GIGABIT WIRELESS TRANSMISSION IN DISPERSIVE ENVIRONMENTS

Channel Characterization and Signal Processing Algorithms

SEYRAN KHADEMI

GIGABIT WIRELESS TRANSMISSION IN DISPERSIVE ENVIRONMENTS

Channel Characterization and Signal Processing Algorithms

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K.Ch.A.M. Luyben,
voorzitter van het College van Promoties,
in het openbaar te verdedigen op dinsdag 22 november 2016 om 10:00 uur

door

Seyran KHADEMI
Master of Science in Electrical Engineering, Chalmers University of Technology,
Gothenburg, Sweden

geboren te Kermanshah, Iran

This dissertation has been approved by the
promotor Prof. dr. ir. A.J. van der Veen and copromotor Dr. ir. G.J.M. Janssen.

Composition of the doctoral committee:

Rector Magnificus	Chairman
Prof. dr. ir. A.J. van der Veen	Delft University of Technology, Promotor
Dr. ir. G.J.M. Janssen	Delft University of Technology, Copromotor

Independent members:

Prof. dr. ir. G.J.T. Leus	Delft University of Technology
Prof. dr. A.G. Yarovoy	Delft University of Technology
Prof. dr. ir. F.M.J. Willems	Eindhoven University of Technology
Dr. ir. R. van Nee	Qualcomm, Breukelen
Prof. dr. U. Mitra	University of Southern California, USA

The research in this thesis was supported by STW under project 10551 “FASTCOM”.

ISBN 978-94-6186-744-5

Cover designed by SEYRAN KHADEMI.

To all brave women in my country
who are fighting for their freedom and rights.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Wireless communications has become an essential part of modern life and industry, covering a great variety of applications ranging from satellite communications and cellular networks to the networks of wireless sensors. Consequently, there are various communication links with different ranges, data rates and latencies which are designed for diverse applications and specifications. A peaceful coexistence of all these wireless devices which share the same medium is not possible without careful accessibility regulations. The standardization of wireless technologies provides a map to the existing products and illustrates the borders and boundaries of the telecommunications industry. A summary of these standards indicates the edges and highlights the research areas of the wireless technology.

1.1 Trends in Wireless Communications

Wireless standards are often categorized according to their range and data rates for different applications. A map for a number of wireless standards in terms of their operational range and data rates is shown in Fig. 1.1 [1]. Wireless personal area network (WPAN) technologies target low-cost and low-power applications within a short range up to tens of meters, while a wireless local area network (WLAN) covers greater distances up to hundreds of meters, but requires more expensive hardware and has a higher power consumption [2].

Indeed, short-range wireless communication (up to 10 m) is one of the evolving areas among current wireless applications to provide connections between various wireless devices at close distances. Home appliances, industrial and even medical

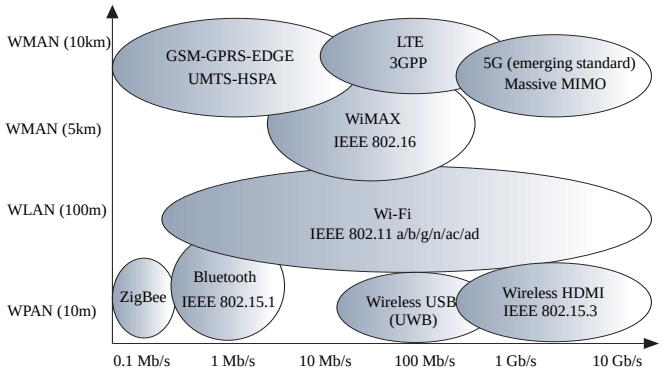


Figure 1.1: Conventional data rates and operational ranges for wireless standards.

devices are shifting towards the use of wireless connections for the sake of easing the mobility, installation and maintenance. Some standards and technologies have specifically been developed to cover such short-distance applications with limited throughput such as Bluetooth and ZigBee. Accordingly, wireless USB which operates in the frequency range of 3.1 to 10.6 GHz, is built upon ultra-wideband (UWB) technology and it is capable of sending up to 0.5 Gbps, to support compressed video streaming. Further improvements will not be seen in the near future for wireless USB, due to the restrictions on the transmit power levels imposed by regulatory bodies.

To cover the real-time streaming of data such as video and music, the IEEE 802.15.3 (WPAN) standard was established for a high data rate and high quality of service. Particularly, the 802.15.3c sub-group was launched to design a WPAN standard for a multi-gigabyte transmission on a millimeter-wave carrier which operates in the unlicensed band of 57-64 GHz defined by the Federal Communications Commission (FCC). The choice of millimeter-wave enables the simple coexistence with other microwave WPAN standards. This is the latest IEEE standard on WPAN and it enables the streaming of high-quality video and other contents between servers and portable devices. This includes applications such as high speed internet access, streaming content download (video on demand, HDTV, home theater, etc.), real time streaming, and wireless data bus for cable replacement [3].

Modulating the signal at millimeter-wave frequencies facilitates the positioning of many antennas in a small area [4, 5], which enables multiple-input multiple-output (MIMO) systems with very large antenna arrays and consequently higher data rates. The availability of broadband spectrum in the frequency band around 60 GHz provides a great opportunity for ultra-high data rate short-range wireless

communications. This frequency band was quite recently proposed for outdoor cellular communications in the emerging 5G standard (visible on the right top of the wireless standards map in Fig. 1.1) [6]. The growing trend towards 60 GHz transmission motivates a great deal of study and research in this area from hardware and integrated circuit (IC) design to signal processing and algorithm development and standardization. However, the huge data throughput and the ultra-high carrier frequency give rise to serious challenges for the low-cost and reliable radio design. Challenges involve aspects of channel propagation issues, baseband modulation schemes, antennas and IC technologies [7].

This classification leads us to spot the challenges in wireless technology and to proactively shift towards modern telecommunications systems which are capable of delivering superior data rates reliably and fast. This is in line with the actual consumer urge which is the main motivation engine behind the development of new technologies and underlying research activities, and this research work is no exception in this regard.

1.2 Research Motivation

Inside mechatronic and industrial machinery, the required wiring is an increasing concern, as it comes with issues like reliability, space efficiency, and flexibility. It thus becomes interesting to replace the wires by wireless connections. On the one hand, using multiple cables inside a dense area to connect moving parts within a confined space can significantly complicate the design and maintenance of the system. A wired connection to a moving part affects the dynamics and may cause cable jams and frequent damage to such machineries. On the other hand, current wireless technology does not meet the data rates and latency offered by wired standards like gigabit Ethernet. In fact, the required specifications for many industrial applications, including gigabit rate and low latency plus high reliability, are nowhere near the existing wireless standards as discussed in Section 1.1.

Specifically, wireless sensor networks (WSN) are currently of growing interest for industrial usage and they are mainly categorized as short-range wireless technologies due to battery limitations. Generally, these sensors are distributed to observe environmental and ambient conditions including temperature, sound, vibration, pressure, motion, etc., and they are widely used within mechatronic systems where human interaction is limited if not impossible [8]. A wide range of WSNs demand high data rates and extremely reliable connections between sensor nodes and possible control units.

This thesis is based on a collaboration between Delft University of Technology and industrial partners and poses real and demanding research questions related to the design and development of a wireless link inside a lithography machine. Lithography systems play a critical role in the development and manufacturing of ICs. The lithography process requires extremely accurate mask and substrate positioning. This task is performed via several sensors and actuators on a moving platform, which are typically connected to the control units via flat-cable wires. The moving platform experience a very high acceleration. The stiffness of the cables causes undesired disturbances to the system, which leads to inaccurate positioning. Also, the trend towards increasing numbers of moving sensors makes the design of the wiring system prohibitively complex, therefore the replacement of the cables is of interest.

The lithography system of interest consists of several $(20-30)$ moving sensors and one fixed central unit within a closed metal environment inside the lithography machine. The operating distance varies between $0.5-300$ cm and the maximum velocity of the moving sensors is $10\frac{\text{m}}{\text{s}}$. The sensors collect the data and send them over the wireless channel within the enclosed environment to the central processing unit to be used in a wideband control loop. A data block is sent every 50 μs seconds, but only a small portion of time is dedicated to transmission, while most of the time is reserved for control processing. The main specifications of the system include high (peak) data rates (Gbps) and a very low latency (μs).

Accordingly, the initial problem statement and the main motivation behind this thesis can be formulated in one question: **How to design a highly reliable short-range gigabit wireless link within a confined metal environment subject to a rigid latency requirement**.

Obviously, for delivering a comprehensive working model, many interconnected design levels have to be considered which clearly takes more than one PhD thesis to be accomplished. During the progress of the thesis, the initial research question has been generalized and diversified which is discussed in the next section.

1.3 Thesis Contributions and Outline

The central research question is partitioned into sub-questions which are addressed in this thesis.

Research Question 1 *What limitations are imposed on the wireless link performance and therefore the design criteria, when the communication system is confined to a closed metal environment which is the common case for industrial machineries?*

A fundamental difference between typical indoor and outdoor wireless applications with industrial systems arises in terms of very distinguished propagation environments. Accurate and viable wireless channel models are of vital importance to design a realistic and functioning wireless system. Therefore, to move towards a reliable and fast wireless connection for industrial usage, many efforts have been made to provide suitable and inclusive channel models. For very small-scale applications such as inter chip connections [9] or board-to-board communications [10, 11], a noticeable difference, in terms of channel properties, has been reported in the literature compared to the typical indoor and UWB channels [12, 13, 14, 15, 16]. Also, Ohira et al. studied propagation characteristics inside information communication technology (ICT) equipment such as a printer, vending and automated teller machine (ATM) [17]. However, there seems to be no literature on channel models for closed metal environments, particularly for the millimeter-wave band which is chosen to be the most suitable frequency band for very high-rate wireless applications as motivated in Section 1.1. Therefore, a channel measurement campaign has been conducted to provide statistics on the radio frequency (RF) behavior in a metal enclosure which emulates the environment within a lithography machine. The measurement results have been processed carefully to establish the foundations for further system design. The results and a comprehensive channel model for a wideband 60 GHz wireless system is presented in **Chapter 2**. A frequency domain channel sounding technique is used for obtaining channel impulse responses for multiple locations of the receive antennas on a fine grid map within metal closets of several dimensions. The results indicate that the channel impulse response within a closed metal cabinet is significantly longer in time compared to the reported channels in the literature, i.e. having an extremely dispersive environment to conduct a wireless connection.

Research Question 2 Long and fading-prone channels require fading resistant modulation and equalization techniques. *What are competitive equalization options which are capable of taming the extremely dispersive wireless channel and will the available techniques admit the high data rate, great reliability and low latency requirements of industrial applications?*

A review on existing equalization techniques for dispersive channels is given in

Chapter 3. This includes time domain and frequency domain techniques as well as block processing and serial equalization algorithms. An explicit comparison for different equalization techniques with latency and complexity evaluations provides a clear measure for choosing the most proper equalization technique for particular specifications.

Wideband transmission techniques are briefly reviewed within the system model of interest, in **Chapter 4**. Interestingly, almost all wideband wireless standards use a multi-carrier technology known as Orthogonal-frequency-division-multiplexing (OFDM), where the band is divided into many narrowband channels, i.e., one of the prime candidates for transmission in highly dispersive channels. A key benefit of OFDM is that it can be efficiently implemented using the fast Fourier transform (FFT), and that the receiver structure becomes simple since each channel or sub-carrier can be treated as narrowband instead of a more complicated wideband channel [18]. A detailed OFDM system design is sketched for a wideband application based on the measured channel impulse responses from Chapter 2 and the BER performance and latencies are simulated and illustrated for different possible settings. This gives an initial hint on the achievable data rates and latency of an OFDM system in such a dispersive and hostile environment.

This is the end of the first part of this thesis, which considers a single-input single-output (SISO) communication system, and more specifically is dedicated to the channel characterization and system design for the lithography device of interest. The second part of the thesis, which covers a more general problem than the "proposed system in metal box", starts with **Chapter 5**. This chapter, is dedicated to review material on multiple-input multiple-output (MIMO) system and the associated capacity boost with respect to a SISO system. The system model is also renewed here to accommodate the MIMO setting considered in the subsequent chapters.

Research Question 3 Although OFDM has impressive fading-resistance properties, a well known drawback is the high peak-to-average-power-ratio (PAPR). If the maximum amplitude of the time domain signal is large, it may push the transmit amplifier into a non-linear region which leads to an erroneous detection and degrades the overall performance of the system dramatically. *The major challenge is how to reduce the PAPR efficiently and effectively in OFDM systems, particularly for multiple antenna systems which have been less studied in the literature.*

PAPR reduction techniques have been developed over the past decade to address this important problem of OFDM systems, however there is always a notice-

able trade-off between reducing the PAPR metric and the sacrificed bandwidth or complexity even in popular techniques. We propose a novel and effective signal processing technique in **Chapter 6** which can be implemented at the transmitter side of MIMO-OFDM systems with minimal complexity overhead. A major competence of the proposed algorithm is its transparency to the receiver which enables the independent implementation in current working wireless systems without a concern to modify the (mobile) receivers.

Research Question 4 To push the boundaries on throughput and performance of wireless systems, MIMO systems have widely been researched over the past decade. The emerging 60 GHz technology sheds new light on MIMO systems by enabling a large number of antennas in a limited space. *How to optimally use multiple antennas and transceivers with respect to their hardware constraints* is the subject of the last research question that is covered in this thesis.

It is no secret that the capacity of the wireless channel can be increased linearly with the minimum number of the transmitters and receivers in a MIMO system [19]. However, the hardware complexity of the system is increased respectively as there are more RF chains including expensive non-linear components. The online complexity of the system can even grow combinatorially as optimal detection is required to be performed on a vector of the received data from multiple antennas rather than a single output. An increased signal processing burden such as higher order equalization, beamforming etc. are other aspects of MIMO systems and this leads to a more complex and susceptible system due to hardware and software imperfections. We aim to limit the complexity and yet benefit from the diversity and multiplexing gains offered by a MIMO system. **Chapter 7** of this thesis is dedicated to formulate and solve for an optimal precoder subject to complexity constraints such as a limitation on the number of RF chains or a per antenna power limit, in this context. This is achieved by jointly defining a precoding and antenna selection pre-processor. The original problem is shown to be extremely difficult to solve and an alternative sub-optimal approach is proposed to solve a relaxed version of the problem.

Besides the revisited and proposed techniques and the application-oriented research questions that were summarized here, the signal processing tools to formulate the problems and the solution mappings are of high importance and are considered as thesis contributions. Optimization theory and techniques are among the most used signal processing tools that have been considered in this work. In particular, we have developed non convex optimization algorithms and convex relaxation

techniques for the problems involving quadratic power expressions that appear frequently in communications, motivated by the common power constraints that are posed in order to avoid unstable systems or to protect hardware components. Also, most of the quantitative measures for performance evaluations are linear or non-linear functions of power, such as channel capacity or interference measures. Some interesting examples, are presented throughout this thesis in **Chapter 6** and **Chapter 7** and many are left for further investigation. **Chapter 8** concludes the thesis by reviewing the presented topics and introducing the related problems and future directions for continuation of this work.

The general notations throughout this thesis are as follows: bold upper case and bold lower case symbols indicate matrices and vectors, respectively and $\mathbf{I}_N$ denotes an identity matrix of size N. The conjugate transpose, conjugate, and transpose of a matrix $\mathbf{A}$ are denoted as $\mathbf{A}^H$, $\mathbf{A}^*$ and $\mathbf{A}^T$. Statistical expectation of vector $\mathbf{a}$ is denoted by $\mathbb{E}\{\mathbf{a}\}$. More specific notations are explained in time. Each chapter follows its own notation, in the sense that the symbols are not globally defined throughout the thesis. In turn, the abbreviations are introduced at each chapter.

1.4 List of Publications

Journals

1. **S. Khademi** and A. J. van der Veen, "Constant Modulus Algorithm for Peak-to-Average Power Ratio (PAPR) Reduction in MIMO OFDM/A," Signal Processing Letters, IEEE , vol.20, no.5, pp.531-534, May 2013,
   doi: 10.1109/LSP.2013.2254114.

2. **S. Khademi**, S. Prabhakar Chepuri, Z. Irahhauten, G. Janssen and A. J. van der Veen, "Channel Measurements and Modeling for a 60 GHz Wireless Link Within a Metal Cabinet," Wireless Communications, IEEE Transactions on, vol.14, no.9, pp.5098-5110,
   doi: 10.1109/TWC.2015.2432755.

3. **S. Khademi**, G. Leus and A. J. van der Veen, "Convex Optimization for Joint Antenna Selection and Precoder Design in Multi-user MISO Systems," Submitted to Signal Processing, Elsevier, April 2016.

Conferences

1. **S. Khademi**, A. J. van der Veen and T. Svantesson, “Precoding technique for peak-to-average-power-ratio (PAPR) reduction in MIMO OFDM/A systems,” Acoustics, Speech and Signal Processing (ICASSP), IEEE 38th International Conference on, pp.3005-3008, March 2012, doi: 10.1109/ICASSP.2012.6288547.

2. **S. Khademi**, S. Prabhakar Chepuri, G. Leus and A. J. van der Veen, “Zero-forcing pre-equalization with transmit antenna selection in MIMO systems,” Acoustics, Speech and Signal Processing (ICASSP), IEEE 39th International Conference on, pp.5046-5050, 26-31, May 2013, doi: 10.1109/ICASSP.2013.6638622.

3. **S. Khademi**, S. Prabhakar Chepuri, Z. Irahhauten, G.J.M. Janssen and A. J. van der Veen, “Channel characterization for wideband 60 GHz wireless link within a metal enclosure,” Antennas and Propagation (EuCAP), 8th European Conference on, pp.1575-1579, April 2014, doi: 10.1109/EuCAP.2014.6902085.

4. **S. Khademi**, E. DeCorte, G. Leus, G. and A. J. van der Veen, “Convex optimization for joint zero-forcing and antenna selection in multiuser MISO systems,” Signal Processing Advances in Wireless Communications (SPAWC), IEEE 15th International Workshop on, pp.30-34, June 2014, doi: 10.1109/SPAWC.2014.6941311.

5. N. Bakhshi Zanjani, **S. Khademi** and G. Leus, “Gradient-based solution for hybrid precoding in MIMO systems,” submitted to Acoustics, Speech and Signal Processing (ICASSP), IEEE 42th International Conference on, March 2017.

Part I

Wideband Channel Characterization and System Design

Chapter 2

60 GHz Channel Measurements and Modeling Within a Metal Cabinet

This chapter presents the channel measurements performed within a closed metal cabinet at 60 GHz covering the frequency range 57- 62 GHz. Two different volumes of an empty metal cupboard are considered to emulate the environment of interest (an industrial machine). Furthermore, we have considered a number of scenarios like line-of-sight (LOS), non LOS (NLOS), and placing absorbers. A statistical channel model is provided to aid short-range wireless link design within such a reflective and confined environment. Based on the measurements, the large scale and small scale parameters are extracted and fitted using the standard log-normal and Saleh-Valenzuela (SV) models, respectively. The obtained results are characterized by a very small path loss exponent, a single cluster phenomenon, and a significantly large root-mean-square (RMS) delay spread. The results show that covering a wall with absorber material dramatically reduces the RMS delay spread. Finally, the proposed channel model is validated by comparing the measured channel with a simulated channel, where the simulated channel is generated from the extracted parameters.

This chapter is published as “Channel Measurements and Modeling for a 60 GHz Wireless Link Within a Metal Cabinet,” Wireless Communications, IEEE Transactions on, vol.PP, no.99, pp.1-1, doi: 10.1109/TWC.2015.2432755.

2.1 Introduction

2.1.1 Problem Context

Inside mechatronic and industrial machinery, the required wiring is an increasing concern, as it comes with issues like reliability, space efficiency, and flexibility. It thus becomes interesting to replace the wires by wireless connections. Literature refers to a so-called "wireless harness" for the communication between components inside machinery devices where the propagation distances are in the order of a few meters or less [20]. On the one hand, using multiple cables inside a dense area to connect moving parts within a confined space can significantly complicate the design and maintenance of the system. A wired connection to a moving part affects the dynamics and may cause cable jams and frequent damage to such machineries. On the other hand, current wireless technology does not meet the data rates offered by wired standards like gigabit Ethernet. To move towards a reliable and fast wireless connection for industrial use, many efforts have been made to provide suitable channel models for the wireless harness applications. In very small-scale applications such as inter chip connections [9] or board-to-board communications [10, 11], a noticeable difference, in terms of channel properties, has been reported in the literature compared with the typical indoor and UWB channels [12, 13, 14, 15, 16]. Furthermore, Ohira et al. studied the propagation characteristic inside the information communication technology (ICT) equipments such as a printer, vending and automated teller machine (ATM) [17] which is the most relevant work in spirit to this chapter as the channel is measured inside a metal enclosure. Also, a simple communication system is tested for ICT devices and associated results are reported in [21].

The unlicensed multi-GHz spectrum available around 60 GHz has gained a lot of interest in the past decade for both indoor and outdoor applications [22, 23, 24]. Specifically, this millimeter-wave band has the ability to support short-range high data rates in the order of Gbps. Both 802.11ad and 802.15c are evolving standards based on this alternative bandwidth (BW) [2, 3]. As a result, many measurements have been conducted to model the propagation environment at 60 GHz. While the literature is mostly concentrated on indoor channel characterization at this band [25, 26, 27, 28], channel models for outdoor implementation of wireless systems based on millimeter-wave have also been investigated [29]. However, there are numerous issues for long-distance communications in this band due to the large attenuation of radio waves because of oxygen absorption. A good survey on channel measurements in 60 GHz can be found in [30].

Channel characterization results for short-range wireless links in the 60 GHz band, have been reported in [31, 32, 3], however, the channel characterization for the so-called wireless harness application [1] is not yet reported. The physically available BW (at least 5 GHz) and small antenna size makes the 60 GHz band very appealing for wireless harness applications. Furthermore, the integration of antennas on small chips [33] can facilitate the deployment of the recently introduced large-MIMO systems [34] which could be a milestone in boosting the data rate in wireless systems.

The main contribution of this chapter is to provide a statistical channel model for applications in the 60 GHz band that operates inside a metal enclosure.

2.1.2 Applications and Motivations

Lithography systems play a critical role in the development and manufacturing of integrated circuits (ICs). The lithography process requires extremely accurate mask and substrate positioning. This task is performed via several sensors and actuators, which are typically connected to the control units via flat-cable wires. In this chapter, we investigate the propagation environment for millimeter-waves inside a lithography system for developing a very high data rate (peak data rate up to a few tens of Gbps) wireless link between the positioning sensors and the control unit. This is fundamental for replacing the wired connections with wireless links.

The sensors and actuators are mounted on moving platforms that experience very high accelerations. The stiffness of the cables causes undesired disturbances to the system which leads to inaccurate positioning. Also, the trend towards increasing numbers of moving sensors makes the design of the wiring system prohibitively complex, therefore the replacement of the cables is of interest.

As we had limited access to an actual lithography machine, the measurements have been conducted inside a metal cabinet that was empty except for some cables, antennas and stand holders. The reproducible setup emulates the propagation environment in a wafer stage section within the lithography machine. This can be described as a metal drawer which is placed in the lithography device and includes two moving wafer stages as illustrated in Fig. 2.1.

This environment contains rather large amounts of open space, in contrast to the compact scenarios found in ICT devices, as investigated in the literature [17]. The initial experiments for establishing the wireless link within the metal enclo-

[1] Kawasaki et al. studied the millimeter propagation environment for internal I/O connections in [9].

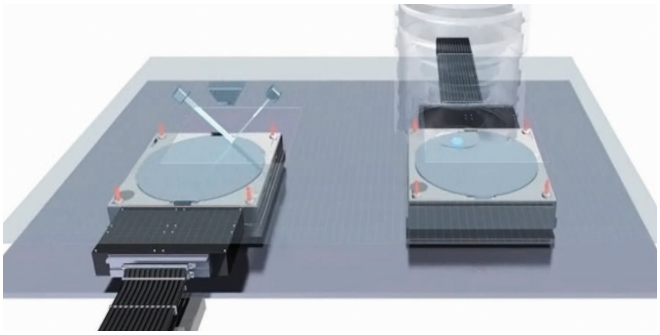


Figure 2.1: An illustration of two moving wafer stages with their cables in a lithography system. The considered measurement scenarios emulate such lithography machines.

sure show an extremely fading environment due to the reflections from the walls, which limits the data rate. Thus, the lack of proper channel models for such hollow and confined environments motivates the considered measurement campaign and modeling.

Apart from lithography machines, there are other systems that can benefit from this work, e.g., scenarios with wireless connections for possible sensors or devices inside an empty elevator or telecabine shaft. The empty cupboard can be viewed as an extreme case of a general metal enclosure. With absorbing objects inside the confined space, one can expect fewer reflections and shorter channel impulse responses.

2.1.3 Outline

In the context of this chapter, we have made extensive measurements of channel frequency responses using a channel frequency domain sounding technique within the 57-62 GHz band. This has been done by placing the receiver on a pre-designed spatial grid, step by step, while the transmitter is fixed. The power delay profile and multipath components are extracted by post processing. Two different volumes of the metal cupboard are used and the measurements are provided for both the LOS and NLOS scenarios. The results indicate that the environments within metal enclosures are highly reflective, and the resulting “long” wireless channels will make wireless communications very challenging. Also, the fading properties change depending on the volume of the cupboard rather than the LOS and NLOS situations. We have also used absorbers to cover a metal wall for one scenario which resulted in a significant reduction in the root-mean-square (RMS) delay spread (RDS) and

this consequently affects the fading properties of the channel.

Both small-scale and large-scale channel model parameters are extracted from the measurements, based on the well-known Saleh-Valenzuela (SV) [35] and log-normal model [36, 19], respectively. Accordingly, a comprehensive statistical channel model is provided to simulate similar fading channels. Random channel instances are generated based on the extracted parameters for arrival time, time decay constant, and number of paths. Next, the RDS properties of the simulated and measured channels are compared. The purpose of this verification is two fold. Firstly, it assures whether the number of measurements is sufficient for extracting the parametric statistical channel model. Secondly, it validates the accuracy of the model itself. Together with the Doppler frequency change (time variance property), the proper channel instances can be simulated via the Matlab channel modeling toolbox [37] or other off-the-shelf simulation software based on SV or stochastic tap-delay-line models [13, 38].

The remainder of this chapter is organized as follows. In Section 2.2.1, we describe the measurement set-up and explain the measurement procedure. In Section 2.2.2, we provide details regarding data processing to extract parameters required for channel modeling. Based on these parameters, large-scale (path loss) and small-scale channel models (RDS) are presented in Sections 2.3 and 2.4, respectively. The proposed statistical channel parameters based on the SV model (time decay constant and arrival rates) are given in Section 2.5. The proposed channel model is validated together with the coherence time and bandwidth of the system in Section 2.6. Also, we compare the statistical parameters for the measured channels with the SV channel model suggested for the IEEE 802.15 standard and other related measurements in the literature. Final remarks are made in Section 2.7.

2.2 Measurement Set-up and Procedure

In this section, the channel measurement procedure and details of the equipment used for the measurements are explained.

Channel characterization can be performed in either time domain or frequency domain [39]. In the measurements provided in this chapter, a frequency domain sounding technique is used. The scattering parameters (i.e., S_{11}, S_{12}, S_{21}, and S_{22}) are measured using a vector network analyzer (VNA) by transmitting sinusoidal waves at discrete frequencies. The frequency spacing, Δf_s, and the scanned BW, B_w, determines the maximum measurable excess delay, τ_{max}, and the resolution

of the captured multipaths, τ_{res}, respectively, and they are given as

$$\Delta f_s = \frac{B_w}{N_s - 1}, \quad \tau_{max} = \frac{1}{\Delta f_s}, \quad \tau_{res} = \frac{1}{B_w}, \tag{2.1}$$

where N_s is the number of transmitted sinusoidal waves.

The frequency domain S_{21} parameter is generally referred to as channel frequency response. The channel impulse response (CIR) is obtained from the measured channel frequency response by taking the inverse fast Fourier transform (IFFT). A Hann window is applied to reduce the effect of side lobes.

2.2.1 Measurement Set-up

The measurement BW is set to $B_w = 5$ GHz, and the channel is sampled from 57 GHz to 62 GHz at $N_s = 12001$ frequency points. This results in a frequency spacing of $\Delta f_s = 0.416$ MHz, so that the time resolution is $\tau_{res} = \frac{1}{B_w} = 0.2$ ns and the maximum measurable excess delay is $\tau_{max} = 2400$ ns. The channel frequency response is measured using a PNA-E series microwave VNA E8361A from Agilent. An intermediate frequency BW of $B_{IF} = 50$ Hz is chosen to reduce the noise power within the measurement band, which improves the dynamic range. This is the receiver BW for single sinusoid in a VNA; the smaller intermediate frequency BW leads to a larger signal to noise ratio. Also each measurement is repeated 50 times to further average out the noise.

Due to the losses inside the VNA and 60 GHz co-axial cables, the measured signal at the receiver is weak. A 60 GHz solid state power amplifier (PA) from QuinStar Inc. (QGW-50662030-P1) is used to compensate for the losses and to further improve the dynamic range. An illustration of the measurement set-up is provided in Fig. 2.2. For the transmit and receive antennas, we have used two identical open waveguide antennas operating in 50-75 GHz frequency band with aperture size 3.759×1.880 mm^2. The beam pattern of the antennas is shown in Fig. 2.3. The gain of the open waveguide antenna is about 4.6 dBi (see [4] for details on computing the gain).

The near field distance for the antenna is calculated based on the Fraunhofer distance and it is found to be less than 3 mm from the antenna aperture. Therefore, all the measurements are taken in the far field, and hence, there is no near field effect considered here. Two holders are used to fix and elevate each antenna to avoid coupling between the antenna and metal surface of the metal enclosure.

To investigate the channel behavior within the empty metal cabinet, we have considered the following four scenarios. *Scenario 1* is an LOS scenario where we

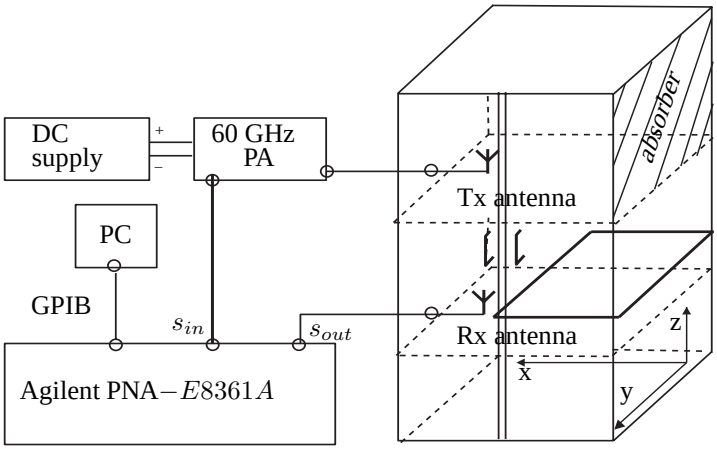


Figure 2.2: Measurement setup for channel sounding inside the metal cabinet. The solid parallelogram just above the first level shows the metal plate that has been used in the NLOS scenario. The top right wall is covered with absorber for *scenario 4* (small size cabinet).

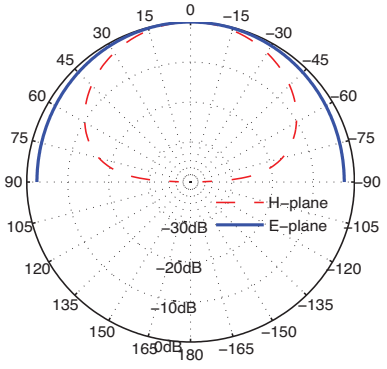


Figure 2.3: Field radiated by the TE_{10} mode in open waveguide antenna with respect to θ angle.

Table 2.1: Receive antenna co-ordinates

	x-axis	y-axis	z-axis
Scenario 1	15-85 cm; 8 steps	5-30 cm; 6 steps	150,165 cm; 2 steps
Scenario 2	15-85 cm; 8 steps	5-30 cm; 6 steps	40,145 cm; 2 steps
Scenario 3	15-40 cm; 6 steps	5-30 cm; 6 steps	40,145 cm; 2 steps
Scenario 4	15-35 cm; 5 steps	5-30 cm; 6 steps	150,165 cm; 2 steps

use a metal enclosure of dimension $100 \times 45 \times 45$ cm^3. *Scenario 2* is an LOS scenario with a metal enclosure of a larger dimension, i.e., $100 \times 45 \times 180$ cm^3. *Scenario 3* is a NLOS scenario with the dimensions $100 \times 45 \times 180$ cm^3. *Scenario 4* is an LOS scenario as in *Scenario 1* except that one of the side walls is covered with an absorber (see the illustration in Fig. 2.2). Absorbers are an alternative physical solution to reduce the channel length which will simplify the required channel equalization. Note that the volume of the metal enclosure in *scenario 2* and *scenario 3* is four times larger than the volume of the metal enclosure used for *scenario 1* and *scenario 4*. To block the LOS path, and create the NLOS scenario, a 50×45 cm metal separation plate is used in *scenario 3* as illustrated in Fig. 2.2.

The transmit and receive antennas were placed on a styrofoam (polystyrene) sheet, which acts as vacuum for radio waves and has a negligible effect on the channel behavior. The transmit and receive antennas were supported using clamps (stand holders) with sufficient clearance from the metal surface. The co-axial cables were drawn into the metal cabinet by means of small holes which are just sufficiently large to pass the cable.

For all scenarios, the location of the transmit antenna was kept fixed. The channel was measured at various locations in 3 dimensions, i.e., x, y, z-axes, as specified in Table 2.1. This produced 96, 96, 72 and 60 receiver locations for scenario 1, 2, 3 and 4, respectively. Two elevation steps were used in z-axis, 6 steps in y-axis and 8, 6 and 5 steps in x-axis for different scenarios as shown in Table 2.1.

In *scenario 1* and *scenario 4* the transmit antenna was fixed at co-ordinate $(x_t, y_t, z_t) = (65, 15, 135)$ cm, and in *scenario 2* and *scenario 3* the transmit antenna was located at $(x_t, y_t, z_t) = (15, 15, 130)$ cm. The position of the metal plate was at $z \approx 60$ cm and $z \approx 140$ cm for the first and second steps in z-axis in *scenario 3*. In *scenario 4*, the bulky absorbers were limiting the space so less measurements were taken in this scenario and only the RDS spread property has been extracted. The minimum and maximum distances between Tx and Rx are in the range of 1.5 m to 15 cm.

2.2.2 Data Processing

Post-processing of the data is required to extract the CIR from the measured frequency domain signals. In principle, this involves an inverse discrete Fourier transform (IDFT). The IDFT includes a window; the resulting impulse response is thresholded to remove paths with small amplitudes. Prior to the IDFT, we cancel the antenna and instrument responses by using an inverse filtering technique

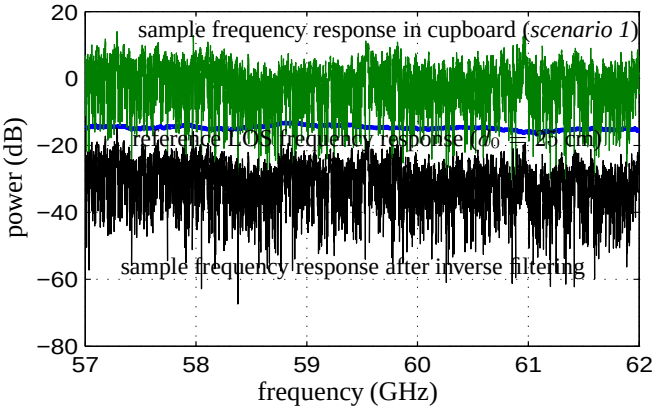


Figure 2.4: Sample channel frequency response from *scenario 1* before (lower channel frequency response) and after inverse filtering (upper channel frequency response) and reference channel frequency response with $d_0 = 25$ cm (line in the middle).

[40, 41] which is briefly explained in Appendix 2.A.

Fig. 2.4 shows the original frequency domain response of a sample measurement from *scenario 1*, the frequency domain response after inverse filtering and the frequency domain signal of the truncated reference measurement. The effect of inverse filtering can be observed after calibration plot where the sample channel frequency response is normalized by $R_{fl}(f)$ which is the channel frequency response for free space without reflections or obstructions consists of a single LOS path. $R_{fl}(f)$ is parametrized by an attenuation and a simple delay equal to the time- of-flight of the signal between the transmit and receive antenna. We can make a recording of the received signal at a known reference distance in free space, and after time gating we obtain $r_{fl}(f)$ which is the CIR corresponding to $R_{fl}(f)$. The change in the power levels after inverse filtering is due to the compensation of antenna and instrument responses.

For model parameters that do not depend on the absolute power (i.e. the small-scale channel model considered in Section. 2.4), we have normalized the received signal to have a maximum value at 0 dB. The dynamic range of the received signal is in the order of 70 dB, where we assume that the noise level is at -70 dB after normalization.

For estimating statistics of the individual link parameters, it is useful to truncate the duration of the channel. We compute the threshold taking into account the noise level, amount of total received power and relevant multipath components [42, 43]. By setting a threshold at 30 dB below the strongest path, more than 98% of the

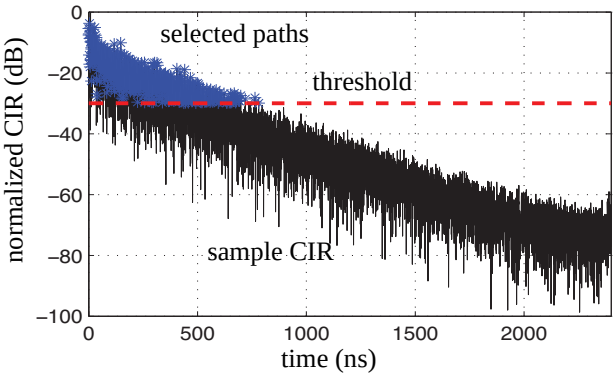


Figure 2.5: Sample CIR with 30 dB threshold and received paths for *scenario 1*.

total power is captured. This threshold is still well above the noise level. As an illustration, Fig. 2.5 shows a normalized received CIR with a threshold at -30 dB. The duration of this channel is still about 800 ns.

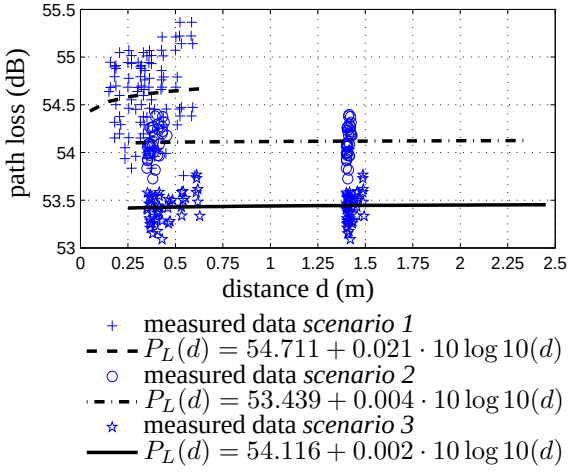
2.3 Path Loss Model

The large-scale channel model, specifically the path loss model, is essential for any wireless system design to calculate its link budget. For a conventional channel (outdoor or indoor), the path loss model suggests that the average received power decreases exponentially with increasing distance between the transmitter and receiver. This is generally expressed in logarithmic scale as

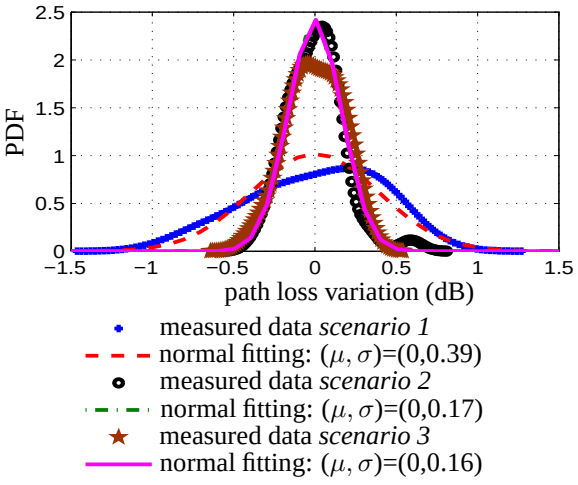
$$P_L(d)_{dB} = P_L(d_0)_{dB} + 10\alpha \log_{10}(\frac{d}{d_0}) + X_\sigma. \tag{2.2}$$

where $P_L(d)_{dB}$ is the signal power loss at a distance d (m) relative to an arbitrary reference distance d_0 (m), α represents the path loss exponent, and X_σ is a zero-mean Gaussian random variable with standard deviation σ reflecting the attenuation (in dB) caused by shadowing [36, 19]. In fact, the first two terms in (2.2) together represent the expected path loss and the last term represents the random variations of this model. Based on the measurements, first the parameters of the statistical model are identified for the average received power and the path loss exponent and later the shadowing parameters are determind.

Using the measurements of the received power for different distances between the transmit and receive antennas, we can estimate the path loss exponent α. Accordingly, for each measurement the distance related path loss term in dB ($P_t - P_r$)



(a) Path-loss as function of distance



(b) PDF of the path loss variation X_σ

Figure 2.6: Path-loss model parameter estimation.

is calculated, based on the known transmit power (-68 dB), as shown in Fig. 2.6a which shows that the path loss exponent α is very small (around 0.02-0.002). The reference distance is taken as 1 meter similar to common indoor environments. This suggests that in such a closed metal environment there is nearly no loss in the received power as function of distance. The same phenomenon is reported in [10] for the environment inside a computer case. Other measurements for NLOS wireless personal area network (WPAN) reported α in the range of $0.04 - 0.09$ [44, 42], while α in the range $1.6 - 6$ is common for typical indoor systems [36]. According to the Friis formula, the path loss for conventional indoor environments should be larger for transmissions at 60 GHz compared to lower carrier frequencies. However, this is not the case for highly reflective environments such as metal enclosures.

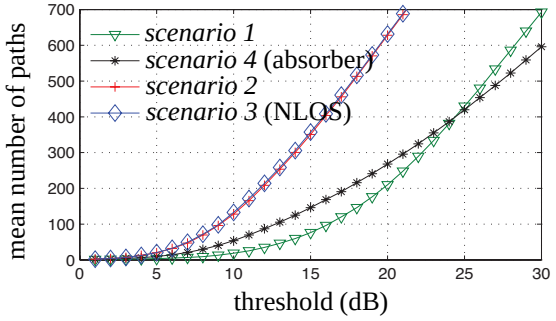
An ideal metal enclosed environment acts as a semi-conservative physical system where the only sources of absorption are the antennas, cables and stand holders. The waves keep bouncing back and forth, and when the distance between the antennas is increased, the received power does not fluctuate because most of the energy reaches the receive antenna either directly or as multipath reflection in the metal cabinet.

Fig. 2.6b shows the probability density function (PDF) of X_σ, i.e., the fluctuation of the path loss around the regression line in Fig. 2.6a. It is seen that the PDF approximately follows a normal distribution, with a standard deviation of 0.16–0.39 dB. Among the considered scenarios, the NLOS case (*scenario 3*) shows the smallest variation, and this is due to the larger distances (volume) and the obstructed LOS path. In general there is no noticeable shadowing effect in the environment even in NLOS case, since the reflected paths are almost as strong as the LOS path in the metal enclosure.

