The limits of MIMO with Large Antenna Arrays

Pedro Valério Catarino Miguel

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisors

Prof. António José Castelo Branco Rodrigues

Prof. Francisco António Taveira Branco Nunes Monteiro

Examination Committee

Chairperson: Prof. Fernando Duarte Nunes

Supervisor: Prof. António José Castelo Branco Rodrigues

Member of Committee: Prof. Francisco Alberto Sena da Silva

October 2014

Abstract

This thesis engages different detection methods for MIMO systems, in an attempt to test and expand the limits of MIMO antenna arrays.

This work begins by studying the concepts of diversity and multiplexing which are embedded in MIMO transmissions. Lattices will then be introduced as a helpful tool for optimizing the performance of several detectors.

One of the concerns of this dissertation was to look for efficient detection algorithms that could allow decent results, while keeping a low complexity so that it could be possible to achieve large antenna arrays. With this in mind, this work presents the most prominent detection techniques for MIMO, starting with the linear receivers Zero-Forcing and MMSE, and then a Successive interference cancelation and lattice reduction-aided algorithms.

An alternative that requires channel state information at the receiver was studied, noting that this configuration allowed pre-processing to be done on the channel matrix, leading to advantageous communication conditions.

Finally a study is made on the properties of large MIMO transmission channels, realizing several interesting facts that are a key motivation for the scalability of MIMO systems.

Keywords

MIMO detection, lattices, Zero-Forcing, MMSE, Successive interference cancelation, lattice reduction

Resumo

Esta tese explora diferentes abordagens de detecção de sistemas MIMO com o intuito de testar e dentro dos possíveis expandir a dimensão dos agregados de antenas utilizados.

Numa primeira fase estudaram-se os conceitos de diversidade e multiplexagem que estão presentes em sistemas MIMO, para que posteriormente fossem identificadas oportunidades de expansão e melhoria dos detectores mais tradicionais. De seguida aprofundou-se o desenho conceptual de lattices que são uma ferramenta extremamente útil e eficaz na redução da complexidade dos receptores, aumentando a performance dos detectores.

Tendo como objectivo principal a obtenção de sistemas MIMO a funcionar para grandes agregados de antenas, houve sempre uma preocupação de procurar iniciativas com bons resultados mas que ao mesmo tempo não fossem demasiado exigentes computacionalmente, para que os seus algoritmos possam ser escalados para grandes dimensões. Com isto em mente testaram-se várias sinergias com os detectores lineares mais conhecidos, Zero-Forcing e MMSE, começando por testar uma abordagem de cancelamento sucessivo de interferência (SIC) e posteriormente uma alternativa com a redução de lattices (LR).

Uma alternativa que pressupõe o conhecimento prévio das condições do canal de transmissão é abordada e estudada, realçando que esta abordagem proporciona condições vantajosas de comunicação através da decomposição do canal e criando canais de comunicação paralelos.

Por último é feito um estudo às propriedades de canais de propagação de grandes dimensões, observando acontecimentos interessantes, que são um factor motivacional para a escalabilidade dos sistemas MIMO.

Palavras-chave

Detecção MIMO, lattices, Zero-Forcing, MMSE, cancelamento sucessivo de interferência, redução de lattices

Acknowledgements

I thank Professor António Rodrigues for the opportunity of working in this thesis.

I would also like to thank all the support given by Professor Francisco Monteiro throughout this year, which made this dissertation possible.

I am thankful to Flávio Brás and Filipe Ferreira for the initial collaboration on the transversal work.

I would also like to thank my family for supporting me whenever I needed.

Contents

List of Figures

List of Tables

Acronyms

3G - Third-generation

4G - Fourth-generation

AWGN - Additive white Gaussian noise

BS - Base station

CDMA - Code Division Multiple Access

CLLL - Complex LLL

CSI - Channel state information

CSIR - Channel state information at the receiver

CVP - Closest vector problem

DL - Downlink

FDMA – Frequency division multiple access

GSM – Global system for mobile communications

ICT - Information and communications technology

IEEE – Institute of Electrical and Electronic Engineers

Iid – Independent and identically distributed

ITU - International Telecommunications Union

LLL - Lenstra-Lenstra-Lovász

LRA - Lattice reduction-aided

LTE - Long Term Evolution

MF - Matched filter

MIMO - Multiple-Input Multiple-Output

MISO - Multiple-Input Single-Output

ML - Maximum likelihood

MLD - Maximum Likelihood Detection

MMSE - Minimum mean square error

M-QAM – (M-ary) quadrature amplitude modulation

MRC - Maximal ratio combining

NP – Non-deterministic polynomial time

OFDM - Orthogonal frequency division multiplexing

OSIC - Ordered successive interference cancelation

QAM - Quadrature amplitude modulation

QPSK - Quadrature phase shift keying

SER - Symbol error rate

SIC - Successive interference cancellation

SIMO - Single-Input Multiple-Output

SISO – Single-Input Single-Output

SNR – Signal to noise ratio

SVD - Singular value decomposition

TDMA - Time division multiplexing

UL - Uplink

WiMAX – Worldwide interoperability for microwave access

ZF - Zero-forcing

ZMSW - Zero-mean spatially white

1 Introduction to MIMO communications

This chapter presents an overview on telecommunications and introduces some basic concepts of MIMO communications.

1.1 Telecommunications progresses

The field of information and communications technology (ICT) is facing a paradigm shift where portable devices and mobile Internet are steadily growing over personal computers and wired Internet services.

Cellular technology started migrating towards data and Internet services with the introduction of third-generation (3G) and fourth-generation (4G) services, allowing new anytime/anywhere computing and multimedia applications that go from simple internet browsing up to mobile video streaming.

According to International Telecommunications Union (ITU), the mobile market will have over nine billion subscriptions in 2020, representing 113% of the projected world population, but with currently seven billion mobile-cellular telephone subscriptions it is now reaching a saturation point as it represents 95,5% of the world's population. Even though the forecast points towards a stalling point, the telephone now has a surprisingly big penetration on the world's population and has become a necessary tool for everyone's life (figure 1).

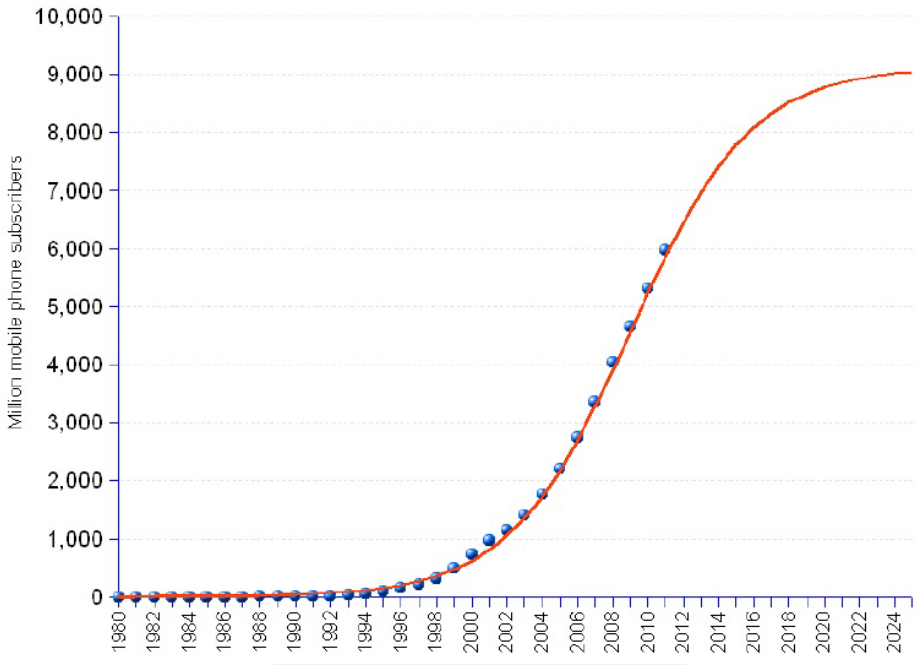


Figure 1 – ITU forecast on mobile phone market 2012-2015 (graph taken form in [1]).

This combined with cellular applications that demand more and more bandwidth and data rates has led to a rapid growth on mobile data services.

To demonstrate the quick evolution of the wireless technology in the recent past, we can observe the figure 2 and realize that in ten years there has been an expansion of new radio equipment at the hardware platform level, going from 2G to 4G including WiFi, Bluetooth, open mobile handsets, software-defined radio, and most recently, open virtualized access points and base stations. At the radio physical layer, the cellular radio link speed has increased from about 2 Mb/s with early 3G systems in the year 2000 to 100 Mb/s with 4G (LTE and WiMax) systems using multiple-input–multiple-output (MIMO) radio technology which is the main focus in this work.

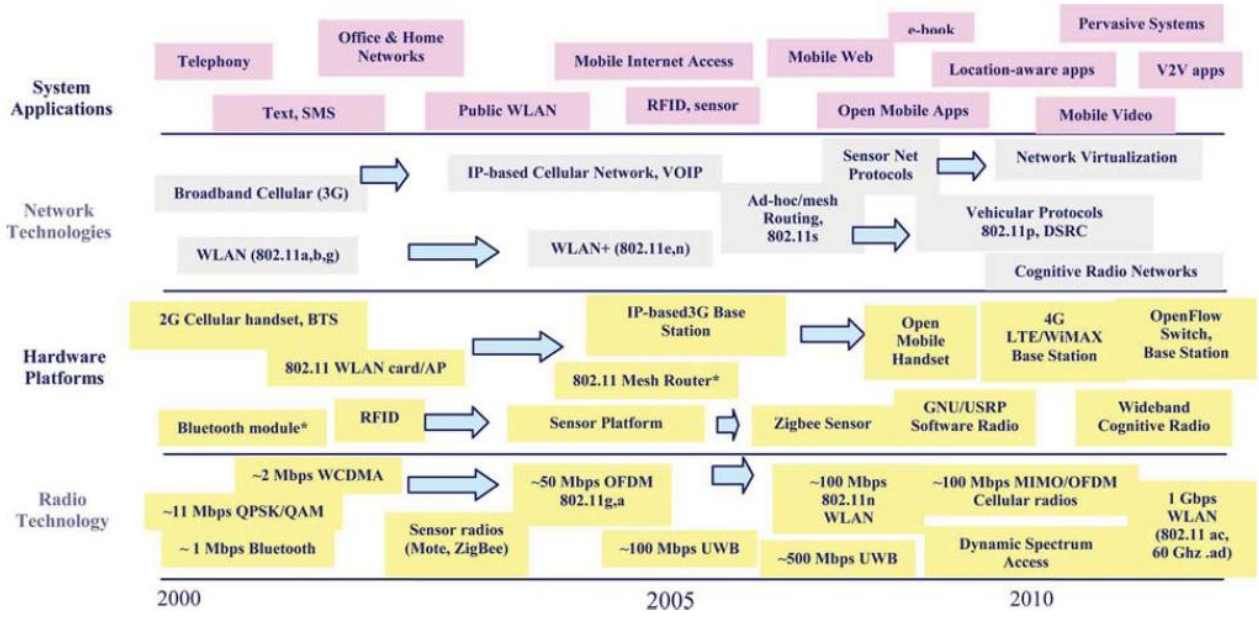


Figure 2 - Wireless Technology roadmap (taken from [2]).

A key player in the networking industry [3] predicts that mobile generated traffic will exceed the traffic of fixed personal computers by 2015, as a result of the migration of most ICT services to mobile devices in the recent years.

With all these trends, large-scale delivery of internet applications on mobile devices will require faster radio access bit rates, improved spectrum efficiency, higher access system capacity and improved security on the open radio medium, among other needs. Most of the research and development on the wireless field today is focused on addressing these aspects, and this thesis will briefly discuss some of those.

One of the problems on the migration into mobility is that mobile phones have limited batteries, which forces all the wireless communications technologies to be computationally

efficient in order to minimize the power consumption. Another challenge is the fact that cellular networks keep getting populated with all the emerging smartphones, burdening the networks with heavy and complex data resulting in the need to distribute bandwidth efficiently.

In order to tackle these problems, multiple access techniques were developed where users shared resources while trying to minimize interference between them (although interference management is undergoing a major paradigm shift with the advent of physical layer network coding and full-duplex based in advanced interference suppression).

Time division multiplexing (TDMA) was used in GSM, allowing the users to share the same frequency, dividing the signal into multiple time-slots. Another multiplexing option is dividing the signal in the frequency domain, giving each user a unique frequency (FDMA).

In CDMA (code division multiple access) the information signals of different users are modulated by orthogonal or non-orthogonal spreading codes. The resulting spread signals simultaneously occupy the same time and bandwidth, and the receiver uses the spreading code structure to separate out the different users.

All these multiplexing strategies described above have evolved into the spatial multiplexing, allowing increased spectral efficiency, and MIMO (multiple-input multiple-output) had an important role in this improvement.

MIMO is a technique that uses multiple antennas at the transmitter and multiple antennas at the receiver, and these multiple antennas at either the transmitter or receiver can be used to achieve diversity, an array power gain, resulting in performance and spectral efficiency gains.

If the MIMO system has access to the channel gain matrix, i.e. through a pilot test, independent signalling paths can be obtained by exploiting the structure of the channel matrix, allowing spatial multiplexing where multiple parallel streams of data are transmitted.

MIMO can then be divided into three categories; precoding, spatial multiplexing and diversity coding.

Precoding in more general terms is all the spatial processing that occurs at the transmitter, and can be looked as multi-stream beam forming. In single stream beam forming, each transmit antenna emits the same signal with appropriate phase and gain weighting such that the signal power is maximized at the receiver input. If the signals emitted from different

antennas add up constructively, the multipath fading effect is reduced leading to an increased received signal gain. In line-of-sight propagation, beam forming results in a well-defined directional pattern. However, conventional beams are not a good analogy in cellular networks, which are mainly characterized by multipath propagation. When the receiver has multiple antennas, the transmit beam forming cannot simultaneously maximize the signal level at all of the receive antennas, and precoding with multiple streams is often beneficial. However, precoding requires knowledge of channel state information (CSI) at the transmitter and the receiver, which might limit its usability.

Spatial multiplexing requires a multiple-antenna setup. In spatial multiplexing, a high-rate signal is split into multiple lower-rate streams and each stream is transmitted from a different transmit antenna in the same frequency. If these signals arrive at the receiver antenna array with sufficiently different spatial signatures and the receiver has accurate CSI, it can separate these streams into (almost) parallel channels. Spatial multiplexing is a very powerful technique for increasing channel capacity at higher signal-to-noise ratios (SNR). The maximum number of spatial streams is limited by the smaller of the number of antennas at either the transmitter or the receiver. Spatial multiplexing can be used without CSI at the transmitter, but can be combined with precoding if CSI is available. Spatial multiplexing can also be used for simultaneous transmission to multiple receivers, known as space-division multiple access or multi-user MIMO, or also as the broadcast channel (in the information theory community), in which case CSI is required at the transmitter.

Diversity coding techniques are used when there is no channel knowledge at the transmitter. In diversity methods, a single stream (unlike multiple streams in spatial multiplexing) is transmitted, but the signal is coded using techniques called space-time coding. The signal is emitted from each of the transmit antennas with full or near orthogonal coding. Diversity coding exploits the independent fading in the multiple antenna links to enhance signal diversity. Because there is no channel knowledge, there is no beam forming or array gain from diversity coding. Diversity coding can be combined with spatial multiplexing when some channel knowledge is available at the transmitter.

In this dissertation, space-time codes and diversity coding will not be further described. Note that nowadays, the emphasis is put on the spectral efficiency at the “MIMO physical layer”, as diversity can be provided by other techniques such as the modulation itself (e.g., OFDM).

A few issues arise in communication over fading channels. Fading occurs due to

variations in the magnitude and/or phase of one or more frequency components in time, caused by the time-varying characteristics and time-dispersive nature of the propagation environment. It can cause a loss of signal power without reducing the power of the noise resulting in a decreased performance in a communication system.

This signal loss can be over some or all of the signal bandwidth. Fading can also be a problem as it changes over time: communication systems are often designed to adapt to such impairments, but the fading can change faster than the adaptations can be made, which is usually a concerning problem in real time communications.

Fading can be divided into large and small scale signalling fading. The first is mostly caused by environmental elements like shadowing by buildings and natural features or rain, meaning that in order to overcome large fading, solutions tend to be at the system protocol level e.g. base station positioning, power control, etc. This thesis will focus on enhancing performance through signal processing techniques applied at the physical layer to mitigate the effects of small scale fading.

The small scale signalling fading in the receiver antennas is caused by signal scattering off objects in the propagation environment. This scattering leads to a phenomenon known as multipath propagation, where the composite received signal is distorted due to being comprised of a number of delayed copies of the transmitted waveform. These copies are also referred to as multipaths, since they propagate along different paths to reach their destination.

