Signal Processing in Massive MIMO Systems Realized with Low Complexity Hardware

MICAELA BORTAS
SARA GUNNARSSON
MASTER´S THESIS
DEPARTMENT OF ELECTRICAL AND INFORMATION TECHNOLOGY |
FACULTY OF ENGINEERING | LTH | LUND UNIVERSITY

Signal Processing in Massive MIMO Systems Realized with Low Complexity Hardware

Micaela Bortas and Sara Gunnarsson

Department of Electrical and Information Technology
Lund University

Department of Electrical Engineering
KU Leuven

Supervisors: Liesbet Van der Perre and Ove Edfors

Co-supervisors: Yanxiang Huang and Cheng-Ming Chen

February 7, 2017

Printed in Sweden
E-huset, Lund, 2017

Abstract

The global mobile data traffic, as well as the energy consumption, for networks is constantly increasing. For this reason, techniques which can provide higher spectral efficiency while being more energy efficient are needed. Massive Multiple-Input Multiple-Output (MIMO) is a technique which can bring these two together and will play an important role in future wireless networks.

The most power consuming part is the base stations and for the complexity in the digital signal processing, the per antenna functions are dominating. More specifically, in an Orthogonal Frequency-Division Multiplexing (OFDM) based system, this includes the filter and the inverse fast Fourier transform (IFFT).

In this thesis the possibilities of implementing the per antenna functions in low complexity hardware is investigated. Both low accuracy and scaling down the supply voltage to the integrated circuits, exploring the error resilience of the system, are considered. Due to this, a remarkable amount of energy can be saved.

Results in this thesis show that for a certain communication system implemented with low accuracy and allowing errors to occur, 97% of the signal processing power can be saved at a Signal-to-Noise Ratio (SNR) degradation of only 2 dB. Concluding this work, massive MIMO can provide high spectral efficiency and implemented with low accuracy hardware it can still be error resilient and lead to higher energy efficiency.

Acknowledgements

First of all we would like to thank Sofie Pollin and all the PhD students in her research group Networked Systems at KU Leuven for welcoming us and making us feel as a part of the group from day one. Your support and your help with general questions has been very valuable to us and made our stay at KU Leuven to the best possible.

We would also like to thank Cheng-Ming Chen and Yanxiang Huang, our daily supervisors, for all the help during this master's thesis. Thanks for always taking time to discuss our questions and problems and for your great support during the whole project.

At last we would also like to thank our supervisors Liesbet Van der Perre and Ove Edfors. Thanks for all the support and guidance which has made this master's thesis possible. Thanks for challenging us, giving us valuable feedback and for always believing in us.

Micaela Bortas and Sara Gunnarsson

Table of Contents

Preface

This master's thesis has been a co-operation between Lund University and KU Leuven. All the work has been conducted at KU Leuven in Belgium, where we spent about five months during the summer and autumn 2016. The idea of doing a master's thesis within massive MIMO appeared during the spring 2015 as a result of an increased interest in the subject due to frequent news reporting about the progress at Lund University. The opportunity to work with very progressive research has been a very valuable experience and we have learned a lot along the way.

In this thesis work, Sara initially worked on the outer simulation framework, the channel and the per-user processing. Micaela initially worked on the per-antenna processing in both the transmitter and receiver. Through this thesis work though, most parts have been shared in some way. If one part was created from scratch by one student, the other might later have worked on refinements, complementary implementations or speed-ups of that part.

One central reference for this thesis is the paper by Y. Huang et al., "Massive MIMO processing at the semiconductors edge: exploiting the system and circuit margins for power savings" [8], which this thesis partly is a further exploration of.

List of Figures

List of Tables

List of acronyms

- **5G** Fifth Generation
- **ACI** Adjacent-Channel Interference
- **ASK** Amplitude Shift Keying
- **AWGN** Additive White Gaussian Noise
- **BER** Bit Error Rate
- **CP** Cyclic Prefix
- **CSI** Channel State Information
- **DL** DownLink
- **FDD** Frequency Division Duplex
- **FFT** Fast Fourier Transform
- **FSK** Frequency Shift Keying
- **IC** Integrated Circuit
- **ICI** Inter-Carrier Interference
- **IFFT** Inverse Fast Fourier Transform
- **ISI** Inter-Symbol Interference
- **ITU** International Telecommunication Union
- **ITU-R** International Telecommunication Union Radiocommunication sector
- **LDPC** Low-Density Parity Check
- **LOS** Line Of Sight
- **LTE** Long-Term Evolution
- **MAMMOET** MAssive MiMO for Efficient Transmission
- **MIMO** Multiple-Input Multiple-Output
- **MRT** Maximum ratio transmission

- **MU-MIMO** Multi-User Multiple-Input Multiple-Output
- **MaMI** Massive MIMO
- **OFDM** Orthogonal Frequency Division Multiplexing
- **PA** Power Amplifier
- **PDP** Power Delay Profile
- **PSK** Phase Shift Keying
- **QAM** Quadrature Amplitude Modulation
- **QPSK** Quadrature Phase Shift Keying
- **RF** Radio frequency
- **RX, Rx** Receiver
- **SINR** Signal-to-Interference-Noise Ratio
- **SNR** Signal-to-Noise Ratio
- **TDD** Time Division Duplex
- **TX, Tx** Transmitter
- **UE** User Equipment
- **UL** UpLink
- **VOS** Voltage Over-Scaling
- **ZF** Zero-Forcing

Popular Science Summary

Massive MIMO (Multiple-Input Multiple-Output) is a promising technique for future 5G systems, but several challenges with this new technology still need to be addressed.

The basic idea behind massive MIMO is to use a large number of antennas relative to the number of users. Massive MIMO has been shown to achieve high capacity since huge improvements in both throughput and energy efficiency can be made. This is essential since the global mobile data traffic and network power consumption is increasing rapidly.

One of several challenges that needs to be addressed with this new technology is the risk of overwhelming energy consumption due to the large number of antennas. One way of addressing this problem is to use low complexity hardware, which has been shown sufficient to build a massive MIMO system with good performance.

Since the dominating part of the digital complexity lies in the antenna processing, this is where the focus for this master's thesis has been. Two methods of scaling down the complexity have been implemented: low accuracy hardware and scaling down power to the integrated circuits and allow for errors to occur.

By representing the information with less number of bits, the accuracy can be scaled down significantly and a lot of power can be saved. Certain parts of the antenna processing have also been able to be made less complex than others, which has been further investigated in this thesis. The conclusion made is that very low accuracy implementations are possible, while still maintaining good performance. Utilizing more antennas have also been shown to achieve better performance for low complexity signals than systems with fewer antennas.

Since massive MIMO has been shown to average out errors well over the antennas, a lot of power can also be saved by utilizing this property. More specifically, 43% power can be saved by introducing just a few errors on the antennas.

A conclusion from this master's thesis is that low complex solutions are possible in massive MIMO systems. With the combination of the two methods of bringing down the complexity, incredible 97% power can be saved. However, there is a trade-off between saving energy by decreasing the complexity and maintaining a good system performance. Other variables in the system have also been shown to be sensible to low accuracy signals and errors, which have affected the overall system performance, even though the positive effect of the many antennas can still be seen.

Chapter 1

Introduction

The use of mobile data in today's society grows more and more and is predicted to grow even faster in the coming years. Also, Internet will depend more on the mobile networks. This causes the need of more developed mobile access technologies.

Massive MIMO is a promising technology for the future of the next generation of wireless systems. Massive MIMO opens up a new dimension of wireless communication which will have a higher capacity due to better spectral efficiency. It will also be more energy efficient.

Progress has been made within the area during the last years and both the research community and the industry have agreed on that massive MIMO will be important for communication systems in the near future. Before massive MIMO can be standardized and deployed, many challenges need to be addressed.

1.1 Background and motivation

The basic idea behind massive MIMO is to use a large number of antennas relative to the number of active terminals. The technique relies on spatial multiplexing and uses linear processing in a time-division duplex mode [2]. It can offer high quality-of-service in different environments and also serve terminals with less beneficial placements, where large-scale fading or being too close to the cell border limits performance.

Due to the large number of antennas in massive MIMO systems there is a risk that the complexity and energy consumption at the base station will be overwhelming. But it is known that even low complexity hardware can be good enough to build a massive MIMO system with good performance. Dominating parts of the energy consumption are the analog and radio frequency (RF) components, which still are lower than in earlier network generations. The system can therefore operate at a much lower overall transmitted RF power.

It has also been shown that the dominating factor for the complexity in the digital signal processing is the per-antenna functions. Therefore, it is relevant to investigate new solutions for these per-antenna functions, both analog and digital, which can further reduce the overall complexity. The goal is to make the system more energy efficient, without significantly reducing performance. Example of possible solutions are digital transmitters [3] and approximate computing solutions [4].

When using more recent Integrated Circuit (IC) technology (65 nm and smaller), there are some challenges. The challenges are the increased circuit variability caused by process, voltage and temperature variability [8], which is the cause of the potentially erroneous behavior created by the low complexity hardware.

1.2 Project aims and main challenges

The aims of this master's thesis work are to first investigate the potential performance loss due to low accuracy in the signal processing in massive MIMO systems. Thereafter, the aim is to investigate how error resilient this low accuracy system is by simulating hardware errors and evaluate the performance losses which occurs. Finally, the whole system needs to be evaluated in order to find a feasible design. The duality of this master's thesis, including both exploration of the massive MIMO system and investigations of the circuits, is shown in Figure 1.1.

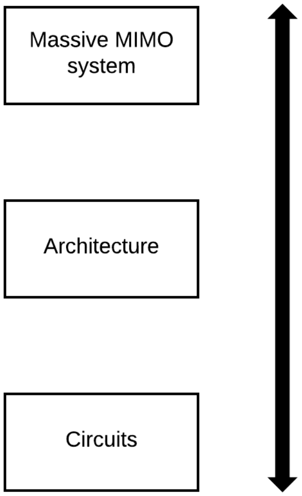

Figure 1.1: Scheme with the duality of combining a low accuracy massive MIMO system with low complex circuits in order to get a more energy efficient overall system.

A challenge in massive MIMO is to find a solution with as low energy consumption as possible at the base station. It is possible to find solutions with low energy consumption but these solutions can result in additional signal distortions and potentially erroneous behavior in the circuits, which could lead to performance loss.

1.3 Approach and methodology

The approach and methodology used in this master's thesis can be seen as a flow diagram, visualized in Figure 1.2, and will be further described in this section.

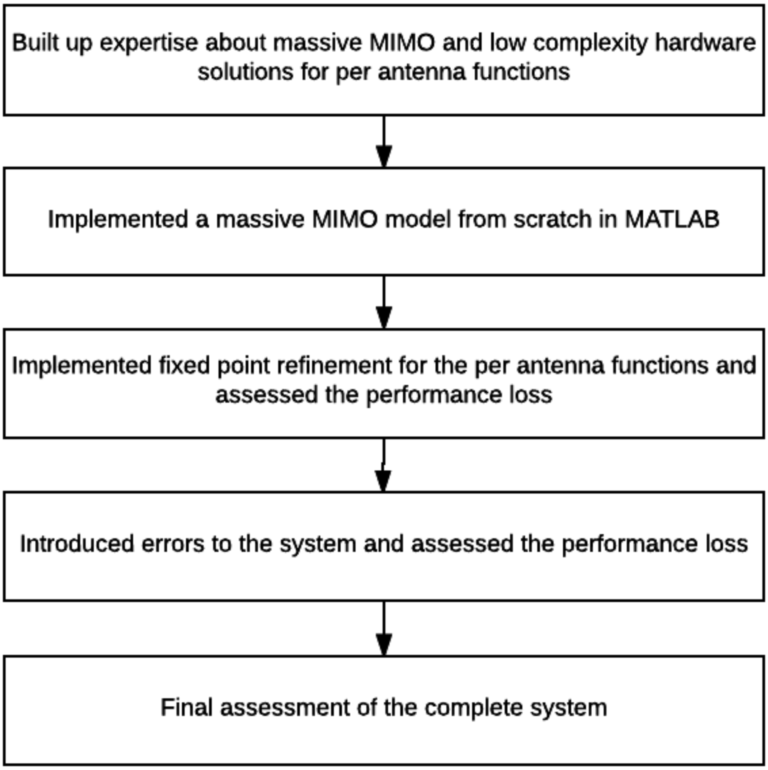

Figure 1.2: Flow diagram of the methodology.

This thesis project first consisted of a phase where expertise about massive MIMO and low-complexity hardware solutions, for per antenna functionality, was built up. In order to do this, several scientific papers were studied. This included both more general papers about massive MIMO [2], [5], [9], [15], [16] but also papers investigating low complex solutions [3], [4]. In this way an understanding of massive MIMO and knowledge about low complex solutions was achieved.

To be able to investigate the aims of this master's thesis, a massive MIMO model was needed. The model was developed from scratch and necessary functions were implemented in MATLAB, partly using built-in MATLAB functions. This resulted in a simulation environment used to perform the investigations needed to reach the goals of this thesis. During this process a good understanding of the simulation framework, system model and communications theory behind it was built up.

With an implemented massive MIMO model ready, the next step was to develop it so that it also could model a low accuracy massive MIMO system. To do this, fixed-point versions of the per-antenna functions were implemented. This made it possible to assess the performance loss caused by lowering the number of bits used in the signal processing.

After implementing the fixed-point per-antenna functions, the next goal was to assess the performance losses caused by non-ideal hardware in combination with the performance losses caused by the low-accuracy signal processing in the system. To be able to do this, errors were introduced to the system in order to simulate the non-ideal hardware.

The last part was the final analysis and assessment of the complete system, including both the massive MIMO system and the effects caused by the non-ideal ICs.

1.4 Previous work

There is a lot of research going on within the area of massive MIMO right now and to mention some of the work which has been particular important for this master's thesis project is the work done within the MAMMOET FP7 EU project [1]. Also, research from the department of electrical engineering (ESAT) at KU Leuven, research from the Electrical and Information Technology (EIT) department at Lund University and research being conducted at imec have been important for this work.

1.5 Advancements and outcome

The result of this project has lead to advancements within future massive MIMO systems concerning improved energy efficiency, cost reduction and increased understanding of how massive MIMO can allow the use of less complex signal processing.

The outcome has been a master's thesis report and a presentation. Another significant outcome is a MATLAB simulation framework for a massive MIMO system implemented in downlink, which can be used by others for further exploration of massive MIMO systems. The project has also resulted in input to the MAMMOET FP7 EU [1] and other massive MIMO projects, where investigations of the topic will continue.

1.6 Resources

The resources used during this master's thesis project are the working space in an office with a desk each provided by KU Leuven in the building for the department of electrical engineering (ESAT). Two computers were also provided, one computer was using Windows and one Linux. During this project all work has been done in MATLAB with an academic license provided by KU Leuven. Version control was done using an SVN server provided by the Electrical and Information Technology (EIT) department at Lund University.

1.7 Delimitations

The decided delimitations of this work are described in this section. The first thing is that the system is only implemented in downlink, which is the most essential part when considering energy efficiency in the per antenna functions. Therefore the system does not use channel estimation in the uplink, but rather relies on perfect channel state information (CSI) in the precoder. Perfect synchronization is also assumed in the system, instead of relying on pilots. The system also supports MIMO simulations but only for symmetric constellations, where the number of antennas at the base station side is the same as on the receiver side.

The signal processing in the simulation framework is limited to one type of channel coding with one block size and four code rates, four types of symbol mapping, one type of precoder, IFFT and FFT, up-sampling and down-sampling, one changeable type of filter and one type of equalizer. There are six different options for the channel, including a channel with only AWGN and five channels which also experience Rayleigh fading.

The focus has mainly been on the filter and the simulations have been using 64-QAM and one specific Rayleigh fading channel. This has been the case in order to be able to compare the results without too many parameters changing. The developed simulation framework supports more variations in the system and can be used for investigations beyond those performed in this work.

1.8 Thesis disposition

After this introduction, the next chapter contains some fundamentals about communications theory, briefly describing everything from the telecommunication aspects to the implementation aspects of the hardware, needed for the understanding of this thesis. Following the theory there is a chapter about the developed system model including the simulation framework surrounding it. The chapter after that includes the work and implementation of the fixed point refinement and modeling of the low complexity hardware. The thesis ends with chapters for the results, conclusions and future work.

Chapter 2

Fundamentals

In this chapter the fundamentals of wireless communication and relating implementation theory is described. General concepts which needs to be taken into consideration when designing a wireless communication system as well as methods used to cope with, or exploit, these concepts are further explained. Also, implementation complexity aspects and theory about the hardware are described.

The first part of the chapter will describe the telecommunication aspect and the most general of these concepts and methods. The second part will be about the design of low-complexity transceivers used in wireless communication systems.

2.1 Wireless communication

As society and new technologies are developing, mobile data traffic increases. According to Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020 [6] mobile data traffic will increase rapidly in the coming years as visualized in Figure 2.1.

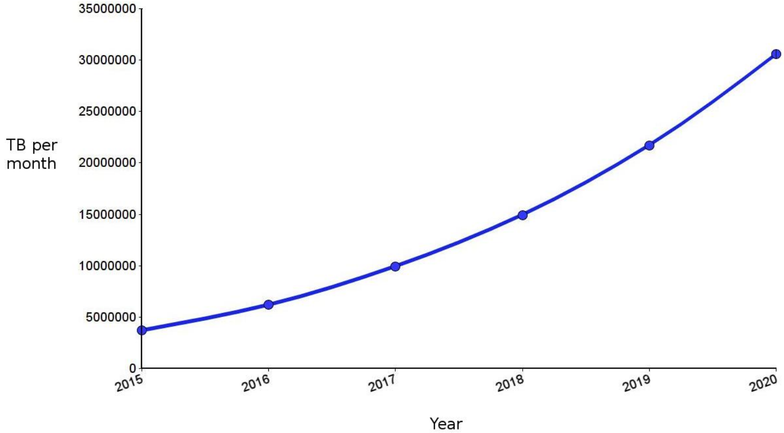

Figure 2.1: Global Mobile Data Traffic, 2015-2020 [6] with TeraByte per month on the y-axis and year on the x-axis.

With increased mobile data traffic many challenges appears as how to use the limited spectrum as efficient as possible, spectral efficiency, and how to make the wireless communication systems more energy efficient, both analog and digital.

