TÉCNICO LISBOA

Interference Suppression using MIMO and Physical Layer Network Coding

Flávio André Silva Brás

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.Francisco António Taveira Branco Nunes Monteiro

Member of Committee: Prof. Mário Alexandre Teles de Figueiredo

April 2014

To my loved ones

Abstract

This thesis chiefly deals with the concept of interference cancellation when a receiver knows the channel that each of the interfering messages have gone through. This may be the case in "pure" MIMO or when merging the concept of successive interference cancelation (SIC) with physical layer network coding (PLNC), interpreting PLNC as a first stage of SIC.

This work begins by introducing the concept of PLNC through different approaches and proceeds providing an overview of lattices, including basic and useful properties, since lattices take an important place in this dissertation. MIMO detection problem is deeply discussed as a closest vector problem (CVP) in a lattice and several low-complexity sub-optimal receivers are analysed, assessing their performances.

Finally, this dissertation presents a set of new strategies combining MIMO with PLNC in scenarios that move beyond the traditional two-way relay channel (TWRC), including the *two time-slots strategy with MIMO terminals* in a network with a MIMO relay that can always be generalized to scenarios with N terminals, each of which equipped with $N-1$ antennas, and a MIMO relay equipped with N antennas, allowing the exchange of messages in just two time-slots.

Keywords

Lattices, MIMO Detection, Physical Layer Network Coding, Successive Interference Cancelation

Resumo

Esta dissertação lida principalmente com o conceito de cancelamento de interferência quando um receptor sabe o canal pelo qual cada uma das mensagens que interferem passaram. Este pode ser o caso de MIMO "puro" ou de quando se junta o conceito de cancelamento sucessivo de interferência (SIC) com o de *physical layer network coding* (PLNC), interpretando PLNC como uma primeira fase do SIC.

Este trabalho começa com a introdução ao conceito de PLNC através de diferentes abordagens e prossegue dando uma visão geral de *lattices*, incluindo propriedades básicas e úteis, uma vez que estes ocupam um lugar de destaque nesta dissertação. O problema de detecção em MIMO é profundamente discutido como um problema de vector mais próximo (CVP) num *lattice* e vários receptores sub-óptimos de baixa complexidade são analisados, avaliando os seus desempenhos.

Por último, este trabalho apresenta um conjunto de novas estratégias que combinam MIMO com PLNC em cenários que vão para lá do *two-way relay channel* (TWRC), incluindo a estratégia de dois *time-slots* com terminais MIMO numa rede com um *relay* MIMO que pode ser generalizada para cenários com N terminais, cada um dos quais equipado com $N-1$ antenas, e um *relay* MIMO equipado com N antenas, permitindo a troca de mensagens entre os terminais em apenas dois *time-slots*.

Palavras-chave

Lattices, Detecção MIMO, Physical Layer Network Coding, Cancelamento Sucessivo de Interferência

Acknowledgments

First of all, I would like to thank Professor António Rodrigues for giving me the opportunity to work on such a recent research topic, for his support and his good mood.

I am especially grateful to Professor Francisco Monteiro for being always available for any issue concerning this work, for all the many weekly hours of supervision, for the numerous insightful discussions that have so greatly enriched my knowledge of the field, for his contributions and advices which greatly improved the quality of this dissertation and for his time spent reviewing this manuscript as well as for providing me with his valuable feedback. This thesis could not exist without his endless support.

A word of gratitude goes to my peer Filipe Ferreira for his shown interest, push forward attitude and help to materialize the path followed by this dissertation.

I would not be at this stage of my life if it were not for my family. It is not enough to say thank you for all the support but I hope I have made you feel proud.

Along the last years the passage of the days would have been neither so quick nor so enjoyable without a number of friends that I met along the way. They know who they are and surely know that I am close to the deadline. Nevertheless I would like to thank to Miguel Pereira and Igor Duarte for helping me with the review of this dissertation, to Nelson Rodrigues and Pompeu Santos for borrow me their laptops for simulations, to André Vieira and Bruno Neves. Seriously it has been a great time.

Contents

List of Figures

List of Tables

Acronyms

BER	Bit error rate
BLAST	Bell Laboratories Layered Space-Time Architecture
CDMA	Code-division multiple access
CF	Compute-and-forward
CLLL	Complex LLL
CSIR	Channel state information at the receiver
CVP	Closest vector problem
DS	Direct sequence
EM	Electromagnetic
FDMA	Frequency division multiple access
FH	Frequency hopping
GSM	Global System for Mobile Communications
KZ	Korkine-Zolotareff algorithm
LLL	Lenstra Lenstra Lovász algorithm
LRA	Lattice reduction-aided
LST	Layered Space-Time
LTE	Long term evolution
MIMO	Multiple-input multiple-output

ML	Maximum likelihood
MMSE	Minimum mean square error
OFDM	Orthogonal frequency division multiplexing
OSIC	Ordered successive interference cancelation
PLNC	Physical layer network coding
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase-shift keying
SDMA	Space-division multiple access
SIC	Successive interference cancellation
SNR	Signal-to-noise ratio
TDMA	Time division multiple access
TWRC	Two-way relay channel
UMTS	Universal Mobile Telecommunications System
V-BLAST	Vertical Bell Laboratories Layered Space-Time Architecture
WiMAX	Worldwide Interoperability for Microwave Access
ZF	Zero-forcing
ZMSW	Zero-mean spatially white

Symbols and Notation

Symbols

$\mathbf{B}$	Lattice generating matrix
$\mathbf{B}_{\mathrm{red}}$	Reduced lattice matrix
C	Constellation
d	Diversity order
E_s	Average energy of the complex symbol
g_T	Transmit filter
g_R	Receiver filter
$\mathbf{H}$	Channel matrix
$\bar{\mathbf{H}}$	Extended matrix $\mathbf{H}$
$\mathbf{H}_{\mathrm{red}}$	Reduced channel matrix
h_{ij}	Channel fading coefficient
h^{-1}	Matched filter
$\mathbf{I}_N$	Identity matrix $N \times N$
k	Rank of the lattice
L	Number of terminals
$\mathbf{M}$	Unimodular matrix
n	Dimension of the lattice

N_R	Number of receive antennas
N_T	Number of transmit antennas
n	Additive noise vector
T_s	Symbol transmission period
U	Unitary matrix
x	Transmitted vector
$\hat{\mathbf{x}}$	Estimated vector with a certain detection technique
W	Receive filter for a certain detection technique
w_i	Messages from transmitter i
$\boldsymbol{y}$	Received vector
$\bar{\mathbf{y}}$	Extended vector **y**
ϵ	Small real number
Λ	Lattice
σ_n^2	Noise variance
σ_x^2	Variance (or power) of the constellations symbols

Notations

$\mathbb{C}^n$	n-dimensional complex space
$\mathbb{R}^n$	n-dimensional real space
$\mathbb{Z}^n$	n-dimensional vectors with integer coordinates
$\mathcal{L}(\boldsymbol{B})$	Lattice generated by basis **B**
$\det(\mathbf{B})$	Determinant of **B**
$(.)^{\mathcal{H}}$	Hermitian operator

$(.)^T$	Transposition
$P(\mathbf{B})$	Fundamental region of a lattice
$Q_C[.]$	Quantisation to constellation C
$\mathrm{vol}(\Lambda)$	Volume of a lattice
$\nu(\Lambda)$	Voronoi region of a lattice
$\Im(.)$	Imaginary part of a complex
$\Re(.)$	Real part of a complex
$\mathbf{H}^{\dagger}$	Pseudo-inverse matrix
$\mathbf{H}^{\mathcal{H}}$	Hermitian matrix

Chapter 1

Introduction to Physical Layer Network Coding

This chapter gives a brief overview of the work. It is presented an introduction to Physical Layer Network Coding and its basic concept followed by a short overview over the Compute-and-Forward strategy.

1.1 Overview

In recent years mobile communications have known great technological developments that have had important social and economical impacts. This importance has resulted, among many other things, in an exponential growth of the number of wireless devices as well as in increasingly richer multimedia applications which have leading these devices to require higher and higher data rates. These trends, combined with the fact of a limited spectrum, point to the conclusion that interference between devices will be one of the dominant bottlenecks in wireless networking.

At the physical layer of wireless networks all data are transmitted through electromagnetic (EM) waves and this means that when a wireless node transmits, the EM signals are often received by more than one node. At the same time, a receiver may be receiving EM signals by a set of others nodes simultaneously. These characteristics may cause interference among signals and traditionally communications systems designs try to either reduce or avoid it. In Wi-Fi networks, for example, when multiple nodes transmit together, packet collisions occur and none of the packets can be received correctly.

Cellular networks are wireless networks distributed over land areas called cells, each served by at least one base station. In this multiuser communications scenario the spectrum that can be used is limited and, additionally, spectrum licensing is very expensive. Therefore, the design of wireless systems has to be as spectrally efficient as possible. As it was emphasized in the previous paragraph, at the physical and medium access layers the main issue in wireless networks is the management of multiple access and interference. In this scenario the overall resources have to be shared by the users and for doing that there are two natural strategies for separating resources between them, based on the orthogonality principle (in either time or frequency domain), thus avoiding interference between them. One of these strategies is *time division multiple access* (TDMA), which separates the transmissions to and from different users in time, introducing the concept of time-slot as the finest divisible resource allocated to a user. The other one is *frequency division multiple access* (FDMA) which achieves the separation in the frequency domain. These two strategies assure a multiple access of the channel by the users and also separate the uplink (from the terminals to a base station) from the downlink (from the base station to the terminals).Standards such as GSM are a prime example of the implementations of these two strategies.

Transmissions in the same band and overlapping in time are also possible. *Code-division multiple access* (CDMA) base *direct sequence* (DS) spread spectrum consists in assigning different orthogonal spreading sequences to different users, and thus each user ends up using all the bandwidth available. Note that this is *just* another manner of implementing orthogonal signalling. When a user in CDMA is demodulating its data, the other users' signals appear as pseudo white noise. Universal frequency reuse is a key property of CDMA systems because all cells use the same spectrum which

eliminates the need for frequency reuse cell planning. Another way for implementing CDMA is by means of *frequency hopping* (FH). FH is an alternative spread spectrum technique to DS where a signal periodically changes its carrier frequency according to a pseudo-random sequence of different frequencies.

All the strategies mentioned above were initially applied to single-carrier systems. Overtime most wireless systems are adopting *orthogonal frequency division multiplexing* (OFDM) as the modulation scheme, where a symbol stream is parallelised over a given bandwidth using adjacent orthogonal frequencies. This is particularly beneficial for frequency selective wireless channels because each one of the sub-streams is transmitted over a narrow-band where the fading is almost flat.

Long term evolution (LTE) combines ideas related to TDMA and FDMA combined with OFDM using the concept of time-frequency *resource-blocks* that a user can use over time [1].

The advent of *multiple-input multiple-output* (MIMO) was the key technique to increase the spectral efficiently in wireless transmission. These systems take advantage of space-dimension which lead to the concept of *space-division multiple access* (SDMA). Along with physical layer network coding (PLNC), MIMO detection will be the main focus of this dissertation, furthermore a set of new strategies combining these two concepts will be proposed in Chapter 4.

Table 1 summarises which are the wireless strategies used in the most common commercial wireless data standards [2].

Table 1. Wireless data standards and radio strategies used for multiplexing.

Wireless data standard	**Wireless techniques used**
GSM	TDMA / FDMA
UMTS	CDMA / FDMA / MIMO
LTE	OFDM / MIMO / SC-FDMA
Wi-Fi	OFDM / MIMO

The strategies introduced until now face interference as a difficulty to communications. Nevertheless it is actually possible to enable more efficient communications over a network making use of interference in many scenarios. A new concept to further enhance the capacity of wireless networks has recently emerged, it is the so-called physical layer network coding (PLNC). It appears to have been independently proposed by several research groups in 2006: Zhang, Liew and Lam [3],

Popovski and Yomo [4], and Nazer and Gastpar [5]. It was presented as a way to exploit the network coding that occurs in Nature when multiple EM waves come together within the same physical space and they add. This mixing of EM waves is indeed a form of network coding, this time performed by Nature. So these authors instead of considering interference as a difficulty to be avoided, they rather put it to a good use in order to improve throughput.

The scheme in [3] assumes a very simple channel model for intermediate nodes, in which the received signal is a sum of two binary-modulated signals plus a Gaussian noise and intermediate nodes try to decode the modulo-two sum of the transmitted messages. It is proved that this simple strategy significantly improves the throughput of the two-way relay channel (TWRC).

In [4] the main idea is the same but a more general channel model is taking into account and the received signal y is given by $y = h_1x_1 + h_2x_2 + z$, where x_1 and x_2 are the signals, z is the Gaussian noise and h_1 and h_2 are known complex-valued channel gains that captures the effects of fading and imperfect phase alignment. It is shown that, in a large range of signal-to-noise ratios (SNR), the strategy in [4] outperforms conventional relaying strategies (such as amplify-and-forward and decode-and-forward) for a two-way relay channel.

The framework by Nazer and Gastpar [5] moves beyond two-way relay channel and will be presented in the next section.

The TWRC is just one of the many scenarios where the broadcast property of the wireless medium can be exploited via network coding and now it will be used to better illustrate the idea of PLNC, this example first appeared in a paper by Wu et al. in 2004 [6].

Let us consider two nodes which cannot hear the transmissions of each other and in order to communicate they are helped by a relay that can hear and transmit to both (Figure 1.1). A practical example of this configuration is a satellite network in which nodes 1 and 2 are the ground stations, and the relay is the satellite. It is assumed that the nodes share the same frequency band and it is imposed a half-duplex constraint which means that each terminal just can send or receive during a single time-slot.

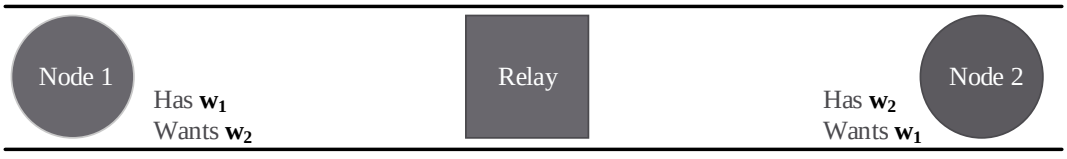

Figure 1.1. Two-way relay channel scheme.

The usual proposal of this situation is the following: node 1 wants to send the message w_1 to node 2 and that node 2 wants to send a message w_2 to node 1, or shortly node 1 and node 2 wants to exchange messages. With the assumptions made if the nodes transmit at the same time the relay will observe a superposition of the two signals corrupted by noise. The traditional scheme with a design

principle that tries to avoid interference and without the use of network coding requires four time-slots to exchange messages as illustrated in Figure 1.2.

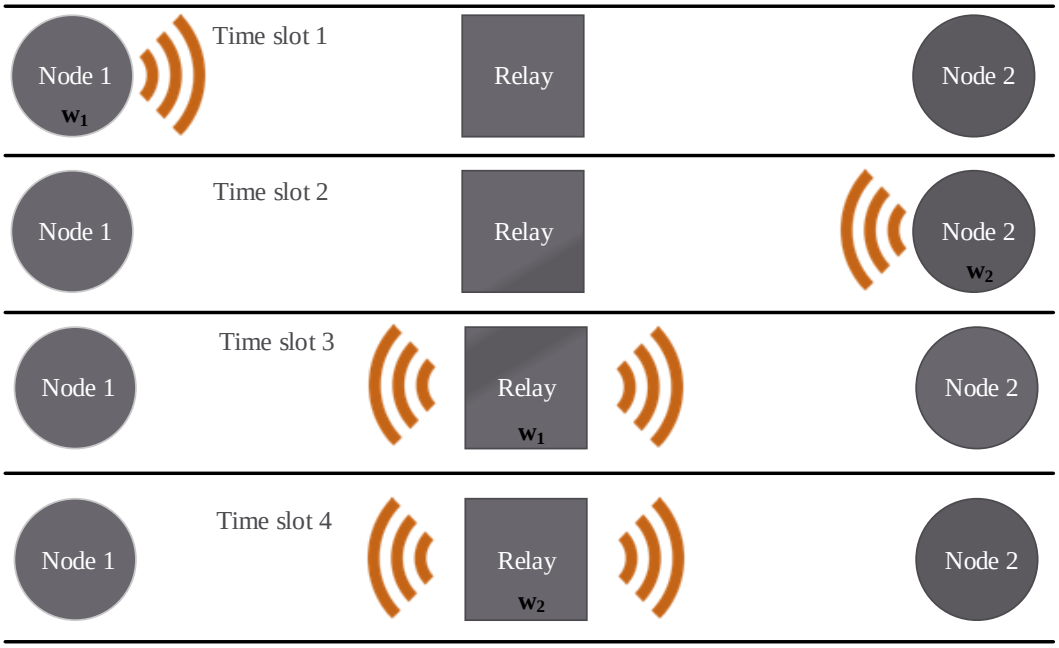

Figure 1.2. A traditional scheme for TWRC.

In this four time-slot strategy each terminal takes one time-slot to transmit its message and then the relay takes more two time-slots to distribute the messages.

By applying pure network coding, the number of time-slots can be reduced to three as illustrated in Figure 1.3. By reducing the number of time-slots from four to three, the use of network coding has a throughput improvement of 33% over the traditional scheme.

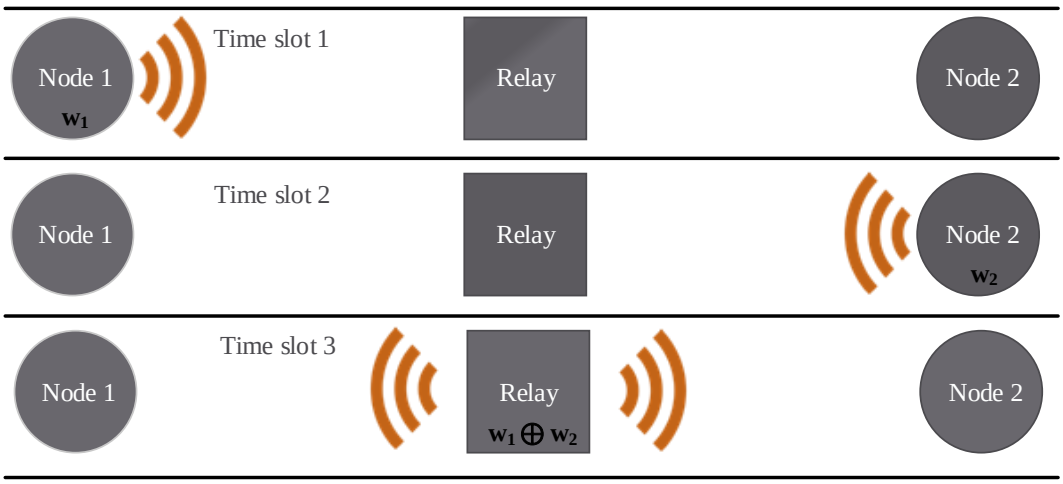

Figure 1.3. A network coding strategy for TWRC.

This strategy is a straightforward way of applying network coding. Initially node 1 sends its message to the relay on time-slot one and then node 2 also sends its message to the relay on the second time-slot. Then the relay compute the sum of the messages ($w_1 \otimes w_2$) and in the third time-slot it sends this sum to both nodes. So its visible that is more efficient if the relay sends the sum of the messages in the broadcast phase what suggest the idea that if it would be possible save time-slots in the multiple-access phase too. Since the relay only needs the sum of the signals we can simultaneously transmit them to the relay at the same time. This is done using PLNC and it reduces the number of time-slots to two as shown in Figure 1.4. It allows nodes 1 and 2 to transmit in the same time-slot and exploits the network coding operation performed by nature where the transmitted signals are added up on the wireless channel. This property can be exploited to send the sum or another linear function to the relay in a single time-slot. By doing so, PLNC can improve the performance of TWRC by 100%.

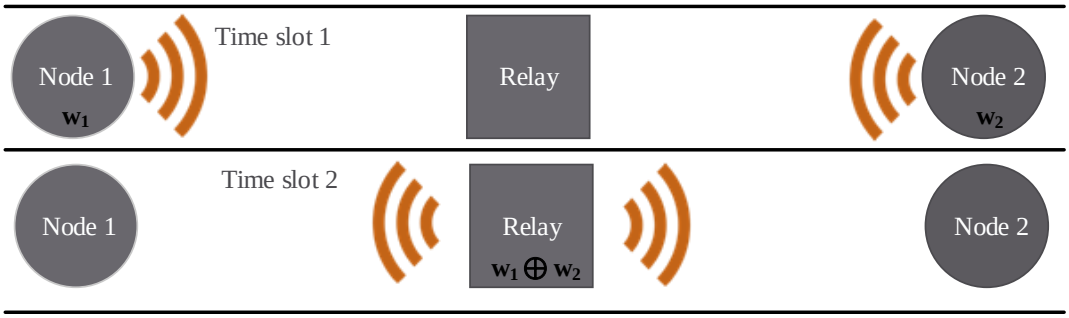

Figure 1.4. A PLNC strategy for the TWRC.

In this strategy of PLNC it is assumed symbol-level and carrier-level synchronization, and also the use of power control in order to packets from nodes 1 and 2 arrive at relay with the same and amplitude.

To date, most works on PLNC have focused on TWRC. However many investigations on its extension to the multi-way relay channel have been made where a relay or a system of relays interconnects more than two end nodes. An example of that is the work of Chen Feng, Danilo Silva and Frank R. Kschischang [7].

As a last note it is important to refer that although PLNC have been originally conceived for application in wireless networks, network coding operations abound in nature. In fact, any physical phenomenon in which an output is the function of a number of inputs can be exploited in the network coding construct. So the application of PLNC could potentially be extended to many other domains such as optical networks [8].

1.2 Compute and Forward

The strategy proposed by Nazer and Gastpar [5] is another branch of PLNC. It is called *compute-and-forward* (CF) and enables relays to decode linear equations of the transmitted messages using noisy linear combinations provided by the channel. If a destination gets sufficiently linear combinations it can solve these linear equations in order to obtain the desired messages.

Compute-and-forward simultaneously affords protection against noise and the opportunity to exploit interference for cooperative gains. This strategy relies on codes with a linear structure, specifically nested lattices codes, which ensures that integer combinations of codewords are themselves codewords [9]. The lattice code should have some form of modulo arithmetic so that we can map between the linear combination taken by the channel and our desired combination over the messages. This property is satisfied by nested lattices which are a subset of another lattice, called the fine lattice. The nested lattice can be replicated tilling the entire fine lattice. Lattices will be presented in Chapter 2, however nested lattices are not discussed so Figure 1.5 illustrates a simple example of its concept through a system were two transmitters send a message at the same time and frequency to a receiver. In Figure 1.5 each transmitter maps its finite-field message into an element of the nested lattice code and sends this vector on the cannel. Here, the channel coefficients (h_1 and h_2) are taken to be equal to 1. Therefore, the receiver observes a noisy sum of the transmitted vectors and determines the closest lattice point. After taking a modulo operation with respect to the nested lattice, the receiver can invert the mapping and determine the modulo sum of the original messages [10].