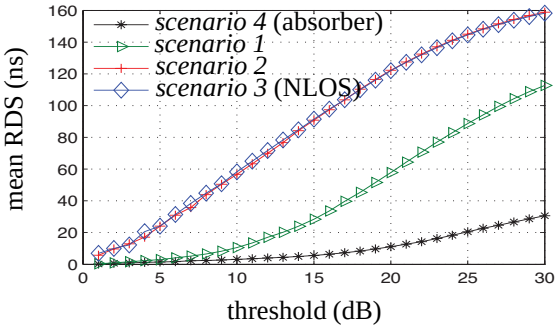
Accordingly, the large scale properties of the channel has been fitted to the well-known log-normal model in (2.2), and can be used for the wireless system design within empty (not-dense) metal enclosures.

2.4 RMS Delay Spread (RDS)

Besides path-loss, the channel can be further characterized by its small-scale properties caused by reflections in the environment, which are modeled as multipath components [36, 19]. We do not consider fading on individual delay paths since the measurements show that there are few multipath components in each resolvable time bin (over the measurement grids), and hence, they are not considered



(a) Number of received paths



(b) mean RDS

Figure 2.7: Number of received paths and RDS for different thresholds.

directly in our model. Instead, we consider the statistics of the model parameters for the (normalized) power delay profiles (power delay profiles) obtained over all the spatial grids i.e. power delay profile$^{(g)}(\tau) = |\mathbf{h}^{(g)}(\tau)|^2$, where g denotes the grid (position) point [16]. For example, $g = 1, 2, \cdots, G = 96$, for *scenario 1* and *scenario 2*. The nth multipath component denoted by nth entry of $\mathbf{h}^{(g)}(\tau)$, and it is described by its power a_n^2 and arrival time t_n.

Multipath leads to small-scale fading (variations over short distances due to constructive and destructive additions). The most important model parameters that describe a multipath channel variations are the RDS and fading properties that can be modeled as the time decay constant and the multipath arrival times in the SV model. These aspects are studied next.

Delay spread describes the time dispersion effect of the channel, i.e., the distribution of the received power in time. A large delay spread causes severe intersymbol interference (ISI) and can deteriorate the system performance. The RDS is a commonly used parameter to characterize this effect [19]. The RDS is obtained

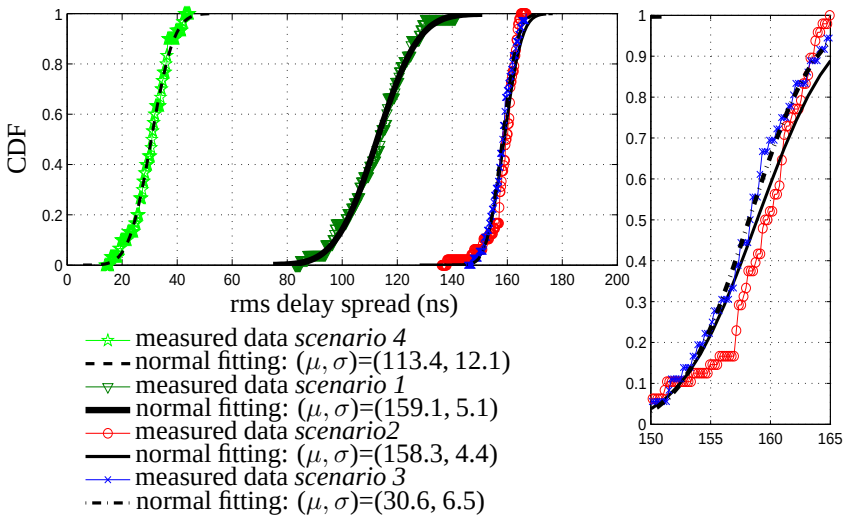


Figure 2.8: Cumulative distribution function for RDS of measured channels.

by first estimating the individual path parameters $\{(a_n^2, t_n)\}$ for each observation, and then computing

$$t_{rms} = \sqrt{\bar{t^2} - (\bar{t})^2}, \quad \bar{t^\varrho} = \frac{\sum_{n=1}^{N} a_n^2 t_n^\varrho}{\sum_{n=1}^{N} a_n^2},$$

where $\bar{t}$, $\bar{t^2}$ and $\bar{t^\varrho}$ are the first, second and ϱ moment of the delay spread, respectively.

Fig. 2.7a shows the number of received paths for different power thresholds. As expected, the number of received paths (N) increases with increasing threshold level. The received paths are saturated more quickly in *scenario 4* due to the absorbers. In the same way, the RDS increases as the number of collected paths increases (Fig. 2.7b). At a threshold of 30 dB, the curves saturate and we used the corresponding value as the estimated RDS. Fig. 2.8 shows the cumulative distribution function (CDF) of the estimated RDS values for all the four scenarios. The figure also shows the fit to a normal distribution. The mean values of the normal distribution, obtained after fitting, reveals the average length of the channel, and they are 113.4 ns (*scenario 1*), 159.1 ns (*scenario 2*), 158.3 ns (*scenario 3*), and 30.6 ns (*scenario 4*). These mean RDS values for empty metal enclosures are significantly larger than the conventional indoor channels, which are typically between $4 - 21$ ns.

These large values will impact the system design and signal processing within

such environments, e.g., the channel equalization and residual inter block interference (IBI) after equalization, and hence, the achievable data rates.

Note that the estimated mean RDS is almost the same for *scenario 2* and *scenario 3*, which shows that there is a clear relation between the volume of such metal enclosures and RDS, independent of LOS and NLOS cases. Also, in *scenario 4* the RDS is reduced by more than 3.5 times as compared to the empty cupboard in *Scenario 1*. These are very interesting results and indicates that even covering one wall with the absorber can reduce the channel length and fading almost to that of a typical indoor environment.

2.5 Saleh-Valenzuela (SV) Model Parameters

Most current IEEE standard channel models [45, 2] and MIMO channel characterizations [28] for millimeter-wave are based on the extended SV model [46, 47]. In this model, the multipaths are considered as a number of rays arriving within different clusters, and separate power decay constants are defined for the rays and the clusters. This is a very well-known and well-validated model for wireless channels with multipath which was proposed to cover the shortcoming from the traditional Rayleigh (Nakagami) models to describe the statistical power delay profile. For instance in UWB channel when only the superposition of few multipath components falls within each resolvable delay, the central limit theorem does not hold anymore. This also is the case in our measurements as the high resolution in time makes it less probable to find many multipath component within each time bin (channel tap) to derive the fading parameters [16] over each path. Accordingly, we use the SV model by extracting the corresponding statistical parameters from the measurement data.

Furthermore, these parameters can be used to generate channel instances with identical statistical properties by defining the average power delay profile based on the extracted parameters together with the Doppler frequency information. We only derive the SV model parameters for the empty cupboard in *scenarios 1-3* and not for *scenario 4* as the focus of the work is on the empty metal enclosure.

2.5.1 Time Decay Constant

A cluster is defined as a group of arrival paths that are reflected from the objects with the same angular profile. One of the common and basic methods to identify the clusters in the channel impulse response (CIR) is by visual observation. We

carefully observed the CIRs that were obtained at different positions. Our observation do not show that the multipath components come form multiple clusters i.e. the power in CIRs is exponentially decaying over the channel length time. This has been observed visually over the measured CIR and verified by the estimated decay parameters. A physical justification comes from the fact that multipath reflections are coming from the (same) walls. Note that if paths from different clusters arrive with the same delay, then the observation technique can not resolve this ambiguity.

In this case, the average power delay profile is defined by only one decay parameter γ rather than the common SV model with two decay parameters. Therefore, the proposed model can be given as:

$$\bar{a}_n^2 = \bar{a}_0^2 \, \exp\left(-t_n/\gamma\right), \tag{2.3}$$

where $\bar{a}_0^2$ and $\bar{a}_n^2$ are the (statistical) average power of the first and nth multipath component, respectively, over all different positions and γ is the power decay time constant for arriving rays, assumed as a random variable. To find the decay parameters first we compute the normalized logarithmic power delay profiles for each measurement. We estimate γ_k for each measurement (each position indicated by index k) in every scenario using a least-squares curve fitting on $\log(a_n^2)/\log(a_0^2)$, as shown by the examples in Fig. 2.9. Time delay instances on the x-axis indicate the arrival time for multipath component with respect to the first path.

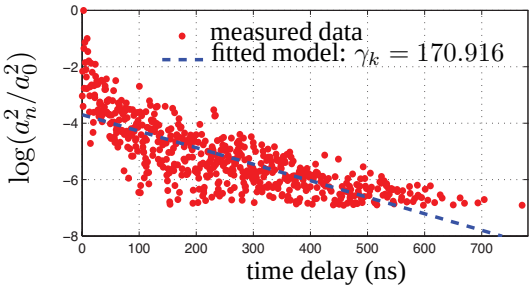
Based on these estimates for the γ_ks which are different realizations for random variable γ, the PDF for γ is plotted and fitted to Gaussian, Gamma, and Weibull distributions for each considered scenarios, as shown in Fig. 2.10. These distributions are commonly used to statistically model γ [44, 42].

The best fitted model is chosen as the argument which minimizes the Akaike Information Criterion (AIC) i.e., the distribution that maximizes the log likelihood function in the estimation problem. Accordingly, a Gamma distribution has been chosen as the best fit for the γ distribution in *scenario 1* and *scenario 2* while Weibull distribution is the best candidate in *scenario 3* in the sense that we loose less information by using these models rather than real data.

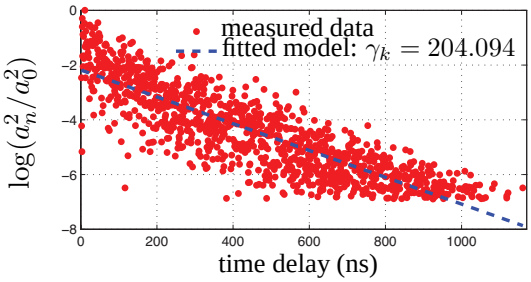
We use the statistically estimated γ in the rest of the chapter. The Gamma distribution is given by

$$f(x|\delta,\beta) = \frac{x^{\delta-1}}{\beta^{\delta}\mathcal{E}(\delta)} \, \exp(-\frac{x}{\beta}), \tag{2.4}$$

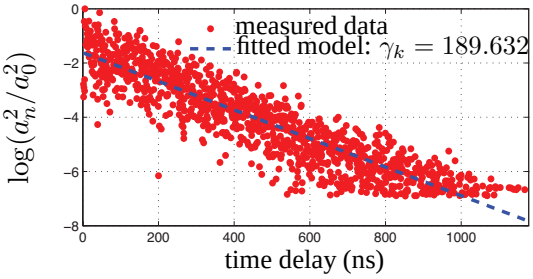
where $\mathcal{E}(\delta)$ is a Gamma function, and the parameters δ and β are computed for all



(a) sample measurement in *Scenario 1*

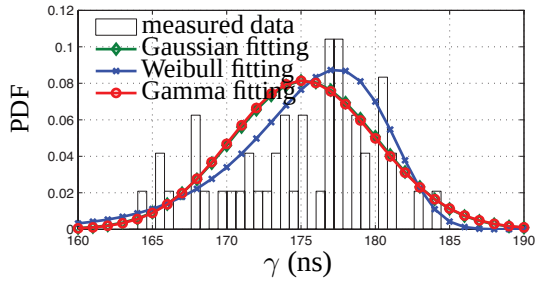


(b) sample measurement in *Scenario 2*

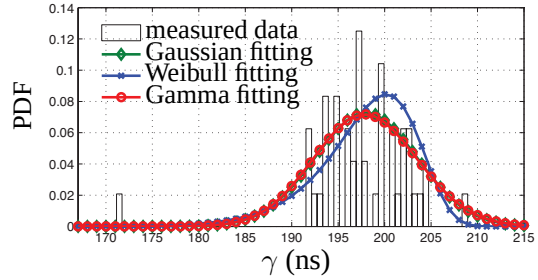


(c) sample measurement in *Scenario 3*

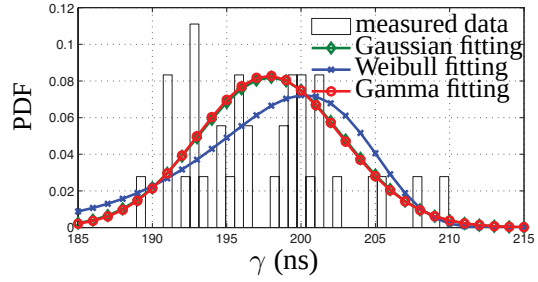
Figure 2.9: LS fit for time decay constant γ_k for each measurement.



(a) *Scenario 1*
Gaussian$(\mu, \sigma) = (175.2, 4.901)$
Gamma$(\delta, \beta) = (1281, 0.137)$
Weibull $(\zeta, k) = (177.5, 42.28)$



(b) *Scenario 2*
Gaussian$(\mu, \sigma) = (197.9, 5.481)$
Gamma$(\delta, \beta) = (1265, 0.156)$
Weibull $(\zeta, k) = (200.3, 46.05)$



(c) *Scenario 3*
Gaussian$(\mu, \sigma) = (197.9, 4.865)$
Gamma$(\delta, \beta) = (1689, 0.117)$
Weibull $(\zeta, k) = (200.4, 39.37)$

Figure 2.10: PDF fittings for time decay constant γ.

scenarios from the empirical data. The Weibull distribution is expressed as

$$f(x|\zeta,k) = \begin{cases} \frac{k}{\zeta^k}\, x^{k-1}\, \exp\big(-(\frac{x}{\lambda})^k\big) & \text{if } x \geqslant 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{2.5}$$

where the scale and shape parameters are ζ and k, respectively.

There are more accurate techniques to estimate the cluster decay which are specially developed for mm-wave channels when the dynamic range of the system is limited due to the high path-loss, that are not applicable for our measurements [28].

2.5.2 Multipath Arrival Times

Next we consider a statistical model on the multipath arrival times (t_n) in order to offer a complete channel model. This gives insight about how dense or sparse the channel is in terms of multipath components and is calculated based on the time difference between two consecutive multipath components. The inter arrival times $t_n - t_{n-1}$ gives the time between the events of multipath arrivals. The multipath arrival times t_n would be typically modeled as a single Poisson process within each cluster. Having one extended cluster as we observe in our measurements cannot be suitably expressed with a single Poisson process. This is due to the fact that the Poisson parameters are considered unrelated to the delays and are treated independently, which does not reflect the reality, so we use different Poisson models for different delay areas.

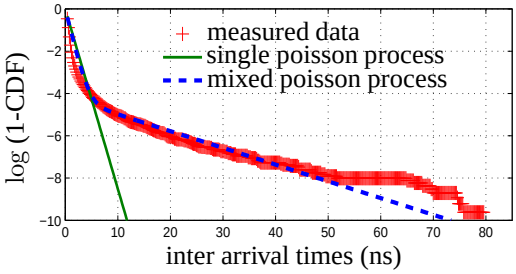
For a single Poisson process, the inter arrival times $t_n - t_{n-1}$ are modeled by an exponential PDF as

$$p(t_n|t_{n-1}) = \lambda \exp\big(-\lambda(t_n - t_{n-1})\big), \tag{2.6}$$

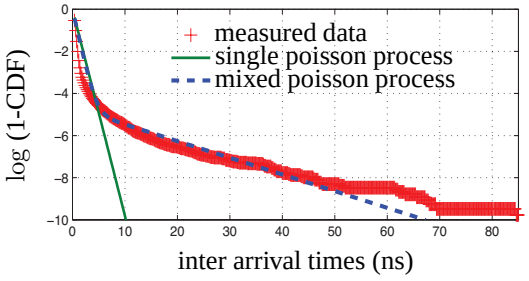
where λ is the mean arrival rate of the multipath components. It is motivated in [48, 42] that when the measured arrival times deviate too much from the single Poisson model, a mixture of two Poisson processes is more suitable for modeling their arrival times. The mixture of two Poisson processes can be expressed as

$$\begin{aligned} p(t_n|t_{n-1}) &= b\,\lambda_1 \exp\big(-\lambda_1(t_n - t_{n-1})\big) \\ &+ (1-b)\,\lambda_2 \exp\big(-\lambda_2(t_n - t_{n-1})\big), \end{aligned} \tag{2.7}$$

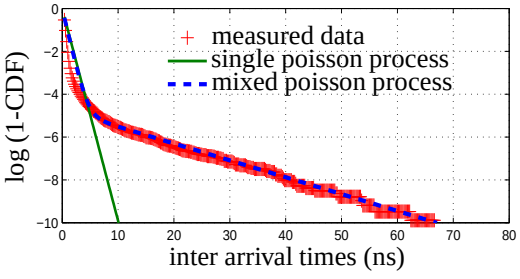
where λ_1 and λ_2 are the arrival rates and parameter $0 \leq b \leq 1$ is the mixing probability.



(a) *Scenario 1*: $\lambda = 0.985$,
$(\lambda_1, \lambda_2, b) = (0.083, 1.180, 0.015)$.



(b) *Scenario 2*: $\lambda = 1.037$,
$(\lambda_1, \lambda_2, b) = (0.059, 1.219, 0.008)$.



(c) *Scenario 3*: $\lambda = 1.094$,
$(\lambda_1, \lambda_2, b) = (0.084, 1.235, 0.009)$.

Figure 2.11: Logarithm of the complementary CDF of the inter arrival times.

Fig. 2.11 shows the corresponding estimated parameters. The inter arrival times are indicated on the x-axis while the logarithmic complementary CDF is shown on the y-axis as it is more informative due to the exponential nature of the Poisson process. As seen, the mixed Poisson process provides a much closer fit to the measured data than the conventional single Poisson process. In fact, parameters b, λ_1 and λ_2, that are estimated and stated in Fig. 2.11, are used further to generate random arrival time values to be used in the production of the channel instances via simulations.

Similar results are reported in IEEE 802.15.4 [48] for device to device communication for ranges less than 10 m (WPAN). Apparently, if the RDS or channel length is large, the arriving paths appear over a wide range of time differences which makes it difficult to be represented by only one Poisson parameter. The results indicate that the inter arrival times are smaller, in general, compared to conventional indoor channels reported in [48, 3]. This indicates the richer scattering environments of the examined metal enclosure.

2.6 Validation and Evaluation

In this section, we validate our proposed statistical model via Matlab simulations and subsequently we study the behavior of the channel with respect to time. The coherence bandwidth of the measured channel is calculated based on the RDS parameters extracted in Sec. 2.4. Finally, channel model parameters from related measurements are compared with extracted model parameters to give an analogy between different environments and applications.

2.6.1 Validation of the Proposed Model via Simulations

We use the estimated SV parameters all the previous section to simulate CIRs and later to compare the properties of these model based simulated channels with the measured channel. This is a straightforward way to validate the proposed statistical channel model. In order to generate a CIR, we need the time instances of multipath arrivals and the energy associated with each path, which are both random variables that are estimated with λ and γ in Sec. 2.5, respectively. Also, we need to define the number of paths for each channel instance which is a normal random variable itself with certain mean and standard deviation. Having these statistical properties we are able to generate random CIR. Note that the quality of the fit for the power delay profiles are examined implicitly through the simulation of the RDS parameters as

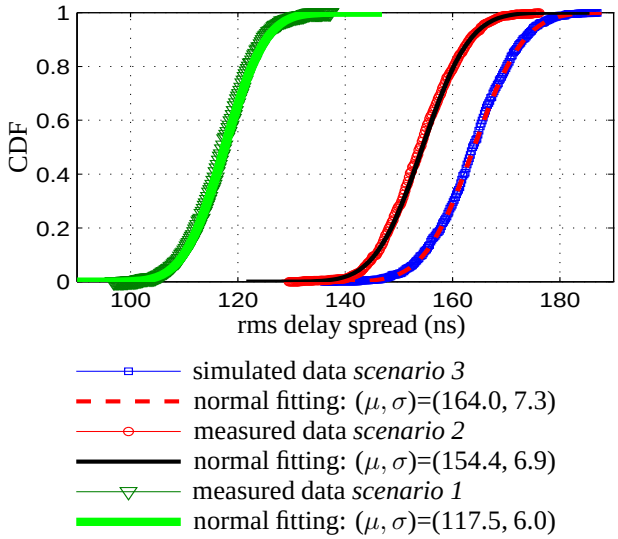


Figure 2.12: Cumulative distribution function of RDS based on 1000 simulated channel instances, from left to right are *scenario 1* to *scenario 3* with the fitted model on top of each scenario.

the random power delay profiles are generated for the simulation of each scenario using the estimated statistical values in Fig. 2.10. We use the RDS for the validation phase as it comprehensively includes all the parameters of the proposed model.

We have simulated 1000 channels using the proposed model parameters for all three scenarios within the empty metal enclosure. The RDS is calculated for these channel instances and the CDF curves with a fitted mean and variance are illustrated in Fig. 2.12.

In *scenario 1* and *scenario 3* there is a small (almost 5 ns) overestimation (4.5% and 3% error) and in *scenario 2*, an underestimation (3% error) of the mean RDS, in comparison to the measured values which shows an acceptable model estimation error. As a result, the proposed model parameters are valid and can be used to simulate random channels for link design and other studies that require the channel model.

2.6.2 Coherence Time and Bandwidth

A good channel model describes the statistical channel strength over both time and frequency domains. The time varying nature of the channel is characterized by the Doppler frequency shift. The resulting coherence time is directly defined

by the relative movement (speed) between transmitter and receiver so this is an application specific parameter [19]. The under-test lithography system is part of a mechatronic device in a closed metal environment in which sensors and actuators on a moving platform have to communicate to a controller on the fixed platform. Since movements that occur outside the enclosure do not affect the channel, we expect a slowly time-varying channel with a sufficiently long coherence time. The Doppler shift is defined as $\Delta f_D = \frac{\nu f_c}{c}$, where ν is the relative speed between transmitter and receiver, c is the speed of light, and f_c is the carrier frequency. If we assume a maximum relative speed of $10\ \mathrm{ms}^{-1}$, then the Doppler frequency range is $\Delta f_D = 2$ kHz, and the coherence time of the channel is $\frac{1}{\Delta f_D} = 0.5$ ms.

The coherence BW denoted as B_c gives a sensible insight into the wideband fading model of the system and is directly estimated from the RDS of the channel. A general approximation is $B_c \approx \frac{\iota}{\mu_c}$, where ι depends on the shape of the power delay profile and μ_c is the so-called mean RDS extracted in Fig. 2.8. A 90% coherence bandwidth is defined as the separation in frequency such that the cross correlation between two frequency samples of the channel is 0.9 i.e., $B_c \approx \frac{0.02}{\mu_c}$ [19]. This is the so called 90% approximation of the mean coherence BW which for different measured scenarios are reported as 176.4, 125.7, 126.3, and 653.6 kHz for scenarios 1 to 4, respectively. The 50% coherence BW is 10 times the 90% values.

The coherence BW for the empty metal box is extremely small, this is visually clear from Fig. 2.4 where the sample channel frequency response shows the dynamic range of almost 30 dB while the reference measurement outside the cupboard is mostly a constant. Note that equalization for such an extreme frequency selective environment is very complex if not impossible. Moreover, despite the general understanding of the 60 GHz propagation environment in outdoor and typical indoor places, the channel does not follow the sparse model in the time domain but it can be considered relatively sparse in the frequency domain.

2.6.3 Comparison to Other Channel Models

To the best of our knowledge, there are no 60 GHz channel models for very short-range wireless communications (wireless harness) prior to this work. However channel modeling has been done for the IEEE 802.15c standard, for small indoor environment such as cubic offices and kiosks which we discuss here for the sake of comparison. We also compare our obtained results with the channel characterization of a room with metal walls [49] as well as a reflective environment when

metallic cabinets are located in the middle of the room [50].

For lower frequencies (3-5 GHz) the results in [17] are interesting for comparison because of the application similarity but the parameters for path loss are expressed in terms of the customized three part model (near, transition and far field) which do not comply with our log-normal model. However, there are some other interesting measurement results for short range wireless applications that we summarize here.

- In [49], path loss and RDS are studied for a 2 GHz band centered at 58 GHz for different room dimensions and properties. In two scenarios, rooms with metal walls are considered with dimensions $44.7 \times 2.4 \times 3.1$ m^3 and $9.9 \times 8.7 \times 3.1$ m^3. For a reference distance of $d_0 = 1$ m, $P_L(d_0)$ around 80 dB and $\alpha < 0.5$ have been reported. Also, the RDS in order of 100 ns is measured which is very close to the results from the metal cabinet.

- In [50], a 60 GHz measurements have been conducted in a room with dimensions of $11.2 \times 6.0 \times 3.2$ m^3 with metal reflectors such as metal walls within the room for LOS and NLOS scenarios as well as for different antenna settings. $P_L(d_0)$ with $d_0 = 1m$, for the Tx-antenna heights of 1.4, 1.9, and 2.4 m are 56.1, 66.8 and 73.1 dB (71.1, 75 and 77.7 dB) for LOS (NLOS), respectively. Path loss exponents of 1.17, 0.18 and 0.61 (5.45, 3.82 and 2.67) are reported for the different Tx elevations for LOS (NLOS) scenarios. As can be seen, small αs in the LOS cases are similar to the ones from the metal cabinet.

- In [51], channel characterization is provided for elevator shafts at 5 GHz with 50 MHz BW, the mean RDS values are reported as 14-60 ns for a still elevator, at different locations (buildings) and the maximum RDS is recorded between 144-152 ns when it is moving (different scenarios with the receiver inside the elevator car and outside are tested). RDS values similar to our measurements, are observed here. The derived log distance models show the path loss exponent in the range of $2.75-6.66$ when the elevator door is closed and $2.40 - 5.76$ when it is open. Also, the shadowing normal distribution exhibits a standard deviation (σ_{P_L}) of $1.89-6.08$ dB (door closed) and $2.37-5.52$ dB (door open).

- In [11], measurements have been conducted in a computer case at 3.1-10.6 GHz (7.5 GHz BW) for a wireless chip area network (WCAN) application. Param-

eters α, $P_L(d_0)$ and σ_{P_L} are 1.607, 23.78 dB, and 0.548 dB (2.692, 25.27 dB, and 1.908 dB) for case closed (case open), respectively, for $d_0 = 62$ cm.

- In a similar work in [10], for board-to-board communication in two computer cases (both dense and sparse) the path loss exponent was reported to be negligible where $P_L(d_0)$ and σ_{P_L} appeared as 29.1 dB and 1.4 dB (28.7 dB and 1.4 dB), for the dense (sparse) case, respectively. The 50% coherence BW of the channel and γ are reported as 79 MHz and 3.49 ns (51 MHz and 5.44 ns) for dense (sparse) case. One γ parameter is considered in this work similar to a single cluster model in this paper. The results show a greater coherence BW and consequently smaller time decay constant compared to our estimated parameters mostly due to the a small volume of the computer case and many absorbing objects inside the metal box. Some losses also can be related to the ventilation holes in the case.

The estimated parameters for our proposed channel model are summarized in Table 4.1, together with the channel model parameters for IEEE 802.15. In this table, the listed parameters are:

$P_L(d_0)$: path loss at reference distance d_0 (m)

α : path loss exponent

σ_{P_L} : path loss log-normal standard deviation

$\bar{L}$: mean RDS

Λ : cluster arrival rate

λ : ray arrival rate (single Poisson fit [2])

Γ : power decay constant for clusters

γ : power decay constant for rays

σ_Γ : cluster power decay log-normal standard deviation

σ_γ : ray power decay log-normal standard deviation

[2]The single Poisson parameter is shown here since we want to compare it to other models which use single Poisson fit.

Table 2.2: Comparison of various channel parameters of the measured channels, compared to IEEE 802.15.3c channel models. Abbreviation "NA" stands for not available and "-" means not applicable here.

		Metal cabinet			IEEE 802.15.3c	
		Sc. 1	Sc.2	Sc. 3	CM4	CM9
Parameter	Unit	LOS	LOS	NLOS	NLOS	LOS
$P_L(d_0)$	dB	54.711	53.439	54.116	56.1	NA
α		0.02	0.004	0.002	3.74	NA
σ_{PL}	dB	0.39	0.17	0.16	8.6	NA
$\bar{L}$	ns	113.4	158.3	159.1	NA	NA
Λ	1/ns	-	-	-	0.07	0.044
λ	1/ns	0.985	1.037	1.094	1.88	1.01
Γ	ns	-	-	-	19.44	64.2
γ	ns	175.23	197.99	197.93	0.42	61.1
σ_Γ	ns	-	-	-	1.82	2.66
σ_γ	ns	4.90	5.48	4.86	1.88	4.39

The numbers for IEEE 802.15c are taken from [45], which provides models for wideband (9 GHz BW) channels at 60 GHz carrier frequency. The reported parameters are selected from the CM4 and CM9 channel models suggested in this document and obtained from measurements in office areas in a NLOS scenario, and within a kiosk with a LOS scenario, respectively.

It can be seen from Table 4.1 that the measured channel in our tested metal enclosure differs significantly from the typical wireless channels, as expected. The main distinctions are:

1. The path loss exponents are very small in both LOS and NLOS cases.

2. The RDS depends on the metal enclosure volume rather than LOS or NLOS.

3. The channel length is significantly long according to the estimated RDS.

4. The arriving rays do not form clusters.

5. The arrival rate is modeled as a mixed Poisson process.

2.7 Conclusion

In this chapter, a comprehensive channel model (large and small scale) is provided for 60 GHz transmission inside a metal enclosure, which is taken as a generic model for the environment inside a lithography system. The frequency domain channel sounding technique with a resolution of 0.2 ns for resolving multipaths and maximum measurable excess delay of 2400 ns is employed to obtain accurate data. A total BW of 5 GHz with a center frequency of 59.5 GHz is used.

The well-known Saleh-Valenzuela model is used to fit the model parameters, which is widely used and validated in the community. Moreover, channel instances are simulated based on the proposed model parameters and the RDS values are shown to comply, in good extent, with the ones from the measured channel. This can serve as a verification of the suggested model.

Distinguishing features of the considered (rather non-conventional) environment are, first of all, the significantly long channels, in the order of 1 μs, together with very rich multipath reflected from the metal walls (small inter arrival times). A statistical model suggests a single cluster nature of the arriving multipath components and the best model fit is proposed as Gamma and Weibull for different scenarios. Further, we observed relatively sparse channel frequency responses with coherence bandwidths of less than 200 kHz, which relates to the high frequency selectivity of the propagation environment. This is a rare phenomenon that has not been observed in other channels before.

The RMS delay spread is shown to be increased by a 40% when the volume of the metal enclosure is increased 4 times, accordingly, this leads to 40% decrease in the coherence bandwidth in a larger metal box. The accurate relationship between the enclosure volume/geometry, and the channel parameters yet needs to be verified in future work. Even though this could be performed by extensive measurements and processing, other analytical approaches such as ray tracing can be employed for further investigation in such a confined environment. Ray tracing may provide more accurate parameters and enable us to study a variety of scenarios without the hassle of sensitive and complex 60 GHz measurements [52, 53]. In our investigation the direction of the antenna does not impact the channel behavior as the open waveguide shows a negligible directivity. Also, the environment of test is somehow symmetrical around a fixed transmitter as the only reflectors are identical metal walls so the expectation is that the power angle profile is almost uniform for the measured channels. However, the angular profile of the channel is of a great interest for MIMO applications.

The purpose of this work is to replace cable connections inside a metal enclosed mechatronic system to ease the installation and integration of the machine and also to improve the accuracy and reliability. High data rate and low latency are two critical requirements for lithography devices due to the fast control feedback loop. The latency of the wireless system is determined by the long CIR and the long cyclic prefix in case of OFDM systems. Physical remedies include an absorber coating inside the metal enclosure [54], if restrictions on installing such bulky materials inside the mechatronic system are permitted. The measurement results with an absorber coating suggest significant channel shortening.

The proposed statistical channel model helps to understand the challenges related to the wideband wireless communications. We believe that the outcome of this chapter contributes to enrich our understanding of the millimeter-wave propagation properties. However, in signal processing, a simplified version of such comprehensive channel models are used for the sake of conciseness. This approximation of the communication channel keeps the signal model mathematically trackable. In the next chapter we introduce a simplified channel model that we use for the signal processing tasks in the subsequent chapters.

2.A Inverse Filtering and Channel Recovery

In this Appendix, we document the selected process of channel estimation from the observed channel frequency responses. Let $x(t)$ be the transmitted signal, which is impaired by the measurement system and the antennas. The received signal $r(t)$ is given by

$$r(t) = x(t) * h_{tx}(t) * h_{sys}(t) * h(t) * h_{rx}(t), \tag{2.8}$$

where $h_{tx}(t)$ and $h_{rx}(t)$ are the impulse responses of the transmit and receive antennas, $h_{sys}(t)$ is the transfer function of the measurement system and $h(t)$ is the CIR of interest.

The CIR for free space without reflections or obstructions consists of a single LOS path, parametrized by an attenuation and a simple delay equal to the time-of-flight of the signal between the transmit and receive antenna. We can make a recording of the received signal at a known reference distance in free space, and after time gating obtain a reference signal $r_{fl}(t)$, given by

$$r_{fl}(t) \approx x(t) * h_{tx}(t) * h_{sys}(t) * h_{rx}(t)\,, \tag{2.9}$$

so that

$$r(t) \approx r_{fl}(t) * h(t). \tag{2.10}$$

More specifically, $r_{fl}(t)$ in (2.9) absorbs the effect of the antennas and the system (this is not entirely accurate as the directionality of the antennas is ignored). The CIR is obtained from (2.10) via inverse filtering. Equivalently, in frequency domain, we can obtain the channel frequency response $H(f)$ by

$$H(f) = \frac{R(f)}{R_{fl}(f)}\,. \tag{2.11}$$

The CIR is then obtained by taking the (windowed) IFFT of $H(f)$ and correction for the delay and attenuation (normalization).

We have obtained a reference LOS signal $r_{fl}(t)$ by placing the transmitter and receiver at a distance of 25 cm outside the metal cabinet (free space). The LOS path was retrieved by time gating the measured signal and truncating it after 50 ns, so as to remove noise and multipaths beyond the direct line of sight.

Chapter 3

Preliminaries on Wireless Channel Estimation and Equalization

3.1 Introduction

Based on the results from Chapter 2, the long and fading-prone wireless channel within a metal enclosure requires an equalization treatment to allow a reliable communication link. In this chapter, the most pertinent equalization techniques for fading channels are reviewed, in the context of linear processing for the single-input single-output (SISO) point to point (single user) wireless system and a brief comparison is given for complexity and latency of different methods. This is the first step towards the feasibility study of the desired wireless link for the mechatronic system of interest, which was introduced in Chapter 2. We aim to identify the existing equalization techniques and their limitations for this purpose. In this thesis, we consider a linear system model and a wireless channel with finite elements.

3.2 Wireless Channel Model

In general, a wireless communication channel is characterized by different factors and phenomena including the path loss, shadowing and small scale fluctuations due to fading. Path loss is a result of wave attenuation and is proportional to the frequency and inverse of the distance between the transmitter and receiver. Shadowing or large scale fading causes random variations due to the blockage from objects in the signal path. These variations are also caused by changes in reflecting surfaces

and scattering objects [19]. The large scale effects are mostly studied in wide-range transmission systems like cellular communications for budget calculation. In short-range transmission we are only interested in the small scale fading properties. Considering this simplification, a wireless channel is expressed in terms of multipath delays and the corresponding fading gains for a wideband communication system as

$$h_a(t) = \sum_{l=0}^{L-1} a_l e^{-j\phi_l} \delta(t - \tau_l)\,, \tag{3.1}$$

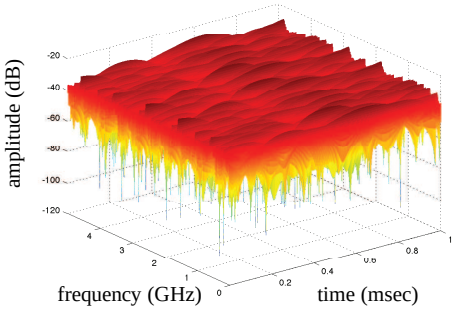
where $h_a(t)$ is the impulse response of the physical channel and in general is a function of time and delay (excitation time), however, here we consider channels that are time-invariant over each transmission block i.e., the channel impulse response for a block is given by (3.1). The number of paths is denoted as L, $\delta(t)$ is the Dirac delta function and (a_l, ϕ_l, τ_l) corresponds to the triple of channel gain (determined using the time decay constant), phase and delay (related to the ray arrival time) for the lth received path. The phase of the lth resolvable delay path is given by $\phi_l = 2\pi f_c \tau_l - \phi_{D_l}$ which is a function of Doppler phase ϕ_{D_l} and carrier frequency of f_c. In fact, when $f_c\tau_l \gg 1$ then a small change in the path delay τ_l can lead to a large phase change in the lth multipath component and this leads to a constructive and destructive addition of multipath components. This phenomenon is called fading and it has a random nature due to the randomly changing parameters a_l, ϕ_l and τ_l. The model in (3.1) can be simplified to a statistical model of

$$h_a(t) = \sum_{l=0}^{L-1} \alpha_l \delta(t - \tau_l)\,, \tag{3.2}$$

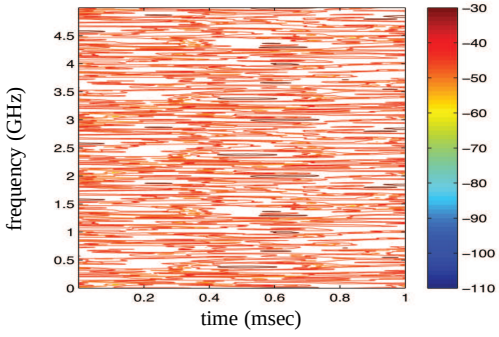
by replacing $a_l e^{-j\phi_l}$ with a complex random variable α_l which represents the fading complex gain of the lth path. Different models are possible for the distribution of the $\{\alpha_l\}$ and $\{\tau_l\}$ variables, which highly depend on the environment. For the $\{\alpha_l\}$, Nakagami distributions and lognormal distributions are the most accepted ones to fit the measurement data in different environments for $\{\alpha_l\}$. A Poisson process is considered to be a reasonable model for excess delay parameters $\{\tau_l\}$ [19, 55]. In Chapter 2 of this thesis these parameters were estimated and validated, for an extremely reflective environment with metal walls, based on acquired experimental data. This is a fundamental step towards a realistic system design for the application of interest.

Simulations of a Wireless Channel

Once the stochastic properties of the wireless channel are known for a particular environment, different realizations of the random channel can be generated using computer simulations. The simulated channel is further used to evaluate the performance of the designed wireless system and plays an important role as a test bed for comparison and verification of algorithms at different levels. A wireless channel simulator needs to model both the time selectivity of the channel due to the Doppler spread, and the frequency selectivity (time spreading) due to the fading. The time dynamics (fading) of a wireless channel is an important feature which is not consider in the model in (3.2). The most common model used in the literature, to simulate the fading process is the Clarke's model [56].



(a) 3D plot of channel gains in dB.



(b) 2D contour plot of channel gains.

Figure 3.1: Simulated channel gains for slowly time-variant and dispersive channel over frequency and time using the tapped delay line (TDL) method, with sampling period of $T_s = 0.2$ ns.

Here, an example of a 3D representation of the channel gains showing both time

and frequency domain fluctuations is given in Fig. 3.2. The channel has been simulated based on the tap delay line (TDL) set up where each tap (path) is constructed using the estimated time decay constants from Chapter 2 with Jakes' Doppler spectrum [56]. This is customized for a slowly time-variant channel of interest in the metal enclosure. Fig. 3.1a shows the channel gains in dB in both time and frequency axis and the corresponding Fig. 3.1b maps the projection of these channel gains on a 2D surface. The time difference in Fig. 3.1b between two horizontal points with relatively different amplitude reveals the coherence time of the channel and respectively the coherence bandwidth can be determined finding two consecutive uncorrelated points in the frequency (vertical) axis.

Reception Model

In the absence of noise, the received signal $y_a(t)$ after convolution with the linear channel introduced in (3.2) can be expressed as

$$y_a(t) = \sum_{l=0}^{L-1} \alpha_l x_a(t - \tau_l)\,, \tag{3.3}$$

where $x_a(t)$ is the analogue transmit signal which can be a simple quadrature amplitude modulation (QAM) signal or a signal generated by more complicated modulation techniques like multicarrier code division multiple access (CDMA), orthogonal frequency division multiplexing (OFDM), etc. Digital signal processing which is the focus of this thesis is performed after sampling. Assuming the $\{\tau_l\}$ can be interpolated by integers between 0 and $L-1$, the received sequence for a discrete time signal model is given by

$$y(n) = \sum_{l=0}^{L-1} \alpha_l x(n-l), \tag{3.4}$$

in the absence of noise. For wideband systems, we usually consider the channels static over each data block (i.e., block time-invariant)

The signal model in (3.4) represents a finite impulse response (FIR) filter with a finite complex impulse response $\mathbf{h} = [h(0), h(1), \cdots, h(L-1)]^T$ and order L which is defined relative to the symbol time such that each element in $\mathbf{h}$ corresponds to an α_l at the sampling moment. The received signal is thus expressed as a convolution of the channel impulse response with the transmitted signal $x(n)$ plus

noise samples denoted by $e(n)$, i.e., given by

$$y(n) = \sum_{l=0}^{L-1} h(l)x(n-l) + e(n), \tag{3.5}$$

The channel gain $h(l)$ in (3.5) caters for the wireless channel effects including the pathloss and shadowing. There are more basic channel models that are used in array processing which assume far field reception, so the channel can be parameterized only by the direction of signal arrivals. The former model is considered in this thesis.

A major consequence of a wideband (convolutive) channel is the inter-symbol interference (ISI) due to the short symbol periods (transmitted pulses) with respect to the channel length. This is equivalent to having a wideband signal with respect to the channel coherence bandwidth. To successfully detect the received sequence, the distractive effect of the wireless fading channel needs to be canceled first. This process is referred to as channel equalization.

3.3 Channel Estimation

The availability of channel state information (at the transmitter and/or receiver) is often required for the design of signal processing algorithms. In fact, to fully compensate for the inter-symbol interference (ISI) phenomenon, the receiver and/or transmitter need to know the instantaneous channel impulse response. Thus, channel estimation is inevitable and therefore, a brief introduction to common channel estimation techniques is given here.

A common way to perform the channel estimation is to transmit a number of known symbols (pilot/training sequence) and measure the time domain channel impulse response based on the received sequence. In contrast, *blind processing* techniques look at the inherent structure of the data for system identification. The channel impulse response is changed over time as discussed before and needs to be estimated each time the channel decorrelates. This is the result of relative movement between the transmitter and receiver (Doppler).

Channel estimation can be done either in time domain or in frequency domain. Here, we look at the conventional training based model for the impulse response estimation. A sequence of known training symbols of length k is sent to the receiver and in the simplest scenario, the standard least square (LS) algorithm is used to estimate the channel based on the received vector without assuming any

properties for the noise or signal. Generally, the sequence of known symbols $\mathbf{x} = [x(0), x(1), x(2), \cdots, x(K-1)]^T$ is inserted periodically in a block of data symbols where the time between two consecutive pilot sequences is determind by the coherence time of the wireless channel. The transmit signal after processing is denoted by $\mathbf{x}$, and the unprocessed data sequence is referred to as $\mathbf{d}$. Depending on the transmitter, these two could be equal or not. If there is no processing involved in transmission then the transmitter transfer matrix is an identity and $\mathbf{x} = \mathbf{d}$.