2.1.1 Spectral efficiency

Spectral efficiency is the data rate which can be achieved over a specific bandwidth. Spectral efficiency, also called spectrum efficiency, is measured in bits/s/Hz. When developing new generations of wireless communications systems, the spectral efficiency has increased as displayed in Figure 2.2. The current world record on spectral efficiency in massive MIMO systems is 145.6 bits/s/Hz but is expected to increase a lot more before 5G is standardized and deployed [14].

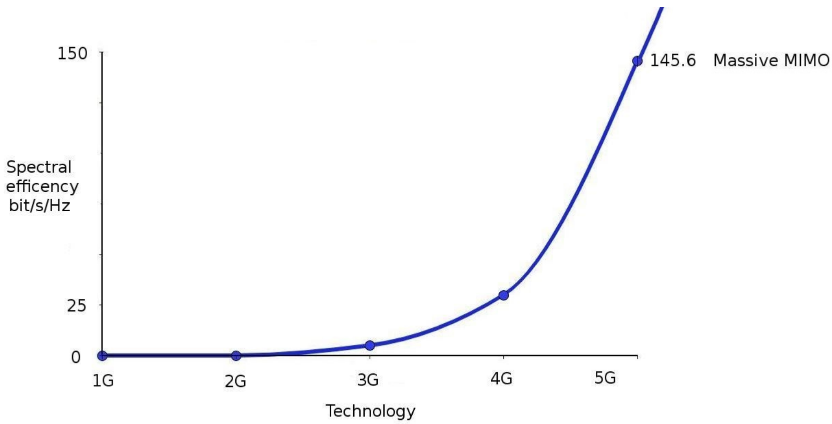

Figure 2.2: Increase of spectral efficiency in the different generations of wireless systems including the last set world record [14].

2.2 Energy consumption in wireless communications systems

Since energy consumption for networks is increasing faster than the total worldwide electricity use, as shown in Figure 2.3, it is essential that new techniques are developed to bring the network power consumption down.

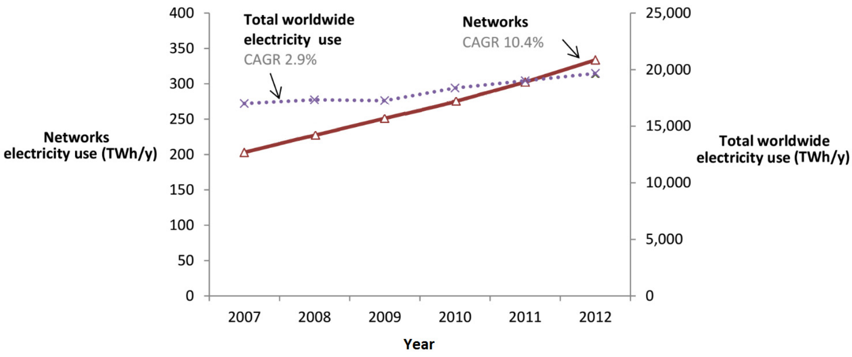

Figure 2.3: Increase of electricity use in networks compared to the total increased electricity worldwide [18].

The base station is responsible for about 80% of the power consumption in today's mobile network systems [25]. More specifically, most of the digital signal processing complexity lies in the antenna functions, as can be seen in Figure 2.4. In massive MIMO systems, where the number of antennas is large, the antenna functions are important to focus on when considering techniques which can bring down the energy consumption.

Even though the analog signal processing stand for most of the processing complexity, roughly 70%, the focus for this thesis is in the digital signal processing.

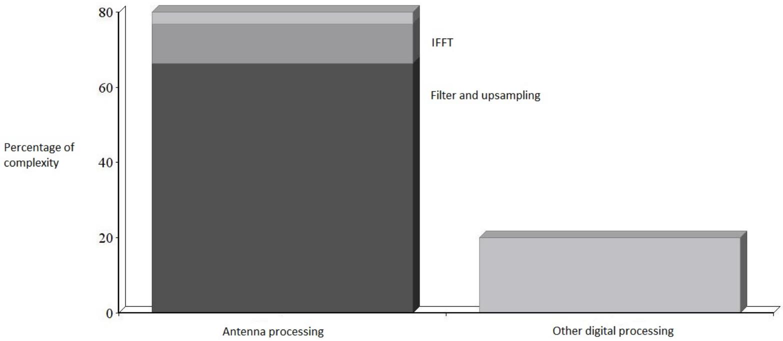

Figure 2.4: The base station complexity lies in the antenna processing [17].

Also seen in Figure 2.4, the most complex parts of the antenna processing lies in the filter and up-sampling, but also in the IFFT [17]. These two parts of the antenna processing are therefore essential to look further into.

2.3 Channel models

To be able to design and build a wireless communication system it is essential to have some understanding of different kinds of channel models and their properties. Channel models are used when modeling networks in order to see how the system design works in different situations and under certain circumstances. One example is fading which describes the varying attenuation which a signal experience in different propagation environments. This section describes some fundamental and important properties and gives a brief overview of the behavior of different channels.

2.3.1 Narrow- and wideband channels

The band is the range of frequencies which is used by the system, the so called bandwidth. Depending on how narrow or broad this range is, the channel is called narrow- or wideband.

Narrowband

When it is a narrowband the system bandwidth is smaller than the channel coherence bandwidth, i.e. the range of frequencies where two different frequencies can be considered to experience the same fading. This is also called flat fading and means that the channel gain for all practical purposes is constant for all frequencies in the interval.

Wideband

In wideband the range of frequencies used by the system is bigger than the coherence bandwidth of the channel. Therefore, a wideband system will experience both frequency selective fading and delay dispersion. Frequency selective fading means that the different frequency components in the signal will experience different fading. Delay dispersion, also called time dispersion, means that the different signal paths will arrive at the receiver at different times and will therefore result in a channel impulse response with a number of noticeable delays caused by the different multi-path components of the channel.

2.3.2 Additive White Gaussian Noise

Additive White Gaussian Noise (AWGN) is a statistical model which gives the distribution of noise, which naturally occurs in a communication system. There are several sources to AWGN, e.g. thermal noise in the receiver circuits and external noise sources picked up by the antenna. It is additive because it is added to the received signal. It is white because the power of the noise is independent of the frequency and therefore the same for the whole spectrum. It is Gaussian because its distribution is Gaussian.

The probability density function for a Gaussian random variable is:

$$pdf(r) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(r-\mu)^2}{2\sigma^2}}$$

where σ^2 is the variance, r the amplitude and μ the mean.

2.3.3 Fading

Fading can be divided into different categories and there are also several probability density functions which are used to describe different kinds of fading. The concepts and distribution relevant for this report are described below.

Large- and small-scale fading

Large-scale fading, also called shadowing, is when a large object is between the transmitter and receiver and therefore affects the received signal over larger-scale distances. The term large is in comparison to a wavelength.

Small-scale fading occur because the signal is being reflected due to obstacles between the transmitter and receiver which causes multi-path propagation. This means that signals from the different multi-paths will arrive with different delays at the receiver and can add up either constructively or destructively, i.e. in the first case the self interference will result in a gain of the signal and in the latter case, the signal eliminates itself. The term small is in comparison to a wavelength.

Rayleigh fading

Rayleigh fading is a statistical model for a small-scale fading channel when there is no line of sight (LOS) component, which is typical in urban areas. A Rayleigh distribution describes the amplitude variations of two uncorrelated orthogonal Gaussian random variables.

The probability density function for the Rayleigh fading distribution is:

$$pdf(r) = \frac{r}{\sigma^2} \mathrm{e}^{-\frac{r^2}{2\sigma^2}}$$

where r is the amplitude and σ^2 the variance.

2.3.4 Power Delay Profile

The power delay profile (PDP) describes how the input energy to the channel is spread over time on its output. This is an effect of the channel impulse response of the channel. In this work, power delay profiles defined by the International Telecommunication Union is primarily used.

International Telecommunication Union Radiocommunication sector standard

The International Telecommunication Union Radiocommunication (ITU-R) sector has certain recommendations for evaluation of radio transmission technologies. For example, they have four different power delay profiles which are displayed in Table 2.1 and Table 2.2, the first one concerning power delay profiles for pedestrian scenarios and the latter for vehicular scenarios [10]. The tables give an overview of when the energy arrives, measured in ns, and with which magnitude, measured in dB.

Tab	Pedestrian A		Pedestrian B	
	Relative delay [ns]	Average power [dB]	Relative delay [ns]	Average power [dB]
1	0	0.0	0	0.0
2	110	-9.7	200	-0.9
3	190	-19.2	800	-4.9
4	410	-22.8	1200	-8.0
5			2300	-7.8
6			3700	-23.9

Table 2.1: Power delay profiles: ITU-R model, Pedestrian [10].

Tab	Vehicular A		Vehicular B	
	Relative delay [ns]	Average power [dB]	Relative delay [ns]	Average power [dB]
1	0	0.0	0	-2.5
2	310	-1.0	300	0.0
3	710	-9.0	8900	-12.8
4	1090	-10.0	12 900	-10.0
5	1730	-15.0	17 100	-25.2
6	2510	-20.0	20 000	-16.0

Table 2.2: Power delay profiles: ITU-R model, Vehicular [10].

2.4 Multi-carrier systems

The idea of multi-carrier modulation (MCM) is to divide data into several low-rate bit streams and sending them over separate carrier signals. This is a way of making better use of the existing RF-band, where orthogonal frequency division multiplexing is the most natural MCM-technique for massive MIMO systems.

2.4.1 Orthogonal frequency-division multiplexing

Orthogonal frequency division multiplexing (OFDM) is a digital multi-carrier modulation method where multiple orthogonal subcarriers are used to carry data at a slow symbol rate. Each transmitted subcarrier results in a sinc-function in frequency domain, where the line up of each individual peak takes place at the zero-crossing of the other subcarriers thanks to the orthogonality between the symbols. An illustration of four OFDM subcarriers can be seen in Figure 2.5, where the orthogonality is clearly shown. Rather than transmitting a high-rate stream of data on a single carrier, the slowly modulated narrowband subcarriers are each exposed to flat fading rather than frequency selective fading. This can easily be corrected with a very simple per-subcarrier equalization in the receiver. Despite the slow symbol rate in OFDM transmission, data rates in the range of conventional single-carrier modulation schemes can still be achieved thanks to the large number and compact way of organizing the subcarriers.

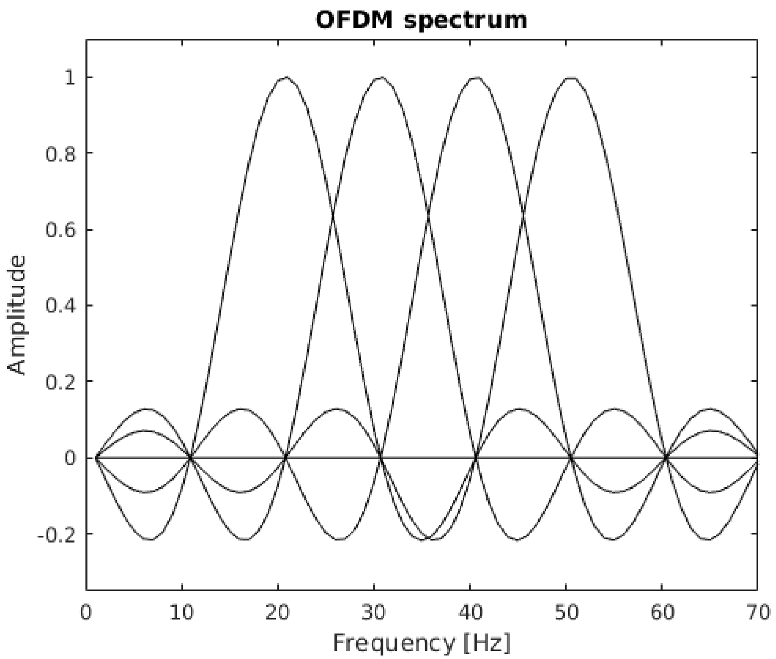

Figure 2.5: Four orthogonally spaced subcarriers.

Inter-symbol interference

In OFDM systems, Inter-symbol interference (ISI) occurs in the time-domain when a delayed version of an OFDM symbol overlaps with an adjacent symbol. One way of minimizing ISI is to separate the OFDM symbols in time by a distance longer than the length of the channel impulse response. If this space between OFDM symbols is filled with a copy of the last samples of the symbol, all subcarriers remain orthogonal also after passing through a frequency selective channel. This part is called the cyclic prefix (CP) and can be seen in Figure 2.6.

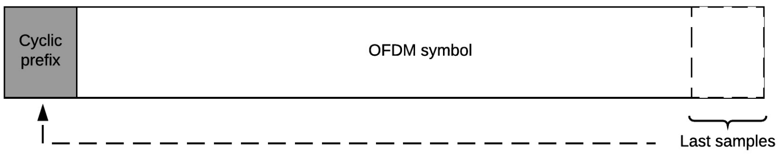

Figure 2.6: Copies of the last samples are added at the beginning of the OFDM symbol to create the cyclic prefix.

Inter-carrier interference

Inter-carrier interference (ICI) occurs when the channel is not constant during the transmission of one OFDM symbol. For example in rapidly varying channels [7], where the mobility and the corresponding Doppler spread in the channel is the source of this interference. The rapid movements in the channel will cause frequency offsets, i.e loss of orthogonality between the subcarriers, and the subcarriers will interfere with each other.

Adjacent-channel interference

To avoid energy from adjacent channels to leak into the pass-band of other channels, i.e. Adjacent-channel interference (ACI), guard bands, or so called guard subcarriers, are used. The guard bands can be seen on the outskirts of the symbol in Figure 2.7, where 1200 subcarriers in the middle are used to carry data and the remaining 848 subcarriers are null carriers used as guard bands on each side of the data, according to LTE standards.

ACI can also be caused by inadequate filtering, where the interfering signal is not filter out properly.

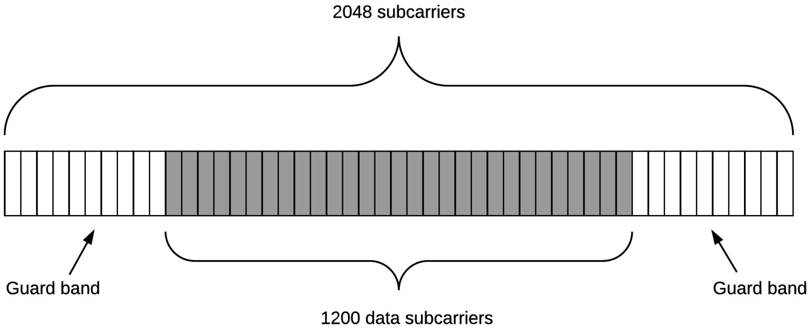

Figure 2.7: Guard bands on each side of the data are utilized to avoid ACI.

2.5 Modulation

Modulation is the process when information is added to a carrier signal in order to transfer information. Demodulation is the opposite, i.e. the process when the receiver filters out the information sent by the transmitter.

A radio signal can be expressed as $s(t) = A(t)\cos(2\pi f_c t + \phi(t))$, where A is the amplitude, f_c the carrier frequency and ϕ the phase. By changing these parameters it is possible modulate the signal to carry information. The basic digital modulation techniques, based on the variation of these parameters, are Amplitude Shift Keying (ASK), Frequency Shift Keying (FSK) and Phase Shift Keying (PSK).

2.5.1 Quadrature Amplitude Modulation

Quadrature Amplitude Modulation (QAM) is a modulation scheme which can be described as a combination of ASK and PSK. It consists of two carrier signals with the same frequency, but 90° out of phase, where the amplitude is modulated by ASK and the phase is modulated by PSK. When these signals are added and combined they give the final signal with both variation in amplitude and phase.

Using QAM in a communication system will lead to the possibility to transmit higher data rates since it consists of more bits per symbol. With higher order modulation formats it is possible to transmit even more bits per symbol, such us 64-QAM and 256-QAM which corresponds to six respectively eight bits per symbol. The drawback with using higher order modulation formats is that the signal gets more sensitive to noise and errors because the constellation points are closer together.

2.5.2 Inverse-/Fast Fourier Transform

In OFDM, an Inverse Fast Fourier Transform (IFFT) is performed to convert the signal to the time domain and a Fast Fourier transform (FFT) to convert the signal to the frequency domain. After the IFFT is performed in the transmitter, the signal is a discrete time domain signal.

Butterfly

A butterfly is the smallest stage in which the FFT algorithm decomposes the discrete Fourier transform (DFT) into. The results of the smaller stages is then combined into a larger DFT. The name butterfly comes from the shape of the data-flow diagram in Figure 2.8 which, in its simplest form, resembles the wings of a butterfly.

Each butterfly takes two complex numbers p and q and computes two new numbers from them, as seen in the right side of the figure.

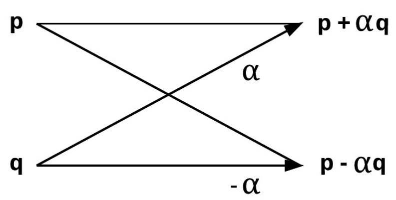

Figure 2.8: Butterfly operation of size 2.

The variable α is called the twiddle factor, or the root of unity. The twiddle factor is a rotation factor:

$$\alpha = \mathrm{e}^{-i\frac{2\pi}{N}}$$

and depends on the FFT/IFFT length N and takes advantage of redundancies and symmetries on the unit circle. These factors are the foundation that makes the FFT possible.

The total number of butterfly stages in a larger FFT is dependent on the number of samples according to: $\log_2(N)$, where each stage consists of $N/2$ butterfly operations. The IFFT is calculated in a similar manner, but using the inverse twiddle factor.

2.5.3 Up/down-sampling and filtering

Up-sampling and filtering is done in the transmitter to create a smoothly varying sequence of samples and also to get rid of discontinuities between the symbols.

The up-sampling creates a sequence of the data were each sample is separated by zeros. The number of zeros between the original samples depends on the up-sampling factor. The filter then replaces the zeros and smooths out the discontinuities between the symbols. The same matched filter is then applied in the receiver (to maximize the signal to noise ratio (SNR)) and finally, down-sampling is performed to recover the transmitted samples back to its digital form.

Square root-raised-cosine filter

For the pulse shaping, described above, a square-root-raised-cosine filter is used. The name derives from the resemblance to a cosine function in frequency domain, and the impulse response exhibits odd symmetry about $\frac{1}{2T}$, as seen in Figure 2.9, where T is the duration time of the filtered signal.