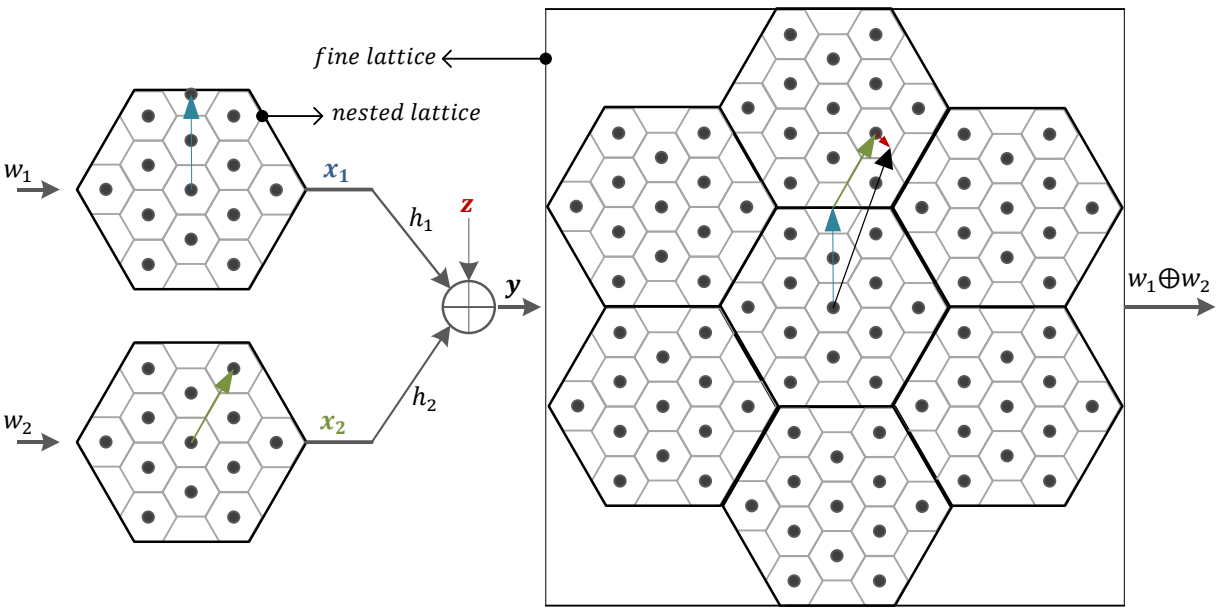

Figure 1.5. Nested lattices in CF strategy.

More generally as depicted in Figure 1.6, considering L terminals and a relay this strategy can be roughly described as follows: each one of the terminals takes messages from a finite field, map them onto nested lattice points, and transmit these across the channel. The relay observes a linear combination of these lattice points and attempts to decode an integer combination of them. This equation of lattice points is finally mapped back to a linear equation over a finite field. When the relay gets L different linear equations it can decode the desired messages [11].

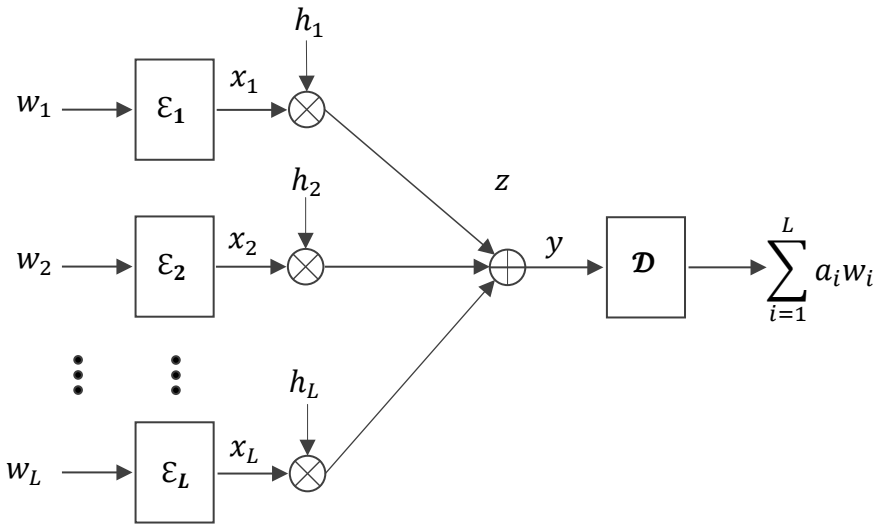

Figure 1.6. Reliably decoding an integer combination of the transmitted messages in CF strategy.

Chapter 2

Lattices

This chapter provides an overview of lattices including basic and useful properties in a more formal definition.

2.1 Context

Lattices are regular arrangements of points in Euclidean space, or in other words, they are a set of points in n-dimensional space with a periodic structure, such as the ones illustrated in Figure 2.1. Three dimensional lattices occur naturally in many settings such as crystals or in a stack of oranges. Historically, lattices were investigated since the late 18th century by mathematicians such as Lagrange, Gauss and later Minkowski, and was about the publication of the last one, Hermann Minkowski, called *Geometrie der Zahlen* in 1896 that lattices have become a standard tool in number theory, especially in the areas of algebraic number theory and the arithmetic theory of quadratic forms for instance. Going forward in time, a significant advance in the algorithmic theory of lattices of general rank occurred in the early 1980's, with the development of the powerful lattice basis reduction algorithm that came to be called the LLL (Lenstra Lenstra Lovász) algorithm that has found numerous applications in both pure and applied mathematics. A complex-value version of this famous algorithm will be presented later on in this chapter.

Lattices have now many applications in computer science and mathematics. Among many others applications they are helpful to the solution of integer programming problems, cryptanalysis, diophantine approximations and the design of error correcting codes for multi-antenna systems. More recently, one of the most promising applications and that has attracted much attention is the use of lattices as a source of computational hardness for the design of secure cryptographic functions.

Another promising application of lattices is its use as basis for coding schemes, and that application is the one that will be explored. In the present work lattices take a prominent place as it will become clearer in the next chapters so it is important to define this structure and understand its main properties.

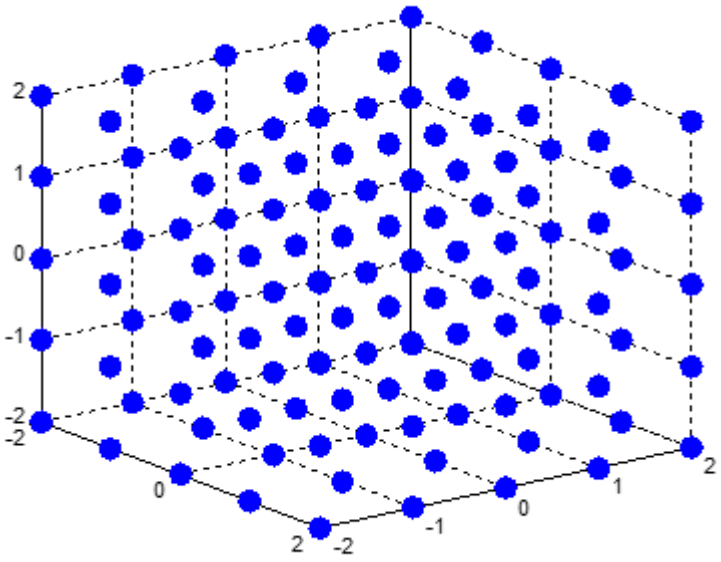

Figure 2.1. Lattice illustrations lattice in $\mathbb{R}^3$

2.2 Basic Definitions

In this subsection it will be presented a more formal definition of lattice as well as its most important mathematical properties based on [12] and [13].

2.2.1 Lattice

There are several ways of specifying a lattice. One of the definitions says that a lattice is a discrete additive subgroup of $\mathbb{R}^n$, i.e., it is a subset $\Lambda \subseteq \mathbb{R}^n$ satisfying the following properties:

- (group) Λ is closed under addition and subtraction,
- (discrete) There is an $\epsilon > 0$ such that any two distinct lattice points $x \neq y \in \Lambda$ are at distance at least $\|x - y\| \geq \epsilon$.

Note that not every subgroup of $\mathbb{R}^n$ is a lattice. For example the subgroup $\mathbb{Q}^n$ is not a lattice because it not fulfils the second property while the group $\mathbb{Z}^n$ is a lattice because integer vectors can be added and subtracted obtaining again an integer vector, and clearly the distance between any two integer vectors is at least one.

An equivalent definition of lattices can be obtained from $\mathbb{Z}^n$ by applying a linear transformation. To illustrate this definition let's consider a matrix $\mathbf{B} \in \mathbb{R}^{n \times k}$ that has full column rank, what means that the columns of this matrix are linearly independent, and then $\mathbf{B}(\mathbb{Z}^n) = \{\mathbf{Bx}: \mathbf{x} \in \mathbb{Z}^n\}$ is also a lattice. Obviously this set is closed under addition and subtraction, besides it is also discrete. Moreover all lattices can be expressed as $\mathbf{B}(\mathbb{Z}^n)$ for some $\mathbf{B}$ what leads to the following definition.

Let $\mathbf{B} = [\mathbf{b_1}, \dots, \mathbf{b_k}] \in \mathbb{R}^{n \times k}$ where $\mathbf{b_i}$ are linearly independent vectors in $\mathbb{R}^n$. The lattice generated by $\mathbf{B}$ is the set of all the integer linear combinations of the columns of $\mathbf{B}$,

$$\mathcal{L}(\mathbf{B}) = \{\mathbf{Bx} : \mathbf{x} \in \mathbb{Z}^k\} = \left\{\sum_{i=1}^{k} x_i b_i : x_i \in \mathbb{Z}\right\}. \tag{2.1}$$

The matrix $\mathbf{B}$ is called a basis for the lattice $\mathcal{L}(\mathbf{B})$. The integer n is called the dimension of the lattice; the integer k is the rank of the lattice and in the case of $n = k$, $\mathcal{L}(\mathbf{B})$ is called a full rank lattice.

In this second definition lattices can be represented by a basis matrix $\mathbf{B}$, containing integer or rational entries, which generate the lattice. It is important to keep in mind this definition because it will be present during all the work. The definition can be extended to complex lattices, however it is possible to transform any complex lattice into a real lattice through the real equivalent model, so the descriptions about lattices will continue with real lattices.

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

$$\Lambda = \mathcal{L}(\mathbf{B}) = \{\mathbf{Bx} : \mathbf{x} \in \mathbb{Z}^k\} \tag{2.2}$$

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

$$\text{span}(\mathbf{B}) = \{\mathbf{By} : \mathbf{y} \in \mathbb{R}^k\}. \tag{2.3}$$

The central difference is that in a lattice *only* integer coefficients are allowed, resulting in a discrete set of points. As vectors $\mathbf{b_1}, \dots, \mathbf{b_n}$ are linear independent, any point $y \in \text{span}(\mathbf{B})$ can be written as a linear combination $\mathbf{y} = \mathbf{x_1 b_1} + \cdots + \mathbf{x_n b_n}$ in a unique way. Therefore $y \in \mathcal{L}(\mathbf{B})$ if and only if $\mathbf{x_1}, \dots, \mathbf{x_n} \in \mathbb{Z}$.

Notice that the definition $\mathcal{L}(\mathbf{B}) = \{\mathbf{Bx} : \mathbf{x} \in \mathbb{Z}^m\}$ can be extended to matrices B 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.

2.2.2 Examples

In order to illustrate the next definitions it will be shown some examples. In the following examples the fulfilled points are the ones that belong to the lattice. Besides these examples only contain a part of the lattice but is intuitive to imagine how they span to the remaining space. Figure 2.2 a) represents a lattice generated by the vectors $(1,0)^{\mathrm{T}}$ and $(0,1)^{\mathrm{T}}$ which is the lattice of all integers' points, $\mathbb{Z}^2$. As it is possible to see in Figure 2.2 b) the basis $(1,1)^{\mathrm{T}}$ and $(2,1)^{\mathrm{T}}$ also generate $\mathbb{Z}^2$ moreover these bases are not unique, fact that will be explained later on in this chapter on section 2.2.7 (Bases). On the other hand Figure 2.2 c) doesn't generate $\mathbb{Z}^2$, its basis $(1,1)^{\mathrm{T}}$ and $(2,0)^{\mathrm{T}}$ generate instead a lattice of all integer points whose coordinates sum to an even number. Apart from the other examples that are full-rank lattices Figure 2.2 d) represents a lattice of dimension 2 and rank 1 generated by the base $(2,1)^{\mathrm{T}}$.

2.2.3 Fundamental Region

For any lattice basis B the fundamental region or parallelepiped is defined as

$$P(\mathbf{B}) = \{\mathbf{Bx} \mid \mathbf{x} \in \mathbb{R}^n, \forall i: 0 \leq \mathbf{x}_i < 1\}. \tag{2.4}$$

Notice that $P(\mathbf{B})$ depends on the basis **B** and it is easy to imagine that if we place one copy of $P(\mathbf{B})$ at each lattice point in $\mathcal{L}(\mathbf{B})$ we obtain a tiling of the entire $\text{span}(\mathcal{L}(\mathbf{B}))$ as shown in Figure 2.3.

In Figure 2.2 and Figure 2.3 the fundamental regions are represented by the shaded areas. In Figure 2.3 are used two different shades just for a better illustration of the tilling of the space but they represent the same area, i.e. the same fundamental region.

Based on the definition of the fundamental region it is possible to determine if a set of vectors forms a basis of a lattice. As it shown above in the examples, not every set of n linearly vectors in $\mathbb{Z}^n$ is a basis of $\mathbb{Z}^n$ and this is possible to determine following the next *lemma*: the fundamental region generated by the vectors should not contain any lattice points, except the origin. For example, notice that the fundamental regions in Figure 2.2 a) and Figure 2.2 b) do not contain any nonzero lattice points and so they are bases of $\mathbb{Z}^2$, however the fundamental region of Figure 2.2 c) contains the lattice point $(1,0)$ and so the vectors that generate this lattice are not a basis of $\mathbb{Z}^2$.

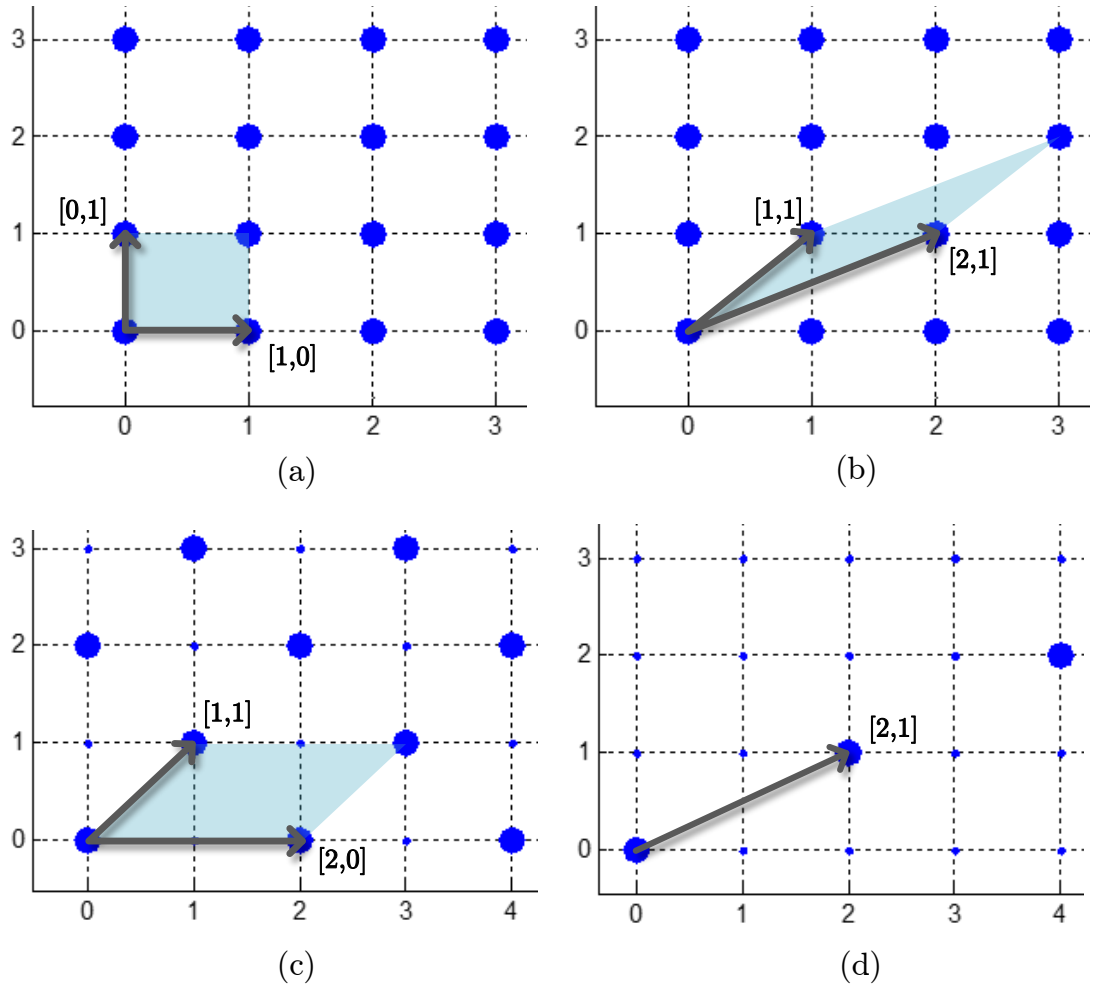

Figure 2.2. Examples of lattices in $\mathbb{R}^2$. (a) Lattice generated by basis $(1,0)^T$ and $(0,1)^T$. (b) Lattice generated by basis $(1,1)^T$ and $(2,1)^T$. (c) Lattice generated by basis $(1,1)^T$ and $(2,0)^T$. (d) Lattice generated by basis $(2,1)^T$.

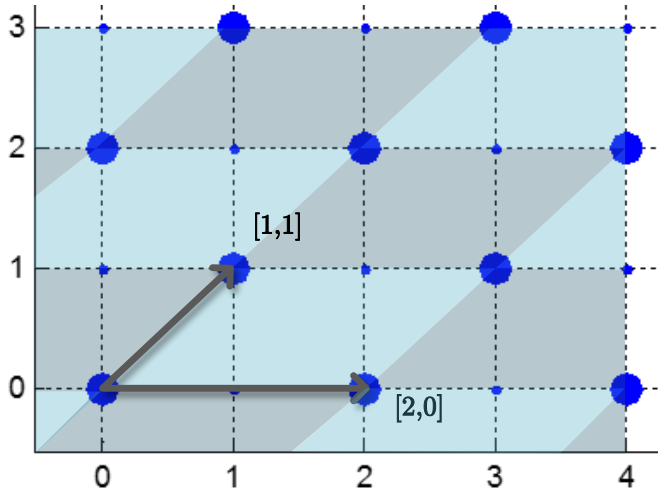

Figure 2.3. Tilling of the span$(\mathcal{L}(\mathbf{B}))$ with the fundamental region $P(\mathbf{B})$.

2.2.4 Voronoi Region

The Voronoi region is defined by

$$\nu(\Lambda) = \{z \in \text{span}(\mathbf{B}) : \|x - z\| < \|y - z\|, \forall y \in \Lambda\}. \tag{2.5}$$

This equations defines the Voronoi region, which consists of the space where the lattice exists 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. This region is a characteristic of the lattice and independent of any particular generating matrix and it is the most interesting fundamental region that tiles the entire space once it constitutes the optimal decision region for the *closest vector problem* (CVP) in a lattice, a problem that will be discussed in the next chapter.

Figure 2.4 shows the tilling of two distinct lattices with its respective Voronoi region. As it is intuitive from the figure the Voronoi regions are limited by the solid lines and they can assume many forms.

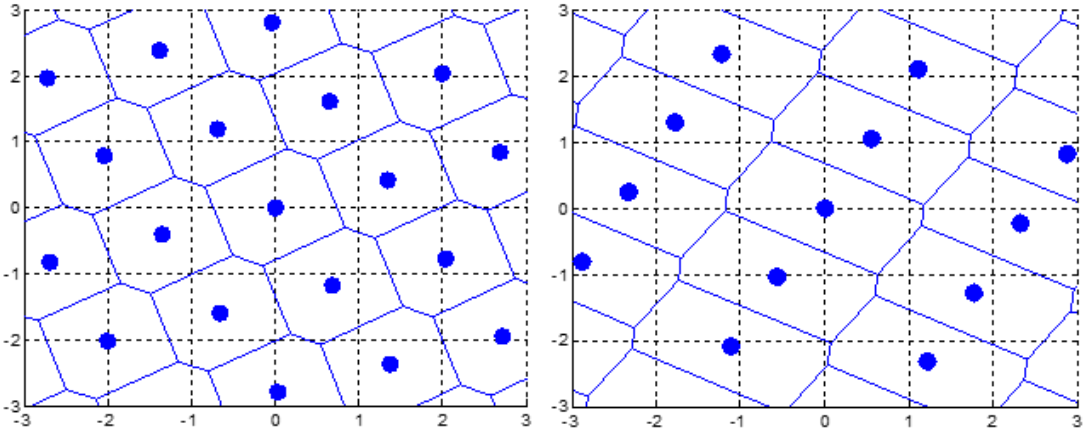

Figure 2.4. Illustrations of Voronoi regions of two distinct lattices.

2.2.5 Volume

In the case of full-rank lattices the volume of the lattice (that it is the same of the volume of any of its many possible fundamental regions) is

$$\mathrm{vol}(\Lambda) = |\det(\mathbf{B})|, \tag{2.6}$$

however, in general the following expression is required:

$$\mathrm{vol}(\Lambda) = \sqrt{\det(\mathbf{B}^{\mathrm{T}}\mathbf{B})}. \tag{2.7}$$

In the complex case the Hermitian operator replaces transposition in the equation above. The volume of the lattice is an invariant of the lattice, i.e., is independent of the choice of basis.

2.2.6 Determinant

Let $\Lambda = \mathcal{L}(\mathbf{B})$ be a lattice of rank n. The determinant of Λ, denoted $\det(\Lambda)$, as the n-dimensional volume of the fundamental region $P(\mathbf{B})$. In symbols, this can be written as $\det(\Lambda) := \sqrt{\det(\mathbf{B}^{\mathrm{T}}\mathbf{B})}$. In the special case that Λ is a full rank lattice, B is a square matrix, and it stays that $\det(\Lambda) = |\det(\mathbf{B})|$.

The determinant of a lattice is an invariant of the lattice, in the sense that it is also independent of the choice of basis $\mathbf{B}$. Indeed, if $\mathbf{B}_1$ and $\mathbf{B}_2$ are two bases of Λ, then as it will be presented next $\mathbf{B}_1 = \mathbf{B}_2\mathbf{U}$ for some unimodular matrix $\mathbf{U}$. Hence,

$$\sqrt{\det(\mathbf{B}_2^{\mathrm{T}}\mathbf{B}_2)} = \sqrt{\det(\mathbf{U}^{\mathrm{T}}\mathbf{B}_1^{\mathrm{T}}\mathbf{B}_1\mathbf{U})} = \sqrt{\det(\mathbf{B}_1^{\mathrm{T}}\mathbf{B}_1)} \tag{2.8}$$

The determinant of a lattice is inversely proportional to its density, what means that the smaller the determinant, the denser the lattice is.