Diversity can be used to minimize the effects of fading, and the diversity techniques presented here operate over time, frequency or space. The main idea is the same, by sending signals that carry the same information through different paths, multiple independently faded replicas of data symbols are obtained at the receiver end and more reliable detection can be achieved.

The more intuitive way to exploit the existence of multipath fading is by replicating the transmission of a signal over multiple antennas. More intelligent schemes exploit channel diversity and, at the same time, efficiently use the degrees of freedom in the channel. Compared to repetition coding, they provide coding gains in addition to diversity gains.

1.2 MIMO in the recent standards

Evolving from the concepts approached in the last section, the use of multiple antennas at both the transmit and receive nodes have become very popular due to the advantages they promise to offer, including high data rates and transmit diversity. It has become one of the most important technologies for the development of existing and emerging wireless communications systems.

The physical layer of the fourth generation (4G) wireless networks is based on these two relatively new concepts; orthogonal frequency division multiplexing (OFDM) and MIMO. Along with larger channel bandwidth, new standards like IEEE 802.16 and Long Term Evolution (LTE) were released.

In its first releases, LTE relied on MIMO mostly for the downlink (i.e., from the base station (BS) to user equipment), using 4 layers in the downlink (DL). In the latest release 10 (known as LTE-Advanced), the role of MIMO also became important in the uplink (UL). Indeed, MIMO is utilised in both uplink and downlink of the IEEE 802.16m standard (WiMAX profile 2.0).

Presently, MIMO is entering the vast domestic market via 802.11n [4], [5], the latest generation of Wi-Fi, designed for a peak rate of 600 Mbps (using 40MHz bandwidth and 4×4 antennas), which is the first commercial product to be based in MIMO and OFDM.

It is known that the capacity of MIMO channels grows linearly with the minimum of the number of antennas on the transmitter and receiver sides [6], and that factor was the base of this thesis, where increasing the number of antennas at the communication terminals now allows high spectral efficiencies and high diversity orders in wireless transmissions.

2 Diversity and Multiplexing

This chapter presents some introductory problems on MIMO systems and studies the conceptual differences between Diversity and Multiplexing.

2.1 MIMO system model

A MIMO system can be modelled by an input vector **x** (with N_T antennas at the transmitter), an output vector y (another vector but with N_R antennas) that is obtained from **x** by means of a linear matrix transformation **H**, which represents the environment channel. In a real system, the detection of y is usually perturbed by a noise vector **n** of the same size. The system can then be resumed to

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

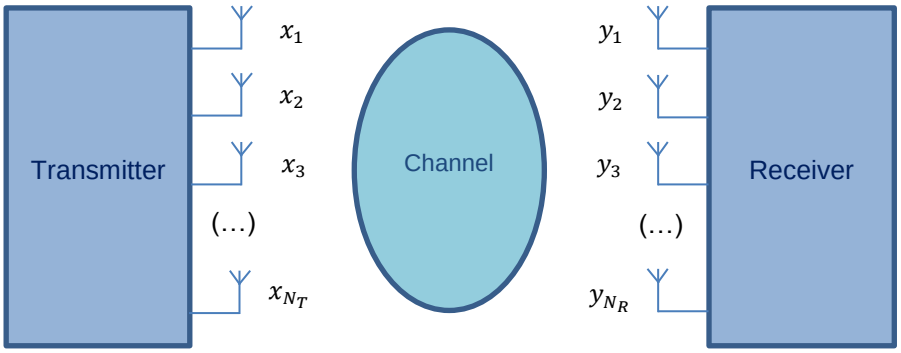


Figure 3 - Wireless MIMO link.

The MIMO scheme illustrated above (figure 3) is the general communication system where N_T and N_R are both bigger than 1. The origin of MIMO relates to the conventional radio system communications with one antenna at the transmitter and one at the receiver (SISO). Some specific cases of MIMO will be analysed in order to note the evolution from SISO to MIMO, including SIMO (Single-Input Multiple-Output) and MISO (Multiple-Input Single-Output).

The main goal is to describe the advantages of increasing the number of antennas at the transmitter, receiver or both, leading to the central question in this dissertation, which is to assess how many antennas can be placed at the while maintaining significant gains with affordable complexity. This study is mostly on a signal processing point of view, meaning that even if a certain number of antennas at the terminals is achieved, electronics might not have yet the technology to create antennas small enough to fit in either a mobile terminal or a

base station. One should also bear in mind that as the spacing between antennas diminishes, they become increasingly correlated and the capacity of the system diminishes [6], but as mentioned this problem is beyond the scope of this thesis.

2.2 Shannon capacity and the advent of MIMO

In 1948 Shannon defined capacity as the mutual information maximized over all possible input distributions [7], and this limit establishes the maximum data rates that can be achieved over wireless channels. One could then have a binary transmission without errors even when in the presence of additive white Gaussian noise (AWGN) (figure 4), as long as the transmission rate was kept below the Shannon's capacity. This meant that a signal could now be detected and regenerated in a wireless transmission.

A discrete-time additive white Gaussian noise channel with channel input/output relationship, at the output of a matched filter, is $y[i] = x[i] + n[i]$, where $x[i]$ is the channel input at time i, $y[i]$ is the corresponding channel output, and $n[i]$ is a white Gaussian noise random process. The channel capacity is then given by Shannon's well-known formula (2), where the channel *SNR* is the power in $x[i]$ divided by the power in $n[i]$

$$C = \log_2(1 + SNR) \tag{2}$$

A Signal to Error Ratio curve of a standard AWGN in a SISO link with 4 QAM was simulated in order to calibrate the standard capacity curve with the literature.

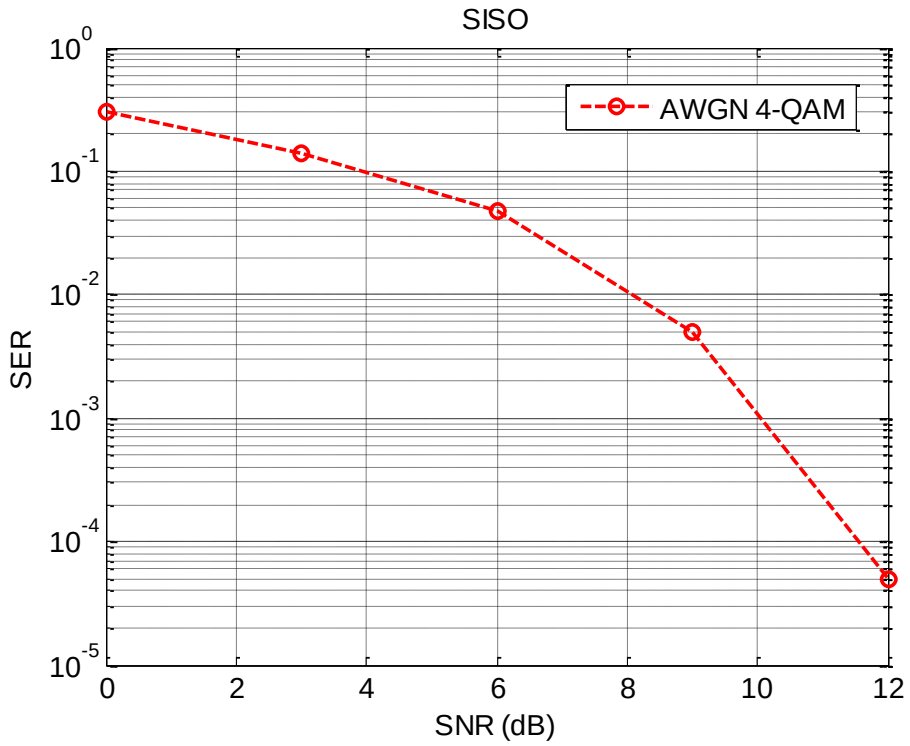


Figure 4 - AWGN SER curve of a SISO model.

Receiver diversity is known to improve the performance of wireless communications in fading channels. The main advantage of receiver diversity is that it mitigates the fluctuations due to fading by averaging them in a way that the channel appears more like an AWGN channel. Antenna diversity, or spatial diversity, can be obtained by placing multiple antennas at the transmitter and/or the receiver. If the antennas are placed sufficiently apart, the channel gains between different antenna pairs fade more or less independently, and independent signal paths are created.

In this case, receive diversity will be analysed with multiple antennas at the receiver, known as single input, multiple output (SIMO) channels. One can observe the diversity gain, as increasing the number of antennas at the receiver greatly reduces the symbol error rate. When the number of antennas tends to infinity the error curve would tend to the one of the simple AWGN case where only the noise impairs the reception. This is because the power of the wireless link is one, modelled by the complex coefficient $\mathbf{H}$, whose average is one, i.e., $\mathrm{E}[\mathbf{H} = 1]$.

A fading channel with one transmit antenna and m receive antennas can be modelled as follows (figure 5):

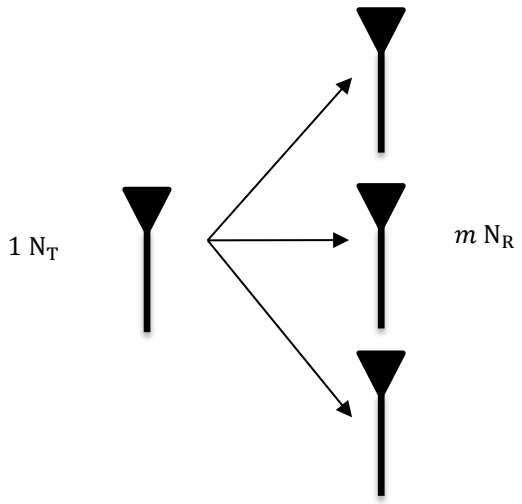


Figure 5 - Single Input Multiple Output antenna configuration.

$$\mathbf{y}_m = \mathbf{H}_m\mathbf{x} + \mathbf{n}_m \tag{3}$$

Where $\mathbf{y} = [y_1, \dots, y_M]^{\mathrm{T}} \in \mathrm{C}^{M\times 1}$ is the received signals vector and $\mathbf{x}$ is the transmitted signal.

$\mathbf{H} \in \mathrm{C}^{M\times 1}$ is the channel matrix representing the links between the transmit and the receive antennas. The noise vector $\mathbf{n}_{\mathrm{m}} = [\mathrm{n}_1, \ldots, \mathrm{n}_M]^{\mathrm{T}} \in \mathrm{C}^{M\times 1}$ represents the noise vector that is added to the incoming signal vector. The entries of n are random variables taken from an independent circularly symmetric complex Gaussian with zero average and variance σ_{n}^2 , so that both its real and imaginary components have variance $\sigma_{\mathrm{n}}^2/2$.

If the antennas are spaced sufficiently far apart, then we can assume that the gains $\mathbf{H}_m$ are independent Rayleigh, and we get a diversity gain of d. (4). This parameter describes how the SER decreases with the increase of SNR, i.e., when plotting the SER against the SNR this diversity order is simply the slope of the SER curve and it can be obtained by

$$d = \frac{\log(\mathrm{SER})}{\log(\mathrm{SNR})} \tag{4}$$

The maximal ratio combining (MRC) is the optimum detector of multiple channels, where the output is a weighted sum of all branches (5)

$$\sum_{i=1}^{N_r} h_i^* \, h_i \boldsymbol{x_{mf}} = \sum_{i=1}^{N_r} |h_i|^2 \boldsymbol{x_{mf}} \tag{5}$$

Which are detected by applying a simple matched filter (MF) to the channel (6). Note that the MF detector simply ignores noise

$$\mathbf{x_{mf}} = \mathbf{H^H y} \tag{6}$$

Applying the MRC detector, one can now explore the diversity slope studied before. In this case, we simulated a SIMO system with 1, 2, 4 and 8 antennas at the receptor over 16 QAM, obtaining the following results (figure 6):

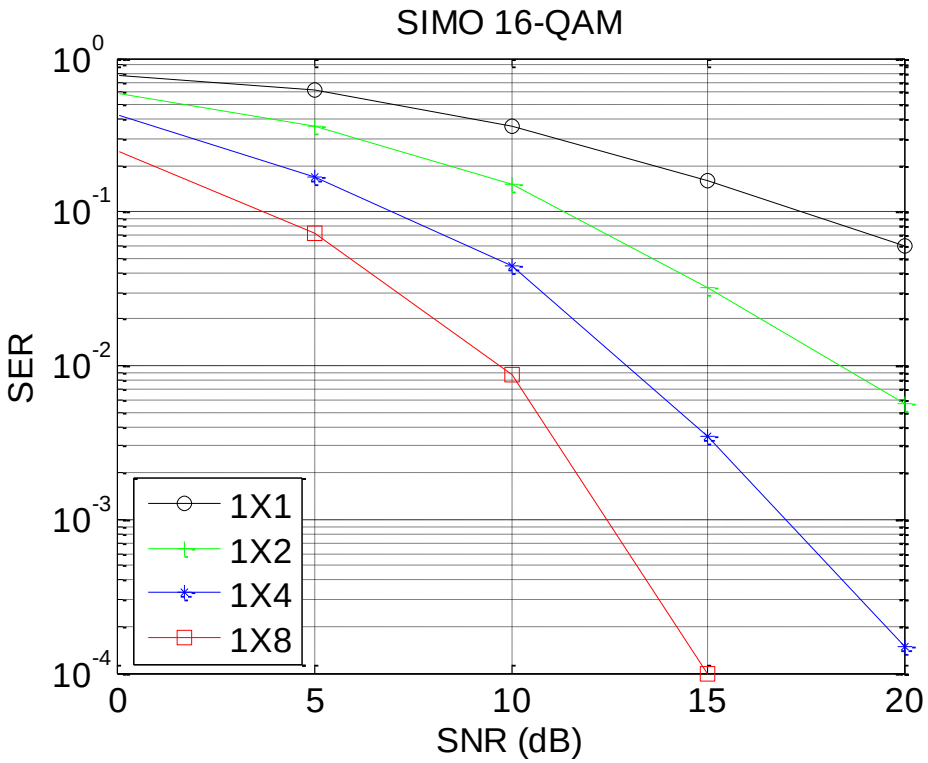


Figure 6 - SIMO curve with 1, 2, 4 and 8 antennas at the receiver.

As we can observe in figure 6, the slope gets more inclined as the number of antennas increases, which is on pair with our predictions.

One should note that the maximal ratio combining is actually an instance of zero-forcing, detection (to be described in chapter 4) when the pseudo-inverse filter is applied.

2.3 The parallelisation of channels via SVD

As we have seen in the previous chapter, having multiple antennas at the transmitter or receiver can be used for diversity gain. When both the transmitter and receiver have multiple antennas, there is another mechanism for performance gain called multiplexing gain.

In figure 7, one sees a source with data sequence 110 to be sent over a MIMO system with three transmitters. In the diversity form of MIMO, the same data, 110 is sent over three different transmitters. If each path is subject to different fading then the likelihood is high that one of these paths will lead to successful reception. This system has a diversity gain of three. The second form uses spatial-multiplexing techniques. In a diversity system, we send the same data over each path. To demonstrate diversity and multiplexing we will analyse a pseudo data transmission. Let's suppose we multiplex the data 1,1,0 over three channels. Each channel carries different data, similar to the idea of an OFDM signal. By multiplexing the data we have increased the data throughput or the capacity of the channel, but we have lost the diversity gain. The multiplexing has tripled the data rate, so the multiplexing gain is 3 but diversity gain is now 0. Whereas in a diversity system the gain comes in form of increased reliability, here the gain comes in form of increased data rate. In this chapter we will represent the multiplexing gain by r.

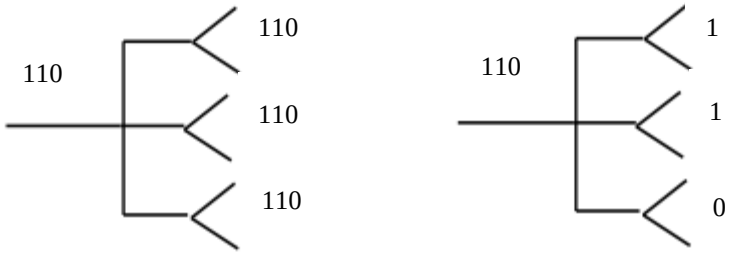


Figure 7 - MIMO Diversity system and MIMO multiplexing system.

For each r the optimal diversity gain $d_{opt}(r)$ is the maximum the diversity gain that can be achieved by any scheme. It is shown in [8] that if the fading block length exceeds the total number of antennas at the transmitter and receiver, then

$$d_{opt}(r) = (M_t - r)(M_r - r), 0 \leq r \leq min(M_t, M_r) \quad (7)$$

In this section, the number of antennas is noted as M_t and M_r which is the notation adopted in several textbooks (e.g., Goldsmith (2003)), however, the notation used throughout this dissertation for the number of antennas at the terminals is N_T and N_R for the transmitter and receiver respectively.