The received signal is the convolution of $\mathbf{x}$ with the wireless channel $\mathbf{h}$. The convolution operation can be represented by a matrix multiplication. To this end, the data vector $\mathbf{x}$ is converted into a Toeplitz matrix $\mathbf{X} \in \mathbb{C}^{(K+L-1)\times L}$ and the received sequence is given by

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{e}\,, \tag{3.6}$$

where $\mathbf{X}$ is

$$\mathbf{X} = \begin{bmatrix} x(0) & x(-1) & x(-2) & \cdots & x(-L+1) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ x(L-1) & x(L-2) & x(L-3) & \cdots & x(0) \\ x(L) & x(L-1) & x(L-2) & \cdots & x(1) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ x(K-2) & x(K-3) & x(K-4) & \cdots & x(K-L-1) \\ x(K-1) & x(K-2) & x(K-3) & \cdots & x(K-L) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ x(K+L-2) & x(K+L-2) & x(K+L-3) & \cdots & x(K-1) \end{bmatrix}. \tag{3.7}$$

The first L samples of $\mathbf{y} = [x(0), \cdots, x(L-1), \cdots, x(K+L-2)]^T$ are the transient part of the convolution since they are contaminated with previous unknown symbols ($[x(-1), \cdots, x(-L+1)]$) and the last L symbols are also affected by unknown data so the valid output vector has $K-L-1$ elements which corresponds to the middle part of the $\mathbf{X}$, denoted by $\hat{\mathbf{y}} = [y(L), y(L+1), \cdots, y(K-2)]^T$.

In order to estimate the channel impulse response, only the known symbols are taken so the matrix $\hat{\mathbf{X}}$ is formed by removing the first and last L rows in $\mathbf{X}$. The channel can then be estimated in least square (LS) sense by

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \|\hat{\mathbf{y}} - \hat{\mathbf{X}}\mathbf{h}\|^2\,. \tag{3.8}$$

The solution to the LS optimization problem in (3.8) is given by

$$\hat{\mathbf{h}} = \hat{\mathbf{X}}^{\dagger}\hat{\mathbf{y}} = (\hat{\mathbf{X}}^{H}\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}^{H}\hat{\mathbf{y}}\,, \tag{3.9}$$

and requires a cubic complexity $O(L^3)$ to compute, with respect to the channel length L. Note that L is itself relative to the symbol time T_s (the inverse of the signal bandwidth). For instance, in a channel lasting over $200\,ns$ and $T_s = 0.1\,ns$, the channel length can be expressed as $L = \frac{200}{0.1} = 2000$. Therefore, for a single transmit antenna, the complexity of the matrix inversion is in order of 10^9 multiplier - accumulator (MAC) operations. In fact the inversion operation in (3.9) can be calculated with complexity $O(L^2)$ because of the Toeplitz structure of $\mathbf{X}$ once a block time-invariant channel model is assumed. Moreover, to have a full column rank matrix for inversion in (3.8), the number of training symbols needs to be sufficient so $K \geq 2L + 1$. The optimal design for the training sequence can be derived by solving an optimization problem when the mean square error (MSE) of the channel estimation is minimized under a total training power constraint of the system [57]. There are alternative less complex frequency domain channel estimation techniques that are based on diagonalization of the matrix $\mathbf{X}$ by means of the Fourier transform. Similar techniques are discussed next in the context of frequency domain equalization.

3.4 Wireless Channel Equalization

Optimal and Suboptimal Receivers

An optimal receiver for a digital communication system employs maximum likelihood sequence estimation (MLSE) for detecting the information sequence from the samples of the received symbols. The MLSE for a channel with ISI has a computational complexity that grows exponentially with the length of the wireless channel, i.e., for a channel of length L and a symbol alphabet with M members, the Viterbi algorithm computes M^{L+1} matrices for each new received symbol. This is simply impractical in real systems with a limited latency requirement. Instead, suboptimal receivers employ linear or nonlinear equalizers to remove or reduce the effect of the channel on the transmitted information. There are different types of equalizers, each with different performance and computational complexity, they can be listed as:

- Linear equalizers including zero forcing (ZF), minimum mean square error (MMSE) and adaptive approaches to update a linear equalizer like least mean

square (LMS) or recursive least square (RLS) algorithms.

- Nonlinear equalizers including the decision feedback equalizer (DFE) and maximum a posterior (MAP) equalizer with a BCJR algorithm or Turbo decoding which treats the channel as a convolutional code.

We focus on linear equalizers here, where the first choice of linear equalizers try to minimize the least square error (LSE) cost function with no probabilistic assumptions, which leads to the best linear unbiased estimator (BLUE) when the noise is white. The LSE criterion gives the minimum-variance unbiased (MVU) estimator when the noise is assumed to be white and Gaussian. However, the performance of such equalizers (e.g., the ZF equalizer) highly depends on noise characteristics. Alternatively, the MSE cost function can be minimized taking into the account the stochastic property of the noise which trades off the noise, and ISI at the output of the equalizer.

In contrast, adaptive algorithms corresponding to both criteria are used for time-variant channels, where filters are trained periodically over time to track the channel state and updating algorithms are used, like RLS and LMS [58]. Frequency domain equalizer (FDE) types [59] are an alternative low-complexity approach to ISI mitigation, where the digital transmission is carried out block-wise and equalization relies on DFT operations which makes them closely related to OFDM systems.

3.4.1 Time Domain Equalization

Time domain equalization (TDE) removes the channel effect by filtering the received signal in time domain with an FIR filter (inverse filtering). This is implementable as serial equalization as well as block equalization. The main difference in the problem formulation appears in the representation of the equalization parameters which are given as a vector $\mathbf{w}$ for a serial equalizer while this is a matrix $\mathbf{W}$ for a block equalizer.

Serial Equalizers

The received sequence $\mathbf{y}$ passes through the equalizer filter with impulse response $\mathbf{w}$ to reconstruct the desired original sequence $\mathbf{x}$. So we have $\hat{\mathbf{x}} = \mathbf{Y}\mathbf{w}$, where $\mathbf{Y} \in \mathbb{C}^{P\times(2L+1)}$ the received symbols is collected in a Toeplitz matrix similar to $\mathbf{X}$ in (3.7). For serial equalizers, the LS criterion minimizes the squared error between $\hat{\mathbf{x}}$ and the desired sequence which leads to a ZF equalizer. Generally,

the filter coefficients are derived directly based on the known training sequence $\mathbf{x}$ introduced in Section 3.3, so the least square filtering is given by

$$\mathbf{w}_{ZF} = \arg\min_{\mathbf{w}} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \arg\min_{\mathbf{w}} ||\mathbf{x} - \mathbf{Y}\mathbf{w}||^2 = (\mathbf{Y}^H\mathbf{Y})^{-1}\mathbf{Y}^H\,\mathbf{x}\,. \quad (3.10)$$

The same approach can be followed to derive the MMSE equalizer, but this time the stochastic mean of the squared error is minimized, so the noise distribution is also taken into account. Accordingly the MMSE equalizer is derived as

$$\mathbf{w}_{MMSE} = (\mathbf{Y}^H\mathbf{Y} + \frac{1}{\gamma}\,\mathbf{I}_{2L+1})^{-1}\mathbf{Y}^H\,\mathbf{x}\,, \quad (3.11)$$

where $\frac{\sigma_x^2}{\sigma_e^2}$ is the signal to noise ratio (SNR). Here, it is assumed that the transmit signal and noise are zero mean.

Both of these equalizers have the same number of taps $(2L+1)$, where L is the channel length, so they introduce the same delay to the system. The matrix inversion operation in (3.10) and (3.11) has a complexity of order $O(L^3)$ and also each received symbol needs to be multiplied with $(2L+1)$ filter coefficients. As a result, for the data sequence of length N, we require $N(2L+1)$ flops to perform the equalization.

Block Linear Equalizers

For block linear equalizer, a block of data is transmitted followed by a guard interval. The receiver decodes the full block after collecting it. The ISI is assumed to be within a data block once a long enough time guard is inserted between two consecutive transmissions. Note that the latency of a block transmission system depends on the block length unlike the serial equalizer. The channel can be explicitly estimated at the receiver side as explained in Section 3.3, and the channel coefficients can be used directly to update the equalizer taps.

The guard interval is inserted between different data blocks to isolate the ISI in each block and is selected to be a cyclic prefix of data or just a stretch of zeros of at least length L. Subsequently, the received vector is given by

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{e}\,, \quad (3.12)$$

where $\mathbf{G} \in \mathbb{C}^{N\times(N+L)}$ is a Toeplitz convolution matrix defined by,

$$\mathbf{G} = \begin{bmatrix} h(L-1) & \cdots & h(0) & 0 & 0 & \cdots & 0 \\ 0 & h(L-1) & \cdots & h(0) & 0 & \cdots & 0 \\ 0 & 0 & h(L-1) & \cdots & h(0) & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & h(L-1) & \cdots & h(0) \end{bmatrix}. \quad (3.13)$$

The transmit sequence is denoted by the vector $\mathbf{x} = [x(-L), x(-L+1), \cdots, x(0), x(1), \cdots, x(N-1)]^T$ after appending the time guard. The vector $\mathbf{y} = [y(0), y(1), \cdots, y(N-1)]^T$ includes the received sequence after the wireless channel and after the removal of time guard (the first L samples). In turn, $\mathbf{e}$ is a vector containing samples of zero-mean complex Gaussian noise. For a block time-invariant wireless channel, $\mathbf{G}$ is a banded Toeplitz matrix and (3.12) models the single-input and single-output block transmission communication system in a Gaussian channel.

For a block time-invariant channel model, the equalization matrix $\mathbf{W}$ is a Toeplitz matrix, containing the coefficients of a filter $\mathbf{w}$. Again, both ZF and MMSE equalizers can be applied in block mode. When the channel is known, the ZF equalizer can be derived by forcing the cross correlation between the error sequence $\boldsymbol{\varepsilon} = \mathbf{x} - \hat{\mathbf{x}}$ and the desired information sequence to be zero where the equalized signal is

$$\hat{\mathbf{x}} = \mathbf{W}\hat{\mathbf{G}}\mathbf{x} + \mathbf{W}\mathbf{e}\,, \tag{3.14}$$

and the Toeplitz version of the estimated channel vector is denoted as $\hat{\mathbf{G}}$. The cross correlation of the error and the signal is given by

$$\mathbb{E}\{\boldsymbol{\varepsilon}\mathbf{x}^H\} = \mathbb{E}\{(\mathbf{x} - \hat{\mathbf{x}})\mathbf{x}^H\} = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\} - \mathbb{E}\{\mathbf{W}\hat{\mathbf{G}}\mathbf{x}\mathbf{x}^H\}\,, \tag{3.15}$$

assuming that the noise and data are uncorrelated.

To force (3.15) equal to zero, $\mathbf{W}$ needs to be the inverse of the channel matrix

$$\mathbf{W}_{ZF} = \hat{\mathbf{G}}^{\dagger} = (\hat{\mathbf{G}}^H\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}^H\,. \tag{3.16}$$

The noise enhancement is the price we pay for using the ZF equalizer. The SNR scales inversely with the noise variance and to keep the performance unaltered, the SNR needs to be increased according to the noise enhancement factor [60].

For the MMSE equalizer, the square error between the transmitted sequence and the convolution of the received sequence with the equalizer filter is minimized. Then the objective function is given by

$$\mathbb{E}\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^H\} = {\sigma_x}^2(\mathbf{W}\hat{\mathbf{G}} - \mathbf{I}_{N+L})(\mathbf{W}\hat{\mathbf{G}} - \mathbf{I}_{N+L})^H + {\sigma_e}^2\mathbf{W}\mathbf{W}^H. \tag{3.17}$$

The derivative of the expression in (3.17) is put to zero to obtain the optimal solution of (3.17), which follows

$$\mathbf{W}_{MMSE} = (\hat{\mathbf{G}}^H\hat{\mathbf{G}} + \frac{1}{\bar{\gamma}}\mathbf{I}_{N+L})^{-1}\hat{\mathbf{G}}^H\,. \tag{3.18}$$

Given a full matrix $\mathbf{W}$, the multiplication of $\mathbf{Wy}$ requires N^2 operations. However, there exist more efficient algorithms to perform the block equalization by exploiting the Toeplitz structure on $\mathbf{W}$. Note that to achieve a better performance, fractionally spaced sampling is required to cancel out the ISI, but this increases the complexity linearly by oversampling factor. Moreover, for gigabyte transmissions, the bandwidth of ADC/DAC is a serious issue even at a Nyquist sampling rate, so to do the oversampling we need to reduce the data rate to cope with limited bandwidth of ADC/DACs.

3.4.2 Frequency Domain Equalization

In a block transmission mode, the time guard block was until now used to insert zero symbols in the data stream. Alternatively, each block can be circularly extended by inserting the repetition of at least L of its last symbols as a time guard (cyclic prefix). This overhead data transmission introduces an elegant mathematical property of periodicity which can be exploited to form structured circular channel matrices. This results in a significant complexity reduction which simplifies the equalization process which is exploited in OFDM systems and frequency domain equalization [61].

The received signal vector $\mathbf{y}$ satisfies $\mathbf{y} = \mathbf{Gx} + \mathbf{e}$ as introduced in (3.12). By inserting a cyclic prefix in $\mathbf{x}$, we can write (3.12) in terms of the original data sequence excluding the cyclic prefix, $\mathbf{x}' = [x(0), x(1), \cdots, x(N-1)]$, as

$$\mathbf{y} = \mathbf{G}_c\mathbf{x}' + \mathbf{e}\,, \tag{3.19}$$

where a circulant matrix $\mathbf{G}_c \in \mathbb{C}^{N\times N}$ is

$$\mathbf{G}_c = \begin{bmatrix} h(0) & 0 & \cdots & h(L-1) & \cdots & h(1) \\ \vdots & h(0) & 0 & 0 & \ddots & \vdots \\ h(L-1) & \cdots & h(0) & 0 & 0 & h(L-1) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & h(L-1) & \cdots & h(1) & h(0) \end{bmatrix}. \tag{3.20}$$

In FDE, the noisy received signal is first transformed by a DFT matrix ($\mathbf{F}$) at the receiver and is transformed back by multiplying an IDFT matrix ($\mathbf{F}^H$) after the equalization process. The original data vector is denoted by $\mathbf{x}'$ and its estimate at the receiver is shown as $\hat{\mathbf{x}}$. In turn, the representation of $\mathbf{y}$ is given by

$$\hat{\mathbf{x}} \;=\; \mathbf{F}^H\mathbf{WFy} = \mathbf{F}^H\mathbf{WFG}_c\mathbf{x}' + \mathbf{F}^H\mathbf{Wv}\,, \tag{3.21}$$

where $\mathbf{v}$ is the Fourier transform of the noise vector. Finally, the data vector $\mathbf{x}'$ can be written as $\mathbf{F}^H\mathbf{F}\mathbf{x}'$, then (3.21) is expressed as

$$\hat{\mathbf{x}} = \underbrace{\mathbf{F}^H\mathbf{W}}_{\mathbf{R}}\mathbf{Q}\mathbf{F}\mathbf{x}' + \mathbf{v}' \,, \tag{3.22}$$

where $\mathbf{Q} = \mathbf{F}\mathbf{G}_c\mathbf{F}^H$ is diagonal for the considered block time-invariant channel. This exploits the properties of circulant matrices. An equivalent noise vector after multiplication with frequency domain equalizer $\mathbf{R}$ is denoted as $\mathbf{v}'$. The diagonal entries of $\mathbf{Q}$ correspond to the Fourier transform of the channel i.e., $Q(k) = \sum_{l=0}^{L-1} h(l)\, e^{-j2\pi kl/N}$.

Both ZF and MMSE criteria can be used, similar to TDE, to derive the FDE by minimizing the mean square error between the decision vector $\hat{\mathbf{x}}$ and the original signal $\mathbf{x}'$, where the error is expressed as

$$\boldsymbol{\varepsilon} = (\mathbf{R}\mathbf{Q}\mathbf{F} - \mathbf{I}_N)\mathbf{x}' + \mathbf{v}' \,. \tag{3.23}$$

The required matrix inversion takes N flops to calculate since the matrix $\mathbf{Q}$ to be inverted is diagonal, versus N^2 flops in TDE. The complexity of the DFT implemented as a fast Fourier transform (FFT) is $O(N\log(N))$ operations and to apply the equalization, N MACs are required unlike the block linear TDE which spends N^2 MACs.

3.4.3 Comparison of Computational Complexity

To conclude the equalization discussion, in Table 3.1 the complexity of the equalization algorithms is summarized in terms of the required flops for N samples in one block (or symbols if oversampling is not the case). Note that the latency is considered as the time between the moment that data is received until it is available for detection, excluding the training (channel estimation) time.

As shown in Table 3.1, the complexity of the time domain block equalization varies between $O(N)$ and $O(N^2)$ i.e., depending on the structure of matrix $\mathbf{G}$. The inversion of a symmetric or unsymmetric Toeplitz matrix can be performed in $O(N^2)$ flops, however the matrix needs to be sufficiently well conditioned in unsymmetric mode. Banded matrices can also save in terms of computational complexity using a Cholesky decomposition. The computational complexity can be reduced to order $O(N)$ if the bandwidth of the banded matrix is small enough and for a regular Toeplitz matrix it is of order $O(N^2)$ [56].

If a continuos transmission system is considered and the channel is highly time-variant then the training sequence needs to be sent frequently so the equalizer can

Table 3.1: Comparison between the complexity and convergence time of different equalization algorithms. The length of one block is N, T_s is the symbol time and L is the channel length.

	algorithm	order of flops per block	latency
block	ZF	$O(N) - O(N^2)$	NT_s
	MMSE	$O(N) - O(N^2)$	NT_s
	FDE	$O(N)$	NT_s
serial	ZF	$O(NL^2)$	$(2L+1)T_s$
	MMSE	$O(NL^2)$	$(2L+1)T_s$
adaptive	RLS	$O(NL^2)$	$\sim (2L+1)T_s$
	LMS	$O(NL)$	$\sim 10LT_s$
	fast Kalman DFE	$O(NL)$	$(2L+1)T_s$
	square root LS DFE	$O(NL^2)$	$(2L+1)T_s$

track the channel variation fast enough. In this case, direct matrix inversion is not an efficient way to update the equalizer coefficients and adaptive algorithms are employed to update the coefficients. RLS and LMS are among the most common adaptive algorithms. Their complexity is listed in Table 3.1. In order to learn the channel coefficient fast enough, the training process needs to converge before the channel changes, in other words the convergence time of the algorithm to the optimal coefficients needs to be smaller than the channel coherence time [19]. Nonlinear equalizers are also considered in Table 3.1 i.e., decision feedback equalizers (DFE) and square root LS DFE. These commonly use a feedback loop to enhance the estimation quality.

Finally, the elements of the original data $\mathbf{d}$ that we are trying to estimate are limited to the discrete values of the digital modulation scheme; in general a vector $\mathbf{d}$ of size $N \times 1$ can takes 2^{MN} values where 2^M is the alphabet size of the modulation format. This indicates a detection problem rather than an estimation problem and the maximum likelihood detector is proved to be the optimal detector. Unfortunately it is not computationally efficient or even feasible to perform a Viterbi search to find the optimal sequence. maximum likelihood detectors with low complexity have been proposed during the past decades, e.g., the received sequence is preprocessed prior to the maximum likelihood detector by passing it through a linear equalizer or a decision feedback equalizer to reduce the channel length to a desired channel impulse response of L_0. This reduces the computational complexity of the Viterbi algorithm to M^{L-L_0} [58].

3.5 Conclusion

In this chapter, we introduced the wireless channel model that is considered in the subsequent chapters to design the signal processing algorithms for a wideband communication system. This is a simplified model compared to the one introduced in Chapter 2. The knowledge of the instantaneous complex channel gains is acquired, by transmitting a short training sequence (known data). Common criteria for receiver design and channel equalization were reviewed with respect to their complexity and latency and the importance and effectiveness of the linear equalizers were highlighted in this context.

Note that, even though the equalization process was studied as part of the receiving process in this chapter, the same linear processing can be performed at the transmitter side to pre-equalize the wireless channel. This is beneficial when there is more processing power available at the transmitter rather than the receiver, e.g, in cellular systems. OFDM is an example of transferring a part of the processing to the transmit side, i.e., by performing the IDFT before transmission of the signal, and later applying the FDE at the receiver. An example of an OFDM design for a wideband system is presented in the next chapter.

Chapter 4

Wideband System Design Example

4.1 Introduction

In high-speed communications, the propagation environment is the main source of disturbance of the original signal and the main task of the transmitter and receiver, respectively, is to prepare and recover the signal before and after passing it through the wireless channel. In general, flat-fading channels (narrowband) are much easier to handle, while wideband channels introduce inter-symbol interference (ISI) which can be regarded as noise that is correlated with the desired signal and is difficult to handle. The main strategy for successful transmission in wideband systems is to translate the channel to a set of parallel narrowband subchannels. In this way, the effective existing transmission and reception techniques for narrowband communications can be applied to each sub channel.

Orthogonal frequency division multiplexing (OFDM) is an effective modulation scheme, and is indeed considered for most of the existing wideband wireless standards including WiMAX, LTE, WiFi, and also for the upcoming new standard for 60 GHz WPAN, i.e., IEEE 802.15.3c. In general, the frequency band is divided into several subcarriers such that each subcarrier experiences a flat-fading channel. In this chapter we introduce, in more detail, the lithography system of interest and we design an OFDM system that can meet the requirements of the system. Basically, this is to implement a wireless link inside a tight feedback control loop.

As introduced in Chapter 1, the main characterization of the communication system of interest for the lithography machine is high peak data rate and low latency. The sensors send a data block every T seconds, but only a small portion of

this interval is dedicated to transmission. Most of the time is reserved for control processing so $T = kT_t$ where T_t is the actual transmission period. The processing time at the receiver is T_{p_r} which is used for demodulation, decoding etc. and defines the processing delay at the receiver. This setting is illustrated in in Fig. 4.1. The peak data rate is given by the ratio of the number of transmitted bits over T_t.

All information bits need to be sent in the transmission period T_t and processed in a very limited processing time (T_{p_r}) according to the system requirements. In fact, latency is a critical measure for the system performance in this kind of application and it is required to be in the order of micro-seconds which is far less than the millisecond requirements on, e.g., video applications. Latancy is defined as the time difference between the moment that the data is ready to be sent at the transmitter and once it is available as data symbols at the receiver. However, some signal processing can be performed during the non-transmitting time slot of duration T_{p_t}, e.g. training symbols transmission and channel estimation.

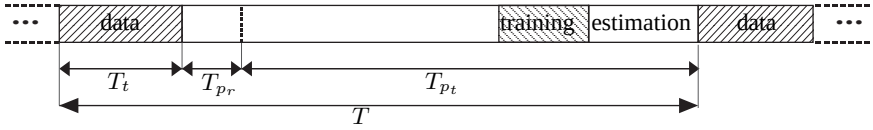


Figure 4.1: Time diagram for data transmission and processing. The time interval T represents the time between the two data transmission periods where T_t is the available time for sensors to send information and T_{p_r} is the available processing time at the receiver. Training symbols are sent before the next transmission and the estimation of the channel coefficients is done during the specified estimation time.

The receiver (control unit) is located at a fixed position so the speed of the moving platform defines the time variation of the wireless channel. By assuming a relative speed of $10\,\frac{m}{s}$ and a frame time of $T = 50\,\mu$s, the transmitter has moved by $0.5\,mm$ between two transmission periods. This is less than 10% of the wavelength at a carrier frequency of 60 GHz. Hence, the channel estimation can be performed after the channel decorrelates which is approximately every 5 frames at 60 GHz and every 100 frames at a 3 GHz carrier frequency which indicates a slowly time variant channel.

This leads to the choice of a frequency selective block time-invariant channel model and requires a careful choice of the equalizer to combat the fading without exaggerating the latency of the system. Moreover, the system uses a wideband burst transmission mode which accumulates the data symbols in a block before transmission rather than continuous transmission.

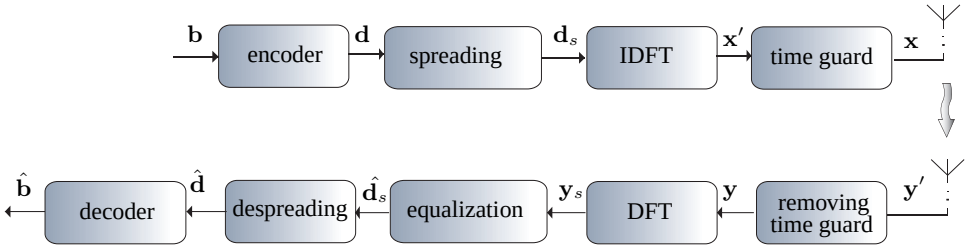


Figure 4.2: A typical block diagram of the transmitter and receiver for a point-to-point wideband communication system with frequency domain equalization at the receiver.

4.2 Block Transmission Model

Systems for wideband transmission can generally be divided into two possible settings: 1) multicarrier transmission including OFDM and multicarrier code division multiple access (multi carrier-CDMA), or 2) single carrier transmission. Note that single carrier transmission is commonly performed with frequency domain equalization for wideband systems to eliminate the frequency selectivity of the wideband channel. Frequency domain equalization is an alternative equalization technique to the classical time domain equalization, which was initially developed for ISI mitigation in wireline channels like dial up modems and in general for narrowband systems [62].

A common diagram for a block-wise wideband transmission system is illustrated in Fig. 4.2. In the transmitter, after mapping the data bits vector $\mathbf{b}$ into some form of digital modulation like quadratic phase shift keying (QPSK) or QAM, the complex symbols are grouped into data blocks of size N so $\mathbf{d} = [d(0), d(1), \cdots, d(N-1)]^T$, $d \in \mathbb{C}^{2^M}$, where M is the alphabet size of the modulation scheme. Each block is linearly transformed by a code spreading matrix $\mathbf{C}$ and the spread data signal is given by $\mathbf{d}_s = \mathbf{C}\mathbf{d}$, where the spreading code selection defines the choice of transmission as follows:

- single carrier with frequency domain equalization: $\mathbf{C} = \mathbf{F} \in \mathbb{C}^{N \times N}$ is a Fourier matrix.
- OFDM: $\mathbf{C} = \mathbf{I}_N \in \mathbb{R}^{N \times N}$ is an identity matrix.
- multi carrier-CDMA: $\mathbf{C} = \mathbf{K} \in \mathbb{C}^{N \times N}$ is a Walsh-Hadamard matrix.

Basically, any unitary matrix satisfying $\mathbf{C}\mathbf{C}^H = \mathbf{I}_N$ can be used for the purpose of code spreading. Then, the spread signal is modulated onto N narrowband subcarriers in a block which is denoted by $\mathbf{x}'$. Based on the model in Fig. 4.2, the transmit

signal before adding the time guard is given by

$$\mathbf{x}' = \mathbf{F}^H \mathbf{C}\mathbf{d}. \tag{4.1}$$

The final transmit signal $\mathbf{x}$ is a concatenation of vector $\mathbf{x}'$ and the time guard sequence which is discussed next for an OFDM system. In turn, the received signal after equalization and despreading is

$$\hat{\mathbf{d}} = \mathbf{C}^H \mathbf{W}\mathbf{F}\mathbf{G}_c\mathbf{F}^H\mathbf{C}\mathbf{d} + \mathbf{e}, \tag{4.2}$$

where $\mathbf{G}_c$ and $\mathbf{W}$ are the circulant channel and equalization matrix which are defined in equations (3.20) and (3.14), respectively, and $\mathbf{e}$ is the additive noise. We know that a circulant matrix is orthogonalized by a Fourier transform, so $\mathbf{F}\mathbf{G}_c\mathbf{F}^H$ is an orthogonal matrix containing the Fourier transform of the channel impulse response and the equalization matrix inverses the channel effect, ideally. Therefore, the received signal is the transmit signal plus zero mean white Gaussian noise and the optimal receiver is the matched filter in this case.

4.3 Orthogonal Frequency Division Multiplexing (OFDM)

OFDM modulation is a multi-carrier transmission scheme with overlapping subcarriers which increases the spectral efficiency and has become very popular because of its ISI rejection property in frequency selective wireless channels. The frequency band is divided into several subcarriers while each subcarrier experiences a flat fading channel [19].

In general, an OFDM system uses a cyclic prefix to isolate ISI within each block. The cyclic prefix is the time guard which is created by concatenating the last N_{cp} samples to the beginning of the block. The length of the cyclic prefix is commonly chosen to be longer than two times the rms delay spread of the wireless channel. Therefore, in OFDM, a long channel can waste bandwidth and limit the data rate. There are techniques to reduce the channel length including a channel shortening filter which uses the same approach as time domain equalization. The channel shortening filter is designed by minimizing the mean square error between the desired channel impulse response and the one after filtering [63]. Nevertheless, deep fading in subchannels results in erroneous decisions on corresponding modulated symbols in those OFDM subcarriers. If some consecutive subchannels are in deep fade, the channel is useless in a portion of the frequency band. In other words, OFDM does not benefit from multipath diversity. There are some classic solutions for this issue which can be classified as:

Interleaving can be performed at the receiver to obtain the multipath diversity by diversity combining. At each transmission period in time, all subchannels do not experience the same fading across the band and this can be used to achieve diversity gain by sending the same OFDM symbol over multiple subchannels at the cost of reducing the data rate. The receiver combines the received signal in such a way that the total SNR is increased at the output [19].

Coding is the simplest way to correct for random errors in the received block of data. The combination of coding and interleaving can be used to avoid burst errors in OFDM due to fading. In fact, symbols in time domain are coded and interleaved before IDFT modulation, so those adjacent subcarriers which experience fading over the wireless channel are relocated at the deinterleaver block and decoded afterward. Therefore, burst errors are spread over the received sequence and can be recovered by some form of error correcting codes.

Adaptive modulation is another way to improve the performance of OFDM system by allocating more complicated mapping scheme or more power in subchannels with better SNR and vice versa. This needs explicit channel state information at the transmitter and is an optimal resource allocation technique to maximize the mutual information (capacity) in a communication link.

Zero padding OFDM addresses the diversity issue in OFDM by taking into account the redundant part of the channel. This was first proposed in [64] based on trailing zeros rather than a cyclic prefix to eliminate the inter-block interference. In this approach the appended zeros are not removed as in cyclic prefix but they are used in the equalization process. If the number of appending zero symbols equals the cyclic prefix length then zero padding OFDM has the same spectral efficiency as cyclic prefix OFDM. The price paid here for symbol recovery is the receiver complexity, since the single FFT in cyclic prefix version is replaced by FIR filtering [65].

OFDM suffers from other issues including a large peak-to-average power ratio (PAPR) which seriously limits the efficiency of the power amplifier. There are several methods to combat PAPR, we discuss this in Chapter 6. The other problem with OFDM is the sensitivity to carrier frequency offset. There are literatures dedicated to comparing OFDM to single carrier with frequency domain equalization, and many simulations are presented on the performance of coded and uncoded

OFDM [61]. Next, we discuss a simple OFDM design for the mechatronic system of interest according to the channel specifications that are introduced in Chapter 2.

4.4 OFDM Design Example for the Dispersive Channel

In this section the BER performance of an OFDM system using the measured channel is compared to that using a simulated Rayleigh fading channel to answer to some extent important system design questions such as

- What are suitable modulation schemes and equalization techniques for such dispersive (rich scattering) and extremely long channels?
- Can we benefit from the diversity gain offered by these types of channels with an acceptable computational complexity or should the channel be shortened using e.g. absorbers?
- What is the channel capacity for such highly reflective environments?

For the numerous existing wireless channel models, these questions are well studied. A straightforward approach is to relate the proposed channel model in Chapter 2 to the available models and modify the system design including the modulation, coding and equalization in order to cope with the new circumstances.

Here we design a wideband orthogonal frequency division multiplexing (OFDM) system with a zero forcing (inverse filtering) frequency domain equalization for a lithography system with certain latency and rate requirements. The proposed system design is simulated in Matlab using the measured channel in the 60 GHz band from Chapter 2 within a closed metal cabinet which emulates the environment inside the lithography machine. In contrast to conventional indoor channels at 60 GHz, the channel in the metal enclosure is highly reflective resulting in a rich scattering environment with a significantly large root-mean-square (RMS) delay spread which makes high data-rate communications a challenging task. The bit error rate (BER) performance is evaluated and compared to two long and short simulated Rayleigh channels and also theoretical BER performance for Rayleigh channels. In general the BER performance is almost 3 dB better when more subcarriers are used in one OFDM block as the ratio of cyclic prefix to the information symbols and consequently the power spread is reduced.

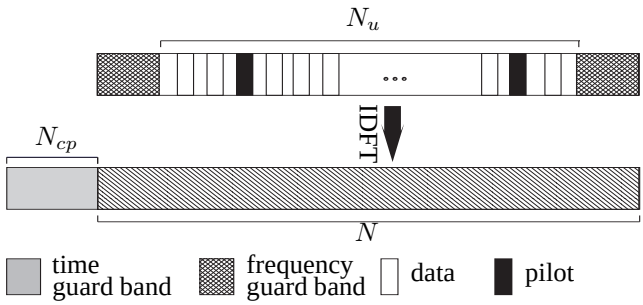


Figure 4.3: A data sequence with frequency guard bands (null subcarriers in frequency domain) and pilot subcarriers is converted to the transmitted time domain OFDM block by taking an inverse discrete Fourier transform (IDFT) and appending the time domain guard (cyclic prefix) to the block.

4.4.1 Design Parameters

The design example is based on a single antenna, standard OFDM modulation without coding (see Fig. 4.3). An OFDM block with bandwidth B_w is split into N subcarriers, consisting of guard bands and N_u 'user' subcarriers (data and pilots). In time domain, the corresponding N samples are augmented with a cyclic prefix of N_{cp} samples. The symbol duration is $T_s = 1/B_w$ and the maximum delay spread is denoted as $t_{\max}$. Straightforward equalization requires that the time duration of the cyclic prefix is larger than the length of the wireless channel:

$$N_{cp}T_s > t_{max} \tag{4.3}$$

This is used to isolate the inter symbol interference (ISI) within each block of the OFDM symbol so that the ISI can be eliminated separately in each block by frequency domain equalization.

We assume M-ary modulation, with $M = 2^m$, so that a symbol consists of m bits. The bandwidth efficiency is

$$\kappa = \frac{m\,N_u}{N_{cp} + N}$$

and the resulting data rate is κ/T_s bits per second.

The data rate can be increased by increasing m but this will require a better SNR or lead to a higher bit-error rate (BER). We can also increase N_u (hence also N) until the bandwidth efficiency saturates to m. Finally, the symbol duration T_s can be shortened by increasing the available bandwidth B_w.

Latency is often also a consideration, and this leads to a limitation on the size of a data packet. The latency is at least equal to the duration of one OFDM block, which is $(N_{cp} + N)T_s$ and our desired system specification poses a maximum to this.

Several other limitations are in place. Apart from practical limitations and computational limitations, the number of subcarriers (N) that can be allocated in one OFDM block is limited by the requirement that the channel is constant over the duration of the OFDM symbol, i.e., the coherence time of the channel should be larger [19]. The coherence time is defined as $\frac{1}{\Delta f_D}$ where Δf_D is the range of possible Doppler frequencies of the channel, and the requirement becomes

$$(N + N_{cp})\,\Delta f_D\,T_s \ll 1\,. \tag{4.4}$$

Another requirement is that each subcarrier experiences flat fading. In frequency domain, the distance between fades is related to $1/t_{max}$ [19]. This leads to

$$\frac{B_w}{N} \gg \frac{1}{t_{max}} \quad \text{or} \quad N \gg B_w t_{max}. \tag{4.5}$$

As an example, let us design a system at $f_c = 60$ GHz with an available bandwidth $B_w = 5$ GHz. For the sake of simplicity, we consider BPSK modulation, which leads to $m = 1$ and $T_s = \frac{1}{B_w} = 0.2$ ns. The channel follows the metal enclosure in Chapter 2 and we take $t_{max} = 1\mu$s, i.e., the maximum length of the measured channels in average.

The proposed system is part of a mechatronic system in a closed metal environment in which a moving platform with sensors and actuators has to communicate to a controller which is fixed. Since movements that occur outside the enclosure do not affect the channel, we expect a slowly time-varying channel with a sufficiently long coherence time. The Doppler shift is defined as $\Delta f_D = \frac{\nu f_c}{c}$, where ν is the relative speed between transmitter and receiver, c is the speed of light, and f_c is the carrier frequency. If we assume a maximum relative speed of $10\ \mathrm{ms}^{-1}$, then the Doppler frequency range is $\Delta f_D = 2$ kHz, and the coherence time of the channel is $\frac{1}{\Delta f_D} = 0.5$ ms.

As discussed earlier, the spectral efficiency increases with the number of subcarriers. Considering (4.4) for $\Delta f_D T_s = 0.4 \times 10^{-6}$, the upper bound on the length of an OFDM block is given by $N_{cp} + N \ll 2.5 \times 10^6$ subcarriers. To satisfy this, we consider as constraint on the size of a transmission block

$$N_{cp} + N < 2.5 \times 10^5 \ \text{ symbols}.$$

The cyclic prefix N_{cp} should satisfy (4.3). For a channel of length 1 μs, this leads to

$$N_{cp} \geq 5000\,,$$

and the exact number depends on the scenario. Finally, condition (4.5) leads to

$$N \gg 5000\,.$$

We consider two versions of the system: (A) minimal latency; (B) maximal data rate. For minimal latency, we take $N_{cp} = 5000$, and a block size of $N = 2^{13} = 8192$ of which we take $N_u = 6720$ data/pilot symbols, and $2 \times N_{\text{guard}} = 1472$ null subcarriers for frequency guards at both ends of a block. This is different for different systems, we take less than 10% of the total subcarriers here. This leads to a spectral efficiency of $\kappa \approx 0.5$. The duration of one data packet becomes 2.64 μs and the data rate is 2.547 Gbps.

For maximal data rate, we choose $N = 2^{17} \approx 1.3{\times}10^5$, of which we take $N_u = 107520$ data/pilot symbols, and $2 \times N_{\text{guard}} = 23552$ null subcarriers. By taking less subcarriers as a frequency guard, the spectral efficiency (data rate) will increase, respectively. In our setting the same amount of frequency guard is used for both proposed OFDM block length, however since the bandwidth of the subchannels is smaller in the longer OFDM block, more subchannels are reserved that is for a fair comparison between the two proposed OFDM designs. The duration of one data packet is about 27.2 μs and the data rate is 3.95 Gbps. Thus, the latency increases more than 10 times by taking $N = 2^{17}$ compared to $N = 2^{13}$ whereas the data rate increases by less than 50%.

For the measured channels, the channel length t_{max} varies depending on the scenario, which alters the cyclic prefix and consequently the block length. Table 4.1 summarizes the parameters that are taken in our design and used in the following simulation. As usual, higher data rates could be obtained by considering higher-level modulation as well as multiple antennas.

4.4.2 Simulation Setup and Results

For the designed BPSK-OFDM systems, we will compare the BER performance for both simulated Rayleigh fading channels and the measured channels for the various scenarios. The Rayleigh fading channel is based on a tapped delay line setup where each path is assumed to be a Rayleigh fading process without considering any specific power delay profile.

		A: $N = 2^{13}$		B: $N = 2^{17}$	
	N_{cp}	rate (Gbps)	latency (μs)	rate (Gbps)	latency (μs)
Simulated 1	11	4.096	1.640	4.101	26.216
Simulated 2	4001	2.756	2.438	3.980	27.015
Scenario 1	3903	2.778	2.418	3.983	26.994
Scenario 2	5812	2.399	2.801	3.927	27.377
Scenario 3	5617	2.433	2.762	3.933	27.338

Table 4.1: Data rates and latency for (A) a low-latency system, and (B) a high-rate system.

The considered performance measure is BER as function of E_b/N_0, where E_b is the transmit energy per bit, and $\frac{N_0}{2}$ is the two-sided noise power spectral density (PSD). In the simulation, we first convert E_b to the energy per symbol E_s, taking into account the number of bits per symbol l and the overhead by the cyclic prefix N_{cp}, resulting in

$$\frac{E_s}{N_0} = \frac{E_b}{N_0} \times \frac{lN_u}{N_{cp} + N} \,. \tag{4.6}$$

The measured and simulated channels are normalized to unit power and convolved with the transmit sequence. In that case, the transmitted E_s/N_0 is equal to the received signal to noise ratio (SNR), and we add white Gaussian noise of suitable power to obtain the specified SNR.

For such dispersive channels, several subcarriers experience fading so that the BER is usually not very good. Various well-known techniques could be introduced to combat the fading subchannels, e.g., channel coding, interleaving, rake receiver design, as well as single input multiple output (SIMO) systems with diversity combining schemes [19]. However, for improved interpretation of the results, we will not consider these in the simulation. In the simulation, we assume that the receiver has perfect knowledge of the channel.

Fig. 4.4 shows the BER as a function of E_b/N_0 for the various channels. As expected, the performance is generally limited by the fading channel as the symbols in channel nulls cannot be recovered by frequency domain equalization. For the larger block size (design (B)), the BER performance is slightly (almost 3 dB) better than the shorter OFDM block, as expected, due to the spectral efficiency and less spread of transmit power over the cyclic symbols.

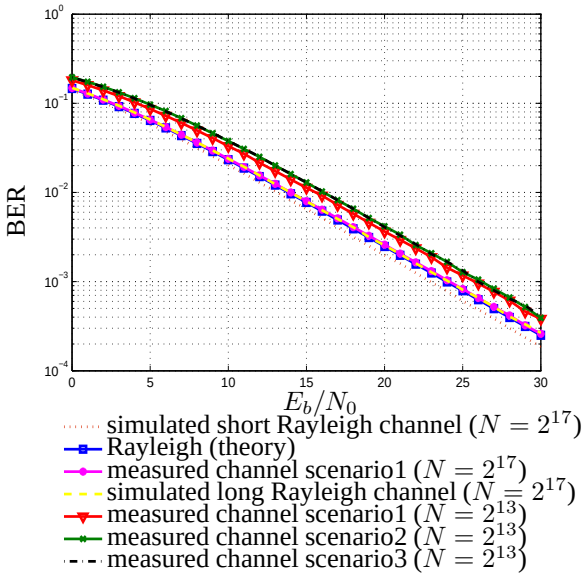


Figure 4.4: BER performance of an uncoded OFDM system over the measured channels for different scenarios with block size $N = 8192$ and $N = 2^{17}$. The BER plot of OFDM block of size $N = 2^{17}$ is shown for *scenario1*, (*scenario2* and *3* have the same BER curves as *scenario1*), and also for two simulated Rayleigh fading channels with $L = 10$ and $L = 4000$ taps. A theoretical BER curve for a narrowband signal (no ISI) is plotted as a reference.

4.5 Conclusion

A simple OFDM system design example for a highly dispersive wireless channel that is measured in a metal enclosure is introduced in this chapter. The results indicate that the frequency domain equalization can work sufficiently good and the most important source of system degradation comes from the fact that we need a long time-domain guard (cyclic prefix) to avoid inter-block interference. The loss is two-fold: increased latency and reduced spectral efficiency. Moreover, the power spread due to the extra power that is put in the redundant symbols, is a waste of power resources in a long cyclic prefix. The latter can be addressed with a zero padded time guard which has been proposed in [65], i.e., by sending nothing in the time guard between two consecutive blocks. However, the detection of the signal is more complicated with zero padded OFDM.