2.2.7 Bases

In the sub-section 2.2.3 (Fundamental Region) it was presented a *lemma* to verify if a basis can be a basis of a certain lattice based on the fundamental region that the vectors of the basis create. Now let us present it in a more formal way.

Let Λ be a lattice of rank n, and let $\mathbf{b}_1, \ldots, \mathbf{b}_n \in \Lambda$ be n linearly independent lattice vectors. Then $\mathbf{b}_1, \ldots, \mathbf{b}_n$ form a basis of Λ if and only if $P(\mathbf{b}_1, \ldots, \mathbf{b}_n) \cap \Lambda = \{0\}$.

Proof. Assume first that $\mathbf{b}_1, \ldots, \mathbf{b}_n$ form a basis of Λ. Then, by definition, Λ is the set of all their integer combinations. Since $P(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ is defined as the set of linear combinations of

$\mathbf{b}_1, \ldots, \mathbf{b}_n$ with coefficients in the range $[0,1[$, the intersection of the two sets is $\{0\}$.

For the other direction, assume that $P(\mathbf{b}_1, \ldots, \mathbf{b}_n) \cap \Lambda = \{0\}$. Since Λ is a rank n lattice and $\mathbf{b}_1, \ldots, \mathbf{b}_n$ are linearly independent, we can write any lattice vector $\mathbf{x} \in \Lambda$ as $\sum \mathbf{y}_i \mathbf{b}_i$ for some $\mathbf{y}_i \in \mathbb{R}^n$. Since by definition a lattice is closed under addition, the vector $\mathbf{x}' = \sum(\mathbf{y}_i - \lfloor \mathbf{y}_i \rfloor)\mathbf{b}_i$ is also in Λ. By our assumption, $\mathbf{x}' = 0$. This implies that all $\mathbf{y}_i$ are integers hence $\mathbf{x}$ is an integer combination of $\mathbf{b}_1, \ldots, \mathbf{b}_n$.

Note that different set of vectors may generate the same lattice but all these different bases are related by unimodular transformations as it will be described below.

Let B_1 and B_2 be two bases. Then $\mathcal{L}(\mathbf{B}_1) = \mathcal{L}(\mathbf{B}_2)$ if and only if there exists a unimodular matrix $\mathbf{U}$ such that $\mathbf{B}_1 = \mathbf{B}_2\mathbf{U}$.

A matrix $\mathbf{U} \in \mathbb{Z}^{n \times n}$ is called unimodular if it is a square matrix with integer entries and determinant ± 1. The inverse of a unimodular matrix is also unimodular.

Proof. First assume $\mathbf{B}_1 = \mathbf{B}_2\mathbf{U}$ for some unimodular matrix $\mathbf{U}$. Notice that if $\mathbf{U}$ is unimodular, then $\mathbf{U}^{-1}$ is also unimodular. In particular, both $\mathbf{U}$ and $\mathbf{U}^{-1}$ are integer matrices, and $\mathbf{B}_1 = \mathbf{B}_2\mathbf{U}$ and $\mathbf{B}_2 = \mathbf{B}_1\mathbf{U}^{-1}$. It follows that $\mathcal{L}(\mathbf{B}_1) \subseteq \mathcal{L}(\mathbf{B}_2)$ and $\mathcal{L}(\mathbf{B}_2) \subseteq \mathcal{L}(\mathbf{B}_1)$, i.e., the two bases $\mathbf{B}_1$ and $\mathbf{B}_2$ are equivalent and they generate the same lattice.

Now assume $\mathbf{B}_1$ and $\mathbf{B}_2$ are two bases for the same lattice $\mathcal{L}(\mathbf{B}_1) = \mathcal{L}(\mathbf{B}_2)$. Then, by definition of lattice, there exist integer square matrices $\mathbf{V}$ and $\mathbf{W}$ such that $\mathbf{B}_1 = \mathbf{B}_2\mathbf{W}$ and $\mathbf{B}_2 = \mathbf{B}_1\mathbf{V}$. Combining these two equations appears $\mathbf{B}_1 = \mathbf{B}_1\mathbf{V}\mathbf{W}$, or equivalently, $\mathbf{B}_1(\mathbf{I} - \mathbf{V}\mathbf{W}) = 0$. Since vectors $\mathbf{B}_1$ are linearly independent, it must be $\mathbf{I} - \mathbf{V}\mathbf{W} = 0$, i.e., $\mathbf{V}\mathbf{W} = \mathbf{I}$. In particular, $\det(\mathbf{V}) \cdot \det(\mathbf{W}) = \det(\mathbf{V} \cdot \mathbf{W}) = \det(\mathbf{I}) = 1$. Since matrices V and W have integer entries, $\det(\mathbf{V}), \det(\mathbf{W}) \in \mathbb{Z}$, and it must be $\det(\mathbf{V}) = \det(\mathbf{W}) = \pm 1$.

A simple way to obtain a basis of a lattice from another is to apply (a sequence of) elementary column operations, as defined bellow. It is easy to see that elementary column operations do not change the lattice generated by the basis because they can expressed as right multiplication by a unimodular matrix. Elementary column operations are:

- Swap the order of two columns in $\mathbf{B_1}$.
- Multiply a column by -1.
- Add an integer multiple of a column to another column: $\mathbf{b}_\mathrm{i} \leftarrow \mathbf{b}_\mathrm{i} + \mathrm{a} \cdot \mathbf{b}_\mathrm{j}$ where $i \neq j$ and $a \in \mathbb{Z}$.

Moreover, any unimodular transformation can be expressed as a sequence of elementary integer column operations.

2.2.8 Successive Minima and Shortest Vector

One basic parameter of a lattice is the length, meaning the Euclidean norm, of the shortest nonzero vector in the lattice. This parameter is denoted by λ_1.

An equivalent way to define λ_1 is the following. It is the smallest r such that the lattice points inside a ball of radius r span a space of dimension 1 as is shown in Figure 2.5. This definition leads to the following generalization of λ_1 known as successive minima.

Let Λ be a lattice of rank n. For $i \in \{1, \dots, n\}$ it is define the ^{ith} successive minimum as

$$\lambda_i(\Lambda) = \inf \left\{ r | \dim \left(\text{span}\left(\Lambda \cap \overline{\mathbf{B}}(0, r)\right)\right) \geq i \right\} \tag{2.9}$$

where $\overline{\mathbf{B}}(0, r) = \{\mathbf{x} \in \mathbb{R}^m | \|\mathbf{x}\| \leq r\}$ is the closed ball of radius r around 0.

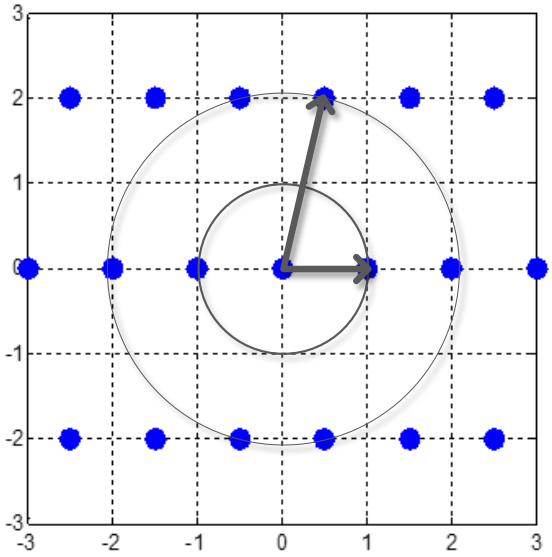

Figure 2.5. Shortest vector $\lambda_1(\Lambda) = 1$

2.3 Lattice Reduction

The idea of lattice reduction consists in changing a basis **B** of a lattice Λ into a shorter basis $\mathbf{B}_{\text{red}}$ such that Λ remains the same. This process can be used to solve the shortest vector problem however for high rank basis there is no known algorithm that finds the shortest vector in polynomial time.

Lattice reduction 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 [14] and is a polynomial time lattice reduction algorithm. LLL usually obtains an approximation for the shortest vector but as it was mentioned there is not any efficient algorithm to solve this problem anyway the approximation obtain by the LLL is enough for many applications. In this dissertation a complex version of the original LLL algorithm (CLLL) will be used to assist a detection process in a receiver (see section 3.5).

In sub-section 2.2.7 were presented simple techniques to obtain different bases that generate the same lattice, where one of those is to swap the order of two vectors in **B** which is equivalent to apply a unimodular transformation. Roughly speaking what LLL algorithm does is to perform successive orthogonal projections and if necessary uses the technique of swapping two consecutives vectors of **B**, in order to get a reduced or near orthogonal basis $\mathbf{B}_{\text{red}}$. So the output of the algorithm is a new basis $\mathbf{B}_{\text{red}}$ consisting of near-orthogonal vectors and a unimodular matrix **M** such that

$$\mathbf{B}_{\text{red}} = \mathbf{BM}.$$

In this dissertation, as mentioned, and based on [15] a complex version of this algorithm was adopted since it reduces the complexity of the implementation without sacrificing any performance. The detailed pseudo-code, using *Matlab* notation, of the algorithm implemented is presented in Table 2. Notice that parameter δ controls the performance and the complexity of the algorithm, i.e., higher values of δ corresponds to higher complexity leading to a better performance. At the simulations performed along this dissertation δ was assumed to be 0.75 as advised in literature. Since in the following chapters basis **B** of the lattice will be considered as the matrix of the channel coefficients **H** (see section 3.1) the pseudo-code will use the latter nomenclature.

Table 2. Pseudo-code of the complex LLL algorithm

Input: $\mathbf{H}$

Output: $\mathbf{M}$ and $\mathbf{H}_{\mathrm{red}}$

$[\mathbf{Q}, \mathbf{R}] =$ QR decomposition $(\mathbf{H})$
$\delta = 0{,}75$
$m = \mathrm{size}(\mathbf{H}, 2)$
$\mathbf{M} = \mathbf{I}_{\mathrm{m}}$
$k = 2$
while $k \leq m$
    for $n = k - 1 : -1 : 1$
        $u = \mathrm{round}((\mathbf{R}(n,k)/\mathbf{R}(n,n))$
        if $u \sim = 0$
            $\mathbf{R}(1:n,k) = \mathbf{R}(1:n,k) - u \cdot \mathbf{R}(1:n,n)$
            $\mathbf{M}(:,k) = \mathbf{M}(:,k) - u \cdot \mathbf{M}(:,n)$
        end
    end
    if $\delta|\mathbf{R}(k-1,k-1)|^2 > |\mathbf{R}(k,k)|^2 + |\mathbf{R}(k-1,k)|^2$
        swap the $(k-1)^{\mathrm{th}}$ and k^{th} columns in $\mathbf{R}$ and $\mathbf{M}$
        $\boldsymbol{\Theta} = \begin{bmatrix} \alpha^* & \beta \\ -\beta & \alpha \end{bmatrix}$ where $\alpha = \frac{\mathbf{R}(k-1,k-1)}{\|\mathbf{R}(k-1:k,k-1)\|}$ and $\beta = \frac{\mathbf{R}(k,k-1)}{\|\mathbf{R}(k-1:k,k-1)\|}$
        $\mathbf{R}(k-1:k, k-1:m) = \boldsymbol{\Theta}\mathbf{R}(k-1:k, k-1:m)$
        $\mathbf{Q}(:, k-1, k) = \mathbf{Q}(:, k-1:k)\boldsymbol{\Theta}^{\mathcal{H}}$
        $k = \max(k-1, 2)$
    else
        $k = k + 1$
    end
end
$\mathbf{H}_{\mathrm{red}} = \mathbf{QR}$

Chapter 3

MIMO Detection

This chapter presents the MIMO detection problem as a closest vector problem (CVP) in a lattice and describes and analyses several low-complexity sub-optimal receivers, assessing their performances.

3.1 MIMO Spatial Multiplexing

As it was described in Chapter 1 system designers are facing a number of challenges. The limited availability of the radio frequency spectrum, a complex space-time varying wireless environment and the increasing demand for higher data rates are some of these challenges. In recent years *Multiple-Input Multiple-Output* (MIMO) systems have emerged as one of the most promising technologies to answer to these challenges. MIMO communications systems can be defined by the use of multiple antennas in the transmitting node as well as in the receiving node. The core idea behind this strategy is that signals sampled in the spatial domain at both ends are combined in such a way that they either create effective multiple parallel spatial data pipes, increasing the data rate, and/or adding diversity to improve the quality of the communications by reducing the *bit-error rate* (BER).

MIMO spatial multiplexing has indeed allowed unprecedented spectral efficiencies in wireless fading channels achieving high date-rates. However this gain of performance comes at a price that is the high complexity of the detection in the receivers. This complexity has posed a great challenge to decoder's implementation and the present chapter will focus on the most part 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) receiver and finally receivers with lattice reduction, but before entering into the specifications of both the receivers and the work developed itself it is important to present a background on MIMO as well as some models and problems definitions.

The use of multiple antennas at both the transmit and receive nodes has become one of the most important paradigms for the deployment of existing and emerging wireless communications systems. The importance of MIMO systems is witnessed by their presence in many recent standards. MIMO along with *orthogonal frequency division multiplexing* (OFDM) technology sustains the physical layer of the fourth generation (4G) wireless networks such as LTE and also the LTE-Advanced release, the latest generation of Wi-Fi, IEEE 802.11.n and Worldwide Interoperability for Microwave Access, known as WiMAX (IEEE 802.16m standard).

Clearly, the benefits from multiple antennas arise from the use of the space dimension that comes as a complement to time and because of that MIMO technology is also known as *space-time* wireless or *smart* antennas. Still MIMO concept is not a new idea. Until the 1990s, the use of antenna arrays at one node of the link was mainly oriented to the estimation of directions of arrival as well as diversity, leading to beamforming and spatial diversity. Beamforming is a powerful technique which increases the link signal-to-noise ratio (SNR) through focusing the energy into desired directions. The concept of spatial diversity is that, in the presence of random fading caused by multipath propagation, the SNR is significantly improved by combining the output of decorrelated antenna elements. In the early 1990s new proposals for using antenna arrays to increase the capacity of wireless links enhanced

enormous opportunities beyond diversity alone. It turned out that diversity was only the first step towards mitigating multipath propagations. With the emergence of MIMO systems, multipath was effectively converted into a benefit for the communications system. MIMO takes advantage of random fading, and possibly delay spread, to multiply transfer rates [2].

One of the first examples of practical application of MIMO is the patent of Paulraj and Kailath [16] which introduced a technique for increasing the capacity of a wireless link using multiple antennas at both ends for application to broadcast digital TV.

Another relevant contribution appeared in 1996 in the paper of Foschini [17] where he introduces the concept of *Layered Space-Time* (LTS) architecture. This architecture was later referred as *Bell Laboratories Layered Space-Time Architecture* (BLAST) and it was designed for a point-to-point MIMO communication system where the data stream generated by the source was divided in several branches and encoded without sharing any information with each other. To solve some limitations of its first release a new version of BLAST architecture was proposed in [18] and it was called *Vertical Bell Laboratories Layered Space-Time Architecture* (V-BLAST). One of the most important results of this kind of architecture is that, under some fading conditions, independent data streams can be simultaneously transmitted over the matrix channel **H** [19]. Motivated by the results of Foschini, a large number of papers appeared in the open literature addressing the different aspects of the MIMO architectures. Further progress in MIMO concepts came up with the paper of Viswanath, Tse, and Anantharam [20], which was one of the first contributions addressing MIMO multiple-access channels.

Other important topics in wireless communications have been investigated since 2000 but the interest in MIMO topics has always been high. While earlier MIMO work focused on single user applications (transmitter to single receiver or vice versa), MIMO are now expanding to multiuser (transmitters to multiple receivers) and network applications (multi-transmitters to a single receiver). These applications offer new challenges in coding channel spanning models to transmit as well as decoding techniques at the receiver's end [21].

Nowadays *large* MIMO systems and *massive* MIMO systems are a promising concept that has attracted the research community, where the number of antennas is large (> 8) in the former and very large in the latter. This approach uses compact antennas and claims to support huge performance gains while still allowing fast iterative receiver decoding [22].

The type of MIMO system that will be discussed in this chapter is a point-to-point communications MIMO, as illustrated in Figure 3.1. 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.

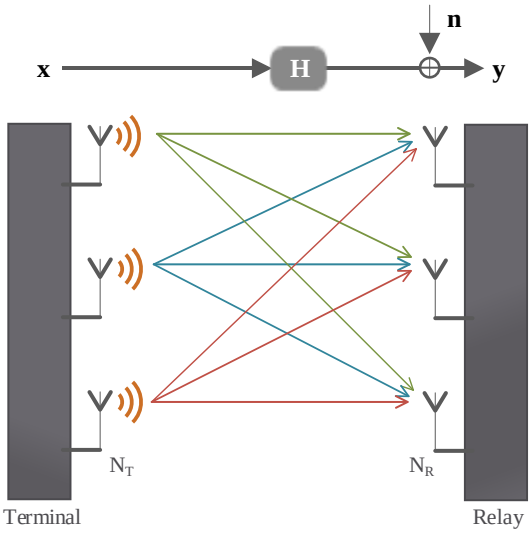

Figure 3.1. Point-to-point MIMO system.

This kind of setup is chiefly considered for semi-mobile local-area wireless data communications. As an example, the reader could think of a laptop computer equipped with a set of antennas communicating with an access point, also equipped with several antennas.

In this MIMO communications system with N_T transmit antennas and N_R receive antennas (with $N_R \geq N_T$ 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

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{3.1}$$

where $\mathbf{y} = [y_1, \dots, y_{N_R}]^T \in \mathbb{C}^{N_R \times 1}$ is the received signals vector and $\mathbf{x} = [x_1, \dots, x_{N_T}]^T \in \mathbb{C}^{N_T \times 1}$ is the transmitted signals vector. The radio links between each pair of transmit and receive antennas are represented by the channel matrix $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ 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 $\mathbf{n} = [n_1, \dots, n_{N_R}]^T \in \mathbb{C}^{N_R \times 1}$ 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 [23].

Full channel knowledge at the receiver is required in the model assumed in this chapter, usually known as *channel state information at the receiver* (CSIR). However, acquiring this knowledge is not a straightforward task, especially in rapidly time-varying channels, but this matter is beyond the scope of this work.

Throughout this dissertation, square quadrature amplitude modulation (QAM) constellations are assumed. Specifically it will be considered M-QAM constellations with $M = 4, 16, 64$ (see Figure 3.2), 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

$$C = \{-(\sqrt{M} - 1), \dots, -3, -1, +1, +3, \dots, +(\sqrt{M} - 1)\}. \tag{3.2}$$

The simulations conducted in this work are also valid for quadrature phase-shift keying (QPSK) modulation since the resulting modulated radio waves are exactly the same of 4-QAM constellation, although the root concepts of QPSK and 4-QAM are different.

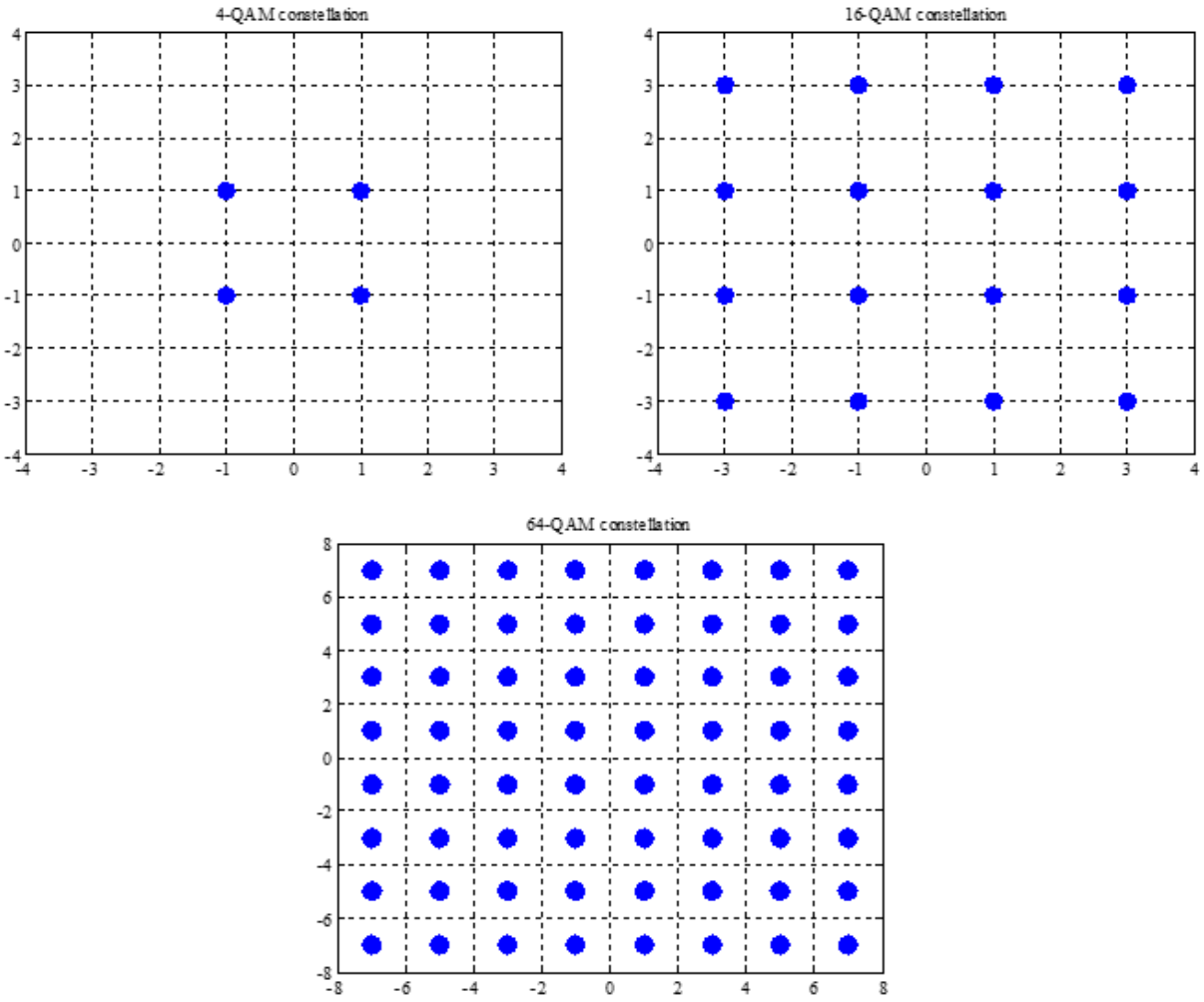

Figure 3.2.Illustration of the M-QAM constellations.