The formula (7) implies that in a MIMO system, if we use all transmit and receive antennas for diversity, we get an error probability proportional to $\mathrm{SNR}^{-M_t M_r}$ and that, moreover, we can use some of these antennas to increase data rate at the expense of diversity gain. To better understand this Diversity versus Multiplexing trade off, the function (7) was plotted into figure 8.

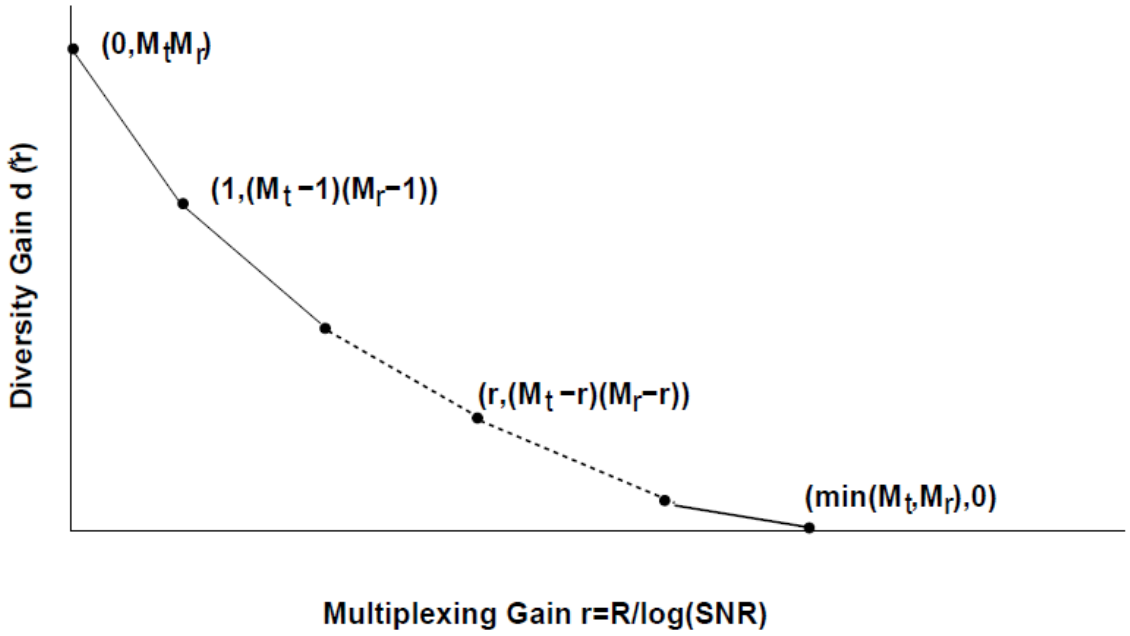


Figure 8 - Diversity-Multiplexing Trade-off for High SNR Block Fading (taken from [8]).

One can then design systems to maximise one or the other: maximising diversity leads to space-time codes; maximising multiplexing gain leads to spatial multiplexing systems (that is the focus of this thesis). Multiplexing is nowadays more important than diversity because there is plenty of diversity being introduced by other techniques such as OFDM.

In this chapter, we consider a MIMO system where CSI is fully available both at the transmitter and the receiver. It is known that precoding techniques can provide large performance improvements in such scenarios. A popular precoding approach is based on

singular value decomposition (SVD) of the channel so that the MIMO channel can be seen as parallel channels.

In slow fading scenarios, channels are subject to block fading. Without rate and power adaptation, outages cannot be avoided. In such scenarios, a popular measure of reliability is the diversity order achieved by a given transmit-receive scheme. We consider SVD precoding for MIMO systems, which transforms the MIMO channels into parallel sub channels.

Let us consider a MIMO channel with $M_r \times M_t$ channel gain matrix $\mathbf{H}$ known to both the transmitter and the receiver. Let R_H denote the rank of $\mathbf{H}$. From matrix theory, for any matrix $\mathbf{H}$ we can obtain its SVD as:

$$\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^{\mathbf{H}}, \tag{8}$$

where the $M_r \times M_t$ matrix $\mathbf{U}$ and the $M_t \times M_t$matrix $\mathbf{V}$ are unitary matrices and Σ is an $M_r \times M_t$ diagonal matrix of singular values $\{\sigma_i\}$ of $\mathbf{H}$. These singular values have the property that $\sigma_i = \sqrt{\lambda_i}$ for λ_i the i^{th} eigenvalue of $\mathbf{H}\mathbf{H}^{\mathbf{H}}$, and R_H of these singular values are nonzero. Since R_H cannot exceed the number of columns or rows of $\mathbf{H}$, $R_H \leq \min(M_r \times M_t)$. If $\mathbf{H}$ is full rank, which is sometimes referred to as a rich scattering environment, then $R_H = \min(M_r \times M_t)$. Other environments may lead to a low rank $\mathbf{H}$: a channel with high correlation among the gains in $\mathbf{H}$ may have rank 1.

The transmit precoding and receiver shaping transform the MIMO channel into R_H parallel single-input single-output channels with input $\tilde{\mathbf{x}}$ and output $\tilde{\mathbf{y}}$, since from the SVD, we have that:

$$\begin{aligned}\tilde{\mathbf{y}} &= \mathbf{U}^{\mathrm{H}}(\mathbf{H}\mathbf{x} + \mathbf{n}) \\ &= \mathbf{U}^{\mathrm{H}}(\mathbf{U}\Sigma\mathbf{V}\mathbf{x} + \mathbf{n}) \\ &= \mathbf{U}^{\mathrm{H}}(\mathbf{U}\Sigma\mathbf{V}\mathbf{V}^{\mathrm{H}}\tilde{\mathbf{x}} + \mathbf{n}) \\ &= \mathbf{U}^{\mathrm{H}}\mathbf{U}\Sigma\mathbf{V}\mathbf{V}^{\mathrm{H}}\tilde{\mathbf{x}} + \mathbf{U}^{\mathrm{H}}\mathbf{n} \\ &= \Sigma\tilde{\mathbf{x}} + \tilde{\mathbf{n}},\end{aligned} \tag{9}$$

where $\tilde{\mathbf{n}} = \mathbf{U}^{\mathbf{H}}\mathbf{n}$ and $\boldsymbol{\Sigma}$ is the diagonal matrix of singular values of $\mathbf{H}$ with σ_i on the i^{th} diagonal. Note that multiplication by a unitary matrix does not change the distribution of the noise, i.e. $\mathbf{n}$ and $\tilde{\mathbf{n}}$ are identically distributed. Thus, the transmit precoding and receiver shaping transform the MIMO channel into $\mathbf{R_H}$ parallel independent channels where the i^{th} channel has input $\tilde{x}_i$, output $\tilde{y}_i$, noise $\tilde{n}_i$, and channel gain σ_i. Note that each σ_i are related

since they are all functions of **H**, but since the resulting parallel channels do not interfere with each other, we say that the channels with these gains are independent, linked only through the total power constraint.

One known technique to take advantage of the channel knowledge, beyond the SVD channel parallelisation is the waterfilling algorithm.

Considering the power of the symbol transmitted of the i^{th} parallel channel by P_i, the capacity is

$$C = \max_{\boldsymbol{P}:\sum_{j=1}^{N_r} P_j \leq 1} \sum_{i=1}^{N_r} \log(1 + \rho\, \lambda_i P_i) \tag{10}$$

With $\boldsymbol{P} = [P_1, P_2, \dots, P_{N_R}]$. This capacity can only be obtained if the inputs are independent complex Gaussian.

The optimization problem then presents us with the following known solution:

$$P_{i,opt} = \left(\mu - \frac{1}{\rho\lambda_i}\right)^+ \tag{11}$$

where $(z)^+ = \max[z, 0]$ and μ is the solution of

$$\sum_{i=1}^{N_r} \left(\mu - \frac{1}{\rho\lambda_i}\right)^+ = 1, \tag{12}$$

and the resulting capacity is:

$$C = \sum_{i=1}^{N_r} (\log(\mu\rho\lambda_i))^+ \tag{13}$$

This algorithm distributes the power according to the strongest channels obtained through the SVD process, in a way that a channel pipe with a corresponding higher Eigen value will be given more power than a less efficient pipe. This results in a better and more efficient power allocation, leading to an increase in capacity.

In figure 9 we can observe the capacity of a precoding scheme, where a plain 4x4 MIMO system is simulated again through a MF decoder, but this time, knowing the channel matrix **H** beforehand, we can precode it to obtain the channel pipes, and then optimally distribute the power.

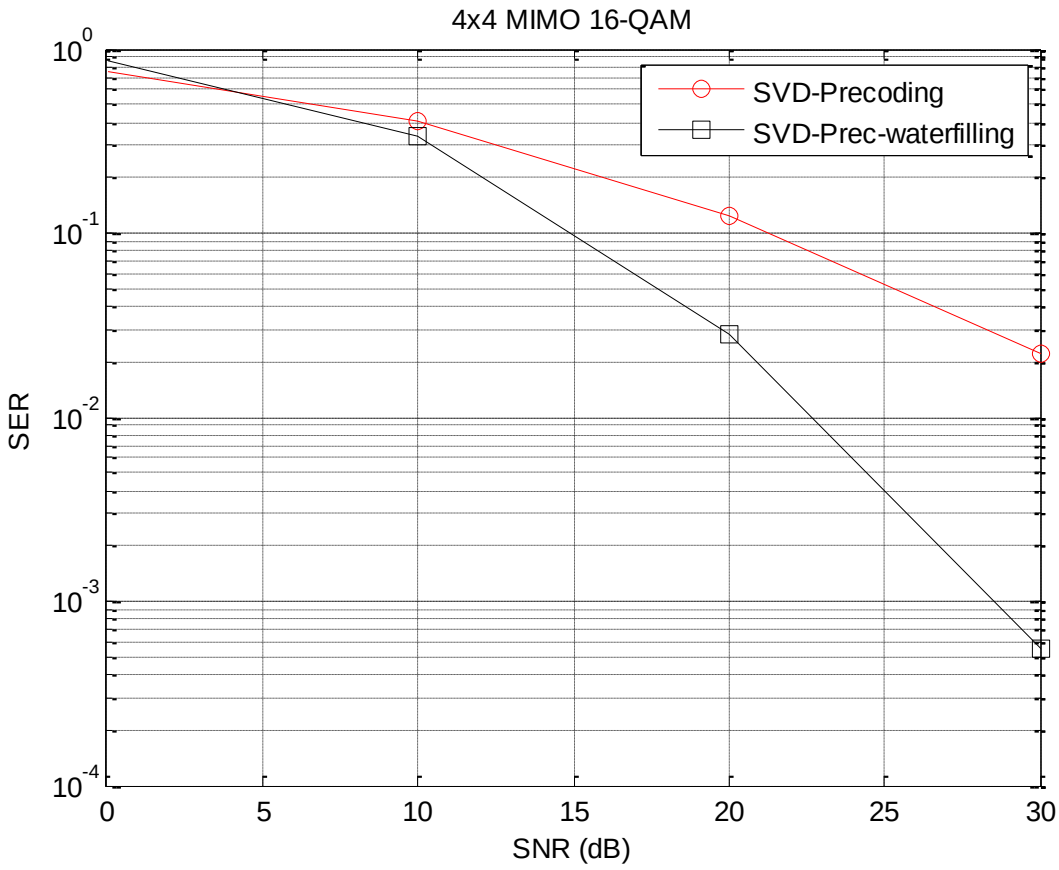


Figure 9 - Detection 4x4 antennas with 16-QAM using SVD precoding and SVD precoding with water filling optimization.

3 Lattices

This chapter introduces the concept and fundamentals of lattices to posteriorly relate them with MIMO.

3.1 Context

A lattice is a discrete set of vectors in real Euclidean N-space forming a regular group of points through ordinary vector addition. In a lattice, the sum or difference of any two vectors in the lattice is also in the lattice. A simple lattice is illustrated on figure 10.

The concept of lattice was introduced in the 18^{th} century by famous mathematicians such as Gauss, Lagrange and Minkowski, but it was the latter that had a major role in making lattices a standard and helpful tool in number theory through its publication "Geometrie der Zahlen in 1896". Lattices are related with problems in the integer domain, such as: continued fractions, simultaneous Diophantine equations, and several other fundamental problems in number theory, and in integer programming. However, in the early 1980s algorithmic theory of lattices had a major development with the introduction of new lattice reduction techniques such as the Lenstra-Lenstra-Lovász (LLL) that has found several uses in both pure and applied mathematics, while creating new implications in many problems.

Lattices still remain a useful tool nowadays. In communication theory point-to-point transmission problems are evolving into network coding ideas, opening a few windows for lattices. In computer science and mathematics, lattices are also rising as their capacity to solve integer programming problems is helpful in many challenges. Despite its apparent simplicity, lattices are related to many of the most difficult algorithmic problems with NP-hard complexity, allowing lattices to assume an important role in cryptography as a source of computational hardness for the design of new barriers to an attacker.

Lattices can also be used as a base for coding schemes and that application will be explored here. In the present work lattices take a prominent place as it will become clearer in the next chapters so it is important to define its structure and understand its main properties.

Most of the properties and definitions in this chapter were taken from [9] [10]

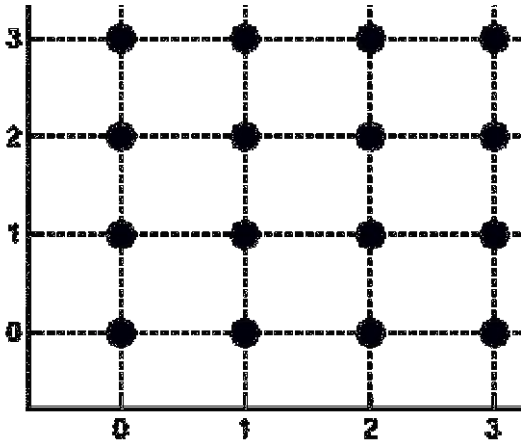


Figure 10 - Lattice illustration in $\mathbb{R}^2$.

3.2 Basic definitions

A lattice can be defined as a subset $\Lambda \subseteq \mathbb{R}^n$ closed under addition and subtraction, and discrete, meaning that there is an $\epsilon > 0$ such that any two distinct lattice points $x \neq y \in \Lambda$ are at distance at least $\|x - y\| \geq \epsilon$.

One subgroup of $\mathbb{R}^n$ that is a lattice is $\mathbb{Z}^n$ because integer vectors can be added and subtracted remaining an integer vector and the distance between any two integer vectors is at least one.

This $\mathbb{Z}^n$ lattice approach will be commonly used in this thesis and so it is important to understand this definition as it is assumed throughout the work.

A lattice is the span of a finite set of vectors in a Euclidean space:

$$\Lambda = L(\mathbf{H}) = \{\mathbf{Hx} \mid \mathbf{x} \in \mathbb{Z}^k\}. \tag{14}$$

Notice the similarity between the definitions of a lattice and the span of a set of vectors **B**:

$$\text{span}(\mathbf{H}) = \{\mathbf{Hy} \mid \mathbf{y} \in \mathbb{R}^k\}. \tag{15}$$

The central difference is that in a lattice only integer coefficients are allowed, resulting in a discrete set of points. As vectors $\mathbf{h_1}, \ldots, \mathbf{h_n}$ are linear independent, any point $y \in \text{span}(\mathbf{H})$ can be written as a linear combination $\mathbf{y} = \mathbf{x_1 h_1} + \cdots + \mathbf{x_n h_n}$ in a unique way (although different sets of vectros, i.e., a basis, can generate the same lattice). Therefore $y \in L(\mathbf{H})$ if and

only if $\mathbf{x_1}, \ldots, \mathbf{x_n} \in \mathbb{Z}$.

Notice that the definition $\mathrm{L}(\mathbf{H}) = \{\mathbf{Hx} | \ \mathbf{x} \in \mathbb{Z}^m\}$ can be extended to matrices **H** whose columns are not linearly independent. However, in this case, the resulting set of points is not always a lattice because it may not be discrete.

3.2.1 Fundamental Region

The fundamental region of a lattice basis **H** is defined as

$$\mathrm{P}(\mathbf{H}) = \{\mathbf{H}x \mid \mathbf{x} \in \mathrm{R}^n, \forall i: 0 \leq x_i < 1\}. \tag{16}$$

The fundamental region must not contain any lattice point inside of it. It could not be represented by linear combination of integer vectors if it had one point inside, where the generating vectors are the limits of the fundamental region. In figure 11 it is possible to observe the lattice and the fundamental region defined by the two generating vectors.

One shall keep in mind that a lattice may be generated by different sets of vectors, and that there are infinite admissible bases for a lattice, which can be inferred from figure 11, as it is possible to select points further distant from the origin to replace a generator vector and still have a fundamental region without any lattice point in it.