In a more general context, the summary of the literature suggests that the single carrier with frequency domain equalization can outperform the plain OFDM in

BER criteria but coded OFDM together with adaptive modulation still keeps the OFDM technique as one of the most promising solutions for wideband systems [66]. OFDM is proposed in this thesis as a competent solution for the short-range and high data-rate system of interest inside the industrial machine. However, this application requires a very careful and unique design of the OFDM system for such a specific environment and requirements. More on the implementation issues of an OFDM system will be presented in coming chapters.

This concludes the first part of this thesis providing detailed specifications and a design example for the mechatronic system of interest which was the starting point of this research project. The first three chapters opened up the challenges of transmission in an extremely reflective environment of confined metal enclosure which can generalize the propagation environment in mechatronic systems. We showed that the classical equalization techniques and modulation schemes can fairly cater for ISI resulting from the dispersive environment while the latency is the most challenging issue afterwards.

The rest of this thesis is dedicated to more general topics in signal processing for wideband communication systems. Multiple antenna systems and multidimensional processing are discussed in the following chapters. MIMO-OFDM systems are emphasized as an ultimate combinational systems that simultaneously exploit different dimensions of the signal in frequency, time and space. The difficulties and bottlenecks in MIMO-OFDM systems are introduced and novel signal processing solutions are developed to move the performance of these state of the art techniques towards their optimum.

Part II

MIMO Transmitter Design

Chapter 5

Preliminaries on Multiple-Input Multiple-Output (MIMO) Systems

5.1 Introduction

We start the second part of this thesis by a brief introduction to a multiple-input multiple-output (MIMO) system that is one of the existing parts of almost any modern wireless system nowadays. This extends the discussed single stream signal and channel model in the previous chapters to the multi-dimensional one. In this chapter, common MIMO system models and some related techniques on precoding and beamforming are briefly reviewed. Also the channel capacity in MIMO is reviewed and is compared to the single channel system.

MIMO techniques are widely used in wireless communications to improve the system performance by gaining diversity (SNR gain) and/or multiplexing (capacity). This exploits independent fading paths in wireless channels by means of an antenna array for which the elements are separated in space. However, the increased performance comes at the expense of additional complexity costs including

- Extra antennas and radio frequency (RF) chains: This includes high-power amplifiers, mixers, analog to digital convertors (ADCs) and digital to analog convertors (DACs).

- Space required to separate the antenna elements: The common spacing criteria assume a spacing of at least 0.38λ or approximately half a wavelength to achieve independent paths [67], hence the required distance between two

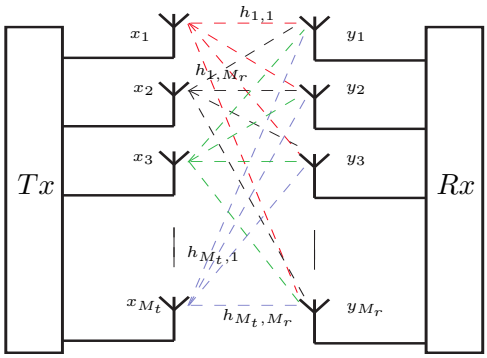


Figure 5.1: A point to point MIMO system with M_t and M_r transmit and receive antennas, respectively. The $h_{i,j}$ value represents the communication channel from the ith transmitter to the jth receiver.

elements is 5 cm for $f_c = 3$ GHz and 0.25 cm for $f_c = 60$ GHz.

- Circuit power: For each antenna additional components are used to receive and process the signal which translates to extra power consumption.

- Multidimensional signal processing: The channel, receive and transmit signals are expanded due to the additional signaling dimension. For example the equalization process is more complex.

The ultimate goal of MIMO system design is to maximize either the capacity (data-rate) or the diversity gain (SNR) subject to one or more of the aforementioned costs or conversely to minimize a cost subject to constraints on minimum SNR or capacity gain.

5.2 FIR-MIMO Channel Model

To establish the basics for the following chapters on MIMO transmission, a commonly used channel model is considered. This model is based on a finite impulse response (FIR) channel i.e., we assume a channel with a finite length similar to the corresponding SISO channel model in Chapter 3.

A point to point MIMO communication system of M_t transmit antennas and M_r receive antennas is illustrated in Fig. 5.1 which can be represented by the fol-

lowing discrete time model

$$\begin{bmatrix} y_1(n) \\ y_2(n) \\ y_3(n) \\ \vdots \\ y_{M_r}(n) \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,M_t} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,M_t} \\ h_{3,1} & h_{3,2} & \cdots & h_{3,M_t} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M_r,1} & h_{M_r,2} & \cdots & h_{M_r,M_t} \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \\ x_3(n) \\ \vdots \\ x_{M_t}(n) \end{bmatrix} + \begin{bmatrix} e_1(n) \\ e_2(n) \\ e_3(n) \\ \vdots \\ e_{M_t}(n) \end{bmatrix} . \quad (5.1)$$

By assuming an instantaneous channel model, this can be formulated in a compact form given by

$$\mathbf{y}(n) = \mathbf{H}\mathbf{x}(n) + \mathbf{e}(n). \quad (5.2)$$

The elements of the instantaneous channel matrix $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ correspond to the complex channel gain between the transmit signal $x(n)$ and the receive signal $y(n)$ as illustrated in Fig. 5.1. An additive noise vector at the receive antennas is denoted by $\mathbf{e}(n)$. The signal model in (5.1) is valid for a frequency-flat MIMO (narrowband) channel with no inter-symbol interference (ISI).

In a frequency selective (wideband) MIMO channel, the received signal is a convolution of a transmitted sequence with a matrix of channel coefficients. Therefore, each received vector $\mathbf{y}(n)$ at time index n is a superposition of the current transmit vector $\mathbf{x}(n)$ and the delayed versions of L previous transmitted vectors where L is the channel length, so

$$\mathbf{y}(n) = \sum_{l=0}^{L-1} \mathbf{H}(l)\mathbf{x}(n-l) + \mathbf{e}(n)\,. \quad (5.3)$$

Here $\mathbf{H}(l)$ is the channel matrix for the lth received path. Equation (5.3) is the MIMO extension of (3.5) to represent a wideband system where the computation is changed to matrix operations. Consequently, the channel estimation complexity grows with order of ${M_t}^3$, for each antenna at the receiver side.

For a wideband system of interest, a block transmission mode is considered as explained in Chapter 4. Accordingly, samples on M_r receive antennas are stacked in a vector $\mathbf{y}$ and the samples on M_t transmit antennas are collected in a vector $\mathbf{x}$. Assuming a block of $N + L$ samples is transmitted, define

$$\mathbf{x} = [\mathbf{x}^T(-L), \mathbf{x}^T(-L+1), \ldots, \mathbf{x}^T(N-1)]^T, \quad (5.4)$$

$$\mathbf{y} = [\mathbf{y}^T(0), \mathbf{y}^T(1), \ldots, \mathbf{y}^T(N-1)]^T. \quad (5.5)$$

In a time-invariant channel, the instantaneous channel matrix $\mathbf{H}$ is repeated along the diagonal of a convolution matrix $\mathbf{G}$, which leads to a block Toeplitz matrix as follows

$$\mathbf{G} = \begin{bmatrix} \mathbf{H}(L-1) & \cdots & \mathbf{H}(0) & 0 & 0 & \cdots & 0 \\ 0 & \mathbf{H}(L-1) & \cdots & \mathbf{H}(0) & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{H}(L-1) & \cdots & \mathbf{H}(0) & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \mathbf{H}(L-1) & \cdots & \mathbf{H}(0) \end{bmatrix}. \quad (5.6)$$

In fact, (5.6) is the MIMO format of (3.4.1), however the dimensions are expanded here as $\mathbf{G} \in \mathbb{C}^{M_r N \times M_t(N+L)}$.

The compact formulation of (5.3) for the received data block, similar to (3.12) yet with expanded dimensions, is given by

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{e}\,. \quad (5.7)$$

The first L blocks of $\mathbf{y}$ are contaminated with the L previous transmit blocks which can be avoided by putting a time guard interval between the consecutive transmit blocks, on all antennas.

The considered system model is very general and can describe different settings for MIMO systems, e.g., the possibility of having only one transmitter or receiver antenna which leads to single-input multiple-output (SIMO) and multiple-input single-output (MISO) systems, respectively. Also the data block sent over each transmit antenna could be independent data to increase the throughput or it could be a repetition of the same information to achieve diversity.

5.3 MIMO-OFDM System Model

A combination of the OFDM model in Chapter 3.4.2 and the MIMO matrix representation leads to an equivalent MIMO-OFDM system model, where the received data vector in time domain is given by

$$\mathbf{y} = \mathbf{G}_c \mathbf{F}_b^H \mathbf{x} + \mathbf{e}\,, \quad (5.8)$$

where $\mathbf{F}_b = \mathbf{I}_{M_t} \otimes \mathbf{F}$ is a block-diagonal matrix with unitary discrete Fourier transform (DFT) matrices $\mathbf{F} \in \mathbb{C}^{N \times N}$ along its diagonal. The transmit signal is represented by $\mathbf{x} \in \mathbb{C}^{N \times 1}$ (excluding the cyclic prefix) where $\mathbf{x}_F = \mathbf{F}_b^H \mathbf{x}$ is the transmit vector in time domain. The signal model in (5.8) is similar to (3.19) but

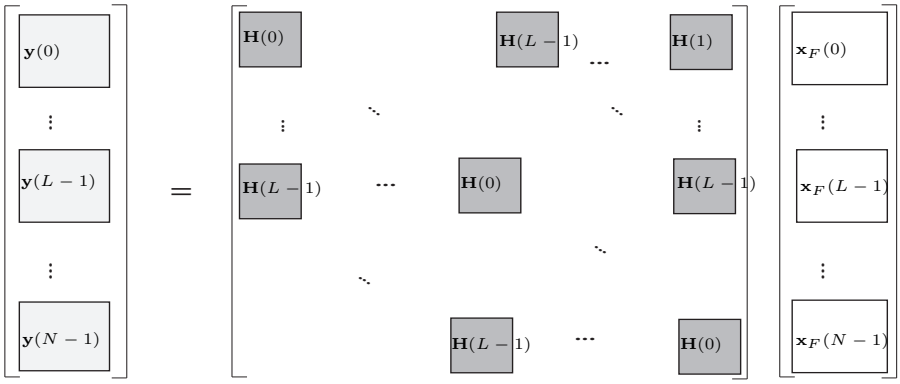


Figure 5.2: A noise-free MIMO-OFDM system model with a circulant channel matrix after discarding cyclic prefix (time guard) data.

differently, the received and transmit vectors are the result of stacking time samples on multiple antennas in MIMO setting. This is illustrated in Fig. 5.2 (ignoring the noise). The insertion of the cyclic prefix as time guard between the transmit blocks enables us to transform the Toeplitz matrix of (5.6) to a block circulant matrix in Fig. 5.2.

At the receiver, the received signal is multiplied by $\mathbf{F}_b$. We know that circulant matrices can be diagonalized using a Fourier transform, so this multiplication leads to a block diagonal channel matrix, with narrowband channel matrices on the diagonal. This transformation is expressed as

$$\mathbf{F}_b\mathbf{y} = \underbrace{\mathbf{F}_b\mathbf{G}_c\mathbf{F}_b^H}_{\mathbf{Q}}\mathbf{x} + \mathbf{v}\,, \tag{5.9}$$

where $\mathbf{v}$ is the Fourier transform of the noise vector, and $\mathbf{Q} = \mathbf{F}_b\mathbf{G}_c\mathbf{F}_b^H$ is a block diagonal channel matrix owing to the property of circulant matrices. This is equivalent to a block diagonal flat fading channel. This simple transformation reduces the interference cancelation task to only a spatial separation by taking care of the temporal interference (ISI). For the rest of this chapter, we consider the narrowband MIMO channel of (5.1) assuming that the frequency selective MIMO channel is flattened, e.g., by the discussed MIMO-OFDM technique.

5.4 Capacity of a MIMO Channel

Multiple antenna technology is one of the most rewarding techniques to increase the wireless channel capacity linearly (same effect as increasing the bandwidth). This

makes the wireless channel outperform the classical AWGN channel at the cost of additional complexity. The capacity gain obtained from multiple antennas heavily depends on the available channel state information (CSI) at either the receiver or transmitter, the SNR and the correlation between the channel gains at each antenna element.

In general, the MIMO channel $\mathbf{H}$ is a stochastic variable (the static channel will be discussed later) and having no prior information about the distribution of the channel, commonly, the elements in $\mathbf{H}$ are modeled as complex jointly Gaussian random variables. More specifically, a zero-mean, spatially white (ZMSW) channel model is considered where the channel mean is zero (over different channel realizations in long-term) and the channel covariance is modeled as white (identity matrix), i.e., the channel elements are assumed to be i.i.d., zero mean and unit variance random variables. This model admits a Rayleigh flat fading channel model [58].

In general, the mutual information between the input and the output is maximized over all possible input distributions, to obtain the channel capacity [19] for all types of channels including both SISO and MIMO channels. The capacity is given here in units of bits (nats) per second per Hz for the logarithm of base two (natural) which represents the spectral efficiency. Hence, multiplying it by bandwidth gives the data rate in bits (nats) per second, i.e., for a MIMO system is defined as

$$C = \max_{P_x(\mathbf{x})} I(\mathbf{x};\mathbf{y}) = \max\left[H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x})\right] \tag{5.10}$$

where $P_x(\mathbf{x})$ is the probability density function (pdf) of the input with the covariance matrix of $\mathbf{R}_x = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\}$ [68]. The entropy $H(\mathbf{y}|\mathbf{x})$ is equivalent to the noise entropy, so maximizing the mutual information is equivalent to maximizing the entropy of $\mathbf{y}$ ($H(\mathbf{y})$) which is a function of SNR and hence of the output covariance matrix. Assuming additive zero mean Gaussian noise with identity covariance ($\sigma_e^2\mathbf{I}_{M_r}$), the output covariance matrix is

$$\mathbf{R}_y = \mathbb{E}\{\mathbf{y}\mathbf{y}^H\} = \mathbf{H}\mathbf{R}_x\mathbf{H}^H + \sigma_e^2\mathbf{I}_{M_r}\,, \tag{5.11}$$

so the optimization variable in (5.10) is reduced to $\mathbf{R}_x$, therefore, for jointly Gaussian processes, the maximum mutual information is given by

$$C = \max_{\mathbf{R}_x} \log\left(\det[\frac{1}{\sigma_e^2}\mathbf{H}\mathbf{R}_x\mathbf{H}^H + \mathbf{I}_{M_r}]\right). \tag{5.12}$$

Shannon proved for an AWGN channel, that for all given output realizations, the entropy is maximized when the input distribution matches the noise distribu-

tion. Therefore, the optimal input distribution is a zero mean complex Gaussian for an AWGN channel [69]. Respectively, for a ZMSW complex Gaussian MIMO channel model, the ideal input is ZMSW complex Gaussian distributed which is matching the channel. Note that the capacity is a stochastic variable since the channel is considered to be stochastic, so the notion of average (ergodic) capacity is commonly used. Hence the optimal input covariance matrix $\mathbf{R}_x$ is the scaled identity (equal power allocation) and the average capacity with a power limit, over all channel realizations is given by

$$C = \mathbb{E}_{\mathbf{H}} \left\{ \log \left(\det[\frac{P}{M_t \sigma_e^2} \mathbf{H}\mathbf{H}^H + \mathbf{I}_{M_r}] \right) \right\} \text{ bits/sec/Hz}, \tag{5.13}$$

in units of bits per second per Hz, i.e., by considering the rate per channel use (spectral efficiency). In turn, P is the total transmit power and σ_e^2 is the noise power for the ZMSW channel model [70]. Unfortunately, the capacity is only known, in a closed form, for a few channel types due to the complexity of capacity evaluations. A useful performance metric for non-ergodic channels is the probability that the capacity is below some value for a specific percentage of channel realizations, which is referred to as the outage capacity [19]. In non-ergodic channels the transmission blocks are short compare to the channel variations, so the ensemble average over the output sequence does not represent the stochastic mean of the random process.

Note that we commonly assume the availability of CSI at the receiver (CSIR), as this is quite easy to obtain by sending pilot symbols to the receiver unlike the CSI at the transmitter which is rather expensive to learn. Once the channel matrix model deviates from ZMSW, then equal power allocation is no longer optimal in terms of capacity and there is a bias that needs to be exploited when the input covariance is adapted (the Gaussian model is still valid). In this situation, assuming channel state information at the transmitter (CSIT), the channel capacity is not known explicitly and is obtained by solving the following optimization problem

$$C = \mathbb{E}_{\mathbf{H}} \left\{ \begin{array}{cc} \max & \log \left(\det[\frac{1}{\sigma_e^2} \mathbf{H}\mathbf{R}_x\mathbf{H}^H + \mathbf{I}_{M_r}] \right) \\ \mathbf{R}_x : \text{Tr}(\mathbf{R}_x) = P & \end{array} \right\}, \tag{5.14}$$

with a short term power constraint over the channel realizations i.e., for each channel realization the power constraint is to be satisfied. The optimization problem in (5.14) gives the average capacity of the channel with respect to the instantaneous total power constraint [70].

An alternative capacity is given by imposing a long term power constraint, i.e., the total power constraint is to be satisfied in average over many channel realizations where the transmit power for each channel realization is $P_{\mathbf{H}}$ and the constraint

is denoted as $\mathbb{E}\{P_{\mathbf{H}}\} = P$. In fact, the transmit power can be larger than P for a specific channel realization as long as in average the constraint is satisfied. In this context, power allocation techniques aim to redistribute the power optimally over all antennas, which will be discussed in Chapter 7. More on this topic can be found in [71, 72, 73].

Interestingly, when the number of transmit antennas is large, we have

$$\lim_{M_t \to \infty} \frac{1}{M_t} \mathbf{H}\mathbf{H}^H = \mathbf{I}_{M_r}\,, \tag{5.15}$$

such that the capacity in (5.13) asymptotically reaches a constant value of $C = M_r \log(1 + \frac{P}{\sigma_e^2})$ [34]. In case of no knowledge of CSIT and large M_t, the capacity is the same as the one with CSIT, i.e, the optimal input covariance is the identity matrix. This is one of the most interesting features of MIMO system, that the capacity grows linearly even if nothing is known about the communication channel at the transmitter.

Note that, even though the availability of CSIT does not increase the capacity for a large number of transmit antennas, it can significantly reduce the decoding task at the receiver [19]. For the sake of comparison and to highlight the superiority of MIMO design with respect to SISO systems, a summary of channel capacities for both SISO and MIMO channels is collected in Table 5.1. The derivation of all the mentioned capacities is out of the scope of this thesis, therefore references are provided for the details.

Static MIMO channels or deterministic channels are time-invariant and accordingly, the parameters are fixed so the capacity is deterministic given by a constant value. If there is no CSIT then the mutual information maximization problem subject to the total power constraint, yields equal power allocation to all transmitters, similar to the stochastic channel. This means $\mathbf{R}_x$ is deterministic and is a scaled identity matrix such that the capacity boils down to

$$C = \log\left(\det[\frac{P}{M_t\sigma_e^2}\mathbf{H}\mathbf{H}^H + \mathbf{I}_{M_r}]\right) = \sum_{m=1}^{R=\text{rank}(\mathbf{H})} \log\left(1 + \frac{\delta_m P}{M_t\sigma_e^2}\right). \tag{5.16}$$

where δ_m is the mth eigenvalue value of $\mathbf{H}\mathbf{H}^H$.

When CSIT is assumed,

$$C = \max_{\mathbf{R}_x:\text{Tr}(\mathbf{R}_x)=P} \log\left(\det[\mathbf{I}_{M_r} + \frac{P(m)}{P}\gamma(m)]\right), \tag{5.17}$$

leads to the waterfilling power allocation once a total power constraint is imposed [19]. The waterfilling solution allocates less power to more noisy channels (low SNR) and vise-versa, and is given by

$$\frac{P(m)}{P} = \begin{cases} \frac{1}{\gamma_0} - \frac{1}{\gamma(m)}, & \gamma(m) \geq \gamma_0 \\ 0, & \gamma(m) \leq \gamma_0 \end{cases} \tag{5.18}$$

where the SNR on the mth channel $\gamma(m), m = 1, 2, \ldots, M$, is defined as $\gamma(m) = \frac{P\delta_m}{\sigma_e^2 M_t}$ and γ_0 is found numerically by the total power constraint equation which implies that no data is sent when the channel SNR is below this threshold. The solution is stated in Table 5.1, and for more details and derivation of the waterfilling solution see [19]. The static channel model is useful to model block fading channels where each block experiences a flat fading channel. Note that the capacity evaluation in this chapter is performed by considering the total power constraint, mainly due to the complexity of the capacity evaluations. Indeed, even for a simple power constraint, the capacity is known for just a few channel types.

In Table 5.1 we summarize the capacity expressions for various SISO and MIMO channels with respect to the availability of the channel information at the transmitter and/or receiver. If the CSIT is available, the waterfilling algorithm is used to optimally allocate the power resources among the channels. Hence, the capacities for SISO and static MIMO channels are obtained by waterfilling with γ_0 as the threshold SNR below which no data is transmitted. In contrast, a uniform power allocation is considered optimal if only CSIR is available. The capacity formulas are given subject to the total power constraint where two types of total constraint, are considered for time-variant MIMO channels: The short-term power constraint and the long-term constraint which have been discussed before. Note that the SNR is fixed for an AWGN channel (no fading) and is given by $\gamma = \frac{P}{\sigma_e^2}$, however for the fading channels this is a random variable which is represented by its pdf $p_\gamma(\gamma)$. The time-variant channel is assumed to be continuous where the discrete time-variant channel can be modeled as a block fading channel with each block experiencing the flat fading (static) channel.

Table 5.1: The capacity for the SISO and MIMO channels under different assumptions on the CSI and fading are given subject to the total transmit power constraint of P. The indices $k = 1, 2, \ldots, N$ denote the frequency bins for frequency selective channels, while $m = 1, 2, \ldots, R$ denote the spatial channels where $R = \text{rank}\{\mathbf{H}\}$. The time-variant MIMO channel capacity knowing the CSIT for short-term power constraint, and long-term power constraint over channel realizations is marked with $*$ and $**$ respectively. The random variable $P_{\mathbf{H}}$ is the transmit power for a channel realization.

CSIT	SISO	time-variant	$\int_{\gamma_0}^{\infty} \log(\frac{\gamma}{\gamma_0}) p_\gamma(\gamma) d\gamma$ [19]
		frequency selective	$\sum_{k:\gamma(k)\geq\gamma_0} \log(\frac{\gamma(k)}{\gamma_0})$ [68]
		doubly selective	$\sum_{k:\gamma(k)\geq\gamma_0} \int_{\gamma_0}^{\infty} \log(\frac{\gamma(k)}{\gamma_0}) p_\gamma(\gamma(k)) d\gamma(k)$ [19]
	MIMO [70]	static	$\sum_{m:\gamma(m)\geq\gamma_0} \log(\frac{\gamma(m)}{\gamma_0})$
		time-variant	$\mathbb{E}_{\mathbf{H}}\{\max_{\mathbf{R}_x:\text{Tr}(\mathbf{R}_x)=P} \log(\det[\frac{1}{\sigma_e^2}\mathbf{H}\mathbf{R}_x\mathbf{H}^H + \mathbf{I}_{M_r}])\}^*$ $\max_{P_{\mathbf{H}}:\mathbb{E}\{P_{\mathbf{H}}\}=P} \mathbb{E}_{\mathbf{H}}\{\max_{\mathbf{R}_x:\text{Tr}(\mathbf{R}_x)=P_{\mathbf{H}}} \log(\det[\frac{1}{\sigma_e^2}\mathbf{H}\mathbf{R}_x\mathbf{H}^H + \mathbf{I}_{M_r}])\}^{**}$
CSIR	SISO	time-variant	$\int_{0}^{\infty} \log(1+\gamma) p_\gamma(\gamma) d\gamma$ [74]
	MIMO [70]	static	$\log(\det[\frac{P}{M_t\sigma_e^2}\mathbf{H}\mathbf{H}^H + \mathbf{I}_{M_r}])$
		time-variant	$\mathbb{E}_{\mathbf{H}}\{\log(\det[\frac{P}{M_t\sigma_e^2}\mathbf{H}\mathbf{H}^H + \mathbf{I}_{M_r}])\}$

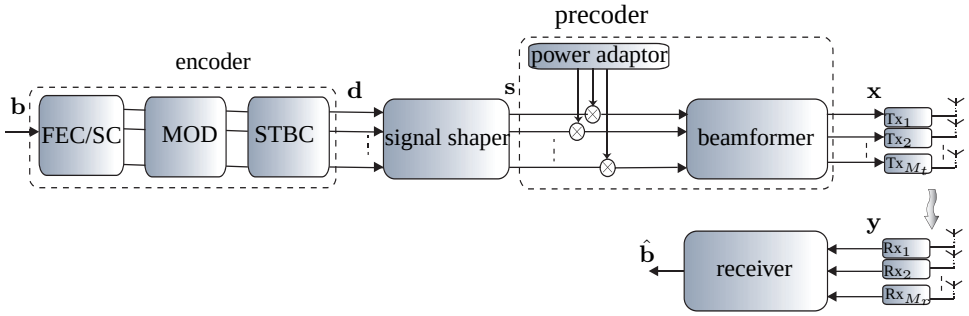


Figure 5.3: A block diagram of a narrowband MIMO system including the encoder and precoder.

5.5 MIMO Precoding Design

In this section we focus explicitly on well-known transmitter designs and common blocks in a MIMO transmitter for the purpose of capacity maximization assuming a flat fading (instantaneous) MIMO channel. The MIMO transmitter (precoding) design is the subject of the next two chapters of this thesis.

5.5.1 MIMO System Schematic

In completion of the previous SISO model in Fig. 4.2, a typical MIMO system with focus on pre-processing blocks at the transmitter side is illustrated in Fig. 5.3. Data streams at bit level are fed to the encoder which divides them into R parallel and independent symbol sequences. The encoder includes a forward error correction coder (FEC) such as LDPC or Reed Solomon code, a source coder which maps the data into symbols, a modulator which prepares the symbols for transmission (such as multicarrier modulation or spreading), and perhaps a space-time coder block which schedules the data streams in time and space for transmission.

The space-time coding is a solution for the multiplexing-diversity trade-off and is categorized into two main types: STBC (space-time block codes) and STTC (space-time trellis codes) which are obtained based on different code design criteria, see [19, 75, 76] and references therein for further information. Considering the spatial multiplexing for $M_r \leq M_t$ [1], the codeword is a vector $\mathbf{d} \in \mathbb{C}^{M_r \times 1}$ and the

[1] It is a common assumption for a spatial multiplexing systems that the number of transmit and receive antennas is the same, however, we include a more general scenario of $M_r \leq M_t$ with M_r streams.

codeword covariance matrix is given by

$$\mathbf{\Phi} = \mathbb{E}\{\mathbf{d}\mathbf{d}^H\}, \tag{5.19}$$

where the expectation is over the codeword distribution [75].

5.5.2 Pre-processing at the Transmitter

Processing at the transmitter includes two major steps at the symbol level: signal shaping and precoding. The former requires no CSIT while the latter exploits the channel information available at the transmitter. Based on this model, the transmit signal is given by

$$\mathbf{x} = \mathbf{Z}\mathbf{d}, \tag{5.20}$$

where $\mathbf{Z} \in \mathbb{C}^{M_t \times M_r}$ is a linear pre-processor which consists of a signal shaper, power allocator, and the beamformer. The singular value decomposition of the pre-processor matrix $\mathbf{Z}$ is given by

$$\mathbf{Z} = \mathbf{U}_Z \mathbf{\Gamma} \mathbf{V}_Z^H; \quad \mathbf{x} = \underbrace{\mathbf{U}_Z \mathbf{\Gamma}}_{\mathbf{W}} \underbrace{\mathbf{V}_Z^H \mathbf{d}}_{\mathbf{s}}, \tag{5.21}$$

where $\mathbf{V}_Z \in \mathbb{C}^{M_r \times M_r}$ is the input shaping matrix that mixes the signal coming from the encoder to form the input, $\mathbf{s} = \mathbf{V}_Z^H \mathbf{d}$, to the precoder. The input shaping matrix is designed with respect to the codeword selection, while unitary matrix $\mathbf{U}_Z \in \mathbb{C}^{M_t \times M_t}$ forms the beam patterns towards the different receive antennas which creates non-interfering transmission paths. Commonly, there is a power constraint that needs to be satisfied, this leads to criteria to determine $\mathbf{\Gamma} \in \mathbb{C}^{M_t \times M_r}$. The precoder is referred to the matrix $\mathbf{W} = \mathbf{U}_Z \mathbf{\Gamma}$, in this context. As a result, the input covariance matrix is given by

$$\mathbf{R}_x = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbb{E}\{\mathbf{W} \underbrace{\mathbf{V}_Z^H \mathbf{d}\mathbf{d}^H \mathbf{V}_Z}_{\mathbf{R}_s} \mathbf{W}^H\}. \tag{5.22}$$

Consider the flat fading MIMO channel in (5.1), with the singular value decomposition (SVD) given by

$$\mathbf{H} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^H\,, \tag{5.23}$$

where $\mathbf{U} \in \mathbb{C}^{M_r \times M_r}$ and $\mathbf{V} \in \mathbb{C}^{M_t \times M_t}$ are unitary matrices and $\mathbf{\Sigma} \in \mathbb{C}^{M_r \times M_t}$ is a diagonal matrix containing the singular values (δ_m) of $\mathbf{H}$. The received signal after the channel is given by

$$\mathbf{y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \mathbf{W}\mathbf{V}_Z\mathbf{d} + \mathbf{e}. \tag{5.24}$$

The signal shaper $\mathbf{V}_Z$ and the precoder $\mathbf{W}$ are designed with respect to different criteria and constraints. The channel capacity is one of the common metrics to be maximized in a stochastic sense, which leads to a matrix $\mathbf{V}_Z$ matching the eigenvectors of the codeword covariance matrix [77]. Commonly for spatial multiplexing it is assumed that $\mathbf{\Phi} = \mathbf{I}_{M_r}$ and hence $\mathbf{V}_Z$ is an identity. We discuss next the criteria to determine the precoder.

Parallel Decomposition of the MIMO Channel

Parallel decomposition of the MIMO channel ensures the maximum achievable rate (capacity) in a point to point MIMO system. The multiplexing gain of a MIMO system comes from the fact that the channel can be decomposed into parallel independent channels. However, this requires the knowledge of the channel and the joint processing at the transmitter as well as the receiver. It is referred to as transmitter beamforming and receiver shaping or *eigen beamforming*, in the literature [75].

The transmitter precoding and receiver shaping can transform the static MIMO channel into SISO channels. This can be implemented by multiplying the transmit sequence by the matrix $\mathbf{V}$ and the receive vector by the matrix $\mathbf{U}^H$. Using this technique, for a channel matrix of rank R, the complexity of the maximum likelihood (ML) detection at receiver becomes linear with R, while in general an exhaustive search is required over all possible vector combinations [77]. This complexity reduction is the direct result of the extra information which is available at the transmitter.

To perform the decomposition, the matrix $\mathbf{U}_Z$ is chosen to match the right eigen vectors of the channel so $\mathbf{U}_Z = \mathbf{V}$. The transmit vector then is given by

$$\mathbf{x} = \mathbf{V}\,\mathbf{\Gamma}\,\mathbf{d}\,, \tag{5.25}$$

for the eigen beamforming problem. In turn, a similar unitary beamformer at the receiver is required to complete the channel decomposition task, i.e.,

$$\mathbf{y} = \mathbf{U}^H(\underbrace{\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H}_{\mathbf{H}}\mathbf{V}\,\mathbf{\Gamma}\,\mathbf{d} + \mathbf{e}) = \mathbf{\Sigma}\mathbf{\Gamma}\,\mathbf{d} + \mathbf{U}^H\mathbf{e}, \tag{5.26}$$

which by taking $\mathbf{\Gamma} = \mathbf{\Sigma}^{-1}$, implying no power constraint, (5.26) yields

$$\mathbf{y} = \mathbf{d} + \mathbf{U}^H\mathbf{e}. \tag{5.27}$$

This is equivalent to R parallel SISO channels with a additive noise.

Optimal Precoder

The parallel channel decomposition using a unitary precoder requires transmitter and receiver cooperation to complete the decomposition, however, this is not realistic in multi-user systems where the receivers are non-collaborative entities in general. Consequently, it is preferable that the precoding is completed at downlink. The optimal capacity achieving precoder maximizes the mutual information between $\mathbf{x}$ and $\mathbf{y}$. To derive the optimal precoder we need to calculate the output covariance as

$$\mathbf{R}_y = \mathbf{H}\mathbf{R}_x\mathbf{H}^H + \sigma_e^2\mathbf{I}_{M_r} = \mathbf{H}\mathbf{W}\mathbf{W}^H\mathbf{H}^H + \sigma_e^2\mathbf{I}_{M_r}. \tag{5.28}$$

Therefore, the average channel capacity problem in (5.14) for a time-variant channel is reduced to finding the optimal linear precoder as follows

$$C = \mathbb{E}_{\mathbf{H}} \left\{ \begin{array}{cc} \max & \log\left(\det[\frac{1}{\sigma_e^2}\mathbf{H}\mathbf{W}\mathbf{W}^H\mathbf{H}^H + \mathbf{I}_{M_r}]\right) \\ \mathbf{W} : \mathrm{Tr}(\mathbf{W}) = P & \end{array} \right\}. \tag{5.29}$$

The total power constraint is imposed by $\mathrm{Tr}(\mathbf{W}) = P$ assuming $\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_{M_r}$. Other constrains rather than the total power constraint can be used to define the capacity which will be discussed in the coming chapters of this thesis. For a static MIMO channel, the expectation in (5.29) is dropped for capacity evaluations.

5.6 Conclusion

We have extended our initial SISO channel and system model, introduced in Chapters 3 and 4, to a MIMO system in this chapter. The basics of MIMO transceiver design have been reviewed, i.e., just a tip of the MIMO iceberg that has grown enormously during the last two decades. There are interesting books that look more into detailed algorithms and their modifications and extensions such as [19, 75] which are the main references of this chapter.

The MIMO-OFDM system model was introduced that removes the ISI in a wideband MIMO system. Having the temporal interference removed, the next concern is how to optimally design the transmitter in order to maximally exploit the spatial dimension. A simple sketch of the common blocks at the transmitter was introduced and their relations and design criteria were briefly explained. The encoder is responsible for the modulation, source and channel coding. Differently from SISO encoder, in MIMO the space time coder allocates data steams to different antennas and time slots. Afterwards, the signal shaper delivers the optimally

shaped data streams to the precoder for further processing based on the CSIT. This is the topic for the coming chapters which focus on the precoder design with respect to the hardware constraints. Furthermore, multi-user MIMO precoding will be discussed later which has not been covered in this chapter.

In the next chapter, we look at the well-known problem of peak-to-average power ratio (PAPR) in MIMO-OFDM systems. The unpredictable OFDM waveform is manipulated by reformulating the precoder design problem and defining new design parameters that can be tuned to adjust the OFDM waveform towards a more hardware-friendly waveform. The proposed signal processing algorithm for the PAPR reduction purpose is very interesting since it exploits an efficient minimization technique to solve the underlying design problem.

Even though the precoding problem is well studied for the total power constraint over all transmitters, the problem is less considered when the power constraint is imposed on each individual antenna. The latter is more realistic since each transmitter has its own RF chain and power amplifier. The problem of precoding design with such per antenna power constraints and a limited number of transmitters (RF chain) for multi-user MIMO is studied in Chapter 7 of this thesis.

Chapter 6

Joint Precoding and PAPR Reduction in MIMO-OFDM Systems

MIMO technology offers a great increase in data rate and reliability of communication system at the cost of extra signaling dimensions. Therefore the prospect of an effective transceiver requires the management of the interference in time, space and frequency. This is getting more complicated in multi-user systems with extra sources of interference. As explained in Chapter 3, orthogonal frequency division multiplexing (OFDM) can reduce a frequency selective channel to a flat fading one with the aid of the mathematical property of circulant matrices. OFDM is extended to a MIMO-OFDM system which translates a frequency selective MIMO channel to a flat-fading MIMO channel which only contains the spatial interference. This is a significant progress to eliminate the interference in temporal dimension and makes OFDM the most popular technique for almost all wideband wireless standards. However, the OFDM scheme suffers from implementation complications due to its unconventional waveform which is not hardware-friendly.

In fact, hardware imperfections can significantly influence the performance of

Part of this chapter is published as "Constant Modulus Algorithm for Peak-to-Average Power Ratio (PAPR) Reduction in MIMO OFDM/A," Signal Processing Letters, IEEE, vol.20, no.5, pp.531-534, May 2013, doi: 10.1109/LSP.2013.2254114.

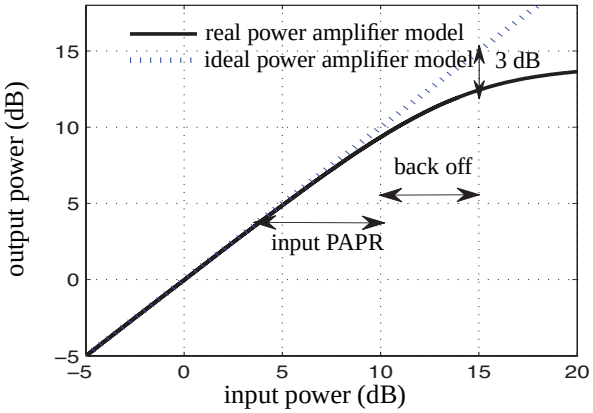


Figure 6.1: Input-output diagram of a typical nonlinear amplifier.

a wireless system. These are required to be modeled properly, particularly for an emerging millimeter-wave systems, in order to develop effective compensating signal processing algorithms. A well-known drawback of OFDM is that the amplitude of the time domain signal varies strongly with the transmitted symbols modulated on the subcarriers in the frequency domain, resulting in a 'peaky' signal that shows a rather wide dynamic range. The measure for evaluating the dynamic range of the system is the peak-to-average power ratio (PAPR) of the signal.

6.1 Introduction

What is PAPR? High power amplifiers (HPA) are inevitable parts of almost all communication systems. These devices are very sensitive to nonlinearity and require precise operating points to be set. In single-carrier modulation, the signal amplitude is deterministic to a good extent, so the operating point in the amplifier can be determined easily while for multi-carrier systems like OFDM, the envelope of the time domain signal will change with different data symbols. The direct result of this fluctuation is a distortion in the HPA, that appears in form of noise at the receiver, and a signal constellation rotation due to phase conversion. Moreover, the orthogonality between subcarriers may be destroyed which leads to a dramatic BER degradation. Fig. 6.1 shows the input-output relation in a typical power amplifier for an ideal and real device and the required input back-off to be considered in order to avoid nonlinearity problems. Clearly, this back-off value decreases the amplifier efficiency so there is a demand to make the output OFDM signal more

robust in terms of envelope fluctuations [78].

To quantify the distortion caused by the amplitude fluctuations, a metric is defined to measure the dynamic range of the multicarrier signal. The PAPR is a direct measure of the severity of the distortion and shows how far the operation point needs to be set from the saturation point. Since the signal amplitude is a stochastic variable, the PAPR of the signal is commonly presented with its probabilistic properties such as cumulative distribution function [79, 80].

PAPR reduction techniques have been studied well since two decades ago when OFDM became popular, so there are many signal processing techniques developed to limit the dynamic range of the OFDM signal. There are many factors that should be considered before a specific PAPR reduction technique is chosen. These factors include PAPR reduction capability, power increase in transmit signal, bit error rate (BER) increase at the receiver, loss in data rate, computational complexity, etc.

Careful attention must be paid to the fact that some techniques result in other harmful effects. For example, amplitude clipping techniques clearly remove the time domain signal peaks, but this results in in-band distortion and out-of-band radiation [81, 82, 83]. Some techniques require a power increase in the transmit signal after using PAPR reduction techniques. For example, tone reservation (TR) requires more signal power because some of its power must be used for the peak reduction carriers. Tone injection (TI) uses a set of equivalent constellation points for an original constellation point to reduce PAPR, but since all the equivalent constellation points require more power than the original constellation point, the transmit signal will have more power after applying TI [84]. Other techniques may have an increase in BER at the receiver if the transmit signal powers is fixed or equivalently may require larger transmit signal power to maintain the BER after applying the PAPR reduction technique; active constellation extension (ACE) is one example [85].

Some techniques require the data rate to be reduced. For example, the block coding technique requires part of the information symbols to be dedicated to control the PAPR. In selection mapping (SLM) and partial transmit sequence (PTS), the data rate is reduced due to the side information used to inform the receiver of processing at the transmitter [86, 87]. In these techniques the side information may be received in error unless some form of protection such as channel coding is employed. Once channel coding is used, the loss in data rate due to side information is increased further. Computational complexity is another issue since most PAPR reduction techniques such as PTS use many iterations to perform an exhaustive

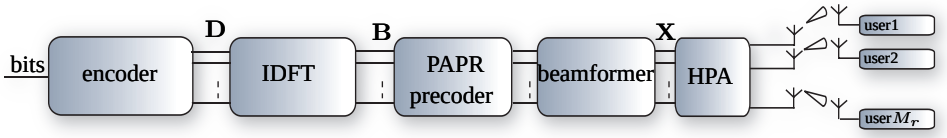


Figure 6.2: A block diagram of a transmitter in multi-user MISO-OFDM system with specific precoding block to reduce the PAPR.

search algorithm. For a more detailed literature survey see [78, 88].

PAPR reduction in MIMO-OFDM is a quite recent topic compared to the traditional PAPR reduction in SISO systems. Although there are many effective PAPR reduction techniques for OFDM, the majority of them are developed for SISO-OFDM and cannot be applied to MIMO-OFDM systems without proper modifications. Moreover, the capability of traditional PAPR reduction techniques to overcome the PAPR problem in MIMO systems is not well verified yet. In multiuser systems the choices are even more limited since the user devices are not collaborating in general, so any precoding technique that relies on intelligent global decoding at the receiver side will not work. Even if each user can invert the precoding effect, this is extremely power inefficient given limited battery life of mobile devices since this needs to be performed for each OFDM symbol (online processing). Additionally, there is an extra dimension in MIMO systems that can be exploited to develop more effective and efficient PAPR reduction techniques and ignoring it leads to loss of performance.

In general, the transmitter in a MIMO-OFDM system, includes an IDFT block in addition to the encoder. A Fast Fourier transform (FFT) is used in practice which makes OFDM very efficient in terms of complexity. In Fig. 6.2 a multi-user MIMO-OFDM system is illustrated, we focus on the downlink direction since the PAPR is more severe in downlink transmissions. Note that in multi-user systems, the users cannot collaborate in general so synchronized uplink processing is impractical and therefore an effective interference cancelation at downlink is of vital importance. Single-antenna users are considered here for simplicity without loss of generalization. The high PAPR is commonly a transmitter problem and we consider multiple antennas at the transmitter side for the downlink, however the same technique can potentially be used for uplink transmissions.

The block diagram shown in Fig. 6.2 is very similar to the one in Fig. 5.3 except

for the IDFT block for OFDM modulation, and the proposed PAPR precoding block that is added for pre-processing the OFDM signal such that the dynamic range of the output waveform is limited. The beamformer directs and weights the transmit signal to match the channel conditions (eigen-beamforming) whereas the encoder, IDFT and PAPR precoding units are designed regardless of the channel conditions and merely depend on the input data.

On the one hand, the PAPR is measured at the input of the HPA after beamforming and on the other hand the beamforming may not be affected by the PAPR precoding which is a data dependent design. These specifications make it hard to design an effective PAPR precoder. In this chapter, we introduce a MIMO-OFDM system model which is equipped with a PAPR-reduction scheme. Furthermore, an interesting algorithm is proposed in this context to overcome the shortcomings of conventional PAPR reduction techniques.

