}_{N-1}\right) \\
&= \left(\mathbf{e}_i \otimes \mathbf{C}_0\right)+\left(\mathbf{e}_{i+1} \otimes \mathbf{C}_1\right)+\ldots+\left(\mathbf{e}_N \otimes \mathbf{C}_{N-i}\right)+ \\
&\quad +\left(\mathbf{e}_1 \otimes \mathbf{C}_{N-i+1}\right)+\ldots+\left(\mathbf{e}_{i-1} \otimes \mathbf{C}_{N-1}\right) \\
&= \mathbf{R}^{i-1}\mathbf{K} \\
&= \mathbf{C}\left(\mathbf{e}_i \otimes \mathbf{I}_B\right).
\end{aligned}
\tag{A.28}
$$

Thus, the i-th block of columns of $\mathbf{P}$ equals the i-th block of columns of $\mathbf{C}$, and since this holds for all i, it is shown that $\mathbf{P}$ equals $\mathbf{C}$. With this knowledge, we can rewrite (A.23) as

$$
\begin{aligned}
\mathbf{\Delta} &= \left(\mathbf{F} \otimes \mathbf{I}_A\right)\mathbf{C}\left(\mathbf{F}^{-1} \otimes \mathbf{I}_B\right) \\
&= \left(\mathbf{F} \otimes \mathbf{I}_A\right)\left(\left(\mathbf{I}_N \otimes \mathbf{C}_0\right)+\left(\mathbf{E} \otimes \mathbf{C}_1\right)+\ldots+\left(\mathbf{E}^{N-1} \otimes \mathbf{C}_{N-1}\right)\right)\left(\mathbf{F}^{-1} \otimes \mathbf{I}_B\right) \\
&= \left(\mathbf{F}\mathbf{F}^{-1} \otimes \mathbf{C}_0\right)+\left(\mathbf{F}\mathbf{E}\mathbf{F}^{-1} \otimes \mathbf{C}_1\right)+\ldots+\left(\mathbf{F}\mathbf{E}^{N-1}\mathbf{F}^{-1} \otimes \mathbf{C}_{N-1}\right).
\end{aligned}
\tag{A.29}
$$

Note that $\mathbf{E}$ is a circulant matrix and, as shown in [108], this can be diagonalised by $\mathbf{F}$. Here, we will recall the proof. To show that $\mathbf{E}$ can be diagonalised by $\mathbf{F}$, we have to prove that

$$
\mathbf{F}\mathbf{E} = \mathbf{D}\mathbf{F},
\tag{A.30}
$$

where $\mathbf{D} = \operatorname{diag}((1 \ W \ W^2 \ \ldots \ W^{N-1})^T)$. To that end, element (i,k) of $\mathbf{F}\mathbf{E}$ is compared with the same element of $\mathbf{D}\mathbf{F}$:

$$
\left(\mathbf{F}\mathbf{E}\right)_{ik} = \begin{pmatrix} 1 & W^{(i-1)} & W^{2(i-1)} & \cdots & W^{(N-1)(i-1)} \end{pmatrix}\mathbf{e}_{k+1} = W^{k(i-1)},
\tag{A.31}
$$

$$
\left(\mathbf{D}\mathbf{F}\right)_{ik} = W^{i-1}\mathbf{e}_i^T \begin{pmatrix} 1 \\ W^{(k-1)} \\ W^{2(k-1)} \\ \vdots \\ W^{(N-1)(k-1)} \end{pmatrix} = W^{i-1}W^{(i-1)(k-1)} = W^{k(i-1)}.
\tag{A.32}
$$

This of course holds for all i and k ($1 \le i, k \le N$). When applying this result to (A.29), our theorem on block circulant matrices can be proved as follows:

$$
\begin{aligned}
\left(\mathbf{F} \otimes \mathbf{I}_A\right)\mathbf{C}\left(\mathbf{F}^{-1} \otimes \mathbf{I}_B\right) &= \left(\mathbf{F}\mathbf{F}^{-1} \otimes \mathbf{C}_0\right)+\left(\mathbf{F}\mathbf{E}\mathbf{F}^{-1} \otimes \mathbf{C}_1\right)+\ldots+\left(\mathbf{F}\mathbf{E}^{N-1}\mathbf{F}^{-1} \otimes \mathbf{C}_{N-1}\right) \\
&= \sum_{n=0}^{N-1} \mathbf{F}\mathbf{E}^n\mathbf{F}^{-1} \otimes \mathbf{C}_n \\
&= \sum_{n=0}^{N-1} \mathbf{D}^n \otimes \mathbf{C}_n = \mathbf{\Delta}.
\end{aligned}
\tag{A.33}
$$

And the i-th matrix on the block diagonal $\mathbf{\Delta}$, $1 \le i \le N$, can be shown to be

$$\sum_{n=0}^{N-1} (\mathbf{D})_{ii}^{n} \mathbf{C}_{n} = \sum_{n=0}^{N-1} \mathbf{C}_{n} W^{(i-1)n} = \sum_{n=0}^{N-1} \mathbf{C}_{n} e^{-j2\pi \frac{n \cdot (i-1)}{N}} , \tag{A.34}$$

which is the DFT on the blocks of $\mathbf{K}$. This concludes the proof.

A.2   Multivariate Complex Gaussian Distribution

The probability density function of an M-dimensional *real* Gaussian (i.e., normal) random vector $\mathbf{x}$, with a covariance matrix $\mathbf{Q}_x$ and mean vector $\boldsymbol{\mu}_x$, is given by ([7])

$$p(\mathbf{x}) = \frac{1}{\sqrt{\det(2\pi \mathbf{Q}_x)}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_x)^T \mathbf{Q}_x^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)\right). \tag{A.35}$$

Following the analysis of [117], an M-dimensional *complex* random variable $\mathbf{z} = \mathbf{x} + j\mathbf{y}$, with $\mathbf{x}$ and $\mathbf{y}$ also being M-dimensional, is said to be multivariate *circularly-symmetric* (or spherically invariant) complex Gaussian distributed when the covariance matrix of real vector $\mathbf{z}'$,

$$\mathbf{z}' = \begin{pmatrix} \mathrm{Re}(\mathbf{z}) \\ \mathrm{Im}(\mathbf{z}) \end{pmatrix} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}, \tag{A.36}$$

has the special structure

$$\mathbf{Q}'_z = E[(\mathbf{z}'-\boldsymbol{\mu}'_z)(\mathbf{z}'-\boldsymbol{\mu}'_z)^H] = \frac{1}{2}\begin{pmatrix} \mathrm{Re}(\mathbf{Q}_z) & -\mathrm{Im}(\mathbf{Q}_z) \\ \mathrm{Im}(\mathbf{Q}_z) & \mathrm{Re}(\mathbf{Q}_z) \end{pmatrix}. \tag{A.37}$$

With

$$\mathbf{Q}_z = E[(\mathbf{z}-\boldsymbol{\mu}_z)(\mathbf{z}-\boldsymbol{\mu}_z)^H] = (\mathbf{Q}_x + \mathbf{Q}_y) + j(\mathbf{Q}_{yx} - \mathbf{Q}_{xy}), \tag{A.38}$$

we see that

$$\mathbf{Q}_x = \mathbf{Q}_y = \tfrac{1}{2}\mathrm{Re}(\mathbf{Q}_z) \text{ and } \mathbf{Q}_{xy} = -\mathbf{Q}_{yx} = -\tfrac{1}{2}\mathrm{Im}(\mathbf{Q}_z), \tag{A.39}$$

where $\mathbf{Q}_x$, $\mathbf{Q}_y$ and $\mathbf{Q}_{xy}$ are the covariance and cross-covariance matrices of the vectors $\mathbf{x}$ and $\mathbf{y}$. The fact that the covariance matrices of $\mathbf{x}$ and $\mathbf{y}$ are equal, explains the circular symmetry of $\mathbf{z}$.

To transform the real multivariate normal distribution of $\mathbf{z}'$ to the complex one of $\mathbf{z}$, we need a number of properties that exist for the mapping:

$$\mathbf{d} \rightarrow \mathbf{d}' = \begin{pmatrix} \mathrm{Re}(\mathbf{d}) \\ \mathrm{Im}(\mathbf{d}) \end{pmatrix} \text{ and } \mathbf{A} \rightarrow \mathbf{A}' = \begin{pmatrix} \mathrm{Re}(\mathbf{A}) & -\mathrm{Im}(\mathbf{A}) \\ \mathrm{Im}(\mathbf{A}) & \mathrm{Re}(\mathbf{A}) \end{pmatrix}. \tag{A.40}$$

These properties are:

$$
\begin{array}{ll}
(1) & \mathbf{C}=\mathbf{AB} \Leftrightarrow \mathbf{C'}=\mathbf{A'B'}, \\[4pt]
(2) & \mathbf{C}=\mathbf{A}^{-1} \Leftrightarrow \mathbf{C'}=\mathbf{A'}^{-1}, \\[4pt]
(3) & \det(\mathbf{A'})=\left|\det(\mathbf{A})\right|^{2}=\det(\mathbf{AA}^{*}), \\[4pt]
(4) & \mathbf{e}=\mathbf{Ad} \Leftrightarrow \mathbf{e'}=\mathbf{A'd'}, \\[4pt]
(5) & \operatorname{Re}(\mathbf{d}^{H}\mathbf{e})=\mathbf{d'}^{T}\mathbf{e'}.
\end{array}
\qquad (A.41)
$$

The properties (1), (4) and (5) are easily verified. The second property follows from the first and the fact that

$$
\mathbf{I}_{M} \Leftrightarrow \mathbf{I'}=\mathbf{I}_{2M}. \qquad (A.42)
$$

The third property follows from

$$
\begin{aligned}
\det(\mathbf{A'}) &= \det\!\left(\begin{pmatrix} \mathbf{I}_{M} & j\mathbf{I}_{M} \\ \mathbf{0}_{M} & \mathbf{I}_{M} \end{pmatrix}\mathbf{A'}\begin{pmatrix} \mathbf{I}_{M} & -j\mathbf{I}_{M} \\ \mathbf{0} & \mathbf{I}_{M} \end{pmatrix}\right) \\[6pt]
&= \det\!\left(\begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \operatorname{Im}(\mathbf{A}) & \mathbf{A}^{*} \end{pmatrix}\right)=\det(\mathbf{A})\det(\mathbf{A})^{*}.
\end{aligned}
\qquad (A.43)
$$

where the * stands for the conjugate of a vector/matrix (and not the Hermitian transpose of a vector/matrix).

Using (A.37), we can say that $\mathbf{Q'}_{z}=\mathbf{A'}/2$ and that $\mathbf{Q}_{z}=\mathbf{A}$. Now, to make the transformation more clear, we split the probability density function of the real multivariate normal distribution of $\mathbf{z'}$, given by (A.35) with $\mathbf{x}=\mathbf{z'}$, in two terms. The first term to transform is the determinant and the second is the exponent. The transformation of the determinant term is based on the third property of (A.41):

$$
\det(2\pi\mathbf{Q'}_{z})^{-1/2}=\det(\pi\mathbf{A'})^{-1/2}=\left(\left|\det(\pi\mathbf{A})\right|^{2}\right)^{-1/2}=\det(\pi\mathbf{Q}_{z})^{-1}. \qquad (A.44)
$$

where the last equality is obtained using the fact that the covariance of the complex vector $\mathbf{z}$, $\mathbf{Q}_{z}$, is a nonnegative definite matrix.