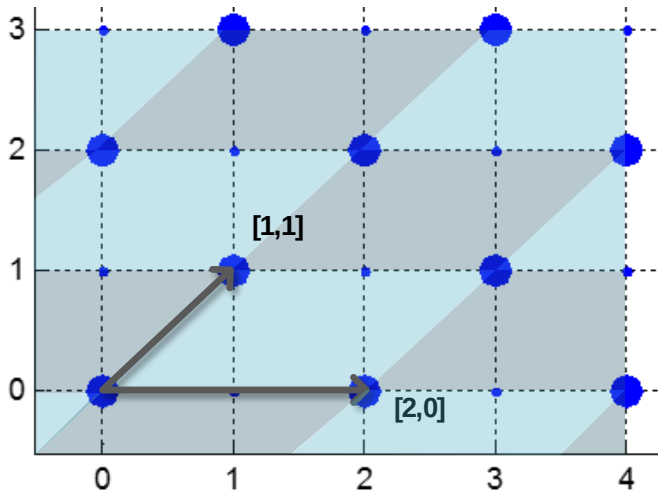


Figure 11 - Span of $\mathrm{L}(\mathbf{H})$ with the fundamental region $P(\mathbf{H})$.

3.2.2 Voronoi Region

The Voronoi region is the area of the lattice that contains all the points in the span of the lattice which are closer to a given lattice point *x* than to any other point in the lattice, and

is defined by:

$$\nu(\Lambda) = \{z \in \text{span}(\mathbf{H}): \|x - z\| < \|y - z\|, \forall y \in \Lambda\}. \tag{17}$$

This region is a characteristic of the lattice and independent of any particular generating matrix and it is the most interesting fundamental region that constitutes the optimal decision region for the closest vector problem in lattices.

Figure 12 shows the Voronoi region of a lattice. One can denote on the figure that the Voronoi regions are limited by the blue lines.

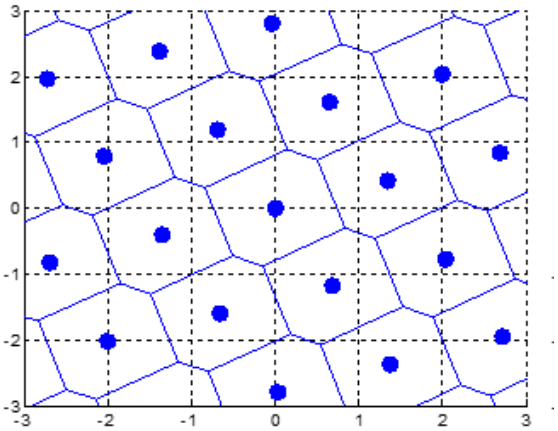


Figure 12 - Illustration of Voronoi regions of a lattice.

3.2.3 Volume

When the channel matrix **H** is non-singular we are in the presence of full-rank lattices. The volume of the lattice (which is independent of the fundamental region) is

$$\mathrm{vol}(\Lambda) = |\det(\mathbf{H})|\,, \tag{18}$$

however, in general the following expression is required:

$$\mathrm{vol}(\Lambda) = \sqrt{\det(\mathbf{H}^H\mathbf{H})} = \sqrt{\det(\mathbf{G})} \tag{19}$$

3.2.4 Shortest Vector and successive minima

The length of a lattice is the Euclidean norm of the shortest nonzero vector in the lattice, and is usually referred to as λ_1.

$\lambda_i$is then the i^{th} successive minimum of a lattice if $\lambda_i$is defined as the smallest real number within the radius of a sphere containing i pairwise independent vectors with norms smaller or equal to i. It is now easy to understand that $\lambda_1$is the shortest vector and has norm 1.

3.3 Lattice Reduction

Lattice reduction is a powerful technique that has several uses and lead to a breakthrough of new and appellative results in the MIMO context, allowing the resolution of problems like finding the shortest vector.

Lattice reduction consists in changing a basis **H** of a lattice Λ into a shorter basis $\mathbf{H}_{\mathrm{red}}$ so that Λ remains the same and it can be implemented by algorithms such as LLL and KZ being the former the most known. The LLL lattice basis reduction algorithm was invented by Arjen Lenstra, Hendrik Lenstra and László Lovász in 1982 [11] and is a polynomial time lattice reduction algorithm.

LLL usually obtains an approximation for the shortest vector but there is not yet an efficient algorithm to solve this problem. However the approximation obtained by the LLL allows many application uses. In this thesis a complex version of the original LLL algorithm (CLLL) will be used in some of the receivers presented on the next chapter.

To obtain a better lattice basis, the LLL algorithm performs successive orthogonal projections and if needed, swaps two consecutive vectors of **H** in an attempt to end up with a near orthogonal basis $\mathbf{H}_{\mathrm{red}}$. The output of the algorithm is a new reduced basis $\mathbf{H}_{\mathrm{red}}$ that has near-orthogonal vectors and a unimodular matrix **M** that follows

$$\mathbf{H}_{\mathrm{red}} = \mathbf{HM}. \tag{20}$$

In the next chapter most receivers will be tested with and without lattice reduction and it will be possible to denote the advantages of this tool. In table 1 we have the pseudo-code of the algorithm implemented. The parameter δ manages the performance and complexity of the algorithm, meaning that higher values of δ lead to a better performance due to higher complexity. At the simulations performed along this dissertation δ was assumed to be 0.75 as it is commonly recommended in the literature as a good trade-off between orthogonality performance and complexity. It is simple to observe now a direct relation between a real lattice basis **H** and a real MIMO channel matrix **H**.

Table 1. Matlab pseudo-code of the complex LLL algorithm

1: $[\mathbf{Q}, \mathbf{R}] =$ QR decomposition ($\mathbf{H}$)
2: $\delta = 0{,}75$
3: $m = \text{size}(\mathbf{H}, 2)$
4: $\mathbf{M} = \mathbf{I}_{\text{m}}$
5: $k = 2$
6: while $k \leq m$
7: for $n = k - 1: -1: 1$
8: $u = \text{round}((\mathbf{R}(n,k)/\mathbf{R}(n,n))$
9: if $u \sim = 0$
10: $\mathbf{R}(1:n,k) = \mathbf{R}(1:n,k) - u \cdot \mathbf{R}(1:n,n)$
11: $\mathbf{M}(:,k) = \mathbf{M}(:,k) - u \cdot \mathbf{M}(:,n)$
12: end
13: end
14: if $\delta|\mathbf{R}(k-1,k-1)|^2 > |\mathbf{R}(k,k)|^2 + |\mathbf{R}(k-1,k)|^2$
15: swap the $(k-1)^{\text{th}}$ and k^{th} columns in $\mathbf{R}$ and $\mathbf{M}$
16: $\mathbf{\Theta} = \begin{bmatrix} \alpha^* & \beta \\ -\beta & \alpha \end{bmatrix}$ where $\alpha = \frac{\mathbf{R}(k-1,k-1)}{\|\mathbf{R}(k-1:k,k-1)\|}$ and $\beta = \frac{\mathbf{R}(k,k-1)}{\|\mathbf{R}(k-1:k,k-1)\|}$
17: $\mathbf{R}(k-1:k, k-1:m) = \mathbf{\Theta}\mathbf{R}(k-1:k, k-1:m)$
18: $\mathbf{Q}(:,k-1,k) = \mathbf{Q}(:,k-1:k)\mathbf{\Theta}^{\mathcal{H}}$
19: $k = \max(k-1,2)$
20: else
21: $k = k + 1$
22: end
23: end
24: $\mathbf{H}_{\text{red}} = \mathbf{QR}$

4 MIMO detection

In this chapter several low-complexity receivers are presented along with results comparing their performance. A lattice approach will be studied in other to achieve better results.

4.1 Fundamentals of Spatial Multiplexing

As described in the second chapter, the main idea behind spatial multiplexing is that signals sampled in the spatial domain at both ends are combined in order to create multiple parallel spatial data pipes (SISO channels), increasing the data rate, while possibly adding diversity to improve the quality of the communications by reducing the bit-error rate

Spatial multiplexing in the context of MIMO has enabled unprecedented spectral efficiencies in wireless fading channels achieving high date-rates. However this gain of performance comes at the price of increased complexity in the detection of the receivers. This complexity presents a great challenge to the decoder's implementation and is usually where most of the research is available and rising nowadays. In this chapter a background study will be done on some of the receivers that have been implemented so far. The performance of linear receivers will be analysed, as well as the successive interference cancellation (SIC) receivers. A lattice approach will be described in an attempt to expand the perception of the detection problem and to obtain better results.

One of the first examples of practical application of MIMO is the patent of Paulraj and Kailath [12] which introduced a technique for increasing the capacity of a wireless link using multiple antennas at both ends for application to broadcast digital TV.

The MIMO system that will be studied in this chapter is a single user point-to-point communication MIMO, which was previously illustrated in Figure 3. This system is based on multiple antenna scenarios where both the transmitter and the receiver use several antennas, each one with separate radio frequency modules, and where the interfering channels are the radio links between all pair of transmit and receive antennas.

In this MIMO communications system with transmit antennas and receive antennas (with so that the linear system it gives rise to is determined) the relation between the transmitted and received signals can be modelled in the baseband as (21)

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{21}$$

where $\mathbf{y} = \left[y_1, \dots, y_{N_R}\right]^{\mathrm{T}} \in \mathrm{C}^{\mathrm{N_R}\times 1}$ is the received signals vector and $\mathbf{x} = \left[x_1, \dots, x_{N_T}\right]^{\mathrm{T}} \in \mathrm{C}^{\mathrm{N_T}\times 1}$ is the transmitted signals vector. The radio links between each pair of transmit and receive antennas are represented by the channel matrix $\mathbf{H} \in \mathrm{C}^{\mathrm{N_R}\times \mathrm{N_T}}$ in which its entries h_{ij} represent the complex coefficient associated with the link between the pair of a i^{th} receive

antenna and the j^{th} transmit antenna. Each h_{ij} is taken from a zero-mean circularly symmetric complex Gaussian distribution with unit variance, which corresponds to having a variance equal to $1/2$ in both real and imaginary components. In order to have an independent and identically distributed Rayleigh fading channel model the phase of each entry h_{ij} is uniformly distributed in $[0\ 2\pi[$ and their amplitude has a Rayleigh distribution. In this model the vector $\mathbf{n} = \left[n_1, \ldots, n_{N_R}\right]^{\mathrm{T}} \in \mathrm{C}^{\mathrm{N_R}\times 1}$ represents the noise vector that is added to the incoming signal vector. The entries of n are random variables taken from an independent circularly symmetric complex Gaussian with zero average and variance σ_n^2 , so that both its real and imaginary components have variance $\sigma_n^2/2$. This noise model is usually called as zero-mean spatially white (ZMSW) noise [13].

In this work square quadrature amplitude modulation (QAM) constellations are used. The simulations are conducted with M-QAM constellations with $M = 4,\ 16,\ 64$ and the input symbols in each transmit antenna are taken from a finite complex constellation C constructed from the Cartesian product $C = C_R \times C_R$, where C_R is the real alphabet

$$C = \left\{-\left(\sqrt{M}-1\right), \ldots, -3, -1, +1, +3, \ldots, +\left(\sqrt{M}-1\right)\right\}. \tag{22}$$

Although QAM is assumed, QPSK (quadrature phase shift keying) can also be used in MIMO, leading to the same results as 4-QAM as the radio waves are modulated the same way. In addition to that, most literature on MIMO spatial multiplexing uses QAM, and it is easier to relate the QAM constellations into lattices.

The average energy of the complex symbol taken from C is given by

$$E_s = \frac{1}{\mathrm{M}} \sum\nolimits_{\mathrm{x_i} \in \mathrm{C}} |\mathrm{x_i}|^2 \tag{23}$$

Assuming, without loss of generality, that the filters at the receiver have impulse response $h(t)$ normalised to $\int |h(t)|^2\ dt = 1$.

Considering that each y_i receives the sum of N_T symbols weighted by unit power random variables, i.e., $\mathrm{E}\left[\left|h_{ij}\right|^2\right] = 1$, on average it is valid to calculate the SNR at the receiver as

$$\frac{E\{\|y\|^2\}}{E\{\|n\|^2\}} = \frac{E\{\|Hx\|^2\}}{E\{\|n\|^2\}} = \frac{E\left[\sum_{i=1}^{N_R}\sum_{j=1}^{N_T}\left|h_{ij}x_j\right|^2\right]}{E\left[\sum_{i=1}^{N_R} n^2\right]} = \frac{N_T N_R \sigma_x^2}{N_R \sigma_n^2} = N_T \frac{\sigma_x^2}{\sigma_n^2}. \quad (24)$$

Table 2 lists the values of the average energy (E_s) for the M-QAM modulations implemented in this work.

Table 2 - Symbol energy for the modulations used.

	4-QAM	**16-QAM**	**64-QAM**
E_s	2	10	42

In the following sections a series of receivers will be analysed, starting with the linear receivers such as zero-forcing (ZF) and minimum mean square error (MMSE); followed by the ordered successive interference cancelation (OSIC) algorithm, and finally the receivers using lattice reduction-aided (LRA).

One shall keep in mind that throughout this chapter, channel knowledge at the receiver is required in the model assumed, previously named as channel state information at the receiver (CSIR). Acquiring the channel knowledge is not an easy task, especially in fast fading channels, but this is not the focus of this thesis.

Throughout this work the performance of a receiver will be noted by plotting the symbol error rate (SER) as a function of the SNR. The diversity gain, or slope d is the metric used to evaluate the performance of the receivers. All receivers here analysed are well calibrated and the results are on pair with the results available in the literature.

4.2 Linear receivers

Linear receivers consist of applying a linear transformation to the received vector followed by a quantization to the symbol alphabet, also known as slicing. They are the simplest of all receivers but also the ones with worse results. However, due to the low complexity, they are scalable to a fair number of antennas at the terminals as it will be studied on the fifth chapter.

One can estimate $\hat{\mathbf{x}}$ from

$$\mathbf{x_{ZF}} = \mathbf{H^{-1}y} = \mathbf{H^{-1}Hx} + \mathbf{H^{-1}n} = \mathbf{x} + \mathbf{H^{-1}n} \tag{25}$$

$$\hat{\mathbf{x}}_{\mathbf{ZF}} = Q_C[\mathbf{x_{ZF}}] \tag{26}$$

which is commonly known as linear *zero-forcing* (ZF) receiver, as the interference caused by $\mathbf{H}$ is forced to be zero. In this thesis the channel matrix $\mathbf{H}$ was defined in the complex domain so in the equation (25), $\mathbf{H^{-1}}$ corresponds to the *pseudo-inverse matrix*, also known as *Moore-Penrose matrix* $\mathbf{H}^{\dagger}$**.**

$$\mathbf{H}^{\dagger} = \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{H}} \tag{27}$$

Superscript $(.)^{\mathbf{H}}$ denotes *Hermitian* operator which is a conjugation followed by transposition or vice-versa. One shall keep in mind that only matrices with non-zero determinants are invertible, and so it is required that $N_R \geq N_T$. Note that the noise is enhanced by the $\mathbf{H}^{\dagger}$ transformation.

$$\hat{\mathbf{x}}_{\mathbf{ZF}} = Q_C\left[\mathbf{H}^{\dagger}\mathbf{y}\right] = Q_C\left[\mathbf{H}^{\dagger}\mathbf{x} + \mathbf{H}^{\dagger}\mathbf{n}\right] = Q_C\left[\mathbf{x} + \mathbf{H}^{\dagger}\mathbf{n}\right] \tag{28}$$

The detected vector $\hat{\mathbf{x}}_{\mathbf{ZF}}$, as obtained from (4.8), is in fact the solution to:

$$\hat{\mathbf{x}}_{\mathbf{ZF}} = \arg\min_{\mathbf{x}\in\mathbb{C}^N} \|\mathbf{y} - \mathbf{Hx}\|^2 \tag{29}$$

This receiver solves the CVP by relaxing it to a search in a continuous neighbourhood instead of computing the distance between the received vector and every point in the lattice. It can also be seen as a linear transformation of the Voronoi regions of the $\mathbb{Z}^N$ by $\mathbf{H}$. The resulting regions are called ZF *decision regions* and correspond to the space where a lattice point will be interpreted as being closer to the lattice point associated with that region [14]

We can minimize the noise enhancement with minimum mean-square error (MMSE) receiver. In this receiver both the interference and the noise are considered in order to minimize the expected error. The main difference between these two linear receivers is that MMSE looks for a linear transformation W_{MMSE} that minimizes the mean square error between the estimated vector and the original vector,

$$\mathbf{W}_{\mathrm{MMSE}} = \arg\min_{\mathbf{W}} E\{\|\mathbf{Wy} - \mathbf{x}\|^2\} \tag{30}$$

That can also be represented as:

$$\mathbf{W}_{\mathrm{MMSE}} = \sigma_x^2 \mathbf{I}_{\mathrm{N}} \cdot \mathbf{H}^{\mathrm{H}} \left(\sigma_n^2 \mathbf{I}_{\mathrm{N}} + \mathbf{H} \cdot \sigma_x^2 \cdot \mathbf{I}_{\mathrm{N}} \cdot \mathbf{H}^{\boldsymbol{H}}\right)^{-1} = \mathbf{H}^{\boldsymbol{H}} \left\{\mathbf{H}\mathbf{H}^{\mathrm{H}} + \frac{\sigma_n^2}{\sigma_x^2} \mathbf{I}_{\mathrm{N}}\right\} \tag{31}$$

$\hat{\mathbf{x}}_{\mathbf{MMSE}}$ can be obtained by a linear transformation W_{MMSE} followed by a quantization step that is the same as in ZF.

$$\hat{\mathbf{x}}_{\mathbf{MMSE}} = Q_C[\mathbf{W}_{\mathrm{MMSE}}\, \mathbf{y}] = Q_C[\mathbf{W}_{\mathrm{MMSE}}\, \mathbf{Hx} + \mathbf{W}_{\mathrm{MMSE}}\mathbf{n}] = Q_C[\mathbf{x} + \mathbf{W}_{\mathrm{MMSE}}\mathbf{n}] \tag{32}$$

It is expected for MMSE to perform better than ZF because it solves CVP problem by relaxing the search in the continuous space where $x \in C^N$ but also for introducing a term that penalises large $\|x\|$ and is proportional to the energy of the noise [14].