6.2 Efficient Low-Peak OFDM

A new PAPR reduction technique based on the proper precoding of the OFDM data is proposed here which is explained throughout the rest of this chapter. Earlier a newly developed technique called CP-PTS was proposed in [18] which is adaptable for different beamforming schemes in standard point to point or multiuser MIMO systems. In this technique, the OFDM subcarriers are grouped into blocks and the phase of each block is changed in a manner similar to the PTS method but without the drawback of sending explicit side information. As long as each block is multiplied with only one phase coefficient, the receiver will perceive this as a channel effect and will compensate for it during the channel equalization process [89]. An extension of CP-PTS to MIMO-OFDM systems is introduced in [90]. In both cases, a sequential quadratic programming (SQP) algorithm is used to solve the phase optimization problem. The computational complexity of this algorithm can be prohibitive for high data rate and/or low latency communication links. The PAPR weights need to be determined again for every OFDM data block, hence the underlying algorithm should be sufficiently efficient to enable real-time processing.

In this section, the same configuration as CP-PTS is used but instead of solving a non-convex optimization problem, an alternative problem formulation is proposed based on a cost function used in constant modulus algorithms (CMAs). Accordingly, the block-iterative SDCMA algorithm [91] is used to find the precoding PAPR weights. The resulting computational complexity is linear in the number of subcarriers. Furthermore, to make sure that the BER performance of the system

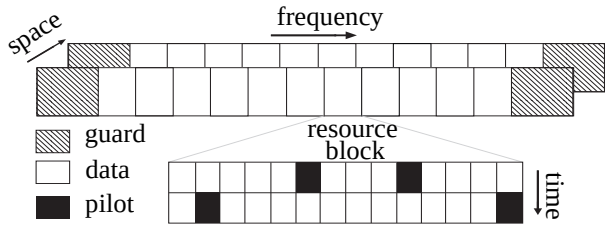


Figure 6.3: Data structure of an OFDM block for a MIMO-OFDM/A downlink.

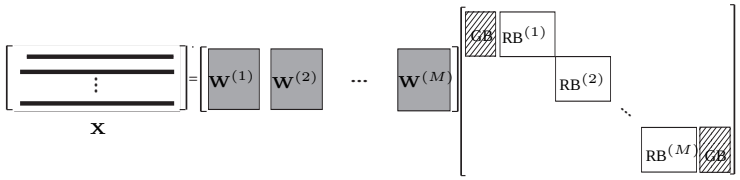


Figure 6.4: Beamformed MIMO transmit data in frequency domain.

is not affected by the PAPR precoding an additional constraint is appended to the CMA objective function which requires the weights to be on the unit circle. Like CP-PTS, the proposed technique is transparent to the receiver; this means that it only affects the base station (BS) and it does not require any signal processing in the mobile station (MS).

The proposed method does not function if the channel estimation algorithm assume the fact that channel coefficients change smoothly over the complete OFDM block. However, this assumption is not valid in the modern multiuser systems based on resource block assignment [89, 92].

6.2.1 Transmit Signal Model for the PAPR Precoding Scheme

Similar to [90] we consider a generic MIMO-OFDM/A downlink scenario with one base station (BS) employing M_t antennas. An OFDM block with N subcarriers is transmitted from each antenna. The N subcarriers include N_u useful subcarriers surrounded by two guard bands with zero energy. The useful subcarriers are further grouped into M resource blocks (RB) each consisting of $N_b = N_u/M$ subcarriers. Data of one or more users is placed in these resource blocks and mapped into the space-time domain using an inverse discrete Fourier transform (IDFT) and space-time block coding (STBC). To allow channel estimation at the receivers (mobile stations), each resource block also contains several pilot subcarriers that act as training symbols. The transmit signal model is illustrated in Fig. 6.3. It is compatible with the WiMAX standard [92].

Let us first describe the MIMO transmit data model in the frequency domain;

for simplicity we consider only a single time block from now on. The data in the q-th resource block is a matrix $\mathbf{D}^{(q)} \in \mathbb{C}^{M_t \times N_b}$, it is premultiplied with a corresponding beamforming matrix $\mathbf{W}^{(q)} \in \mathbb{C}^{M_t \times M_t}$, $q = 1, \cdots, M$, resulting in transmit sequences $\mathbf{X}_F^{(q)} = \mathbf{W}^{(q)H}\mathbf{D}^{(q)}$. Together with guard intervals, they are collected in a matrix $\mathbf{X}_F \in \mathbb{C}^{M_t \times N}$, where the M_t rows of this matrix represent the N symbols to be transmitted from the M_t antennas. The data model is

$$\mathbf{X}_F = \mathbf{W}^H \mathbf{D}\,, \tag{6.1}$$

where $\mathbf{W} = [\mathbf{W}^{(1)H}, \cdots, \mathbf{W}^{(M)H}]^H$, and $\mathbf{D} \in \mathbb{C}^{MM_t \times N}$ is a block-diagonal matrix with structure as in Fig. 6.4, which includes guard intervals as well. Matrix $\mathbf{X}_F$ represents the spatial data in the frequency domain i.e., MIMO without considering the PAPR precoding and OFDM modulation.

The time-domain MIMO-OFDM transmit data model is obtained by taking the IDFT of the beamformed data matrix $\mathbf{X}$, resulting in

$$\mathbf{X}_T = \mathbf{X}\mathbf{F}^H = \mathbf{W}^H\mathbf{D}\mathbf{F}^H\,, \tag{6.2}$$

where $\mathbf{F}^H \in \mathbb{C}^{N \times N}$ denotes the IDFT matrix, and $\mathbf{X}_T \in \mathbb{C}^{M_t \times N}$ contains the resulting transmit OFDM sequences for each of the M_t antennas. Let us further denote the time-domain data matrix $\mathbf{B} = \mathbf{D}\mathbf{F}^H$; this is a full matrix. Accordingly, the beamformed OFDM block can be expressed as

$$\mathbf{X}_T = \mathbf{W}^H\mathbf{B}\,. \tag{6.3}$$

Denote the total power (or energy) in the data matrix $\mathbf{D}$ by $P_d := \|\mathbf{D}\|_F^2 = \|\text{vec}(\mathbf{D})\|^2 =: \alpha N_t$, where $N_t = NM_t$. Function vec$(\mathbf{D})$ creates a column vector whose elements are the columns of the matrix $\mathbf{D}$. N_t is the total number of subcarriers or samples to be sent from all M_t antennas, and α is defined as the average transmit power per sample (including the zero power guard bands). If we assume that the beamforming matrix $\mathbf{W}$ consists of *orthonormal* matrices $\mathbf{W}^{(q)}$, then applying beamforming and the IDFT does not change the total transmit power.

6.3 Constant Modulus Algorithm for PAPR Reduction

6.3.1 Introduction

The IDFT operation in (6.2) leads to a large dynamic range of the resulting time-domain OFDM signal. PAPR is a common metric to measure the distortion caused

by probable high peak of the OFDM signal and for a MIMO-OFDM block $\mathbf{X}_T$ we define

$$\text{PAPR}(\mathbf{X}_T) = \frac{\alpha N_t \|\text{vec}(\mathbf{X}_T)\|_\infty^2}{\|\text{vec}(\mathbf{X}_T)\|_2^2} \,. \tag{6.4}$$

Clearly, the lowest PAPR is achieved for a constant modulus signal, for which the infinity norm is equal to the average power of the sequence.

The main idea in [18, 90] is to design a precoding matrix to transform the OFDM symbols in $\mathbf{X}_T$ to a favorable signal $\mathbf{S}$ with lower PAPR (ideally a constant modulus signal). This precoding matrix $\mathbf{\Omega}$ needs to fulfill the following requirements:

1. Reduce the dynamic range of the OFDM block,

2. Preserve the beamforming property,

3. Be transparent to the receiver,

4. Not impact the bit error rate (BER).

To satisfy the second and third constraint, we are allowed to premultiply each resource block, $\mathbf{D}^{(q)}$, with a diagonal scaling matrix $\mathbf{\Omega}^{(q)}$. To the receiver, this will appear as a fading channel effect. To not affect the BER, the scaling should be unimodular (phase only). Equivalently, a diagonal (unimodular) precoding matrix $\mathbf{\Omega} \in \mathbb{C}^{MM_t \times MM_t}$ is applied to $\mathbf{D}$. The resulting MIMO-OFDM transmit matrix (replacing $\mathbf{X}_T$) is

$$\mathbf{S} = \mathbf{W}^H \mathbf{\Omega} \mathbf{D} \mathbf{F}^H \,. \tag{6.5}$$

If we define $\boldsymbol{\omega} = \text{vecdiag}(\mathbf{\Omega})$, then the PAPR reduction problem is to design $\boldsymbol{\omega}$ as

$$\min_{\boldsymbol{\omega}} \ \|\text{vec}(\mathbf{S})\|_\infty^2 \quad \text{s.t.} \quad \|\text{vec}(\mathbf{S})\|_2^2 = P \tag{6.6}$$

where $P = \alpha N_t$ is a fixed total transmit power. This problem is not convex because nonlinear equality constraints can rarely be expressed in a convex form. The approach in [18, 90] was to solve a series of quadratic convex subproblems iteratively. Although this does not solve the original problem in (6.6) exactly, the results were excellent compared to other techniques, and attractive as the method is transparent to the receiver and does not distort the transmit signals. Unfortunately, this approach is yet too complex for real time applications.

6.3.2 Formulation as a Constant Modulus Problem

Using properties of Kronecker products, we can rewrite $\mathbf{S}$ in (6.5) as

$$\mathbf{s} = \text{vec}(\mathbf{S}) = (\bar{\mathbf{B}} \circ \mathbf{W})^H \text{vecdiag}(\mathbf{\Omega}) =: \mathbf{A}\boldsymbol{\omega}\,, \tag{6.7}$$

where $\mathbf{A} \in \mathbb{C}^{N_t \times MM_t}$, $\mathbf{DF}^H = \mathbf{B} \in \mathbb{C}^{MM_t \times N}$, $\bar{\mathbf{B}}$ denotes the complex conjugate of $\mathbf{B}$, and $\circ$ denotes the Khatri-Rao product (column-wise Kronecker product). The vecdiag$(\mathbf{D})$ creates a column vector whose elements are the main diagonal of the matrix $\mathbf{D}$. The optimization problem (6.6) becomes

$$\min_{\boldsymbol{\omega}} \ \|\mathbf{A}\boldsymbol{\omega}\|_\infty^2 \quad \text{s.t.} \quad \|\mathbf{A}\boldsymbol{\omega}\|_2^2 = \alpha N_t \tag{6.8}$$

We now propose an alternative formulation of this problem, by replacing the infinity norm by the average deviation of the OFDM block from a constant modulus signal. Ideally, the resulting $\mathbf{S}$ will be close to a CM signal, and hence have close-to-optimal PAPR. The corresponding cost function is

$$J(\boldsymbol{\omega}) = \|\mathbf{A}\boldsymbol{\omega} \odot \overline{(\mathbf{A}\boldsymbol{\omega})} - \alpha \mathbf{1}_{N_t}\|_2^2 = \sum_{n=1}^{N_t} (\boldsymbol{\omega}^H \mathbf{a}_n \mathbf{a}_n^H \boldsymbol{\omega} - \alpha)^2\,. \tag{6.9}$$

Here, the vector $\mathbf{a}_n^H,\ n = 1, \cdots, N_t$ represents the n-th row of matrix $\mathbf{A}$, the column vector $\mathbf{1}_{N_t}$ is a vector with all entries equal to 1 and dimension N_t, and $\odot$ denotes the Schur-Hadamard product (pointwise multiplication).

This formulation is similar to the well-known "CMA(2,2)" cost function for adaptive blind equalization or blind beamforming, and can be solved efficiently using available iterative algorithms. The matrix $\mathbf{A}$ plays the role of the data matrix in the usual CMA context, whereas $\boldsymbol{\omega}$ plays the role of the beamforming vector. The original CMA cost function is expressed in terms of an expectation operator; the present "deterministic" formulation is similar to the Steepest Descent CMA (SDCMA) in [91].

6.3.3 Steepest-Descent CMA (SDCMA)

The SDCMA is a block-iterative algorithm in which we act on the full data matrix $\mathbf{A}$ and update $\boldsymbol{\omega}$ until it converges. The derivation of the block SDCMA is straightforward when the statistical expectation in original formula in [91] is replaced by an average over a block. For the i-th iteration, we start from the current estimate

$\boldsymbol{\omega}^i$ and compute:

$$\begin{aligned}
\hat{\mathbf{s}}^i &= \mathbf{A}\boldsymbol{\omega}^i && (6.10)\\
\mathbf{e}^i &= (\hat{\mathbf{s}}^i \odot \bar{\hat{\mathbf{s}}}^i) - \alpha \mathbf{1}_{N_t} && (6.11)\\
\hat{\mathbf{s}}_e &= \hat{\mathbf{s}}^i \odot \mathbf{e}^i && (6.12)\\
\boldsymbol{\omega}^{i+1} &= \boldsymbol{\omega}^i - \mu \nabla J(\boldsymbol{\omega}^i) = \boldsymbol{\omega}^i - \mu \mathbf{A}^T \hat{\mathbf{s}}_e \,. && (6.13)
\end{aligned}$$

Here, μ is a suitable step size, and $\hat{\mathbf{s}}_e$ is the update error. The maximal step size μ could be defined as a scale independent parameter in relation to the signal power in $\mathbf{A}$. To keep the solution unchanged as $\mathbf{A}$ scales, μ needs to be divided by factor α^2, $\mu = \mu'/\alpha^2$. For convergence, the algorithm is initialized with $\boldsymbol{\omega}^0 = \mathbf{1}$ (although other choices are possible). The algorithm should be run until the cost function $J(\boldsymbol{\omega})$ converges; in practice convergence is fast and the algorithm is run for a fixed small number of iterations.

To satisfy the power constraint in (6.6), we can simply scale the resulting $\boldsymbol{\omega}$ after convergence. If $\mathbf{s} = \mathbf{A}\boldsymbol{\omega}$ is indeed a constant modulus signal, then $\|\mathbf{s}\|_2^2 = \alpha N_t$, and the power constraint is inherently already satisfied. Thus, the scaling is expected to be close to 1 and could be omitted in practice (it has no effect on the cost function PAPR($\mathbf{S}$)).

A difference with the standard CMA is that, here, a good solution does not necessarily exist. The usual application of CMA is for a linear combination of constant modulus sources for which, without noise, a perfect beamformer exists. The present situation could be said to correspond to a very noisy source separation situation. Note that, also for other methods, there are no existence results for PAPR reduction.

6.3.4 Unit-Circle CMA (UC-CMA)

In SDCMA, the computed $\boldsymbol{\omega}$ has no constraints and may have some small entries. These are equivalent to a (broad) null in the channel which will affect the BER performance. Ideally, we should restrict the entries of $\boldsymbol{\omega}$ to take only unimodular values: $\omega_m = e^{j\phi_m}$, $m = 1, \cdots, M$, and add this constraint to the optimization problem (6.8).

In order to restrict the solution to be on the unit circle, a normalization step is added to each iteration after (6.13):

$$\boldsymbol{\omega}^{i+1} = \boldsymbol{\omega}^{i+1} \oslash |\boldsymbol{\omega}^{i+1}| \qquad (6.14)$$

where $\oslash$ denotes pointwise division, and $|\cdot|$ takes the absolute value of each entry of the vector argument. This alternative updating algorithm is called Unit Circle CMA (UC-CMA) since (6.14) projects the solution of CMA to a unit circle at each iteration.

6.4 Computational Complexity

The complexity of the SDCMA algorithm in (6.13) is dominated by the matrix products $\mathbf{A}^T\hat{\mathbf{s}}_e$ and $\mathbf{A}\boldsymbol{\omega}^i$. The resulting complexity is approximately $2NMM_t^2$ per iteration (linear in the number of subcarriers). UC-CMA has the same complexity.

In conventional PTS [87], each resource block (sub-block in PTS context) is weighted with a phase shift in such a way that the summation of sub-blocks produce an OFDM sequence with a smaller PAPR [87]. The phase weights are selected by an exhaustive search among a discrete set of phases, and are sent as side information to the receiver. Accordingly, all combinations of the M available phase weights need to be calculated and then multiplied with an IDFT summation matrix, which has the same size as matrix $\mathbf{B}$. Finally, one sequence with the least PAPR metric is chosen with the corresponding phase weights. The complexity of the exhaustive search is calculated for the simplest set of only two phases $\{\pm 1 = e^{\pm j\pi/2}\}$ and M resource blocks as $2^M NM$ multiplications and 2^M comparisons. For CP-PTS , the complexity is $O(M^3)$; the exact expression for complexity is derived in [18].

6.5 Simulation Results

In WiMAX, one resource block spans $N_b = 14$ sub-carriers over two OFDM symbols in time, containing 4 pilots and 24 data symbols. For a 10MHz system, there are a total of $M = 60$ resource blocks [92]. In agreement with this WiMAX setting, the proposed PAPR reduction technique is simulated for an OFDM block of size $N = 1024$ including $MN_b = 840$ data subcarriers with QPSK modulation and 92 guard subcarriers at each end of the band. The number of MIMO transmit antennas is either $M_t = 1$, 2 or 4, as will be indicated. The various techniques are evaluated using the complementary cumulative density function (CCDF), which denotes the probability that the PAPR of a data block exceeds the argument of the function. To avoid the PAPR underestimation, The algorithm is run with four times oversampling so the number of the samples processed in the simulations is $N' = 4N_t$.

A total number of 10,000 OFDM blocks are randomly generated to produce the CCDF curves. For each block, a random complex fading channel is generated,

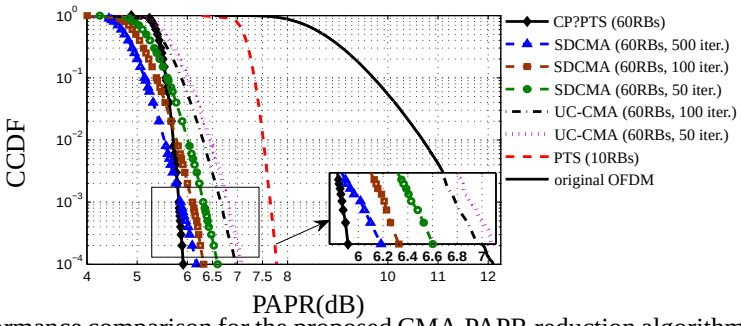


Figure 6.5: Performance comparison for the proposed CMA PAPR reduction algorithm for various number of iterations and $\mu' = 0.05$, CP-PTS with 5 iterations, and conventional PTS with phase alphabet $\{\pm 1\}$ and $M = 10$.

and the beamforming matrices $\mathbf{W}$ are chosen as the right singular vectors of these channel matrices.

In Fig. 6.5, the CCDF performance is shown for SDCMA (various number of iterations), UC-CMA (50 iterations), and compared to CP-PTS [90] and the standard PTS [87]. The latter algorithm is simulated only for $M = 10$ resource blocks due to prohibitive computational complexity for larger M. In this simulation, $M_t = 1$ transmit antenna. The simulations show that the proposed techniques attain a PAPR reduction of up to 6 dB. Although 50 iterations are sufficient for good performance, another 0.5 dB is gained by increasing this to 500 iterations. UC-CMA (50 iterations) is worse by about 0.5 dB. The PAPR reduction for PTS is worse by 1 to 2 dB. The previously proposed CP-PTS outperforms PTS and SDCMA with 50 iterations, however a similar gain is reached by SDCMA with a larger number of iterations. Moreover, the CCDF curves in Fig. 6.5 show the superior performance of SDCMA in 90% and 99.9% of OFDM blocks in 50 and 500 iterations respectively, comparing to the CP-PTS.

The empirical CDF of $|\omega_q|$ values in SDCMA indicates the Rayleigh distribution of PAPR weights which affect the BER performance of the system. Fig. 6.6 shows the BER versus SNR curves for the QPSK-OFDM system without PAPR reduction in a randomly generated Rayleigh fading and AWGN channels compared to the scenarios that SDCMA and UC-CMA weights are applied at the transmitter. In SDCMA and UC-CMA the channel is assumed to be AWGN and the received vector is divided by $\boldsymbol{\omega}$ to equalize the PAPR weights. Where, in Rayleigh fading channel the received vector is divided by the frequency domain channel coefficients. in both cases, the perfect channel recovery is assumed. From Fig. 6.6, the effect of non-modified SDCMA is analogous to a Rayleigh fading channel in terms

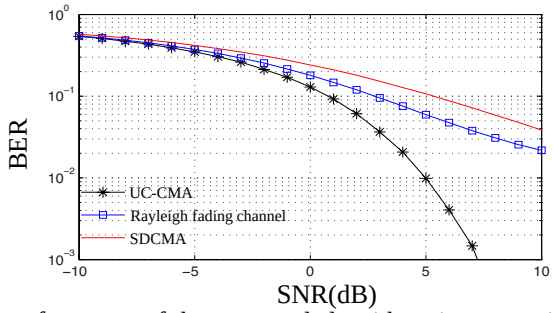


Figure 6.6: BER performance of the proposed algorithms in comparison with AWGN and Raleigh fading channels for single antenna QPSK-OFDM system of size $N = 1024$ and $M = 60$.

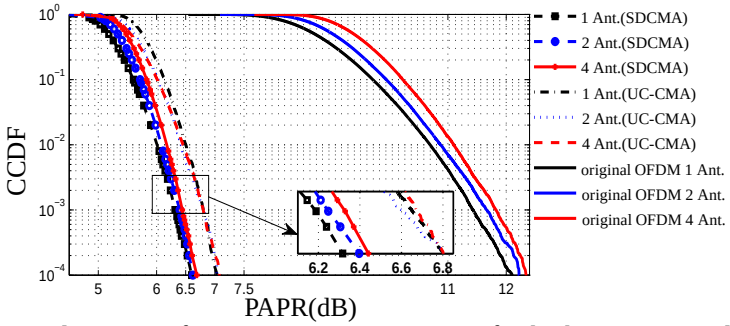


Figure 6.7: PAPR reduction performance in MIMO-OFDM for both SDCMA and UC-CMA with 50 iterations and $\mu' = 0.05$.

of BER performance so the same error correcting codes used for a fading channel can be applied here. As expected the UC-CMA does not influence the BER performance. This motivates the use of UC-CMA technique.

Fig. 6.7 shows the performance of SDCMA and UC-CMA for various number of transmit antennas, $M_t = 1, 2, 4$, and 50 iterations. It is seen that the performance is not a strong function of the number of antennas; small improvements are seen due to more available phase weights or degrees of freedom in the optimization problem.

To demonstrate computational complexity, Matlab runtimes on a standard 2011 laptop are shown in Fig. 6.8 as a function of M (number of resource blocks). In this simulation, $M_t = 1$. It is seen that the proposed CMA algorithms (using 50 iterations) are about a factor 50 faster than CP-PTS, whereas the complexity of PTS is growing exponentially with the number of resource blocks and is quickly not feasible anymore.

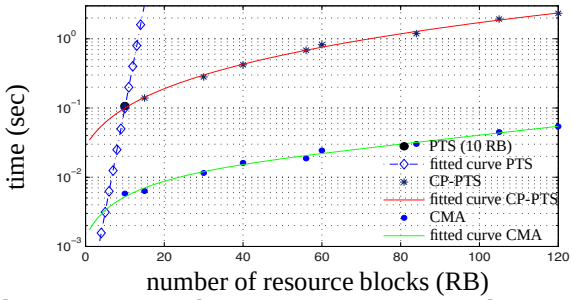


Figure 6.8: Complexity comparison between CMA, CP-PTS and PTS using Matlab runtime evaluation.

6.6 Conclusion

In this chapter a MIMO-OFDM data model is introduced and the famous PAPR problem is investigated in this context. Also, an efficient and effective PAPR reduction algorithm is developed which does not affect the bandwidth efficiency and the receiver design while performs acceptably by means of Matlab simulations. This is one step towards the design of a realistic signal processing algorithm which caters for hardware imperfections.

In the next chapter, another aspect of hardware limitations in the design of precoders in MIMO systems is considered. First of all the power restriction is defined per antenna since each antenna has its own transmitter and amplifier. Also, there are restrictions on the number of available RF chains that can be installed in the system.

Chapter 7

Joint Precoding and Antenna Selection for Multiuser MIMO

Modern multiple-input multiple-output (MIMO) communication networks employ many antennas, and the optimal use of them is a complex problem subject to several practical constraints, such as a maximum number of radio frequency (RF) chains, leading to an antenna selection problem. We consider the downlink of a spatial multiplexing multi-user multiple-input single-output (MU-MISO) system, and propose a unified framework for the joint optimization problem of antenna selection, transmit beamforming, and power allocation with realistic yet complicated to handle, per antenna power constraints. In its original formulation, this is a non-convex and NP-hard problem which poses no elegant solution. The proposed solution is based on appending a group sparsity inducing regularizer (GSIR) which makes it amenable for convex optimization via non-trivial relaxation techniques. Extensive simulation results show a negligible difference in performance between this approach and the optimal (exhaustive search) selection, while the computational complexity is significantly smaller compared to the optimal approach.

This chapter is submitted as “Convex Optimization for Joint Antenna Selection and Precoding”, Signal Processing, Elsevier, April 2016.

7.1 Introduction

7.1.1 Problem Context

The advent of high carrier frequency wireless systems such as the 60 GHz short range standards IEEE.802.15c and IEEE.802.11ad opens up the possibility to radically increase the number of antennas beyond conventional multiple-input multiple-output (MIMO) systems, and perhaps even to integrate all of them on a single chip [33]. A critical factor in increasing the number of antennas is the cost of the radio frequency (RF) chain. Antenna selection techniques are needed to determine the optimal subset of antennas to be connected to the available RF chains, based on the actual channel conditions. Until now, most research was dedicated to single-user (SU) MIMO (i.e., point to point MIMO), whereas antenna selection for multi-user (MU) MIMO has remained rather unexplored. This chapter will focus on MU-MIMO at the downlink transmitter.

A good survey on different criteria and sub-optimal techniques for antenna selection is given in [93] which classifies selection techniques for both the transmitter (Tx) and the receiver (Rx) and gives asymptotic performance results. For SU-MIMO antenna selection at the receiver, it was shown that a subset of optimally selected antennas can achieve the same capacity as a full system [94]. The high complexity of the combinatorial problem has prompted for "greedy" selection techniques [95, 96, 97], whereas [98] proposed a convex optimization approach. In [97], the convergence of greedy algorithms is proved for sub-modular problems with objective functions including the capacity maximization subject to the total power constraint. SU-MIMO antenna selection at the transmitter was studied in [99, 100, 23, 101].

While in SU-MIMO systems the channels are considered known at both sides, MU-MIMO systems consist of a basestation (or network of basestations), and users which are individual entities that in general cannot co-operate. Despite fundamental similarities in selection algorithms, antenna selection techniques for SU-MIMO systems need to be greatly adapted for MU-MIMO due to the resulting differences. In particular for the downlink, antenna selection needs to be combined with other techniques such as beamforming (BF) for interference cancelation and power allocation for the best distribution of available power among streams.

In this chapter, we propose a joint approach to solve for the precoding (beamforming and power allocation) and antenna selection matrices simultaneously, and we refer to this as joint antenna selection and precoding (JASP). For conciseness, we focus on downlink spatial multiplexing MU-MISO systems, where each re-

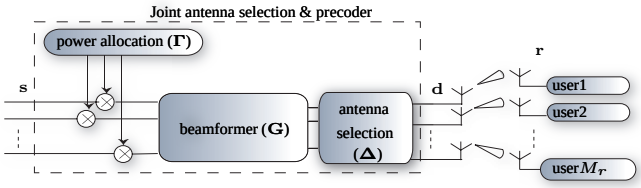


Figure 7.1: Block diagram of MU-MISO link.

ceiver (user) has a single antenna and receives an independent data stream. Fig. 7.1 shows schematically the MU-MISO system of interest. Stacking the received data in a vector $\mathbf{r}$, the data model is of the form

$$\mathbf{r} = \mathbf{H}\mathbf{d} + \mathbf{n}\,, \qquad \mathbf{d} = \mathbf{W}\mathbf{s} \tag{7.1}$$

where the matrix $\mathbf{H}$ contains the channel state information (CSI, considered known[1]), $\mathbf{d}$ is the transmitted data, and $\mathbf{n}$ is additive noise. The raw input data $\mathbf{s}$ is mapped to $\mathbf{d}$ via a matrix $\mathbf{W}$ which consists of three factors: a diagonal selection matrix $\mathbf{\Delta}$, a beamforming matrix $\mathbf{G}$, and a power allocation matrix $\mathbf{\Gamma}$. We write $\mathbf{W} = \mathbf{\Delta}\mathbf{W}'$, where $\mathbf{W}'$ is the usual precoding matrix ($\mathbf{W}' = \mathbf{G}\mathbf{\Gamma}^{\frac{1}{2}}$).

In the plain precoding problem, $\mathbf{\Delta}$ is absent, and $\mathbf{G}$ is designed either as a zero forcing (ZF) precoder ($\mathbf{HG}$ is an identity matrix) or as the minimum mean square error (MMSE) precoder[2]. Note that here the formulation of SINR balancing beamformer is equivalent to MMSE beamformer therefore it is referred to as such, however this is not generally the case. The parameter $\mathbf{\Gamma}$ is designed either by posing a total power constraint (TPC) or a per antenna power constraint (PAPC). We consider PAPC as a more realistic constraint as each antenna power is limited by its corresponding RF chain.

Including the antenna selection matrix $\mathbf{\Delta}$, the joint precoding and antenna selection problem in its general form is to find $\mathbf{W}$ such that

$$\mathbf{W}^{\text{opt}} = \arg \max f(\mathbf{W}) \qquad \text{s.t. constraints} \tag{7.2}$$

where $f(\mathbf{W})$ is a general performance measure, such as capacity. This problem is combinatorial and NP-hard for almost all precoding strategies. Indeed, the optimal

[1]CSI can be acquired by sequential use of the available RF chains and it needs to be updated each time the channel decorrelates, hence the coherence time of the channel needs to be relatively large for the estimate of the channel to be valid.

[2]When there is no interference management constraint, the joint antenna selection and precoding problem is reduced to the sparse power allocation problem. A very similar problem is addressed in [98].

solution $\mathbf{W}^{\text{opt}}$ corresponds to a subset of antennas whose $\mathbf{W}'$ maximizes $f(\mathbf{W}')$ among all possible combinations of antennas. To avoid this enumeration, the disjoint antenna selection approach finds the best antennas ignoring the precoder, e.g., by maximizing $f(\mathbf{\Delta})$ for throughput and independently finding the optimal $\mathbf{W}'$ later for that set of selected antennas (fixing $\mathbf{\Delta}$). This leads to a sub-optimal performance as will be verified here via simulations. Alternatively, sub-optimal sequential optimization was used by fixing $\mathbf{\Delta}$ or $\mathbf{W}'$, successively, and optimizing over the other variable [102, 103]. Other papers consider a known (fixed) power allocation [104, 105] or consider antenna selection at the receiver instead [106, 103, 102].

In this chapter, we consider ZF and MMSE precoders, as they are the most common yet effective linear precoders, subject to per antenna power constraints. We provide sufficient examples for different possible combinations of precoders. However, none of these precoders has closed form solutions [107] nor are they convex in their original form. We will follow the primal approach for solving for the precoder, using iterative convex algorithms developed for quadratically constrained quadratic problems (QCQP) [108]. The most important works for ZF precoding can be listed as [109, 110, 107] whereas MMSE precoding has been studied for [108, 111, 112, 113]. The antenna selection problem can then be introduced by appending a convex (group) sparsity inducing regularizer to the plain precoding problem.

Convex group sparsity inducing regularizers have gained considerable attention [114] as they leverage a relaxation for many old NP-hard selection problems. Precoding is no exception in this regard [115, 116], e.g., in [117] synthesizing a sparse beam-pattern for linear and planar arrays is considered using an ℓ_1-norm penalty. A joint beamforming and base station selection (clustering) in [118] is considered where an iterative group Lasso minimization approach is proposed to for sum-rate utility maximization. In [119] we proposed a convex formulation based on the ℓ_1-norm penalty to handle a simple JASP problem i.e, maximizing the capacity subject to a total power constraint. Joint antenna selection and MMSE precoding is an example of a QCQP with sparsity requirements that is addressed in [120] for broadcast beamforming where it is formulated as a semi-definite program (SDP) using the squared ℓ_1-norm as a sparsity inducing regularizer. Later in [121] the multicast extension was introduced which is the most relevant work for this chapter.

7.1.2 Contributions

- We propose a convex formulation for the JASP problem from a unified vantage point, i.e., different from the existing literature in the sense that we try to solve (7.2) directly rather than finding $\mathbf{\Delta}$ and $\mathbf{W}'$ sequentially or disjointly. The original form of the JASP problem in (7.2) is non-convex w.r.t. the precoder matrix variable, and combinatorial w.r.t. the selection variable.

- Sparse regularization for antenna selection has been proposed recently in the literature. What is missing is a generic problem formulation which describes the joint antenna selection and linear precoding problem in a unified manner regardless of the choice of the sparsity regularizer and the precoding strategy.

- We show that using the proposed regularizer in [121] for SINR balancing precoder, is unnecessarily complicated for the proposed MMSE (SINR balancing) problem in this chapter. In fact, SINR constraints can be readily reformulated as a linear matrix inequality (LMI) or even second order cone (SOC) constraints as proposed in [108, 112]. We show that the plain group Lasso regularizer can be used directly to formulate the sparse MMSE precoding problem. Indeed, the squared transformation as in [121] is only necessary for precoding schemes involving the capacity expression which can not be made convex using the plain group Lasso regularizer. This can be considered as new insight to an existing technique.

- The proposed sparsity regularizations can be used for other purposes as well, e.g., the convex formulation allows to find the minimum number of antennas that is sufficient for a certain quality of service. This is different from antenna selection since the number of required antennas is not given a priori and yields an optimization variable itself. We refer to this approach as the antenna reduction technique (ART).

- We claim that the conventional greedy approaches for antenna selection are not effective when per antenna power constraints are involved since the capacity function is not a sub-modular function anymore of the set of antenna elements in the sense that the effectiveness of an additional antenna element is not diminishing as the number of antennas increases. Hence, popular greedy algorithms such as [95] do not converge to a point close to the optimal solution [97]. This is an important observation which justifies the necessity for developing a “close to optimal” approach for the systems which aim at

capacity maximization with per antenna power constraints.

7.1.3 Notation

Bold upper case and bold lower case symbols indicate matrices and vectors, respectively. The conjugate transpose and transpose of a matrix $\mathbf{A}$ are denoted as $\mathbf{A}^H$ and $\mathbf{A}^T$. $A(i,j)$ is the (i,j)-th entry of matrix $\mathbf{A}$, and a_j is the jth entry of a vector $\mathbf{a}$. The jth column of $\mathbf{A}$ is denoted as $\mathbf{a}(:,j)$, and its ith row is denoted as $\mathbf{a}^T(i,:)$, where $\mathbf{a}(i,:)$ is a column vector. Partitioning $\mathbf{A}$ into blocks, $\mathbf{A}_{i,j}$ denotes the (i,j)-th submatrix. We use $\mathbf{A}_j$ to represent the jth matrix of a set of matrices which share the same structure, and $\|\mathbf{a}\|_p$ denotes the ℓ_p-norm of a vector. $\mathrm{Tr}(\mathbf{A})$ is the trace of $\mathbf{A}$, and $\mathrm{vec}(\mathbf{A})$ vectorizes $\mathbf{A}$ by stacking the columns of the matrix. $\mathbf{I}_N$ is the identity matrix of size N, $\mathbf{1}_N$ and $\mathbf{1}_{N\times M}$ are an $N\times 1$ vector and $N\times M$ matrix consisting of all ones, $\mathbf{0}_N$ and $\mathbf{0}_{N\times M}$ are an $N\times 1$ vector and $N\times M$ matrix of zero entries.

7.2 System Model

Consider a spatial multiplexing MU-MISO system as in (7.1), which we rewrite as

$$\mathbf{r} = \mathbf{HWs} + \mathbf{n}\,, \quad \mathbf{W} = \boldsymbol{\Delta}\mathbf{W}' \tag{7.3}$$

with M_r users (streams) each with a single antenna, and an access point with M_t transmit antennas which incorporates the antenna selection scheme. The received data vector over all users is $\mathbf{r} \in \mathbb{C}^{M_r\times 1}$. The matrix $\mathbf{H} \in \mathbb{C}^{M_r\times M_t}$ is the MU channel which is considered known. The raw data $\mathbf{s} \in \mathbb{C}^{M_r}$ is precoded by $\mathbf{W} = \boldsymbol{\Delta}\mathbf{W}'$, where $\boldsymbol{\Delta}$ is the antenna selection matrix and $\mathbf{W}' = \mathbf{G}\boldsymbol{\Gamma}^{1/2} \in \mathbb{C}^{M_t\times M_r}$ is the plain precoding matrix involving the beamforming matrix $\mathbf{G}$ and the diagonal power allocation matrix $\boldsymbol{\Gamma}$. The selection matrix $\boldsymbol{\Delta} = \mathrm{diag}(\boldsymbol{\delta})$ is a diagonal matrix, where the entries on the diagonal are zero for eliminated antennas and one for selected antennas, so $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_{M_t}] \in \{0,1\}^{M_t}$. The noise $\mathbf{n}$ is assumed to be a complex zero-mean Gaussian random vector. The covariance matrix of the input signal $\mathbf{s}$ to the precoder block is assumed to be an identity matrix, i.e., $E\{\mathbf{ss}^H\} = \mathbf{I}_{M_r}$, and the design parameters in the model are $\boldsymbol{\delta}$ and $\mathbf{W}'$.

Table 7.1: Linear precoding design criteria.

power	PAPC:	$\forall i: \quad \|\delta_i \mathbf{w}(i,:)\|_2^2 \leq p_i^*$
beamforming	ZF:	$\forall j: \quad \sum_{l \neq j} \|\mathbf{h}^H(j,:)\boldsymbol{\Delta}\mathbf{w}'(:,l)\|^2 = 0$
	MMSE:	$\forall j: \quad \gamma_j \geq \gamma_j^*$
quality	total power:	$f_{\mathrm{P}} = \mathrm{Tr}(\mathbf{W}'\boldsymbol{\Delta}\mathbf{W}'^H) = \sum_i \|\delta_i \mathbf{w}(i,:)\|_2^2$
	capacity:	$f_{\mathrm{C}} = \sum_{j=1}^{M_r} \log(1+\gamma_j)$

For the jth user, (7.3) can be written as

$$\begin{aligned} r_j &= \sum_{l=1}^{Mr} \mathbf{h}^H(j,:)\boldsymbol{\Delta}\mathbf{w}'(:,l)s_l + n_j && (7.4)\\ &= \mathbf{h}^H(j,:)\boldsymbol{\Delta}\mathbf{w}'(:,j)s_j + \sum_{l \neq j} \mathbf{h}^H(j,:)\boldsymbol{\Delta}\mathbf{w}'(:,l)s_l + n_j && (7.5)\\ &= \mathbf{h}^H(j,:)\boldsymbol{\Delta}\mathbf{w}'(:,j)s_j + \mathbf{h}^H(j,:)\boldsymbol{\Delta}\bar{\mathbf{W}}'_j\bar{\mathbf{s}}_j + n_j && (7.6) \end{aligned}$$

where $\bar{\mathbf{W}}'_j \in \mathbb{C}^{M_t \times (M_r-1)}$ and $\bar{\mathbf{s}}_j \in \mathbb{C}^{M_r-1}$ are the precoding matrix and data vector corresponding to all users except j, and (with some abuse of notation) $\mathbf{h}^H(j,:)$ is the jth row of $\mathbf{H}$. The first and second term in (7.6) are the desired signal and the interference of user j, respectively. The SINR for user j can therefore be expressed as

$$\gamma_j = \frac{|\mathbf{h}^H(j,:)\boldsymbol{\Delta}\mathbf{w}'(:,j)|^2}{\sum_{l \neq j} |\mathbf{h}^H(j,:)\boldsymbol{\Delta}\mathbf{w}'(:,l)|^2 + \sigma_j^2}\,. \tag{7.7}$$

Clearly the SINR of each user depends on all precoding vectors and selected antennas.

Table 7.1 shows common design criteria in terms of the precoder for the power allocation subject to PAPC. Beamforming is used for interference management, and as beamforming strategies we consider ZF (cancel all interference so $\mathbf{HW}$ is diagonal) and MMSE (guarantee a certain SINR for each user). Finally, the system performance (quality) can be formulated as a total sum-rate of the system (capacity) or total transmit power as follows

$$\begin{aligned} f_{\mathrm{C}}(\mathbf{W}', \boldsymbol{\delta}) &= \sum_{j=1}^{M_r} \log(1+\gamma_j) \\ f_{\mathrm{P}}(\mathbf{W}', \boldsymbol{\delta}) &= \mathrm{Tr}(\mathbf{W}'\boldsymbol{\Delta}\mathbf{W}'^H) = \sum_i \|\delta_i \mathbf{w}(i,:)\|_2^2 \end{aligned} \tag{7.8}$$

In turn, the parameter p_i^* denotes the maximum transmit power for the ith antenna, and $\gamma_j{}^*$ denotes the minimum SINR for user j, which both are used in the constraints.

The precoding problem can be defined in several ways depending on the restrictions and flexibilities of the system, e.g., as maximizing the quality subject to one of the beamforming criteria and a power constraint (*performance maximization*), or alternatively as minimizing the transmit power subject to one of the beamformingcriteria and a quality constraint (*power minimization*). We only consider some of these possible combinations in this chapter covering the necessary underlying optimization algorithms.

Regarding the power constraints, TPC is a relaxation of the PAPC and any feasible solution for the latter is feasible for the former too. We focus on PAPC in this chapter as this has relevance in practice, yet harder to deal with. Equal rates (fairness) are obtained by maximizing the minimum SINR among all users, which is the capacity achieving strategy in MMSE precoding [112]. This is a quasi convex problem for MMSE precoders due to a presence of the fractional term in the objective [67]. We do not consider that problem in this chapter.

Instead, we will define and elaborate three generic problems. The first problem, denoted as $\mathcal{P}$, is the capacity maximization problem subject to PAPC, ZF and antenna selection constraints. Also, the converse problem of power minimization subject to capacity, PAPC and antenna selection constraint is studied. These are referred to as *ZF-JASP* throughout this chapter. The former is specified as a capacity maximization problem and the latter as power minimization problem. The second problem, denoted as $\mathcal{P}'$, is the MMSE precoder from a power minimization subject to the PAPC, SINR and antenna selection constraints named as *MMSE-JASP* problem. The third problem is the ART problem, which is denoted by $\mathcal{P}''$ and introduced considering two set of constraints: 1) PAPC and SINR constraints 2) PAPC, ZF and capacity constraints, referred to as *MMSE-ART* and *ZF-ART*, respectively. We believe these examples sufficiently provide all the required transformation and relaxation techniques, thus extensions for other possible precoders are directly possible.