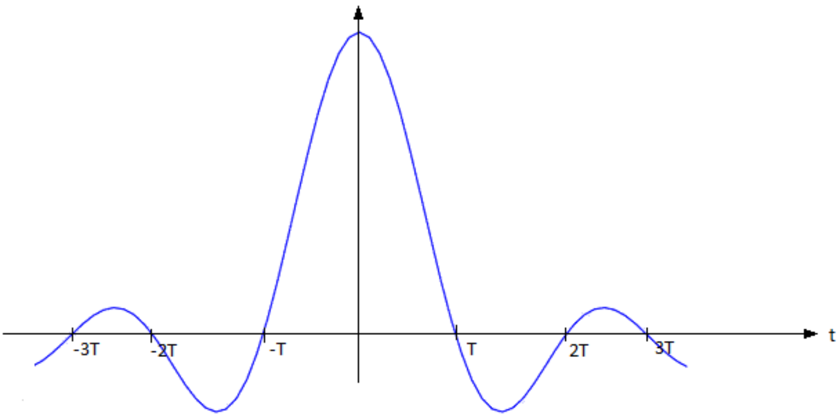

Figure 2.9: The impulse response of the square-root-raised-cosine filter.

2.6 Channel coding

Channel coding is an important factor to be able to achieve efficient and reliable data transmission over an unreliable communication channel, where channel coding is used as an error control technique. There are two main error control techniques, Automatic Repeat reQuest (ARQ) and Forward Error Correction (FEC).

Automatic Repeat reQuest

The idea of ARQ is to re-transmit data if it is too corrupted by errors. This is done utilizing some redundant data as error-detection code. If this error-detection code is not consistent with the rest of the message, the receiver will ask the transmitter to re-transmit the data.

Forward Error Correction

In FEC, rather than re-transmitting data, enough redundant data is transmitted for the receiver to be able to recover from errors all by itself.

2.6.1 Linear block codes

Block codes is a type of FEC code that divides data into blocks and adds check bits to each block. Most block codes used today are linear block code, i.e. the modulo-2 sum of any two code words is also a code word.

A parity check matrix $\boldsymbol{H}$ is used to detect errors in the received code by using the fact that:

$$\boldsymbol{c}\boldsymbol{H}^T = 0$$

where $\boldsymbol{c}$ is the coded data vector and $\boldsymbol{H}$ is the parity check matrix, produced in such a way that the condition is fulfilled if no errors are present. For error patterns with few enough bit errors, correct messages can be decoded.

Low-Density Parity Check codes

The Low-Density Parity Check (LDPC) code is a type of linear block code where the parity check matrix contains mostly zeros and relatively few ones, hence the name Low-Density Parity Check. Even though several disadvantages come with the use of this kind of simple parity check matrix, the low complexity of the decoding scheme more than compensate for this in sense of efficiency. Because of the low complexity, LDPC coding makes a good choice for massive MIMO systems.

2.7 Channel equalization

Channel equalization is done in the receiver to reduce channel effects in order to achieve better demodulation. The equalizer compensate for the gain and phase caused by the channel at the sub channel's frequencies simply by dividing the received signal by the channel coefficient. If

$$y = hx + n$$

where y is the received signal, x the transmitted signal, h the channel and n the channel noise, the equalizer recovers the signal by performing

$$\frac{y}{h} = x + \frac{n}{h}.$$

(perfect knowledge of h assumed here).

The equalization is done at every subcarrier in frequency domain, which are recovered in the receiver. This is referred to as a one-tap equalizer.

2.8 Synchronization and channel estimation

In order to achieve efficient transmission, the transmitter and receiver have to be synchronized to each other. Synchronization may also be referred to as "symbol timing" since it determines the start, or trigger point, of a symbol.

Subcarriers with known data, called pilots, are used for synchronization but also have the important task of channel estimation and detection of frequency offsets.

In case of multiple antennas, where different transmitters need to be synchronized to different receivers, pilot subcarriers are used for synchronization since multiple streams are transmitted at the same time in the same channel bandwidth.

2.9 Bit-error rate

Bit-error rate (BER) is a central concept within wireless communication. BER indicates the fraction of the received bits in a digital data transmission that are in error, i.e the fraction of bits that have been altered due to channel effects, synchronization errors or errors in the hardware.

$$\mathrm{BER} = \frac{\text{Number of received bits with errors}}{\text{Total number of transferred bits}}.$$

2.10 Signal-to-noise ratio

The signal-to-noise ratio (SNR) is a measurement of the signal power relative to the noise power, often expressed in decibels (dB) as:

$$\mathrm{SNR} = \log_{10}(\frac{P_{\text{signal}}}{P_{\text{noise}}}).$$

The SNR is measured per-receive-antenna, which means that the total transmit power in the massive MIMO downlink, for a fixed SNR, is proportional to the number of users.

The BER together with the SNR gives a good overview of the performance of a wireless system. A high SNR often yields a better BER.

2.11 Duplex schemes

An important part of any radio communication system is the way in which radio communications are maintained in both directions. A duplex scheme is necessary for organizing the transmitter and receiver to either talk or listen during communication. The duplex scheme dominated in current cellular systems is Frequency

Division Duplex (FDD), but massive MIMO utilize the benefits of the Time Division Duplex (TDD) scheme to obtain simultaneous communication. More on the challenges of using FDD for massive MIMO can be found in [9].

2.11.1 Time-Division Duplex

Time-division duplex is a multiplexing method where the data streams in different directions share the same frequency resource but are separated in time. If we have channel reciprocity, the benefit of TDD operation is that pilots are only needed in the uplink, which saves a lot of resources needed for channel estimation.

2.12 Multiple-antenna systems

Using multiple antennas in the transmitter and/or receiver can improve the system performance in several important ways. The following techniques are advantageously utilized in different scenarios were multiple antennas are used.

2.12.1 Beamforming

By adjusting the phase of the signal, constructive superposition can be achieved, so called beamforming. This correspond to steering the antenna pattern towards the desired direction, or as often the case in Massive MIMO - in many directions. Beamforming provides improved SNR (and SINR) since the received signal power is increased and interference can be canceled. Two ways of performing beamforming is described below.

Spatial multiplexing

By simultaneously transmitting data to all users on all base station antennas, with amplitude and phase adjusted to the channel between each base station antenna and the corresponding user, multiple channels can be created to users in the same time/frequency resource. This creates spatial multiplexing, where multiple users share the available resource and bit rates are considerably increased.

Spatial diversity

Spatial diversity, in the downlink, also exploit multiple antennas on the transmitter side. The quality and reliability of the signal is increased by transmitting versions of the same data sequence from all the antennas. Appropriate combining in the transmitter is then performed so that all multi-path components of the signal arrive with the same phase at the receiver antenna and are added constructively when arriving at the antenna. Spatial diversity mitigated fading and thereby decreases the BER.

2.12.2 Multiple Input Multiple Output

Multiple Input Multiple Output (MIMO) is a multiple antenna technique where several antennas are used both on the transmitter and on the receiver side. In Figure 2.10 two cases of MIMO can be seen, single user MIMO and multi user MIMO. In both cases, more than one antenna can be found in both the transmitter and receiver side, with an important difference in performance for the two cases.

Single-user MIMO

In single-user MIMO, multiple streams of data can be sent to only one device at a time, employing all available antenna elements.

Multi-user MIMO

In the multi-user MIMO (MU-MIMO) case, several users are served in parallel on the same frequency resource. This has the clear advantage that the time each device has to wait for a signal decreases, and a drastic overall speed-up of the system can be achieved. In other words, the multiplexing gain (or the extra degrees of freedom) can be shared by all users, in contrary to the single user MIMO case.

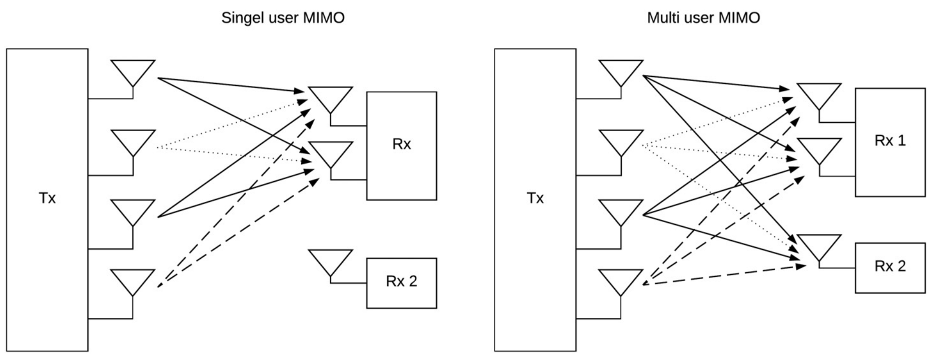

Figure 2.10: In the single user MIMO case, only one user can be served at a time. In the multi user MIMO case, multiple users can be served in parallel.

2.13 Massive MIMO

Massive MIMO (MaMi) is MU-MIMO technique where the number of base station-antennas are much greater than the number of receiving terminals, which are typically assumed to be single-antenna devices. The extra antennas help by providing antenna gain, diversity and eliminating inter-user interference, bringing huge improvements in throughput and radiated energy efficiency. An increase of ten times or more in system capacity can be achieved with MaMi [8].

Unlike many other techniques, MaMi is well employed in rich scattering environments where the number of propagation paths is large. This makes MaMi well suited for urban environments, with many obstacles.

Another great benefit of MaMi systems is the opportunity for extensive use of inexpensive low-power components. MaMi-system have a great ability to average out errors over all the antennas which makes low complexity hardware a feasible solution for energy and cost reductions in MaMi-systems. This concept is the main core of this work.

Furthermore, thanks to the excessive number of base station antennas, simple linear signal processing can be used in the base station in place of more complex non-linear processing without significant loss of performance. This lowers base station complexity and improves energy reductions.

2.13.1 Frame structure

The TDD frame structure, used here, and in which massive MIMO data is divided, is shown in Figure 2.11. Similar to the frame structure in [5], one frame of 10 milliseconds (ms) consists of 10 subframes, each containing two time slots of 0.5 ms. Every time slot consist of seven OFDM symbols, where the first three are used for uplink-transmission and the last three for downlink-transmission. The middle symbol is left empty to act as a guard interval to separate uplink-data from downlink-data, allowing transceivers to switch between transmit and receive.

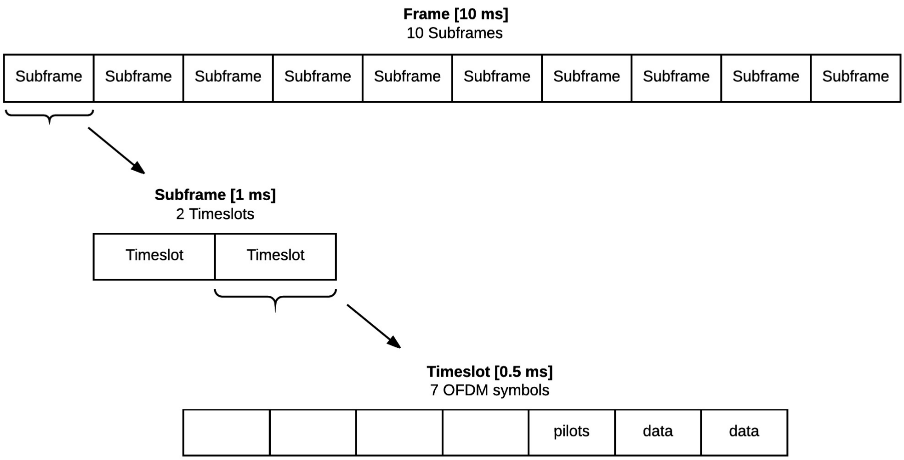

Figure 2.11: Massive MIMO frame structure in which data is divided into.

2.13.2 Massive MIMO precoding

Massive MIMO precoding can be described as a generalized form of beamforming, as described in section 2.12.1, that aims to minimize the error in the receiver output. A signal processing technique is performed in the base station, exploiting transmit diversity by weighting information streams based on the CSI estimated in the uplink.

In massive MIMO systems, linear precoding techniques have turned out to be near optimal even if less complex than non-linear approaches.

First of all, a general expression for the down-link transmission on one of the subcarriers can be described as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{z} + \boldsymbol{n}$$

where $\boldsymbol{y}$ is a vector of the received signals at the K terminals, $\boldsymbol{H}$ the KxM channel matrix between the M base station antennas and the K single-antenna terminals, $\boldsymbol{z}$ the transmitted signals on the M base station antennas and $\boldsymbol{n}$ is the channel noise.

Now using a linear precoder $\boldsymbol{W}$ to calculate the transmit signals on the M base station antennas, from the vector of data points $\boldsymbol{x}$ intended for the K terminals, we have

$$\boldsymbol{z} = \boldsymbol{W}\boldsymbol{x}.$$

The most common linear precoding techniques used in massive MIMO are maximum ratio transmission (MRT) and zero-forcing (ZF). For a single-cell downlink massive MIMO system, ZF has been shown to achieve higher data rates then MRT [13].

Maximum ratio transmission

To be able to compare MRT and ZF precoding, the MRT precoder first needs to be defined further. With the notation introduced above, the MRT precoder simply becomes:

$$\boldsymbol{W}_{\mathrm{MRT}} = \boldsymbol{H}^H$$

where $\boldsymbol{H}$ is the estimated channel matrix.

Zero-forcing precoding

In ZF precoding, the multiple transmit antennas can cancel out the inter user interference. Hence, ZF is referred to as null-steering. The ZF-processing in the base station can be expressed as:

$$\boldsymbol{W}_{\mathrm{ZF}} = \boldsymbol{H}^H(\boldsymbol{H}\boldsymbol{H}^H)^{-1} = \boldsymbol{W}_{\mathrm{MRT}}(\boldsymbol{H}\boldsymbol{H}^H)^{-1}$$

where $\boldsymbol{H}$ is the estimated channel matrix, $\boldsymbol{W}$ a MxK matrix, M is referring to the number of transmit antennas and K to the number of receiving single-antenna users.

2.14 Integrated Circuits

An integrated circuit (IC) is a chip that has millions or billions of components like resistors, capacitors, and transistors compactly fabricated and connected together on a semiconductor material. The two main advantages with IC's over discrete circuits are cost and performance:

Rather than constructing transistors individually, the entire chip is constructed as a single unit, using fine optical techniques to create patterns in the fabric. This reduces the cost remarkably in comparison with discrete circuits.

Because of the size of the IC's and the close connections between the components, the power consumption can be kept low. This results in a high performance. Other benefits of the IC's are: less weight, easy replacement and high reliability.

The IC's have many application areas and are mainly categorized in analog and digital IC's. Analog IC's can handle any inputs and produce outputs of any level, while digital IC's operate with binary digital signals, i.e. 1:s and 0:s. A common use of analog IC's are in multipliers and different types of amplifiers. Digital IC's can be found in any kind of logical application such as in timers, memories chips and multiplexers.

2.14.1 Semiconductors

A semiconductor is a material with specific electrical properties. The most common type of semiconductors are created with silicon, which is likely to remain the basic material for semiconductors used in IC's [11].

2.15 Evolution of Integrated Circuits

When the science and exploration of IC's first started in the late 1950's the goal was to implement more complex functions in a smaller space and with less weight. Two of these early evolved techniques was the active integrated circuits with semiconductors and the more passive technique with thin-film resistors. These different techniques took inspiration from one and another in order to develop and establish on the market as reliable techniques for a lower cost but with better performance than previous electronics [11].

During the beginning of this evolution an observation by Gordon Moore was made. This observation stated that the number of components per integrated function would approximately double every two years. This meaning, that the cost would also decrease rapidly for the IC's [11].

2.15.1 Technology challenges

The dominant factor in the beginning of the evolution of integrated circuits was the scaling. Going from the size range of μm to today's technology in the range of 10 nm has made the ICs starting to reach the fundamental limits making it no longer possible to do further scaling [12].

2.15.2 Design challenges

When technology scaling is approaching its limit, future development will rely more and more on new possible design solutions in order to proceed with further advancements within IC technology. This makes the line between hardware and software harder to distinguish since software then will play a vital role when considering the design aspects of the IC [12].

In later technologies, an increased circuit variability caused by the process, voltage and temperature variations has emerged which threatens the reliability of the circuits. Because of this, circuits have been designed at the worst corner point to cope with this variability. This technique guarantees performance but leads to wasted power consumption [8][12].

The different techniques addressing the issue with power consumption are dynamic scaling techniques, which use variable supply voltages to manage variabilities in circuits, and error resilience, which scales the supply voltage aggressively and allow errors to occur in the IC's. To continue benefiting from technology scaling, techniques to design more errors resilient systems are becoming essential. This refers both to errors originating in design and manufacturing processes and errors which appear while using the hardware. In upcoming designs a more statistical, rather than deterministic, approach is beneficial [8][12]. The different methods for setting the supply voltage and how this affects the IC's, are visualized in Figure 2.12.

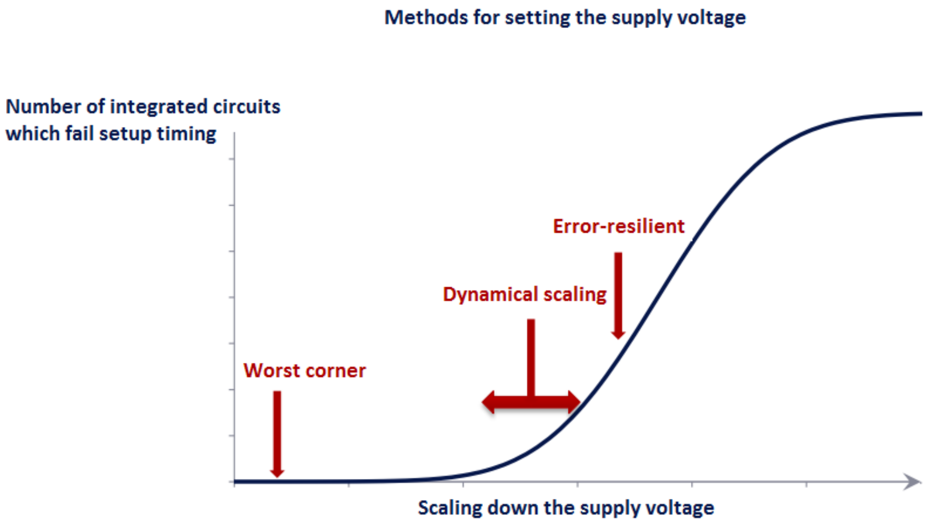

Figure 2.12: Methods for setting the supply voltage [8]

2.16 Low complexity hardware

As discussed above, two aspects of low complexity are considered: Low accuracy and error-resilience power supply scaling that allows for digital hardware errors. In this section, expressions for the complexity will be further investigated for the filter and IFFT.