4.3 Maximum Likelihood Detection

Before posting the results of the linear receivers, one should stablish the standard comparison of all the detectors relating to the MLD curve, the best curve possibly obtained, although mainly computational as its performance is extremely expensive computationally.

The maximum likelihood estimate $\hat{\mathbf{x}}$ [15] for $\mathbf{x}$ given $\mathbf{y}$ is

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\max_{\mathrm{x} \in \mathrm{C}^N} \mathrm{p}(\mathbf{y}|\mathbf{H}, \mathbf{x}) \tag{33}$$

Where assuming the ZMSW noise model, the probability density function of $\mathbf{y}$ given $\mathbf{H}$ and $\mathbf{x}$ can be written as

$$\mathrm{p}(\mathbf{y}|\mathbf{H}, \mathbf{x}) = \frac{1}{(2\pi\sigma_n^2)^{N/2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{Hx}\|^2}{2\sigma_n^2}\right) \tag{34}$$

Leading to

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\max_{\mathrm{x}\in C^N} \frac{1}{(2\pi\sigma_n^2)^{N/2}} \exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{2\sigma_n^2}\right)$$

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{\mathrm{x}\in C^N} \|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2 \tag{35}$$

The exponential parcel induces a growth in the search space for M-QAM constellations thus discouraging the use of brute force maximum-likelihood detection when N increases.

The figures 13 and 14 show the performance of ZF, MMSE and ML detectors for 2x2 MIMO configurations. One can observe several happenings. The ML curve achieves diversity order equal to *N* as stated in the previous chapters, and so it serves as comparison because it achieves the best possible performance. Note that these simulations were run with few antennas as ML is a heavy detection algorithm, hence the 2x2 system. Looking now at ZF and MMSE curves, it is possible to observe that MMSE performs slightly better than ZF, because it solves the CVP problem by relaxing the search in the continuous space but also for introducing a term that penalises large $\|\mathrm{x}\|$ and is proportional to the energy of the noise. Regarding the inclination of the curve, both ZF and MMSE maintain a steady -1 slope which corresponds to a decrease of SER by a factor of 10 for every 10dB of SNR. It can also be noticed that 16-QAM constellation suffers from a power penalty of some dB when comparing to the 4-QAM curve.

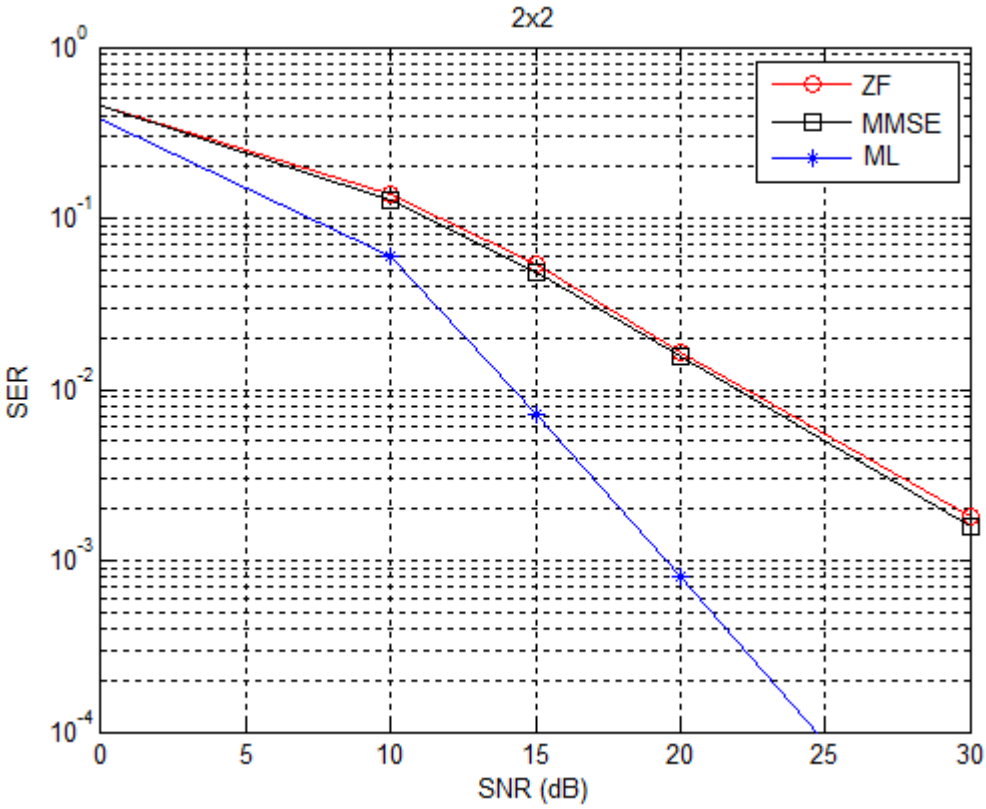


Figure 13 - ZF, MMSE and ML detection with 2×2 antennas for 4-QAM.

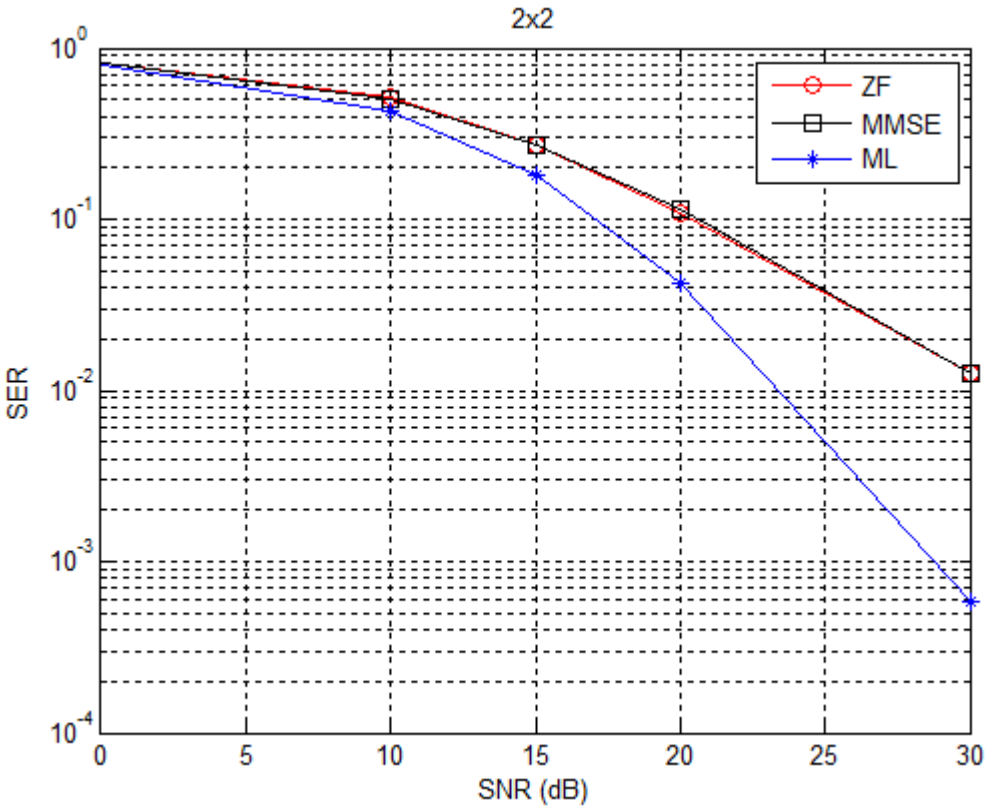


Figure 14 - ZF, MMSE and ML detection with 2×2 antennas for 16-QAM.

4.4 Detection by Ordered Successive Interference Cancellation

Linear receivers are limited as they fail to remove the interference between signals from each antenna. More ingenious detection methods that manage to remove the interference end up being nonlinear. A popular approach is called ordered successive interference cancelation (OSIC) also known as V-BLAST [16].

In this algorithm the detection side of spatial multiplexing requires some attention on its implementation as each transmitted symbol is received by all the receive antennas. OSIC then uses linear detection to infer the first modulated symbol of the layer with least noise and subsequently assumes this detection was correct, continuing to detect the symbols from the remaining layers, replicating the interference created by the detected symbol and subtracting from the following layers. This procedures goes on until all symbols are detected.

The decision criterion at each stage is to select the layer that enhances less the noise power after the linear detection, resulting in a minimised error probability for that layer. Without ordering this selection there would be a performance drop [17].

It is possible to implement an OSIC detector in several ways. In this thesis a QR decomposition method applied on a ZF will be presented, which basically translates the channel matrix **H** into an upper triangular matrix, to then perform the search among the simplified layers.

A **QR** decomposition is performed on the channel matrix as $\mathbf{H} = \mathbf{QR}$, where **Q** is an $N_T \times N_R$ orthonormal matrix and **R** is an $N_T \times N_R$ upper triangular matrix. Then the detection problem can be seen as

$$\hat{\boldsymbol{x}} = \arg\min \left\| \mathbf{Q}^{\mathrm{H}} \boldsymbol{y} - \mathbf{R}\mathbf{x} \right\|^2 \tag{36}$$

Thanks to the upper triangular structure of **R** , the detector performs a layered-search from the N_{th} layer to the 1^{st}layer to find optimal solutions

One problem of this strategy is that if one symbol is incorrectly detected, all the following symbols are likely to be incorrect as well.

Instead of ZF equalization it is also possible perform linear MMSE equalization to obtain a MMSE version of the OSIC receiver, by adjusting the equation 4.16

The performance of the OSIC receivers described above can be seen in figures 15 and 16.The OSIC based on ZF and MMSE is denoted as OSIC-ZF and OSIC-MMSE. One can see

that these linear OSIC receivers still obtain curves with slope -1 as OSIC does not improve diversity but a large power gain. This gain can be observed by comparing figures 17 and 18 with the results obtained in the previous section, however the simulations on figure 15 and 16 were run with a 2x2 MIMO system and here we are analysing a 3x3 MIMO configuration, so the gain is higher than if comparing under the same circumstances.

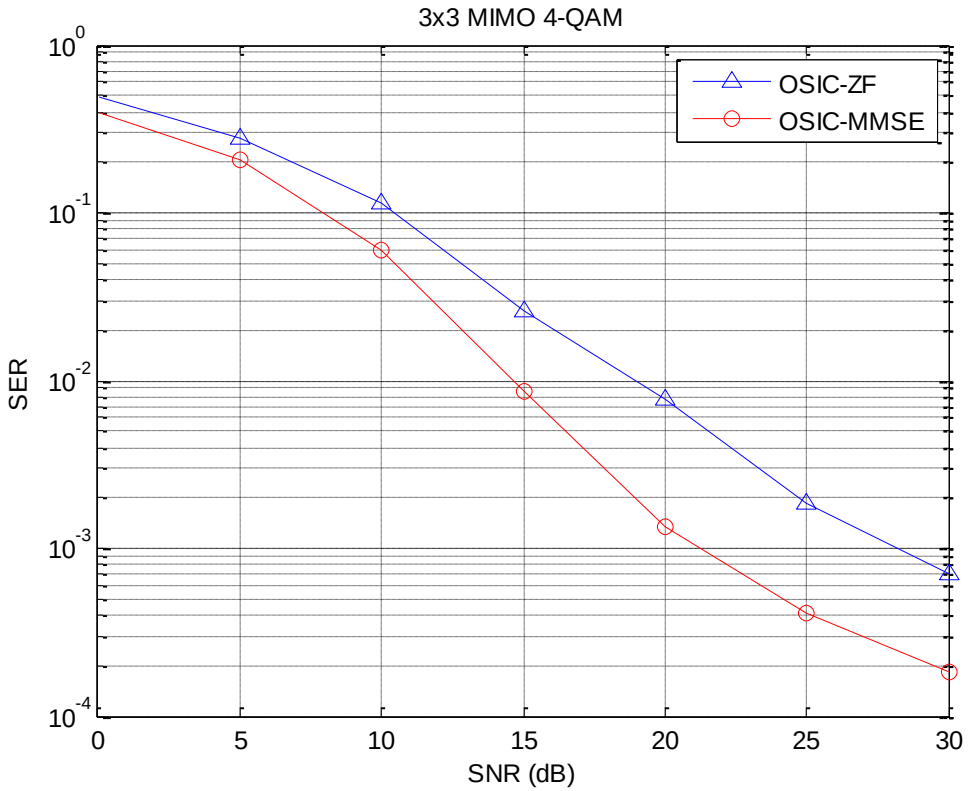


Figure 15 - OSIC-ZF and OSIC-MMSE detection with 3x3 antennas for 4-QAM.

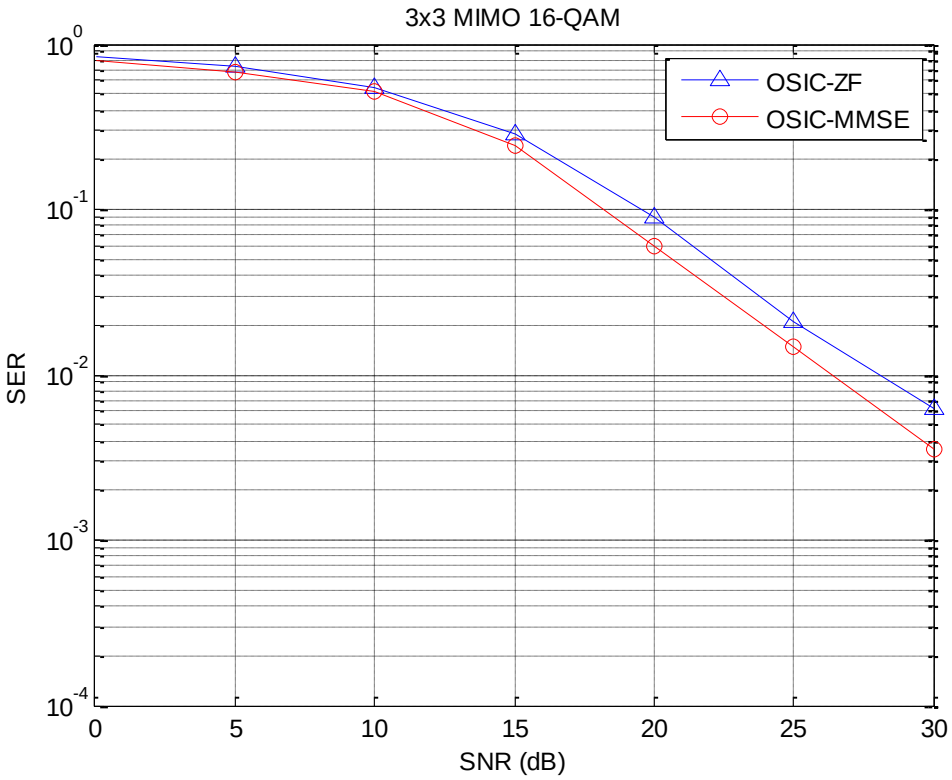


Figure 16 - OSIC-ZF and OSIC-MMSE detection with $\mathbf{3 \times 3}$ antennas for 16-QAM.