7.3 Problem Formulation: JASP

We consider in this section the formulation of the JASP problems including ZF-JASP and MMSE-JASP. In its general form, the JASP problem is written as

$$\boxed{\begin{array}{cc} \text{maximize} & f(\mathbf{W}', \boldsymbol{\delta}) \\ \mathbf{W}' \in \mathbb{C}^{M_t \times M_r}, \boldsymbol{\delta} \in \{0,1\}^{M_t} & \\ \text{subject to} & \mathcal{C}(\mathbf{W}', \boldsymbol{\delta}) \end{array}} \tag{7.9}$$

where the optimization variables are the precoding matrix $\mathbf{W}'$ and the selection vector $\boldsymbol{\delta}$. As shown in Table 7.1, the objective function $f(\mathbf{W}', \boldsymbol{\delta})$ can be either $-f_{\mathrm{P}}(\mathbf{W}', \boldsymbol{\delta})$ for power minimization or $f_{\mathrm{C}}(\mathbf{W}', \boldsymbol{\delta})$ for capacity maximization.

The optimization problem in (7.9) (representing both ZF-JASP and MMSE-JASP) is not tractable in any of its aforementioned forms as finding the optimal pair $(\mathbf{W}', \boldsymbol{\Delta})$ leads to a mixed-integer problem. This is equivalent to solving the plain precoding problem while enumerating over a set that includes all possible matrices $\boldsymbol{\Delta}$, which has $\binom{M_t}{L_s}$ members, L_s being the number of available RF chains. This is a combinatorial problem w.r.t. $\boldsymbol{\Delta}$, and in most cases is NP-hard w.r.t. $\mathbf{W}'$. To enable the use of efficient convex optimization techniques, we introduce a relaxation to this problem. The number of relaxation steps is indicated in the subscript, e.g., $\mathcal{P}_0$ means no relaxation, $\mathcal{P}_1$ shows one relaxation step and so on.

7.3.1 Joint Antenna Selection and Precoding (JASP) Problem

In order to introduce the joint problem of precoding and antenna selection, the two optimization variables are merged by introducing a new variable $\mathbf{W} = \boldsymbol{\Delta}\mathbf{W}'$ which conveys both the precoding and antenna selection. Thus, we can write the joint antenna selection and precoding optimization problem in terms of $\mathbf{W}$, where the objectives in (7.8) become

$$\begin{array}{rcl} f_{\mathrm{C}}(\mathbf{W}) & = & \sum_{j=1}^{M_r} \log(1 + |\mathbf{h}^H(j,:)\mathbf{w}(:,j)|^2) \\ f_{\mathrm{P}}(\mathbf{W}) & = & \sum_i \|\mathbf{w}(i,:)\|_2^2 = \mathrm{Tr}(\mathbf{W}^H\mathbf{W}) \end{array} \tag{7.10}$$

We first consider the ZF-JASP problem ($\mathcal{P}$). Since we require the precoder to zero the interference, the second term in (7.6) is zero, i.e., $\gamma_j = \frac{|\mathbf{h}^H(j,:)\mathbf{w}(:,j)|^2}{\sigma_j}$ in (7.7). For normalization purposes we assume $\sigma_j = 1$, then the capacity expression (f_{C}) defined in (7.8) is simplified to $\sum_{j=1}^{M_r} \log(1 + |\mathbf{h}^H(j,:)\mathbf{w}(:,j)|^2)$, and the ZF-JASP problem is defined as

$$\mathcal{P}_0 : \begin{array}{ll} \text{maximize} & f_{\mathrm{C}}(\mathbf{W}) \\ \mathbf{W} \in \mathbb{C}^{M_t \times M_r} & \\ \text{subject to} & \mathcal{C}_0(\mathbf{W}) : \begin{cases} \sum_{l \neq j} |\mathbf{h}^H(j,:)\mathbf{w}(:,l)|^2 = 0; & j = 1, 2, \ldots, M_r \\ \|\mathbf{w}(i,:)\|_2^2 \leq p_i^*; & i = 1, 2, \ldots, M_t \\ \|\mathbf{W}\|_{0,2} = L_s; & \end{cases} \end{array}$$

We used that $\|\boldsymbol{\delta}\|_0 = \|\mathbf{W}\|_{0,2}$, i.e., the ℓ_0-norm of the vector collecting the ℓ_2-norms of the rows of $\mathbf{W}$, obviously this is a non-convex function on precoder variable. The set of constraints for ZF-JASP is defined as the ZF constraint, the PAPC

constraint, and the antenna selection constraint, respectively. A related power minimization problem is also of high interest where the objective is to minimize the total transmit power subject to the constraints in $\mathcal{C}_0$ plus one extra constraint on f_{C}. Note that ZF-JASP is always feasible as long as $L_s \geq M_r$ for a full-rank channel matrix.

In MMSE precoding, normally the power is minimized while a set of predefined SINR constraints are satisfied, leading to

$$\mathcal{P}_0': \quad \begin{array}{ll} \underset{\mathbf{W} \in \mathbb{C}^{M_t \times M_r}}{\text{minimize}} & f_{\mathrm{P}}(\mathbf{W}) \\ \text{subject to} & \mathcal{C}_0'(\mathbf{W}): \begin{cases} \frac{|\mathbf{h}^H(j,:)\mathbf{w}(:,j)|^2}{\sum_{l \neq j} |\mathbf{h}^H(j,:)\mathbf{w}(:,l)|^2 + 1} \geq \gamma_j^*; & j = 1, \ldots, M_r \\ \|\mathbf{w}(i,:)\|_2^2 \leq p_i^*; & i = 1, \ldots, M_t \\ \|\mathbf{W}\|_{0,2} = L_s; \end{cases} \end{array}$$

where f_{P} is defined in (7.8). Note that $\mathcal{C}_0$ and $\mathcal{C}_0'$ include the same PAPC and antenna selection constraints and only the interference (first) constraints are different for ZF and MMSE precoding.

7.3.2 General Convex Formulation

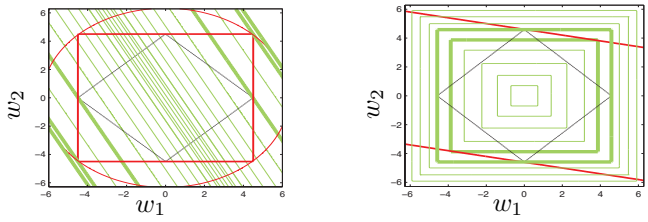
To move towards a convex formulation, an immediate choice for the relaxation (softening) of $\|\mathbf{W}\|_{0,2}$ is to replace it by a group sparsity inducing regularizer. A common group sparsity inducing regularizer is defined as a hybrid $\ell_{1,q}$-norm on the matrix $\mathbf{W}$ with entries $w(i,j)$, defined as

$$\|\mathbf{W}\|_{1,q} = \sum_{i=1} \{ \sum_{j=1} |w(i,j)|^q \}^{\frac{1}{q}}, \tag{7.11}$$

where, $\|\mathbf{W}\|_{1,2}$ is the so-called group Lasso regularization and is the tightest convex relaxation of $\|\mathbf{W}\|_{0,2}$ [122]. However, there is no closed form solution to find the exact counterpart of L_s for the substituted convex term $\|\mathbf{W}\|_{1,2}$. This means that an iterative approach is needed to find a value for $\|\mathbf{W}\|_{1,2}$ which gives exactly L_s non-zero rows in $\mathbf{W}$ for the JASP problem.

According to optimization theory, any constraint can be appended to an objective function with a proper Lagrange multiplier, e.g., problem $\mathcal{P}_0$ ($\mathcal{P}_0'$) can be relaxed as $\mathcal{P}_1$ ($\mathcal{P}_1'$) given a $\lambda' \geq 0$ corresponding to L_s which is found via a simple binary search as explained later, leading to

$$\mathcal{P}_1\,(\mathcal{P}_1'): \quad \begin{array}{ll} \underset{\mathbf{W} \in \mathbb{C}^{M_t \times M_r}}{\text{maximize}} & f(\mathbf{W}) - \lambda' \|\mathbf{W}\|_{1,2} \\ \text{subject to} & \mathcal{C}_1(\mathbf{W})\,(\mathcal{C}_1'(\mathbf{W})) \end{array} \tag{7.12}$$



(a) (b)

Figure 7.2: A 2D illustration of an example JASP problem. The black diamond is the ℓ_1-norm sparsity inducing regularizer, red lines show the TPC (circle) and PAPC (box) constraints. (a) optimizing capacity (green contours); (b) optimizing power (green contours), with a capacity constraint.

where $\mathcal{C}_1$ is formed by omitting the last constraint of $\mathcal{C}_0$ and the same is applicable to form the $\mathcal{C}_1'$ so

$$\mathcal{C}_1(\mathbf{W}) : \begin{cases} \sum_{l\neq j} |\mathbf{h}^H(j,:)\mathbf{w}(:,l)|^2 = 0; & j = 1,2,\ldots,M_r \\ \|\mathbf{w}(i,:)\|_2^2 \leq p_i^*; & i = 1,2,\ldots,M_t \end{cases} \tag{7.13}$$

$$\mathcal{C}_1'(\mathbf{W}) = \begin{cases} \frac{|\mathbf{h}^H(j,:)\mathbf{w}(:,j)|^2}{\sum_{l\neq j} |\mathbf{h}^H(j,:)\mathbf{w}(:,l)|^2+1} \geq \gamma_j^*; & j = 1,\ldots,M_r \\ \|\mathbf{w}(i,:)\|_2^2 \leq p_i^*; & i = 1,\ldots,M_t \end{cases} \tag{7.14}$$

where the relaxation is tight once the exact λ', corresponding to L_s, is found. The parameter λ' controls the number of selected antennas as it weighs the group sparsity inducing regularizer relative to the main objective. The solutions $\mathcal{P}_0^{\text{opt}}$ ($\mathcal{P}_0'^{\text{opt}}$) and $\mathcal{P}_1^{\text{opt}}$ ($\mathcal{P}_1'^{\text{opt}}$) show similar sparsity patterns once a proper λ' is found. The group sparsity inducing regularizer properties of the ℓ_1-norm are well known, nonetheless the following gives a simple example that may provide an illustration of the JASP problem.

Fig. 7.2 illustrates a system with one user, two antennas, and a real beamformer $\mathbf{w} = [w_1, w_2]^T$; the channel is $\mathbf{h} = [1, 0.5]$. For one user, there is no interference and hence no ZF constraint. For TPC we take the interior of a circle $w_1^2 + w_2^2 \leq 9$, and the PAPC is the total power divided by two, $|w_i| \leq 4.5$ for each antenna, which leads to the square box shown in Fig. 7.2. Optimizing for capacity (the slanted green lines), we see that the PAPCs are always satisfied with equality while if the individual constraint for each antenna is larger than 4.5, the TPC is an active constraint and the PAPC becomes inactive (ineffective). The optimal capacity contours are the ticker green lines in the corner in Fig. 7.2a which coincide with the vertices of the PAPC square or are tangent to the circle of the TPC when there is no PAPC considered in the problem.

Taking an ℓ_1-norm on $\mathbf{w}$ leads to a diamond-shaped region, which induces sparse solutions since the vertices are the first points that touch the square of the PAPC and maximize the capacity. In Fig. 7.2a the resulting sparse solution is $w_2 = 0, w_1 = 4.5$ leading to a smaller capacity than for the full system. The figure also shows that the largest capacity is achieved when only a TPC is imposed on the precoder with the full antenna set and the second best is the full system with PAPC.

Fig. 7.2b shows the converse problem of minimizing the total power subject to meeting a certain capacity. The two parallel red lines correspond to the capacity constraint, while the contours of constant equal power for each antenna are shown as squares. Without the sparsity inducing regularizer, the smaller green square is the optimal solution, whereas including the sparsity inducing regularizer (diamond shape touching the capacity constraint) leads to the larger green square power contour with solutions at the vertices of the sparsity inducing regularizer. In this case, more transmit power is needed while fewer (1) antennas are used. More in general (N dimensions), the feasible set is a polytope and since the objective is always growing to the outer direction of this polytope, the solution appears on the vertices [111].

7.4 Proposed Convex Formulation of ZF-JASP

The relaxed problem ZF-JASP ($\mathcal{P}_1$) is still non-convex in $\mathbf{W}$. In this section, we show that the plain group sparsity term in (7.11) cannot be used directly to formulate the sparse ZF-JASP precoding problem.

7.4.1 The Proposed Relaxation Technique

The optimization problem $\mathcal{P}_1$ is a special form of a QCQP which is proven to be non-convex and NP-hard in its original form [123]. We will apply the relaxation technique that is used in [112, 110, 124], to relax the non-convex precoding problem by introducing the positive semi-definite variables $\mathbf{Z}_j = \mathbf{w}(:,j)\mathbf{w}^H(:,j)$. Accordingly, $\mathcal{P}_1$ can be expressed as

$$\boxed{\begin{array}{ll} \text{maximize} & f(\{\mathbf{Z}_j\}) - \lambda g(\{\mathbf{Z}_j\}) \\ \{\mathbf{Z}_j\} \in \mathbb{S}_+^{M_t} & \\ \text{subject to} & \mathcal{C}_1(\{\mathbf{Z}_j\}),\, \text{rank}(\mathbf{Z}_j) = 1;\, \forall j \end{array}} \tag{7.15}$$

where $\mathbb{S}_+^{M_t}$ denotes the cone of positive semi-definite matrices and $g(\{\mathbf{Z}_j\})$ is a group sparsity inducing regularizer (a transformation of $\|\mathbf{W}\|_{1,2}$) with its regular-

ization parameter λ. Obviously, the rank constraints are not convex and later we drop (relax) these constraints. This relaxation technique requires the decomposition of the optimal $\mathbf{Z}_j$ to obtain $\mathbf{w}(:,j)$ at the end, and also increases the number of optimization variables quadratically as well as the complexity of the algorithm. We should not use this relaxation technique unless it is absolutely necessary, due to its complexity. The nonlinear nature of the capacity expression enforces this relaxation technique as used in the QCQP.

W need to redefine the functions for the objective and constraints in terms of the new variables $\{\mathbf{Z}_j\}$. The transmit power of the i-th antenna $\|\mathbf{w}(i,:)\|_2^2$ and the interference terms are readily expressed as

$$p_i := \|\mathbf{w}(i,:)\|_2^2 = \sum_{j=1}^{M_r} Z_j(i,i), \tag{7.16}$$

$$|\mathbf{h}^H(j,:)\mathbf{w}(:,l)|^2 = \mathbf{h}^H(j,:)\mathbf{Z}_l\mathbf{h}(j,:) = \mathrm{Tr}(\mathbf{Q}_j\mathbf{Z}_l), \tag{7.17}$$

$$\mathbf{Q}_j := \mathbf{h}(j,:)\mathbf{h}^H(j,:). \tag{7.18}$$

Similarly, the capacity function in (7.10) can be rewritten in terms of $\mathbf{Z}_j$ and $\mathbf{Q}_j$ as

$$f_{\mathrm{C}}(\{\mathbf{Z}_j\}) = \sum_{j=1}^{M_r} \log(1 + \mathrm{Tr}(\mathbf{Q}_j\mathbf{Z}_j)). \tag{7.19}$$

However, it is not straightforward to find the proper transformation from $\|\mathbf{W}\|_{1,2}$ to $g(\{\mathbf{Z}_j\})$.

We need to rewrite the group sparsity inducing regularizer as a convex function of the quadratic variables $\{\mathbf{Z}_j\}$. This is introduced in [121] for multicasting problem, as $\|\mathbf{W}\|_{1,\infty}^2$. Moreover, a relevant transformation technique is introduced in [111, page 131], which suggests to replace a function by its square for convex relaxation purposes. Here we propose $\|\mathbf{W}\|_{1,2}^2$ as an alternative convex sparsity regularizer as this is more relevant to the classic group Lasso formulation[3]. These two group sparsity inducing regularizers do not necessarily return the same subset of antennas for a fixed channel realization. We also considered $\|\mathbf{W}\|_{1,\infty}^2$ in the simulations as a sparsity regularizer for the ZF-JASP problem, for the sake of comparison. Simulation results indicate no statistical privilege among these two regularizers.

The derivation for the proposed group sparsity inducing regularizer is similar to [121], however, the proposed regularizer requires proper introduction in order to

[3]Using the infinity norm limits the precoder solutions to the ones with equal magnitude rows.

be understood within the new context (capacity problem) so extra discussions and proofs, compared to [121], are provided here.

Substituting the plain group Lasso regularizer in $\mathcal{P}_1$ results in $\mathcal{P}_2$ which is given by

$$\boxed{\begin{array}{lll} & \text{maximize} & f_C(\mathbf{W}) - \lambda(\|\mathbf{W}\|_{1,2})^2 \\ \mathcal{P}_2: & \mathbf{W} \in \mathbb{C}^{M_t \times M_r} & \\ & \text{subject to} & \mathcal{C}_1(\mathbf{W}) \end{array}} \tag{7.20}$$

Following [111], we can say that

$$\exists\,(\lambda', \lambda) \quad : \quad \mathcal{P}_2^{\text{opt}} \equiv \mathcal{P}_1^{\text{opt}}, \tag{7.21}$$

which means that there exist a λ corresponding to λ', such that the problem $\mathcal{P}_2$ is equivalent to $\mathcal{P}_1$.

We make the following definitions. Let $\mathbf{z} = \text{vec}(\mathbf{W})$ and define $\mathbf{Z}_{j,l} = \mathbf{w}(:,j)\mathbf{w}^H(:,l)$

$$\mathbf{Z} = \mathbf{z}\mathbf{z}^H = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{Z}_{1,2} & \cdots & \mathbf{Z}_{1,M_r} \\ \mathbf{Z}_{2,1} & \mathbf{Z}_2 & \cdots & \mathbf{Z}_{2,M_r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_{M_r,1} & \mathbf{Z}_{M_r,2} & \cdots & \mathbf{Z}_{M_r} \end{bmatrix} \in \mathbb{C}^{M_t M_r \times M_t M_r} \tag{7.22}$$

Similarly, we introduce $\mathbf{z}' = \text{vec}(\mathbf{W}^T)$ and define $\mathbf{Z}'_{i_1,i_2} = \mathbf{w}(i_1,:)\mathbf{w}^H(i_2,:) \in \mathbb{C}^{M_r \times M_r}$, and

$$\mathbf{Z}' = \mathbf{z}'\mathbf{z}'^H = \begin{bmatrix} \mathbf{Z}'_{1,1} & \mathbf{Z}'_{1,2} & \cdots & \mathbf{Z}'_{1,M_t} \\ \mathbf{Z}'_{2,1} & \mathbf{Z}'_{2,2} & \cdots & \mathbf{Z}'_{2,M_t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}'_{M_t,1} & \mathbf{Z}'_{M_t,2} & \cdots & \mathbf{Z}'_{M_t,M_t} \end{bmatrix} \in \mathbb{C}^{M_t M_r \times M_t M_r} \tag{7.23}$$

The entries of $\mathbf{Z}$ and $\mathbf{Z}'$ are related as $Z'_{i_1,i_2}(j,l) = Z_{j,l}(i_1,i_2)$. In other words, $\mathbf{Z} = \mathbf{P}\mathbf{Z}'\mathbf{P}^T$ where $\mathbf{P}$ is a permutation matrix i.e., equivalent to $\text{vec}(\mathbf{W}) = \mathbf{P}\text{vec}(\mathbf{W}^T)$. With these definitions we can write the squared mixed norm of $\mathbf{W}$ as

$$\begin{aligned} (\|\mathbf{W}\|_{1,2})^2 = \big(\sum_{i=1}^{M_t} \|\mathbf{w}(i,:)\|_2\big)^2 &= \sum_{i_1=1}^{M_t}\sum_{i_2=1}^{M_t} \|\mathbf{w}(i_1,:)\|_2 \cdot \|\mathbf{w}(i_2,:)\|_2 \\ &= \sum_{i_1=1}^{M_t}\sum_{i_2=1}^{M_t} \|\mathbf{Z}'_{i_1,i_2}\|_F\,. \end{aligned} \tag{7.24}$$

To derive the last step, we have used the following property for arbitrary vectors $\mathbf{a}$ and $\mathbf{b}$: $\|\mathbf{a}\mathbf{b}^H\|_F^2 = \|\mathbf{a}\|_2^2\|\mathbf{b}\|_2^2$. Next, define a matrix $\bar{\mathbf{Z}} \in \mathbb{R}^{M_t \times M_t}$ with scalar

entries $\bar{Z}(i_1, i_2) = \|\mathbf{Z}'_{i_1,i_2}\|_F$, then $(\|\mathbf{W}\|_{1,2})^2 = \text{Tr}(\mathbf{1}_{M_t \times M_t}\bar{\mathbf{Z}})$. Further note that the entries of $\bar{\mathbf{Z}}$ satisfy

$$\bar{Z}(i_1, i_2) = \|\mathbf{Z}'_{i_1,i_2}\|_F = \sqrt{\sum_{j,l} |Z_{j,l}(i_1, i_2)|^2}\ . \tag{7.25}$$

The group sparsity inducing regularizer in (7.24) is a convex function of $\mathbf{Z}_{j,l} = \mathbf{w}(:, j)\mathbf{w}^H(:, l)$, then the JASP problem of $\mathcal{P}_2$ can thus be relaxed to the following convex optimization problem:

$$\mathcal{P}_3: \quad \begin{array}{ll} \text{maximize} & f_C(\{\mathbf{Z}_j\}) - \lambda \text{Tr}(\mathbf{1}_{M_t \times M_t}\bar{\mathbf{Z}}) \\ \mathbf{Z}_j \in \mathbb{S}_+^{M_t}, \bar{\mathbf{Z}} \in \mathbb{R}^{M_t \times M_t} & \\ \text{subject to} & \mathcal{C}_3(\{\mathbf{Z}_j\}, \bar{\mathbf{Z}}) \end{array}$$

where

$$\mathcal{C}_3(\{\mathbf{Z}_j\}, \bar{\mathbf{Z}}) = \begin{cases} \sum_{l \neq j} \text{Tr}(\mathbf{Q}_j \mathbf{Z}_l) = 0; & j = 1, \cdots, M_r \\ \sum_j Z_j(i, i) \leq p_i^*; & i = 1, \cdots, M_t \\ \bar{Z}(i_1, i_2) \geq \sqrt{\sum_j |Z_j(i_1, i_2)|^2} & i_1, i_2 = 1, \cdots, M_t \end{cases}$$

Although $\mathbf{Z}$ was defined as a rank one matrix in (7.22), the rank constraint is not posed. Also the rank constraints on the individual $\mathbf{Z}_j$ have been dropped and what is left is a constraint that the $\mathbf{Z}_j$ are positive semi-definite. Although $\bar{\mathbf{Z}}$ is related to $\mathbf{Z}$, in this formulation $\bar{\mathbf{Z}}$ is treated as an independent variable which is related to $\mathbf{Z}$ only via the last constraint (a standard SOC constraint). The relaxation of $\bar{\mathbf{Z}}$ is tight because the last constraint is always satisfied with equality: the objective function tries to minimize $\text{Tr}(\mathbf{1}_{M_t \times M_t}\bar{\mathbf{Z}})$ which makes the elements of $\bar{\mathbf{Z}}$ as small as possible.

The last constraint and its origin in (7.25), is a function of the entries (i_1, i_2) of $\mathbf{Z}_{j,l}$. These can be split into the diagonal block-entries $\{\mathbf{Z}_j\}$ and the off-diagonal block-entries $\{\mathbf{Z}_{j,l}\}_{j \neq l}$. Since there is no constraint on these off-diagonal matrices (after we drop the non-convex rank constraint on the rank of $\mathbf{Z}$), they are naturally put to zero when the group sparsity inducing regularizer is being minimized, and consequently we could drop them from the optimization problem $\mathcal{P}_3$.

Remark 1. *Solving $\mathcal{P}_3$ promotes a sparse solution for $\bar{\mathbf{Z}}$ and consequently for the $\mathbf{Z}_j$ matrices. If the solution is of rank one, and hence decomposable as $\mathbf{Z}_j = \mathbf{w}(:, j)\mathbf{w}^H(:, j)$, then $\mathbf{W}$ tends to be row-wise sparse.*

Please see Appendix 7.A for further discussion on Remark 2. As a result of Remark 2, the eliminated antennas can be determined by looking at $\bar{\mathbf{Z}}$ at the solution of the optimization problem. Conversely, the set of selected antennas is

$\mathcal{I}^c = \{i : \bar{Z}(i,i) \neq 0,\ i = 1, 2, \cdots, M_t\}$, which is obtained by looking at the nonzero diagonal entries of $\bar{\mathbf{Z}}$.

7.4.2 The Subspace-Aware Formulation

It is known that the ZF constraint can be removed by searching only in the ZF feasible subset of the optimization problem. This leads to an analytical solution for ZF precoding with a total power constraints [107]. Here we adapt this formulation for the case of a PAPC leading to simplifications in the computations. It is known that the ZF constraint can be written as $\bar{\mathbf{H}}_j \mathbf{w}(:,j) = \mathbf{0}_{M_r - 1}$, where $\bar{\mathbf{H}}_j$ is the aggregated interference channel,

$$\bar{\mathbf{H}}_j = [\mathbf{h}(1,:), \cdots, \mathbf{h}(j-1,:), \mathbf{h}(j+1,:), \cdots, \mathbf{h}(M_r,:)]^H .$$

Following [107], the ZF precoder $\mathbf{w}(:,j)$ needs to lie in the null space of $\bar{\mathbf{H}}_j \in \mathbb{C}^{(M_r-1)\times M_t}$. Let $\mathbf{K}_j \in \mathbb{C}^{M_t \times S_n}$ holds an orthonormal basis for the null space of $\bar{\mathbf{H}}_j$. Assuming $\bar{\mathbf{H}}_j$ is full rank, then the size of the nullity is equal to $S_n = M_t - M_r + 1$. The ZF constraint can now be expressed by requiring that $\mathbf{w}(:,j)$ can be written as $\mathbf{w}(:,j) = \mathbf{K}_j (:,j)$, for a vector $(:,j) \in \mathbb{C}^{S_n}$. These vectors are collected in a matrix

$$\mathbf{M} = [(:,1), (:,2), \cdots, (:,M_r)] \in \mathbb{C}^{S_n \times M_r}.$$

We define the counterpart semi-definite quadratic variables $\mathbf{Y}_j = (:,j)^H (:,j),\ j = 1, 2, \cdots, M_r$, to replace the $\mathbf{Z}_j$ matrices. The received power of user j can now be expressed as

$$\begin{aligned} |\mathbf{h}^H(j,:)\mathbf{w}(:,j)|^2 &= \mathbf{h}^H(j,:)\mathbf{K}_j(:,j)^H(:,j)\mathbf{K}_j^H\mathbf{h}(j,:) && (7.26)\\ &= \mathrm{Tr}(\mathbf{K}_j^H\mathbf{h}(j,:)\mathbf{h}^H(j,:)\mathbf{K}_j\mathbf{Y}_j) = \mathrm{Tr}(\mathbf{A}_j\mathbf{Y}_j), && (7.27)\end{aligned}$$

where $\mathbf{A}_j = \mathbf{K}_j^H\mathbf{h}(j,:)\mathbf{h}^H(j,:)\mathbf{K}_j$. Therefore, the objective function can be written as

$$f_{\mathrm{C}}(\{\mathbf{Y}_j\}) = \textstyle\sum_{j=1}^{M_r} \log(1 + \mathrm{Tr}(\mathbf{A}_j\mathbf{Y}_j)). \qquad (7.28)$$

We also need to rewrite the constraints and group sparsity inducing regularizer that involve $\mathbf{Z}_j$ in terms of the $\mathbf{Y}_j$. This is straightforward:

$$\mathbf{Z}_j = \mathbf{w}(:,j)\mathbf{w}^H(:,j) = \mathbf{K}_j^H(:,j)^H(:,j)\mathbf{K}_j, \qquad (7.29)$$

with entry (i_1, i_2) given by

$$Z_j(i_1,i_2) = \mathbf{k}_j^H(i_1,:)(:,j)^H(:,j)\mathbf{k}_j(i_2,:) = \mathrm{Tr}(\mathbf{B}_j^{(i_1,i_2)}\mathbf{Y}_j);$$

$$\mathbf{B}_j^{(i_1,i_2)} := \mathbf{k}_j(i_1,:)\mathbf{k}_j^H(i_2,:).$$

The resulting performance maximization problem with the ZF criterion of $\mathcal{P}_3$ becomes

$$\mathcal{P}_3: \quad \begin{array}{cc} \text{maximize} & f_C(\{\mathbf{Y}_j\}) - \lambda \text{Tr}(\mathbf{1}_{M_t\times M_t}\bar{\mathbf{Z}}) \\ \mathbf{Y}_j \in \mathbb{S}_+^{M_t}, \bar{\mathbf{Z}} \in \mathbb{R}^{M_t\times M_t} & \\ \text{subject to} & \mathcal{C}_3(\{\mathbf{Y}_j\}, \bar{\mathbf{Z}}) \end{array}$$

where the constraint is given by

$$\mathcal{C}_3(\{\mathbf{Y}_j\}, \bar{\mathbf{Z}}) = \begin{cases} \sum_j \text{Tr}(\mathbf{B}_j^{(i,i)}\mathbf{Y}_j) \le p_i^*; & i = 1, \cdots, M_t \\ \bar{Z}(i_1,i_2) \ge \sqrt{\sum_j (\text{Tr}(\mathbf{B}_j^{(i_1,i_2)}\mathbf{Y}_j))^2} & i_1, i_2 = 1, \cdots, M_t \end{cases}$$

One can easily show the last constraint in $\mathcal{P}_3$ as a SOC constraints or in general as LMI form, that is given by

$$\begin{bmatrix} \bar{Z}(i_1,i_2) & \mathbf{v}^{(i_1,i_2)} \\ \mathbf{v}^{(i_1,i_2)} & \bar{Z}(i_1,i_2)\,\mathbf{I}_{M_r} \end{bmatrix} \succeq 0, \tag{7.30}$$

where $\mathbf{v}^{(i_1,i_2)} := [\text{Tr}(\mathbf{B}_1^{(i_1,i_2)}\mathbf{Y}_1), \cdots, \text{Tr}(\mathbf{B}_{M_r}^{(i_1,i_2)}\mathbf{Y}_{M_r})]$, for a specific index pair (i_1, i_2). In summary, we can write

$$\mathcal{P}_0^{\text{fsb}} \subseteq \mathcal{P}_1^{\text{fsb}} \equiv \mathcal{P}_2^{\text{fsb}} \subseteq \mathcal{P}_3^{\text{fsb}}, \tag{7.31}$$

denoting $\mathcal{P}_0^{\text{fsb}}$ as the feasible sets of $\mathcal{P}_0$ and so forth.

In the next Section, we show that the MMSE-JASP problem does not need the aformensioned relaxation technique for its convex transformation and the plain group Lasso regularizer suffices for the convex formulation of the problem. In Sec. 7.6, we discus the algorithms that uses these convex optimization problems to solve the JASP problems.

7.5 Proposed SDP Formulation of MMSE-JASP and ART

7.5.1 JASP Problem with MMSE Precoder (MMSE-JASP)

A formulation of the MMSE precoder as a SOC programming problem has been introduced in [108, 112]. This formulation relies on the fact that the optimal precoder is determined up to a phase rotation, i.e., if $\mathbf{W}$ is optimal then $\mathbf{W}\text{diag}(e^{j\phi_j})$ yields another optimal precoder, where the ϕ_js are arbitrary phases. Accordingly,

$\mathbf{h}^H(j,:)\mathbf{w}(:,j)$ can be taken as real, and with a proper phase rotation the SINR expression can be reformulated as a linear function of $\mathbf{w}(:,j)$.

The SINR constraints can be written in SOC form or in general as LMI constraints. For each user, one LMI is required to pose the SINR constraint. The PAPC can be written as linear constraints using the same argument and the plain group Lasso sparsity inducing regularizer ($\|\mathbf{W}\|_{1,2}$) can be applied directly to formulate the JASP problem as follows

$$\mathcal{P}_1': \begin{array}{ll} \text{minimize} & f_{\mathrm{P}}(\mathbf{W}) + \lambda \sum_i \|\mathbf{w}(i,:)\|_2 \\ \mathbf{W} \in \mathbb{C}^{M_t \times M_r} & \\ \text{subject to} & \mathcal{C}_1'(\mathbf{W}) = \left\{ \begin{array}{ll} \mathbf{C}_j \succeq 0; & j = 1, \ldots, M_r \\ \|\mathbf{w}(i,:)\|_2 \leq \sqrt{p_i^*}; & i = 1, \ldots, M_t \end{array} \right. \end{array}$$

where

$$\mathbf{C}_j = \begin{bmatrix} \frac{1}{\sqrt{\gamma_j}^*}\mathbf{h}^H(j,:)\mathbf{w}(:,j) & [\mathbf{h}^H(j,:)\bar{\mathbf{W}}_j, 1] \\ [\mathbf{h}^H(j,:)\bar{\mathbf{W}}_j, 1]^H & \frac{1}{\sqrt{\gamma_j}^*}\mathbf{h}^H(j,:)\mathbf{w}(:,j)\,\mathbf{I}_{M_r} \end{bmatrix}$$

and $\bar{\mathbf{W}}_j \in \mathbb{C}^{M_t \times M_r - 1}$ is defined similar to $\bar{\mathbf{W}}_j'$ in (7.6). Problem $\mathcal{P}_1'$ can be solved using any SDP solver. The Lagrange multiplier λ in $\mathcal{P}_1'$ takes different value from the one in ZF-JASP problem, however we use the same notation for the sake of generalization.

7.5.2 Antenna Reduction Technique (ART) Problem

We propose another interesting setting for the joint problem, i.e., if there is no rigid number of antennas to be selected and we are interested in finding the minimum number of antennas that can guarantee a set of desired constraints. In this case $L_s = \|\boldsymbol{\delta}\|_0$ is the optimization parameter and the ART problem can be formulated as

$$\mathcal{P}_0'': \begin{array}{ll} \text{minimize} & \|\boldsymbol{\delta}\|_0 \\ \mathbf{W}' \in \mathbb{C}^{M_t \times M_r}, \boldsymbol{\delta} \in \{0,1\}^{M_t} & \\ \text{subject to} & \text{power, BF, and quality constraints} \end{array}$$

The problem $\mathcal{P}_0''$ minimizes the number of transmit antennas subject to given constraint on the power, beamforming criteria and performance quality. Two different ART problems are introduced here: ZF-ART which finds the minimum antennas that can satisfy a set of PAPC, ZF and capacity constraint and MMSE-ART with PAPC and SINR constraints.

The ZF-ART problem with ZF and capacity constraints can be defined as

$$\mathcal{P}''_{3_C}: \begin{array}{ll} \text{minimize} & \text{Tr}(\mathbf{1}_{M_t\times M_t}\bar{\mathbf{Z}}) \\ \mathbf{Y}_j \in \mathbb{S}_+^{M_t}, \bar{\mathbf{Z}} \in \mathbb{R}^{M_t\times M_t} & \\ \text{subject to} & \mathcal{C}_3\,,\ f_C(\{\mathbf{Y}_j\}) \geq f^* \end{array} \tag{7.32}$$

where the feasible set is the intersection of $\mathcal{C}_3$ with an extra constraint on the capacity. Similar to $\mathcal{P}_3$, there are three steps of relaxations to reach the $\mathcal{P}''_{3_C}$ which is denoted by index 3. Although the same relaxation steps are taken, the feasible set for problem $\mathcal{P}''_{3_C}$ is different from $\mathcal{P}_3$ as the performance is not anymore the objective and needs to be satisfied as a constraint ($f_C(\{\mathbf{Y}_j\}) \geq f^*$). The problem need not to be feasible for every f^* or p_i^*.

Proposition 1. *If problem $\mathcal{P}''_{3_C}$ is feasible, there exists an attainable rank one solution for the $\mathbf{Y}_j$.*

Proof. The proof is in Appendix 7.B. □

Likewise, the same problem can be formulated for MMSE precoding. The MMSE-ART problem after one step of relaxation (similar to MMSE-JASP) is formulated as

$$\mathcal{P}''_{1_M}: \begin{array}{ll} \text{minimize} & \sum_i \|\mathbf{w}(i,:)\|_2 \\ \mathbf{W} \in \mathbb{C}^{M_t\times M_r} & \\ \text{subject to} & \mathcal{C}'_1(\mathbf{W}) \end{array} \tag{7.33}$$

which can de readily solved by any SDP solver.

7.6 Proposed Algorithms

7.6.1 Outline

We present algorithms for solving the JASP problems in two major steps. The original ($\mathcal{P}$) problem formulation specifies an exact number of L_s antennas that may be selected. In the relaxed problem, L_s does not appear, but we need to determine the regularization parameter λ that leads to the correct selection of L_s antennas. This will be found iteratively in step 1.

To improve the convergence to the sparse precoder, the iteratively reweighed algorithm of [125] is used. In the n-th iteration of the ZF-JASP algorithm, the all-one matrix $\mathbf{1}_{M_t\times M_t}$ in the group sparsity inducing regularizer of $\mathcal{P}_3$ is replaced by

a more general matrix $\mathbf{U}^{(n)} \in \mathbb{R}^{M_t \times M_t}$ which is updated relative to the entries of $\bar{\mathbf{Z}}$ at each iteration. Likewise, for the MMSE-JASP algorithm, in the n-th iteration, the all-one vector in $\mathcal{P}_1'$ is replaced by $\mathbf{u}^{(n)} \in \mathbb{R}^{M_t}$ which is updated relative to the $\|\mathbf{w}(i,:)\|_2$.

- *Step 1:* Solve the JASP problem, iteratively adapting $\lambda \in (0, 1]$ and $\mathbf{U}^{(n)}$ ($\mathbf{u}^{(n)}$) until the optimal solution for $\bar{\mathbf{Z}}$, called $\bar{\mathbf{Z}}^{\text{opt}}$ ($\mathbf{W}^{\text{opt}}$) has exactly L_s non-zero diagonal elements (rows).

  Note that the non-zero diagonal elements in $\bar{\mathbf{Z}}$ and non-zero rows in $\mathbf{W}$ indicate the selected antennas that correspond to the non-zero powers p_i. Define $\mathbf{p} = [p_1, \cdots, p_{M_t}]^T$, let $\mathcal{I}^c$ be the set of selected antennas, and denote by $\mathbf{H}_s := \mathbf{H}(:, \mathcal{I}^c) \in \mathbb{C}^{M_r \times L_s}$ the corresponding reduced-size channel matrix.

- *Step 2:* Once L_s antennas have been selected, the JASP problem is solved once more for $\lambda = 0$ and the reduced channel $\mathbf{H}_s$. This reduces the problem to the plain precoding problem ($\mathbf{\Delta}$ is fixed) and gives the final optimal precoder $\mathbf{W}_s^{\text{opt}}$.

In the notation, we will use subscript "s" to denote variables related to $\mathbf{H}_s$ and the selected set of antennas. The reason that we still need *Step 2* is clarified by the following remark.

Remark 2. *Let $\mathbf{W}^{opt}$ be the optimal sparse solution to any JASP problem, directly obtained or derived from the rank one factorizations of the optimal $\mathbf{Z}_j$s, where λ is such that $\|\mathbf{W}^{opt}\|_{0,2} = L_s$. Then, there always exists a precoder $\mathbf{W}_s^{opt}$ with the same sparsity pattern as $\mathbf{W}^{opt}$ such that $f(\mathbf{W}_s^{opt}) \geq f(\mathbf{W}^{opt})$.*

Note that there exists a range of $\lambda \in [\lambda_{L_s}, \lambda_{L_s+1})$ which gives the same sparse rows for the precoder in the JASP problem, even if the objective values are different for each value of λ. Within this range, solving for $\lambda = \lambda_{L_s}$ gives a solution with the largest $f(\mathbf{W})$.

Although we do not know λ_{L_s}, we can obtain the optimal solution by fixing the sparsity pattern (fixing $\mathbf{\Delta}$ and reducing $\mathbf{H}$ to $\mathbf{H}_s$), and subsequently removing the sparsity constraint (setting $\lambda = 0$). The resulting optimization problem then only maximizes the performance subject to the corresponding constraints within the set of selected antennas, which leads to the plain precoding problem and gives the optimal precoder $\mathbf{W}_s^{\text{opt}}$.

7.6.2 Algorithm for ZF-JASP

For the ZF-JASP problem which involves the nonlinear capacity expression, we need to solve problem $\mathcal{P}_3$ (or the converse power minimization problem). In general, $\mathcal{P}^{\lambda}$ denotes the problem with a selected λ, and by $\mathcal{P}^{\lambda=0}$ we mean the plain precoding problem in *Step 2* that follows if we set $\lambda = 0$ and fix the selected antennas. In that case, we work with $\mathbf{H}_s$. Let $\mathbf{K}_s \in \mathbb{C}^{L_s \times S_n}$ denote the corresponding orthonormal basis of the nullspace, and define $(\mathbf{B}_s)_j^{(i_1,i_2)}$ accordingly. The plain precoding problem in *Step 2* is then formulated as follows:

$$\mathcal{P}_3{}^{\lambda=0}: \begin{array}{ll} \text{maximize} & f_C(\{(\mathbf{Y}_s)_j\}) \\ (\mathbf{Y}_s)_j \in \mathbb{S}_+^{L_s} & \\ \text{subject to} & \sum_j \text{Tr}((\mathbf{B}_s)_j^{(i,i)}(\mathbf{Y}_s)_j) \leq p_i^*;\ \forall i \in \mathcal{I}^c \end{array} \tag{7.34}$$

(Note that the second and third constraints in $\mathcal{C}_3$ were dropped as they are not relevant for $\lambda = 0$.) This is a convex problem that can be solved using interior point methods, and we omit further details. The optimal beamforming vectors $\mathbf{w}_s(:,j)$ are extracted from $(\mathbf{Y}_s)_j$ using $(\mathbf{Y}_s)_j = \mathbf{w}_s(:,j)\mathbf{w}_s^H(:,j)$. This assumes that $(\mathbf{Y}_s)_j$ is rank one.

Remark 3. *There always is an attainable rank one solution for problem* $\mathcal{P}_3{}^{\lambda=0}$.

From [110] we know that there is always a rank one solution for $\mathcal{P}_3{}^{\lambda=0}$ before invoking the subspace approach. Since there is no alternation of the problem using the proper subspace to search for the solution then the same property is remained after this process. The extended discussion is given in Appendix 7.C.

Algorithm 1 summarizes the iterative algorithm for *Step 1* of $\mathcal{P}_3$. In the algorithm, L denotes the currently selected number of antennas, and $\lambda \in [\lambda_L, \lambda_U]$ is the currently selected regularization parameter. The experimental results show that taking a small value (close to zero) for λ_L and taking $\lambda_U = 1$ commonly gives the proper λ in few iterations (always less than 10 iterations) regardless of the choice for the initial λ. In an inner loop, problem $\mathcal{P}_3^{\lambda}$ is solved and the resulting number of selected antennas L is determined. If this does not correspond to the required number L_s, then the interval bound λ_L or λ_U is adjusted and the problem is solved again with λ set using a bisection technique. The sparsity enhancing weights $\mathbf{U}^{(n)}$ are adjusted in an inner loop, where $\mathbf{U}^{(n)}$ is updated by penalizing rows with smaller norms. The ϵ parameter is commonly used in this context to ensure stability and avoiding the undesired non-zero estimate at $n+1$ iteration due to the zero-valued

$|\bar{Z}(i_1,i_2)^{(n)}|$. This is chosen relative to the smallest value for $|\bar{Z}(i_1,i_2)|$ and is fixed here.