By lowering the complexity, great power savings can be achieved while still maintaining good performance. Low complexity implementations is therefore a very desirable method for power savings.

2.16.1 Low accuracy hardware

In terms of accuracy, digital signal processing can be divided into two formats: floating point and fixed point. When a digit is represented with fixed point, a fixed number of bits are used to represent the integer part and a fixed number of bits are reserved to represent the fractional part. No matter how large the number is, it will always use these fixed numbers of bits for each part. In floating point, the decimal point can 'float', which means that a certain number of bits are used to represent the digit, but the number of bits reserved for the integer and fractional part respectively, are not fixed.

When information is made fixed point, the total word length is defined. Thus, only the total bit representation is considered.

The IFFT and filter are both processes that perform additions and multiplications. To evaluate the complexity of the IFFT and filter, the complexity of additions and multiplications must first be considered.

The complexity of an addition, referred to as C, taken from the complexity of a Ladner-Fischer adder [19], also verified in [20] and [21], can be calculated as:

$$C_{\text{adder}} = n \log_2(n)$$

where n is the maximum input size in bits.

The complexity of a multiplication, taken form the complexity of the Baugh-Wooley (BW) adder [22], verified in [21], can be calculated as:

$$C_{\text{multiplier}} = m \cdot n$$

where m and n are the number of bits that each of the inputs are represented with respectively.

Using this knowledge, the complexity of the IFFT and filter can now be further investigated.

Filter

For a filter with k taps (coefficients), $2k$ adders and multipliers are used. The factor 2 comes from the partitioning of the real and imaginary input.

The complexity of the filter is then calculated as:

$$C_{\text{filter}} = 2 \cdot k \cdot (m+n) \cdot \log_2(m+n) + 2 \cdot k \cdot m \cdot n$$

where n is the number of bits used for the input and m is the number of bits used for the filter coefficients.

IFFT

A 2048-IFFT, as used in this thesis, has $\log 2(2048)$-stages $\cdot \frac{1}{2}$ butterfly per bit. Each butterfly takes 6 adders and 4 multipliers. Therefore, each sample requires:

$$(11\frac{1}{2} \cdot 6 \cdot C_{\text{adder}} + 11 \cdot \frac{1}{2} \cdot 4 \cdot C_{\text{multiplier}}) \cdot 2 = 66 \cdot C_{\text{adder}} + 44 \cdot C_{\text{multiplier}}$$

A multiplication by 2 is done to take the difficult memory management for the IFFT into account.

Finally, the complexity for the IFFT can be calculated as:

$$C_{\text{IFFT}} = 66 \cdot (n) \cdot \log_2(n) + 44 \cdot n \cdot m = 66 \cdot n \cdot \log_2(n) + 44 \cdot n \cdot m$$

where n is the number of bits used for the input and m is the number of bits used to represent the IFFT twiddle factor.

2.16.2 Digital hardware errors

A certain amount of power can be saved by allowing a certain amount of digital hardware errors to occur in the IC's [23]. In [24], power savings in the range of 43%-60% have been achieved for different levels of power supply scaling, which generates the hardware errors. The different values for power savings are based on silicon measurement results. Since the measured results are different for different chips because of variations in the circuits, the range 43-60% is chosen to provide a good estimation on the range of benefits from measurements.

Voltage Over-Scaling

The method of scaling down the supply voltage for the IC's is referred to as Voltage Over-Scaling (VOS). This method is considered as an error-free power saving method as long as the setup timing constraints in the circuits are met, i.e. the constrains on the interval before clock transition, in which data must be stable. But meeting these constrains in terms of power supply scaling is not that straight forward due to variability in the circuits. Therefore, hardware errors must be accounted for. For logical components, not meeting the setup timing means that a signal is mis-captured, and for analog (memory) components, this results in incorrect write/read data/address or complete loss of data [8].

Antenna outage

The worst scenario in terms of hardware errors is complete failure of antennas, i.e. antenna outage. Antenna outage occurs when a circuit controlling signal is corrupted or when the transceiver systems is simply broken [8].

Chapter 3

Massive MIMO system model and simulation framework

In this chapter the massive MIMO system model and simulation framework is explained. To be able to generate results which are relevant for achieving the aims of this master's thesis, a number of changeable parameters were implemented. Some of the parameters have been changed when generating the results in this report, and some of them have been kept constant. More about the parameters and their correlation to the system model is explained in this chapter. Also, the values of the parameters used in the simulations which have generated the results in this report are presented as well as the other available options.

To simulate low complex hardware in the per-antenna functions, filters and IFFTs, and to investigate the impact on the system performance caused by this, fixed-point refinement and simulating hardware errors are necessary to include in the simulation model. The number of bits representing the signal and filter coefficients in the filter and the number of bits representing the signal and the twiddle factor in the IFFT therefore needs to be variable. To see the effects of hardware errors of different magnitudes, a parameter which can change this magnitude is necessary. The more specific implementation of this is explained in the next chapter and the parameters concerning this is described in the section below as they are a part of the defined global variables.

Since there are random parameters involved in the system, such as the generated data and the Rayleigh fading channel, a simulation framework which can run the system model several times is needed to get statistically significant results. This chapter will also further explain this simulation framework.

3.1 Global variables

The global variables are divided into three categories, one for the simulation parameters, one for the system parameters and one for the fixed-point refinement and error insertion. An overview of the parameters which have been implemented into the massive MIMO framework can be seen in Figure 3.1. The developed MATLAB script containing these variable can be found in Appendix A.

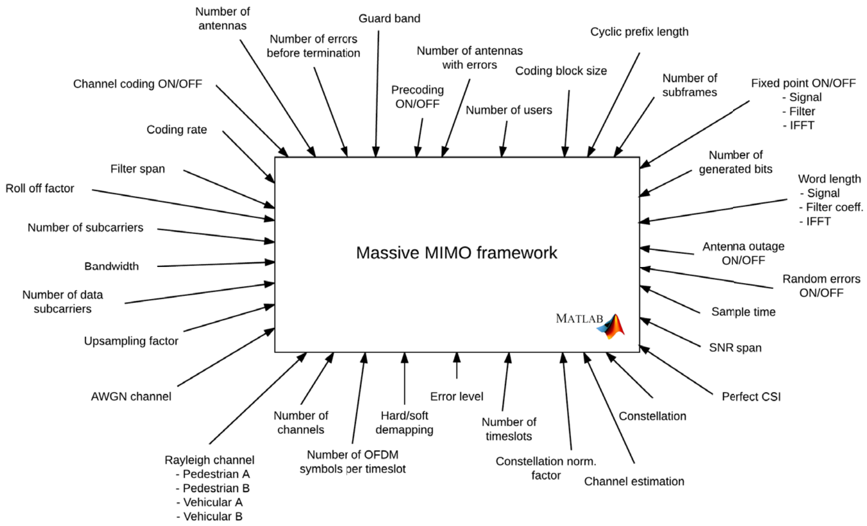

Figure 3.1: The massive MIMO framework with parameters.

3.1.1 Simulation

In the global variable which includes the simulation parameters, it is possible to change all parameters directly related to the simulation framework.

First, it is possible to set the number of channels generated during a simulation, which for this report has been 1000. It is possible to choose any number, where a higher number gives more statistically significant results but takes more simulation time. Secondly, the SNR span can be changed by choosing the start and end SNR, and also the step size in between. These values have been changed a lot and the values for each simulation can be seen in the figures in the result chapter. The needed SNR value for a simulation depends on several other parameters, like the chosen signal constellation and if channel coding is used or not. Thirdly, there is a parameter regulating the number of errors which needs to be found before terminating the SNR loop. In the simulations generated for this report this parameter has been set to 20, after verifying that this generated good enough results. Especially when running simulations which result in lower BERs, meaning that the errors are less frequent, a certain number of errors is needed in order to get an accurate estimate of the BER. At last, there is also a parameter changing the number of generated bits in each turn. In these simulations the parameter has been set to 100000 but any number can be chosen since the system will add zeros after the data in order to make it into a complete number of OFDM symbols and frames.

3.1.2 System

In the global variable which includes the system parameters, it is possible to change all parameters directly related to the system model.

First, a system that can simulate different number of antennas and users is essential in order to see how different antenna-user constellations act when implemented with low complex hardware. These parameters has been varied between the simulations and the chosen values can be seen in the figures in the results.

Then it is possible to set the frame structure in terms of number of subframes, number of timeslots and the number of OFDM symbols in each timeslots. It is possible to simulate the frame structure described in fundamentals, but for the simulations in these results the combination of one subframe with two timeslots, each with three OFDM symbols, has been simulated. The reason for this was to save simulation time.

Furthermore, this global variable includes the system bandwidth, in these simulations set to $20MHz$ according to the LTE standard, and the sample time which in this case corresponds to $50ns$. The up-sampling is also set here, and has in these simulations been set to 2. A higher up-sampling factor would take more overall simulation time and this is why 2 was chosen.

The channel parameters are also set in these global variables. The first option is to choose between an AWGN channel or a channel which includes both AWGN and Rayleigh fading. If the option with Rayleigh fading is chosen, it is possible to choose between four power delay profiles according to Table 2.1 and Table 2.2. In these simulations the Pedestrian A option has been chosen and kept throughout all the simulations in order to evaluate the changing parameters when experiencing the same kind of channel. It is also an option to turn the ZF precoder off. This parameter has been made changeable because of testing, but all presented simulations include includes ZF precoding.

The massive MIMO system investigated is also equipped with relevant channel coding to able to see the effect when this is included in the signal processing, and particularly how it effects the performance. The channel coding implemented is LDPC coding with the fixed block size 672, according to the IEEE 802.11ad standard. The available code rates are $\frac{1}{2}$, $\frac{5}{8}$, $\frac{3}{4}$ and $\frac{13}{16}$. Whether coding is used in the simulations or not can be seen in the figures of the results, and when present, the code rate can also be seen. LDPC coding was chosen as it is a good candidate because of its low-complexity.

The signal constellation is also changeable by choosing between the number of bits per symbol. The allowed options are 2, 5, 6 and 8 bits per symbol resulting in the constellations QPSK, 16-QAM, 64-QAM and 256-QAM. In these simulations 64-QAM has been used. 64-QAM can send more bits per symbol than the lower constellations but is also more sensitive to disturbances.

Other changeable parameters are the number of subcarriers per OFDM symbol, where 2048 has been used, and the number of subcarriers carrying data, which was set to 1200. Furthermore, the guard band, 848 subcarriers at each side, and the length of the cyclic prefix, 144 subcarriers, is defined here. All this according to LTE standards. The filter also have a number of options, where the first one is whether the entire signal should be filtered at once or if the filter should be applied for each OFDM symbol. In these simulations the entire signal has been filtered at once since it took the least simulation time. Other changeable parameters for the filter is the filter span and roll of factor, which determines the width and the slope of the pulse shaping filter. For the simulations in this report the entire signal has been filtered with a filter span of ten, and a roll of factor of 0.25. This pulse-shaping filter was a good suit for removing discontinuities between symbols during the simulations.

At last, it is also possible to choose whether the equalizer in the receiver should be based on perfect CSI or channel estimations. Here, perfect CSI is used since realistic pilots were not implemented in this system model. Using perfect CSI generates better results than a real scenario which uses channel estimation, since in reality perfect CSI is not available.

3.1.3 Fixed-point refinement and error insertion

In the global variable which includes the fixed-point refinement and error insertion parameters, it is possible to change all parameters directly related to the simulations where low-complexity hardware is simulated.

First, there are options for choosing whether fixed-point refinements should be turned on or off for the signal, filter coefficients and twiddle factor. If one of these are turned on, then the desired word length for this option needs to be set. The values of these parameters have been changed a lot during the simulations and the chosen word lengths will be visible in the figures in the results.

Furthermore, there are options for the error insertion simulations. These options starts with the choice to have the error insertion simulations on or off. If these simulations are on, then the choice is between simulating antenna outage or to simulate other errors of different magnitude. There are cases for 'Slight errors', 'Lots of errors' and 'Extreme errors'. The error arrays of these different errors can be seen in Appendix B. The number of antennas with errors is also optional in order to be able to compare different cases with different number of antennas with errors. All these parameters have been changed during this master's thesis work and the figures in the results will indicate whether the results are with or without errors and in the case of errors, the chosen magnitude of errors is displayed in the figure.

3.2 Simulation framework

To run the massive MIMO system model a simulation framework surrounding it is necessary. In order to get statistically significant results, the system model needs to run several times because of randomness involved in the system. This section is further explaining the implemented simulation framework.

One simulation of the program consist of three loops. One which iterates over a certain number of different channels, one which iterates the system for a range of SNR values and one which iterates each SNR value until it reaches a certain number of errors. The overall structure of the simulation can be seen in the block diagram in Figure 3.2.

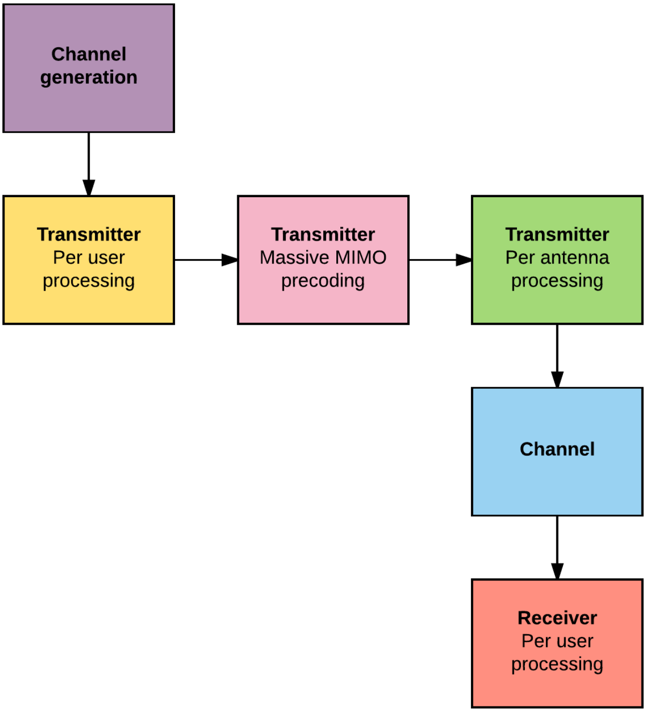

Figure 3.2: Overall simulation model.

The channel generation block runs every time a new channel loop begins. The blocks which represents the transmitter, channel and receiver runs for every new channel iteration and for every SNR value until it reaches a certain number of errors.

3.3 System model

In Figure 3.3 the block diagram of the complete system model is presented. This figure is a refinement of Figure 3.2. The system model is implemented for downlink since this is usually where most of the data is sent. Therefore it is relevant to investigate the downlink case in terms of the relation between the performance and using low-complexity hardware.

The system model assumes perfect synchronization and relies on perfect CSI. The following sections will further explain the structure and functionality of the system model.

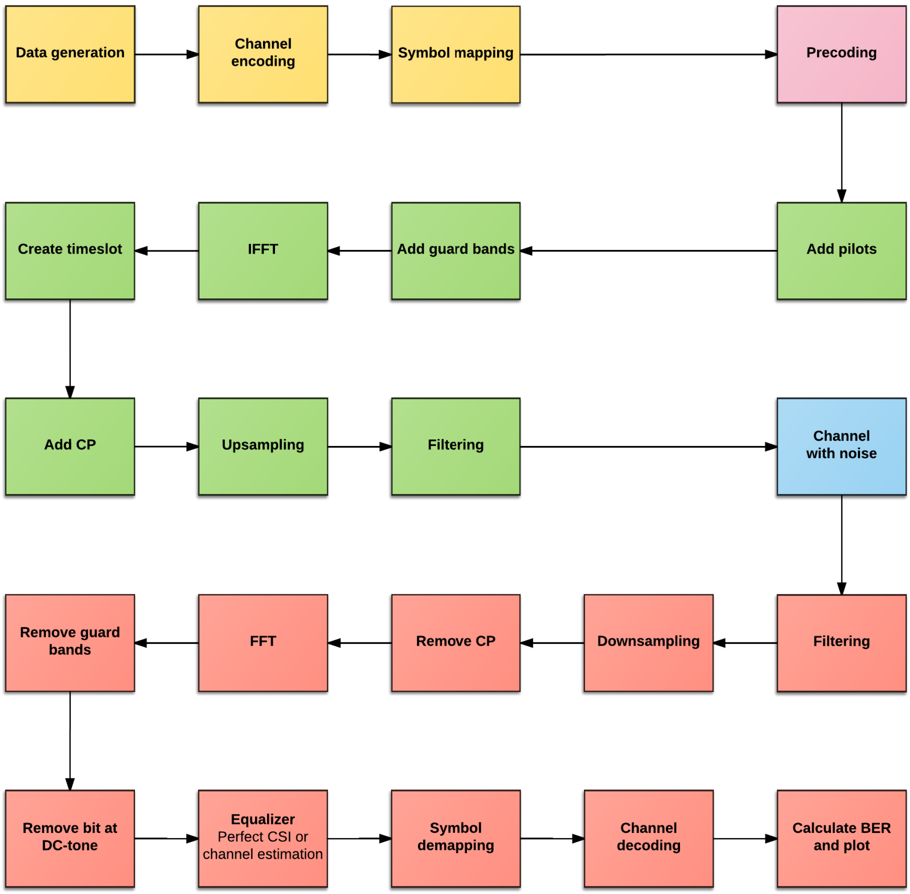

Figure 3.3: Overall process model.

3.4 Making the system model into massive MIMO

Since a massive MIMO system implies having a lot of antennas and several users, some parts in the process model in Figure 3.3 needs to be done for every user in the system, for every antenna in the base station, or for every user at the receiver side.

In Figure 3.4 the concept of a base station serving several users with a large amount of antennas is displayed. First the transmitter does signal processing for every one of the K users. Thereafter, the signal processing consists of a massive MIMO precoder which is based on the channels between the base station and the different users. In this case the massive MIMO precoder is based on perfect CSI. After the massive MIMO precoder, there is signal processing for the M antennas in the base station, before leaving the transmitter and passing through the channel. In the end, the signals reach the receivers where signal processing is done for every user. Every user is assumed to have one antenna.