4.5 Lattice Reduction-Aided Detection

On the previous chapter we discussed about lattice bases generation, and realised that the same lattice can be created by an infinite possibility of bases. However, some bases have properties that enhance the performance of the standard receivers, as their decision regions are closer to the ideal Voronoi regions. Intuitively this opens a door where one wants to preferably use a better lattice base in order to achieve better results, and this is the motivation for the lattice reduction-aided (LRA) detection.

Lattice reduction can be implemented by algorithms like LLL and KZ, and these algorithms try to find a "better" base for a lattice, which consists of a base whose generating vectors are more orthogonal to each other than the original ones. This new lattice is usually referred to as $\mathrm{H_{red}}$and relates to H by

$$\mathbf{H}_{\mathrm{red}} \equiv \mathbf{HM}\,, \tag{37}$$

where $\mathbf{M}$ is an unimodular matrix, meaning that it is composed of integer entries and has $|\det(\mathbf{M})| = 1$. Note that $\mathbf{M}^{-1}$ is also unimodular.

Inserting the concept of lattice reduction in MIMO we can now re-write the standard system model

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n} = \mathbf{HM}\,\mathbf{M^{-1}x} + \mathbf{n} \tag{38}$$

Considering the equation 37 and defining $\mathbf{z} = \mathbf{M^{-1}x}$, we obtain the reduced system

$$\mathbf{y} = \mathbf{H_{red}z} + \mathbf{n} \tag{39}$$

In this model, $\mathbf{z}$ is a modified data vector that can be detected with a lower SER than would $\mathbf{x}$ without Lattice Reduction, and this is true for any type of receivers that follow LR pre-processing, meaning that all of the receivers analysed in this thesis are eligible for LRA detection.

The original data vector $\mathbf{x}$ can be recovered from $\mathbf{z}$ according to

$$\mathbf{z} = \mathbf{M}^{-1}\mathbf{x} \implies \mathbf{x} = \mathbf{Mz} \tag{40}$$

An adjustment has to be done before utilizing LR tools because the M-QAM constellations used in this dissertation are defined without the origin and have non unitary distance between the symbols, so a translation of the constellation is needed to apply lattice reduction. This translation is done by modifying the received vector

$$\mathbf{y}_{\mathrm{red}} = \frac{1}{2}(\mathbf{y} + \mathbf{Hp}) \tag{41}$$

where $\mathbf{p}$ denotes a column vector of n complex elements all equal to $1+j$.

The LLL algorithm was first derived for integer lattices and then applied to real lattices. In this thesis a complex LLL algorithm was implemented, and it helps reducing the complexity of MIMO detectors, even while taking the pre-processing into account.

The pseudo-algorithm used to simulate LRA receivers is shown in table 3.

Table 3 Pseudo-code of LRA detection.

1: Shift and scale the constellation to have zero as a lattice point:

$\mathbf{y}_{\mathrm{red}} = \frac{1}{2}(\mathbf{y} + \mathbf{Hp})$,

$\mathbf{p} = \left[\underbrace{1+j,\ldots,1+j}_{N}\right]^{T}$.

2: Reduce the lattice basis **H** using CLLL:

$[\mathbf{H}_{\mathrm{red}}, \mathbf{M}] = \mathrm{CLLL}(\mathbf{H})$.

3: Apply some detector to the CVP defined by $(\mathbf{H}_{\mathrm{red}}, \mathbf{y}_{\mathrm{red}})$

$\mathbf{z} = \mathrm{Detect}(\mathbf{H}_{\mathrm{red}}, \mathbf{y}_{\mathrm{red}})$.

4: Using $\mathbf{z}$ and $\mathbf{M}$ estimate the symbol in the original coordinate system

$\hat{\mathbf{x}} = 2\mathbf{M}Q_z[\mathbf{z}] - \mathbf{p}$.

The performance of LRA receivers is now presented in different configurations. A first analysis was performed on the linear receivers with lattice reduction which can be seen in figures 17 and 18. Comparing the results of LRA-ZF and LRA-MMSE with the results from the last section OSIC-ZF and OSIC-MMSE, one can conclude that the LRA allows a desirable and visible gain and that gain is even more accentuated as the QAM constellation goes from 4 to 16.

The second analysis of LRA receivers was performed on the linear receivers, but this time with OSIC schemes, leading to figures 19 and 20. Comparing the results of these two figures with figures 17 and 18, one can observe that the introduction of the OSIC approach allows a slight gain, however it is much less noticeable than the gain obtained by the increment of the LRA technique in the first two figures.

Summarising the information introduced hitherto, one should note that LRA-OSIC-MMSE receiver is the one that allows the best performance so far. However, an experimental approach will be made upon LRA-OSIC detection algorithms, which consists of doing a randomized detection of the OSIC algorithm, in an attempt of achieving results beyond LRA-OSIC receivers.

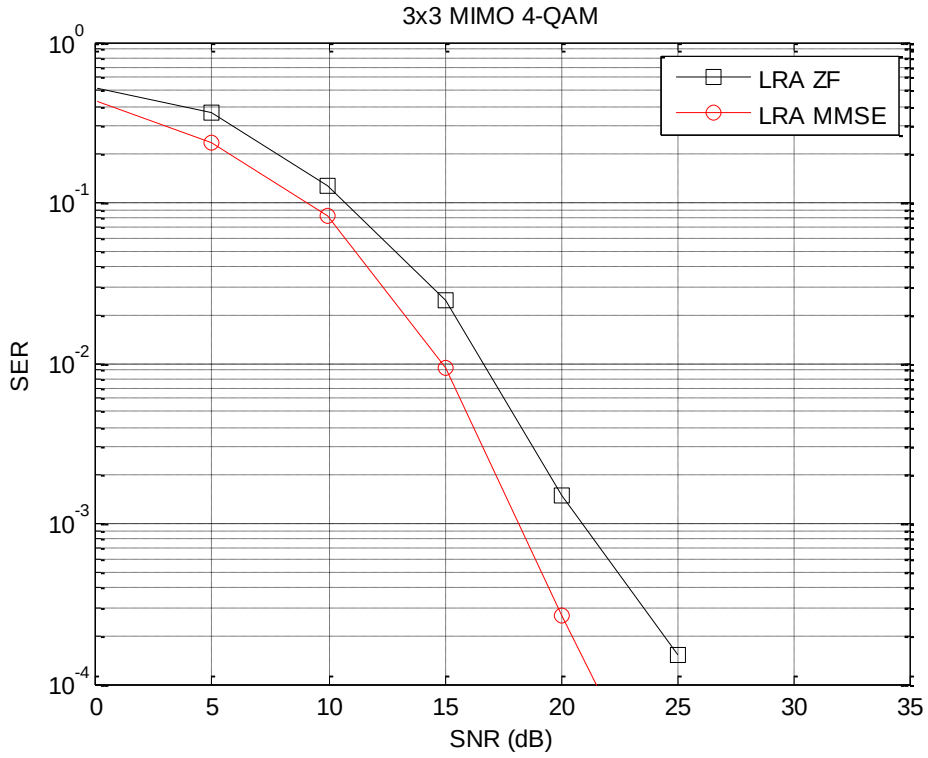


Figure 17 - LRA-ZF and LRA-MMSE detection with 3×3 antennas for 4-QAM.

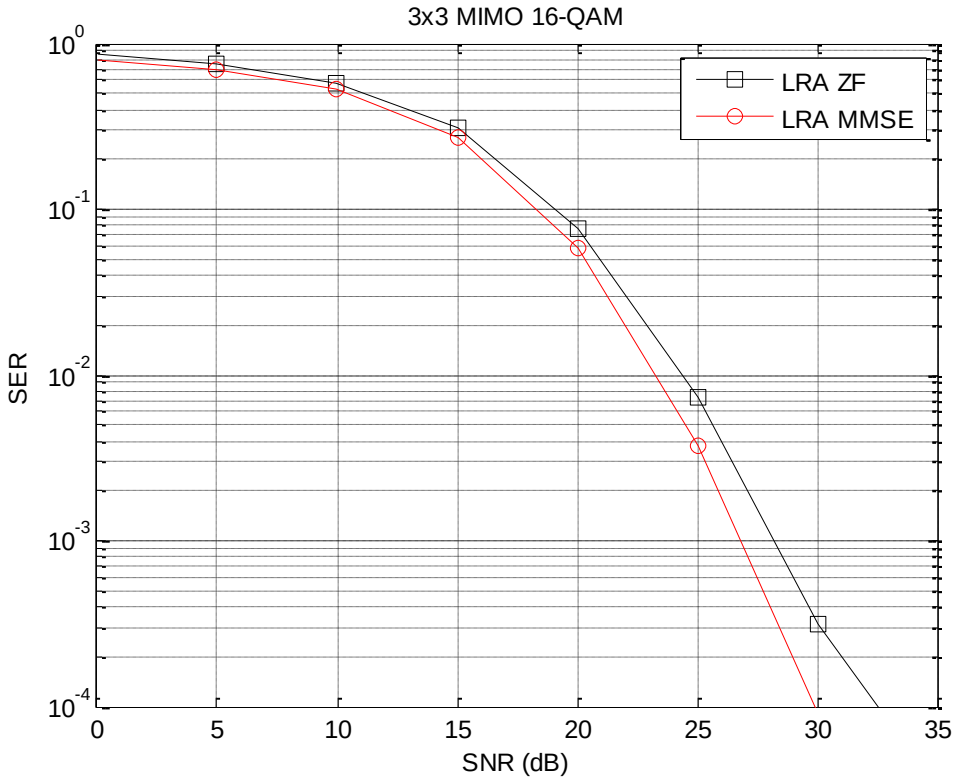


Figure 18 - LRA-ZF and LRA-MMSE detection with **3 × 3** antennas for 16-QAM.

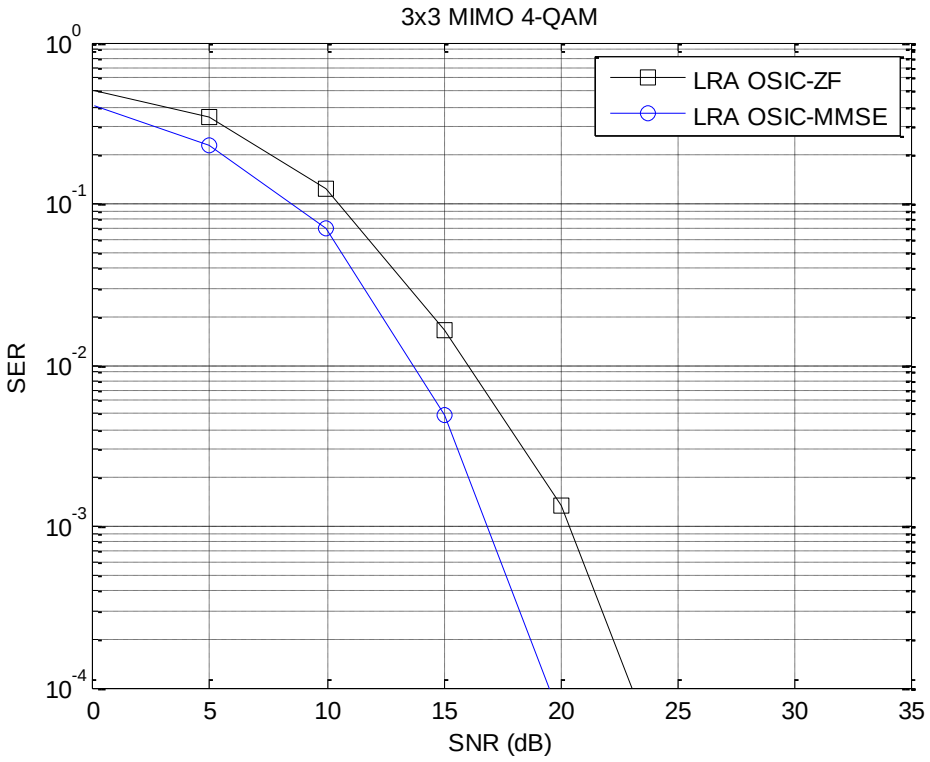


Figure 19 - LRA-OSIC-ZF and LRA-OSIC-MMSE detection with 3 × 3 antennas for 4-QAM

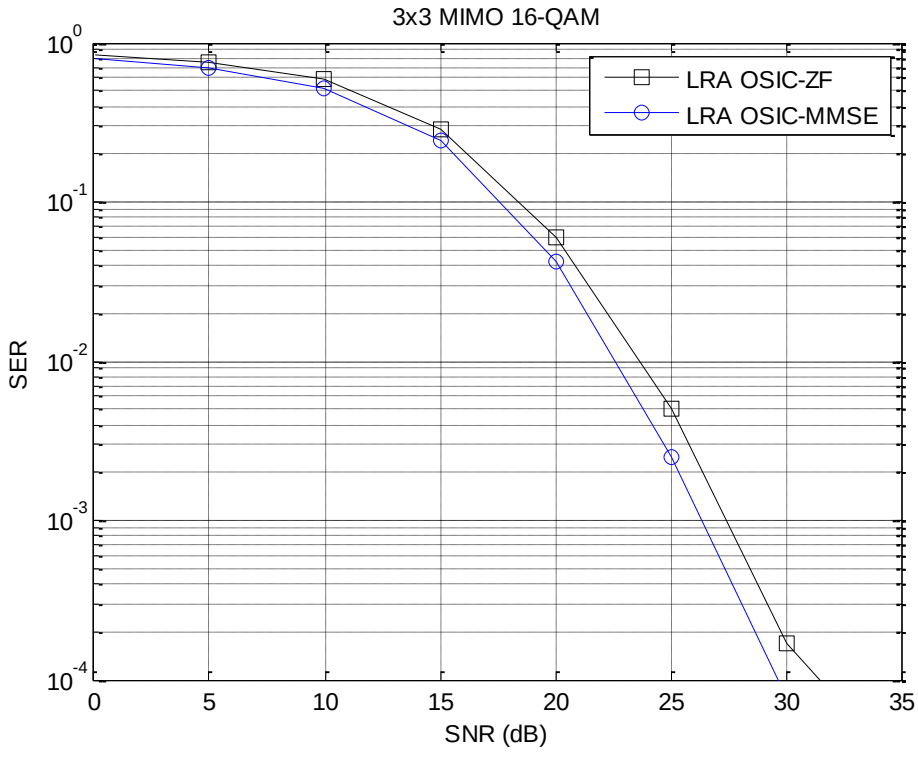


Figure 20 - LRA-ZF and LRA-MMSE detection with **3** × **3** antennas for 16-QAM

4.6 Randomized Decoder

While lattice reduction-aided decoding has a relatively small complexity, its performance exhibits a widening gap to the maximum likelihood curve as the dimension increases. In order to improve its performance, this thesis presents an improved algorithm that can shorten that gap to ML detection, which is a randomized version of SIC, first proposed by Liu et al. (2011), and based on the work by Klein (2000).

Previously, an OSIC algorithm was described where a standard rounding to the nearest Gaussian integers was applied. However, in this algorithm, the standard rounding will be replaced by Klein's randomized rounding, and a few candidate lattice points are sampled from a Gaussian-like distribution over the lattice. To find the closest lattice point, Klein's algorithm is used to sample some lattice points and the closest among those samples is chosen. Lattice reduction increases the probability of finding the closest lattice point, and only needs to be run once during pre-processing [18]. Furthermore, the sampling can operate very efficiently in parallel.

One should calculate

$$\hat{x}_i = randRound\left(\frac{\hat{y}_i - \sum_{j=i+1}^{N_T} r_i \hat{x}_j}{r_{i,i}}\right), \tag{42}$$

where function $randRound\ (r)$ rounds the real and imaginary parts of r (denoted as r_{re} and r_{im}) to integers q_{re} and q_{im}, respectively, according to the discrete Gaussian distribution defined as

$$P(Q = q_{re}) = \frac{e^{-c(r_{re}-q_{re})}}{\sum_{q=-\infty}^{+\infty} e^{-c(r_{re}-q_{re})}} \tag{43}$$

For the real part and

$$P(Q = q_{im}) = \frac{e^{-c(r_{im}-q_{im})}}{\sum_{q=-\infty}^{+\infty} e^{-c(r_{im}-q_{im})}} \tag{44}$$

for the imaginary part. This Gaussian distribution is usually similar to figure 21,

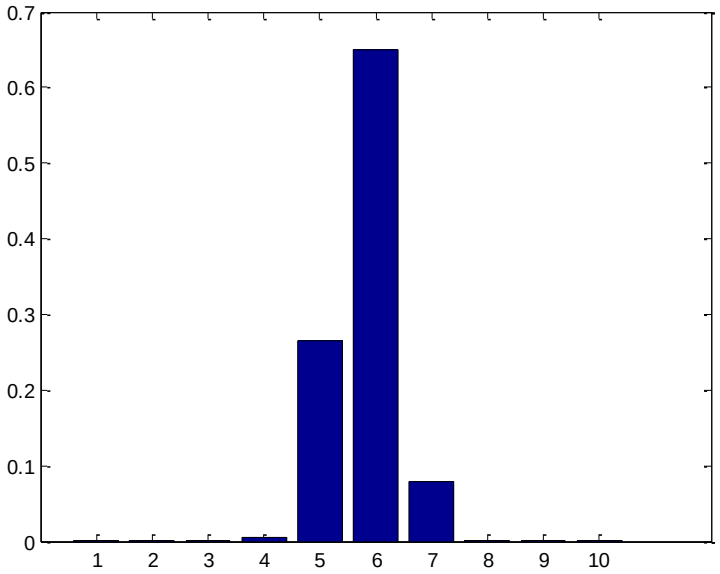


Figure 21 – Discrete Gausssian Normal Distribution obtained from (43) and (44) with test values.