The second term of the probability density function, i.e., the exponent term, can be transformed to the complex case by using the first, second, fourth and fifth property of (A.41) and the substitutions $\mathbf{y'}=\mathbf{z'}-\boldsymbol{\mu'}_{z}$ and $\mathbf{x'}=\mathbf{A'}^{-1}\mathbf{y'}$:

$$
\begin{aligned}
\exp\!\left(-\tfrac{1}{2}(\mathbf{z'}-\boldsymbol{\mu'}_{z})^{T}\mathbf{Q'}_{z}^{-1}(\mathbf{z'}-\boldsymbol{\mu'}_{z})\right) &= \exp\!\left(-\mathbf{y'}^{T}\mathbf{A'}^{-1}\mathbf{y'}\right) \\[4pt]
&= \exp\!\left(-\mathbf{y'}^{T}\mathbf{x'}\right)=\exp\!\left(-\operatorname{Re}(\mathbf{y}^{H}\mathbf{x})\right) \\[4pt]
&= \exp\!\left(-\operatorname{Re}\!\left((\mathbf{z}-\boldsymbol{\mu}_{z})^{H}\mathbf{A}^{-1}(\mathbf{z}-\boldsymbol{\mu}_{z})\right)\right) \\[4pt]
&= \exp\!\left(-(\mathbf{z}-\boldsymbol{\mu}_{z})^{H}\mathbf{Q}_{z}^{-1}(\mathbf{z}-\boldsymbol{\mu}_{z})\right).
\end{aligned}
\qquad (A.45)
$$

where the last equality is obtained from the fact that, for a Hermitian matrix $\mathbf{Q}$ and for all complex vectors $\mathbf{d}$, the number $\mathbf{d}^H\mathbf{Q}\mathbf{d}$ is real. As a proof, we can compute $(\mathbf{d}^H\mathbf{Q}\mathbf{d})^H$. We expect to get the conjugate of the 1 by 1 matrix $\mathbf{d}^H\mathbf{Q}\mathbf{d}$, but instead, we get the same number back: $(\mathbf{d}^H\mathbf{Q}\mathbf{d})^H = \mathbf{d}^H\mathbf{Q}^H(\mathbf{d}^H)^H = \mathbf{d}^H\mathbf{Q}\mathbf{d}$. Therefore, the number must be real.

Summarising, the probability density function (pdf) of a vector $\mathbf{z}$ with circularly-symmetric complex Gaussian distributed elements equals

$$p(\mathbf{z}) = \det(\pi\mathbf{Q}_z)^{-1}\exp\left(-(\mathbf{z}-\boldsymbol{\mu}_z)^H\mathbf{Q}_z^{-1}(\mathbf{z}-\boldsymbol{\mu}_z)\right),\qquad(A.46)$$

where $\mathbf{Q}_z$ represents the covariance matrix and $\boldsymbol{\mu}_z$ the mean of $\mathbf{z}$. Note that this distribution is fully characterised by these two quantities.

Appendix B

Test System Specifications

B.1 MBA-5 5 GHz Miniature Broadband Antenna Datasheet

B.1.1 Introduction to MBA-5

The Ascom MBA-5 (Miniature Broadband Antenna) is a wideband antenna designed primarily for wireless LAN applications in the 5GHz band ([10]). It covers all the US, European and Japanese 5GHz WLAN bands simultaneously. This datasheet gives a complete overview of the performance of the MBA-5 in terms of bandwidth and radiation pattern. Most of the information required for the integration of the antenna into new designs is given. A picture of the antenna is given in Figure B-1.
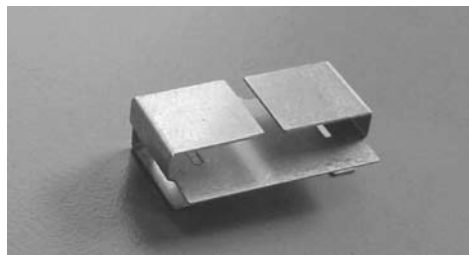
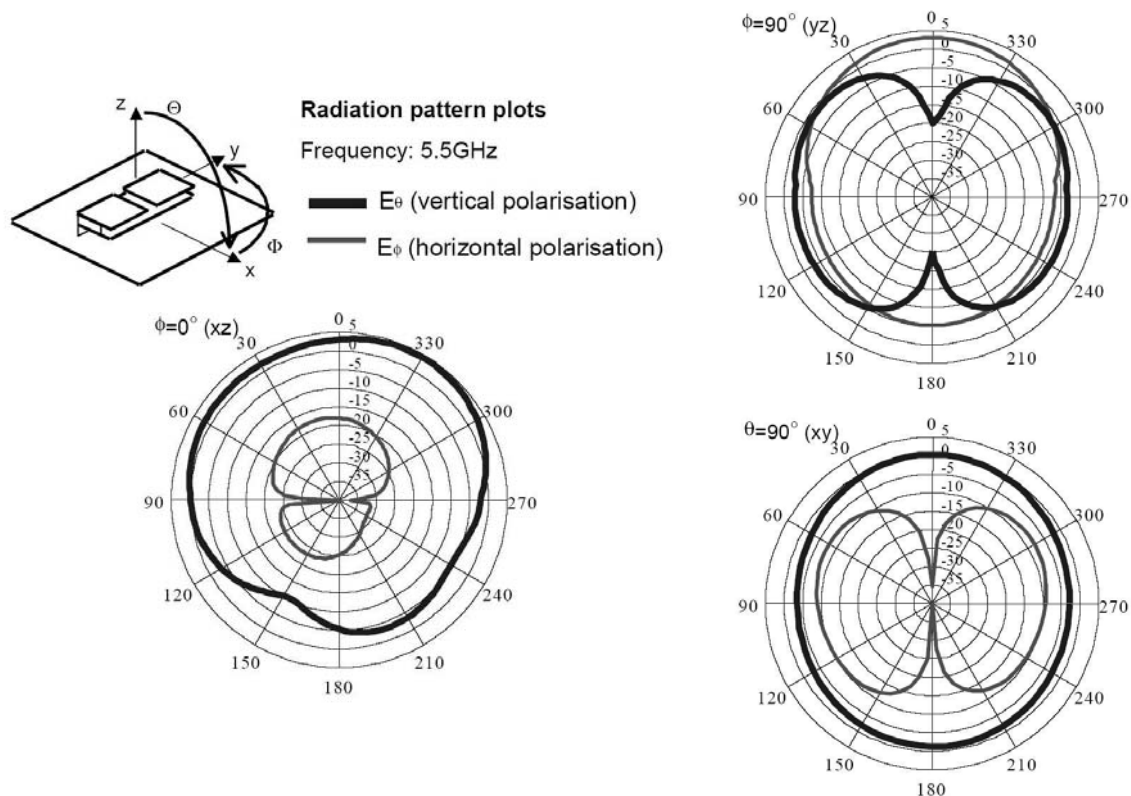
Figure B-1: the MBA-5 antenna.

B.1.2 Advantages

Electrical and Radiation Performance

The measured return loss bandwidth of the MBA-5 (VSWR 2:1 or better) extends from 5.1 to 6.3 GHz for an unpackaged antenna. When packaged into plastic the bandwidth shifts

downwards by about 200 MHz, thus covering 4.9 to 6.1 GHz, i.e. all 5 GHz bands worldwide simultaneously.

The radiation pattern of the MBA-5 in the azimuth plane resembles that of a dipole – omnidirectional with no nulls and a peak gain of between 1dBi and 0dBi depending on frequency (see Figure B-2). Peak gain of around 4dBi occurs in the elevation plane, directed upwards.

Figure B-2: Radiation patterns of the MBA-5 antenna

Mechanical Properties

The MBA-5 is a surface-mount device requiring no special treatment and may be easily integrated into a design flow. Its small dimensions of approximately 15mm × 10mm × 5mm make it perfect for PC-card integration. An array of up to 3 antennas will fit comfortably onto a PC-card extension, thus opening up the possibility for antenna spatial diversity.

How it Works

In brief, the key to the extremely wide bandwidth of the MBA-5 are the two coupled slots inherent in the shape of the antenna. Two separate resonances are excited, which combine to create the exceptionally broad bandwidth.

Performance Optimization

The small size of the Ascom MBA-5 means that the shape of its reflection loss curve is affected quite strongly by its surroundings. Results shown in Figure B-3 are for an MBA-5 on a ground plane of 30 mm x 30 mm.

| MBA-5	4.9 GHz to 6.1 GHz Linear Polarised Miniature Broadband Antenna		
Electrical			
		5.1GHz	5.5GHz	5.8GHz
Return Loss	dB	-11	-27	-19
Radiation Efficiency	%	90	99	80
Peak Gain Azimuth	dBi	-0.9	0.3	-0.2
Peak Gain Elevation	dBi	4.0	4.2	3.0
Mechanical			
Dimensions (length x width x height)	15 x 10 x 5 (mm) 0,59 x 0,39 x 0,20 (inch)		
Feed	Coplanar waveguide or microstrip line		
Packaging	Tape reel or tray		
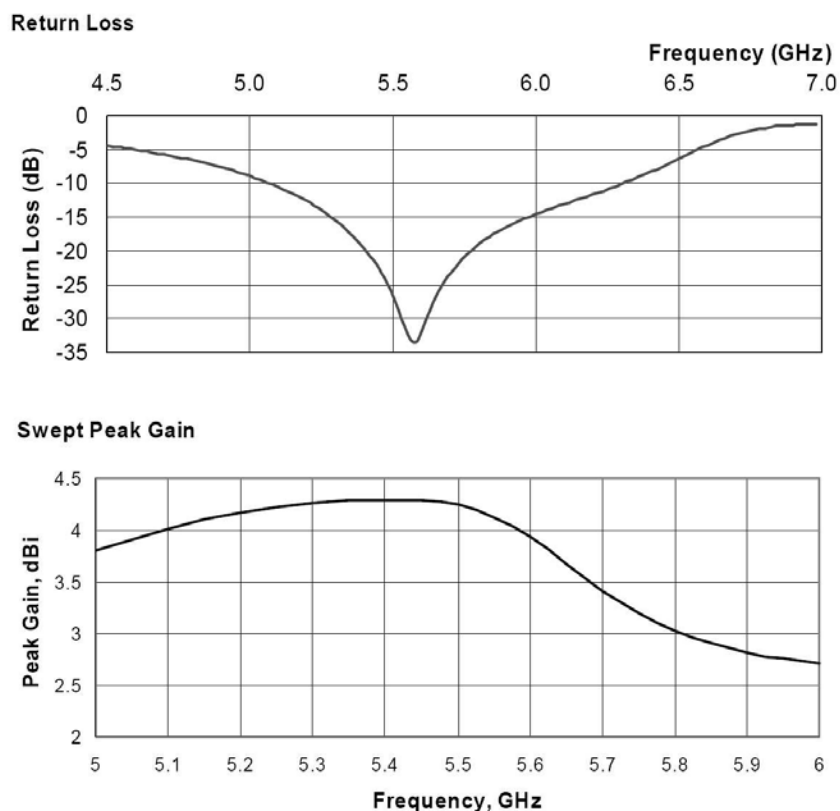
Figure B-3: Results for an unpackaged antenna. When packaged into plastic the bandwidth shifts downwards by about 200 MHz, thus covering 4.9 to 6.1 GHz, i.e., all 5GHz bands worldwide simultaneously.