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

$$E_s = \frac{1}{M}\sum_{x_i \in C} |x_i|^2 \tag{3.3}$$

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., $E\left[\left|h_{ij}\right|^2\right] = 1$, on *average* it is valid to calculate the SNR at the receiver as

$$\frac{E\{\|\mathbf{y}\|^2\}}{E\{\|\mathbf{n}\|^2\}} = \frac{E\{\|\mathbf{Hx}\|^2\}}{E\{\|\mathbf{n}\|^2\}} = \frac{E\left[\sum_{i=1}^{N_R}\sum_{j=1}^{N_T}\left|h_{ij}x_j\right|^2\right]}{E\left[\sum_{i=1}^{N_R} n_i^2\right]} = \frac{N_T N_R \sigma_x^2}{N_R \sigma_n^2} = N_T \frac{\sigma_x^2}{\sigma_n^2}. \tag{3.4}$$

Throughout this dissertation the performance results will be plotted as *symbol error rate* (SER) as function of the SNR defined in (3.4).

In order to evaluate the performance of the receivers it will be considered the *diversity order*, d, or simply the *slope*. This parameter describes how the SER decreases with the increase of SNR, i.e., when plotting the SER as a function of the SNR this *diversity order* is simply the slope of the SER curve and it can be obtained by

$$d = \frac{\log(\text{SER})}{\log(\text{SNR})}. \tag{3.5}$$

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

Table 3. Average symbol energy of modulations.

	4-QAM	**16-QAM**	**64-QAM**
$\mathbf{E_s}$	2	10	42

3.1.1 The Real Equivalent Model

The model for spatial multiplexing described in (3.1) is in the complex domain. In order to use the LLL algorithm, which works under real vector spaces, the implementation of the system model started based on the real equivalent model [23].

This model consists on stacking the real and complex parts of the vectors. Besides that an appropriate construction of the channel matrix is needed. With this the model equation (3.1) can equivalently be described only by real vectors as

$$\begin{bmatrix}\Re(\mathbf{y})\\ \Im(\mathbf{y})\end{bmatrix} = \begin{bmatrix}\Re(\mathbf{H}) & -\Im(\mathbf{H})\\ \Im(\mathbf{H}) & \Re(\mathbf{H})\end{bmatrix}\begin{bmatrix}\Re(\mathbf{x})\\ \Im(\mathbf{x})\end{bmatrix} + \begin{bmatrix}\Re(\mathbf{n})\\ \Im(\mathbf{n})\end{bmatrix} \tag{3.6}$$

where $\Re$ and $\Im$ denote the real and imaginary parts, respectively.

With the implementation of CLLL algorithm the equivalent real model was leaved behind and the simulations of all the receivers were done in the complex domain, since it reduces the complexity of the implementation without sacrificing any performance.

3.1.2 The Closest Vector Problem

One of the main problems considered in this thesis is the detection of a vector **x** given the noisy observation **y**. This topic has been a main research problem in spatial multiplexing and assuming that all vectors **x** are equiprobable this detection can be seen as *maximum likelihood* (ML) detection which will be analysed later on. In lattice theory this problem is known as *closet vector problem* (CVP). As it is intuitive to understand in the model adopted in (3.1), the received **y** is a point of a lattice displaced from its original location by the effect of some noise. Looking back at the definition of a lattice in expression (2.1), it is straightforward to conclude that the basis **B** of the lattice is the matrix **H** and the real and imaginary components of the symbols **x** from a M-QAM constellation C can be made isomorphic to $\mathbb{Z}^k$.

Therefore the CVP can be described as the problem of finding the **x** that *better explains* the observation **y**, which is the one that after the linear transformation (**Hx**) generates the closest vector to the received vector **y**, i.e., it corresponds to the application of the maximum likelihood principle. Figure 3.3 exemplifies this problem in a simply case in $\mathbb{Z}^2$.

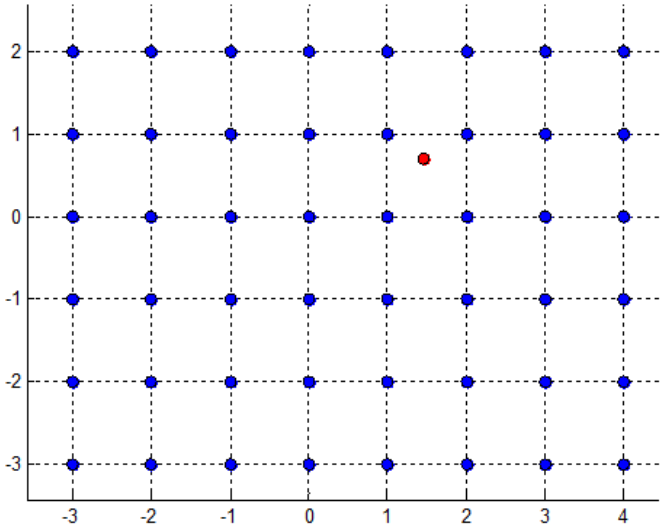

Figure 3.3. Illustration of a simple example of CVP in $\mathbb{Z}^2$.

From an algorithmic complexity point of view CVP is proven to be *NP-hard* which is the worst case scenario in the hierarchy of complexity classes [24]. However, this problem can be solved approximately by means of a number of sub-optimal techniques. Examples of these are *zero-forcing* (ZF) and *successive interference cancelation* (SIC) (the latter proposed by Babai in [25]), or techniques based on *lattice reduction* [26]. This last technique has some good characteristics. In quasi-static fading channels the complexity of lattice reduction is negligible because for a long frame of data the channel remains unchanged and attains a near-optimum performance. In addition, as it was discussed in Chapter 2, there are reduction algorithms such as LLL (or its CLLL, the complex counterpart) algorithm with polynomial complexity and with complexity independent of SNR [27]. These characteristics make *lattice-reduction-aided* (LRA) decoding especially suited for MIMO communications.

The following sections will introduce the most important type of MIMO receivers. After a brief overview their performance will be shown with different number of antennas, keeping $N_T = N_R = N$, and with different M-ary QAM modulations. Specifically, it will be analysed the 2×2 ($N = 2$) and 3×3 ($N = 3$) MIMO systems. The implemented model is the one previously described and the performance of each receiver will be assessed by plotting the SER as function of the SNR.

In the following, the simplest receivers will be introduced first: 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) approach will be brought into discussion.

The performance results shown in this chapter serve to demonstrate that the system model and all the types of receivers considered are well calibrated and equal the results available in the literature.

3.2 Maximum Likelihood Detection

Assuming the zero-mean spatially white (ZMSW) noise model introduced in the section 3.1 the probability density function of **y** given **H** and **x** can be written as

$$\mathrm{p}(\mathbf{y}|\mathbf{H},\mathbf{x}) = \frac{1}{(2\pi\sigma_n^2)^{N/2}} \exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{2\sigma_n^2}\right) \tag{3.7}$$

and consequently the maximum likelihood estimate ($\hat{\mathbf{x}}$) [28] for **x** given **y** is

$$\begin{aligned}\hat{\mathbf{x}}_{\mathrm{ML}} &= \arg\max_{\mathrm{x}\in\mathrm{C}^N} \mathrm{p}(\mathbf{y}|\mathbf{H},\mathbf{x}) \\ &= \arg\max_{\mathrm{x}\in\mathrm{C}^N} \frac{1}{(2\pi\sigma_n^2)^{N/2}} \exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2}{2\sigma_n^2}\right)\end{aligned} \tag{3.8}$$

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\min_{\mathrm{x}\in\mathrm{C}^N} \|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2 \tag{3.9}$$

Therefore the detection problem becomes that of minimizing the exponent of (3.7). The exponential growth of the search space for M-QAM constellations and the increase of dimensions N discourage the use of brute force maximum-likelihood detection in many practical systems given that simply evaluating $\|\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\|^2$ for all possible $\boldsymbol{x} \in C^N$ requires too much time consumption.

The simulation setup implemented is designed to support any number of antennas and M-QAM constellations with $M = 4, 16, 64$. However due to computational constraints in this dissertation only 2×2 and 3×3 MIMO configurations as well as just 4 and 16-QAM constellations will be used to show the performance of the receivers.

Figures 3.4 and 3.5 show the performance of maximum likelihood (ML) detection for 2×2 and 3×3 MIMO configurations. This strategy achieves diversity order equal to N as proven in literature, which means that it captures all the spatial diversity of the configurations, and so the ML performance will be always present in the following results along this chapter as a term of comparison as the best performance attainable. Its performance was simulated for each specific case so there are some variations of its performance caused by the variations of the specific channels considered. Analysing the results one can conclude through equation 3.5 that for 2×2 MIMO the slope is "-2" and for 3×3 MIMO is "-3" as predicted. Besides notice that 16-QAM constellation suffers from a *power penalty* of approximately 8 dB, i.e., it achieves the same SER value from the 4-QAM constellation only in a higher value of SNR.

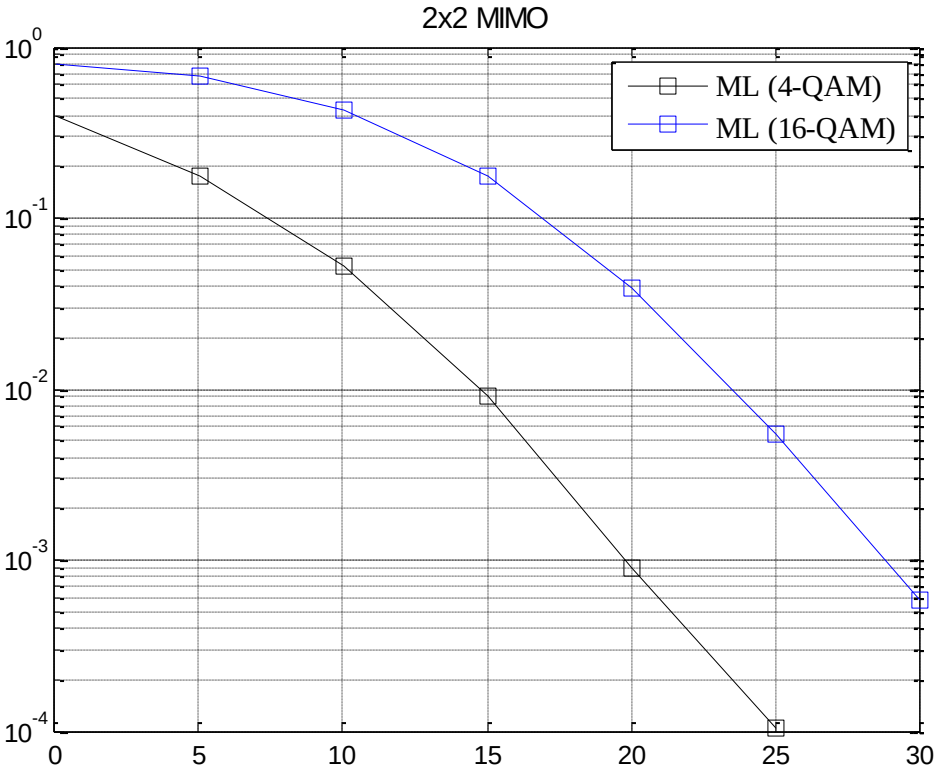

Figure 3.4. ML detection with 2×2 antennas.

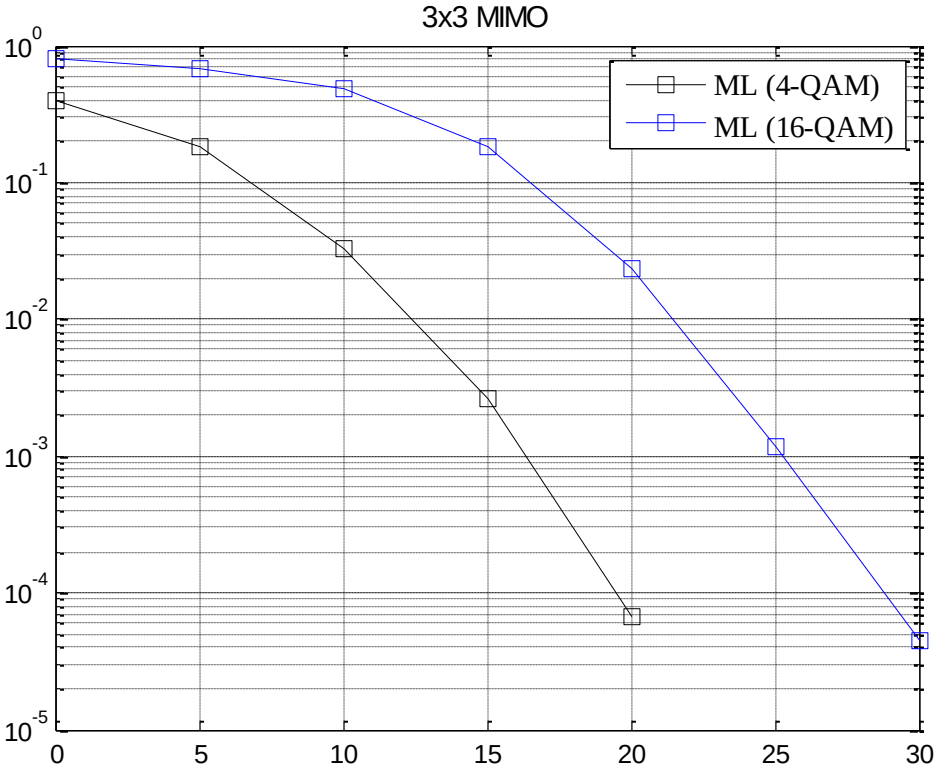

Figure 3.5. ML detection with 3×3 antennas.

3.3 Linear Equalization

Linear receivers consist of applying a linear transformation to the received vector followed by a quantization to the symbol alphabet (slicing). A simple way to obtain an estimate $\hat{\boldsymbol{x}}$ is to form

$$\mathbf{x_{ZF}} = \mathbf{H^{-1}y} = \mathbf{H^{-1}Hx} + \mathbf{H^{-1}n} = \mathbf{x} + \mathbf{H^{-1}n} \tag{3.10}$$

$$\hat{\mathbf{x}}_{\mathbf{ZF}} = Q_C[\mathbf{x_{ZF}}] \tag{3.11}$$

This is the so-called linear *zero-forcing* (ZF) receiver, since the interference caused by $\mathbf{H}$ is forced to be zero. Notice that once equation (3.9) is in the complex domain $\mathbf{H}^{-1}$ corresponds to the *pseudo-inverse matrix*, also known as *Moore-Penrose matrix* $\mathbf{H}^{\dagger}$.

$$\mathbf{H}^{\dagger} = \left(\mathbf{H}^{\mathcal{H}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathcal{H}} \tag{3.12}$$

Superscript $(.)^{\mathcal{H}}$ denotes *Hermitian* operator (conjugation followed by transposition or vice-versa). The inversion of $\mathbf{H}$ is a trivial operation however it can only be defined for *invertible* matrices, i.e. matrices with non-zero determinant. Here resides the need of $N_R \geq N_T$. The filtered noise is transformed by $\mathbf{H}^{\dagger}$, which constitutes a noise enhancement factor.

$$\hat{\mathbf{x}}_{\mathbf{ZF}} = Q_C\left[\mathbf{H}^{\dagger}\mathbf{y}\right] = Q_C\left[\mathbf{H}^{\dagger}\mathbf{x} + \mathbf{H}^{\dagger}\mathbf{n}\right] = Q_C\left[\mathbf{x} + \mathbf{H}^{\dagger}\mathbf{n}\right] \tag{3.13}$$

Summarizing the ZF receiver structure has a linear transformation $\mathbf{W}_{\mathrm{ZF}} = \mathbf{H}^{\dagger}$ followed by a quantization to the symbol alphabet C by threshold decision

The detected vector $\hat{\mathbf{x}}_{\mathbf{ZF}}$, as obtained from (3.13), is in fact the solution to

$$\hat{\mathbf{x}}_{\mathbf{ZF}} = \arg\min_{\mathbf{x}\in\mathbb{C}^N} \|\mathbf{y} - \mathbf{Hx}\|^2 \tag{3.14}$$

Notice that in $\hat{\mathbf{x}}_{\mathbf{ZF}}$ the search is now made in the *continuous* domain $\mathbb{C}^N$ instead of the discrete complex alphabet C as it is in $\hat{\mathbf{x}}_{\mathbf{ML}}$. This is the origin of the sub-optimality of the ZF receiver. After the inverse transformation, all the points in the lattice are matched back to the initial $\mathbb{Z}^N$. The orthogonal geometry of $\mathbb{Z}^N$ eliminates all the interference between dimensions of the lattice, i.e., between the MIMO layers.

ZF solves the CVP by relaxing it to a search in a continuous neighbourhood instead of computing the distance between the *target* (received vector) and every point in the lattice. In a geometrical perspective ZF performs a linear transformation of the Voronoi regions (mentioned in sub-section 2.2.4) of the $\mathbb{Z}^N$, original cubic lattice, 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 [29].

Different bases will output different ZF decision regions that could correspond better or worse to the ideal Voronoi regions of the lattice. The ZF decisions regions are the fundamental regions of the lattice that were shown in the sub-chapter 2.2.3 and the lower the match between ZF and Voronoi decision regions the greater the SER obtained. Taking into account these facts it is possible to conclude that finding bases such that form the same lattice and better approximate their fundamental regions to the Voronoi decision regions will improve the performance of linear receivers. This is the main idea of the LRA receivers that will be discussed latter.

The noise enhancement factor can be minimized with a *minimum mean-square error* (MMSE) receiver. In this receiver it is taken into account both the interference and the noise in order to minimize the expected error. The MMSE linear receiver looks for a linear transformation $\mathbf{W}_{\mathrm{MMSE}}$ that minimizes the mean square error between the estimated vector and the original vector,

$$\mathbf{W}_{\mathrm{MMSE}} = \arg\min_{\mathbf{W}} E\{\|\mathbf{W}\mathbf{y} - \mathbf{x}\|^2\} \tag{3.15}$$

There are other possible expressions for representing $\mathbf{W}_{\mathrm{MMSE}}$ but the one assumed in this work is

$$\mathbf{W}_{\mathrm{MMSE}} = \sigma_x^2 \mathbf{I}_{\mathrm{N}} \cdot \mathbf{H}^{\mathcal{H}} \left(\sigma_n^2 \mathbf{I}_{\mathrm{N}} + \mathbf{H} \cdot \sigma_x^2 \cdot \mathbf{I}_{\mathrm{N}} \cdot \mathbf{H}^{\mathcal{H}}\right)^{-1} = \mathbf{H}^{\mathcal{H}} \left\{\mathbf{H}\mathbf{H}^{\mathcal{H}} + \frac{\sigma_n^2}{\sigma_x^2} \mathbf{I}_{\mathrm{N}}\right\} \tag{3.16}$$

The estimated $\hat{\mathbf{x}}_{\mathbf{MMSE}}$ is obtained by a linear transformation $\mathbf{W}_{\mathrm{MMSE}}$ followed by a quantization step that is the same as the one used in the ZF case.

$$\hat{\mathbf{x}}_{\mathbf{MMSE}} = Q_C[\mathbf{W}_{\mathrm{MMSE}}\, \mathbf{y}] = Q_C[\mathbf{W}_{\mathrm{MMSE}}\, \mathbf{H}\mathbf{x} + \mathbf{W}_{\mathrm{MMSE}}\mathbf{n}] = Q_C[\mathbf{x} + \mathbf{W}_{\mathrm{MMSE}}\mathbf{n}] \tag{3.17}$$

The MMSE performs better than ZF receiver because it solves the CVP problem by relaxing the search in the continuous space where $\mathbf{x} \in C^N$ but also for introducing a term that penalises large $\|\mathbf{x}\|$ and is proportional to the energy of the noise [29].

In order to make an easier implementation an extended system was assumed [15].

$$\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \frac{\sigma_n}{\sigma_x} \mathbf{I}_{\mathrm{N}} \end{bmatrix}, \qquad \bar{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ 0_{N\times 1} \end{bmatrix} \tag{3.18}$$

Therefore the estimated $\hat{\mathbf{x}}_{\mathbf{MMSE}}$ can be easily obtained as

$$\hat{\mathbf{x}}_{\mathbf{MMSE}} = Q_C\left[\, \bar{\mathbf{H}}^{\dagger} \bar{\mathbf{y}}\right]. \tag{3.19}$$

Figures 3.6 to 3.9 show the performance of ZF and MMSE linear receivers in different configurations. The ML performance is also present as term of comparison with the best performance attainable. Analysing the figures it possible observe that although MMSE receiver performs better than the ZF both SER curves settle to a slope of -1 as proven in literature, which corresponds to a decrease of symbol error rate by a factor of 10 for a 10-fold increase of SNR. From Figures 3.6 to 3.9 one can observe that MMSE detection has a *power gain* compared with ZF, i.e., it achieves the same SER value from the 4-QAM constellation in a lower value of SNR, yet for higher constellations this power gain decreases as well as for higher SNR values, becoming minimal.

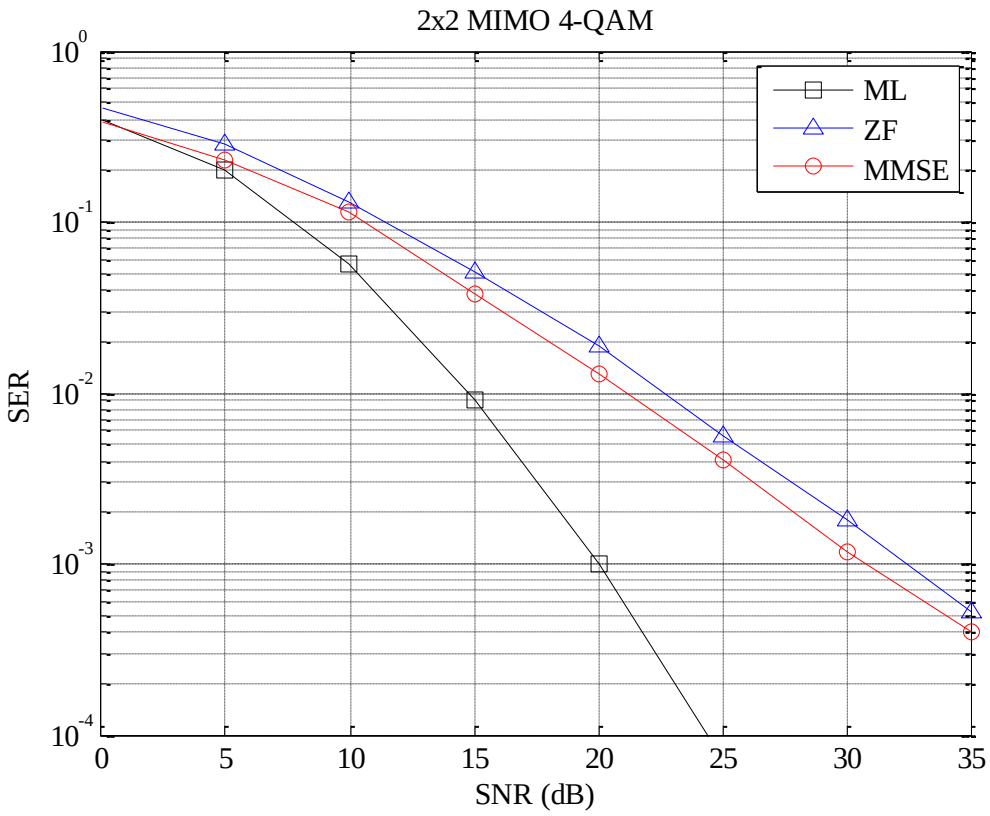

Figure 3.6. Detection 2×2 antennas with 4-QAM using linear receivers.