The parameter c_i is computed as

$$c_i = \frac{\log(\rho)}{\min_j(r^2{}_{i,i})} r^2{}_{i,i} \tag{45}$$

with ρ being another parameter whose optimum value can be obtained from

$$K = (e\rho)^{\frac{4N_T}{\rho}} \tag{46}$$

The number of candidate lattice points that are considered in the algorithm is represented by *K*. The candidate list is then built by repeating *K* times the procedure for computing a lattice point with Equation 42. If the transmission is uncoded the final estimate will correspond to the closest of the *K* lattice point candidates. Due to the randomized nature of the algorithm, to avoid the possibility of the final estimate being further away than the one produced by the OSIC decoder, one of the *K* candidates should be obtained through standard rounding. The final transmitted vector estimate is computed by scaling back the obtained solution into the original lattice.

Figure 22 depicts the performance of randomized decoding in comparison to the traditional ZF inverter and the standard OSIC detector. A 3x3 MIMO system was considered with 4 QAM modulation.

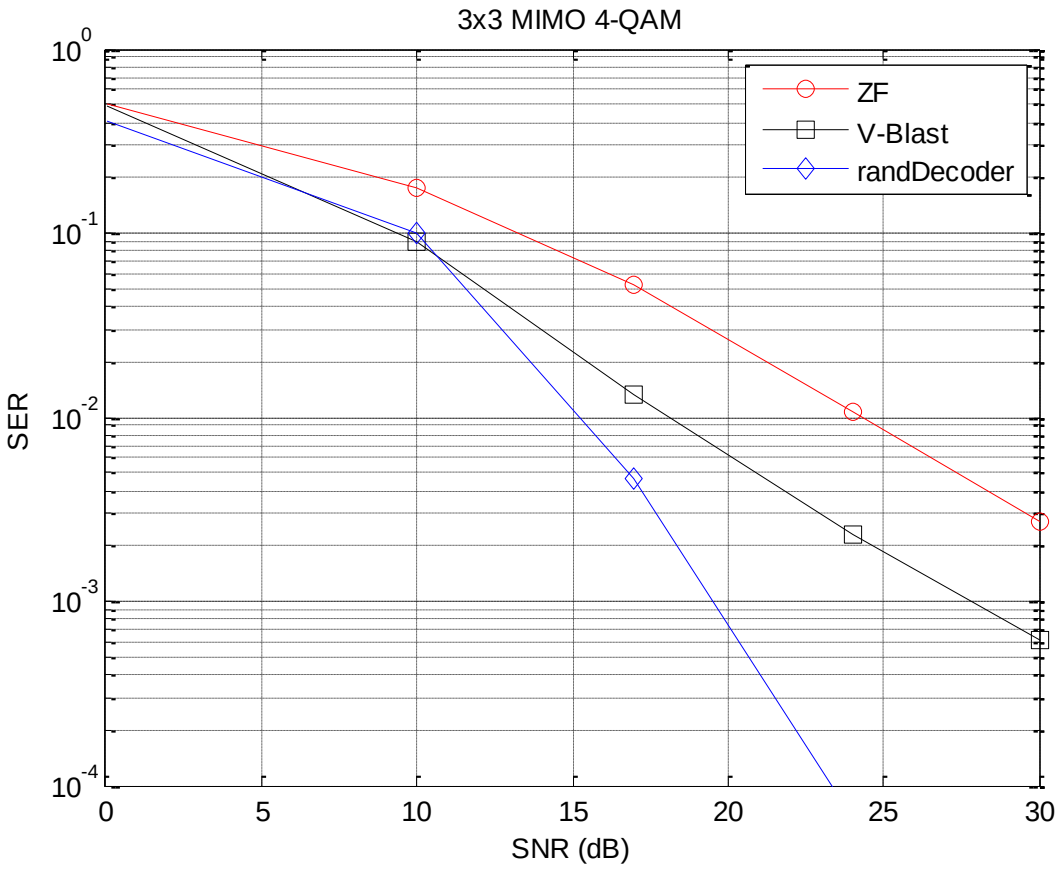


Figure 22 - ZF, MMSE and Randomized Decoder with 3×3 antennas for 4-QAM

One may observe how the randomised version of OSIC captures the full diversity available, in contrast to traditional deterministic OSIC based on deterministic rounding at each iteration. The price to pay is the need to generate *K* candidates to select from, basically multiplying the number of operation by that factor *K*.

5 Expanding the MIMO limits

In this chapter the limits of conventional MIMO receivers are tested for large systems, and some new approaches are made in an attempt to increase the number of antennas at both terminals.

5.1 Scaling up MIMO

From chapter 2 we know that the more antennas the transmitter/receiver is equipped with and the more degrees of freedom the propagation channel can provide, the better the performance in terms of data rate or link reliability. There are a few wireless broadband standards based on this knowledge, like the LTE standard that allows for up to 8 antenna ports at the base station to take advantage of all these features. On a channel that varies rapidly as a function of time and frequency, and where circumstances permit coding across many channel coherence intervals, the achievable rate scales with min (*nt, nr*) log (1+SNR).The gains in multiuser systems are even more impressive, because such systems offer the possibility to transmit simultaneously to several users and the flexibility to select what users to schedule for reception at any given point in time [19].

However, the price to pay for obtaining these new data rates with MIMO is increased complexity of the hardware and energy consumption of the signal processing at both ends. In point-to-point communications the complexity at the receiver requires more attention than the complexity at the transmitter because for instance, the complexity of optimal signal detection grows exponentially with N_T [20] [21]. On the other hand, in multiuser systems the greatest concern is on the complexity at the transmitter, as the coding schemes used to transmit information simultaneously to more than one user while maintaining decent levels of inter-user interference are rather complicated.

A decade ago, the large antenna arrays regime, when N_T and N_R increases, has not got beyond academic study, in that some asymptotic capacity scaling laws are known for ideal situations. However, this view is changing due to a few important system aspects in the large terminals regime have been discovered. For example, [22] showed that asymptotically as $N_T \rightarrow \infty$, and under realistic assumptions on the propagation channel, with a bandwidth of 20 MHz, a time-division multiplexing cellular system may accommodate more than 40 single-antenna users that are offered a net average throughput of 17 Mbits per second both in the reverse (uplink) and the forward (downlink) links, and a throughput of 3.6 Mbits per second with 95% probability! These rates are achievable without cooperation among the base stations and by relatively rudimentary techniques for CSI acquisition based on uplink pilot measurements.

When talking about very large MIMO, one thinks of systems that use antenna arrays with hundreds or more antennas. Very large MIMO entails a considerable number of antennas

simultaneously serving a much smaller number of terminals. The disparity in number emerges as a desirable operating condition and a practical one as well. The number of terminals that can be simultaneously served is limited by the inability to acquire channel-state information for an unlimited number of terminals. Larger numbers of terminals can always be accommodated by combining very large MIMO technology with conventional time and frequency division multiplexing via OFDM. It is expected that in very large MIMO systems, each antenna unit uses extremely low power, in the order of mW. Ideally, the total transmitted power should be kept constant as one increases N_T, meaning that the power per antenna should be proportional to $1/N_T$. It could also be possible to reduce the total transmitted power, but the need for multi-user multiplexing gains, errors in CSI, and interference would prevent these power savings in practice. Even considering these problems, the possibility of saving an order of magnitude in transmit power is desirable because it would open the possibility of achieving a better system performance under the same power constraints.

Another concern when studying very-large MIMO is the energy consumption of the cellular base stations. However, very-large MIMO designs can be made extremely robust to the failure of one or a few of the antenna units, meaning that a punctual failure would not harshly affect the system. A malfunctioning individual antenna can be swapped, in contrast to classical array designs, which use few antennas fed from a high-power amplifier.

5.2 Large dimensional channels

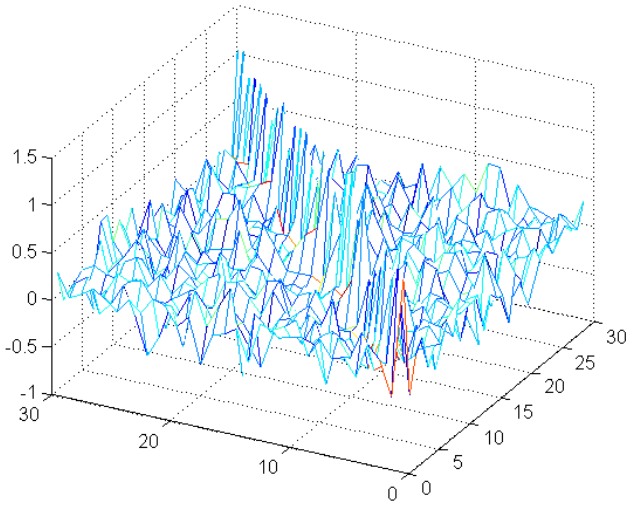
Although the main advantages of large MIMO systems stand upon the increase of the diversity gain and data rate, there are also other advantages given by large dimension systems that cannot be seen in smaller systems, and are a key factor for the existence of large MIMO arrays.

When the number of antennas at the terminals increases, the matrix **H** becomes larger, and the distribution of the singular values becomes less reliant on the entries of the channel matrix as long as they stay independent and identically distributed (iid) random variables.

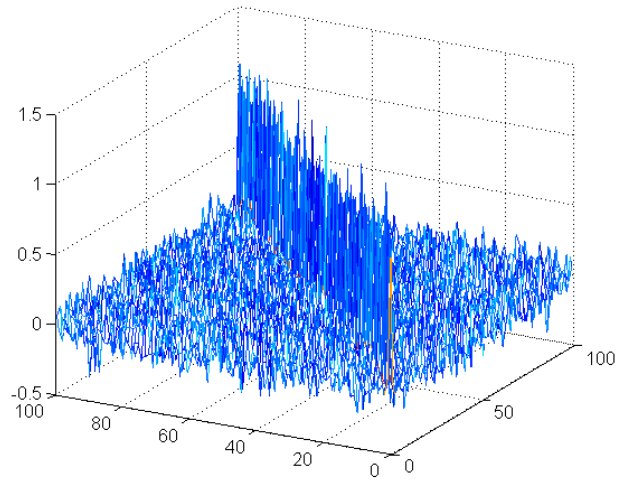
This is a known result of the Marcenko-Pastur law [23] and states that if the entries of an $nr \times nt$ matrix **H** are zero mean iid with variance $1/nr$, then the empirical distribution of the eigenvalues of $\mathbf{H}^{H}\mathbf{H}$ converges almost surely, as N_T, $nr \to \infty$ with $N_T/N_R \to \beta$, to the density function.

A direct result of the Marcenko-Pastur law is that very tall or very wide channel matrices tend to be very well conditioned, meaning that the eigenvalue histogram of a single realisation converges to the average asymptotic eigenvalue distribution, resulting in a more deterministic channel as the channel dimension increases. This is usually called "channel hardening" in the literature [24], and its behaviour at large dimensions can be observed in the figure 23, where $\mathbf{H}^{H}\mathbf{H}$ was plotted with N_T=N_R=30, 100 and 300, the entries of **H** are iid Gaussian entries with zero mean and unit variance.

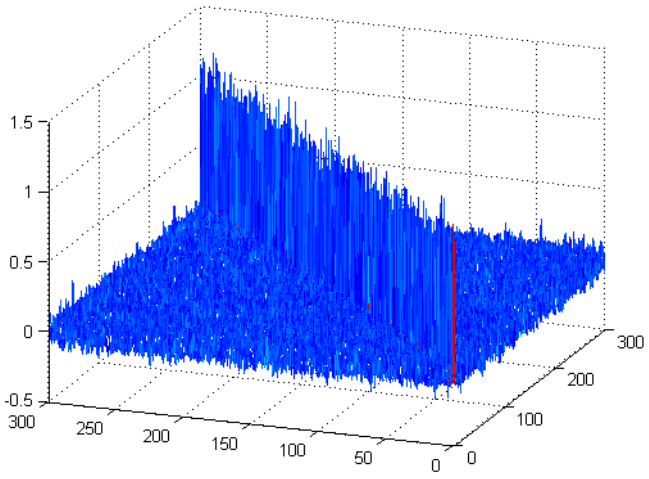
The main property to retain is that as **H** grows in size, the diagonal terms of $\mathbf{H}^{H}\mathbf{H}$ become increasingly larger than non-diagonal terms, resulting in a near-diagonal matrix.



(a)



(b)



(c)

Figure 23 - Intensity plots of $\mathbf{H}^{H}\mathbf{H}$ matrices for (a) 30×30 MIMO, (b) 100 × 100 MIMO, and (c) 300 × 300 MIMO channels.

5.3 Testing the limits of the conventional receivers for large-MIMO systems

The discovery of channel hardening enabled some advantages in signal processing, more specifically in large dimension MIMO. One of these advantages rely on the fact that the inversion of large random matrices can be efficiently approximated with series expansion like Newman Series [25] and deterministic approximations. This is helpful because the linear detectors studied in this thesis like Zero-Forcing and minimum mean square error detectors perform matrix inversions in their algorithmic process, thus making them unviable detectors for very large systems, due to the excessive computation requirement of inverting for example 100x100 matrices.

However, due to its simplicity, we can still obtain decent results for large antenna arrays using these two simple detectors. Channel hardening turns these low complexity detection algorithms suited for large channels.

Figures 24 and 25 show the performance of ZF and MMSE receivers but this time in large MIMO configurations. This time the ML curve was dismissed as it is a really heavy decoder, making it not feasible for comparison. The modulation chosen for this test was also 4-QAM as it's the cheapest modulation, and this test is mainly focusing on hardware performance, hence the focus on cheap algorithms in order to reach large dimensions.

However, one can note that even at a high antenna regime, the linear detectors can maintain a steady slope. This is on pair with the previous results and the literature that states that the linear receivers can only maintain a slope of -1. As expected, increasing the number of antenna terminals does not influence the curve slope but rather the SER, that is slightly degraded by the increasing dimensions.

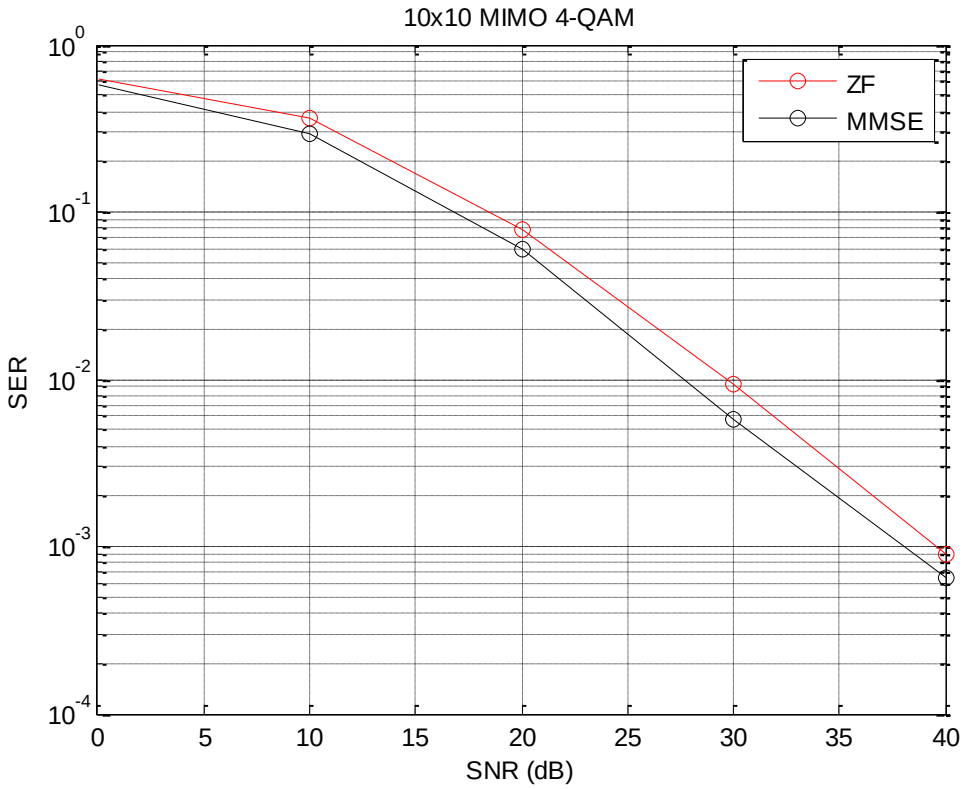


Figure 24 - Detection 10x10 antennas with 4-QAM using linear receivers.