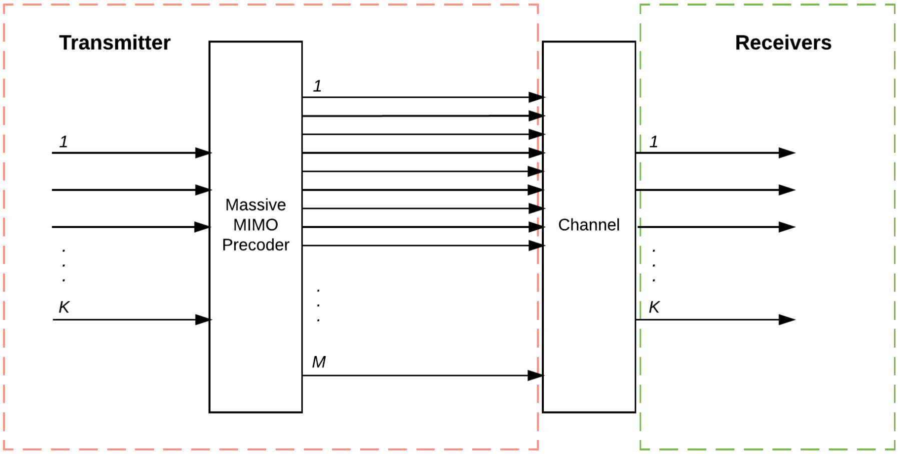

Figure 3.4: Visualization of a massive MIMO system.

To take the system model and develop it to a massive MIMO system, which can be seen in Figure 3.4, and to elaborate on the dimensions of massive MIMO and its implications on the system model, some further explanation is needed. A visualization of the dimensions and how it is connected to the system model can be found in Figure 3.5.

The simulation starts with a per-subcarrier generation of the chosen channel. The number of channels generated is K x M. This means one channel leading from each of the M base station antennas to each of the K users, resulting in a K x M channel matrix.

The next step in the simulation is the data generation, where a number of bits are randomly generated for every one of the K users. To precode the data, it is multiplied with the ZF precoder according to the formula

$$\boldsymbol{z} = \boldsymbol{W}\boldsymbol{x}.$$

This generates a length-M vector of precoded data, where each element represents the transmitted signal on one of the antennas in the base station. The M x K precoder comes from the K x M channel which were previous generated, since the system relies on perfect CSI rather than on channel estimates. The ZF precoder is generated as

$$\boldsymbol{W}_{\mathrm{ZF}} = \boldsymbol{H}^H(\boldsymbol{H}\boldsymbol{H}^H)^{-1}.$$

The precoded data, of the size M x 1, leaves the precoder continuing to the M antenna processes. After antenna processing the final transmit signals are ready to be transmitted. In the antenna processing the signal is converted into the time domain again by the IFFT.

The M x 1 transmit signals then passes through the channel which is computed as a convolution, because being in time domain, between the M x 1 transmit signals and the K x M channels which were generated at the beginning of the simulation. Then noise is added and the total result becomes

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{z} + \boldsymbol{n}.$$

This leads to a matrix with Kx1 received signals, one for each user. These K x 1 received signals are then processed by the K receivers.

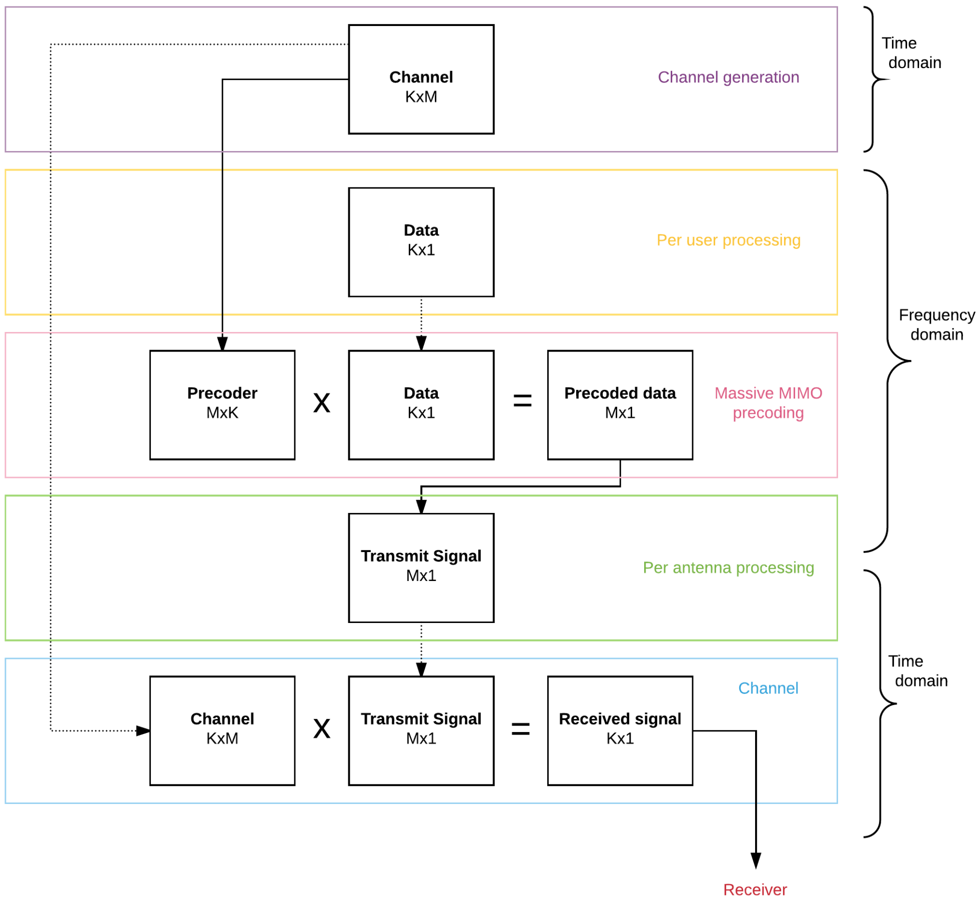

Figure 3.5: Dimensions of the Massive MIMO system.

3.5 Channel generation

As seen in Figure 3.2 the channel generation is the first thing which is in the simulation model. This channel is used when it is convolved with the transmit signal before getting to the receivers. Since the system relies on perfect CSI, this channel matrix is also used when creating the ZF precoder and in the equalizer on the receiver side.

In the system model there are, as previous listed, two different categories of channels, first there is a channel which is an AWGN channel which means that the channel matrix here is only 1, since the noise is added later in conjunction with the channel convolution. The other option is to have a channel with both AWGN and Rayleigh fading, where there are four options of power delay profiles (PDP) according to the ITU-R standard [10]. The different available PDPs can be seen in Figure 3.6.

In the simulations the Pedestrian A option with AWGN has been used. The power and the delays found in Table 2.1 is normalized after being generated in order to have a normalized power and synchronization in the system. Therefore, the Rayleigh fading does not effect the SNR in the simulation, but instead this is regulated via the AWGN, which is added later. One Rayleigh fading channel is generated, according to the Pedestrian A power delay profile, for every channel between the M antennas to the K users.

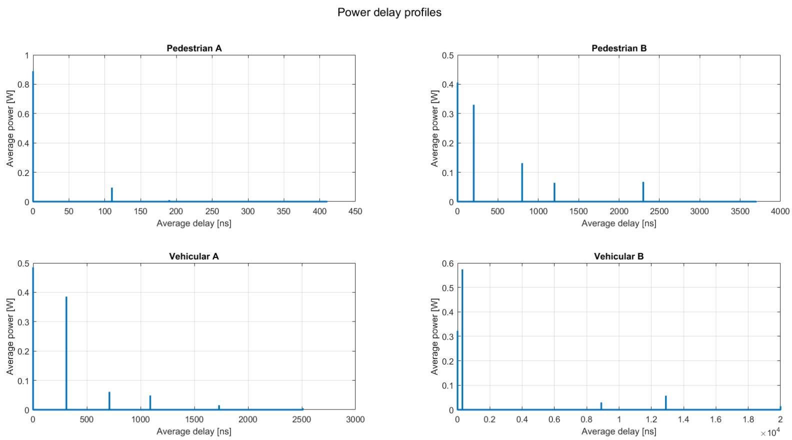

Figure 3.6: Power delay profiles according to the ITU-R standard [10].

3.6 Transmitter - Per-user processing

The signal processing in the transmitter starts with the processing for the K users, the yellow blocks in Figure 3.2 and 3.3. The per-user processing starts with random data generation for the K users.

Thereafter it is possible to choose whether an uncoded signal or a signal coded with LDPC is desired. Having coding optional gives the possibility to evaluate the performance with and without coding.

The last part of the per-user processing is the symbol mapping. An important factor for how robust a system can be against errors, is the signal constellation where large ones are more sensitive to distortions and noise. Because of this, it is important to be able to choose from different signal constellations in the simulations. After this the signal is normalized, with a normalization factor which depends on the chosen signal constellation. This normalization factor is defined in the script and can be seen in Appendix A.

3.7 Transmitter - Massive MIMO precoding

After the per-user processing, the data from the K users is precoded, the pink blocks in Figure 3.2 and Figure 3.3. The pilots, which are added later to the signal, are also precoded here. The precoding implemented is a ZF precoder using perfect CSI from the previously generated channels. It is possible to turn the precoder off and run the simulation framework as a MIMO system without precoding, but in this framework it is only possible to do this when $K = M$. After the ZF precoder the signal is normalized per-subcarrier.

3.8 Transmitter - Per-antenna processing

Following the ZF precoder is the per-antenna processing for the M antennas, the green blocks in Figure 3.2 and Figure 3.3. The per-antenna processing starts with adding the precoded pilots which in this framework, before precoding, is just an OFDM symbol set to all ones. Then for each 1200 data subcarriers, guard bands are added on each side of these data subcarriers. This results in OFDM symbols with 2048 subcarriers each.

Continuing the per-antenna processing, the signal is transformed into the time domain by the IFFT and normalized by multiplying the signal with $\sqrt{2048}$. Time slots are created so that each contains one OFDM symbol with pilots, followed by two OFDM symbols with downlink data. Thereafter the cyclic prefix is added to each OFDM symbol.

As a last part of the processing in the antennas, the signal is up-sampled according to the chosen value of this variable and then filtered with the filter options chosen in the script.

3.9 Channel

The transmit signals coming from the antennas gets convolved with the up-sampled channels in the blue blocks of Figure 3.2 and Figure 3.3. The signals reaching each user from each base station antenna are added together and AWGN is applied to the total signal per-user, giving it a certain SNR value in dB.

3.10 Receiver - Per-user processing

Each user in the system is modeled as having one receive antenna. For every one of the K users the same process is applied, shown as red blocks in Figure 3.2 and Figure 3.3. Generally it is the same processing steps as in the transmitter, but reversed. Therefore, the per-user processing in the receiver starts with filtering, down-sampling, removing of the cyclic prefix, multiplying with $\frac{1}{\sqrt{2048}}$ as a normalization due to the FFT, FFT and removing of the guard bands. Thereafter, the channel compensation in done, in terms of an equalizer. It is optional whether the equalizer should use perfect CSI or a channel estimation based on the sent pilots.

The receiver processing ends with symbol demapping, channel decoding and finally a calculation of the BER for the simulation. The BER generated for each SNR value in the simulation is stored in an array after terminating the SNR loop. If the number of channel loops is bigger than one, then the mean BER for each SNR value is calculated and presented as the BER for that SNR value in the simulation. This is done right after terminating the channel loop.

Chapter 4

Low complexity implementations of per-antenna functions

In this chapter, the low complexity implementations are presented in detail. Methods for implementing low accuracy in the filters and IFFTs are presented as well as the simulation of the hardware errors.

4.1 Low accuracy

In all fixed-point simulations, the signal was first made into fixed-point in the very beginning of each antenna function, where the total word length was set. The fixed-pointed signal was then used in simulations together with the fixed-pointed filter coefficients and the twiddle factor in the IFFT individually.

4.1.1 Filter

The filter was made into fixed point by lowering the accuracy of its filter coefficients. For all simulations, a 21-tap filter was used, completing the equation from section 2.16.1 makes the complexity after filtering:

$$42 \cdot (m+n) \cdot \log_2(m+n) + 42 \cdot m \cdot n$$

where n is the word length of the signal and m is the word length of the filter coefficients.

4.1.2 IFFT

In the IFFT, the twiddle factor, used in every butterfly stage, was made into fixed point. To sum up the fundamentals, section 2.16.1, the complexity of the IFFT, considering the number of adders and multipliers used in every stage of the butterfly, is:

$$66 \cdot n \cdot \log_2(n) + 44 \cdot n \cdot m$$

where n is the word length of the signal and m is the number of bits used to represent the twiddle factor.

4.2 Hardware errors

To simulate hardware errors, corruption of the signal was done at the very end of the antenna functions.

4.2.1 Voltage-Over-Scaling

Three different levels of corruption was simulated: 'Slight errors', 'Lots of errors' and 'Extreme errors', which bring 43%, 54% and 60% power savings respectively. Every level of corruption was done using error rate arrays, which corrupted the the most significant bit (MSB), and also several of the following bits, with a certain probability. More details on the error rate arrays and how they corrupt the signal can be found in Appendix B. The 'Slight error'-case, that resembles only a few errors on the affected antenna, generates an error on the MSB with a probability of 0.0002. For 'Lots of errors' and 'Extreme errors' the corresponding probability of errors on the MSB was 0.254 and 0.5 respectively, in the latter case it is the same probability for the eight MSBs. These values can also be found in the error rate arrays in Appendix B. Because of the additions and multiplications in the filter, the word length of the filter output is longer than the word lengths of the input and filter coefficients. Since the errors are implemented after filtering, the MSB in the error rate arrays are positioned at the 13:th bit, which is the maximum word length of the filter output.

The error arrays are not generated in this thesis, but provided for this work. The exact values were derived during simulations for different voltage conditions for the IC's.

4.2.2 Antenna outage

To simulate complete antenna outage, the maximum value of the signal was set on the affected antenna. Another way of simulating hardware failure is to simply not account for the affected antenna at all, and put the signal value to zero, but in this work the most pessimistic way of simulating hardware failure is considered.

Chapter 5
Results

In this chapter the results generated for this master's thesis will be presented. The results have been generated in order to investigate the aims of this master's thesis and the developed system model with a simulation framework, as described earlier. These results are presented and evaluated below.

This chapter is mainly divided into two parts. The first part consists of the results originating from the fixed-point refinement implementations in the filter and IFFT. There is also further evaluation and assessment of the possible performance loss caused by low-accuracy hardware. Examples of calculations on possible power saving results when using hardware with lower accuracy are also presented.

The second part mainly consists of the results originating from the simulations where errors were introduced in order to simulate hardware errors, like those caused by Voltage-Over-Scaling (VOS) and complete antenna outage. An assessment of the possible performance loss due to these errors is also presented as well as the possible power saving when working in the error resilient area.

5.1 Low accuracy

In this section the results for the simulations based on the fixed-point refinement implementations will be presented. That is, simulations where the signal, filter coefficients and twiddle factor are represented with different number of bits with the goal to lower the accuracy in order to save power in the digital signal processing. First results for the assessment and effects of fixed-point refinement in the IFFT will be presented, followed by the results for the simulations with fixed point refinement in the filter. The effects of the fixed-point refinement will be presented in comparison to the floating point case in the terms of achievable BER for the same SNR values.

5.1.1 IFFT

In the assessment of the low-accuracy implementation in the IFFT, the goal was to find out how many bits that were needed when changing the number of antennas, but keeping the same SNR value and still achieving the same BER. When investigating this, it is possible to make conclusions about how the number of antennas affects the needed complexity in the digital signal processing in the per-antenna functions. To get these results, simulations for different combinations of word lengths for the signal and the twiddle factor was done. For each one of these pairs the simulation was done in order to find the combinations which could achieve the BER 10^{-3} for the SNR 18 dB.

The BER value of 10^{-3} was chosen as a reasonable value of a BER in terms of performance versus simulation time. A lower BER would take more simulation time and a higher BER was considered to be too bad in terms of performance. To get the SNR value which was chosen for the simulations, a system with 4 users and 8 antennas was simulated during the chosen circumstances. The chosen SNR value at 18 dB is the SNR value which in this simulation could achieve the BER of 10^{-3}. The reasoning behind this was that if the combination with 4 user and 8 antennas could achieve this BER for a certain SNR value, then the other combinations with more antennas would also be able to achieve this BER for the same SNR value but with less number of needed bits.

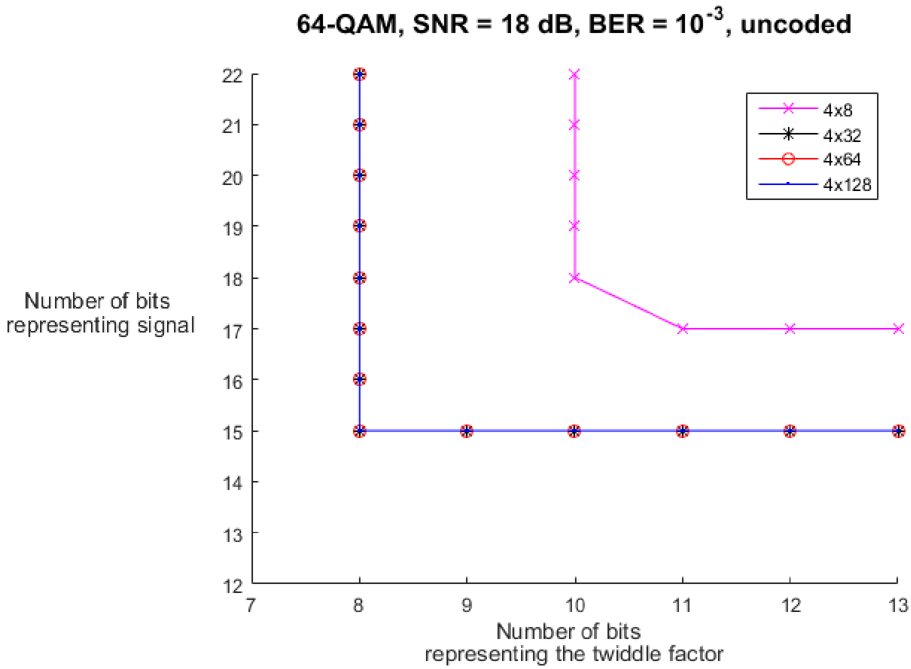

Figure 5.1: Number of bits needed to represent the signal and twiddle factor to achieve the target BER for different number of antennas.