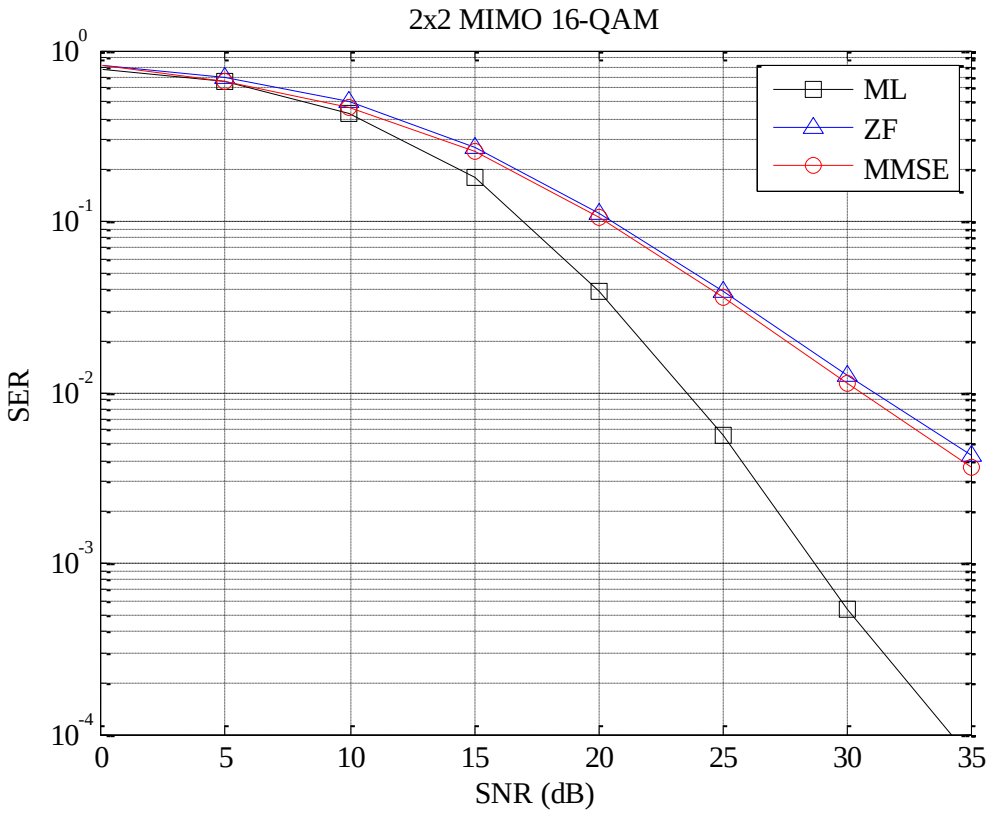

Figure 3.7. Detection 2×2 antennas with 16-QAM using linear receivers.

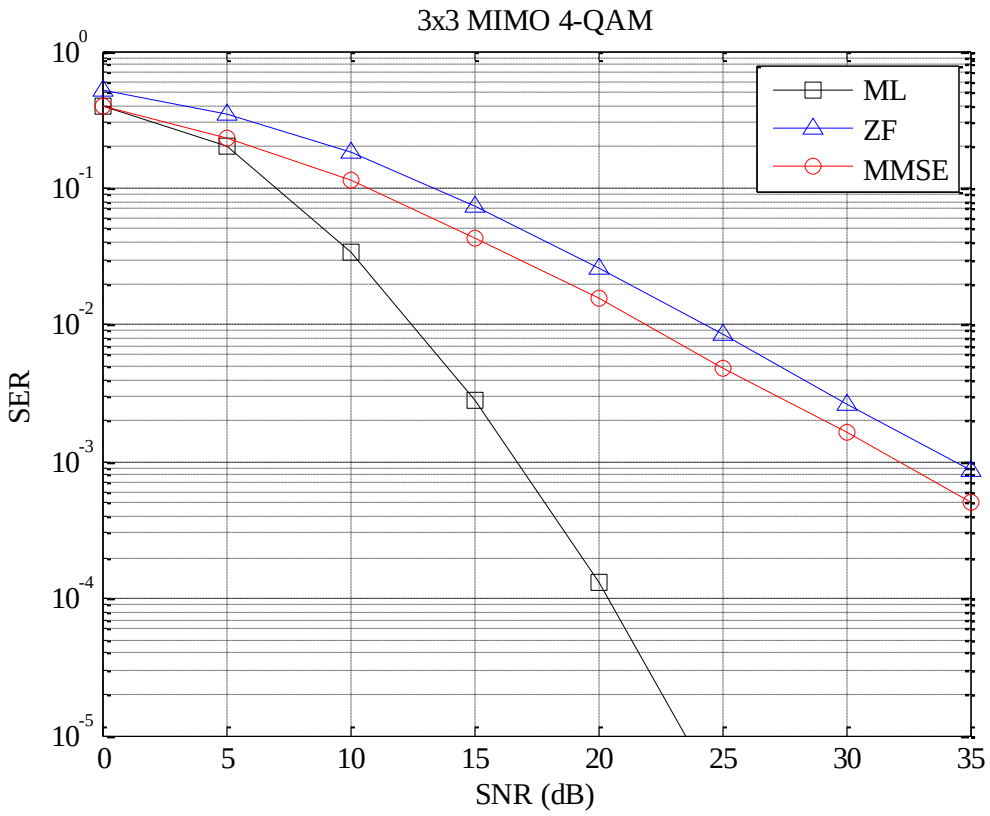

Figure 3.8. Detection 3×3 antennas with 4-QAM using linear receivers.

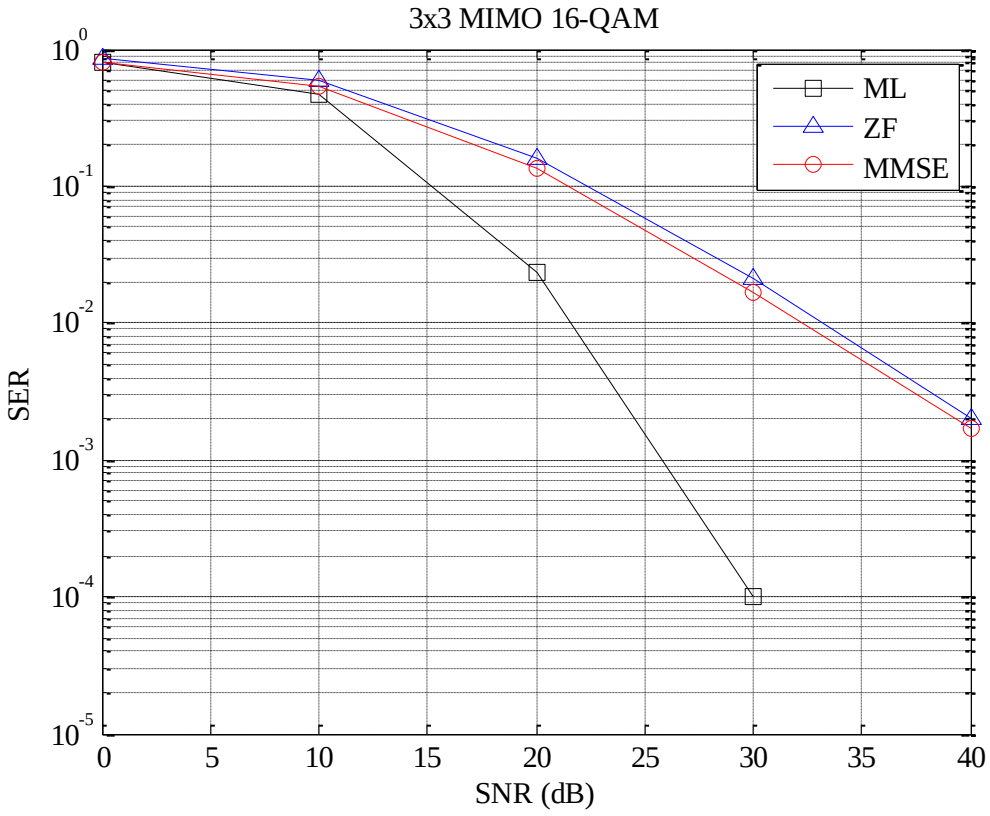

Figure 3.9. Detection 3×3 antennas with 16-QAM using linear receivers.

3.4 Order Successive Interference Cancellation Detection

Nonlinear methods on the receivers have to involve decision feedback somehow in order to remove the interference between the signals from each antenna. In this context the most commonly used approach is called *order successive interference cancelation* (OSIC). OSIC in fact corresponds to the original algorithm proposed for the detection of the original V-BLAST transmission scheme in [30]. As it was described in subsection 3.1, V-BLAST allows for the transmission of independent data streams over the matrix channel **H**.

While the transmission side of spatial multiplexing V-BLAST is straightforward, the detection side is more difficult to implement since the signals from each antenna are received by all the receive antennas. The principle of OSIC is to use linear detection to detect first the modulation symbol of the layer least affected by noise and, subsequently, assuming that the first symbol was correctly detected, the interference created by that symbol is replicated and subtracted from the others layers. The procedure continues detecting the next best signal (again in the sense of the one with the least noise enhancement) and subtracting the interference caused by it to the remaining layers, and repeating this procedure until all symbols are detected.

The optimal criterion at each stage is to select the layer that less enhances the noise power after the linear detection. It was shown in [29] that this corresponds to deciding for the layer that spans the *n*-dimensional lattice by translating parallel hyperplanes of *n*-1 dimensions which have the maximum separation among all possible families of hyperplanes. In doing this, the error probability is minimised for that layer. This corresponds to ordered SIC, without which the performance of SIC is seriously degraded by 3 dB[31].

Since the worst subchannel dominates the average SER this is the optimal criterion if all subchannels use the same signalling rate and power. This result was already found in [30], in the first implementation of the V-BLAST detector.

For a better understanding of the OSIC procedure let us follow a practical example of its application on this work for 2×2 MIMO system through 4-QAM constellation and using ZF detection.

Consider the system of (3.1), where $\mathbf{x} = [-1+1i \quad 1-1i \quad -1-1i \;]^T$ and

$$\mathbf{y^{(1)}} = \begin{bmatrix} -2.43 + 0.08i \\ +1.07 - 0.97i \\ -0.69 - 0.06i \end{bmatrix} = \underbrace{\begin{bmatrix} -0.73 + 0.45i & -0.10 - 0.74i & 0.73 - 0.29i \\ 0.94 - 0.15i & 0.65 - 0.19i & -0.24 + 0.70i \\ -0.30 - 0.61i & -0.21 - 0.31i & 0.72 - 0.21i \end{bmatrix}}_{\mathbf{H}} \mathbf{x} + \mathbf{n}\,. \tag{3.20}$$

By applying the ZF linear transformations it is obtained

$$\hat{\mathbf{x}} = \mathbf{H}^{\dagger}\mathbf{y} = \mathbf{x} + \mathbf{H}^{\dagger}\mathbf{n} \tag{3.21}$$

$$\mathbf{W^{(1)}} = \mathbf{H}^{\dagger} = \begin{bmatrix} -0.18 - 0.58i & 0.65 + 0.53i & 1.04 + 0.38i \\ 0.28 + 0.65i & 1.59 + 1.28i & 1.58 - 1.18i \\ 0.23 - 0.01i & -0.75 + 1.62i & 1.92 + 1.94i \end{bmatrix}. \tag{3.22}$$

The component of $\hat{\mathbf{x}}$ that experience the lowest noise enhancement by $\mathbf{H}^{\dagger}$ can be detected with high reliability. Since ZMSW noise model is assumed this component corresponds to the row of $\mathbf{H}^{\dagger}$ of smallest *length*, which is the Euclidean norm of the row.

$$\|\mathbf{w}_1(1,:)\|^2 = 2.30 \quad \|\mathbf{w}_1(2,:)\|^2 = 8.55 \quad \|\mathbf{w}_1(3,:)\|^2 = 10.69 \tag{3.23}$$

By observation of (3.23) the lowest noise enhancement is associated with the first row, which corresponds to the first antenna (*layer*). The next step is decide the symbol by quantization

$$\hat{\mathbf{x}}(1) = Q_C\left[\mathbf{W^{(1)}}\mathbf{y^{(1)}}\right] = 1 + 1i\,. \tag{3.24}$$

Now the interference due to $\hat{\mathbf{x}}(1)$ is replicated and then subtracted on the others layers

$$\mathbf{y^{(2)}} = \mathbf{y} - \mathbf{H}(:,\mathbf{1})\hat{\mathbf{x}}(1) = \begin{bmatrix} -1.26 + 0.37i \\ -0.02 - 1.76i \\ -1.01 - 0.85i \end{bmatrix} \tag{3.25}$$

Once the symbol $\hat{\mathbf{x}}(1)$ is already decided the first generator vector in the channel matrix $\mathbf{H}$ becomes zero. Thereby the matrix $\mathbf{H}$ took into account to the second round of the algorithm is (3.26) and its associated ZF linear transformation matrix (3.27).

$$\mathbf{H}^{(2)} = \begin{bmatrix} 0 & -0.10 - 0.74i & 0.73 - 0.29i \\ 0 & 0.64 - 0.19i & -0.24 + 0.70i \\ 0 & -0.21 - 0.31i & 0.72 - 0.21i \end{bmatrix} \tag{3.26}$$

$$\mathbf{W^{(2)}} = \begin{bmatrix} 0 & 0 & 0 \\ 0.92 + 1.14i & 0.47 + 1.21i & 0.16 - 0.78i \\ -0.27 + 0.99i & -0.94 + 0.09i & 1.00 + 0.13i \end{bmatrix} \tag{3.27}$$

Follows the calculation of the component of $\hat{\mathbf{x}}$ that experience the lowest noise but this time over the estimates available via $\mathbf{W^{(2)}}$ linear equalization of $\mathbf{H^{(2)}}$.

$$\|\mathbf{w}_2(1,:)\|^2 = 0 \quad \|\mathbf{w}_2(2,:)\|^2 = 4.46 \quad \|\mathbf{w}_2(3,:)\|^2 = 2.96 \tag{3.28}$$

At this stage the third row is the one that less enhances the noise and therefore $\hat{\mathbf{x}}(3)$ is now decided as

$$\hat{\mathbf{x}}(3) = Q_C\left[\mathbf{W^{(2)}}\mathbf{y^{(2)}}\right] = -1 + 1i\,, \tag{3.29}$$

and its interference is subtracted from $\mathbf{y^{(2)}}$

$$\mathbf{y}^{(3)} = \mathbf{y}^{(2)} - \mathbf{H}(:,3)\hat{\mathbf{x}}(3) = \begin{bmatrix} -0.82 - 0.65i \\ 0.44 - 0.82i \\ -0.51 - 0.08i \end{bmatrix} \tag{3.30}$$

and the matrix **H** becomes in (3.31), where its third column is now null.

$$\mathbf{H}^{(3)} = \begin{bmatrix} 0 & -0.10 - 0.74i & 0 \\ 0 & +0.64 - 0.19i & 0 \\ 0 & -0.21 - 0.31i & 0 \end{bmatrix} \tag{3.31}$$

By applying the ZF linear transformations over $\mathbf{H}^{(3)}$ it is obtained $\mathbf{W}^{(3)}$.

$$\mathbf{W}^{(3)} = \begin{bmatrix} 0 & 0 & 0 \\ -0.09 + 0.65i & 0.56 + 0.17i & -0.19 - 0.27i \\ 0 & 0 & 0 \end{bmatrix} \tag{3.32}$$

Finally remains estimate $\hat{\mathbf{x}}(2)$ and the process comes to an end.

$$\hat{\mathbf{x}}(2) = Q_C\left[\mathbf{W}^{(3)}\mathbf{y}^{(3)}\right] = -1 - 1i\,. \tag{3.33}$$

As expected all the symbols were correctly detected. However, one of the basic problems of this strategy is that if one symbol is incorrectly detected, and for bad luck it is the first one to be decided, it leads to a higher probability of wrongly detection to the following symbols.

Instead of ZF equalization it is also possible perform linear MMSE equalization to obtain a MMSE version of the OSIC receiver. The only aspect that changes in what was exposed before is the linear transformations **W** should be replaced by the expression of MMSE equalization in the equation (3.16).

The performance of OSIC receivers in different configurations can be found in Figures 3.10 to 3.13. In these figures are present both versions of OSIC receivers, the OSIC combined with ZF (OSIC-ZF) and with MMSE (OSIC-MMSE) for 2×2 and 3×3 MIMO systems using 4 and 16 QAM constellation. Moreover they can be compared with the ML detection. Although the performance of the OSIC receiver capture the same diversity of ordinary linear receivers, $d = 1$, by inspection of Figures 3.6 to 3.9 one can conclude that it provides a large power gain when compared with the ordinary linear receivers. Besides with the use of OSIC the MMSE equalizer performance presents a higher difference of power gain to OSIC-ZF then the simple MMSE compared with the simple ZF. One can conclude that the use of OSIC enhance more the performance of MMSE than ZF equalizer.

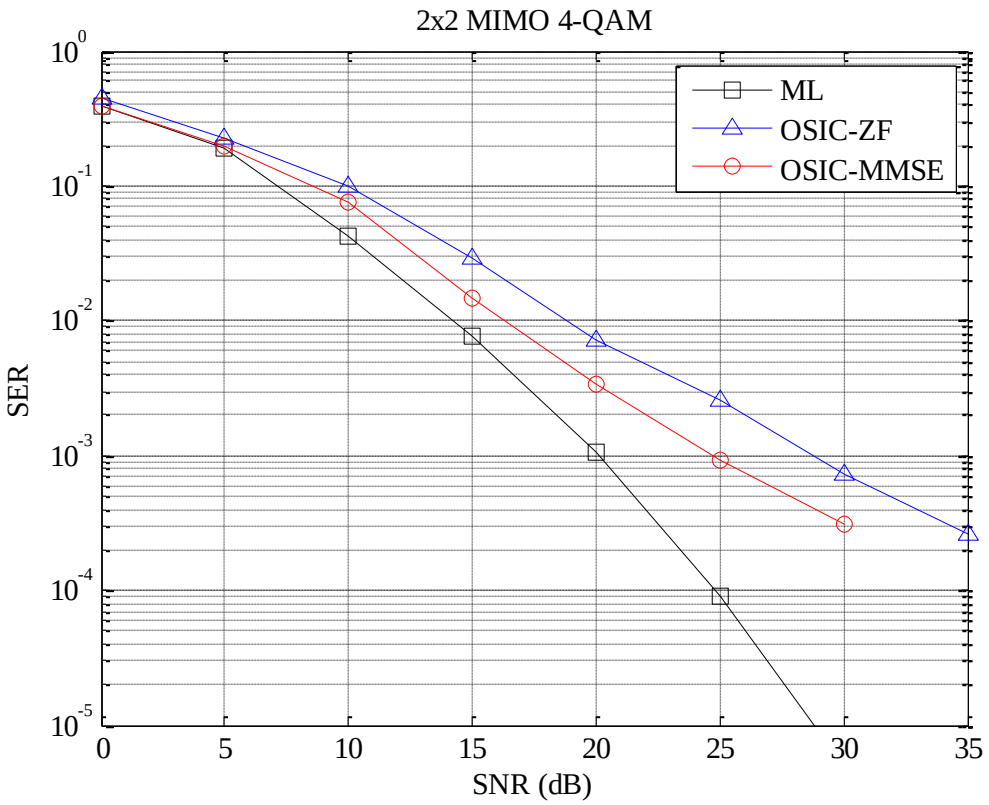

Figure 3.10. Detection 2×2 antennas with 4-QAM using OSIC linear receivers.

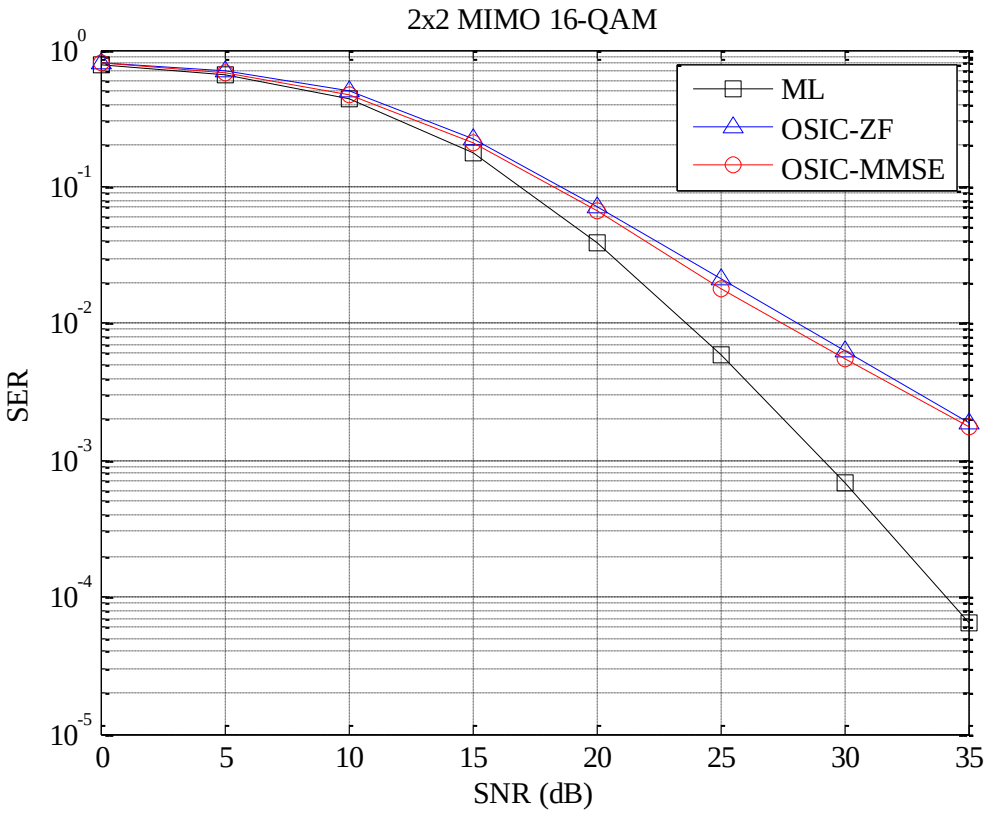

Figure 3.11. Detection 2×2 antennas with 16-QAM using OSIC linear receivers.