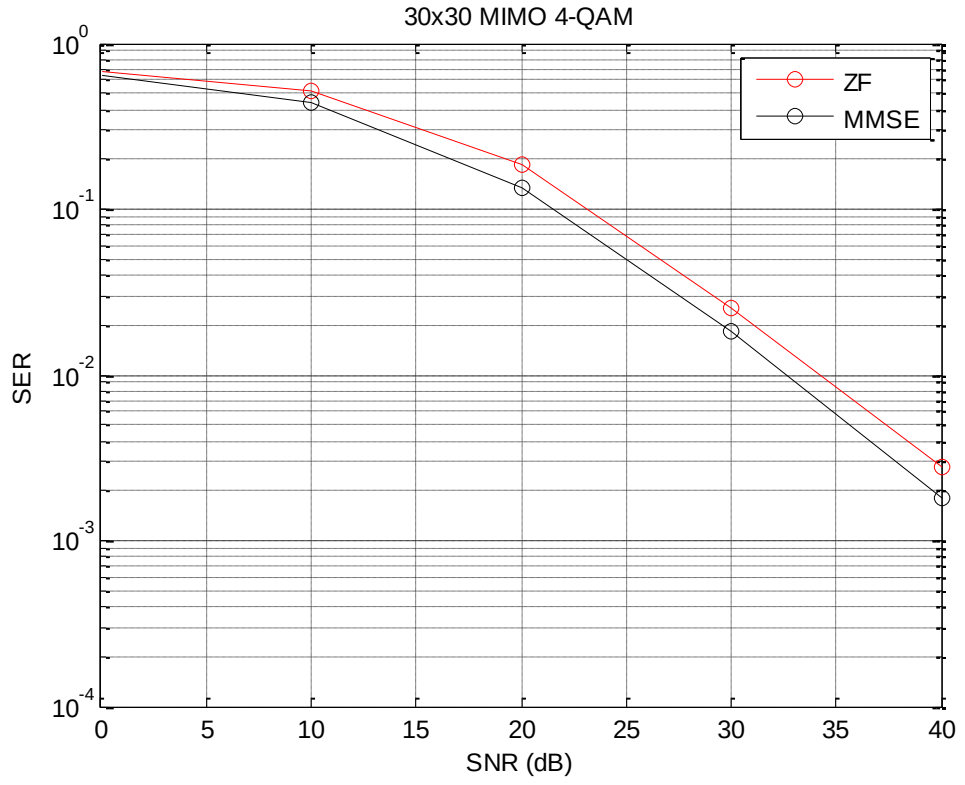


Figure 25 - Detection 30x30 antennas with 4-QAM using linear receivers.

Another direct consequence from channel hardening and that could already be observed from the figure 23 is that when dimensions grow big, the Gram matrix tends to be near diagonal. By plotting the Frobenious distance of a Gram matrix to the correspondent Identity matrix $\|\mathbf{H}^{\mathbf{H}}\mathbf{H} - \mathbf{I}\|_{Frob}$ we got the following figure. The script created random Channel matrices from *nt*=1 up to 400.

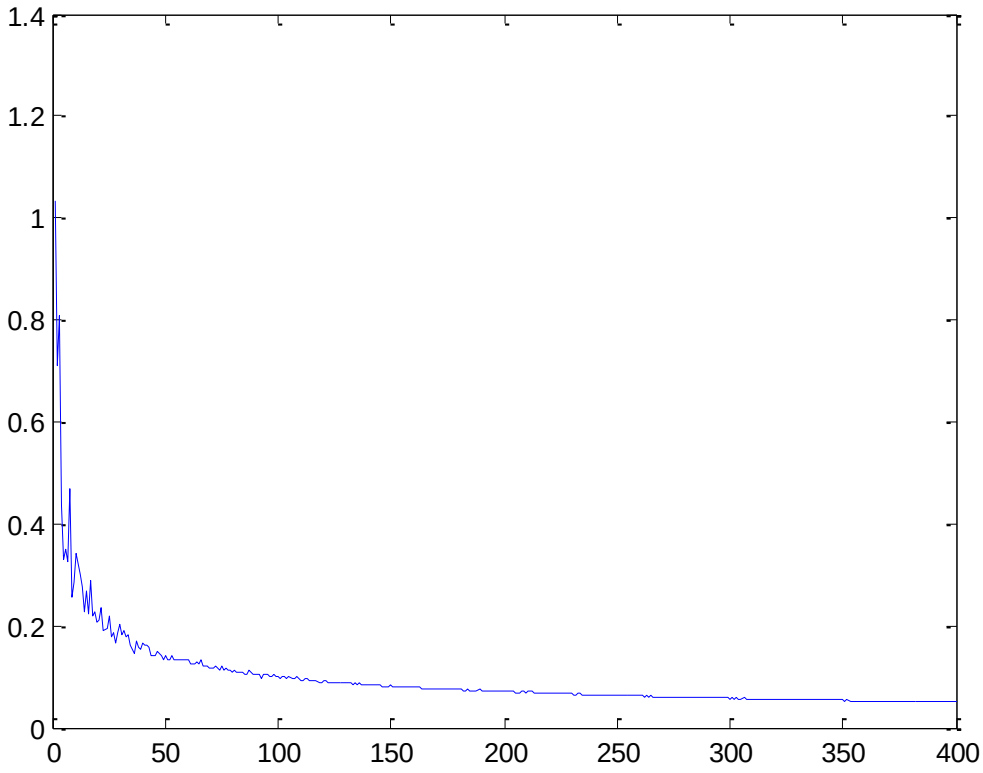


Figure 26 – Plot of the Euclidean distance $\|\mathbf{H}^{\mathbf{H}}\mathbf{H} - \mathbf{I}\|_{Frob}$ of a Gram matrix to an Identity matrix.

Observing this figure 26 we can note that as the number of antennas increases, the distance to the Identity matrix gets smaller, limited by an asymptotic zero.

This similarity to diagonal matrices also allows us to obtain some results by applying the standard SVD precoding to large channels. And again, the results are feasible due to channel hardening.

Note that even though the number of antennas more than doubles, the performance of the detector does not alter much, meaning that it is a good and scalable algorithm.

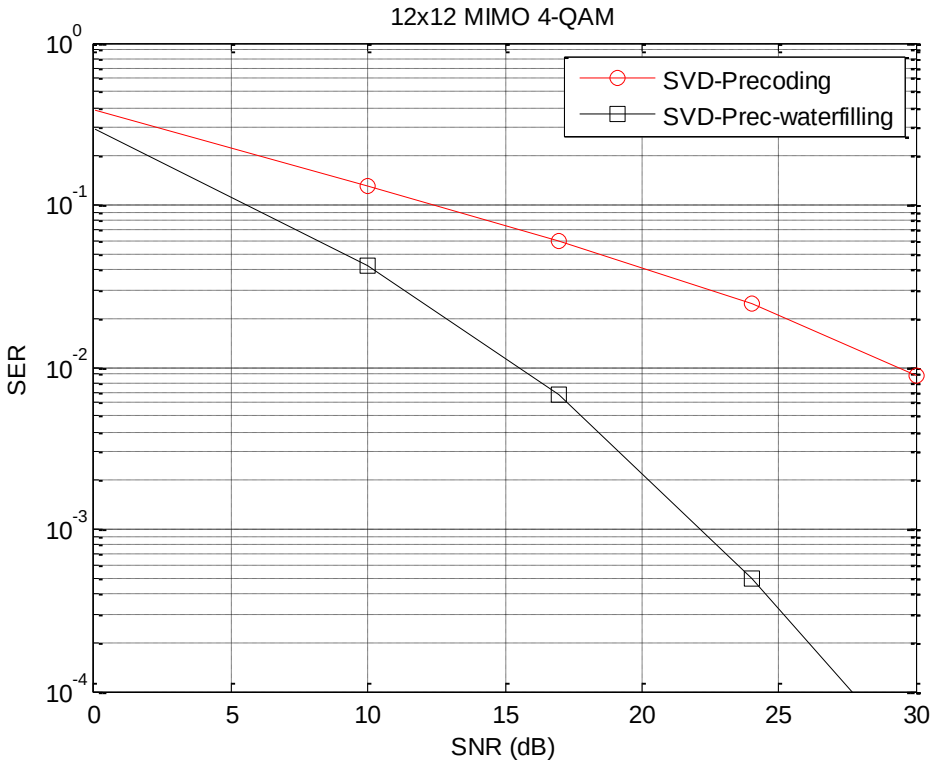


Figure 27 - Detection 12x12 antennas with 4-QAM using SVD channels.

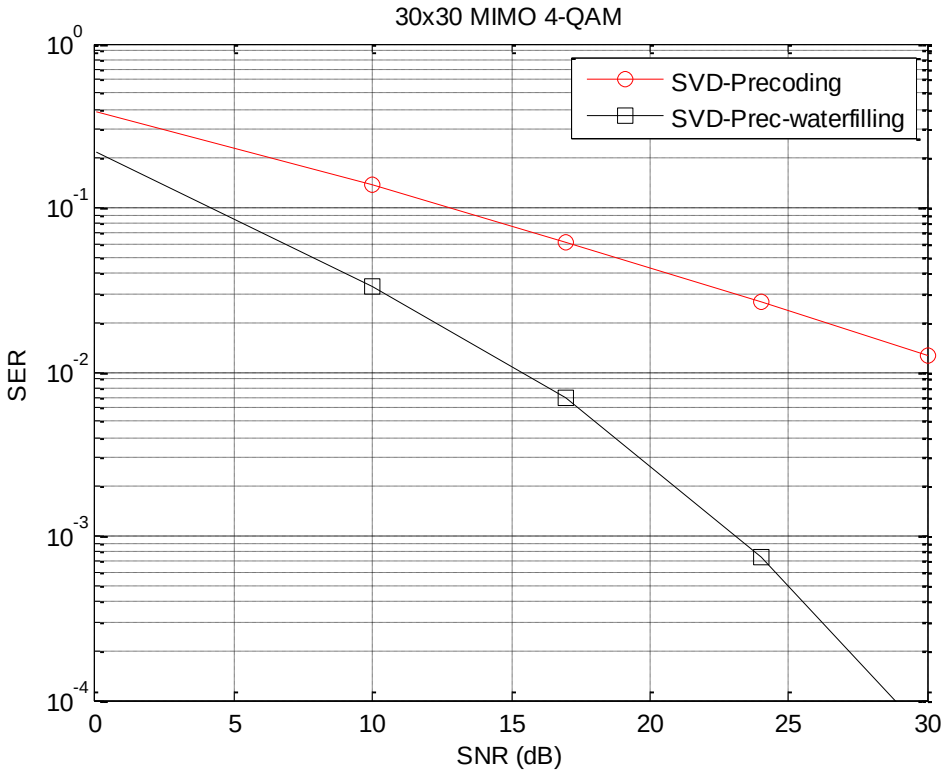


Figure 28 - Detection 30x30 antennas with 4-QAM using SVD channels.

6 Conclusions

In this chapter a wrap up of the thesis is made, and then some areas of possible future research are identified.

One of the objectives of this work was to exploit certain channel characteristics that would make MIMO systems with large antenna arrays possible. The two main aspects that were studied in order to achieve large antenna systems are diversity and spatial multiplexing, which were conceptually described on chapter 2, and then put into practice on chapters 4 and 5, testing popular receivers like ZF, MF and some non-linear variations with OSIC and LRA. Although the receivers that achieve the best performance are LRA-OSIC-MMSE and a randomized version of it, these algorithms ended up being computationally heavy as the number of antennas at the terminals increases, meaning that those were not scalable beyond 10 antennas.

In order to obtain large antenna arrays working for MIMO systems, we had to bet on cheaper algorithms like plain ZF and MMSE, and more ingenious detection methods involving channel information feedback that would allow the processing of the channel information at the receiver. By having access to the channel at the transmitter, it would then be possible to take advantage of the channel properties, creating individual parallel data pipes that are directly related to the channel matrix Eigen values. This channel simplification enables the use of large antenna arrays on MIMO systems, even though channel state information at the transmitter is required.

Throughout this work, some opportunities were identified as possible improvements for further development. In chapter 5, the Newman series expansion was pointed as a way of decreasing the complexity of the inversion of matrices through a numeric approximation. This approximation may allow the inversion of large matrices in a feasible time, possibly enabling the use of linear receivers in very large antenna arrays.

Another possibility for achieving very large MIMO systems is to exploit the randomized decoder properties. This algorithm achieves full diversity at little complexity even for large MIMO configurations, so it might possible to take advantage of this fact and increase the dimension of the arrays used in MIMO systems.

References

[1] http://stats.areppim.com/archives/insight_mobilex2012.pdf

[2] Dipankar Raychaudhuri and Narayan B. Mandayam," Frontiers of Wireless and Mobile Communications", Proceedings of the IEEE | Vol. 100, No. 4, April 2012

[3] Cisco: Global Mobile Data traffic Forecast update, 2009-2014

[4] Alain Sibille, Claude Oestges, and Alberto Zanella, Eds., MIMO: From Theory to Implementation. Amsterdam, Netherlands: Academic Press, 2011.

[5] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," Wireless Pers. Commun., vol. 6, pp. 311-335, March 1998.

[6] Arogyaswami Paulraj, Rohit Nabar, and Dhananjay Gore, Introduction to Space-Time Wireless Communications. Cambridge, Cambridge: Cambridge University Press, 2003.

[7] C. E. Shannon A Mathematical Theory of Communication. Bell Sys. Tech. Journal, pp. 379–423, 623–656, 1948.

[8] Andrea Goldsmith, *Wireless Communications*. Cambridge, UK: Cambridge University Press, 2005.

[9] Oded Revev, "Lattices in Computer Science", Tel Aviv University, Fall 2004.

[10] Daniele Micciancio, "Lattices Algorithms and Applications", University of California, San Diego, Winter 2010.

[11] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, "Factoring polynomials with rational coefficients", Math. Ann. vol. 261, no.4, pp. 515–534, 1982.

[12] A. J. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception," Technical Report U.S. Patent, no. 5,345,599, 1994.

[13] C.Windpassinger, "Detection and precoding for multiple input multiple output channels," PhD dissertation, University of Erlangen-Nürnberg, Erlangen, Germany, 2004.

[14] F. A. Monteiro, “Lattices in MIMO Spatial Multiplexing: Detection and Geometry,” PhD dissertation, Department of Engineering, University of Cambridge, May 2012.

[15] H. L. Van Trees, “Detection, Estimation, and Modulation Theory,” volume I, John Wiley & Sons, Inc., New York, NY, USA, 1968.

[16] G. D. Golden, C. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, “Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture,” IET Electronics Letters, vol. 35, no. 1, January 1999.

[17] A. Gorokhov, D. Gore, A. Paulraj, “Performance bounds for antenna selection in MIMO system Communications,” In Proc. IEEE ICC, vol. 5, 3021 –3025, 2003.

[18] Shuiyin Liu, Cong Ling, and Damien Stehlé, "Randomized lattice decoding," IEEE

Transations on information Theory, September 2011

[19] D. Tse and P.Viswanath, Fundaments of Wireless Communications. Cambridge, U.K.:Cambridge Univ. Press, 2005

[20] E.G. Larsson, “NINO detection methods: How they work”, IEEE Signal Process. Mag., vol.26, no.3, pp. 91-95, May 209

[21] J.Jaldén and B. Ottersten, “On the complexity of sphere decoding in digital communications”, IEEE Trans.Signal. Process., vol. 53, no. 4, pp. 1474-1484, Apr. 2005

[22] T.L.Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas”, IEEE Trans. Wireless. Commun., vol.9, no.11, pp.3590-3600, Nov. 2010

[23] A. Tulino and S. Verdu, Random Matrix Theory and Wireless Communications. Foundations and Trends in Communications and Information Theory. Delft, The Netherlands: Now Publishers, Inc., 2004.

[24] A. Chockalingam,”Low-Complexity Algorithms for Large-MIMO Detection,”in Proc. Communications, Control and Signal Processing (ISCCSP), 2010.

[25] F. Rusek, D. Persson,B. K. Lau, E. G. Larsson, T. L. Marzetta,O. Edfors,F. Tufvesson: Scaling up MIMO: opportunities and challenges with very large arrays IEEE Signal Processing Magazine, Vol. 30, No. 1, pp. 40-60, 2013