For an uncoded signal with 64-QAM, a fixed SNR at 18 dB, a target BER of 10^{-3} and varying number of bits representing signal and twiddle factor, the results are presented in Figure 5.1. The figure includes simulations for the same number of users, but different number of antennas. The curves represent the minimum needed number of bits to achieve a BER of 10^{-3}, compared to the different number of antennas. To the left of and below the curves, this target BER cannot be achieved and above and to the right of the curves the performance will be sufficient or better.

The results for the cases with 32, 64 and 128 antennas are completely overlapping each other. The corner point, the point with the lowest number of needed bits, is at 15 bits for the signal and 8 bits for the twiddle factor. The case with 8 antennas on the other hand is not overlapping the other combinations and has two corner points, which both need a higher number of bits in order to achieve the target BER.

The results shows that when applying more than 32 antennas the number of needed bits do not decrease. This could be because of the several steps in the IFFT, where the results are multiplied with the twiddle factor. This could cause that a certain, minimum, number of bits are needed, and that an increased number of antennas do not compensate more for the information loss caused by the low accuracy. When using 8 antennas, more number of bits are required, probably because the number of antennas is not big enough to compensate for the low accuracy in the IFFT.

5.1.2 Filter

The same simulation as the one which was previous done for the IFFT was also done for the filter, with the difference that this simulation changes the number of bits needed for the filter coefficients instead. Since there are two processes in the antenna functions which take up most of the complexity, the IFFT and the filter, it is relevant to investigate both. This is relevant because then they can be compared and conclusions about the different results can be made. Results show the number of needed bits for these two cases and then conclusions will include complexity and possible energy savings.

For an uncoded signal with 64-QAM, a fixed SNR at 18 dB, a target BER of 10^{-3} and varying number of bits representing the signal and filter coefficients the results in Figure 5.2 are presented. The figure includes simulations for the same number of users but with different number of antennas and shows the minimum needed number of bits to achieve a BER of 10^{-3} for a SNR value at 18 dB. The results for the cases with 64 and 128 antennas are almost completely overlapping each other. The graph for the case with 32 antennas is close to the 64 and 128 cases with the exception that one or two more bits is needed for the signal to achieve the BER at 10^{-3}. There are two corner point for 64 antennas, one is at six bits for the filter coefficients and four bits for the signal. The other one is at four bits for the filter coefficients and six bits for the signal. The first one of these has been used in further simulations. These corner points can also be compared to the result for the IFFT in Figure 5.1 where the corner point for 64 antennas was at 8 bits for the twiddle factor and 15 bits for the signal.

An interesting result is the comparison to the IFFT, where more bits, were needed at the corner points. An interpretation of these results is that it is possible to push the accuracy lower in the filter than in the IFFT. This is the reason why further assessment of low accuracy hardware will focus on the filter. This is also interesting because the filter is more power consuming than the IFFT which makes it even more relevant to focus on the filter.

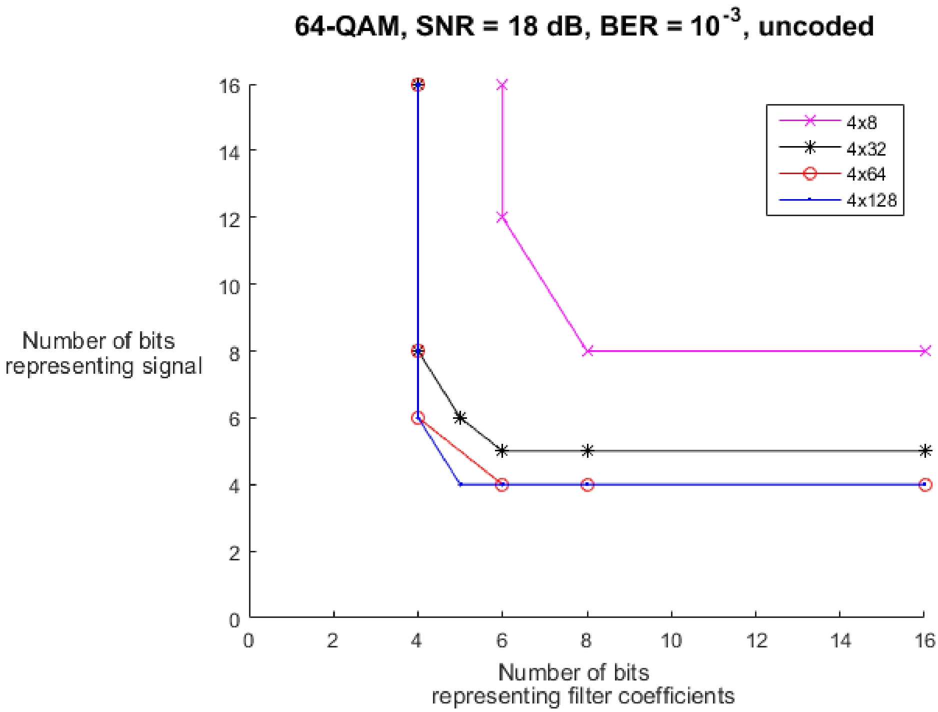

Figure 5.2: Number of bits needed to represent the signal and filter coefficients to achieve the target BER for different number of antennas.

Another result is that once again the difference in number of needed bits differs quite much for the 8 antenna case compared to the other cases. One reason for this is that for this combination of users versus antennas, the system load in relation to the number of antennas is the heaviest one. For this reason, the system does not really experience the complete positive effect which massive MIMO yields, not having significantly more number of antennas relative to the number of users.

Different number of users

With the observation that the word length in the filter could be pushed further than in the IFFT, further assessment of the filter was done. Fixing the number of antennas at 64, but varying the number of users between 4, 8, 16 and 32 leads to results for how different system loads affects the number of needed bits.

With the same conditions as in the previous graph, but with a fixed number of antennas of 64 and varying number of users, the result became as shown in Figure 5.3. The graphs show the minimum number of needed bits to represent the signal and filter coefficients in order to achieve the target BER of 10^{-3} for the SNR value 18 dB. The graphs for the different combinations are almost completely overlapping for the cases with 4, 8, 16 and 32 users.

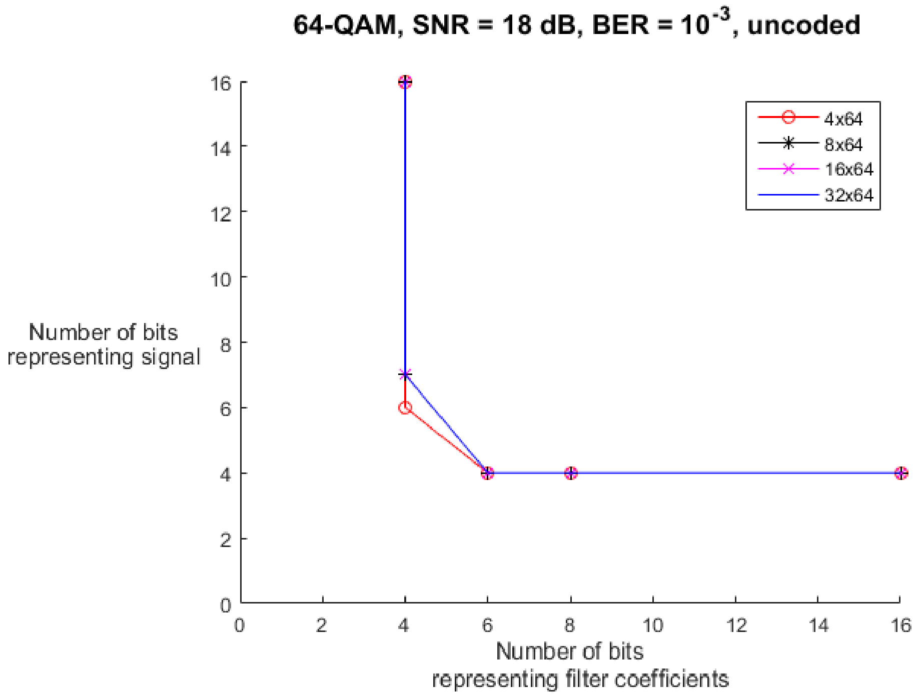

Figure 5.3: Number of bits needed to represent the signal and filter coefficients to achieve the target BER for different number of users.

The different system loads simulated did not result in different requirements, except for one case, on the word lengths for the signal and the filter coefficients. The combination with 32 users and 64 antennas is the combination with the highest system load in this simulation. Required word lengths may increase for higher system loads, but this has not been investigated.

Different BER

Continuing the evaluation, a very important factor in a communications system is the performance and what BER that is possible to achieve in the system. Therefore, simulations when keeping the combination of 4 users and 64 antennas was done in order to investigate how many more bits were needed to achieve different BERs.

For an uncoded signal with 64-QAM, a certain SNR at 18 dB and a fixed combination of 4 users and 64 antennas, the results can be seen in Figure 5.4. The graphs are showing the minimum required bits which can achieve the target BER in the simulation and are partially overlapping, but with some differences which mainly can be seen at the corner points.

To achieve a better BER, e.g. going from 10^{-3} to 10^{-4}, it could be sufficient with just a few more number of bits, meaning that it could be worth to increase

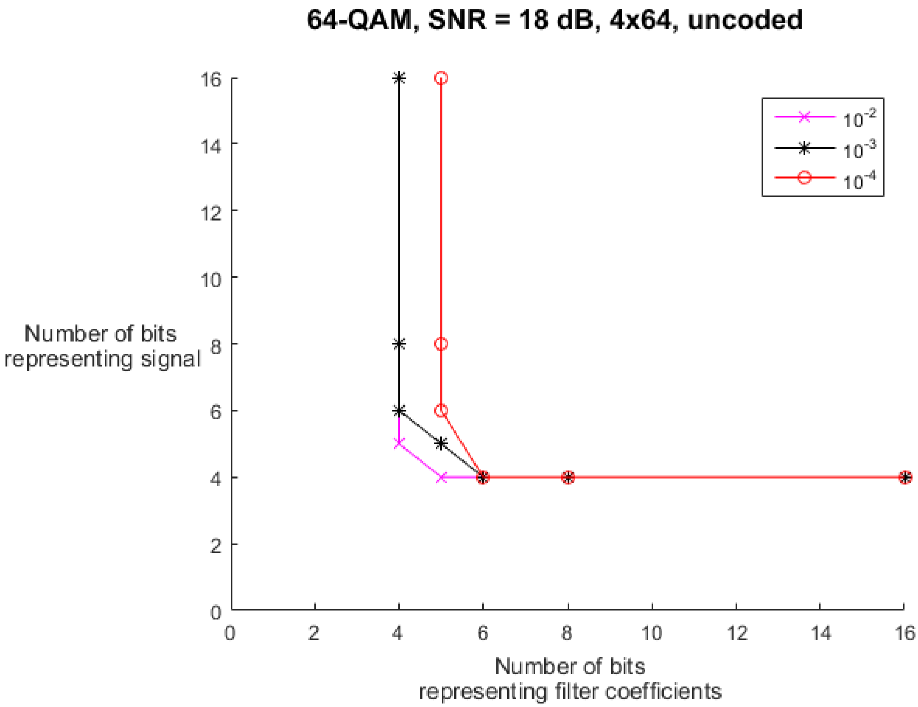

Figure 5.4: Number of bits needed to represent the signal and filter coefficients to achieve different BER with the same number of users and antennas.

the number of bits by just one bit each for the signal and filter coefficients. In this way a much better BER can be achieved while not increasing the complexity of the digital signal processing too much, but still yield an increased energy efficiency.

Evaluation of corner point

In order to see if the corner point with four bits for the signal and six bits for the filter coefficients were good enough to be kept in the further simulations, the performance at this point was evaluated. To do this, the same fixed-point was used and simulated for a range of SNR values to see how the BER was affected.

For an uncoded 64-QAM signal, with the fixed-point word lengths four for the signal and six for the filter, using four users and 64 base station antennas, the BER performance turned out as in Figure 5.5. For higher SNR values the slope of the curve starts to decrease which indicates that the chosen fixed-point values results in an error floor. This means that the BER will not improve much further even when the SNR increases even more.

Reaching an error floor at the BER 10^{-4} is not sufficient, if no coding is used, so for further evaluation, more bits are needed to avoid reaching an error floor and in that way keep continuing to improve the performance even for higher SNR

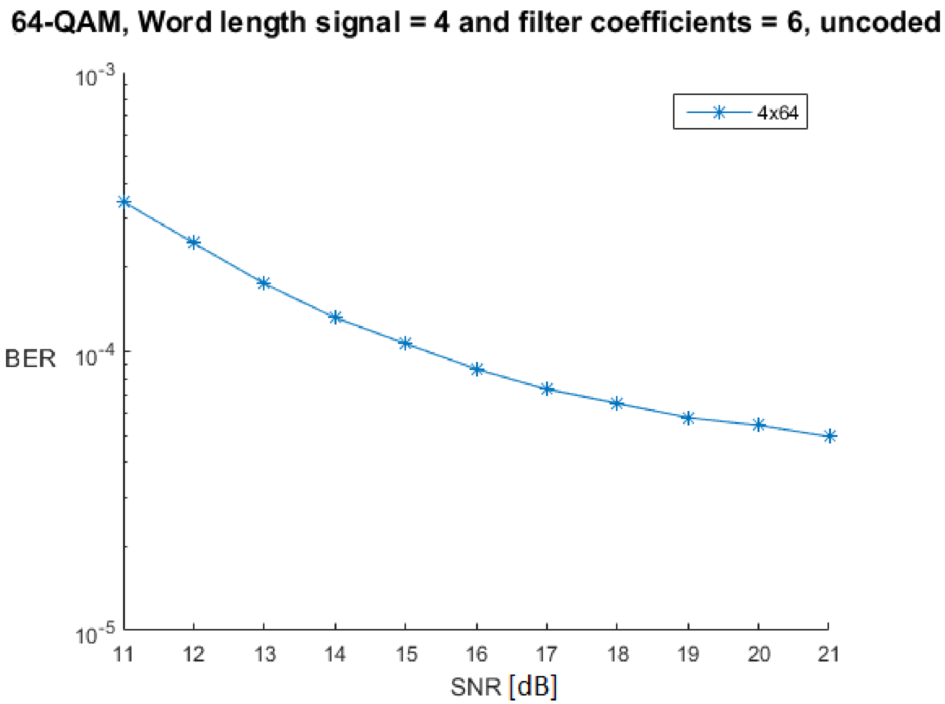

Figure 5.5: The corner point simulated for several SNR values, resulting in an error floor.

values. In Figure 5.4 it was shown that just one or a few more bits for the fixed-point values could be enough to improve the BER remarkably. In communications systems, coding is usually applied. Therefore, an attempt to remove the error floor by applying LDPC coding was done. For the same fixed-point values as before, a curve for several SNR values was simulated with LDPC coding

Figure 5.6 shows the resulting BER for the chosen fixed-point word lengths but when LDPC coding with a coding rate on $\frac{3}{4}$ is applied. It shows a slope which, after a certain SNR point, decreases rapidly for SNR values much lower than the ones in Figure 5.5.

Adding LDPC coding makes the error floor so low that it is no longer visible in the figure and no longer acts as a practical limitation to the system. LDPC coding improves the performance remarkably but increases the complexity of the digital signal processing. In the further simulations, LDPC coding will not be used. Instead the number of bits for the fixed-point value of the signal will be increased by one, i.e. five bits representing the signal. The reason for this is that simulations without channel coding takes less time, but when increasing the number of bits it could still be possible to run simulations without a visible error floor.

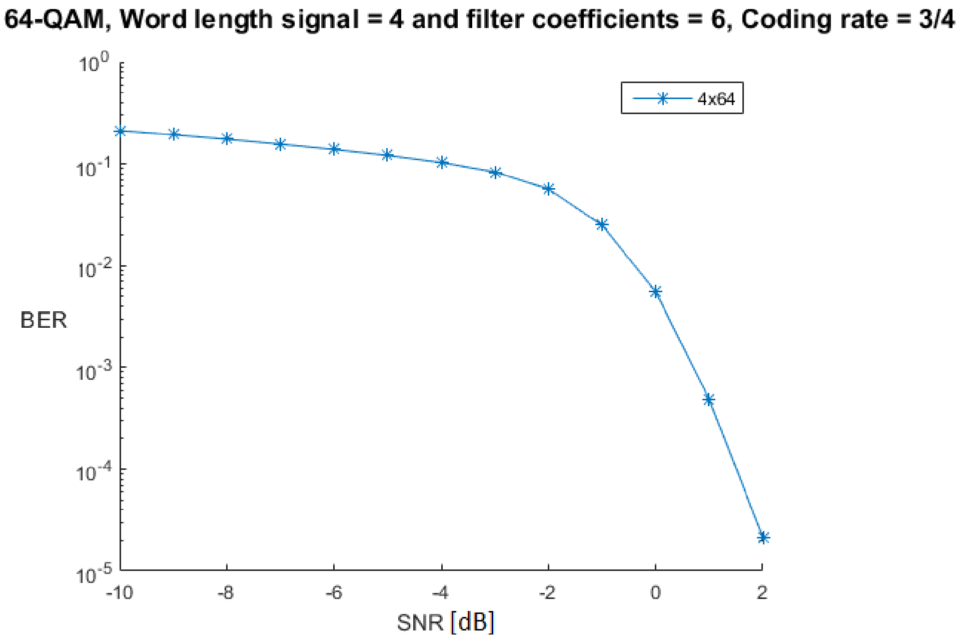

Figure 5.6: The corner point simulated for several SNR values with LDPC coding.

Assessment of the performance loss and power saving

When fixed-point values are set, some performance can be lost if the number of chosen bits representing the signal and filter coefficients are not enough to get a result similar to the floating point case. As a result from the error floor shown in Figure 5.5 the chosen number of bits representing the signal is increased from four to five bits, while keeping the number of bits representing the filter coefficients at six. The curve for this new combination of fixed-point values was simulated for a range of SNR values in order to see the possible performance loss caused by this. When doing so, it is relevant to compare this to the floating point case to see the difference. Furthermore, it is interesting to calculate and compare the complexity for the case when using floating point as well as the case when using fixed-point values. In this way a number for the possible energy savings can be presented.

In Figure 5.7 the floating point case is compared to the fixed-point case. For the lower SNR values the curves are basically the same, but the difference get bigger for higher SNR values. For a BER at 10^{-3} the SNR degradation is about one dB. Though, it might be beneficial to increase the number of bits representing the signal with just one or two bits to get a better accuracy and a smaller SNR degradation compared to the floating point case. The remarkable power saving due to using low accuracy hardware will still be present, but the SNR degradation would probably be smaller.