Algorithm 1 ZF-JASP (performance maximization)

Initialize $\lambda, \lambda_L, \lambda_U$.
$L := M_t$
$N := 0$
while $L \neq L_s$ **and** $N \leq N_{max}$ **do**
  $n := 0$
  $\mathbf{U}^{(0)} := \mathbf{1}_{M_t \times M_t}$
  while $L \neq L_s$ **and** $n \leq n_{max}$ **do**
    Solve $\mathcal{P}_3^{\lambda}$
    Update L based on $\bar{\mathbf{Z}}^{\text{opt}}$
    $U(i_1,i_2)^{(n+1)} := 1/(|\bar{Z}(i_1,i_2)^{(n)}| + \epsilon),\ \forall i_1, i_2$
    $n := n + 1$
  end while
  if $L > L_s$ **then** $\lambda_L := \lambda$
  else if $L < L_s$ **then** $\lambda_U := \lambda$
  end if
  $\lambda := \lambda_L + (\lambda_U - \lambda_L)/2$.
  $N := N + 1$
end while
if $L > L_s$ **then** do brute-force elimination.

For ZF-JASP precoding the problem is always feasible if $L_s \geq M_r$ so there is no need for a feasibility check. If the desired number of antennas has not been achieved within the limited number of iterations then brute-force elimination is performed by sorting the p_is and choosing the L_s most significant ones. Once the proper antenna set is found from Algorithm 1, the reduced size channel matrix is formed as $\mathbf{H}_s$ and it is passed on to *Step 2*. We use the Matlab package CVX [126] to solve the resulting convex optimization problems which uses the SeDuMi and SDP3 solvers but alternative solvers such as YALMIP and MOSEK can be used instead.

7.6.3 Algorithm for MMSE-JASP Problem

Note that, unlike the ZF problem, the MMSE problem may not be feasible even for $L_s \geq M_r$. So first we solve the ART problem related to the MMSE problem (Problem $\mathcal{P}''_{1_M}$) to see what is the minimum number of antennas that can handle the SINR and power constraints simultaneously. If a solution with L_s antennas is not

found then the problem is not feasible with L_s. Subsequently, an algorithm similar to Algorithm 1 is carried out to solve the MMSE problem except for the fact that the problem $\mathcal{P}_1'$ is solved in the fifth step of Algorithm 1 and that we are looking at $\mathbf{W}$ directly to find the eliminated antennas rather than at $\bar{\mathbf{Z}}$.

7.6.4 Algorithm for MMSE-ART Problem

The ART problem admits the simplest algorithm as it does not require a binary search. Accordingly, Algorithm 2 finds an exact optimal MMSE precoder. There is no need to perform *Step 2*. The ZF-ART follows the same path.

Algorithm 2 MMSE-ART

$\mathbf{u}^{(0)} := \mathbf{1}_{M_t}$
$L := M_t$ $n := 0$
while $n < n_{max}$ **do**
  Solve $\mathcal{P}''_{1_M}(\mathbf{u}^{(n)})$
  Update L based on $\mathbf{W}^{\text{opt}}$
  $u(i)^{(n+1)} := 1/(\|\mathbf{w}(i,:)\|_2^{(n)} + \epsilon),\ \forall i$
  $n := n + 1$
end while

7.6.5 Computational Complexity

SDP solvers use interior point methods to solve the resulting convex problems. Interior point methods use Newton's method at each iteration to solve a set of linear equations. The number of Newton iterations mostly depends on the number of linear inequality constraints [127, 128], but this grows only slowly with the problem dimension and for our settings varies between $3-5$ which can be neglected in the analysis. Thus, the complexity is dominated by the flops required per Newton iteration, and this depends strongly on the problem structure and the solver. Without specific structure, a rough approximation on the worst-case complexity states the number of operations as $\xi = \max(\mu^2\eta\,, \mu^3, F)$, where μ is the number of variables in the optimization problem and F is the cost of the first and second order derivative [111, page 8]. Nevertheless, this is not a very useful measure to evaluate the complexity of the problems in our setting where different problems are considered with various structures and variables. Alternatively, we give a relative comparison between the optimal approach and the proposed algorithms.

In the optimal approach, an exhaustive search to solve problem $\mathcal{P}_1$ requires solving $\binom{M_t}{L_s}$ convex optimization problems, each with a complexity of ξ. This is 28, 210 and 8008 times ξ for the selection of 6 antennas out of 8,10 and 16 antennas, respectively. Clearly the problem becomes intractable even for a small increase in the number of antennas. On the other hand, the proposed JASP approach in Algorithm 1 solves at most $n_{\max}$ (typically we take $n_{\max} = 10$) convex optimization problems for each selection step in the binary search, independent of L_s and M_t. E.g., this gives a constant $10 \times \xi$ for Algorithm 1, compared to $(\binom{16}{6} = 8008) \times \xi$ for the optimal approach ($L_s = 6$ and $M_t = 16$). In our experience, the algorithm usually converges in less than 10 iterations, on average about 6 iterations are used.

The ART problem is NP-hard in the form of Problem $\mathcal{P}_0''$, and quickly becomes impossible to solve optimally. Using the proposed approach, the ART problem can be solved efficiently without the need to find λ. However, in general, the expansion of the matrix dimensions in the linearization technique, used in the proposed Algorithm 1 as well as in the plain precoding problem with capacity term, has an important practical ramification; the complexity of the problem is increased (squared). Due to this effect, we can only handle medium size capacity problems with M_t up to 20 with a normal computer and general solvers like CVX and SeDuMi in a reasonable time. For more antennas, more efficient algorithms are required, i.e., perhaps possible by smoothing the non-differentiable constraints or sub-gradient approaches. This is not a problem for MMSE-JASP since the optimization is over a linear variable, owing to the alternative SDP formulation in [112]. In the later we are able to handle on the order of hundreds of antennas.

7.7 Simulation Results

The ZF-JASP problem was relaxed into a convex problem in three steps while the MMSE-JASP requires only one step of relaxation. It is not known how tight the relaxed problems are w.r.t. the original problem. This is evaluated using Monte Carlo simulations.

Each precoding strategy is evaluated using 10,000 random channel realizations. In turn the channel is generated using independent and identically distributed complex Gaussian random variables [4]. From these, a complementary cumulative density function (CCDF) is derived to show the probability that the shown performance

[4]The frequency selectivity of the channel does not directly affect the system model and the problem formulation in the chapter, as the channel can be potentially flattened using equalization techniques in general.

metric is less than or exceeds a certain value.

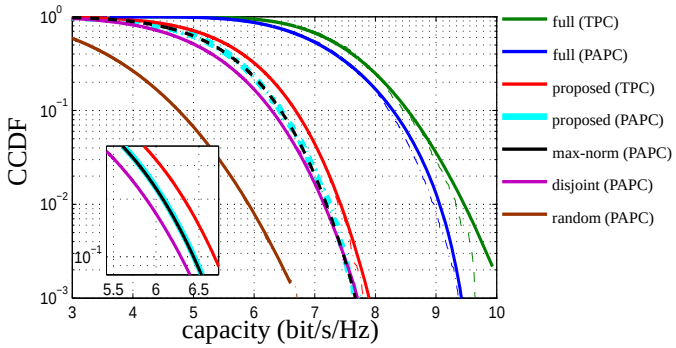
The simulations compare the proposed algorithm to the 'full' case without antenna selection, a 'random selection' of L_s antennas, and a 'disjoint' scheme which finds the antenna selection $\mathbf{\Delta}$ and precoder $\mathbf{W}'$ sequentially: an optimal subset of antennas is selected to maximize the capacity, $\log(\det[\mathbf{I}_{M_r} + \mathbf{H}\mathbf{\Delta}\mathbf{H}^H])$, using an exhaustive search, and subsequently the resulting plain precoding problem $\mathcal{P}_3^{\lambda=0}$ is solved on this 'optimal' subset (which did not take into account all constraints). Globally optimal solutions are calculated by enumerating all $\binom{M_t}{L_s}$ possible choices of L_s antennas, each time solving the resulting plain precoding problem.

Fig. 7.3 shows the resulting CCDF curves for the ZF-JASP problem. In the simulations, the maximum number of iterations is $N_{max} = 5$ and $n_{max} = 10$ in Algorithm 1. These are chosen based on the average number of iterations that the algorithm takes to find the the exact L_s number of antennas. Empirical results are fitted to suitable distributions by examining 15 different distributions and using a log-likelihood test to find the best fit. This is to show the sufficiency of empirical data as there is no significant fluctuations compared to a fitted distribution. We show the CCDF of the full, proposed and the proposed selection with the max-norm regularizer (adapted from [121]) for PAPC in Fig. 7.3a. The result does not show any noticeable difference between the choice of ℓ_∞-norm from [121] and the proposed ℓ_2-norm sparsity regularizer in term of the performance. Moreover, the TPC is considered which gives an insight regarding the performance loss compared to the PAPC scenario.

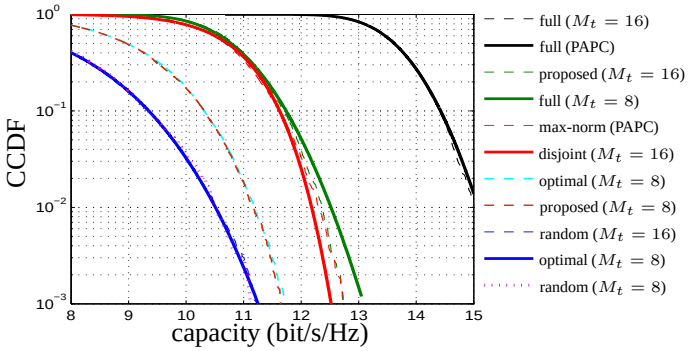
From Fig. 7.3b, it is seen that the optimal solution coincides with the proposed algorithm if $L_s = 6$ antennas are selected out of 8 antennas. Note that the combinatorial nature of the problem makes it impossible to find the optimal solution for Mt=16. In the simulation, the diversity gain (capacity increase for increasing M_t) is almost 2 bits per second per Hertz (bit/s/Hz) when M_t is doubled. The results also show the sub-optimality of the disjoint approach in comparison to the proposed algorithm. Note that even if the 'optimal' subset of antennas is found in the disjoint approach (only taking capacity into account), its performance is always worse than the proposed approach. In turn, random selection does not lead to a performance improvement even if M_t is increased.

We also consider the converse problem of power minimization for ZF precoding where the total transmit power is minimized subject to a predefined constraint on the capacity performance. Algorithm 1 is readily modified to cover this problem.

The average powers are shown as a function of the number of transmit antennas M_t in Fig. 7.4. It is seen that, while the average diversity gain. in terms of the



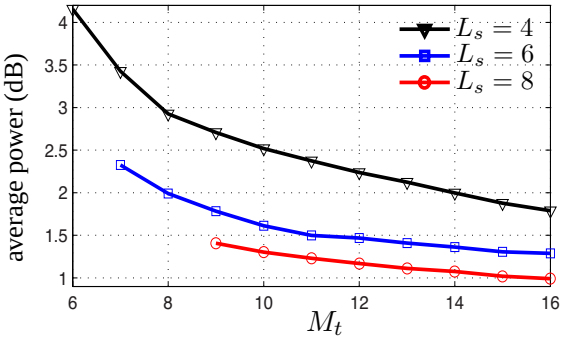
(a) $M_t = 10$, $M_r = 5$, $L_s = 6$ and $P^* = 3$ $p_i^* = 0.5$ dB. CCDF for full and disjoint selection are shown for TPC. The performance of max-norm and ℓ_2-norm regularizers coincides for PAPC.



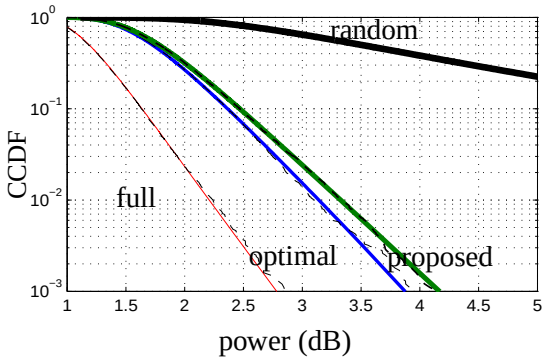
(b) $M_t = 8, 16$, $M_r = 4$, $L_s = 6$, $p_i^* = 2.5$ dB, Optimal selections are shown for $M_t = 8$, and disjoint selections for $M_t = 16$.

Figure 7.3: CCDF curves for ZF-JASP, with capacity objective. The dashed lines represent the empirical data and solid lines are used for the fitted Weibull model.

reduced average power is increased noticeably when M_t is increased from 6 to 10, the reduction becomes less significant (more linear) once more than 12 antennas are available. The diversity gain is not linear when L_s is increased, hence antenna selection leads to a noticeable power decrease if there are only 4 (few) RF chains available. Overall, an average power of almost 2 dB (20% of the total power) can be saved by increasing the number of available antennas from 6 to 16 and performing the proposed JASP algorithm. Defining the PAPC relative to L_s enables us to pose a total power constraint, inherently, if needed, as considered here to be 10 dB.



(a) Average transmit power versus number of available transmit antennas (M_t) for the ZF-JASP problem of power control, where $M_r = 3$, $C^* = 4$ bit/s/Hz, and $p_i^* = \frac{10}{L_s}$ dB.



(b) CCDF for $M_t = 10$ and $M_r = 3$, $L_s = 5$, $p_i^* = 2$ dB and $C^* = 4$ bit/s/Hz. The dashed lines represent the empirical data and solid lines are used for the fitted generalized extreme value models.

Figure 7.4: Power minimization problem for ZF-JASP.

However, average power does not show the probability of high power events which might affect the functionality of the amplifier and other nonlinear hardware components. In Fig. 7.4b, CCDF curves are presented which show how probable it is to get a certain transmit power. It is seen that random antenna selection is extremely inefficient while the proposed algorithm is very close to the optimal approach.

CCDF curves for the MMSE-JASP problem related to $\mathcal{P}_1'$ are shown in Fig. 7.5. Effectively, the gain (relative performance) of increasing M_t from 8 to 128, and

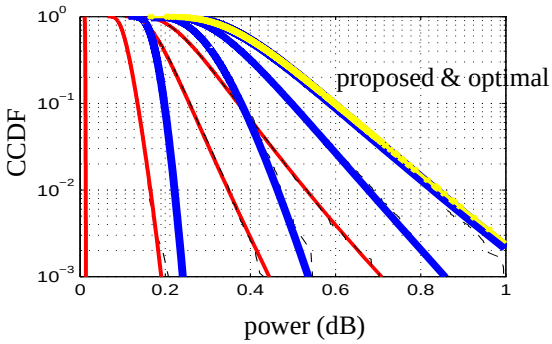


Figure 7.5: The CCDFs for the MMSE-JASP problem, for $M_r = 3$, $L_s = 4$ and $M_t = 8, 10, 16, 128$ respectively from right to left, $\gamma^* = 1$ dB and $p_i^* = 0.4$ dB. Dashed lines represent the empirical data, solid lines are fitted statistical models. The optimal solution is simulated for $M_t = 8$ and it is represented with yellow line. The tick (blue) lines correspond to the proposed technique and the full antenna performance are illustrated with the thin (red) lines.

then performing antenna selection to choose the best L_s antennas, is somehow larger than the relative gain of using the full set of 8 to 128 antennas. Moreover, the low complexity of the SDP implementation of the problem allows for more antennas to be considered and consequently more antenna selection gain to be obtained.

From a statistical point of view, the initial results show that the Weibull distribution gives the best fit for the capacity maximization problem, whereas the generalized extreme value is more suitable for the power minimization problem. However, for any solid conclusion on the statistical behavior of the system, a more dedicated study beyond the scope of this work is required.

Finally, the MMSE-ART problem related to $\mathcal{P}''_{1_M}$ (Algorithm 2) is simulated with $n_{max} = 3$ for different M_t, in Fig. 7.6, by simulating 10,000 channel realizations. The minimum number of antennas are searched once the PAPC, TPC and SINR constraints are fixed. The results show that a larger M_t leads to more selected antennas with a fixed n_{max} while for the smallest $M_t = 12$, most realizations end up with $L_s = 6$ (least possible number of antennas). However, as expected, increasing M_t reduces the chance of getting infeasible problems, specially for small P^* or large γ^*.

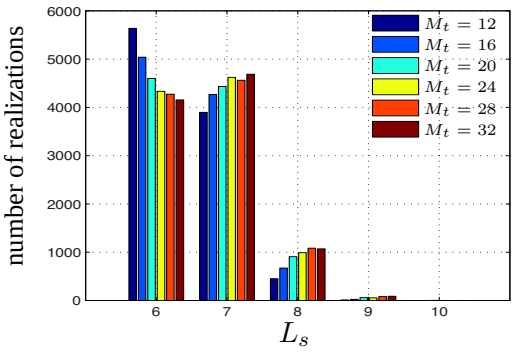


Figure 7.6: The histogram for the MMSE-ART problem, for $M_r = 6$, $M_t = 12, 16, 20, 24, 28, 32$, $\gamma^* = 25$ dB and PAPC with $p_i^* = 1$ dB. The x axis shows the number of selected antennas as $6, 7, 8, 9, 10$ and the y axis is the count of realizations that a certain number of antennas are selected.

7.8 Summary and remarks

The main contribution of this work is the classification of the most pertinent linear precoding strategies together with the proper sparsity inducing regularizers that can provide a convex formulation of the joint antenna selection and precoding problem. Simulation results show that the resulting algorithms yield solutions very close to the optimal selection strategy with far less computational complexity. The disjoint approach is shown to be sub-optimal in this context.

The chapter considered a capacity maximization and power minimization problems with beamforming, power allocation and antenna selection constraints. The map of all steps taken to reach a computationally feasible (convex) problem for capacity involved problem of ZF-JASP is given as

$$\begin{aligned} Step1: &\quad \mathcal{P}_3 \xRightarrow{relaxation} \mathcal{P}_2 \overset{(7.21)}{\Longleftrightarrow} \mathcal{P}_1 \xRightarrow{relaxation} \mathcal{P}_0 \\ Step2: &\quad \mathcal{P}_3{}^{\lambda=0} \overset{[23]}{\Longleftrightarrow} \mathcal{P}_2{}^{\lambda=0} \overset{(7.21)}{\Longleftrightarrow} \mathcal{P}_1{}^{\lambda=0} \end{aligned}$$

where the double-headed arrow ($\Leftrightarrow$) represents equivalence and the single-headed ($\Rightarrow$) arrow means relaxation. Three relaxations were performed to reach the surrogate convex problem when the nonlinear capacity expression is involved. For the JASP-MMSE problem a one-step relaxation is sufficient. A fairness problem which has not been discussed in this chapter can also be handled using the same SDP formulation, i.e., maximizing $f_{\mathrm{F}} = \min(\gamma_j)$ instead of the capacity expression (f_{C}) in ZF-JASP.

The main source of sub-optimality comes from the relaxation which translates the original ℓ_0-norm (counting the number of selected antennas) to a convex group sparsity inducing regularizer, based on the squared ℓ_1-norm on the precoder matrix, which is common in all the discussed problems. Unfortunately, there is no guarantee that this relaxation yields the same solution as the original discrete selection problem even if the ℓ_1-norm is the tightest known relaxation of the problem. However, if the first step (*Step 1*) of the algorithm finds the correct subset of antennas (equal to the optimum set), then the second and final step (*Step 2*) of finding the corresponding precoder is always tight and the exact sparse precoder is guaranteed to be found. In this case, the solution of the JASP algorithm is optimum.

We conclude the chapter by emphasizing that for the capacity involved problems the squared group Lasso regularizer is required to transform the joint antenna selection and precoding problem to a convex problem. Although the proposed sparsity regularizer can be cast as LMI, the problem is yet not a SDP problem due to the presence of the nonlinear capacity expression and is solved using general interior point algorithms. For the joint antenna selection and MMSE precoding problem a common group Lasso sparsity regularizer suffices, and the resulting sparse precoding problem can be solved using any SDP solver.

7.A Discussion on Remark 1

Minimizing the group sparsity inducing regularizer, $g(\bar{\mathbf{Z}}) := \text{Tr}(\mathbf{1}_{M_t \times M_t}\bar{\mathbf{Z}})$, acts as minimizing the sum of M_t^2 values $\bar{Z}(i_1, i_2)$, i.e., the ℓ_1-norm of this sequence. The sparsity property of this norm tends to make each of the $\bar{Z}(i_1, i_2)$ entries equal to zero. If that occurs, then since $\bar{Z}(i_1, i_2) \geq \sqrt{\sum_j |Z_j(i_1, i_2)|^2}$, it follows that the entry (i_1, i_2) of all $\mathbf{Z}_j$ matrices is equal to zero. If moreover the $\mathbf{Z}_j$ matrices converge to rank one, i.e., $\mathbf{Z}_j = \mathbf{w}(:, j)\mathbf{w}^H(:, j)$, then $w(i_1, j)$ and $w(i_2, j)$ are zero for all j, or $\mathbf{w}(i_1, :) = \mathbf{0}$ and $\mathbf{w}(i_2, :) = \mathbf{0}$. Thus, $\mathbf{W}$ tends to be row-wise sparse. The rows that are zero correspond to the eliminated (un-used) antennas.

7.B Proof of Proposition 1

To prove the proposition, first we prove a more general Lemma as follows.

Lemma 1. *Let $\mathbf{\Omega} \in \mathbb{C}^{M \times N}$ and $\mathbf{\Xi} \in \mathbb{C}^{M \times M}$ be given matrices where $M \leq N$ and $\mathbf{\Xi}$ is full rank, and suppose there exists a Hermitian matrix $\mathbf{\Pi} \in \mathbb{C}^{N \times N}$ of*

rank J where $M \leq J \leq N$ such that $\mathbf{\Omega\Pi\Omega}^H = \mathbf{\Xi}$. Then there exists a Hermitian matrix $\mathbf{\Pi}'$ of rank $J' = M$ such that $\mathbf{\Omega\Pi}'\mathbf{\Omega}^H = \mathbf{\Xi}$.

Proof. First of all, conditioning $\mathbf{\Xi}$ to be full rank implies that the rank of $\mathbf{\Omega}$ and $\mathbf{\Pi}$ can not be less than M according to the rank product property i.e., the rank of the product of matrices is always less than or equal to the minimum rank of matrices and hence, $\mathbf{\Omega}$ is full rank. Now, suppose $\mathbf{\Omega}_\perp \in \mathbb{C}^{N\times N} = \mathbf{\Omega}^H(\mathbf{\Omega\Omega}^H)^{-1}\mathbf{\Omega}$ is the orthogonal projection matrix onto the rowspace of $\mathbf{\Omega}$. Then $\mathbf{\Omega\Omega}_\perp = \mathbf{\Omega}$, so $\mathbf{\Omega\Omega}_\perp\mathbf{\Pi\Omega}_\perp^H\mathbf{\Omega}^H = \mathbf{\Xi}$. Note that $\text{rank}\{\mathbf{\Omega}_\perp\mathbf{\Pi\Omega}_\perp^H\} = M \leq J$, since the projection matrix maps $\mathbf{\Pi}$ into a subspace with dimension M as the rank of $\mathbf{\Omega}_\perp$ is always M. Putting $\mathbf{\Pi}' = \mathbf{\Omega}_\perp\mathbf{\Pi\Omega}_\perp^H$ proves the lemma. □

The ZF criterion can be stated as $\mathbf{H\Theta H}^H = \mathbf{\Gamma}$ where $\mathbf{\Gamma}$ is a diagonal (full rank) matrix, and from Lemma 1, there is always a $\mathbf{\Theta}$ of rank M_r which means it is decomposable as $\mathbf{WW}^H$ where $\mathbf{W}$ is the beamforming matrix. This is equivalent to always having rank one solutions for all $\mathbf{Z}_j$ matrices in $\mathcal{P}_3$ and consequently for all $\mathbf{Y}_j$ matrices in $\mathcal{P}_3$. Note that the orthogonal projection is a bounded operator so the low rank solution is always feasible (satisfying the power constraint) if the higher rank solution is feasible. Also, as long as $\mathbf{\Gamma}$ is fixed, the capacity (minimum rate) constraint is still feasible.

7.C Discussion on Remark 4

As proven in [110], we may assume that there always is a rank one solution for the variables $(\mathbf{Z}_s)_j$ in the plain precoding problem ${\mathcal{P}_3}^{\lambda=0}$.

Since $(\mathbf{Z}_s)_j = (\mathbf{K}_s)_j(\mathbf{Y}_s)_j(\mathbf{K}_s)_j^H$, and $(\mathbf{K}_s)_j$ is tall with full column rank, it follows that $(\mathbf{Y}_s)_j$ is also rank one. So the rank relaxation for ${\mathcal{P}_3}^{\lambda=0}$ is tight.

Suppose the optimal solution of ${\mathcal{P}_3}^{\lambda=0}$ with any rank is $\mathbf{Y}_j^{\text{opt}}$ for $j = 1, 2, \cdots, M_r$. Extending the technique proposed in [110] for ${\mathcal{P}_3}^{\lambda=0}$, this solution can be transformed into the desired rank one precoder solution by solving

$$\begin{aligned} &\text{minimize} \quad \Re\{\mathbf{h}^H(j,:)\mathbf{K}_j\mathbf{t}_j\} \\ &\mathbf{t}_j \in \mathbb{C}^{S_n} \\ &\text{subject to} \quad |\mathbf{k}_j^H(i,:)\mathbf{t}_j|^2 \leq \text{Tr}(\mathbf{B}_j^{(i,i)}\mathbf{Y}_j^{\text{opt}}); \quad \forall i \end{aligned} \tag{7.35}$$

for $\mathbf{t}_j = \mathbf{t}_j^{\text{opt}}$. The precoder is given by $\mathbf{w}(:,j) = \mathbf{K}_j\mathbf{t}_j^{\text{opt}}$.

Part III

Conclusion and Future Work

Chapter 8

Conclusion and Future Work

We now revisit the research questions that were introduced in Chapter 1, i.e., to give a concluding remarks on the covered topics. Moreover, possible future directions will be introduced which can be built upon this foundation for further research on this area.

8.1 Summary of Results

This thesis was initiated by an actual industrial demand to establish a fast and reliable wireless link within a mechatronic system, mainly to ease maintenance and to reduce the required space for these machinery devices. By careful investigation of the wireless channel within an enclosed environment (similar to the mechatronic system) via performing measurements, basis for the system design was founded. The first part of the thesis, including three chapters, is dedicated to channel modeling (stochastic approach) and feasibility study of the specific problem of designing a wireless link within a mechatronic device. Furthermore, a thorough overview of possible equalization techniques is provided which motivates the use of frequency domain equalization and orthogonal frequency division multiplexing (OFDM) technique in such dispersive environment. In addition, an example of OFDM design for the wideband communication model was proposed which can be deployed to achieve high data-rates in severe frequency selective channels. In this part, only single-input single-output (SISO) systems are studied.

In the second part, the thesis investigates a more general research area which covers the transmitter design in multiple-input multiple-output (MIMO) systems.

In fact, MIMO systems are an inevitable part of any modern wireless standard that aims for high data rate transmission. First, a brief overview and system model is given in Chapter 5. Building on this, a precoding technique was proposed which takes the OFDM waveform, with potentially a large dynamic range, and delivers a more predictable (close to a constant envelope) signal for the transmission, which is easier to deal with from an implementation point of view. Finally, the precoder design was extended for relatively more complex constraints including per antenna power constraint and limited available radio frequency (RF) chains.

Note that we mainly looked at *linear* signal processing operators in this thesis and the system is *time-invariant* at least over one processing block, so an LTI model was considered here. Further, the instantaneous channel state information was assumed to be available at the transmitter and the receiver. More specifically, numerical algorithms and *iterative optimization techniques* are used as a general tool to solve the resulting optimization problems. Accordingly, the gist of the discussion was centered around how to model, formulate and solve a linear processor given a priori knowledge about the communication channel to effectively address the challenges and opportunities in highly dispersive environments.

8.1.1 Classification of Non-convex Optimization for Communications

Apart from the application oriented problems that are considered in this work, the signal processing tools that are used throughout this thesis to formulate and solve the underlying problems are interesting and deserve more attention in the context of signal processing theory. Summarizing the thesis with this angle leads to the following overview of approaches for solving non-convex optimization problems.

More specifically, the problems that have been addressed in the second part of this thesis share the same structure in terms of the optimization problem to be solved. Commonly, a multivariate nonlinear objective function (performance measure), such as capacity, is to be maximized or a cost function to be minimized subject to a set of equality or inequality constraints. In general, for an arbitrary two complex (vector) variables of $\mathbf{x}$ and $\mathbf{y}$ of length n, an optimization problem can be stated as

$$\begin{aligned} \underset{\mathbf{x},\mathbf{y}}{\text{minimize}} \quad & f(\mathbf{x},\mathbf{y}) \\ \text{subject to} \quad & \mathcal{C}_1 : h(\mathbf{x},\mathbf{y}) = 0 \\ & \mathcal{C}_2 : g(\mathbf{x},\mathbf{y}) \leq 0 \end{aligned} \tag{8.1}$$

Once the objective, $f(\mathbf{x},\mathbf{y}) : \{\mathbf{x},\mathbf{y}\} \in \mathbb{C}^n \rightarrow \mathbb{R}$, is a convex (concave) function

of the variables and the equality and inequality constraints are affine and convex (concave), respectively, the problem is convex and can be transformed to standard programming frameworks including linear, quadratic, semidefinite programming then there are standard optimization toolboxes that can be used to solve these problems. If the problem is convex, the KKT conditions are sufficient and necessary conditions for optimality and even if the problem can not be formulated as aforementioned programming techniques, it is still solvable provided that the KKT equations are solvable [111].

In contrast to convex problems, there is no clear classification of non-convex problems, nor is there a shortcut to summarize the broad range of heuristic algorithms [129]. The KKT equations, in general, may or may not have a closed form solution (system of nonlinear equations) and in particular for non-convex problems, the conditions are necessary but not sufficient for optimality. In communications theory, the existing problems are often non-convex and/or NP hard due to the nonlinear and/or discrete nature of the objective and constraints. Therefore, different applications and examples are of great interest and can contribute to this field of research.

In this thesis, different approaches are provided to handle the non-convex optimization problems that we encountered. Two examples are given in equations (6.9) and (7.9). Based on the author's experience on the topic, possible approaches for non-convex problems are briefly explained here.

1. The first step is transforming the constrained optimization problem to an unconstrained one, using the method of Lagrange multipliers in case of only equality constraints [130]. Gradient based optimization techniques including gradient descent, conjugate gradient and quasi Newton methods, are used to solve many types of optimization problems. This requires that the objective and the constraints are differentiable so the gradient and the Jacobian can be defined explicitly to form the Lagrangian equation (root finding problem). For non-convex problems the Lagrangian equation may have more than one root which corresponds to the local optima of the objective [131]. For general optimization problems with equality and inequality constraints, the KKT conditions need to be solved in order to solve the optimization problem. The solution for KKT conditions may be achieved numerically, e.g., the waterfilling solution for (5.29). For non-convex problems, the solution of the KKT conditions may converge to the global optimum, *most of the time*. This means that the statistical performance of the algorithm is acceptable with respect to the required accuracy. The performance and convergence of such algorithms

to the global optimum are highly dependent on the problem structure, and more specifically is not guaranteed for non-convex objective functions.

2. Proximal formulation of non-convex problems is another technique which approximates the original problem (objective and constraints) to the piecewise linear or quadratic functions which the convergence for is guaranteed. Other transformations of the problem to ones with known convergence properties (even if not convex) are also possible. The PAPR problem in Chapter 6 with the constant modulus algorithm that is used for solving the optimization problem, with some consideration, is an example of such an approach.

3. For multivariate optimization problems, alternating optimization techniques are among the popular algorithms. The idea for alternating programming is to fix one variable and solve the optimization problem with respect to the other variable. This process is repeated until the solution converges to the optimal one. Particularly, when the objective and constraints can be separated in terms of disjoint variables then the solution is equivalent to the original problem [111]. The representation of such separable problems is as follows

$$\begin{array}{lll} \text{minimize} & \text{minimize} & f(\mathbf{x},\mathbf{y}) \\ \mathbf{x} \in \mathcal{C}_1, \mathcal{C}_2 & \mathbf{y} \in \mathcal{C'}_1, \mathcal{C'}_2 & \end{array} \tag{8.2}$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are functions of $\mathbf{x}$ and $\mathcal{C'}_1$ and $\mathcal{C'}_2$ are functions of $\mathbf{y}$. The optimal solution is reached in (8.2) once both inner and outer problems are convex and strictly speaking solvable, see [132] for an example.

4. Sequential (successive) programming is another approach to solve nonlinear problems. Here, the objective and constrains are estimated with their Taylor expansion at each iteration and the approximation of the problem is solved instead of the original problem. The approximation is repeated till the solution converges to the optimum, although this is not guaranteed in general. Sequential quadratic programming (SQP) is classified within this category which is known to be a powerful tool for solving nonlinear optimization problems [133]. An example for the application of SQP can be found in [18]. Moreover, SQP is a strong candidate for solving the hybrid precoding problem that will be introduced in section 8.3. The difference between the the proximal method and the sequential programing is that the former approximate the problem once regardless of the initial guess, while the later uses successive approximation of the problem through the route to the solution.

5. Convex relaxation is a popular technique which projects the non-convex problem set to a larger convex set. The relaxed problem is solved instead of the original problem. The solution is hoped to be close to the solution of the original problem, and the tightness of the relaxation is the measure of the effectiveness of this method [111]. The proposed algorithms in Chapter 7 are categorized as such. This field of research is relatively new and there is an increasing interest on convex relaxation techniques for solving originally non-convex problems, due to availability of various solvers that can handle convex optimization problems.

6. One of the effective tools to approach non-convex problems is to solve the dual problem, which is convex by definition, instead of the primal problem. Lagrangian duality is the most common duality paradigm that is used in this context. However, the solution to the dual problem may not be equal to the primal solution and gives a lower bound on the objective value in the optimization problem. If the primal and dual solutions are identical then there exist a zero duality gap [111].

Note that the aforementioned techniques are not exclusive and are given based on the experience of the author within the scope of this thesis. For more general and complete classification, see [129]. These type of optimization problems can be considered as potential research topics to be explored further in the context of wireless communications and signal processing.

After this summary, we now look back to the research questions that were defined at the beginning of this thesis to evaluate the extent of the provided research and the main contributions.

8.2 Contributions to the Posed Research Questions

Research Question 1 *What limitations are imposed on the wireless link performance and therefore the design criteria, when the communication system is confined in a closed metal environment which is commonly the case for industrial machineries?*

In Chapter 2 of this thesis, a comprehensive stochastic channel model was proposed based on the 60 GHz measurement results within a metal cabinet. The estimated channel length (in the order of 1 μs) indicated an extremely long (slowly damping) channel impulse response for such a confined metal environment. One

distinguishing characterization of the measured channel is the continuous multipath reception, which neither follows the well-known clustering (Saleh-Valenzuela) model nor the common sparse representation of the wireless channel for millimeter waves. These measurement results opened up new opportunities for research in system design and underlying signal processing techniques for highly dispersive environments. This is due the unique properties of the investigated wireless channel which has not been reported before in the literature, to the best of our knowledge. Hence, this work can be referred to as one of the first millimeter wave channel characterizations for confined wireless applications.

Even though accurate channel characterization allows for precise and realistic channel simulations, it is often too complicated to consider in a signal processing model particularly for linear processing of the signal. Therefore a simpler model is established in Chapter 3 that can capture the most influential features of the measured wireless channel, i.e., the channel length and the power delay profile. Accordingly, we believe that the channel characterization which is offered in this work suffices to build a proper channel model for the design and investigation of signal processing algorithms.

Research Question 2 *What are the competitive equalization options which are capable of taming extremely dispersive wireless channels, and will the available techniques admit the high data rate, great reliability and low latency requirements of industrial applications?*

This question was addressed in Chapter 3 by giving a complete overview of competent linear equalization techniques for the considered linear system model. Single carrier modulation with frequency domain equalization at the receiver side or equivalently OFDM technique are concluded to be the most capable techniques to combat the inter-symbol interference (ISI) resulting from the frequency selective channel.

Motivated from the conclusion of Chapter 3, a wideband system model and a basic OFDM system design was proposed in Chapter 4 that can provide data-rates up to a few Gbps for the measured channels of Chapter 2. We believe these three chapters provide enough material to give a clear perspective on the system performance with respect to the different constraints on data-rate, latency and reliability of the system. In a nutshell, with the least complex processing (linear computational complexity), a latency in the order of a few tens of μs are foreseen. Also the data-rate (in the order of a few Gbps) and a reasonable bit-error rate was achieved to a good extent via Matlab simulations. The latency is determined by the block

processing delay (communication delay), and the processing delay which is related to the computational cost of the algorithms. The former is discussed in Chapter 4 and that has a direct relation with the spectral efficiency of the system. The latter is highly affected by the design as well as the implementation of the signal processing algorithms on a DSP, FPGA and etc. and results in variations in the output latency. This is worth investigation by the experts in this area.

Research Question 3 *How to reduce the peak-to-average-power-ratio (PAPR) efficiently and effectively in OFDM systems, particularly for multiple antenna systems which has been less studied in the literature?*

The PAPR is a rather old hardware related problem in digital communications in general, and for multicarrier systems like OFDM in particular. However, the PAPR problem is less investigated for MIMO-OFDM systems. To address the PAPR problem in (multi-user) MIMO-OFDM systems, an interesting PAPR reduction technique was proposed in Chapter 6. This algorithm is shown to be very effective via Matlab simulations. The proposed algorithm is transparent to the receiver and does not impose extra processing at the other end, which makes it unique compared to the existing techniques and is widely applicable to MIMO-OFDM systems of various types. To the best of our knowledge, there is no other algorithm reported in the literature that can offer a comparable PAPR reduction gain with such a low cost in terms of the computations, and no cost of bandwidth, BER and power consumption. However, the proposed technique relies on certain assumptions on the channel estimation and inversion at the receiver side and is not designed for systems which exploit models for channel correlations between consecutive OFDM blocks.

It is worth mentioning that the PAPR reduction is a timely problem and there is still ongoing research in this area specially for the emerging millimeter wave MIMO-OFDM systems with increasing hardware sensitivity. The author believes that the most effective signal processing algorithms to combat the high PAPR, are need to be developed in a close loop with the electronic front-end design and together with an RF design expert. Indeed, this is a multidisciplinary problem and is required to be treated as such, so a valuable follow up of this work will be implementation and evaluation of the actual PAPR reduction gain of the proposed algorithm on a test platform.

Research Question 4 *How to optimally use a MIMO system considering the existing hardware constraints?*

In Chapter 7, an antenna selection technique was studied and an inclusive framework was proposed to formulate this problem in the context of transmit precoding. The problem was stated as a joint antenna selection and precoding problem. This solves for the optimal sparse precoder and leads to using a subset of channels instead of the full channel with the ambition of lowering the signal processing burden at the transmitter as well as the receiver. The main contribution of Chapter 7 is the non-trivial relaxation techniques that are developed to transform the originally non-convex antenna selection problem to one that can be handled by available off-the-shelf convex solvers. The proposed antenna selection scheme can be seen as a convex optimization framework to formulate a general MIMO precoding problem considering the hardware constraints such as per antenna power constraints and limited available number of RF chains.

However, the research question above demands much more extensive research to fully accommodate different types of hardware constraints which can be handled by signal processing techniques. This requires first the identification and proper modeling of the hardware constraints where this process itself is a broad area of research and is impossible without a close collaboration with RF experts. The second step is introducing a convenient formulation and framework to represent these constraints effectively and later to trace appropriate signal processing tools that can solve the given problem. The focus of this thesis was to address the latter aspect for a very specific problem of MIMO precoding. Finally, we sense the necessity of an overarching literature survey that collects all the possible hardware limitations in this regard, particularly for the future millimeter wave systems.

To summarize, the optimal design of a wireless system in general is by solving for all design parameters simultaneously in one optimization problem, as they are all variables of one common objective which is the system performance, accordingly multidisciplinary design is the ultimate approach. Overall, we covered diverse topics in the field of wireless communications and broad range of areas have been explored, from channel modeling to signal processing and optimization techniques. However, there are areas that are not touched in this work regarding our main research question that was introduced in Chapter 1, i.e., **How to design a highly reliable short-range gigabit wireless link within a confined metal environment subject to a rigid latency requirement**. These mainly concern the practical requirements for the implementation of the system and their impact on the estimated performance and latency.

8.3 Future Work

In this section we provide some research directions that can be followed as a result of this thesis. First we introduce direct extensions of the proposed techniques that can be considered to improve or evaluate the performance.

8.3.1 Antenna Selection at Uplink

In traditional MIMO systems, precoding is performed digitally at the baseband where dedicated RF chains are required for each antenna element [67]. Unlike the antenna elements that can be manufactured and deployed in large scales, the RF chain components are expensive and bulky to install and maintain, particularly for emerging 60 GHz technology [134]. One possible solution to reduce the number of RF chains yet benefit from large scale MIMO systems is the hard antenna selection technique that is discussed in Chapter 7.

The proposed model in Chapter 7 aims at downlink antenna selection at the base station, however, it can be extended to include antenna selection in the uplink direction or more specifically at the user device. Nevertheless, the underlying signal processing algorithm to perform antenna selection at the user side needs to be computationally much more efficient compared to the one at the base station. This is due to the fact that user devices are commonly battery operated and require low power consumption. For example, alternative proximal sparsity inducing regularizers which are differentiable and leverage gradient based algorithms can be considered [135].

8.3.2 Hybrid Precoding

An alternative approach with respect to the hard antenna selection that is discussed in Chapter 7, is soft antenna selection, which is widely referred to as hybrid precoding. In hybrid precoding, beside baseband processing in the complex domain, RF processing is considered using analog phase shifters [136, 137, 138, 139]. This is by decomposing the precoder $\mathbf{W} \in \mathbb{C}^{M_t \times M_r}$ into two parts $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{M_t \times L}$ $\mathbf{W}_{\text{BB}} \in \mathbb{C}^{L \times M_r}$, where M_t and M_r are the number of transmit and receive antennas, respectively, and L is the number of available RF chains. The elements in $\mathbf{W}_{\text{RF}}$ is required to be constant modulus since only phase shifters are used to implement the RF precoder. Mathematically, hybrid precoding, for maximizing the capacity, is a hard (non convex) problem to solve. The main investigation is to formulate the problem so it can be solved using known optimization techniques. It is expected

to have better performance results for hybrid precoding technique compared to the antenna selection scheme for the same number of RF chains.

8.3.3 Imperfect Channel State Information

Throughout this thesis, we assumed that perfect channel information is available at the transmitter so the processing can be classified as deterministic approaches. However, in practice this is not the case. Specifically, when the feedback channel from the receiver is used to inform the transmitter about the channel, limited bandwidth with a coarse bit quantizations are used. This can lead to erroneous channel information and potentially degrades the performance of the wireless system. Considering this imperfection and deriving proper bounds to evaluate the model mismatch can be the next research topic for both hard and soft antenna selection techniques. Another type of imperfection can be considered when second order channel statistics such as a channel covariance matrix is available instead of the instantaneous channel information, and accordingly the statistical performance of the algorithms may be investigated with respect to this prior knowledge which is more towards the stochastic system design approach.

The aforementioned topics are the immediate follow ups of the work presented in this thesis. Now we will introduce more generic research topics that are made available as the outcome of this thesis, as well as some uncovered areas that are yet to be investigated.