B.2   Block Diagrams of the IF Stages of the Test System
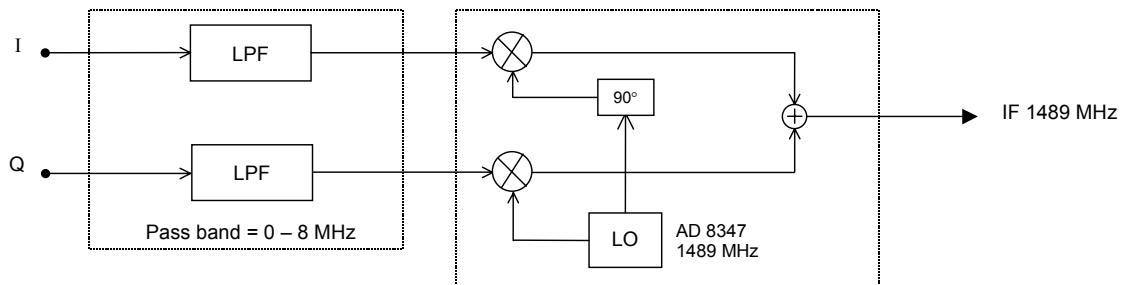
Figure B-4: Block diagram of the baseband to IF conversion per TX branch. The baseband I and Q signals are fed through a Low-Pass Filter (LPF), mixed up to IF, and combined after that.
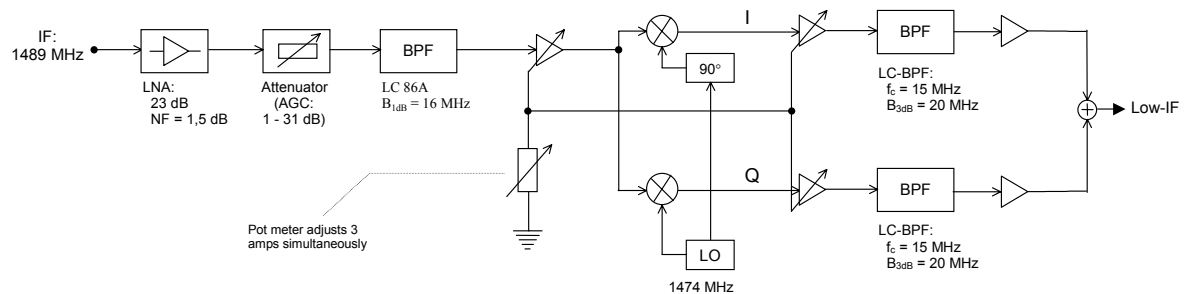
Figure B-5: Block diagram of the IF to low-IF conversion per RX branch. The IF input signal is properly scaled by the LNA and AGC to use the full resolution of the ADC. Then the signal passes a Band-Pass Filter (BPF). Finally, a circuitry of amplifiers, LO, and filters mix the signal down to low-IF.

Appendix C

Complexity Analysis

C.1 Introduction

In this appendix, the complexity is analysed for the SDM algorithms introduced in Chapter 4: ZF, MMSE, ZF with SIC, MMSE with SIC, and MLD. The complexity figures are split in a complexity number for the preamble processing and a complexity number for the payload processing. The reason for this is that it is assumed that data is transferred on a packet-by-packet basis and that during the transmission of a packet the communication channel **H** stays constant (i.e., the MIMO system is operating in a quasi-static environment). Therefore, processes that do not depend on the information in the payload can be performed in the preamble phase.

Before we determine the complexity of the SDM algorithms, we introduce a number of general rules, namely, the complexity of a matrix multiplication, the conversion from complex complexity figures to real complexity figures, the complexity of a slicer, and the complexity of finding a minimum value from a set of values.

The *complexity of a matrix product* is determined as follows. Suppose two matrices **A** and **B** (real or complex) with dimensions $C \times D$ and $D \times E$ are multiplied, then the (i, l)-th element of the resulting matrix is given by

$$\mathbf{a}^i \mathbf{b}_l = \sum_{k=1}^{D} a_{ik} b_{kl} \tag{C.1}$$

where $\mathbf{a}^i$ represents the i-th row of matrix **A**, $\mathbf{b}_l$ denotes the l-th column of **B** and a_{ik} and b_{kl} stand for the k-th element of this row and column, respectively. Thus, in order to obtain one element of the resulting matrix, $D - 1$ additions and D multiplications need to be performed. The resulting matrix is $C \times E$ dimensional and, therefore, a total of $C(D - 1)E$ additions and CDE multiplications are needed to multiply the two matrices **A** and **B**.

To write complex additions and complex multiplications in terms of *real additions and real multiplications*, it is easily verified that one complex addition (C_ADD) consists of two real additions (R_ADDs); the real *and* the imaginary part of the two complex numbers are added. Furthermore, a complex multiplication (C_MUL) can be rewritten in the following two ways ([89]):

$$(a + jb)(c + jd) = (ac - bd) + j(bc + ad),$$ (C.2)

$$(a + jb)(c + jd) = (ac - bd) + j((a + b)(c + d) - ac - bd).$$ (C.3)

The first option consists of 4 real multiplications (R_MULs), *ac*, *bd*, *bc* and *ad*, and 2 R_ADDs, *ac − bd* and *bc + ad*. A subtraction is counted as an addition and the addition before the *j* does not count because the real and imaginary parts are stored separately. The second option has only three real multiplications (*ac*, *bd*, (*a* + *b*)(*c* + *d*)), plus five real additions. Compared with the first case, the total operations count is higher by two, but in a number of hardware implementations, a multiplication is a more complex operation. In the remainder of this appendix, however, we will use the first option.

The *complexity of a slicer* is minimal in terms of additions and/or multiplications. For an *M*-PSK constellation scheme, the phase range $[-\pi, \pi]$ is divided in *M* equal parts. In such a regular structure, we can recursively search in which half of the (remaining) range the phase of the estimated symbol best fits. This results in a complexity equivalent to $\log_2(M)$ comparisons. For an *M*-QAM constellation diagram, we can split the real and imaginary part. Each of these parts is regularly divided in $\sqrt{M}$ slicing ranges. Also in this case, we can recursively search in which half of the (remaining) range the real or imaginary part of the estimated symbol best fits, and the complexity is equal to $\log_2(\sqrt{M})$ comparisons for the real and for the imaginary part, or $2 \cdot \log_2(\sqrt{M}) = \log_2(M)$ comparisons in total. It is reasonable to assume that a comparison is as complex as a real addition and, therefore, the slicing of the N_t-dimensional vector $\mathbf{s}_{est}$ requires at most $N_t \cdot \log_2(M)$ R_ADDs.

In order to find the *minimum* of *N* numbers in hardware, the easiest thing to do is start with the first two elements, subtract the second number from the first, and compare the result with zero. If the result is larger than zero, the second number is the smallest, otherwise the first number is the smallest, etc. Obviously, finding the minimum between two real numbers has the complexity of one real addition. As a result, determining the minimum of *N* values has a complexity of *N* − 1 real additions.

Based on these general rules, the complexity of ZF, MMSE, ZF with SIC, MMSE with SIC, and MLD is determined, respectively, in Appendix C.2, Appendix C.3, Appendix C.4, Appendix C.5, and Appendix C.6.

C.2   Complexity of ZF

As described in Subsection 4.6.1, the Zero Forcing technique is based on calculation of the pseudo-inverse of the channel transfer matrix **H**. Because it is assumed that the MIMO system is operating in a quasi-static environment, i.e., **H** is constant during a packet transmission, the pseudo-inverse of **H** needs to be calculated only once per packet. The

pseudo-inverse can be calculated after the channel training in the preamble processing. During the payload processing, the pseudo-inverse is used for the estimation of every transmitted MIMO vector **s** of the corresponding packet. In this section the complexity of the ZF algorithm in the preamble and payload processing is determined.

Complexity in the preamble processing

During the preamble processing, the pseudo-inverse of the channel matrix **H** is determined. For determining the complexity of the calculation of the pseudo-inverse, we will use the equality that can be deduced from Subsection 4.6.1 and is given by

$$\mathbf{H}^{\dagger} = \left(\mathbf{H}^{H}\mathbf{H}\right)^{-1}\mathbf{H}^{H} . \tag{C.4}$$

The dimensions of $\mathbf{H}^{\dagger}$, $\mathbf{H}$ and $\mathbf{H}^{H}$ are $N_t \times N_r$, $N_r \times N_t$ and $N_t \times N_r$, respectively. To find the pseudo-inverse of **H**, we first need to determine the complexity of the matrix product $\mathbf{H}^{H}\mathbf{H}$. To determine this complexity, we will use the general rule introduced in Appendix C.1 that states that the complexity of the product of two matrices **A** and **B** (real or complex) with dimensions $C \times D$ and $D \times E$ equals $C(D-1)E$ additions and CDE multiplications (real or complex). Hence, the complexity of the matrix product $\mathbf{H}^{H}\mathbf{H}$ yields $N_t^2(N_r-1)$ C_ADDs and $N_t^2 N_r$ C_MULs. The result is a square matrix with dimensions $N_t \times N_t$.

From this square matrix $\mathbf{H}^{H}\mathbf{H}$, the inverse needs to be determined. In [89] it is shown that the direct inversion of a given square matrix **A** (with dimensions $N \times N$) has a complexity in the order of N^3 additions and N^3 multiplications in total. So, inverting $\mathbf{H}^{H}\mathbf{H}$ has a complexity of N_t^3 C_ADDs and N_t^3 C_MULs.

Finally, the inverse of $\mathbf{H}^{H}\mathbf{H}$ (which is $N_t \times N_t$ dimensional) is multiplied by $\mathbf{H}^{H}$. The complexity of this last multiplication is equal to $N_t(N_t-1)N_r$ C_ADDs and $N_t^2 N_r$ C_MULs (see Appendix C.1). This leads to a total complexity of $N_t^3 + N_t^2(N_r-1) + N_t(N_t-1)N_r$ C_ADDs and $N_t^3 + 2N_t^2 N_r$ C_MULs in the training phase, or (based on a general rule introduced in Appendix C.1), the complexity in terms of real operations equals

$$\begin{aligned} 4N_t^3 + 4N_t^2 N_r + 2N_t^2\left(N_r-1\right) + 2N_t\left(N_t-1\right)N_r \\ = 4N_t^3 + N_t^2\left(8N_r-2\right) - 2N_t N_r \ \ \text{R\_ADDs}, \end{aligned} \tag{C.5}$$

and $4N_t^3 + 8N_t^2 N_r$ R_MULs.

Complexity in the payload processing

The payload processing for ZF consists of a matrix-vector multiplication per transmitted vector and a slicing step to translate the estimated elements of **s** to the possible transmitted symbols. Recalling from Subsection 4.6.1, the matrix-vector multiplication is given by

$$\mathbf{s}_{\text{est}} = \mathbf{H}^{\dagger}\mathbf{x} . \tag{C.6}$$

The complexity of this product is equal to $N_t(N_r - 1)$ complex additions and $N_t N_r$ complex multiplications.

As explained in Appendix C.1, the complexity of slicing N_t M-ary constellation points equals $N_t \cdot \log_2(M)$ R_ADDs.

Summarising, the complexity of the ZF algorithm during the payload processing and per transmitted vector **s** is equal to $N_t(N_r - 1)$ C_ADDs + $N_t \cdot \log_2(M)$ R_ADDs and $N_t N_r$ C_MULs, or equivalently, $2N_t N_r + 2N_t(N_r - 1) + N_t \cdot \log_2(M)$ R_ADDs and $4N_t N_r$ R_MULs. When N_s vectors are transmitted within a packet, these numbers must be multiplied by N_s to obtain the complexity per packet.

C.3   Complexity of MMSE

The complexity of the MMSE algorithm is almost equal to the complexity of the ZF method described in the previous section.