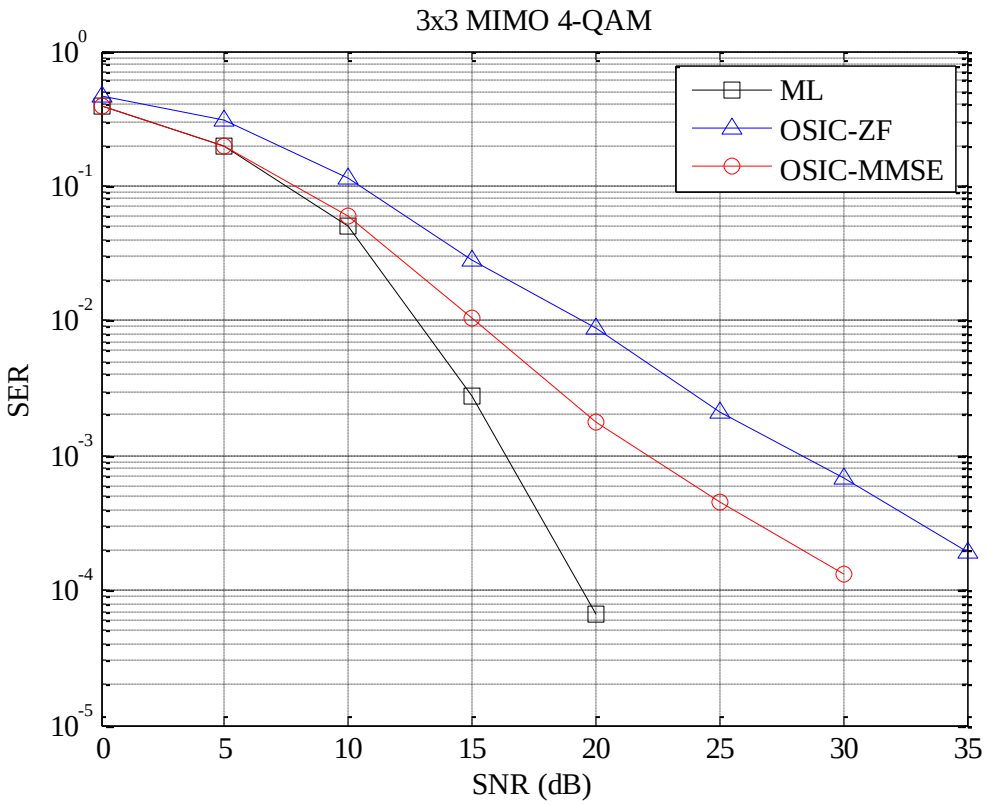

Figure 3.12. Detection 3×3 antennas with 4-QAM using OSIC linear receivers.

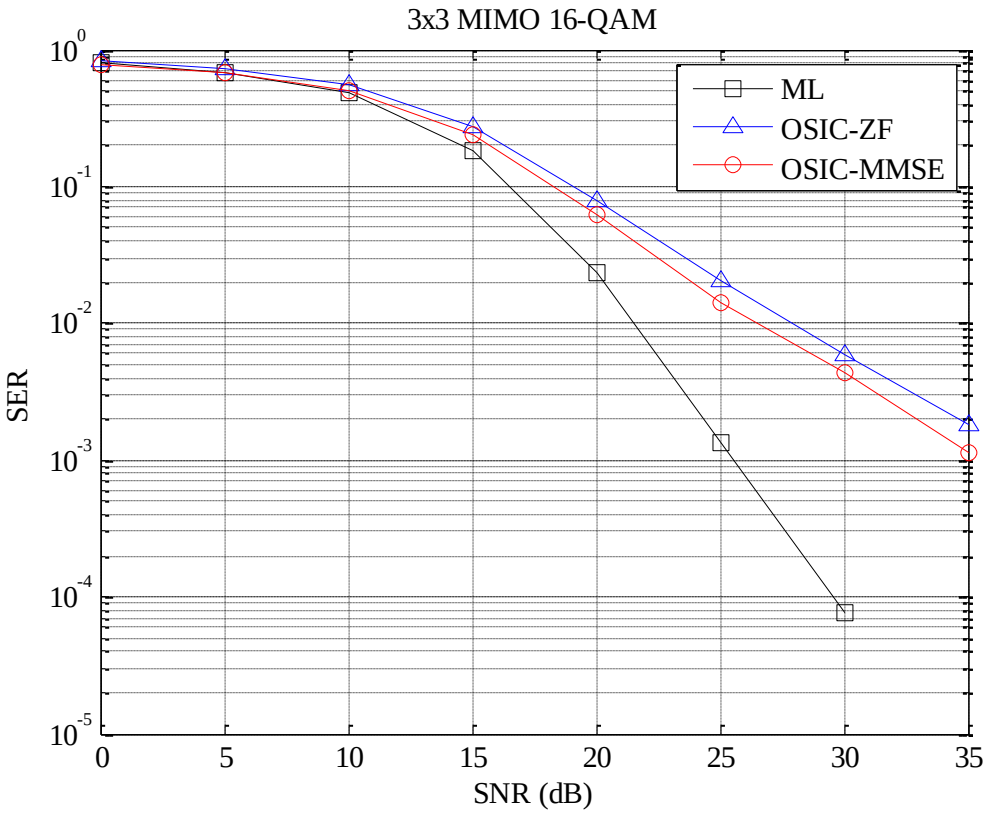

Figure 3.13. Detection 3×3 antennas with 16-QAM using OSIC linear receivers.

3.5 Lattice Reduction-Aided Detection

The concept on *lattice reduction-aided* (LRA) detection arises naturally under this lattice perspective. Considering the detection problem as a CVP, one wants to look for methods that produce good approximate solutions (preferably, in polynomial time). The basic idea behind LRA is to use lattice reduction in conjunction with traditional low-complexity decoders.

Using the lattice detection interpretation, given that the used constellations are QAM constellations, and therefore $\mathbf{x} \in C^N \subset \mathbb{Z}^N$, the noiseless received points in the communication scenario corresponds to points of the lattice $\mathbf{H}\mathbb{Z}^N$.

As was already discussed different bases can generate the same lattice, however there are bases that perform linear equalizer (ZF and MMSE) decision regions which are closer to the ideal Voronoi regions than others. Lattice reduction can be implemented by algorithms such as LLL and KZ. These algorithms optimise the generating matrix of the lattice to obtain a "better" description of the lattice [23]. The lattice points corresponds to integer linear combinations of the columns of $\mathbf{H} = [\mathbf{h_1}, \dots, \mathbf{h}_N]$, and lattice reduction obtains another set of column vectors, collected in the *reduced* matrix $\mathbf{H}_{\text{red}}$, which spans the same set of points, $\mathbf{H}\mathbb{Z}^N \equiv \mathbf{H}_{\text{red}}\mathbb{Z}^N$ while having its generator vectors more orthogonal to each other than the ones in the original basis. Since the matrix $\mathbf{H}_{\text{red}}$ is composed of more orthogonal vectors, the low-complexity detectors will perform better in this lattice because the noise enhancement factor (which for example in linear equalisation amounts to $\mathbf{H}_{\text{red}}^{-1}$), is decreased.

The matrices $\mathbf{H}$ and $\mathbf{H}_{\text{red}}$ are related by

$$\mathbf{H}_{\text{red}} \equiv \mathbf{H}\mathbf{M}\,, \tag{3.34}$$

where $\mathbf{M}$ is a unimodular matrix, i.e., has integer entries and $|\det(\mathbf{M})| = 1$, therefore $\mathbf{M}^{-1}$ is also unimodular and with integer entries.

Using equation (3.34) in (3.1) results in

$$\mathbf{y} = \mathbf{H}_{\text{red}} \underbrace{\mathbf{M}^{-1}\mathbf{x}}_{\mathbf{z}} + \mathbf{n}\,, \tag{3.35}$$

and $\mathbf{z}$ can be interpreted as the noiseless received signal points as the points in the lattice described by $\mathbf{H}_{\text{red}}$. The original data vector $\mathbf{x}$ can be recovered from $\mathbf{z}$ nothing that

$$\mathbf{z} = \mathbf{M}^{-1}\mathbf{x} \implies \mathbf{x} = \mathbf{M}\mathbf{z} \tag{3.36}$$

The M-QAM constellations used in this dissertation are defined without the origin and have non unitary distance between the symbols. In order to apply *lattice reduction* algorithms it is necessary to make a translation of the constellation, by creating the modified received vector

$$\mathbf{y}_{\text{red}} = \frac{1}{2}(\mathbf{y} + \mathbf{Hp}) \tag{3.37}$$

where $\mathbf{p}$ denotes a column vector of N elements all equal to 1.

After the lattice reduction pre-processing stage, any of receivers earlier described in this dissertation may be used concatenated with the lattice reduction block, namely ZF, MMSE and OSIC.

The complex LLL (CLLL) was implemented in the simulations with LRA. Since the CLLL reduces basis which are complex matrices, other components of MIMO detectors, such as the computation of pseudo-inverse and QR decomposition, can also be operated with complex arithmetic without doubling the channel matrix dimension. Thus, the use of CLLL helps reducing the complexity of other parts of the MIMO detector, not just the reduction algorithm. CLLL directly operates on the complex matrix **H**, and because of that it only requires half of the complexity of traditional LLL using the real equivalent model explained in subsection 3.3.1.

Table 4 represents the pseudo-algorithm implemented to simulate the LRA receivers.

Table 4 Pseudo-code of LRA detection.

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

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

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

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

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

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

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

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

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

Figures 3.14 to 3.21 show the performance of LRA receivers in different configurations. These configurations include 2×2 and 3×3 MIMO systems using both 4 and 16 QAM constellations. The LRA receivers capture the same diversity of ML receivers as proved in literature. However they suffer some power penalty. By analysing Figures 3.14 to 3.18 one can conclude that LRA ZF an LRA MMSE in 2×2 MIMO system, taking into account the ML detection for each case, have less power penalty when compared with its use in 3×3 MIMO system. Besides the power penalty difference between LRA ZF and MMSE detection becomes higher in the 3×3 MIMO system, been the latter the one that achieves better performance losing only a few dB from the optimal ML detection in 2×2 MIMO system. In the lattice reduction followed by OSIC strategies, depicted in Figures 3.19 to 3.21, it is possible to observe that the power gain over the LRA linear receivers is large, specially for the case of 3×3 MIMO system. Furthermore the fact observed in section 3.4 is also noticed, once that after the pre-processing stage of lattice reduction the OSIC-MMSE performs much better than the OSIC-ZF detection. For a 2×2 MIMO system using 4-QAM constellation, the LRA OSIC-MMSE receivers have less than 1 dB of power penalty (see Figure 3.18).

Reviewing all the strategies implemented for the MIMO detection the performance results points to the conclusion that LRA OSIC-MMSE receiver is the one that achieves the best performance. It capture the same diversity order of ML detection, $d = N$, and among the other receivers is the one that has the smaller power penalty compared with ML. On the other hand is also the one that needs more computation, so apart from the LRA receivers is OSIC-MMSE receiver whose has the best performance. It has a slope of -1 but is the one that minimizes the power penalty compared with ML detection, again apart from LRA receivers.

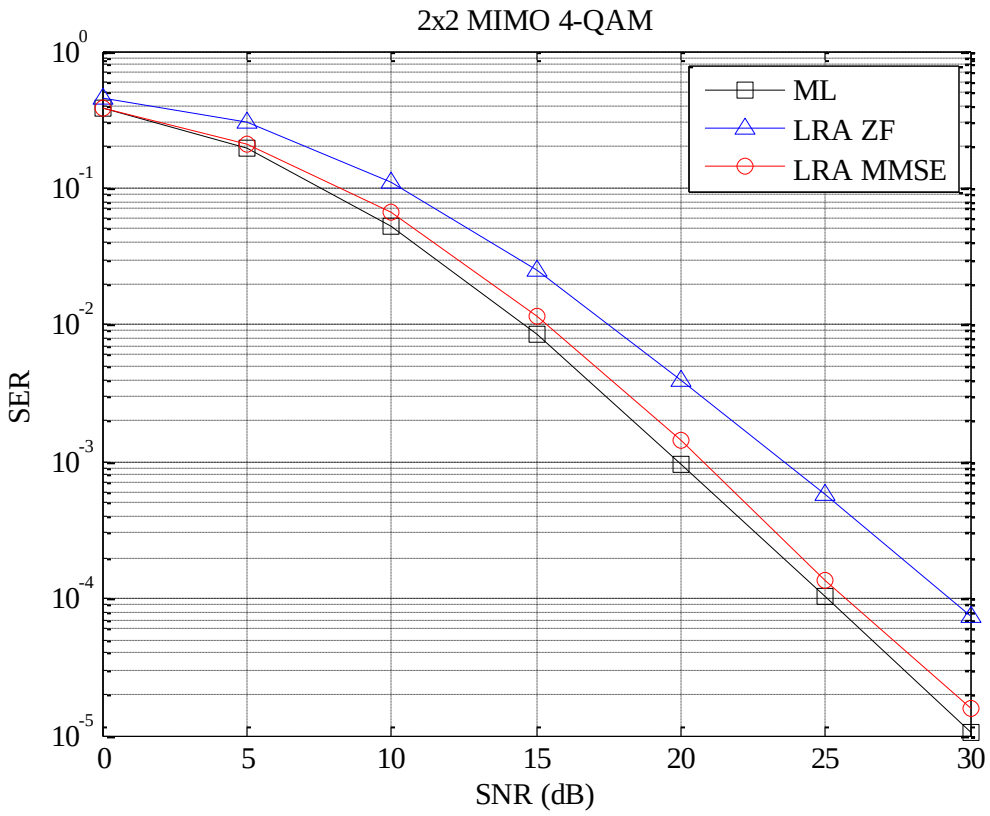

Figure 3.14. Detection 2×2 antennas with 4-QAM using LRA linear receivers.

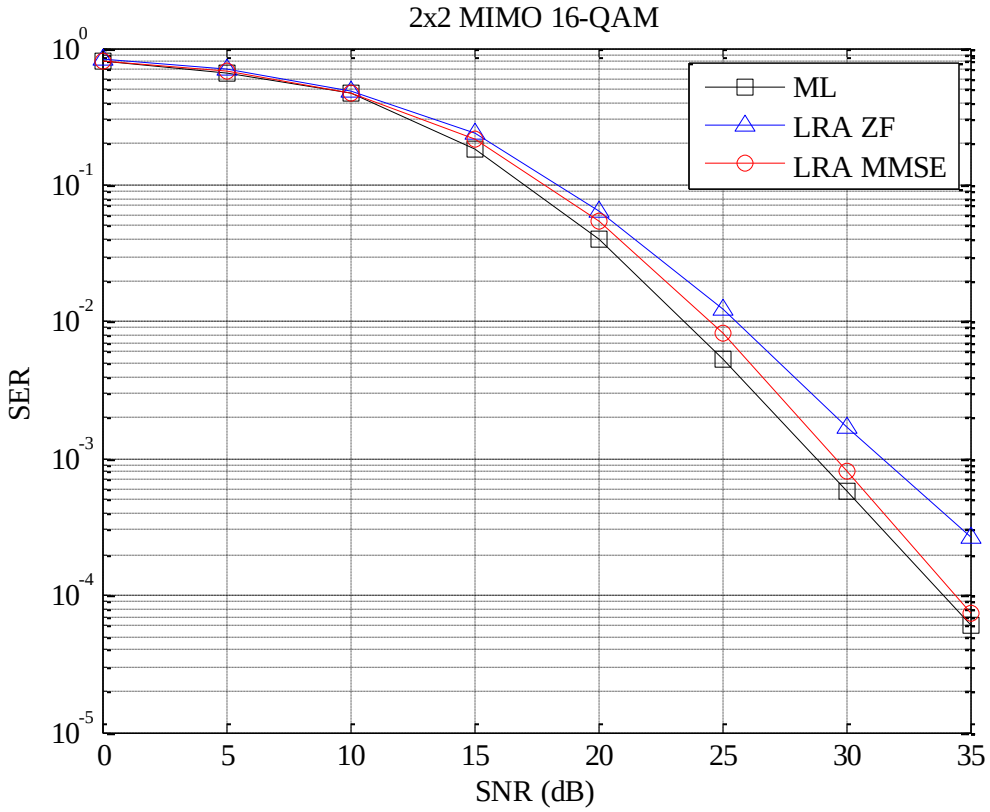

Figure 3.15. Detection 2×2 antennas with 16-QAM using LRA linear receivers.

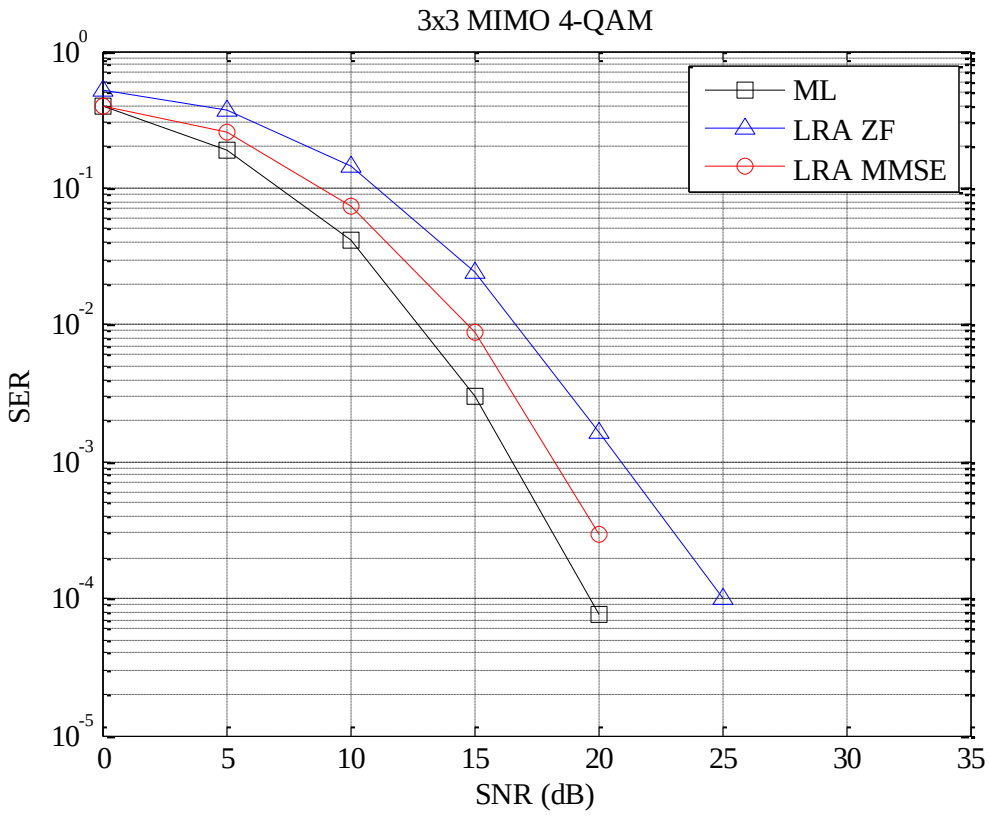

Figure 3.16. Detection 3×3 antennas with 4-QAM using LRA linear receivers.

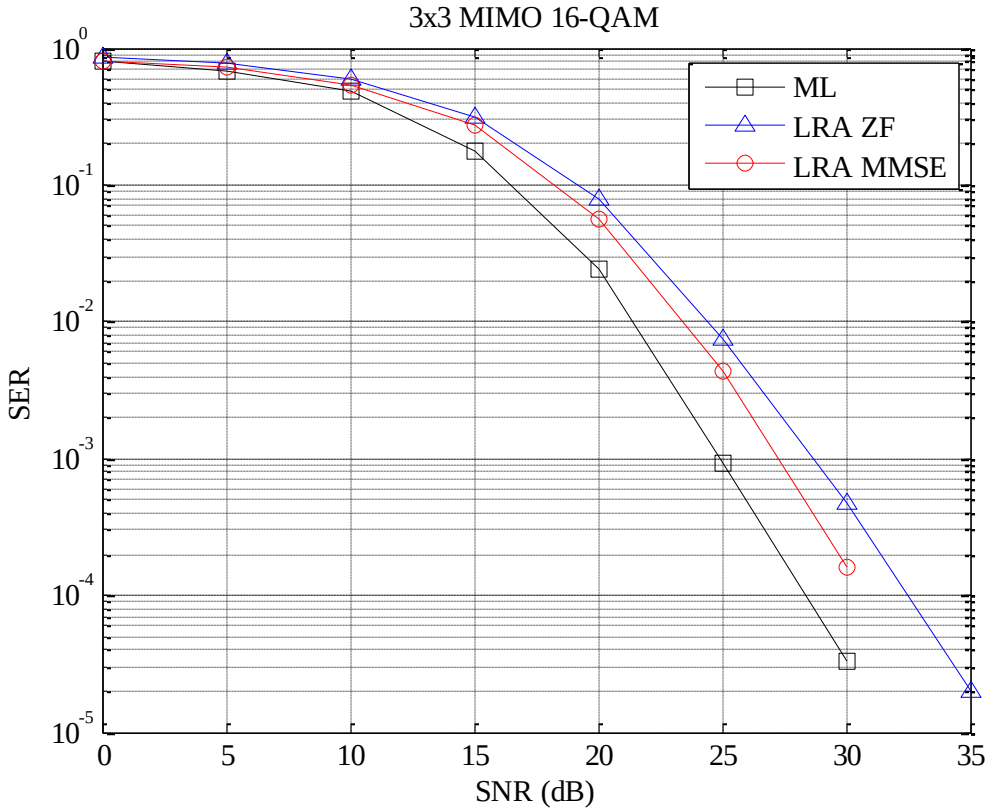

Figure 3.17. Detection 3×3 antennas with 16-QAM using LRA linear receivers.