8.3.4 Capacity Analysis for Highly Dispersive Channels

In Chapter 2 of this thesis, channel characterization was performed for an extremely reflective environment within a metal cabinet. The results are notably different from what has been reported in the literature. The channel frequency response shows quick fluctuations. As a consequence, the coherence bandwidth of the channel is significantly narrow. Initial investigations show that for the 5 GHz bandwidth, centered at 59.5 GHz, the channel is Gaussian distributed over frequency. This leads to a Rayleigh fading distribution in frequency domain and one can show that the Shannon capacity in such frequency selective channels is upper-bounded by the capacity of an AWGN channel with SNR given by $\gamma = \frac{P}{N\sigma_v^2}$, σ_v^2 and P being the noise power in each frequency band and the transmit power, respectively. Indeed. this result follows by averaging the capacity (bit per second per Hz) over relatively

many (N) frequency bands as

$$C = \frac{1}{N}\sum_{k=1}^{N} \log(1 + \frac{P|\beta|^2}{N\sigma_v^2}) = \mathbb{E}\{\log(1 + \frac{P|\beta|^2}{N\sigma_v^2})\}\,, \tag{8.3}$$

where N is the number of frequency bands and β is a Gaussian random variable which characterizes the fading in the frequency domain. Using Jensen's inequality

$$\mathbb{E}\{\log(1 + \frac{P|\beta|^2}{N\sigma_v^2})\} \leq \log(1 + \mathbb{E}\{|\beta|^2\}\frac{P}{N\sigma_v^2}) = \log(1 + \frac{P}{N\sigma_v^2})\,, \tag{8.4}$$

once $\mathbb{E}\{|\beta|^2\}$ is unity. This has not been reported before in the literature to the best of our knowledge and can be exploited to arrive at more profound capacity analysis for such channels. This can be viewed as an extension of the work to address *Research Question 1*.

Furthermore, the capacity of the fading wireless channel in complexity limited systems (including computational complexity) is not studied at all. This is tightly related to latency (delay) constrained systems including the wireless link for industrial machineries. Recently this problem is considered for a special receiver which uses a lattice search [140]. The maximization of the mutual information (to obtain the capacity) is performed subject to the linear complexity of the optimal receiver. This can be further investigated to give a more realistic notion of delay constrained channel capacity. In general, the mutual information framework offers a versatile quantitative measure to evaluate the effectiveness of the communication channel with respect to different choice of processors, which is not explored fairly in the signal processing literature. Therefore the communication rate per channel use is potentially a global measure that can fairly describe the overall as well as step by step efficiency of the system and can replace many of the local measures that are used for system design.

Finally, we did not look at nonlinear signal processing techniques due to their complications and susceptibility to hardware imperfections. This can be further investigated in the context of this thesis. Also, distributed algorithms for ad-hoc implementation of wireless sensors and decentralized approaches are not covered in this thesis and can be considered for future extensions.

We hope that the research directions that are provided in this chapter can open up new ideas to be investigated in the field of wireless communication and signal processing theory. There are still many challenges to be identified and handled to promise impeccable wireless connections, for demanding industrial and secure applications which often take place in alternative and unusual propagation

environments. However, the author believes that this work can introduce a fair example of an academic approach to tackle a rather wide and multidisciplinary research/engineering problem.

Bibliography

[1] H. Yang, "Towards low-cost gigabit wireless systems at 60 GHz channel modeling and baseband design," Ph.D. dissertation, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2008.

[2] IEEE80211ad, "IEEE draft standard for local and metropolitan area networks - specific requirements - part 11: Wireless lan medium access control (MAC) and physical layer (PHY) specifications - amendment 3: Enhancements for very high throughput in the 60 GHz band," *IEEE P802.11ad/D5.0, September 2011, (Draft Amendment based on IEEE P802.11REVmb D10.0) as amended by IEEE 802.11ae D5.0 and IEEE 802.11aa D6.0)*, pp. 1–601, Dec 2011.

[3] IEEE80215c, "IEEE standard for high rate wireless personal area networks (WPANs): Millimeter-wave-based alternative physical layer extension," *IEEE Std 802.15.3c-2009 (Amendment to IEEE Std 802.15.3-2003)*, pp. 1–187, December 2009.

[4] C. A. Balanis, *Antenna theory: analysis and design*. J. Wiley, New York, 1982.

[5] S. J. Orfanidis, *Electromagnetic Waves and Antennas*. ECE Department, Rutgers University, 2004.

[6] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Mag.*, vol. 49, no. 6, pp. 101–107, 2011.

[7] N. Guo, R. C. Qiu, S. S. Mo, and K. Takahashi, "60GHz millimeter-wave radio: Principle, technology, and new results," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, no. 1, pp. 48–48, Jan. 2007.

[8] A. Corovic, L. Kadric, N. Lipa, S. Opanovic, and N. Nosovic, "emuseum - more interesting way of learning from the past," in *MIPRO, 2012 Proceedings of the 35th International Convention*, May 2012, pp. 900–904.

[9] K. Kawasaki, Y. Akiyama, K. Komori, M. Uno, H. Takeuchi, T. Itagaki, Y. Hino, Y. Kawasaki, K. Ito, and A. Hajimiri, "A millimeter-wave intra-connect solution," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, Feb 2010, pp. 414–415.

[10] J. Karedal, A. Singh, F. Tufvesson, and A. Molisch, "Characterization of a computer board-to-board ultra-wideband channel," *Communications Letters, IEEE*, vol. 11, no. 6, pp. 468–470, June 2007.

[11] Z. M. Chen and Y.-P. Zhang, "Inter-chip wireless communication channel: Measurement, characterization, and modeling," *Antennas and Propagation, IEEE Transactions on*, vol. 55, no. 3, pp. 978–986, March 2007.

[12] H. Hashemi, "The indoor radio propagation channel," *Proc. of the IEEE*, vol. 81, no. 7, pp. 943–968, Jul. 1993.

[13] D. Cassioli, M. Win, and A. Molisch, "The ultra-wide bandwidth indoor channel: from statistical model to simulations," *Selected Areas in Communications, IEEE Journal on*, vol. 20, no. 6, pp. 1247–1257, Aug 2002.

[14] J. Karedal, S. Wyne, P. Almers, F. Tufvesson, and A. Molisch, "A measurement-based statistical model for industrial ultra-wideband channels," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 8, pp. 3028–3037, August 2007.

[15] S. Ghassemzadeh, R. Jana, C. Rice, W. Turin, and V. Tarokh, "Measurement and modeling of an ultra-wide bandwidth indoor channel," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1786–1796, 2004.

[16] A. Molisch, J. Foerster, and M. Pendergrass, "Channel models for ultra-wideband personal area networks," *Wireless Communications, IEEE*, vol. 10, no. 6, pp. 14–21, Dec 2003.

[17] M. Ohira, T. Umaba, S. Kitazawa, H. Ban, and M. Ueba, "Experimental characterization of microwave radio propagation in ict equipment for wireless harness communications," *Antennas and Propagation, IEEE Transactions on*, vol. 59, no. 12, pp. 4757–4765, Dec 2011.

[18] S. Khademi, T. Svantesson, M. Viberg, and T. Eriksson, "Peak-to-average-power-ratio (PAPR) reduction in WiMAX and OFDM/A systems," *EURASIP Journal on advances in signal processing*, August 2011.

[19] A. Goldsmith and A. Nin, *Wireless Communications*. Cambridge University Press, 2005.

[20] R. Frank, "Wireless technologies simplify wiring harness," *Auto Electron*, pp. 17–22, July 2007.

[21] N. Nakamoto, H. Ban, T. Oka, S. Kitazawa, K. Kobayashi, N. Kikuchi, H. Hatamoto, S. Shimizu, and M. Hara, "Wireless harness inside ict equipments," in *Advanced Communication Technology (ICACT), 2013 15th International Conference on*, Jan 2013, pp. 135–43.

[22] F. Giannetti, M. Luise, and R. Reggiannini, "Mobile and personal communications in the 60 GHz band: A survey," *Wireless Personal Communications*, vol. 10, no. 2, pp. 207–243, 1999.

[23] R. Daniels and R. Heath, "60 GHz wireless communications: emerging requirements and design recommendations," *IEEE Veh. Technol. Mag.*, vol. 2, no. 3, pp. 41–50, Sept. 2007.

[24] R. Daniels, J. Murdock, T. Rappaport, and R. Heath, "60 GHz wireless: Up close and personal," *IEEE Microw. Mag.*, vol. 11, no. 7, pp. 44–50, Dec. 2010.

[25] P. F. M. Smulders, "Statistical characterization of 60 GHz indoor radio channels," *IEEE Trans. Antennas Propag.*, vol. 57, no. 10, pp. 2820–2829, 2009.

[26] N. Moraitis and P. Constantinou, "Measurements and characterization of wideband indoor radio channel at 60 GHz," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 880–889, 2006.

[27] M.-S. Choi, G. Grosskopf, and D. Rohde, "Statistical characteristics of 60 GHz wideband indoor propagation channel," in *Proc. of PIMRC*, vol. 1, Sept. 2005, pp. 599 –603.

[28] C. Gustafson, K. Haneda, S. Wyne, and F. Tufvesson, "On mm-wave multipath clustering and channel modeling," *Antennas and Propagation, IEEE Transactions on*, vol. 62, no. 3, pp. 1445–1455, March 2014.

[29] T. Rappaport, F. Gutierrez, E. Ben-Dor, J. Murdock, Y. Qiao, and J. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *Antennas and Propagation, IEEE Transactions on*, vol. 61, no. 4, pp. 1850–1859, April 2013.

[30] W. Fu, J. Hu, and S. Zhang, "Frequency-domain measurement of 60 GHz indoor channels: a measurement setup, literature data, and analysis," *Instrumentation Measurement Magazine, IEEE*, vol. 16, no. 2, pp. 34–40, April 2013.

[31] S. Geng, J. Kivinen, X. Zhao, and P. Vainikainen, "Millimeter-wave propagation channel characterization for short-range wireless communications," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 1, pp. 3–13, Jan 2009.

[32] M. Peter, W. Keusgen, A. Kortke, and M. Schirrmacher, "Measurement and analysis of the 60 GHz in-vehicular broadband radio channel," in *Proc. of VTC*, Oct. 2007, pp. 834–838.

[33] K. K. O, K. Kim, B. A. Floyd *et al.*, "On-chip antennas in silicon ICs and their application," *Electron Devices, IEEE Transaction on*, vol. 52, no. 7, pp. 1312–1323, 2005.

[34] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40 –60, Jan. 2013.

[35] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *Selected Areas in Communications, IEEE Journal on*, vol. 5, no. 2, pp. 128–137, February 1987.

[36] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, 1996.

[37] MATLAB, *version 8.2.0.701 (R2013b)*. Natick, Massachusetts: The MathWorks Inc., 2013.

[38] M. Patzold, A. Szczepanski, and N. Youssef, "Methods for modeling of specified and measured multipath power-delay profiles," *Vehicular Technology, IEEE Transactions on*, vol. 51, no. 5, pp. 978–988, Sep 2002.

[39] J. Parsons, D. Demery, and A. Turkmani, "Sounding techniques for wideband mobile radio channels: a review," *Proc. of IPCSV*, vol. 138, no. 5, pp. 437–446, Oct. 1991.

[40] Z. Irahhauten, *Ultra-Wideband Wireless Channel: Measurements, Analysis and Modeling*. Delft University of Technology, Dept. EEMCS, The Netherlands, 2008.

[41] A. Siamarou and M. Al-Nuaimi, "A wideband frequency-domain channel-sounding system and delay-spread measurements at the license-free 57 to 64 GHz band," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 3, pp. 519–526, 2010.

[42] Z. Irahhauten, H. Nikookar, and G. J. M. Janssen, "An overview of ultra wide band indoor channel measurements and modeling," *IEEE Microw. Wireless Compon. Lett.*, vol. 14, no. 8, pp. 386–388, Aug. 2004.

[43] Z. Irahhauten, A. Yarovoy, G. J. M. Janssen, H. Nikookar, and L. Ligthart, "Suppression of noise and narrowband interference in uwb indoor channel measurements," in *Proc. of ICU*, 2005, pp. 108–112.

[44] Z. Irahhauten, A. Yarovoy, G. Janssen, H. Nikookar, and L. Ligthart, "Ultra-wideband indoor propagation channel: Measurements, analysis and modeling," in *Proc. of EuCAP*, Nov. 2006, pp. 1–6.

[45] TG3c, "Task group 3c (TG3c) channel modeling sub-committee final report," IEEE P802.15 working group for wireless personal area networks (WPANs), Tech. Rep., Mar. 2007.

[46] A. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, Jan. 1987.

[47] Q. Spencer, M. Rice, B. Jeffs, and M. Jensen, "A statistical model for angle of arrival in indoor multipath propagation," in *Proc.of VTC*, vol. 3, 1997, pp. 1415–1419.

[48] IEEE802154, "IEEE standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks specific requirements part 15.4: Wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (LR-WPANs)," *IEEE Std 802.15.4/2003*, pp. 1–670, 2003.

[49] P. Smulders and A. Wagemans, "Wideband indoor radio propagation measurements at 58 GHz," *Electronics Letters*, vol. 28, no. 13, pp. 1270–1272, June 1992.

[50] H. Yang, M. H. A. J. Herben, and P. F. M. Smulders, "Impact of antenna pattern and reflective environment on 60 GHz indoor radio channel characteristics," *Antennas and Wireless Propagation Letters, IEEE*, vol. 4, pp. 300–303, 2005.

[51] R. Sun and D. Matolak, "Characterization of the 5-GHz elevator shaft channel," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 10, pp. 5138–5145, October 2013.

[52] J. Liberti and T. Rappaport, "A geometrically based model for line-of-sight multipath radio channels," in *Vehicular Technology Conference, 1996. Mobile Technology for the Human Race., IEEE 46th*, vol. 2, Apr 1996, pp. 844–848.

[53] C.-F. Yang, B.-C. Wu, and C.-J. Ko, "A ray-tracing method for modeling indoor wave propagation and penetration," *Antennas and Propagation, IEEE Transactions on*, vol. 46, no. 6, pp. 907–919, Jun 1998.

[54] O. Hashimoto, T. Abe, Y. Hashimoto, T. Tanaka, and K. Ishino, "Realization of resistive-sheet type wave absorber in 60 GHz frequency band," *Electronics Letters*, vol. 30, no. 8, pp. 657–658, 1994.

[55] T. Kaiser, F. Zheng, and E. Dimitrov, "An overview of ultra-wide-band systems with MIMO," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 285–312, Feb 2009.

[56] C. D. Iskandar, *A MATLAB based Object Oriented Approach to Multipath Fading Channel Simulation*, HiTek Multisystems 7945 Avenue de Cornouailles, Quebec, QC, Canada, G1H 3V9, 2008.

[57] G. Leus and A. van der Veen, "Channel estimation," in *Smart Antennas: State-of-the-Art*, T. Kaiser, A. Bourdoux, H. Boche, J. R. Fonollosa, J. B. Andersen, and W. Utschick, Eds. Hindawi, 2005, pp. 293–320.

[58] J. G. Proakis and M. Salehi, *Fundamentals of Communication Systems*. Prentice Hal, 2005.

[59] T. Walzman and M. Schwartz, "Automatic equalization using the discrete frequency domain," *Information Theory, IEEE Transactions on*, vol. 19, no. 1, pp. 59–68, Jan 1973.

[60] U. Madhow, *Fundamentals of Digital Communication*. Cambridge University Press, 2008, chapter 5.

[61] F. Pancaldi, G. Vitetta, R. Kalbasi, N. Al-Dhahir, M. Uysal, and H. Mheidat, "Single-carrier frequency domain equalization," *IEEE Signal Process. Mag.*, vol. 25, no. 5, pp. 37–56, 2008.

[62] G. Vitetta, D. P. Taylor, G. Colavolpe, F. Pancaldi, and P. A. Martin, *Wireless Communications: Algorithmic Techniques*. J. Wiley, New York, 2013.

[63] I. Barhumi, G. Leus, and M. Moonen, "Time-domain and frequency-domain per-tone equalization for OFDM over doubly selective channels," *Signal Processing*, vol. 84, no. 11, pp. 2055–2066, 2004.

[64] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers part 2: Blind channel estimation, synchronization, and direct equalization," *IEEE Trans. Signal Processing*, vol. 47, no. 7, pp. 2007–2022, July 1999.

[65] B. Muquet, Z. Wang, G. Giannakis, M. de Courville, and P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier transmissions?" *Communications, IEEE Transactions on*, vol. 50, no. 12, pp. 2136–2148, Dec 2002.

[66] J. Tubbax, B. Come, L. V. der Perre, L. Deneire, S. Donnay, and M. Engels, "OFDM versus single carrier with cyclic prefix: a system-based comparison," in *Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th*, vol. 2, 2001, pp. 1115–1119.

[67] L. C. Godara, *Handbook of Antennas in Wireless Communications*. Boca Raton, FL, USA: CRC Press, Inc., 2001.

[68] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, July 2006.

[69] C. E. Shannon, "Communication in the presence of noise," *Proc. Institute of Radio Engineers*, vol. 37, no. 1, pp. 10–21, 1949.

[70] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *Selected Areas in Communications, IEEE Journal on*, vol. 21, no. 5, pp. 684–702, 2003.

[71] A. Narula, M. Lopez, M. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 8, pp. 1423–1436, Oct 1998.

[72] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *Wireless Communications, IEEE Transactions on*, vol. 3, no. 4, pp. 1165–1175, July 2004.

[73] A. Moustakas, S. Simon, and A. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: a (not so) large n analysis," *Information Theory, IEEE Transactions on*, vol. 49, no. 10, pp. 2545–2561, Oct 2003.

[74] M. S. Alouini and A. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *Vehicular Technology, IEEE Transactions on*, vol. 48, no. 4, pp. 1165–1181, Jul 1999.

[75] E. Biglieri, R. Calderbank, A. Constantinides, and A. Goldsmith, *MIMO Wireless Communications*. Cambridge University Press, 2007.

[76] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1073–1096, May 2003.

[77] E. G. Larsson and P. Stocia, *Space-Time Block Coding for Wireless Communication*. Cambridge University Press, 2003.

[78] S. Khademi, *OFDM Peak-to-Average-Power-Ratio Reduction in WiMAX Systems*. Chalmers University of Technology, 2010, M.S. thesis.

[79] M. Sharif, M. Gharavi-Alkhansari, and B. Khalaj, "On the peak-to-average power of ofdm signals based on oversampling," *Communications, IEEE Transactions on*, vol. 51, no. 1, pp. 72–78, Jan 2003.

[80] G. Wunder and H. Boche, "Upper bounds on the statistical distribution of the crest-factor in ofdm transmission," *Information Theory, IEEE Transactions on*, vol. 49, no. 2, pp. 488–494, Feb 2003.

[81] R. O'Neill and L. Lopes, "Envelope variations and spectral splatter in clipped multicarrier signals," in *Personal, Indoor and Mobile Radio Communications, 1995. PIMRC'95. Wireless: Merging onto the Information Superhighway., Sixth IEEE International Symposium on*, vol. 1, Sep 1995, pp. 71–75.

[82] X. Li and L. Cimini, "Effects of clipping and filtering on the performance of ofdm," *Communications Letters, IEEE*, vol. 2, no. 5, pp. 131–133, May 1998.

[83] J. Armstrong, "Peak-to-average power reduction for ofdm by repeated clipping and frequency domain filtering," *Electronics Letters*, vol. 38, no. 5, pp. 246–247, Feb 2002.

[84] J.Tellado, "Peak to average power reduction in multicarrier modulation," Ph.D. dissertation, Stanford University, 2000.

[85] B. S. Krongold and D. Jones, "PAR reduction in OFDM via active constellation extension," in *Acoustics Speech and Signal Processing (ICASSP)*, vol. 4, April 2003.

[86] R. Bauml, R. Fischer, and J. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," *Electronics Letters*, vol. 32, no. 22, pp. 2056–2057, Oct 1996.

[87] S. Muller and J. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *IEEE Electronics Lett.*, vol. 33, no. 5, pp. 368–369, February 1997.

[88] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Communications Mag.*, vol. 12, no. 2, pp. 56–65, April 2005.

[89] Y. Shen and E. Martinez, "WiMAX channel estimation: algorithms and implementations," *Freescale Semiconductor*, 2007.

[90] S. Khademi, A. van der Veen, and T. Svantesson, "Precoding technique for peak-to-average-power-ratio (PAPR) reduction in MIMO OFDM/A systems," in *Acoustics Speech and Signal Processing (ICASSP)*, march 2012, pp. 3005–3008.

[91] J. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals." *IEEE trans. Acoustics, Speech and Signal Processing*, vol. 31, pp. 459–471, April 1983.

[92] J. Wang, Z. Lan, R. Funada, and H. Harada, "On scheduling and power allocation over multiuser MIMO-OFDMA: Fundamental design and performance evaluation in WiMAX systems," *IEEE Int. Symposium on Personal, Indoor and Mobile Radio Communications.*, pp. 2752–2756, 2009.

[93] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *Communications Magazine, IEEE*, vol. 42, no. 10, pp. 68–73, October 2004.

[94] A. F. Molisch and M. Z. Win, "MIMO systems with antenna selection," *IEEE Microwave Magazine*, vol. 5, no. 1, pp. 46–56, March 2004.

[95] A. Gorokhov, D. Gore, and A. Paulraj, "Receive antenna selection for MIMO spatial multiplexing: theory and algorithms," *Signal Processing, IEEE Transactions on*, vol. 51, no. 11, pp. 2796–2807, November 2003.

[96] M. Gharavi-Alkhansari and A. Gershman, "Fast antenna subset selection in MIMO systems," *Signal Processing, IEEE Transactions on*, vol. 52, no. 2, pp. 339–347, February 2004.

[97] R. Vaze and H. Ganapathy, "Sub-modularity and antenna selection in MIMO systems," *Communications Letters, IEEE*, vol. 16, no. 9, pp. 1446–1449, September 2012.

[98] A. Dua, K. Medepalli, and A. Paulraj, "Receive antenna selection in MIMO systems using convex optimization," *Wireless Communications, IEEE Transactions on*, vol. 5, no. 9, pp. 2353–2357, September 2006.

[99] R. Heath, S. Sandhu, and A. Paulraj, "Antenna selection for spatial multiplexing systems with linear receivers," *Communications Letters, IEEE*, vol. 5, no. 4, pp. 142–144, April 2001.

[100] D. Gore and A. Paulraj, "Space-time block coding with optimal antenna selection," in *Acoustics Speech and Signal Processing (ICASSP)*, vol. 4, 2001, pp. 2441–2444.

[101] ——, "MIMO antenna subset selection with space-time coding," *Signal Processing, IEEE Transactions on*, vol. 50, no. 10, pp. 2580–2588, October 2002.

[102] F. Wang and M. Bialkowski, "A joint design of transmit antenna selection and multiuser scheduling for multiuser MIMO systems employing block diagonalization precoding scheme," in *Wireless Communications and Signal Processing (WCSP)*, November 2011, pp. 1–5.

[103] S. Sigdel and W. Krzymien, "Simplified fair scheduling and antenna selection algorithms for multiuser MIMO orthogonal space-division multiplexing downlink," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 3, pp. 1329–1344, March 2009.

[104] S. Park and D. Love, "Capacity limits of multiple antenna multicasting using antenna subset selection," *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2524–2534, June 2008.

[105] P.-H. Lin and S.-H. Tsai, "Performance analysis and algorithm designs for transmit antenna selection in linearly precoded multiuser MIMO systems," *Vehicular Technology, IEEE Transactions on*, vol. 61, no. 4, pp. 1698–1708, May 2012.

[106] M. Sadek, A. Tarighat, and A. Sayed, "Active antenna selection in multiuser MIMO communications," *Signal Processing, IEEE Transactions on*, vol. 55, no. 4, pp. 1498–1510, April 2007.

[107] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *Signal Processing, IEEE Transactions on*, vol. 52, no. 2, pp. 461–471, February 2004.

[108] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *37th Annual Allerton Conference on Communication, Control, and Computing*, 1999, pp. 987–996.

[109] T. Yoo and A. Goldsmith, "On the optimality of multi antenna broadcast scheduling using zero-forcing beamforming," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 3, pp. 528–541, March 2006.

[110] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *Signal Processing, IEEE Transactions on*, vol. 56, no. 9, pp. 4409–4418, September 2008.

[111] S. Boyd and L. Vandenberghe, *Convex Optimization: Convex optimization problems, Equivalent problems*. New York, NY, USA: Cambridge University Press, 2004.

[112] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *Signal Processing, IEEE Transactions on*, vol. 54, no. 1, pp. 161–176, January 2006.

[113] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *Signal Processing, IEEE Transactions on*, vol. 54, no. 6, pp. 2239–2251, June 2006.

[114] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *arXiv*, vol. abs/1108.0775, 2011.

[115] S. J. Kim, S. Jain, and G. Giannakis, "Backhaul-constrained multi-cell cooperation using compressive sensing and spectral clustering," in *Signal Processing Advances in Wireless Communications (SPAWC)*, June 2012, pp. 65 –69.

[116] M. Yukawa and I. Yamada, "Minimal antenna-subset selection under capacity constraint for power-efficient MIMO systems: A relaxed minimization approach," in *Acoustics Speech and Signal Processing (ICASSP)*, March 2010, pp. 3058–3061.

[117] S. E. Nai, W. Ser, Z. L. Yu, and H. Chen, "Beampattern synthesis for linear and planar arrays with antenna selection by convex optimization," *Antennas and Propagation, IEEE Transactions on*, vol. 58, no. 12, pp. 3923–3930, December 2010.

[118] M. Hong, R. Sun, H. Baligh, and Z. Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 226–240, February 2013.

[119] S. Khademi, S. Chepuri, G. Leus, and A. J. van der Veen, "Zero-forcing preequalization with transmit antenna selection in MIMO systems," in *Acoustics Speech and Signal Processing (ICASSP)*, May 2013, pp. 5046–5050.

[120] O. Mehanna, N. Sidiropoulos, and G. Giannakis, "Multicast beamforming with antenna selection," in *Signal Processing Advances in Wireless Communications (SPAWC)*, June 2012, pp. 70–74.

[121] ——, "Joint multicast beamforming and antenna selection," *Signal Processing, IEEE Transactions on*, vol. 61, no. 10, pp. 2660–2674, May 2013.

[122] Friedman, T. Hastie, and R. Tibshirani, "A note on the group Lasso and a sparse group Lasso," *arXiv*, vol. abs/1001.0736v1, 2010.

[123] Z. Q. Luo, W. K. Ma, A.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *Signal Processing Magazine, IEEE*, vol. 27, no. 3, pp. 20–34, May 2010.

[124] E. Karipidis, N. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *Signal Processing, IEEE Transactions on*, vol. 56, no. 3, pp. 1268–1279, March 2008.

[125] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighed ℓ_1-norm minimization," 2007.

[126] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," September 2012.

[127] L. Vandenberghe, S. Boyd, and S. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, pp. 499–533, 1998.

[128] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, pp. 49–95, 1994.

[129] M. Chiang, *Nonconvex Optimization for Communication Networks*. Boston, MA: Springer US, 2009, pp. 137–196.

[130] Y. Nesterov, *Introductory lectures on convex optimization : a basic course*, ser. Applied optimization. Boston, Dordrecht, London: Kluwer Academic Publ., 2004.

[131] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Mathematical Optimization, 1996.

[132] S. Khademi, R. C. Hendriks, and W. B. Kleijn, “Jointly optimal near-end and far-end multi-microphone speech intelligibility enhancement based on mutual information,” in *Acoustics Speech and Signal Processing (ICASSP)*, March 2016, pp. 654–658.

[133] P. T. Boggs and J. W. Tolle, *Sequential quadratic programming*. Acta Numerica, 1995.

[134] S. Khademi, E. DeCorte, G. Leus, and A. J. van der Veen, “Convex optimization for joint zero-forcing and antenna selection in multiuser MISO systems,” in *Signal Processing Advances in Wireless Communications (SPAWC)*, June 2014.

[135] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, “Optimization with sparsity-inducing penalties,” *arXiv*, vol. abs/1108.0775, 2011. [Online]. Available: http://arxiv.org/abs/1108.0775

[136] X. Zhang, A. F. Molisch, and S.-Y. Kung, “Variable-phase-shift-based rf-baseband codesign for MIMO antenna selection,” *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4091–4103, Nov 2005.

[137] F. Sohrabi and W. Yu, “Hybrid digital and analog beamforming design for large-scale MIMO systems,” in *Acoustics Speech and Signal Processing (ICASSP)*, April 2015, pp. 2929–2933.

[138] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, March 2014.

[139] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, “Hybrid precoding for millimeter wave cellular systems with partial channel knowledge,” in *Information Theory and Applications Workshop (ITA)*, February 2013, pp. 1–5.

[140] J. Jalden and P. Elia, “DMT optimality of LR-aided linear decoders for a general class of channels, lattice designs, and system models,” *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4765–4780, October 2010.

Propositions

1. The frequency domain response of the channel within a metal enclosure is sparse.
2. A priori knowledge of a communication channel is essential for an optimal system design as the information sent over a “bad” channel cannot be retrieved by any means of processing.
3. A suboptimal solution is often more valuable in terms of the performance-complexity trade-off, nevertheless knowledge of the optimal solution is a great aid to quantify this trade-off.
4. Linear, quadratic and semidefinite programming are analogues to drawing on a flat, cylindrical and spherical surface, respectively. Solving non-convex optimization problems is like drawing on a crumpled ball of paper.
5. Teaching and learning can be modeled as stochastic (noisy) processes so the mutual information between the teacher and the student is maximized by recognition and characterization of the noise that is present.
6. The ubiquitous availability of wireless connections has introduced the habitual use of smartphones which has absurdly reduced real human interactions, contrary to the genuine objective of communication.
7. In a world played by the rules set by a masculine mentality, it is less likely that people with feminine skills would excel and visa versa.
8. Individualism is an inevitable part of modernism that, if left untamed, could lead humanity to a new level of self-destruction.
9. Self-reference is a paradox in most religious arguments.
10. “... there is no more dreadful punishment than futile and hopeless labor.” Albert Camus, The Myth of Sisyphus

These propositions are regarded as opposable and defendable, and have been approved as such by the promotor, Prof. Alle-Jan ven der Veen.

Summary

The advent of the digital era has revolutionized many aspect of our society and has significantly improved the quality of our lives. Consequently, signal processing has gained a considerable attention as the science behind the digital life. Among different applications for signal processing theory and algorithms, wireless communications remains one of the attractive and popular ones due to the widespread use of mobile devices.

This thesis is dedicated to develop signal processing algorithms to design high-speed wireless transceivers that can perform in highly reflective and harsh environments. The start of this research work initiated as a collaboration between TU Delft and an industrial partner, on a research aimed at short range gigabit wireless link within a lithography machine. The underlying unique wireless environment, together with the challenging specifications of the communication link for mechatronic systems, made this a compelling research project.

The first part of this research work focuses on constructing a reliable propagation model for dispersive environments, based on actual measurements. In our opinion it is crucial to have decent models to build effective theory and applications upon it. We developed a statistical channel model for the 60 GHz band for the extreme case of a confined metal enclosure in order to evaluate and test the existing signal processing algorithms under such pessimistic ambient conditions. This unique experiment opened up new research challenges to look back to popular design paradigms and reevaluate them with respect to the proposed channel model with a delay spread in the order of microseconds.

The concept of orthogonal frequency division multiplexing (OFDM) transmission was revisited and a customized OFDM system was designed which meets the data rate requirements of the mechatronic system of interest. The effectiveness of

the proposed OFDM design was examined via Matlab simulations using the measured and modeled channels. Interestingly, the performance of the OFDM system is not heavily affected by the frequency selectivity of the extreme propagation environment. The loss is mainly due to the time guard that is dedicated to avoid interference between consecutive OFDM blocks, suggesting the use of longer OFDM blocks to minimize the bandwidth loss.

The second part of this thesis is dedicated to multiple-input multiple-output (MIMO) systems versus the single-input single-output (SISO) system which was studied in the first part. The emphasis is on general challenges in high speed (wideband) communication systems rather than the specific wireless link within a mechatronic machine. Challenging research questions are posed regarding the design and implementation of MIMO systems. This part starts with a brief introduction to such systems and redefining our system model with respect to the MIMO setting and it continues by revisiting the timely problem of peak-to-average power-ratio (PAPR) reduction in OFDM systems, which deals with stochastic (data-dependent) OFDM waveforms, and the proposal of an effective algorithm to handle this challenge within the MIMO context . The hard problem of antenna selection for a MIMO system was considered at the end by investigating different linear precoding designs subject to the realistic hardware constraints including per antenna power constraints (rather than a conventional total power constraint) and limited number of RF chains.

The major content of this thesis concerns offering alternative formulations and optimization problems for transmitter design in the context of linear signal processing, to include hardware constraints which are more critical in emerging millimeter wave wireless systems. This requires the reformulation and relaxation of non-convex and hard design problems to make them suitable for available optimization tools, including sub-optimal but less computationally demanding algorithms based on non-convex optimization theory. A short classification of these non-convex optimization techniques is given as part of the conclusion in the last chapter of this thesis.

Samenvatting

De komst van het digitale tijdperk heeft vele aspecten van onze samenleving veranderd en heeft de kwaliteit van leven significant verbeterd. Als de wetenschap achter het digitale leven staat signaalbewerking in de belangstelling. Onder de toepassingen van signaalbewerkings-theorie en algoritmes is draadloze communicatie een van de aantrekkelijkste en populairste, dit vanwege het wijdverbreide gebruik van mobiele apparaten.

Dit proefschrift behandelt signaalbewerkings-algorimes voor het ontwerp van hoge-snelheid draadloze transceivers die kunnen opereren in moeilijke omgevingen met veel reflecties. Dit onderzoek begon als een samenwerking van TU Delft met een industriele partij, gericht op een giga-bit draadloze verbinding over korte afstanden binnen een lithografie-machine. De onderliggende unieke draadloze omgeving, samen met de uitdagende specificaties van de vereiste communicatieverbinding voor mechatronische machines, maakte dit een interessant onderzoeksproject.

Het eerste deel van dit onderzoek richt zich op het verkrijgen van een betrouwbaar propagatiemodel voor dispersieve omgevingen, gebaseerd op echte metingen. Naar onze mening is het cruciaal om geschikte modellen te hebben om daarop effectieve theorie en toepassingen te kunnen baseren. We ontwikkelden een statistisch kanaalmodel voor de 60 GHz band voor het extreme geval van een dichte metalen kast om daarmee de bestaande signaalbewerkings-algoritmes te kunnen testen onder zulke negatieve omstandigheden. Dit unieke experiment gaf aanleiding tot nieuwe onderzoeksvragen rond populaire ontwerptechnieken en het evalueren hiervan in het licht van het voorgestelde kanaalmodel met een spreiding in de tijd in de orde van microseconden.

Het OFDM concept voor het versturen van data is opnieuw bekeken en een aangepast OFDM systeem is ontworpen waarmee de benodigde data-snelheden

van het beschouwde mechatronische systeem gehaald kunnen worden. De effectieviteit van het voorgestelde OFDM ontwerp is onderzocht middels Matlab simulaties gebruikmakend van de gemeten en gemodelleerde kanalen. Interessant genoeg blijkt dat de prestatie van het OFDM systeem niet sterk wordt beïnvloed door de frequentie-selectiviteit van de extreem reflectieve omgeving. Het verlies aan bandbreedte is hoofdzakelijk een gevolg van de ongebruikte tijd-intervallen die toegevoegd zijn om storing tussen opeenvolgende OFDM blokken te vermijden. Het gebruik van langere OFDM blokken zou dit verlies kunnen verminderen.

Het tweede deel van dit proefschrift behandelt multi-antenne (MIMO) systemen, in tegenstelling tot het enkele-antenne (SISO) systeem dat in het eerste deel is bekeken. De nadruk ligt op algemene uitdagingen rond hoge-snelheids (breedband) communicatiesystemen, in plaats van de specifieke draadloze verbinding binnen een mechatronisch systeem. Uitdagende onderzoeksvragen rond het ontwerp en de implementatie van MIMO systemen worden gesteld. Dit deel begint met een korte inleiding voor zulke systemen, en het herdefinieren van ons systeemmodel naar deze MIMO situatie, en het vervolgt met het opnieuw bekijken van het actuele probleem van het reduceren van de verhouding van de piek tot het gemiddelde vermogen (PAPR) in OFDM systemen, en het afleiden van een effectief algoritme om dit in de context van MIMO systemen te verbeteren. Tot slot is het moeilijke probleem van het kiezen van de beste deelverzameling van antennes in een MIMO systeem, in combinatie met verschillende lineaire precoding technieken met realistische beperkingen op de hardware, waaronder een vermogenslimiet voor iedere antenne apart (in plaats van de gebruikelijke beperking op het totale vermogen), en een beperkt aantal RF ontvangers.

Het merendeel van dit proefschrift gaat over het verkrijgen van alternatieve formuleringen en optimalisatieproblemen voor zenderontwerp, beschreven door middel van lineaire algebra, en het rekening houden met beperkingen aan de hardware die des te belangrijker worden in toekomstige millimeter-golf draadloze systemen. Dit vereist het herformuleren en afzwakken van niet-convexe moeilijke optimalisatieproblemen om deze geschikt te maken voor beschikbare optimalisatieprogramma's, waaronder sub-optimale maar minder complexe algoritmes, gebaseerd op niet-convexe optimalisatietheorie. Een kort overzicht van deze niet-convexe optimalisatietechnieken is gegeven als onderdeel van de conclusie in het laatste hoofdstuk van dit proefschrift.

Acknowledgments

It has been a long journey and luckily I have not been alone along the road. I owe my achievements and successes to all the people that educated me and supported me thus far. I try my best to show my gratitude to those who took part in my journey but, undoubtedly, I would not be able to remember and acknowledge everyone.

Alle-Jan, it is my honor to be your student and I will always be proud of it. Indeed, you taught me the essentials of writing a scientific document, and I know it has not always been easy. You have read my drafts with patience and care so many times trying to find the pitfalls and summarizing the overlong paragraphs. Although, nothing can pass below your standard of perfection, you manage to be a friendly and easygoing leader for the group and a supporting supervisor for the students. Beyond your intelligence, I admire you for your genuine passion for science, work ethics and integrity. Thank you for all your help, from the first day that I arrived here in Delft up to now.

I am happy to be a part of Circuit and Systems (CAS) group of TU Delft for the past six years where it has always been an inspiring and friendly environment with excellence spirit of research and collaboration. I had the privilege to collaborate with knowledgeable individuals and distinguished scientists. I would like to thank *Geert* for his timely feedback, constructive criticism and suggestions which improved the quality of my work as a researcher. I am grateful for the guidance and help that I received during my PhD from *Gerard*. I owe *Richard* for his supervision during my post-doc appointment. I learned a lot from our discussions and his acute and to the point attitude, this thesis would have been delayed without his understanding and support.

I would like to extend my appreciation to *Marco Spirito* and the people in Electronics Research Laboratory (ELCA) of TU Delft for providing the equipment,

labratories and for sharing their valuable experience and knowledge for the experimental part of this thesis. I am grateful to *Zoubir Irahhauten* for coaching me during the measurement campaign in TU Delft laboratory and his contribution to my research. I want to mention the people at ASML including *Iwan Akkermans* and STW committee members. I am grateful to *Richard Heusdens*, *Jos Weber*, *Rene van Leuken*, *Bastiaan Kleijn*, *Mats Viberg*, *Andreas Molisch* for their advice and inputs.

During my appointment at TU delft, I shared office with great individuals: *Milad* thanks for all the technical and non-technical discussions and for your friendship, *Rocio* thanks for all the nice moments that we had and for your care and company, *Andrea* thanks for all the interesting discussions that we had on optimization theory and your high spirit and motivation, *Venkat* thanks for being such a nice officemate and for your empathy and of course for the apples! I also shared memorable office hours with *Yiyin*, *Yu*, *Yan* and *Shahzad*.

I am thankful to *Sundeep* for his help in finalizing this thesis and for sharing his experience in arranging the PhD defense and many other things. *Thomas* and *Matthew* thanks for proofreading the English for some parts of the thesis. I cannot forget my kind colleagues and fellows including: *Raj*, *Sumeet*, *Jorge*, *Jeroen*, *Jorn*, *Rob*, *Georg*, *Shahrzad*, *Andreas*, *Jie*, *Elvin*, *Nikolay*, *Amir*, *Christos*, *Hadi*, *Hamid*, *Sina*, *Tao*, *Esteban*, *Augusto*, *Zijian*, *Minaksie*, *Rosario*, *Simon*, *Wim*, *Alexander* and *Antoon*.

I enjoyed the company of priceless friends during my four-years in Delft, who brought unforgettable moments of joy and happiness to my life. I start with my favorite Sepidehs in the world. *Sepideh B.* thanks for simply being there for me every time that I needed you. *Sepideh G.* thanks for your friendship, wise advices and all the nice moments that we have shared together. *Sepideh T.* thanks for all the funs, chit chats and laughters. *Haleh* thanks for the sweet surprises and all the chearful memories. I will not forget *Amar*, *Bahar*, *Hamed*, *Milad*, *Mina*, *Pooyan*, *Sara*, *Sima*, *Samaneh*, *Sami*, *Sogol* and *Samad*.

Last but not the least, I want to show my humble gratitude to my beloved family who have always supported me in good and bad days. If I am here now, it is because of all the sacrifices you made for me. I feel very lucky and grateful to have you in my life. *Maman*, you stood by me no matter what, and you are always there whenever I need you. I cherish you for your dignity, your principles, your high spirit and never-ending source of motivation and energy. You are the strongest person that I know. *Baba*, you taught me how to ride a bike, to swim, and even how to repair a punctured car tire! You were my first Math and English teacher

and I never ever will forget the joy of reading poems from *Shahnameh* and *Hafez*, to the *Wonders of eighty days around the world*, as early as I was four. You are the most goodhearted person that I know. *Saman* you are my little sister and the first person that I felt I am responsible for, we shared so much laughter, gossip and secret missions and so many screams because of the cockroaches. Well you always are the bravest one! *Khale Bahar* and *Amoo Reza* you are the best parents in-law that I could have asked for. You treated me as your own child and loved and supported me all these years. I am wholeheartedly thankful for that. *Noushin* you are the coolest sister in-law ever.

Nader you are my rock, you are my muse, you are my soulmate. *Daria* my beautiful little angel, you gave me the most amazing life experience ever and I love you forever. You did your share for this thesis too, not to say that it would have been finished earlier if it was not for you! Indeed the cover page of this book is decorated with your voice calling me “mama”, the sweetest thing that I have ever heard.

Seyran
Utrecht, October 2016

Biography

Seyran Khademi was born and grown up in Kermanshah, located in the western part of Iran. She received the BSc. degree in Electrical Engineering, in 2005 from University of Tabriz with dissertation in the field of satellite communications. After her bachelor studies she hold a position as RF design engineer in a private company in Tehran performing design and test of RF boards for radio transceivers. She got her MSc degree in Communications Engineering from Chalmers University of Technology in Gothenburg, Sweden in 2010. She did her MSc thesis titled as " PAPR reduction in OFDM WiMAX systems" under the supervision of Prof. Mats Viberg, Prof. Thomas Eriksson and Dr. Thomas Svantesson (ArrayComm in San Jose, California). She started her PhD at Circuits and Systems (CAS) group at Delft University of Technology (TUD), The Netherlands in January 2011. She is appointed as a postdoctoral researcher in CAS group at TUD, from February 2015 and is working on audio and speech processing. Her research interests include applied mathematics, information theory and optimization techniques, in general. She has been working on signal processing for wireless communications, 60 GHz technology, audio and speech processing and intelligibility enhancement for speech communication systems.