In the preamble-processing phase, the following MIMO processing matrix needs to be determined (see Subsection 4.7.1):

$$\mathbf{W} = \left(\alpha \mathbf{I}_{N_t} + \mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^H . \tag{C.7}$$

The calculation of this matrix has almost the same complexity as the determination of the pseudo-inverse in case of the ZF algorithm. Since α is real, the only additional complexity consists of the N_t real additions of α (i.e., the addition of α to the real part of the diagonal elements of $\mathbf{H}^H \mathbf{H}$). This leads to a total complexity in the preamble-processing phase of $4N_t^3 + N_t^2(8N_r - 2) - 2N_t N_r + N_t$ R_ADDs and $4N_t^3 + 8N_t^2 N_r$ R_MULs.

The complexity of MMSE during the payload processing is equal to that of ZF and consists of a matrix-vector product with the same dimensions and slicing. Recalling from the previous section, the payload complexity equals $2N_t N_r + 2N_t(N_r-1) + N_t \cdot \log_2(M)$ R_ADDs and $4N_t N_r$ R_MULs for every transmitted vector **s** to decode.

C.4   Complexity of ZF with SIC

The processing of the ZF with SIC algorithm can be divided into two parts: the processing during the preamble and processing of the payload. Based on the assumption that the MIMO channel is static during a packet transmission, the ordering and the weight vectors can be determined during the preamble processing. During the payload processing the actual detection and SIC is performed.

Complexity in the preamble processing

In order to find the weighting vectors, an iterative algorithm that consists of two steps can be performed. First the steps are described and then the complexity will be determined:

1. Compute the pseudo-inverse of **H**, $\mathbf{H}^\dagger$. Find the minimum squared length row of $\mathbf{H}^\dagger$. This row is a weight vector. Permute it to be the last row and permute the columns of **H** accordingly.

2. (While $N_t - 1 > 0$) go back to step 1, but now with:

$$\mathbf{H} \to \mathbf{H}^{(N_t - 1)} = \begin{pmatrix} \mathbf{h}_1 & \cdots & \mathbf{h}_{N_t - 1} \end{pmatrix} \text{ and } N_t \to N_t - 1. \tag{C.8}$$

The complexity of calculating the pseudo-inverse is already determined in Appendix C.2. For an $N_r \times N_t$ dimensional matrix $\mathbf{H}$, it equals $4N_t^3 + N_t^2(8N_r - 2) - 2N_tN_r$ R_ADDs and $4N_t^3 + 8N_t^2N_r$ R_MULs.

The next steps are the calculation of the squared length of all rows of $\mathbf{H}^\dagger$ and the determination of the minimum squared length row. Note that, according to Subsection 4.8.1 finding the minimum squared length row of $\mathbf{H}$ is equal to finding the minimum element $\mathbf{P}_{pp}$ on the diagonal of $\mathbf{P}$, $p = 1, \ldots, N_t$. Since $\mathbf{P}$ is obtained through the computation of the pseudo-inverse of $\mathbf{H}$, the complexity of these steps consists only of finding the minimum. As explained in Appendix C.1, finding a minimum of N_t values N_t values has a complexity of $N_t - 1$ real additions.

The permutations of step 1 are considered to have no complexity. The only thing that needs to be done is exchanging the memory pointers that respectively point to the two rows of $\mathbf{H}^\dagger$ and the two columns of $\mathbf{H}$ that need to be permuted.

Since the algorithm is an iterative algorithm, and the dimensions of the used matrices scale down, the complexity per iteration is reduced. To take along this reduction in complexity during the iterations, the total complexity can be written by using series. The final number of real additions can be shown to be equal to

$$
\begin{aligned}
4\sum_{p=1}^{N_t} p^3 &+ (8N_r - 2)\sum_{p=1}^{N_t} p^2 - 2N_r \sum_{p=1}^{N_t} p + \sum_{p=1}^{N_t}(p-1) \\
&= 4\sum_{p=1}^{N_t} p^3 + (8N_r - 2)\sum_{p=1}^{N_t} p^2 + (1 - 2N_r)\sum_{p=1}^{N_t} p - N_t \\
&= N_t^2(N_t + 1)^2 + (8N_r - 2)\frac{N_t(N_t + 1)(2N_t + 1)}{6} + (1 - 2N_r)\frac{N_t(N_t + 1)}{2} - N_t \\
&= \frac{1}{6}N_t\left(6N_t^3 + 8N_t^2(1 + 2N_r) + 3N_t(1 + 6N_r) + 2N_r - 5\right) \text{R\_ADDs}.
\end{aligned}
\tag{C.9}
$$

The total number of real multiplications of the preamble phase of ZF with SIC equals

$$
\begin{aligned}
4\sum_{p=1}^{N_t} p^3 &+ 8N_r \sum_{p=1}^{N_t} p^2 \\
&= N_t^2(N_t + 1)^2 + 8N_r \frac{N_t(N_t + 1)(2N_t + 1)}{6} \text{R\_MULs}.
\end{aligned}
\tag{C.10}
$$

Complexity in the payload processing

During the data processing the weighting vectors are used to first estimate the best element of the transmitted vector $\mathbf{s}$. The result is sliced to find a hard-decision value of the

transmitted constellation symbol and then this symbol is used in the feedback loop in order to find the next estimate. The following steps represent this iterative process:

1. Form the estimate of the best component p of $\mathbf{s}$. Due to the permutation the corresponding weight vector equals the N_t-th row of permuted $\mathbf{H}^\dagger$. In case of ZF:

$$\left(\mathbf{s}_{\text{est}}\right)_p = \mathbf{w}^{N_t}\mathbf{x} . \tag{C.11}$$

   Slice $(\mathbf{s}_{\text{est}})_p$ to the nearest constellation point $(\mathbf{s}_{\text{est,sliced}})_p$.

2. (While $N_t - 1 > 0$) go back to step 1, but now with

$$\mathbf{x} \rightarrow \mathbf{x} - \mathbf{h}_{N_t}\left(\mathbf{s}_{\text{est,sliced}}\right)_p \text{ and } N_t \rightarrow N_t - 1 . \tag{C.12}$$

The complexity of the first step of this iterative algorithm equals $N_r - 1$ C_ADDs and N_r C_MULs, because two N_r-element vectors are multiplied. The slicing step for an M-ary constellation has a complexity of $\log_2(M)$ R_ADDs as is explained in Appendix C.1. Step 2 consists of a scalar-vector product and a vector subtraction. The scalar-vector product has a complexity that is equal to N_r C_MULs and the complexity of the vector subtraction is N_r C_ADDs, since the vectors have N_r elements.

Because above steps are performed N_t times, it can be said that the complexity of the payload processing of Zero Forcing with Successive Interference Cancellation equals $2N_t(4N_r - 1) + N_t \cdot \log_2(M)$ R_ADDs and $8N_tN_r$ R_MULs per transmitted vector $\mathbf{s}$.

Finally, it may be interesting to note that an efficient low-complexity implementation of the SIC principle can be found in [49].

C.5   Complexity of MMSE with SIC

The complexity of MMSE with SIC can be determined in the same way as is done for Zero Forcing with SIC in the previous section. Compared to ZF with SIC, there is a slight difference in the preamble processing, namely in the determination of the weight vectors. In case of MMSE, the iterative process of the weight calculation is given by

1. Compute the weight matrix $\mathbf{W} = \mathbf{P}\mathbf{H}^H$, with $\mathbf{P} = (\alpha\mathbf{I} + \mathbf{H}^H\mathbf{H})^{-1}$. Find the smallest diagonal entry of $\mathbf{P}$ and suppose this is the p-th entry. Permute the p-th row of $\mathbf{W}$ to be the last row and permute the columns of $\mathbf{H}$ accordingly. The permuted row of $\mathbf{W}$ is a weight vector.

2. (While $N_t - 1 > 0$) go back to step 1, but now with:

$$\mathbf{H} \rightarrow \mathbf{H}^{(N_t-1)} = \left(\mathbf{h}_1 \quad \cdots \quad \mathbf{h}_{N_t-1}\right) \text{ and } N_t \rightarrow N_t - 1 . \tag{C.13}$$

Compared to ZF with SIC, the complexity of step 1 is slightly higher, because of the addition of $\alpha\mathbf{I}$ to $\mathbf{H}^H\mathbf{H}$. Since α is real, the only additional complexity consists of the N_t

real additions of α (i.e., the addition of α to the real part of the diagonal elements of $\mathbf{H}^H\mathbf{H}$). This leads to a complexity of $4N_t^3 + N_t^2(8N_r - 2) - 2N_tN_r + N_t$ R_ADDs and $4N_t^3 + 8N_t^2N_r$ R_MULs. Taking all iterations of the algorithm into account, this leads to a total complexity (including the complexity of finding the minimal diagonal element of $\mathbf{P}$) of

$$
\begin{aligned}
4\sum_{p=1}^{N_t} p^3 &+ \left(8N_r - 2\right)\sum_{p=1}^{N_t} p^2 + \left(1 - 2N_r\right)\sum_{p=1}^{N_t} p + \sum_{p=1}^{N_t}\left(p-1\right) \\
&= \frac{1}{3}N_t\left(3N_t^3 + 4N_t^2\left(1 + 2N_r\right) + 3N_t\left(1 + 3N_r\right) + N_r - 1\right) \text{ R\_ADDs},
\end{aligned}
$$

(C.14)

and

$$
\begin{aligned}
4\sum_{p=1}^{N_t} p^3 &+ 8N_r\sum_{p=1}^{N_t} p^2 \\
&= N_t^2\left(N_t + 1\right)^2 + 8N_r\frac{N_t\left(N_t + 1\right)\left(2N_t + 1\right)}{6} \text{ R\_MULs}.
\end{aligned}
$$

(C.15)

For the MMSE algorithm with SIC, the payload processing is equivalent to the ZF technique with SIC, except that in the former case the weight vectors are rows of the processing matrix $\mathbf{W}$ in stead of rows of the pseudo-inverse of $\mathbf{H}$. The last fact is irrelevant for the complexity, thus, the complexity of the payload processing of MMSE with SIC equals $2N_t(4N_r - 1) + N_t \cdot \log_2(M)$ R_ADDs and $8N_tN_r$ R_MULs per transmitted vector $\mathbf{s}$ (see Appendix C.4).

C.6 Complexity of MLD

Since the search in MLD goes over all possible transmitted vectors $\mathbf{s}_i$, with $i = 1, \ldots, I$, the complexity of MLD is proportional to the number of candidates I. Furthermore, the search is performed in the "$\mathbf{x}$-space" and therefore each candidate $\mathbf{s}_i$ has to be multiplied by the channel matrix $\mathbf{H}$. Note that we can write this matrix-vector product of the i-th candidate as

$$
\sum_{p=1}^{N_t} \mathbf{h}_p \left(\mathbf{s}_i\right)_p ,
$$

(C.16)

where $\mathbf{h}_p$ denotes the p-th column of $\mathbf{H}$. Since all elements of $\mathbf{s}_i$ are taken from the set of M constellation points and, thus, a given element is taken from the set $\{s_1, \ldots, s_M\}$, we can obtain all MLD candidates following the tree of Figure C-1.

When the channel can be assumed static during a packet transmission, it is efficient to store all candidates in the memory during the preamble processing and use them in the payload processing. From Figure C-1 it can be observed, however, that the amount of candidates grows exponentially with the number of TX antennas. When the memory is not large enough, not all candidates can be stored and they have to be calculated for every transmitted vector to perform MLD. So, clearly the complexity of MLD is largely

dependent on the amount of memory that is available or that one is willing to use. Two extreme cases will be compared in this section:

- *Minimum amount of memory*: just enough memory is available to store the products of the columns of **H** times the constellation points. For instance, for BPSK enough memory should be implemented to store ($\mathbf{h}_p$, $-\mathbf{h}_p$) for all $1 \leq p \leq N_t$. Note that in this case it is not really necessary to store $-\mathbf{h}_p$, because it can be deduced easily from $\mathbf{h}_p$ and, thus, two times less memory is required. Also for other constellation diagrams their symmetry can be exploited. Furthermore, note that in this case for every vector within the payload, all possible combinations of $\mathbf{H}\mathbf{s}_i$, with $1 \leq i \leq I$, have to be determined over and over again, leading to a significant complexity penalty.