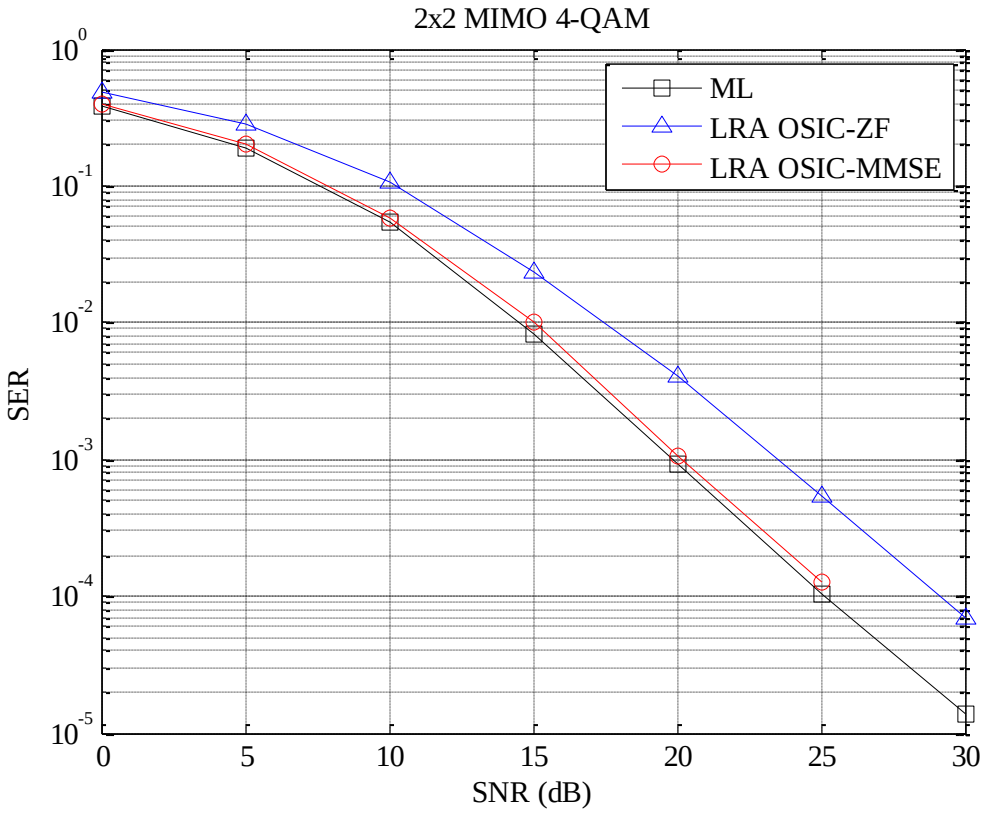

Figure 3.18. Detection 2×2 antennas with 4-QAM using LRA OSIC linear receivers.

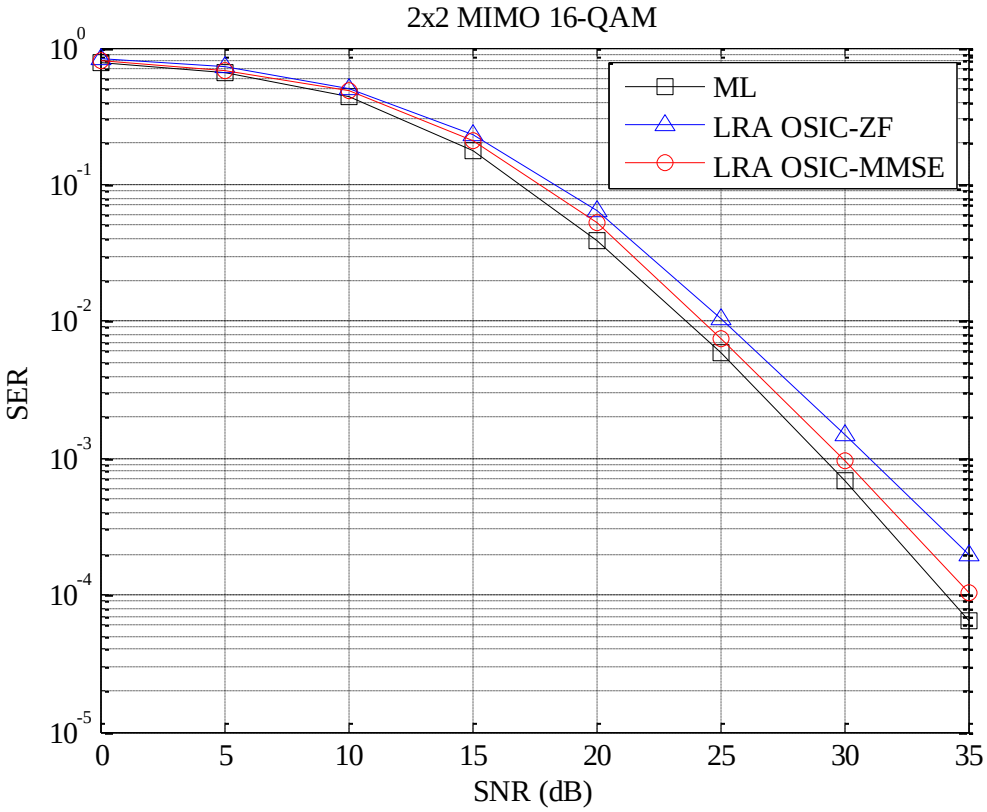

Figure 3.19. Detection 2×2 antennas with 16-QAM using LRA OSIC linear receivers.

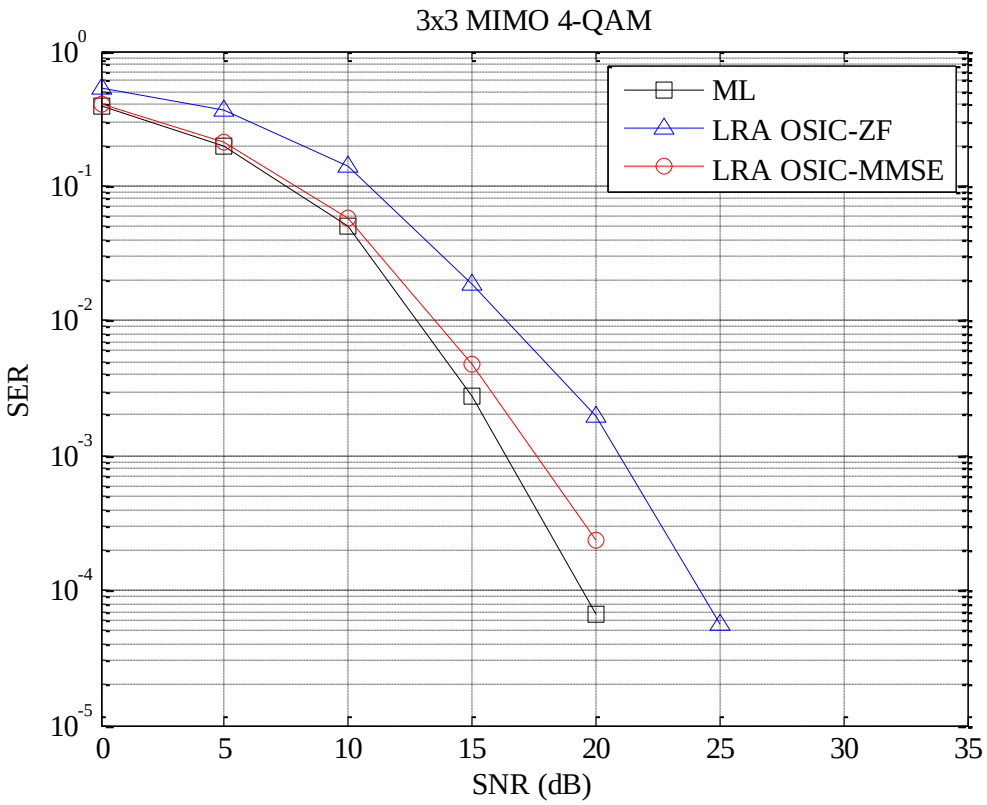

Figure 3.20. Detection 3×3 antennas with 4-QAM using LRA OSIC linear receivers.

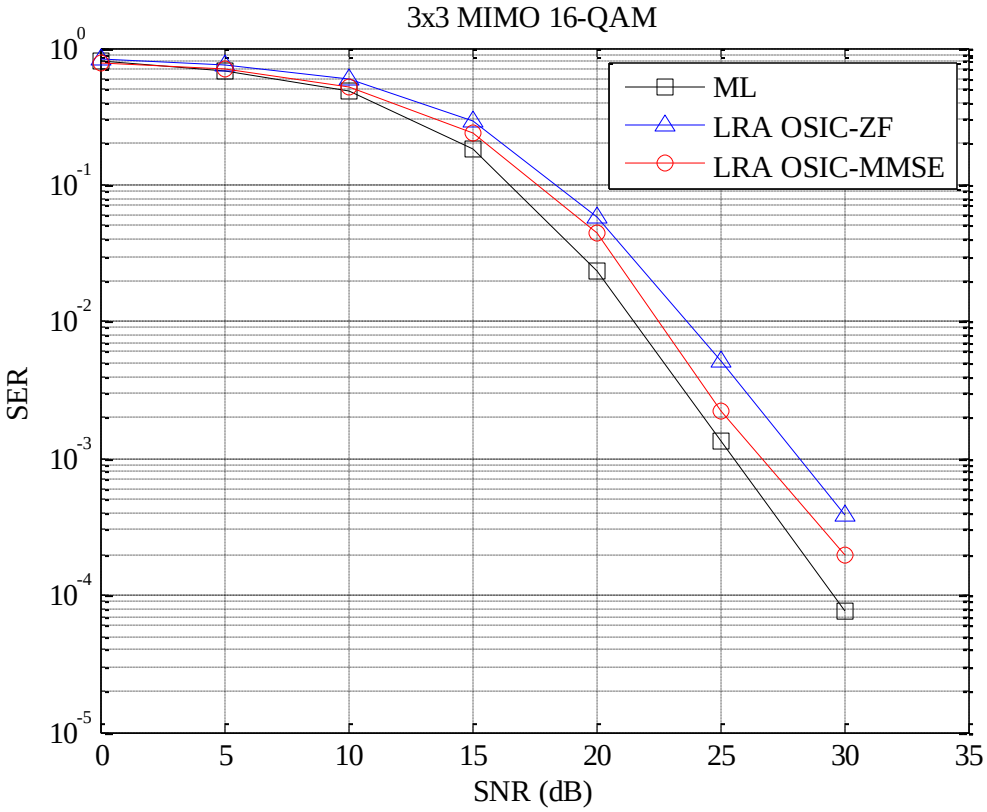

Figure 3.21. Detection 3×3 antennas with 16-QAM using LRA OSIC linear receivers.

Chapter 4

MIMO combined with PLNC

The present chapter proposes a set of new strategies combining MIMO with PLNC in scenarios that move beyond the traditional TWRC.

4.1 System Model

Chapter 3 introduced the concept of MIMO systems as well as the most part of the strategies used to solve the CVP. Using these MIMO techniques in a context involving PLNC is the objective of this chapter. One suitable name for the following described scenarios may be *PLNC in a network with a MIMO relay*. Different communication strategies and different scenario configurations will be considered.

The main idea of PLNC exploited in this dissertation is to create a mechanism similar to network coding, however implemented at the very bottom of the physical layer that deals with EM signal reception and modulation. Through a proper modulation and demodulation technique at the nodes additions of EM signals can be properly defined, so that the interference becomes part of the arithmetic operation of the network coding.

Understanding the structure of the physical layer is of utmost importance for the proposals that will be presented in this chapter. The physical layer considered in this dissertation is the wireless medium and thus its properties will be reviewed, despite the fact that some of them have already been considered implicitly in the previous chapter. Based on [10] in a PLNC perspective there are three key observations to be made:

1) *Signal fading is linear:*

   Between the transmitter and a receiver, an EM signal experiences a linear transformation. This transformation is primarily caused by reflections and multipath propagation. Assuming band-limited communication the respective signals can be represented uniquely by complex-valued discrete-time samples. Then the received signal at any point in space can be expressed as the convulsion of the transmitted signal with an impulse-response function that characterizes the signal propagation. So a usual approach is to model this impulse response function in a statistical way instead of modeling exactly the physical surroundings.

   Assuming flat (i.e., frequency non-selective) fading, as it would be appropriate for narrow band communication, this impulse response function reduces to a delayed Dirac delta function whose height and time-delay characterizes the signal. This time delay may be viewed as a phase-shift in the discrete-time simulation model. This particularly simple model is the one considered in the previous chapter and is now applied to a network. Taking into account the concept of time-slot, the induced signal can be expressed as $hx[t]$ where $x[t]$ is the sequence of complex symbols transmitted at the symbol transmission period T_s. The prevalent model in the literature assumes a random fading h taken from a zero-mean circularly symmetric complex Gaussian distribution with unit variance, as it

was also considered in Chapter 3. As well as in the previous chapter the strategies discussed below require CSIR, which is motivated by signal measurements that can be acquired at the receiver. Notice that the transmitter is oblivious to h.

2) *Multiple signals interfere in a linear additive way:*

Instead of considering one transmitter only, let us now consider L nodes transmitting simultaneously. According to *observation 1*, the induced signal at any point in space can be expressed as

$$\sum_{l=1}^{L} h_l x_l[t] \,. \tag{4.1}$$

It is commonly assumed that the respective fading coefficients h_l are independent from each other since each node transmits from a different position in space. Throughout this chapter all those coefficients follow also a zero-mean circularly symmetric complex Gaussian distribution with unit variance and, in order to have an independent and identically distributed Rayleigh fading channel model, the phase of each h is uniformly distributed in $[0\ 2\pi[$, and its amplitude has a Rayleigh distribution (cf. chapter 3).

3) *Noise is independent of the signal and added at the receiver:*

At any receiving antenna one has to consider some additive noise, denoted by n which is taken from an independent circularly symmetric complex Gaussian with zero average and variance σ_n^2 corresponding, usually dubbed as the ZMSW noise model.

Taking into account these three observations along with several other detailed considerations [31] it is possible to define the commonly used model for the signal at a particular receiver when multiple L nodes are simultaneously transmitting as

$$y[t] = \sum_{l=1}^{L} h_l x_l[t] + n[t]. \tag{4.2}$$

Notice that the errors caused by the noise always accumulate over the various stages of the network, which is an undesirable element for linear network coding at the physical layer.

In order to better explain the model implemented other detailed considerations have to be made, especially for the understanding of the matched filter. As previously mentioned, this model works with the complex baseband representation of discrete time signals. Figure 4.1 shows the communication system diagram assumed in this dissertation and Figure 4.2 the full system diagram in continuous time.

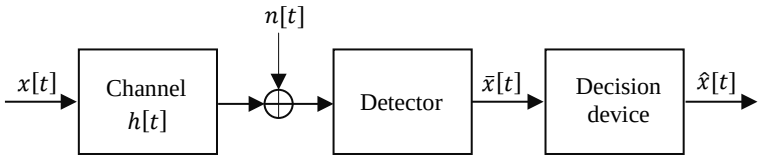

Figure 4.1. Simplified system diagram in discrete-time.

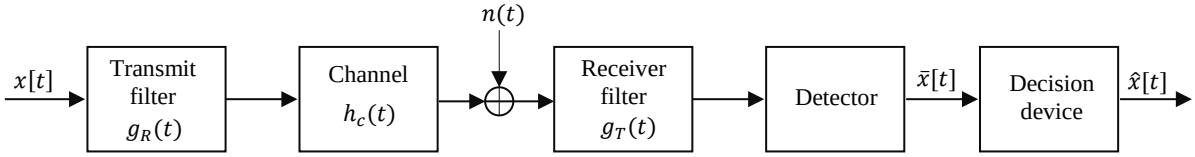

Figure 4.2. Full system diagram in continuous-time.

In Figure 4.2 the real and imaginary components of $x[t]$ are sequences of pulse amplitudes, which are each passed through a transmit filter g_T. The filtered signals are then modulated, giving rise to both in-phase and quadrature components transmitted over the channel by the carrier. The channel is modelled as a linear filter $h_c(t)$ and the Gaussian noise, n_c, is added at the receiver, as considered in *observation 3)*. Note that the modulation and demodulation stages are not represented in the figures. After demodulation, the base-band signal is passed through a receiver filter g_R and then sampled for further discrete-time processing. Assuming that the samples are taken at the optimal detection instant, the discrete-time channel impulse response may equivalently be defined by the convolution (denoted by the symbol "$*$"), as shown for example in [32],

$$h(t) = g_T(t) * h_c(t) * g_R(t). \tag{4.3}$$

In the conventional communications approach, the receive filter $g_R(t)$ is matched to the convolution of the transmitted pulse $g_T(t)$ and the channel impulse response $h_c(t)$, i.e.,

$$c(t) = g_T(t) * h_c(t)\,, \tag{4.4}$$

while the matched filter is defined as

$$g_R(t) = G_R c^*(-t)\,, \tag{4.5}$$

where G_R denotes some real gain and $c^*(t)$ is the conjugate of $c(t)$ [32]. This matched filter weights the received signal according to the signal amplitude and also aligns its phase to maximize the output SNR [2] and this maximises the SNR at some optimal sampling instant. The signal then becomes

$$h^* y[t] = h^* h x[t] + h^*\,, \tag{4.6}$$

and therefore $\bar{x}[t]$ (see Figure 4.1) can be obtained as,

$$\bar{x}[t] = \frac{h^* y[t]}{\|h\|^2} = \frac{h^* h x[t]}{\|h\|^2} + \frac{h^* n}{\|h\|^2} = x[t] + \frac{h^* n}{\|h\|^2}. \tag{4.7}$$

Using this matched filter, which, when generalised to the detection of vector messages, is known as the *maximal ratio combiner*, the effects of the channel are typically compensated. In the discrete-time domain this can be interpreted as applying a normalized detection filter [32],

$$h_{MF} = \frac{h^*}{\|h\|^2}. \tag{4.8}$$

In short, the estimate obtained by the matched filter can be represented as,

$$\hat{x}[t] = Q_C\left[h^{-1} y[t]\right]. \tag{4.9}$$

Notice that the compensation of the effects of the channel performed by this matched filter is different to the ones achieved by $\mathbf{H}^{\dagger}$, which was introduced in section 3.3. This is because the matched filter cannot deal with interference, treating everything as noise. On the other hand, zero forcing focuses on removing the interference only and then considers noise. The matched filter can only be used when *only* one message is expected to exist in the signal, cleared from all other interference coming from other users or antennas, which is seldom the case in the systems tested in this work. Notice too that h should be denoted by $h[t]$, however, given that in the remainder of this dissertation only slow fading channels are considered, meaning that they stay constant over many transmitted symbols, these coefficients were simply denoted by h.

Perfect synchronization between the nodes is assumed in this dissertation, however, synchronization is for long time an active research problem in wireless networks. Symbol time and carrier-frequency synchronizations, which are both needed in PLNC, have been actively investigated by researchers in fields such as OFDMA [33].

There are three areas of PLNC research: communication-theoretic, information-theoretic and networking aspects, out of which, networking has received the least attention. Nevertheless, this issue becomes even more important when we extend the application of PLNC beyond TWRC. PLNC in TWRC have been intensively researched in the recent PhD dissertation [34], where a strategy exploring the intersection of PLNC, MIMO, and the Alamouti coding scheme are proposed. However, that research did not venture beyond the TWRC. As soon as the theoretical understanding of this scenario matures, it is most likely that the research focus will move towards the application of PLNC to general topologies. Up until now, the literature on this area is scarce and it is difficult to find examples to compare with.

The goal of this chapter is to propose some strategies to implement PLNC beyond TWRC, while including MIMO links and detection strategies. Notwithstanding a first attempt to combine PLNC with MIMO will be start with the TWRC. More specifically, the assumed scenario consists of a certain number of terminals that want to exchange messages among them, while knowing that there is no direct link between them. Therefore they have to revert to a communication made through a relay. All the nodes are considerer half duplex, i.e., they do not transmit and receive at the same time.

In all the strategies proposed, the uplink (from the terminals to the relay) consists of decentralized transmitters (mobile terminals) and, on a first iteration of the communication protocol, a central receiver (MIMO relay), which in the following phase becomes a central transmitter. Considering terminals equipped with a single antenna and a relay equipped with a number of antennas equal to the number of terminals, it can be considered as a distributed MIMO system (multipoint-to-point). The uplink direction of any multiuser mobile communication system is an example of a *distributed MIMO* system. The role of the base station is to recover the individual users' signals from its received signal, and because a number of users transmit at the same time in the same band, this received signal is the superposition of all the active users' signals [23]. Figure 4.3 illustrates the scenario considered in the uplink.

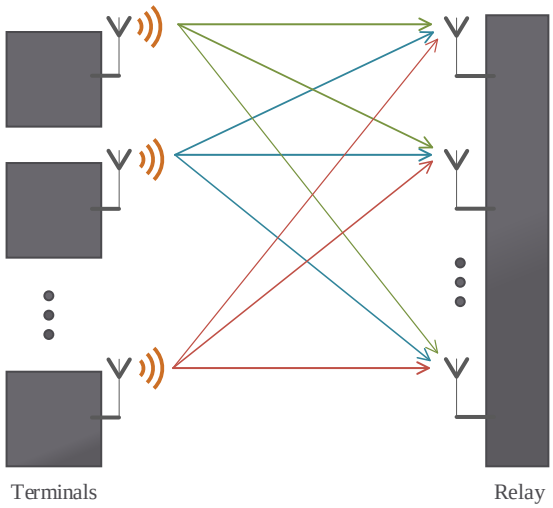

Figure 4.3. Distributed MIMO at the uplink phase.

Given that the uplink scenario is a MIMO system, although it is a distributed one, all the detection strategies discussed in Chapter 3 can be applied.

The MIMO relay receives the sum of the signals coming from the different terminals and detects the message sent by each user. Then, in the downlink, it chooses a strategy to send the messages back to the terminals in order to succeed the exchange of messages between them. After receiving the sum of the signals sent by the MIMO relay each terminal attempts to obtain the

messages from the other terminals. Very importantly, in this dissertation all channels are considered reciprocal, i.e., they are the same for the uplink and downlink phase.

The strategies used at the receivers in the downlink are based in a PLNC perspective since they take advantage of the fact that propagation in wireless medium results in a signal that is a linear combination of the signals sent by the MIMO relay, i.e., the receivers know which are those linear combinations once it is considered CISR and so they can cancel the channel effect. However the decoding results may not be precise due to the induced noise (see equation (4.2)).

Furthermore the signal that each terminal receives in the downlink contains, in the majority of cases discussed, the message that the terminal itself sent in the uplink phase. Consequently each terminal can subtract the contribution of its message from the downlink received signal, what can be seen as self interference cancelation. Then each terminal tries to decode the messages from the other ones using this "reduced" signal, performing a detection strategy that will be described later on for each specific scenario.

Besides the traditional TWRC configuration, this dissertation proposes to combine MIMO techniques with PLNC in a scenario with three users and a relay. While in the TWRC case the traditional scheme takes 4 time-slots to accomplish the exchange of messages, as it was explained in Chapter 1, it is straightforward to intuit that in the new scenario, with three terminals and a relay, the permutation of the messages takes 6 time-slots.

There will be considered two strategies for the exchange of messages in the scenario with three terminals: one that uses just 2 time-slots, improving the throughput in 150%, and another one that uses 3 time-slots, which have an improvement of 100% compared to the traditional scheme.

For all the cases discussed, one can assume that the MIMO relay has a high processing capacity and because of that the uplink will always use LRA OSIC-MMSE detection and the uplink takes just one time-slot making use of the distributed MIMO system. This option was taken because LRA OSIC-MMSE was the detection strategy that achieved the best performance among the others in the previous chapter, excluding ML detection that has exponential time consumption.

The SER of the downlink phase it is obtained comparing the messages decoded by each terminal with the original messages sent by them in the uplink phase. Notice that the downlink performance suffers from errors due to the previous uplink detection in the relay. This reinforces the choice of LRA OSIC-MMSE in the MIMO relay as an attempt to minimize the impact in downlink performance.