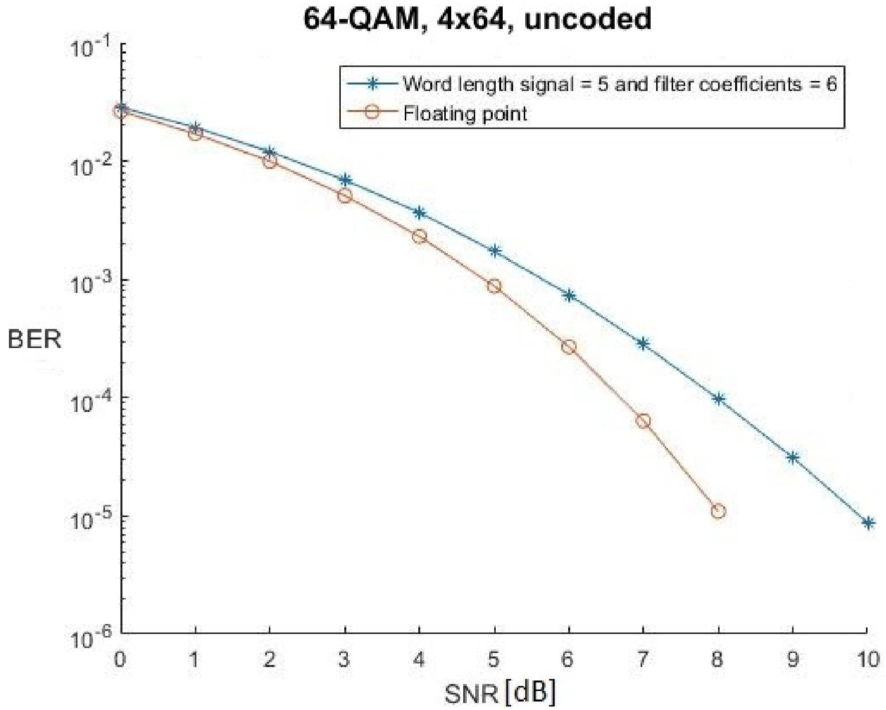

Figure 5.7: Comparison between floating point and fixed point.

When using the formula to calculate the complexity the result becomes

$$42 \cdot (6+5) \cdot \log_2(6+5) + 42 \cdot 6 \cdot 5 = 2858.3.$$

In comparison to the case where 32 bits are used to represent both the signal and filter coefficients the result is

$$42 \cdot (32+32) \cdot \log_2(32+32) + 42 \cdot 32 \cdot 32 = 59136.$$

Comparing these two leads to

$$1 - \frac{2858.3}{59136} = 0.9517.$$

This shows that the energy savings for this fixed-point case will be 95%, as compared to the case when using 32 bits to represent both the signal and filter coefficients, but for a BER of 10^{-3} there will be a SNR degradation at one dB.

5.2 Hardware errors in the filter

The further simulations will introduce errors of different magnitude, representing VOS, and simulate failures on one or several antennas. Since IC's traditionally are designed at their worst corners, which is not very energy efficient, it is important to investigate how a system performs when operating in the error resilient area in order to increase the energy efficiency. This, in combination with previous results on low accuracy hardware opens up for greater power savings.

5.2.1 Voltage-Over-Scaling

To simulate cases with VOS, errors of different magnitudes were introduced into the system. Allowing a bigger amount of errors can save more energy but will have a bigger impact on the performance. Therefore, it is necessary to investigate the possible performance loss and to calculate the possible energy savings in order to get a picture of the benefits and losses. How the errors are introduced was described in the previous chapter and the error matrices for the different cases can be seen in Appendix B.

Slight errors

First, an error case with only slight errors was considered and simulated in order to assess the performance loss. Even for slight errors it is possible to save 43% power which makes it reasonable to look closer into.

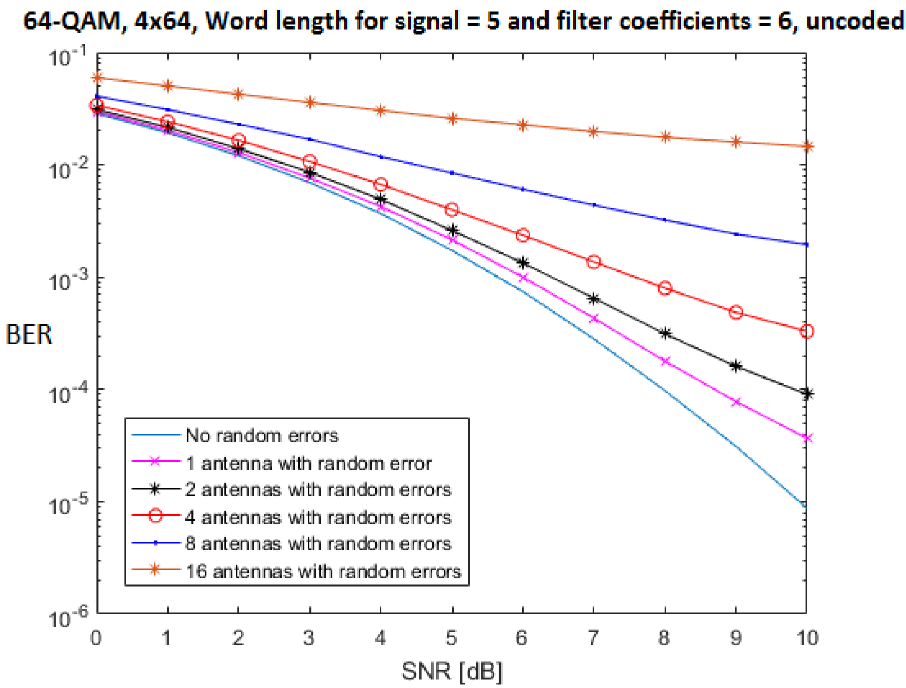

Figure 5.8: Simulation of a system where slight errors has been introduced. The labels in the figure corresponds to the graphs from bottom and up.

Figure 5.8 shows the simulation results for an uncoded 64-QAM signal with 4 users, 64 base station antennas and the previous chosen fixed-point word lengths. The different curves represent the case were no antennas have errors and the cases were 1, 2, 4 8 and 16 antennas have errors. For just a few antennas with errors the curves follow the no error case but with a bit worse BER for higher SNR values. When increasing the number of antennas with errors the result gets even worse and for the case with 16 antennas the curve is getting almost completely flat at higher SNRs.

Allowing just slight errors on a few antennas can be possible to handle. When using just a few more bits to represent the signal, maybe the system can recover even more efficiently from slight errors. Using a smaller signal constellation might also be more error resilient.

Assessment of the performance loss and power saving

When using fixed-point values it has been shown that it is possible to save power. When also scaling down the supply voltage and allowing errors to occur it is possible to save even more power. Both fixed-point values and errors causes a performance loss. To assess this performance loss a case with errors was compared to the no error case and associated to the theoretical power saving.

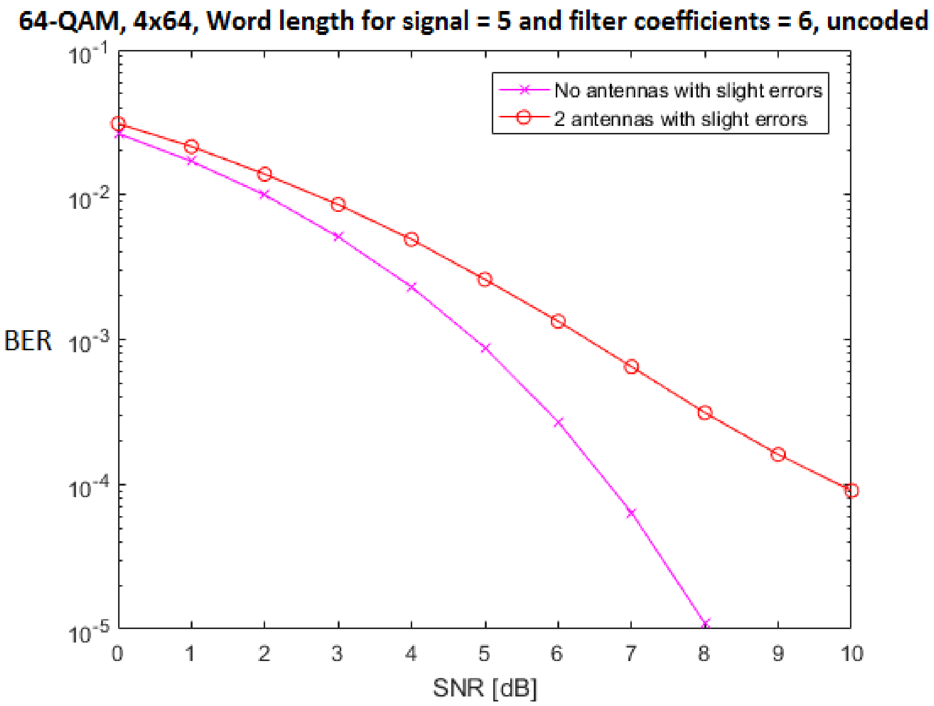

Figure 5.9: Comparison between fixed point with no errors and fixed point with slight errors on two antennas.

Comparing the no error case to the case with slight errors on two antennas gives two curves which are close at low SNR values, as seen in Figure 5.10, but differs more for higher SNR values.

According to [24] it is possible to save 43% power when allowing slight errors in the system. Combining this with the possible 95% power saving due to the fixed-point refinement this results in

$$1 - ((1 - 0.95) * (1 - 0.43)) = 0.9715.$$

This shows a 97% power saving compared to the case when using 32 bits to represent both the signal and the filter coefficients. For a BER of 10^{-3}, the SNR degradation between the fixed-point with no errors and the fixed-point with slight errors on two antennas means a SNR degradation on approximately one dB.

Lots of errors

Further the error case with lots of errors, about 25% probability that an error occurs on the MSB, was considered. Allowing more errors in the system increases the possible energy savings and for this case 54% power saving is possible [24].

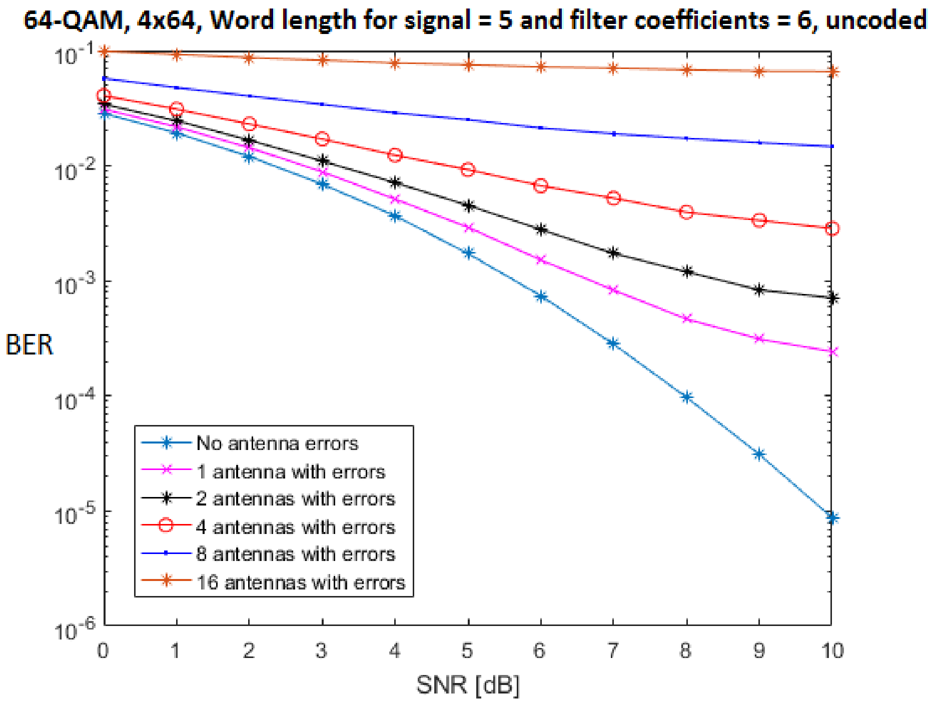

Figure 5.10: Simulation of a system where lots of errors has been introduced. The labels in the figure corresponds to the graphs from bottom and up.

Figure 5.10 shows the simulation results for an uncoded 64-QAM signal with 4 users and 64 antennas and the previously chosen fixed-point values. The different curves represent the case were no antennas have errors and the cases were 1, 2, 4, 8 and 16 antennas have errors. Even for a few number of antennas with errors the difference between the no error case and the error case is quite big, especially for the higher SNR values.

The difference between the no error case and the different error cases is bigger in Figure 5.10 than in Figure 5.8 which is reasonable since the probability of having an error on the MSB is bigger in the first one. The curves become almost flat quite fast which indicates that the magnitude of errors is hard to handle for this system.

Extreme errors

The last error case is the extreme error case, with about 50% chance of an error on the eight MSBs. Allowing even more errors in the system increases the possible energy savings and for this case 60% power saving is possible [24].

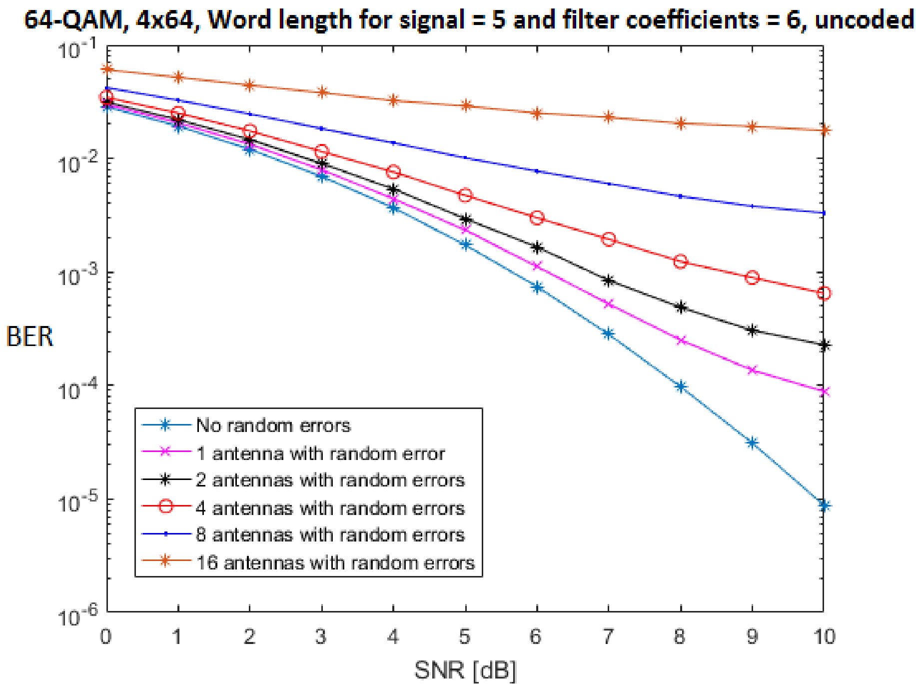

Figure 5.11: Simulation of a system where extreme errors has been introduced. The labels in the figure corresponds to the graphs from bottom and up.

Figure 5.11 shows the simulation results for an uncoded 64-QAM signal with 4 users and 64 antennas and the previously chosen fixed-point values. The different curves represent the case were no antennas have errors and the cases were 1, 2, 4, 8 and 16 antennas have errors. In comparison to Figure 5.8 the curves with errors are only slightly worse and in comparison to Figure 5.10 the curves with errors are slightly better.

Intuitively, the results should get worse when having a higher probability of errors. Despite that, the performance in Figure 5.11 is slightly better than the one in Figure 5.10. Investigations of the system has been made in order to find an explanation for this inconsistency. Preliminary, the explanation is that it is a result caused by the power control in the transmitter, which is the power scaling in this model. Further investigations will be made about this inconsistency, but is out of scope of this master's thesis.

5.2.2 Antenna outage

It is also interesting to simulate the worst case scenario which in this case is a complete failure of one or several antennas. The antennas with complete outage was simulated as putting the whole transmit signal from those antennas to the maximum value of that signal.

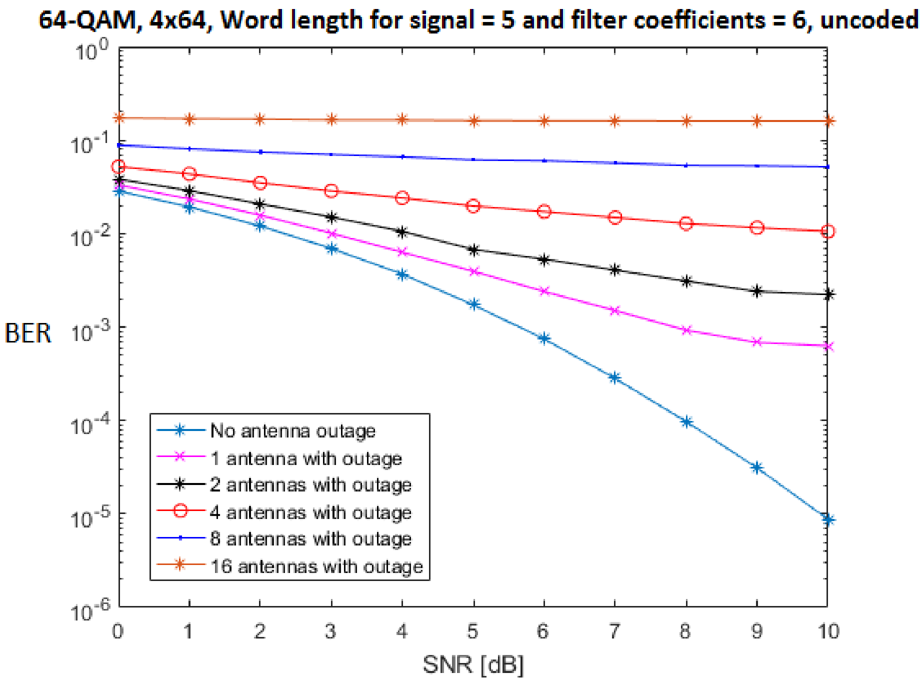

Figure 5.12: Simulation of a system where antenna outage has been introduced. The labels in the figure corresponds to the graphs from bottom and up.