- *Maximum amount of memory*: enough memory is available to store all possible combinations of $\mathbf{H}\mathbf{s}_i$, with $1 \leq i \leq I$, during the training phase. Note that, since I grows exponentially with N_t, a large memory may be required.

Next, the complexity of above cases will be determined both for the preamble processing and for the payload processing.
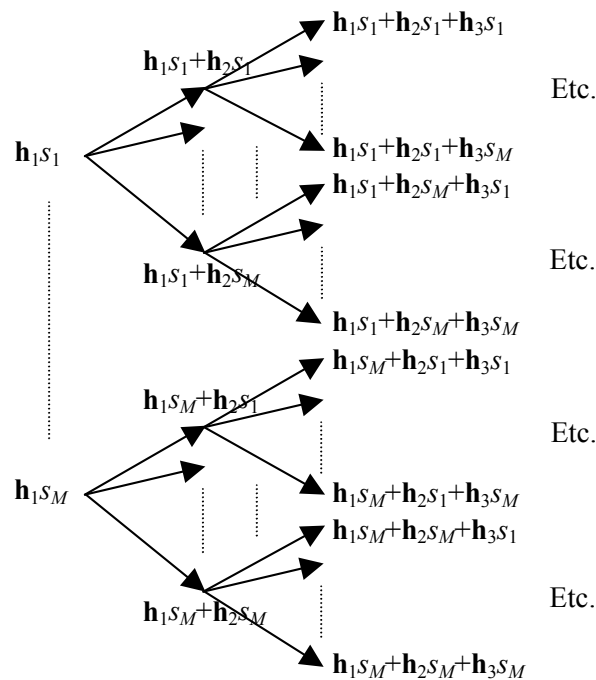
Figure C-1: All candidates for MLD.

Complexity in the preamble processing (minimum amount of memory)

In the case that hardly any memory is used for Maximum Likelihood Decoding, a large part of the processing takes place in the data phase. The only processing needed in the training phase with respect to MLD is the calculation of the products of the constellation values with the columns of **H**. If there are M constellation points, then the complexity of determining these products is MN_rN_t C_MULs or $2MN_rN_t$ R_ADDs and $4MN_rN_t$ R_MULs.

Complexity in the payload processing (minimum amount of memory)

As shown in Subsection 4.10.1, the maximum likelihood detection per transmitted vector is given by

$$\mathbf{s}_{\mathrm{ml}} = \arg \min_{\mathbf{s}_i \in \{\mathbf{s}_1, \ldots, \mathbf{s}_I\}} \left\| \mathbf{x} - \mathbf{H}\mathbf{s}_i \right\|^2 = \arg \min_{\mathbf{s}_i \in \{\mathbf{s}_1, \ldots, \mathbf{s}_I\}} \left\| \mathbf{x} - \sum_{p=1}^{N_t} \mathbf{h}_p (\mathbf{s}_i)_p \right\|^2 . \quad (C.17)$$

Because we assume that enough memory is present to at least store all the products of the constellation points with the columns of $\mathbf{H}$, the $\mathbf{h}_p s$'s, only complex additions are required to find $\mathbf{x} - \mathbf{H}\mathbf{s}_i$. Starting at $\mathbf{x}$ and then with $\mathbf{x} - \mathbf{h}_1 s_1$ to $\mathbf{x} - \mathbf{h}_1 s_M$, we can further use the tree as shown in Figure C-1 to calculate all $\mathbf{x} - \mathbf{H}\mathbf{s}_i$, $i = 1, \ldots, I$. The complexity equals

$$\sum_{p=1}^{N_t} M^p N_r = N_r M \frac{M^{N_t} - 1}{M - 1} \text{ C\_ADDs.} \quad (C.18)$$

Calculating the squared norm of the result can be shown to have a complexity of $2N_r$ real multiplications and $2N_r - 1$ real additions. Since we have I candidates, the calculation of the squared norm needs to be performed I times. Finally, the minimum of the I possible squared norms must be obtained which introduces a complexity of $I - 1$ extra real additions (see Appendix C.1). So, in total this yields a complexity per transmitted vector of $2N_r I$ R_MULs and

$$2N_r M \frac{I - 1}{M - 1} + 2N_r I - 1 \text{ R\_ADDs.} \quad (C.19)$$

If a packet consists of N_s vectors, the total complexity of the payload processing is obtained by multiplying above numbers by N_s.

Complexity in the preamble processing (maximum amount of memory)

If there is enough memory available to store all candidates, they can be determined during the preamble processing and used in the data phase. The complexity of the matrix-vector product $\mathbf{H}\mathbf{s}_i$ can be obtained using the tree of Figure C-1. Starting with $\mathbf{h}_1 s_1$ to $\mathbf{h}_1 s_M$, we can further use the tree to calculate all $\mathbf{H}\mathbf{s}_i$, $i = 1, \ldots, I$. Thus, first, the products of every column of $\mathbf{H}$ with every constellation point value should be determined. This has a total complexity of $2MN_r N_t$ R_ADDs and $4MN_r N_t$ R_MULs. After that, the additions as shown in the tree have to be performed. This has a complexity of

$$\sum_{p=2}^{N_t} M^p N_r = M^2 N_r \frac{M^{N_t - 1} - 1}{M - 1} \text{ C\_ADDs.} \quad (C.20)$$

This gives a total complexity of $4MN_r N_t$ R_MULs and

$$2M^2 N_r \frac{M^{N_t - 1} - 1}{M - 1} + 2MN_r N_t \text{ R\_ADDs.} \quad (C.21)$$

Complexity in the payload processing (maximum amount of memory)

During the data phase the vector subtraction $\mathbf{x} - \mathbf{H}\mathbf{s}_i$ and the squared norm of the result have to be determined for the I possible $\mathbf{s}$ vectors. The next step is to obtain the minimum of the squared norms. The vector subtraction $\mathbf{x} - \mathbf{H}\mathbf{s}_i$ is performed for $i = 1, \ldots, I$ and has a complexity equal to $N_r I$ C_ADDs. Next, the norms of the I results have to be determined. This has a complexity of $(2N_r-1)I$ R_ADDs and $2N_r I$ R_MULs. After that, the minimum of the I (real) norms must be obtained which has a complexity equal to $I - 1$ R_ADDs. So, this yields a total complexity of $4N_r I - 1$ R_ADDs and $2N_r I$ R_MULs per transmitted vector.

To obtain the complexity for the entire packet, these complexity numbers have to be multiplied by the number of spatial vectors within a packet, N_s. Note that these complexity figures increase linearly with the number of receiving antennas and exponentially with the number of transmit antennas.

It is possible to reduce this complexity considerably without too much loss in performance. From above analysis, it is clear that one of the significant terms in the complexity calculation for MLD is the determination of the norm of $\mathbf{x} - \mathbf{H}\mathbf{s}_i$ (commonly known as the squared l_2 norm ([54]): $\|\mathbf{x} - \mathbf{H}\mathbf{s}_i\|^2$). In order to reduce the complexity, an approximation of the l_1 norm can be used:

$$\|\mathbf{x} - \mathbf{H}\mathbf{s}_i\| \approx$$
$$\left|\text{Re}\left(x_1 - \mathbf{h}^1\mathbf{s}_i\right)\right| + \left|\text{Im}\left(x_1 - \mathbf{h}^1\mathbf{s}_i\right)\right| + \ldots + \left|\text{Re}\left(x_{N_r} - \mathbf{h}^{N_r}\mathbf{s}_i\right)\right| + \left|\text{Im}\left(x_{N_r} - \mathbf{h}^{N_r}\mathbf{s}_i\right)\right|, \quad \text{(C.22)}$$

where $\mathbf{h}^q$ stands for the q-th row of $\mathbf{H}$. This approximation consists only of real additions (and no multiplications) making the MLD algorithm less complex. The drawback of MLD with the approximated norm is that the BER performance deteriorates by approximately 0.5 dB (see, e.g., Figure 4-18 and [129]).

The new norm definition does not influence the complexity of the preamble processing. The complexity of the payload phase, however, is significantly reduced since the calculation of the approximated l_1 norm results in a complexity of $(2N_r - 1)I$ real additions and no multiplications. As a result, the overall complexity per TX vector for this reduced complexity case equals $4N_r I - 1$ R_ADDs. Clearly, this is a significant reduction in complexity compared to the previous numbers, but the complexity still increases exponentially with N_t. More ideas for complexity reduction are published, e.g., in [11].

References

[1]   A. Adjoudani, et al, "Prototype experience for MIMO BLAST over third-generation wireless system", *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, April 2003, pp. 440-451.

[2]   D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri, "Space-time coded OFDM for high data-rate wireless communication over wideband channels", in *Proc. of the 48th IEEE Vehicular Technology Conference (VTC) 1998*, vol. 3, 1998, pp. 2232-2236.

[3]   Airgo Networks, "Airgo launches the next generation in WLAN", Nov. 2003, http://www.airgonetworks.com.

[4]   D. Aktas, H. El Gamal and M. P. Fitz, "Towards optimal space-time coding", in *Proc. of the 36th Asilomar Conference on Signals, Systems and Computers 2002*, vol. 2, Nov. 2002, pp. 1137-1141.

[5]   S. M. Alamouti, "A simple transmit diversity technique for wireless communications", *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, Oct. 1998, pp. 1451-1458.

[6]   J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends", *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 11, Nov. 2000, pp. 2172-2178.

[7]   T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Second edition, New York, John Wiley & Sons, 1984.

[8]   S. L. Ariyavisitakul, "Turbo space-time processing to improve wireless channel capacity", *IEEE Transactions on Communications*, vol. 48, no. 8, Aug. 2000, pp. 1347-1358.

[9]   A. G. Armada and M. Calvo, "Phase noise and subcarrier spacing effects on the performance of an OFDM communication system", *IEEE Communication Letters*, vol. 2, no. 1, Jan. 1998, pp. 11-13.

[10]   Ascom, "WLAN miniature broadband antenna MBA-5", November 2003, http://www.ascom.com.

[11]   G. Awater, A. van Zelst, and R. van Nee, "Reduced complexity space division multiplexing receivers", in *Proc. of the IEEE 51st Vehicular Technology Conference (VTC) 2000 Spring*, vol. 1, May 2000, pp. 11-15.

[12]   B4, "BraBant BreedBand", November 2003, http://www.brabantbreedband.nl.

[13]   L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate", *IEEE Transactions on Information Theory*, March 1974, pp. 284-287.

[14]   I. Barhumi, G. Leus, and M. Moonen, "Optimal training sequences for channel estimation in MIMO OFDM systems in mobile wireless channels", in *Proc. of the International Zurich Seminar on Broadband Communications, 2002, Access, Transmission, Networking*, pp. 44-1 - 44-6.

[15]   M. Bengtsson, P. Laspougeas, S. Mayrargue, and B. Ottersten, "Single and Dual Multi-Sector Channel Characterisation − Analysis and Models", IST-1999-10322 SATURN D523, part 1.

[16]   C. Berrou, A. Glavieux and, P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1", in *Proc. of the IEEE International Conference on Communications (ICC) 1993*, vol. 2, May 1993, pp. 1064-1070.

[17]   C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes", *IEEE Transactions on Communications*, vol. 44, no. 10, Oct. 1996, pp. 1261-1271.