Figure 5.12 shows the simulation results for an uncoded 64-QAM signal with 4 users and 64 antennas and the previously chosen fixed-point values. The different curves represent the case were no antennas are in outage and the cases were 1, 2, 4, 8 and 16 antennas are experiencing antenna outage. When adding antennas with outage the performance quickly gets much worse which is probably a result of how the outage was modeled. If the simulation model for antenna outage had been constructed in another way, the results may have looked differently.

This system is too sensitive to handle antenna outage, even for a few number of antennas. Increasing the number of bits representing the signal or changing the signal constellation to a constellation which is more robust towards errors could increase the ability to tolerate one or several antennas experiencing complete failure.

Chapter 6
Conclusions

Conclusion which can be made from this master's thesis is that low-complexity solutions are possible in massive MIMO systems. Ever for short word lengths, it is possible for massive MIMO systems to recover from some errors. However, there is a trade-off between saving energy by decreasing the number of bits used to represent the signal and the filter coefficients and the possible SNR degradation which results from this. The results have shown that by using only a few more bits a significantly better BER can be achieved, meaning that it could be worth to increase the complexity with these few bits in order to decrease the SNR degradation.

The simulations also show that it is possible to push the word lengths further down in the filter than in the IFFT. This conclusion can be drawn from the results that for 4 users and 64 base station antennas one of the corner point for the filter was at four bits for signal and six bits for filter coefficients while the corner point for the IFFT was at 15 bits for the signal and eight bits for the twiddle factor.

Consistently, throughout this report, the results are based on a signal with a 64-QAM constellation. Two reasons why the results for the simulated VOS and antenna outage did not get better is partly that the fixed-point word length for the signal could have needed one or a few more bits and partly that 64-QAM is more sensitive to errors than smaller signal constellations. The conclusions from this is that 64-QAM represented with these fixed-point word lengths is a bit too sensitive to sufficiently recover from errors and antenna outage.

Furthermore, simulations have shown that it can be better with more extreme errors where the error rate on the eight MSBs is 50% than having a 25% probability on the MSB. This is an interesting result but unfortunately a clear explanation for this has not been found. A preliminary explanation is that it is caused by the power control in the transmitter, which is the power scaling in this model.

Another conclusion drawn is that large power saving are possible for the digital signal processing in massive MIMO systems. More specifically 97% can be saved when using the fixed-point word lengths five for the signal and six for the filter coefficients and allowing slight errors on two antennas. For a BER of 10^{-3} this would mean a SNR degradation of about two dB. The possible power savings, resulting from the low-accuracy hardware implementation, is in comparison to the case where 32 bits are used both for signal and filter coefficients. Since the simulation with LDPC coding showed that uncoded BERs around 10^{-2} - 10^{-3} can be quite adequate to get a good result with coding, the most interesting part of the plot is around 2-4 dB. In this area the difference between the curves are smaller than for the higher SNR values.

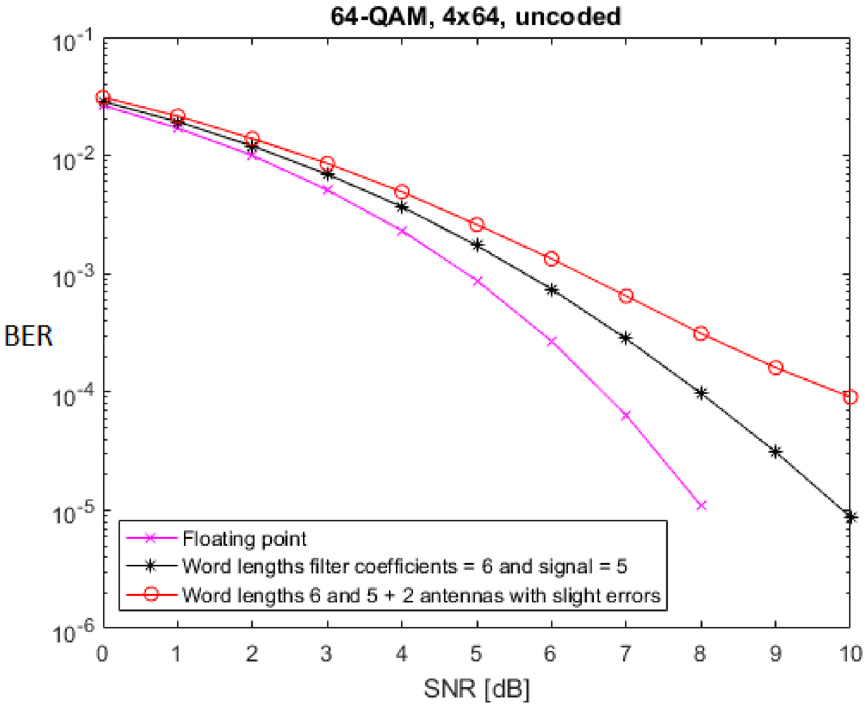

Figure 6.1: Comparison between floating point, fixed point and fixed point with slight errors on two antennas.

In Figure 6.1 the curves for these cases are shown and the SNR degradation at circa two dB at 10^{-3} is visualized as the difference between the curve showing the floating point and the curve which is with the fixed-point implementation and slight errors on two antennas. The third curve in between is the case with the fixed-point implementation but no antenna errors.

Summarizing this master's thesis leads to the final overall conclusion that massive MIMO can provide high spectral efficiency and implemented with low-accuracy hardware it can still be error resilient leading to high energy efficiency.

Chapter 7
Future work

There are a lot of possibilities of future continuation of this master's thesis. First of all, it would be interesting to do the same kind of evaluation with fixed-point refinement and introducing errors for different signal constellations such as for example QPSK and 16-QAM. Since these lower constellations are less sensitive to errors in comparison to 64-QAM, these systems would probably turn out to be more error resilient than 64-QAM. Lower fixed-point values could also be possible for another lower signal constellation. In this evaluation of other signal constellations, it would also be interesting to do calculations on the possible energy savings for these constellations and compare this to the 64-QAM case. A comparison between these systems, when taking the possible throughput and the possible energy saving into account, would be valuable.

In the error simulations in this master's thesis, only the case for 4 users and 64 base station antennas has been considered. Therefore simulations with other combinations using more antennas could generate results about how bigger massive MIMO systems handle these errors and if it would be possible to use 64-QAM and get better results for these cases when more antennas are used.

In this master's thesis the focus has been on the filter since conclusions were made, based on the simulations, that it was here the biggest possibilities for energy saving were. The same simulations and assessment as has been done for the filter could also be done for the IFFT.

Other possible continuations of this work is to investigate and develop digital signal processing functions which can discover and correct errors during run-time in order to build an even more error resilient system.

References

[1] MAMMOET - Massive MIMO for Efficient Transmission, `mammoet-project.eu/`

[2] E. G. Larsson, F. Tufvesson, O. Edfors and T. L. Marzetta, "Massive MIMO for Next Generation Wireless Systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186-195, February 2014

[3] P. Nuyts, P. Reynaert and W. Dehaene, "Frequency-Domain Analysis of Digital PWM-Based RF Modulators for Flexible Wireless Transmitters", *IEEE Transactions on Circuits and Systems I*, vol. 61, no. 1, pp. 238-246, January 2014

[4] Y. Huang, M. Li, C. Li and L. Van der Perre, "Partial Computation-skip Scheme for Power Supply Voltage Scaling", *First Workshop on Approximate Computing (WAPCO), in conjunction with the HiPEAC Conference*, Netherlands, January 2015

[5] J. Vieira, S. Malkowsky, K. Nieman, Z. Miers, N. Kundargi, L. Liu, I. Wong, V. Owall, O. Edfors and F. Tufvesson, "A Flexible 100-antenna Testbed for Massive MIMO", *Globecom 2014 Workshop - Massive MIMO: From Theory to Practice*, pp. 287-293, USA, December 2014

[6] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020 `http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html`

[7] M. Simko, C. Mehlfuhrer, T. Zemen and M. Rupp, "Inter-Carrier Interference Estimation in MIMO OFDM Systems with Arbitrary Pilot Structure", *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, Hungary, May 2011

[8] Y. Huang, C. Desset, A. Bourdoux, W. Dehaene and L. Van der Perre, "Massive MIMO Processing at the Semiconductors Edge: Exploiting the System and Circuit Margins for Power Savings", *The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing*, USA, March 2017

[9] E. Bjornson, E. G. Larsson and T. L. Marzetta, "Massive MIMO: Ten Myths and One Critical Question", *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114-123, February 2016

[10] Recommendation (1997) ITU-R M.1225. Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000

[11] G. E. Moore, "Cramming More Components onto Integrated Circuits", *Electronics*, pp. 114-117, April 1965

[12] J. M. Rabaey and S. Malik, "Challenges and Solutions for Late- and Post-Silicon Design", *IEEE Design & Test of Computers*, pp. 296-302, July/August 2008

[13] T. Parfait, Y. Kuang and K. Jerry, "Performance Analysis and Comparison of ZF and MRT Based Downlink Massive MIMO Systems", *in the proceedings of 2014 Sixth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 383-388, China, July 2014

[14] Lund University, "Researchers Set New 5G-world Record and Earn Multiple Awards", `www.lunduniversity.lu.se/article/researchers-set-new-5g-world-record-and-earn-multiple-awards`

[15] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin and R. Zhang, "An Overview of Massive MIMO: Benefits and Challenges", *IEEE journal of selected topics in signal processing*, vol. 8, no. 5, pp. 742-758, October 2014

[16] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays", *IEEE Signal processing magazine*, pp. 40-60, January 2013

[17] B. Debaillie, C. Desset and F. Louagie, "A Flexible and Future-Proof Power Model for Cellular Base Stations", *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*, Scotland, May 2015

[18] W. V. Heddeghem, M. C. Parker, S. Lambert, W. Vereecken, B. Lannoo, D. Colle, M. Pickavet and P. Demeester, "Using an Analytical Power Model to Survey Power Saving Approaches in Backbone Networks", *Networks and Optical Communications (NOC), 2012 17th European Conference*, Spain, June 2012

[19] R. E. Lander and M. J. Fischer, "Parallel Prefix Computation", *Journal of the Association for Computing Machinery*, vol. 27, no. 4, pp. 831-838, October 1980

[20] G. Dimitrakopoulos and D. Nikolos "High-speed Parallel-prefix VLSI Ling Adders", *IEEE Transactions on Computers*, vol. 45, no. 2, pp. 225-231, January 2005

[21] Y. Huang, A. Kapoor, R. Rutten and J. P. de Gyvez, "A 13 Bits 4.096 GHz 45 nm CMOS Digital Decimation Filter Chain with Carry-Save Format Numbers", *Microprocessors & Microsystems*, vol. 39, no. 8, pp. 869-878, November 2015

[22] C. R. Baugh and B. A. Wooley, "A Two's Complement Parallel Array Multiplication Algorithm", *IEEE Transactions on Computers*, vol. C-22, no. 12, pp. 1045-1047, December 1973

[23] Y. Huang, M. Li, C. Li, P. Debacker and L. Van der Perre, "Computation-skip Error Mitigation Scheme for Power Supply Voltage Scaling in Recursive Applications", *Journal of Signal Processing Systems*, vol. 84, no. 3, pp. 413-424, January 2016

[24] M. Fojtik, D. Fick, Y. Kim, N. Pinckney, D. Harris, D. Blaauw and D. Sylvester, "Bubble Razor: An Architecture-Independent Approach to Timing-Error Detection and Correction", *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International*, pp. 488-490, February 2012

[25] F. E. Salem, A. Tall and Z. Altman, "Energy Consumption Optimization in 5G Networks Using Uultilevel Beamforming and Large Scale Antenna Systems", *Wireless Communications and Networking Conference (WCNC), 2016 IEEE*, Qatar, April 2016

Appendix A

MATLAB script

Below is the code for the MATLAB start-up script which has been developed and used in the simulation framework. This start-up script is where the parameters are set. The rest of the program runs when calling mainMaMiDL at the end of this start-up script.

```
% Massive MIMO script
%
% Master thesis: Signal processing in Massive MIMO systems realized
% with low complexity hardware
% By: Sara and Micaela
% Last updated: 2016-08-31

clear
clear global
%close all
%clf

%% Define the global variables
global sim sys expr

%% The global variable for the simulation parameters
sim = struct(...
    ... % Channel  loop
    'nChannels' , 1000 , ... % Number of different channel generations
    ... % SNR loop
    'snrPointsStart' , 0 , ... % First SNR value
    'snrPointsEnd' , 10 , ... % Last SNR value
    'snrStepSize' , 1 , ... % Step size in SNR loop
    ... % Error loop
    'nLoopErrors' , 20 , ... % Number of errors needed before
                         ... % terminating while-loop
    'maxErrorLoops' , 100 , ... % Maximum number of iterations in the
                            ... % while loop (OBS, NOT USED)
    ... % Data generation
    'nGeneratedBits' , 100000 ... % Number of data bits to be generated
);
```

```
%% The global variable for the system parameters
sys = struct(...
    ... % Configuration
    'nUsers' , 4 , ... % Number of users, K
    'nAntennas' , 64 , ... % Number of antennas, M
    'diversityOrder' , 4 , ... % Diversity order of the system
    ... % Frame structure
    'nSubframes' , 1 , ... % Number of subframes in one frame
    'nTimeslots' , 2 , ... % Number of timeslots in one frame
    'nOfdmInTimeslot' , 3 , ... % Number of OFDM symbols per timeslot
    ... % Synchronization
    'bandwidth', 20e6 , ... % Bandwidth
    'sampleTime' , 5e-8 , ... % Sample time (1/bandwidth)
    'upsamplingFactor' , 2 , ... % Upsampling factor
    ... % Channel
    'channelType' , 1 , ... % 0: AWGN channel, 1: Rayleigh fading
    'powerDelayProfile' , 1 , ... % Power deley profile, Rayleigh channel
                              ... % (0, 1, 2, 3, 4 for
                              ... % 1 tap, PedA, PedB, VehA, VehB)
    'precoding' , 1 , ... % 0: Precoding off, 1: Precoding on
    ... % Channel coding
    'channelCoding' , 0 , ... % 0: uncoded, 1: coded with LDPC
                          ... % (for uncoded change to hard demapping,
                          ... % for coded change to soft)
    'channelCodingRate' , 3/4 , ... % Channel coding rate
                                ... % (1/2, 5/8, 3/4, 13/16)
    'blockSize' , 672 , ... % Coding block size for channel coding (fixed)
    ... % Modulation
    'nBitsPerSymbol' , 6 , ... % Number of bits per symbol
                           ... % (2, 4, 6, 8 for
                           ... % QPSK, 16-QAM, 64-QAM, 256-QAM)
    'qamNormFactor' , 42 , ... % Normalization factor because of modulation
                           ... % (2, 10, 42, 170 for
                           ... % QPSK, 16-QAM, 64-QAM, 256-QAM)
    'mappingType' , 0 , ... % 0: Hard demapping, 1: Soft demapping
    ... % OFDM
    'nSubcarriers' , 2048 , ... % Number of subcarriers
    'nDataSubcarriers' , 1200 , ... % Number of subcarriers containing data
    'guardBand' , 848 , ... % Guard band between subcarriers
    'lengthCP' , 144 , ... % Length of Cyclic Prefix
    ... % Filter
    'filterType' , 1 , ... % 0: Filter every symbol and sum overlap,
                       ... % 1: Filter entire signal
    'rollOff' , 0.25 , ... % Roll of factor (0.2, 0.25, 1)
    'filterSpan' , 10 , ... % Filter span
    ... % Equalizer
    'channelEstimation' , 0 ... % 0: Perfect CSI, 1: Channel estimation
);
```

```
%% The global variable for the experimental parameters
% (fixed point and error insertion)
expr = struct(...
    ... % Fixed point signal
    'fixedPointSignal' , 1 , ... % 0: Fixed point signal exp. off,
                             ... % 1: Fixed point signal exp. on
    'wlSignal' , 5 , ... % Word length for the signal
    ... % Fixed point filter
    'fixedPointFilter' , 1 , ... % 0: Fixed point filter exp. off,
                             ... % 1: Fixed point filter exp. on
    'wlFilterCoeff' , 6 , ... % Word length for the filter coeff.
    ... % Fixed point IFFT
    'fixedPointIfft' , 0 , ... % 0: Fixed point IFFT exp. off,
                           ... % 1: Fixed point IFFT experiments on
    'wlIfft' , 8 , ... % Word length for the IFFT
    ... % Antenna errors (fixed point filter)
    'antennaError' , 1 , ... % 0: Antenna error exp. off,
                         ... % 1: Antenna error experiments on
    'nAntennaErrors' , 2 , ... % Number of antennas with errors
    'errorLevel' , 3 , ... % 0: No error case, 1: Slight error case,
                       ... % 2: 1/4 MSB errors, 3: 1/2 MSB errors
    'antennaOutage' , 0  ... % 0: No antenna outage, just errors,
                         ... % 1: Antenna outage, results set to max valu
);

%% Run the massive MIMO downlink function
tic
mainMaMiDL
toc
```

Appendix B

Error rate arrays

Below is the code for the error arrays used in the simulations. Case 0 is the 'Slight error'-case, case 1 is the 'Lots of errors'-case and case 3 is the 'Extreme error'-case. The Least Significant Bit (LSB) is at the first position in the array. The Most Significant Bit (MSB) is at the 13th position in the error array. The bits 14-32 are signed bits. The numbers in the arrays represents the probability that an error occur on that specific bit. In these cases it means that the MSB in case 1 has the probability 0.02% of error, case 2 has the probability 25.4% and case 3 has the probability 0.5%.

```
%% Choose error rate matrix
switch errorLevel
    case 0 % No error case
        errorRateMatrix = ...
            [0 0 0 0 0 0 0 0 0 0 0 0 0 ...
             0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
    case 1 % 2/10000 MSB errors
        errorRateMatrix = ...
            [0 0 0 0 0 0 0 0 0 0 0 0 0.0002 ...
             0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
    case 2 % 1/4 MSB errors
        errorRateMatrix = ...
            [0 0 0 0 0 0 0 0 0 0 0 0.022 0.254 ...
             0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
    case 3 % 1/2 MSB errors
        errorRateMatrix = ...
            [0 0.008 0.04 0.26 0.41 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
             0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0];
    otherwise
        error('Error level not implemented');
end
```

Printed by Tryckeriet i E-huset, Lund 2017

LUND UNIVERSITY

Series of Master's theses
Department of Electrical and Information Technology
LU/LTH-EIT 2017-557

http://www.eit.lth.se