[18]   E. Biglieri, G. Taricco and A. Tulino, "Performance of space-time codes for a large number of antennas", *IEEE Transactions on Information Theory*, vol. 48, no. 7, July 2002, pp. 1794-1803.

[19]   H. Bölcskei, M. Borgmann, and A. J. Paulraj, "Space-frequency coded MIMO-OFDM with variable multiplexing-diversity tradeoff", in *Proc. of the IEEE International Conference on Communications (ICC) 2003*, vol. 4, May 2003, pp. 2837-2841.

[20]   H. Bölcskei and A. J. Paulraj, "Space-frequency coded broadband OFDM systems", in *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC) 2000*, vol. 1, 2000, pp. 1-6.

[21]   N. Chayat, "Tentative Criteria for Comparison of Modulation Methods", Tech. Rep. IEEE P802.11-97/96.

[22]   D. Chizhik, F. Rashid-Farrokhi, J. Ling, and A. Lozano, "Effect of antenna separation on the capacity of BLAST in correlated channels", *IEEE Communications Letters*, vol. 4, no. 11, Nov. 2000, pp. 337-339.

[23]   L. J. Cimini Jr., "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing", *IEEE Transactions on Communications*, vol. com-33, no. 7, July 1985, pp. 665-675.

[24]   CommsDesign, "Agere demos 162-Mb/s wireless LAN", Nov. 2003, http://www.commsdesign.com/story/OEG20021118S0040.

[25]   Da-Shan 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, March 2000, pp. 502-513.

[26]   DAB, "The world DAB forum", Aug. 2003, http://www.worlddab.org.

[27]   G. Dolmans, B Vandewiele, M. Collados, A. van Zelst, I. Modonesi, M. Jevrosimovic and P. Smulders, "Smart antenna concepts; prototype design choices and field-tests (DCP 3.1)", *BTS Project B4 BroadbandRadio@Hand, BTS01063,* Deliverable 3.1, May 2002.

[28]   Dongzhe Cui and A. M. Haimovich, "Error performance analysis of turbo space-time coded modulation over fading channels", in *Proc. of the IEEE International Conference on Communications (ICC) 2001*, vol. 9, 2001, pp. 2809-2813.

[29]   DVB, "The standard for the digital world", Aug. 2003, http://www.dvb.org.

[30]   ETSI, "Broadband radio access networks (BRAN), HIPERLAN Type 2; Physical (PHY) layer", Tech. Rep. ETSI TS 101 475, v1.1.1, April 2000.

[31]   ETSI, "ETSI – Telecom standards", Nov. 2003, http://www.etsi.org.

[32]   B. Farhang-Boroujeny and C. Schlegel, "Efficient multicarrier realization of full-rate space-time orthogonal block coded systems", in *Proc. of the International Conference on Communications (ICC) 2003*, vol. 4, May 2003, pp. 2267-2271.

[33]   F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal BLAST processing with multiple-access interference", in *Proc. of the 52nd IEEE VTS Vehicular Technology Conference (VTC) 2000 Fall*, vol. 1, 2000, pp. 87-91.

[34]   FCC, "Amendment of the commission's rules to provide for operation of unlicensed NII devices in the 5-GHz frequency range", memorandum option and order, ET docket no. 96-102, June 24, 1998.

[35]   G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas", *Bell Laboratories Technical Journal*, vol. 1, no. 2, autumn 1996, pp. 41-59.

[36]   G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas", *AT&T-Bell Labs Internal Tech. Memo*, Sept. 1995. Published in *Wireless Personal Communications*, vol. 6, no. 3, March 1998, pp. 311-335.

[37]   D. C. Garrett, L. M. Davis, and G. K. Woodward, "19.2 Mbit/s 4×4 BLAST/ MIMO detector with soft ML outputs", *IEEE Electronics Letters*, vol. 39, no. 2, Jan. 2003, pp. 233-235.

[38]   G. Ginis and J. M. Cioffi, "On the relation between V-BLAST and the GDFE", *IEEE Communications Letters*, vol. 5, no. 9, Sept. 2001, pp. 364-366.

[39]   J.-C. Guey, M. P. Fitz, M. R. Bell, and W.-Y. Kuo, "Signal design for transmit diversity wireless communication systems over Rayleigh fading channels", in *Proc. of the IEEE Vehicular Technology Conference (VTC) 1996*, vol. 1, May 1996, pp. 136-140.

[40]   J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications", in *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM) 1989*, Nov. 1989, pp. 47.1.1-47.1.7.

[41]   J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and conventional codes", *IEEE Transactions on Information Theory*, vol. 42, March 1996, no. 2, pp. 429-445.

[42]   Haiquan Wang and Xiang-Gen Xia, "Upper bounds of rates of space-time block codes from complex orthogonal designs", in *Proc. of the IEEE International Symposium on Information Theory, 2002*, July 2002, p. 303.

[43]   G. Halls, "HIPERLAN radio channel models and simulation results", Tech. Rep. RES10TTG 93/58.

[44]   J. S. Hammerschmidt, "Adaptive Space and Space-Time Signal Processing for High-Rate Mobile Data Receivers", Ph.D. dissertation, VDI Verlag GmbH, Düsseldorf, 2001.

[45]   A. R. Hammons Jr. and H. El Gamal, "On the theory of space-time codes for PSK modulation", *IEEE Transactions on Information Theory*, vol. 46, no. 2, March 2000, pp. 524-542.

[46]   J. Hansen and M. Nold, "Analytic calculation of the power delay profile for single room wireless LAN environments", *IEEE Global Telecommunications Conference (GLOBECOM) 2000*, vol. 1, 2000, pp. 98-102.

[47] D. A. Harville, *Matrix Algebra from a Statistician's Perspective*, New-York, Springer-Verlag, Sept. 1997.

[48] H. Hashemi, "The indoor propagation channel", *Proceedings of the IEEE*, vol. 81, no. 7, July 1993, pp. 943-968.

[49] B. Hassibi, "An efficient square-root algorithm for BLAST", Aug. 2003, http://mars.bell-labs.com/cm/ms/what/mars/index.html.

[50] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time", *IEEE Transactions on Information Theory*, vol. 48, no. 7, July 2002, pp. 1804-1824.

[51] R. W. Heath and A. J. Paulraj, "Switching between multiplexing and diversity based on constellation distance", *38th Annual Allerton Conference on Communications, Control and Computing*, Monticello IL, Oct. 2000.

[52] J. Heiskala and J. Terry, *OFDM wireless LANs: a theoretical and practical guide*, Sams Publishing, Indianapolis, USA, 2002.

[53] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel", *IEEE Transactions on Communications*, vol. 51, no. 3, March 2003, pp. 389-399.

[54] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, Cambridge University Press, 1985.

[55] Hsuan-Jung Su and E. Geraniotis, "Space-time turbo codes with full antenna diversity", *IEEE Transactions on Communications*, vol. 49, no. 1, Jan. 2001, pp. 47-57.

[56] IEEE 802.11, "IEEE P802.11, the working group for wireless LANs", Aug. 2003, http://grouper.ieee.org/groups/802/11/.

[57] IEEE, "Supplement to standard for telecommunications and information exchange between systems – LAN/MAN specific requirements – part 11: wireless MAC and PHY specifications: high speed physical layer in the 5 GHz band", ISO/IEC 8802-11:1999, 1999/Amd 1:2000(E), 2000.

[58] In-Stat/MDR, "In-Stat/MDR - Report Information", Nov. 2003, http://www.instat.com/abstract.asp?id=160&SKU=IN030902WL

[59] ITU, "World telecommunication development report 2002", Oct. 2003, http://www.itu.int/itu-d/ict/publications/wtdr_02/

[60] H. Jafarkhani, "A quasi-orthogonal space-time block code", *IEEE Transactions on Communications*, vol. 49, no. 1, Jan. 2001.

[61]   L. M. A. Jalloul, K. Rohani, K. Kuchi, and J. Chen, "Performance analysis of CDMA transmit diversity methods", in *Proc. of the IEEE Vehicular Technology Conference (VTC) 1999,* vol. 3, pp. 1326-1330, 1999.

[62]   G. J. M. Janssen, P. A. Stigter, and R. Prasad, "Wideband indoor channel measurements and BER analysis of frequency selective multipath channels at 2.4, 4.75 and 11.5 GHz", *IEEE Transactions on Communications*, vol. 44, no. 10, Oct. 1996, pp. 1272-1288.

[63]   Jibing Wang, E. Biglieri, and Kung Yao, "Asymptotic performance of space-frequency codes over broadband channels", *IEEE Communications Letters*, vol. 6, no. 12, Dec. 2002, pp. 523-525.

[64]   Kai 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 *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM), 2001*, vol. 1, pp. 156-160.

[65]   Kai Yu and B. Ottersten, "Models for MIMO propagation channels: a review", *Wireless Communications and Mobile Computing 2002*, John Wiley & Sons, no. 2, pp. 653-666.

[66]   T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, New Jersey, Prentice-Hall, March 2000.

[67]   T. Keller and L. Hanzo, "Orthogonal frequency division multiplex synchronisation techniques for wireless local area networks", in *Proc. of the Personal, Indoor and Mobile Radio Communications (PIMRC) 1996*, Taipei, Taiwan, pp. 963-967.

[68]   J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation", *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, Aug. 2002, pp. 1211-1226.

[69]   M. A. Khalighi, K. Raoof, and G. Jourdain, "Capacity of wireless communication systems employing antenna arrays, a tutorial study", *Wireless Personal Communications*, vol. 23, 2002, pp. 321–352.

[70]   M. Kiessling, J. Speidel, N. Geng, and M. Reinhardt, "Performance analysis of MIMO maximum likelihood receivers with channel correlation, colored gaussian noise, and linear prefiltering", in *Proc. of the IEEE International Conference on Communications (ICC) 2003*, vol. 5, pp. 3026-3030.

[71]   J. Kivinen, X. Zhao, and P. Vainikainen, "Wideband indoor radio channel measurements with direction of arrival estimations in the 5 GHz band", in *Proc. of the 50nd IEEE VTS Vehicular Technology Conference (VTC) 1999 Fall*, Sept. 1999, pp. 2308-2312.

[72] Lei He and Hongya Ge, "QPSK orthogonal space-time coding scheme with full-rate and full-diversity for system with four transmit antennas", in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003*, vol. 4, April 2003, pp. IV_337-IV_340.

[73] Lizhong Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels", *IEEE Transactions on Information Theory*, vol. 49, no. 5, May 2003, pp. 1073-1096.

[74] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix", *IEEE Communications Letters*, vol. 5, no. 9, Sept. 2001, pp. 369-371.

[75] A. Lozano, "Surveying in Literature on Multi-Transmit Multi-Receive Wireless Communications", Internal Report, Lucent Technologies, Feb. 20, 2002.

[76] MMAC, "Multimedia Mobile Access Communication Systems", Aug. 2003, http://www.arib.or.jp/mmac/e/.

[77] J. Medbo, H. Hallenberg, and J.-E. Berg, "Propagation characteristics at 5 GHz in typical radio-LAN scenarios", in *Proc. of the 49nd IEEE VTS Vehicular Technology Conference (VTC) 1999 Spring*, May 1999, pp. 185-189.

[78] A. N. Mody and G. L. Stüber, "Synchronization for MIMO OFDM systems", in *Proc. IEEE Global Communications Conference 2001*, vol. 1, Nov. 2001, pp. 509-513.

[79] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction", *IEEE Transactions on Communications*, vol. 42, no. 10, Oct. 1994, pp. 2908-2914.

[80] R. R. Müller, "A random matrix model for the antenna array channel with decaying power delay profile", in *Proc. of the IEEE Information Theory Workshop 2001*, 2001, pp. 144-146.

[81] S. H. Müller-Weinfurtner, "On the optimality of metrics for coarse frame synchronization in OFDM: a comparison", in *Proc. of the Personal, Indoor and Mobile Radio Communications (PIMRC) 1998*, Sept. 1998, pp. 533-537.

[82] A. F. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "A space-time coding modem for high-data-rate wireless communications", *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, Oct. 1998, pp. 1459-1478.

[83] C. B. Papadias and G. J. Foschini, "A space-time coding approach for systems employing four transmit antennas", in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2001*, vol. 4, May 2001, pp. 2481-2484.

[84]   A. J. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission / directional reception", U.S. Patent No. 5,345,599, 1994.

[85]   A. J. Paulraj, D. Gore, and R. U. Nabar, "Performance limits in fading MIMO channels", in *Proc. of the 5th International Symposium on Wireless Personal Multimedia Communications (WPMC) 2002*, vol. 1, 2002, pp. 7-11.

[86]   K. I. Pedersen, J. B. Andersen, J. P. Kermoal, and P. Mogensen, "A stochastic multiple-input-multiple-output radio channel model for evaluation of space-time coding algorithms", in *Proc. of the 52nd IEEE VTS Vehicular Technology Conference (VTC) 2000 Fall*, vol. 2, pp. 893-897.

[87]   K. I. Pedersen and P. E. Mogensen, "A simple downlink antenna array algorithm based on a hybrid scheme of transmit diversity and conventional beamforming", in *Proc. of the 53rd IEEE VTS Vehicular Technology Conference (VTC) 2001 Spring*, vol. 1, pp. 58-62.

[88]   D. Porrat, P. Kyritsi, and D. C. Cox, "MIMO capacity in hallways and adjacent rooms", in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM) 2002*, vol. 2, Nov. 2002, pp. 1930-1934.

[89]   W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C, The Art of Scientific Computing*, Second Edition, Cambridge, Cambridge University Press, 1992.

[90]   J. G. Proakis, *Digital Communications*, Third Edition, New York, McGraw-Hill, 1995, McGraw-Hill Series in Electrical and Computer Engineering.

[91]   Radio Frequency Identification (RFID), "Radio Frequency Identification (RFID) homepage", Jan. 2004, http://www.aimglobal.org/technologies/rfid/.

[92]   T. S. Rappaport, *Wireless Communications, Principles and Practice*, New Jersey, Prentice-Hall, 1996.

[93]   G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication", *IEEE Trans. on Communications*, vol. 46, no. 3, March 1998, pp. 357-366.

[94]   G. G. Raleigh and T. Boros, "Joint space-time parameter estimation for wireless communication channels", *IEEE Transactions on Signal Processing*, vol. 46, no. 5, May 1998, pp. 1333-1343.

[95]   L.G. Roberts, "Internet traffic measurement 2000 and 2001", Jan. 16, 2001, http://www.cibcwm.com/conferences/hour/foundingfather/roberts.pdf

[96]   P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms in the log domain", in *Proc. of the International Conference on Communications (ICC) 1995*, June 1995, pp. 1009-1013.

[97] A. A. M. Saleh and R. A. Valenzuela, "A statistical model for indoor multipath propagation", *IEEE Journal on Selected Areas in Communications*, vol. SAC-5, no. 2, Feb. 1987, pp. 128-137.

[98] J. Salz, "Digital transmission over cross-coupled linear channels", *AT&T Technical Journal*, vol. 64, no. 6, July-Aug. 1985, pp. 1147-1159.

[99] K. Sam Shanmugam, *Digital and Analog Communication Systems*, J. Wiley & Sons, Inc., New York, 1985.

[100] H. Sampath, "Linear precoding and decoding for multiple input multiple output (MIMO) wireless channels", Ph.D. dissertation, Department of Electrical Engineering, Stanford University, April 2001.

[101] S. Sandhu and A. Paulraj, "Space-time block codes: a capacity perspective", *IEEE Communications Letters*, vol. 4, no. 12, Dec. 2000, pp. 384-386.

[102] T. C. W. Schenk, "Synchronisation of Multiple-Input Multiple-Output OFDM", M.Sc. Thesis, Eindhoven University of Technology, Faculty of Electrical Engineering, Telecommunication Technology and Electromagnetics, June 2002.

[103] T. C. W. Schenk and A. van Zelst, "Frequency synchronization for MIMO OFDM wireless LAN systems", in *Proc. of IEEE Vehicular Technology Conference (VTC) Fall 2003*, Orlando (FL), Oct. 2003, paper 05D-03.

[104] T. C. W. Schenk and P. Mattheijssen, "Analysis of the influence of phase noise in MIMO OFDM based WLAN systems", in *Proc. of the 10th IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT 2003)*, November 2003.

[105] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM", *IEEE Transactions on Communications*, vol. 45, no. 12, Dec. 1997, pp. 1613-1621.

[106] M. Sellathurai and S. Haykin, "TURBO-BLAST for high-speed wireless communications", in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC) 2000*, vol. 1, Sept. 2000, pp. 315-320.

[107] C. E. Shannon, "A mathematical theory of communication", *Bell Systems Technical Journal*, vol. 27, 1948, pp. 379-423, 623-656.

[108] E. Sjostrom, "Circulant Matrices", March 2003, http://www.mai.liu.se/~evlun/pub/lic/node9.html.

[109] M. Speth, "OFDM receivers for broadband-transmission", April 1999, http://www.ert.rwth-aachen.de/Projekte/Theo/OFDM/www_ofdm.html.

[110] M. Speth, A. Senst, and H. Meyr, "Low complexity space-frequency MLSE for multi-user OFDM", in *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM) 1999*, vol. 5, Dec. 1999, pp. 2395-2399.

[111] A. Stefanov and T. M. Duman, "Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: system model, decoding approaches, and practical considerations", *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, May 2001, pp. 958-968.

[112] M. Stege and G. Fettweis, "Multistratum-permutation codes for MIMO communication", *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, June 2003, pp. 774-782.

[113] G. Strang, *Linear Algebra and its Applications*, Third Edition, San Diego, Harcourt Brace Jovanovich, Publishers, 1988.

[114] V. Tarokh, A. Naguib, and A. R. Calderbank, "Combined array processing and space-time coding", *IEEE Transactions on Information Theory*, vol. 45, no. 4, May 1999, pp. 1121-1128.

[115] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs" *IEEE Transactions on Information Theory*, vol. 45, July 1999, pp. 1456-1467.

[116] 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. 3, March 1998, pp. 744-756.

[117] I.E. Telatar, "Capacity of multi-antenna gaussian channels", Lucent Technologies, Internal Report, BL011217-950615-07TM, June 1995, Published in *European Transactions on Telecommunications*, vol. 10, no. 6, Nov./Dec. 1999, pp. 585-595.

[118] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes", *IEEE Transactions on Communications*, vol. 49, no. 10, Oct. 2001, pp. 1727-1737.

[119] S. ten Brink, "Convergence of iterative decoding", *Electronics Letters*, vol. 35, no. 13, June 1999, pp. 1117-1119.

[120] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation", in *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM) 1998*, vol. 1, Nov. 1998, pp. 579-584.

[121] S. Thoen, "Transmit optimization for OFDM/SDMA-based wireless local area networks", Ph.D. dissertation, Katholieke Universiteit Leuven, Department of Electrical Engineering, ESAT, May 2002.

[122] O. Tirkkonen, A. Boariu, and A. Hottinen, "Minimal non-orthogonality rate 1 space-time block code for 3+ tx", in *Proc. of the International Symposium on Spread Spectrum Techniques and Applications,* Sept. 2000, pp. 429-432.

[123] A. M. Tonello, "Space-time bit-interleaved coded modulation with an iterative decoding strategy", in *Proc. of the 52nd IEEE Vehicular Technology Conference (VTC) 2000 Fall*, vol. 1, Sept. 2000, pp. 473-478.

[124] D. Tujkovic, M. Juntt, and M. Latva-Aho, "Space-frequency-time turbo coded modulation", *IEEE Communications Letters*, vol. 5, no. 12, Dec. 2001, pp. 480-482.

[125] R. van Nee, A. van Zelst, and G. A. Awater, "Maximum likelihood decoding in a space division multiplex system", in *Proc. of the IEEE 51st Vehicular Technology Conference (VTC) 2000 Spring*, vol. 1, May 2000, pp. 6-10.

[126] R. van Nee and R. Prasad, *OFDM for Mobile Multimedia Communications*, Boston, Artech House, 2000.

[127] R. van Nee, G. Awater, Masahiro Morikura, Hitoshi Takanashi, M. Webster, and K. Helford, "New high rate wireless LAN standards", *IEEE Communications Magazine*, Dec. 1999.

[128] R. van Poppel, "Analysis and measurements of wideband MIMO radio channels in indoor environments at 5.2 GHz – Towards the validation of a single coefficient spatial correlation model for MIMO radio channels", M.Sc. Thesis, Eindhoven University of Technology, Faculty of Electrical Engineering, Telecommunication Technology and Electromagnetics, May 2003.

[129] A. van Zelst, "Extending the capacity of next generation wireless LANs using space division multiplexing combined with OFDM", M.Sc. Thesis, Eindhoven University of Technology, Faculty of Electrical Engineering, Telecommunication Technology and Electromagnetics, Oct. 1999.

[130] A. van Zelst, "Per-antenna-coded schemes for MIMO OFDM", in *Proc. of the IEEE International Conference on Communications (ICC) 2003*, vol. 4, May 2003, pp. 2832-2836.

[131] A. van Zelst, "Physical interpretation of MIMO transmissions", in *Proc. of the 10th Symposium on Communications and Vehicular Technology in the Benelux (SCVT 2003)*, November 2003.

[132] A. van Zelst, "Space division multiplexing algorithms", in *Proc. of the 10th Mediterranean Electrotechnical Conference (MELECON) 2000*, vol. 3, May 2000, pp. 1218-1221.

[133] A. van Zelst, "A compact representation of spatial correlation in MIMO radio channels", submitted to *International Conference on Communications (ICC) 2004*, Paris, June 2004.

[134] A. van Zelst and J. S. Hammerschmidt, "A single coefficient spatial correlation model for multiple-input multiple-output (MIMO) radio channels", in *Proc. of the 27th General Assembly of the International Union of Radio Science (URSI)*, Maastricht, The Netherlands, Aug. 2002.

[135] A. van Zelst and T. C. W. Schenk, "Implementation of a MIMO OFDM based wireless LAN system", *IEEE Transactions on Signal Processing*, Vol. 52, No. 2, Feb. 2004.

[136] A. van Zelst and T. C. W. Schenk, "Throughput enhancement for wireless communications systems using the multiple antenna technique MIMO", in *Proc. of the IEEE International Conference on Consumer Electronics (ICCE) 2003*, June 2003, pp. 300-301.

[137] A. van Zelst, R. van Nee, and G. A. Awater, "Space division multiplexing (SDM) for OFDM systems", in *Proc. of the IEEE 51st Vehicular Technology Conference (VTC) 2000 Spring*, vol. 2, May 2000, pp. 1070-1074.

[138] A. van Zelst, R. van Nee, and G. A. Awater, "Turbo-BLAST and its performance", in *Proc. of IEEE Vehicular Technology Conference (VTC) 2001 Spring*, vol. 2, May 2001, pp. 1282-1286.

[139] U. Wachsmann, J. Thielecke, and H. Schotten, "Exploiting the data-rate potential of MIMO channels: multi-stratum space-time coding", in *Proc. of the IEEE VTS 53th Vehicular Technology Conference (VTC) 2001 Spring*, vol. 1, May 2001, pp. 199-203.

[140] Weifeng Su and Xiang-Gen Xia, "Two generalized complex orthogonal space-time block codes of rates 7/11 and 3/5 for 5 and 6 transmit antennas", *IEEE Transactions on Information Theory*, vol. 49, no. 1, Jan. 2003, pp. 313-316.

[141] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform", *IEEE Transactions on Communication Technology*, vol. com-19, no. 5, Oct. 1971, pp. 628-634.

[142] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems", *IEEE Transactions on Communications*, vol. 42, no. 2, Feb./Mar./Apr. 1994, pp. 1740-1751.

[143] K. Witrisal, "OFDM air-interface design for multimedia communications", Ph.D. dissertation, Department of ITS, University of Delft, April 2002.

[144] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel", in *Proc. of the URSI International Symposium on Signals, Systems, and Electronics (ISSSE) 1998*, Pisa, 29 Sept. - 2 Oct. 1998, pp. 295-300.

[145] Won-Joon Choi, R. Negi, and J. M. Cioffi, "Combined ML and DFE decoding for the V-BLAST system", in *Proc. of the IEEE International Conference on Communications (ICC) 2000*, vol. 3, pp. 1243-1248.

[146] Xiaodong Li, H. Huang, G. J. Foschini, and R. A. Valenzuela, "Effects of iterative detection and decoding on the performance of BLAST", in *Proc. of the IEEE Global Telecommunications Conference (GLOBECOM) 2000*, vol. 2, pp. 1061-1066.

[147] Zhiqiang Liu, Yan Xin and G.B. Giannakis, "Space-time-frequency coded OFDM over frequency-selective fading channels", *IEEE Transactions on Signal Processing*, vol. 50, no. 10, Oct. 2002, pp. 2465-2476.

[148] Zhuo Chen, Branka S. Vucetic, Jinhong Yuan and Ka Leong Lo, "Space–time trellis codes for 4-PSK with three and four transmit antennas in quasi-static flat fading channels", *IEEE Transactions on Communications*, vol. 50, no. 2, Feb. 2002, pp. 67-69.

Samenvatting

Breedbandige applicaties – zoals hoge snelheid computer netwerken, multimedia diensten, of data netwerken in ziekenhuizen voor telediagnose met behulp van digitale informatie – en de vraag naar flexibiliteit zijn de aanjagers van de behoefte naar breedbandige draadloze communicatiesystemen. Omdat het beschikbare frequentiespectrum schaars is, zullen toekomstige systemen aanmerkelijk efficiënter moeten omgaan met het spectrum ter verbetering van de link datasnelheid en netwerk capaciteit. Een veelbelovende manier is het gebruik van meerdere antennes bij zowel de zender als de ontvanger (een zogenaamd Multiple-Input Multiple-Output (MIMO) systeem). Met een dergelijk systeem kan de datasnelheid worden verhoogd door verschillende datastromen simultaan te zenden op verschillende zendantennes maar op dezelfde draaggolf. Hoewel deze parallelle datastromen mixen in de lucht kunnen ze toch herwonnen worden bij de ontvanger door gebruik te maken van meerdere ontvangantennes en corresponderende signaalbewerkingsalgoritmen, onder de voorwaarde dat het MIMO kanaal goed geconditioneerd is. Over het algemeen is dit het geval in omgevingen waarin het radiosignaal in hoge mate wordt verstrooid, zoals ruimtes binnenshuis. Deze techniek wordt Space Division Multiplexing (SDM) genoemd.

De combinatie van de datasnelheidsverbetering van SDM met de robuustheid van Orthogonal Frequency Division Multiplexing (OFDM) tegen smalbandige interferentie en frequentieselectieve fading, veroorzaakt door ernstige multipadverstrooiing, wordt gezien als een veelbelovende basis voor toekomstige radiocommunicatiesystemen met hoge datasnelheid (voor gebruik binnenshuis). SDM OFDM is de focus van dit proefschrift en de belangrijkste bijdragen, in de logische volgorde van fundamenteel begrip, theoretische analyse tot praktische metingen, worden hieronder opgesomd.

Ten eerste, door middel van een fysische interpretatie is een fundamentele en intuïtieve uitleg gegeven van de spectrale efficiëntie en stabiliteit van een draadloos MIMO systeem in omgevingen waarin het radiosignaal in hoge mate wordt verstrooid, zoals binnenshuis.

Ten tweede is een generiek breedbandig MIMO kanaalmodel voor binnenshuis voorgesteld dat mede een Line-of-Sight (LOS) component en ruimtelijke-correlatie modelleert. Dit

model comprimeert de typisch grote hoeveelheid kanaalparameters tot een klein aantal bewust gekozen parameters.

Ten derde zijn verschillende smalbandige SDM algoritmen beschreven. Deze beschrijving bevat tevens schema's die betrouwbaarheidsindicaties meegeven aan het uitgangssignaal voor situaties waarin externe codering en decodering wordt gebruikt. De foutenkans prestatie en de complexiteit van de algoritmen is geëvalueerd voor verschillende antenneconfiguraties, constellatiegroottes, en kanaaleigenschappen (inclusief ruimtelijke correlatie en een LOS component), met en zonder codering. Het is aangetoond dat Maximum Likelihood Detection (MLD) beter presteert dan de andere technieken. De complexiteit van MLD is echter het grootst en groeit exponentieel met het aantal zendantennes. Minder complexe alternatieven zijn gevonden met slechts een beperkt prestatieverlies.

Ten vierde observeerden we door middel van een generieke kijk op (gecodeerde) MIMO technieken dat de beste prestatie behaald kan worden door een evaluatie te doen over alle punten van het niet-redundante rooster dat alle mogelijke ruimte-tijd codewoorden bevat. De complexiteit van een zodanige zoektocht groeit echter exponentieel met het aantal rooster punten. Het in dit proefschrift geïntroduceerde turbo SDM schema behaalt een significante complexiteitsreductie terwijl de prestatie erg dicht bij die van volledige evaluatie ligt. De complexiteit is gereduceerd door the ruimte- en tijdbewerking te scheiden. De goede prestatie wordt bereikt door iteratie tussen deze bewerkingsstappen uit te voeren. Deze aanpak is afgeleid van het zogenaamde decoderingsprincipe van turbo-coding.

Ten vijfde, omdat OFDM al de basis vormt van de bestaande draadloze Local Area Network (LAN) standaarden IEEE 802.11a en g, wordt de combinatie van SDM en OFDM gezien als een aantrekkelijke oplossing voor toekomstige draadloze LANs met hoge snelheden. OFDM splitst, over het algemeen gesproken, een breedbandig frequentieselectief kanaal in een aantal smalbandige kanalen (elk met frequentie onafhankelijke fading). Dit betekent dat per subkanaal de gepresenteerde smalbandige SDM algoritmen toegepast kunnen worden. De combinatie SDM OFDM is geëvalueerd in theorie, met simulaties ter evaluatie van de prestatie en met metingen. De theoretische evaluatie is uitgevoerd met behulp van een algemene ruimte-frequentie foutenkansanalyse. Het is aangetoond dat de maximale diversiteitwinst gelijk is aan het product van het aantal zend- en ontvangantennes en de effectieve lengte van de impuls responsie van het kanaal. SDM OFDM schema's met codering zijn voorgesteld, die een groot deel van deze diversiteitwinst behalen.

Ten zesde, in de praktijk moet het systeem kunnen omgaan met systeemverstoringen zoals frequentie offset, offset in de tijdsbepaling, fase ruis, DC offset, etc. Om deze verstoringen voor SDM OFDM in de draadloze LAN context te kunnen aanpakken hebben we training en synchronisatiealgoritmen voorgesteld die een uitbreiding zijn op die van IEEE 802.11a systemen. Om deze algoritmen en het algemene SDM OFDM concept te valideren is in Agere Systems een testsysteem gebouwd met drie zend- en drie ontvangantennes en gebaseerd op IEEE 802.11a parameters. Meetresultaten met dit testsysteem in een kantooromgeving laten succesvolle transmissies met snelheden tot 162 Mb/s zien. Dit is drie keer de data snelheid van een "gewoon" IEEE 802.11a OFDM systeem.

Tenslotte concludeerden we dat SDM OFDM, hoewel er ruimte is voor verbetering, een aantrekkelijke en praktische oplossing is om de datasnelheid en/of robuustheid van draadloze systemen, bijvoorbeeld gebaseerd op IEEE 802.11a, aanzienlijk te verbeteren.

Acknowledgements

Obtaining a Ph.D. degree is a long, hard, but very interesting "journey". Inspiring company and a stimulating environment together with a good vision and corresponding tasks and planning are crucial to succeed. Without the intention to forget someone, I would like to express my gratitude to those who accompanied me during the journey and directly or indirectly provided support and advice on the definition of the goals and the execution of the corresponding tasks.

It started more or less with prof. Gert Brussaard and dr. Peter Smulders getting me into the exciting world of wireless (digital) communication systems and helping me to find a M.Sc. project in a company. Dr. Richard van Nee and dr. Geert Awater made this possible in the very stimulating environment of Bell Labs, Lucent Technologies, Nieuwegein. After finishing my M.Sc. thesis, the enthusiasm of above persons and the full financial support of Lucent Technologies (and later Agere Systems), which is gratefully acknowledged, led to the definition of my Ph.D. project. I am very thankful for the rather unique opportunity I got by starting my Ph.D. work in combination with the Radiocommunication group of the Eindhoven University of Technology (TU/e) and Bell Labs.

Moreover, I would like to thank prof. Gert Brussaard and dr. Peter Smulders for their supervision, advice, feedback, and valuable suggestions during my Ph.D. project, and for being my first promotor and copromotor, respectively. Special thanks for giving me the freedom to do the project in a company. Sincere credits go to dr. Richard van Nee and dr. Jochen Hammerschmidt who coached me from the company side, respectively the first year and the last two and a half years; Richard for his enthusiasm, creativity, and great suggestions, Jochen for his advice, thorough review of my work, for being particularly hard to convince, and the role he played in the 6 weeks internship in Agere in Murray Hill. Having worked in the environment where, among others, the transistor was invented and Claude Shannon did his fundamental studies which form a basis of this work, was very inspiring and unforgettable. In this context I would also like to thank Aon Mujtaba.

Curriculum Vitae

Allert van Zelst was born in Waalwijk, The Netherlands, on March 19, 1976. After attending the secondary school at the Willem van Oranje College in Waalwijk, he started Electrical Engineering (EE) at the Eindhoven University of Technology (TU/e) in Eindhoven, The Netherlands, in September 1994. About five years later, he received the M.Sc. degree in EE at TU/e after finishing a graduation project on data-rate enhancements of Wireless Local Area Networks by exploiting the spatial dimension through multiple-antenna techniques. This project was carried out at Bell Labs, Lucent Technologies, in Nieuwegein, The Netherlands. Based on this work, he won the second prize of the IEEE Region 8 Student Paper Contest 2000 for a paper entitled "Space Division Multiplexing Algorithms". In November 1999, A. van Zelst continued with research on this topic to pursue the Ph.D. degree at the radio communications chair of TU/e, which resulted in this dissertation. The Ph.D. project was sponsored by the Dutch cooperative research project B4 Broadband Radio@hand and Agere Systems (formerly the Microelectronics Group of Lucent Technologies) in Nieuwegein, The Netherlands. The main part of the research was performed within the Wireless Systems Research department of Agere Systems in Nieuwegein. In the framework of the Ph.D. project, A. van Zelst also visited a part of this department in Murray Hill, New Jersey, USA, for an internship of 6 weeks. As of January 2004, he joined Airgo Networks, Breukelen, The Netherlands.