As it was stated in Chapter 3 the newly implemented strategies performance results will be plotted as *symbol error rate* (SER) as function of the SNR defined in (3.4) and both terminals and

relay will make use of the same M-QAM constellations. Notice that for simplicity of results the plots will just have one curve for the uplink, regarding the MIMO relay, and another one for the downlink representing just one terminal since the others will present a quite similar performance.

The following section will present in detail the detections methods implemented at the downlink for the TWRC and for the three terminals with a relay scenario. As was already mentioned the latter will be divided in the two time-slot and three time-slot approaches. One last strategy presented is a two time-slot strategy where the terminals are MIMO terminal equipped with two antennas.

4.2 MIMO combined with PLNC in TWRC

The MIMO combined with PLNC in TWRC strategy is illustrated in Figure 4.4 and Figure 4.5. The two terminals are equipped with a single antenna and the relay with two antennas. The channel coefficients h are also depicted in both Figures and the messages exchanged are denoted by $x \in C$.

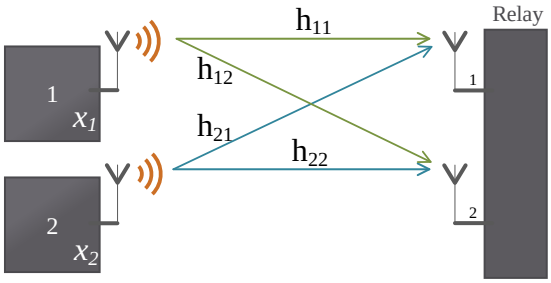

Figure 4.4. Uplink phase on MIMO combined with PLNC in TWRC.

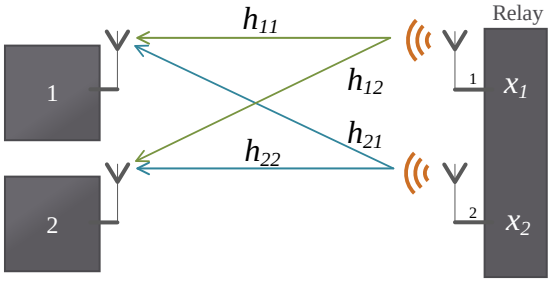

Figure 4.5. Downlink phase on MIMO combined with PLNC in TWRC.

In the uplink terminals 1 and 2 send their messages to the relay. As in an ordinary 2×2 distributed MIMO system the relay receives the messages x_1 (from terminal 1) and x_2 (from terminal 2). Then the relay sends x_1 in one antenna and x_2 in the other one. This transmitting configuration is

assumed to be known in advance by the terminals and therefore, accordingly with what was described in section 4.1, the received signal, for example, of terminal 1 is given by

$$y_1 = h_{11}x_1 + h_{12}x_2 + n\,. \tag{4.10}$$

The terminal possesses channel state information (CSIR) and knows its own sent uplink message. Hence the first step of the decoding process amounts to canceling its own contribution from the received signal, resulting in

$$y_1' = y_1 - h_{11}x_1. \tag{4.11}$$

Making use of a simple matched filter that cancels the channel linear transformation experienced by message x_2, terminal 1 can obtain a noisy version of the message x_2 and then apply a quantization step,

$$\hat{x}_2 = Q_C[h_{12}^{-1}\, y_1']\,. \tag{4.12}$$

The aforementioned procedure is followed by terminal 2 in an equivalent way. The performance of this strategy is presented hereafter in Figures 4.6 and 4.7 for the 4 and 16-QAM constellations, respectively, recalling that uplink performance is achieved by a LRA OSIC-MMSE receiver. The performance results of this strategy are lower than the ones presented in [34] for TWRC, since here Alamouti coding scheme was not used. However the throughput of the present strategy is improved by 100% compared with the traditional scheme as it was discussed in Chapter 1, obtaining a reduction of communications from 6 time-slots to 2 time-slots.

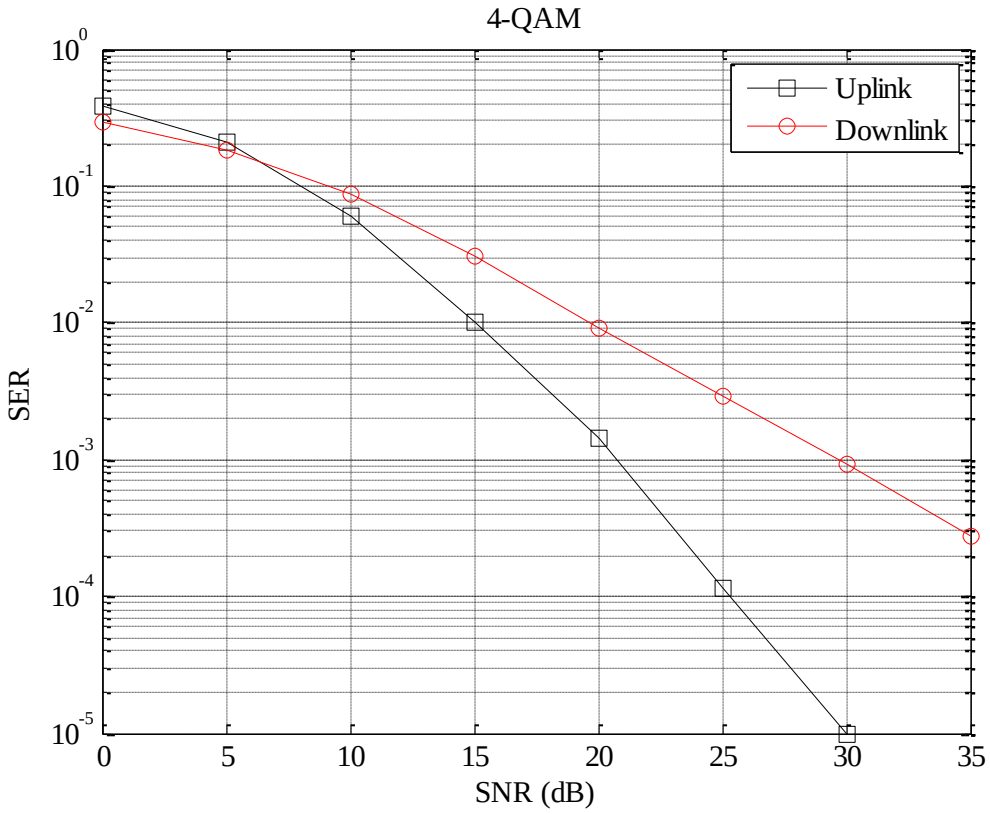

Figure 4.6. Detection MIMO combined with PLNC in TWRC using 4-QAM.

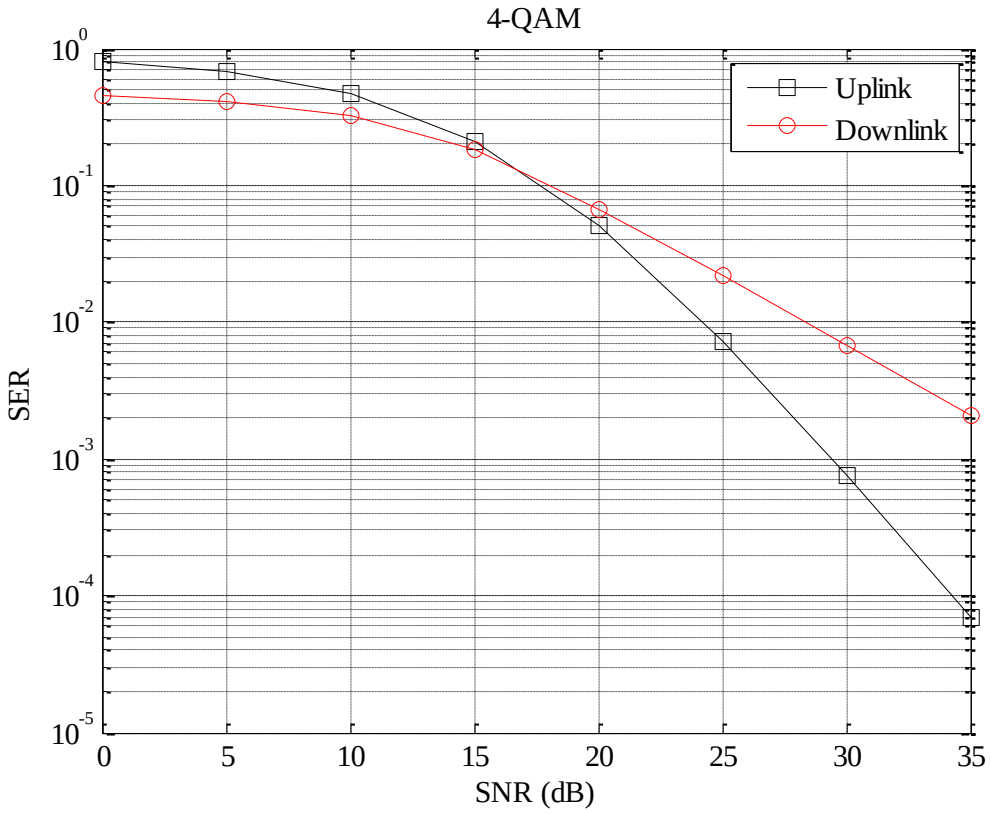

Figure 4.7. Detection MIMO combined with PLNC in TWRC using 16-QAM.

4.3 PLNC in a Network with a MIMO Relay

The scenario assumed in the following subsections consists of three terminals which aim to exchange messages between them. The messages are symbols $x_i \in C$, where the index i refers to the terminal from where the symbol was originally transmitted in the uplink. There is a direct link between the terminals and the relay, which means that the terminals do not receive any signal from other terminals and can only communicate through the relay. The uplink consists of a simple case of a distributed 3×3 MIMO system with a relay equipped with three antennas while the terminals only use one antenna to transmit their messages to the relay. This stage takes only one time-slot and the MIMO relay makes use of a LRA OSIC-MMSE receiver to decode the messages sent by the terminals. The uplink phase is equal for all the strategies discussed along this section, and therefore only the different downlink phases will be presented in the following sub-sections. Figure 4.8 illustrates the uplink phase and the channel coefficients considered.

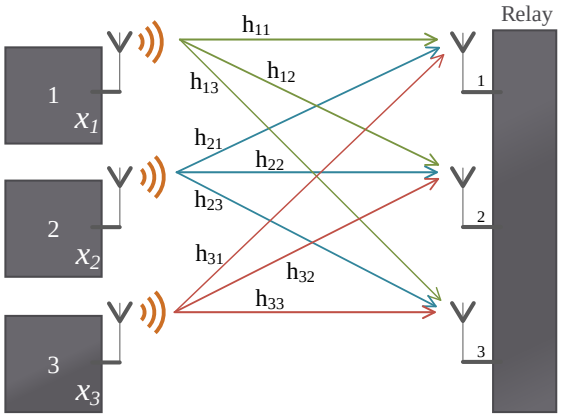

Figure 4.8. Uplink phase with 3 terminals.

4.3.1 Two Time-slots Strategy

In the uplink the MIMO relay decodes the messages x_1, x_2 and x_3 coming from the three terminals. In the two time-slots strategy the downlink phase, depicted in Figure 4.9, consists in the transmission of those messages where each antenna of the relay transmits one of the received messages, i.e., antenna 1 of the relay transmits x_1, antenna 2 transmits x_2 and antenna 3 transmits x_3, all at the same time in the second time-slot.

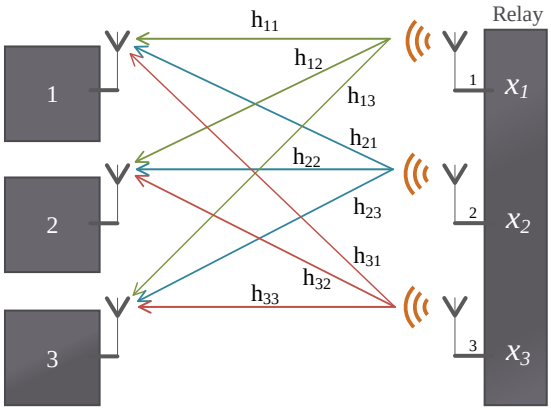

Figure 4.9. Downlink phase on two time-slot strategy with 3 terminals.

The above-mentioned transmitting configuration is assumed to be known in advance by the terminals and therefore, after the linear transformation performed by the wireless medium described in section 4.1, the received signal at each terminal is given by

$$y_1 = h_{11}x_1 + h_{12}x_2 + h_{13}x_3 + n\,, \tag{4.13}$$

$$y_2 = h_{21}x_1 + h_{22}x_2 + h_{23}x_3 + n\,, \tag{4.14}$$

$$y_3 = h_{31}x_1 + h_{32}x_2 + h_{33}x_3 + n\,. \tag{4.15}$$

The coefficients h are the ones represented in Figure 4.9 and y_i denote the received signal at the i^{th} terminal. Since all terminals possess CSIR and knows its own sent uplink message, the first step of the decoding process amounts to canceling its own contribution from the received signal, resulting in

$$y_1' = y_1 - h_{11}x_1\,, \tag{4.16}$$

$$y_2' = y_2 - h_{22}x_2\,, \tag{4.17}$$

$$y_3' = y_3 - h_{33}x_3\,. \tag{4.18}$$

The y_i' obtained are now decoded making use of brute force (ML) detection discussed in section 3.2 allowing each terminal to estimate the other two messages. In this first configuration, it is very important to note that this ML detection is made using the projections of the signals in just one dimension, since the y_i' are just a scalar and not vectors, as in section 3.2. Therefore, signal detection will be much affected by noise given that in this projective space the Euclidean distance between symbols is not lower bounded which eventually contributes in a decisive way to the poor performance

of this strategy. It is also worth pointing out that the downlink performance suffers from errors due to the previous uplink detection in the relay.

This strategy was simulated for 4 and 16-QAM modulation and the performance results are presented in Figures 4.10 and 4.11. One can conclude that the performance of such strategy, apart from the higher complexity due to ML detection, is lower than the one obtained in MIMO combined with PLNC in TWRC case, even though it concerns three terminals. Nevertheless, the throughput of the present strategy is improved by 150% compared with the traditional scheme, reducing the communications from 6 time-slots to 2 time-slots.

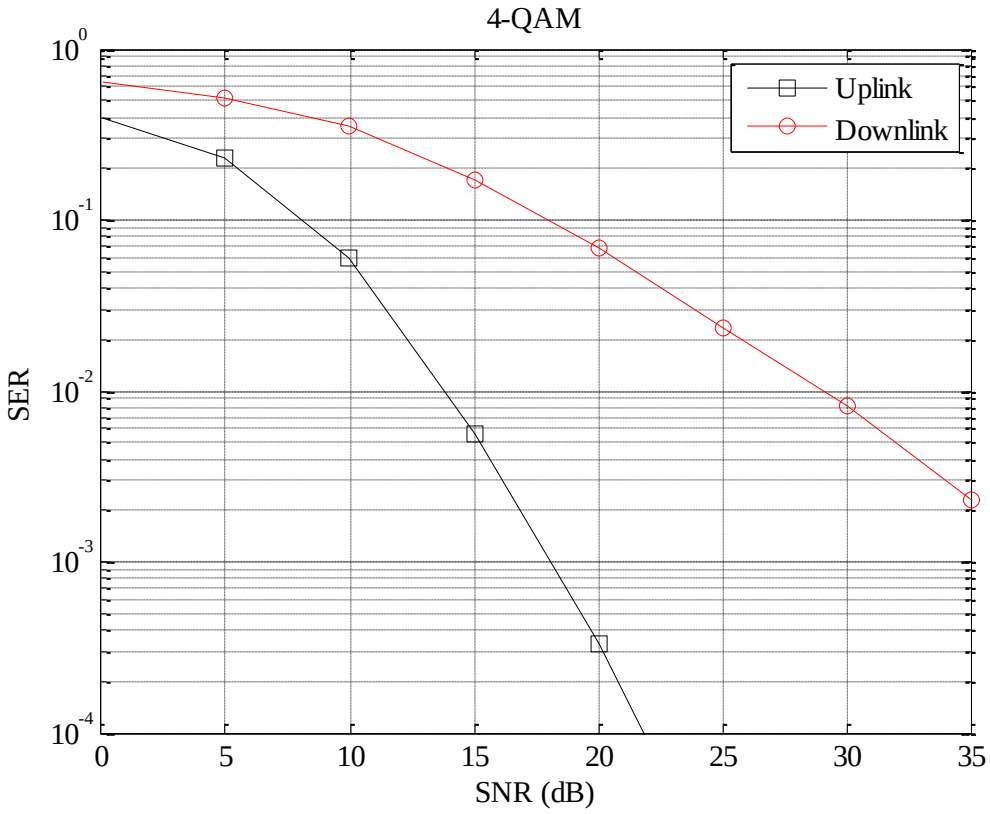

Figure 4.10. Detection two time-slot strategy with 3 terminals using 4-QAM.

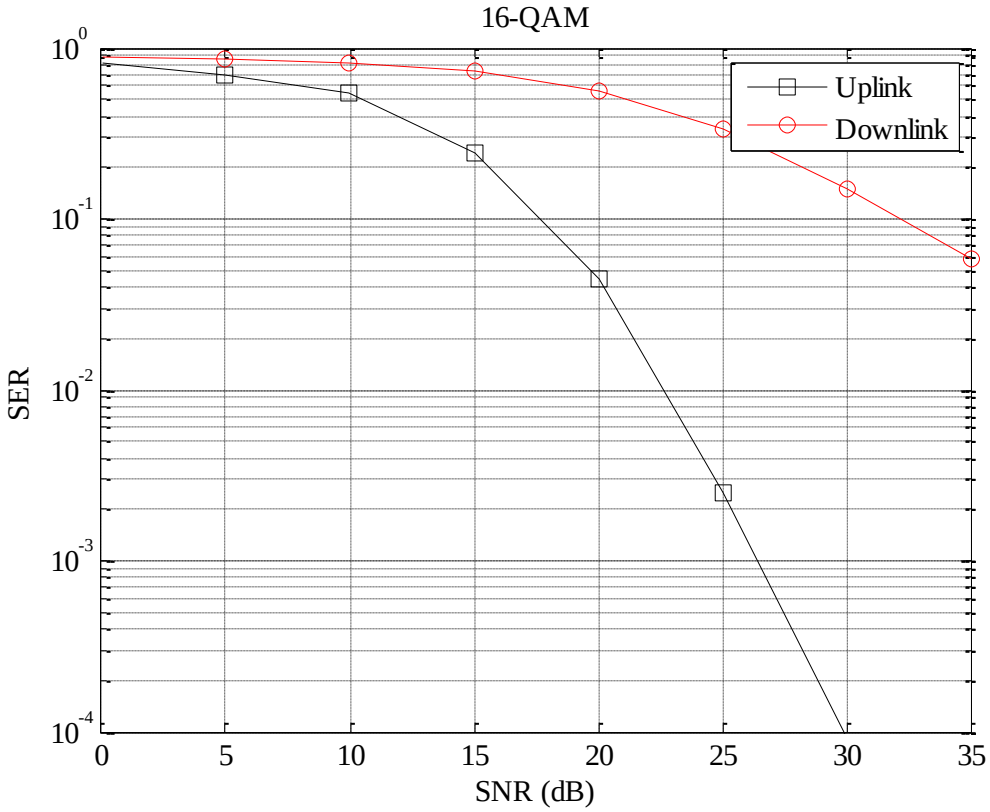

Figure 4.11. Detection two time-slot strategy with 3 terminals using 16-QAM.

4.3.2 Three Time-slots Strategy

Considering the scenario described for these sub-sections in the uplink the MIMO relay decodes the messages x_1, x_2 and x_3 in one time-slot. This three time-slot strategy uses 2 time-slots in the downlink phase, as illustrated in Figures 4.12 and 4.13. So in the second time-slot, which corresponds to the first one in the downlink phase, the relay transmits the messages x_1 and x_2 on antenna 1 and 2, respectively.

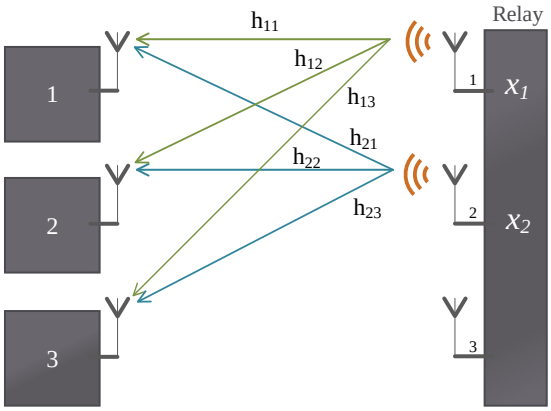

Figure 4.12. First time-slot from downlink phase on three time-slot strategy with 3 terminals.

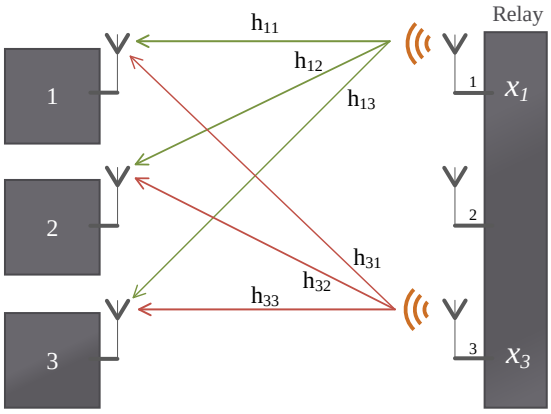

Figure 4.13. Second time-slot from downlink phase on three time-slot strategy with 3 terminals.

The above-mentioned transmitting configuration and the one considered for the second time-slot of the downlink phase are assumed to be known in advance by the terminals and therefore, taking into account the model described in section 4.1, the received signal at each terminal at the first time-slot of the downlink phase is given by

$$y_1[2] = h_{11}x_1 + h_{12}x_2 + n\,, \tag{4.19}$$

$$y_2[2] = h_{21}x_1 + h_{22}x_2 + n\,, \tag{4.20}$$

$$y_3[2] = h_{31}x_1 + h_{32}x_2 + n\,, \tag{4.21}$$

where the coefficients h are the ones represented in Figure 4.12 and $y_i[t]$ denote the received signal at the i^{th} terminal at the time-slot t. Notice that each terminal possesses CSIR.

At this point both terminal 1 and 2 know that they are receiving a signal that contains the message that they sent in the uplink phase so the first step of the decoding process amounts to canceling its own contribution from the received signal, resulting in

$$y_1'[2] = y_1[2] - h_{11}x_1\,, \tag{4.22}$$

$$y_2'[2] = y_2[2] - h_{22}x_2\,, \tag{4.23}$$

and making use of simple matched filter terminal 1 estimates $\hat{x}_2$ from (4.24) and terminal 2 estimates $\hat{x}_1$ from (4.25).

$$\hat{x}_2 = Q_C\left[h_{12}^{-1}\, y_1'[2]\right] \tag{4.24}$$

$$\hat{x}_1 = Q_C\left[h_{21}^{-1}\, y_2'[2]\right] \tag{4.25}$$

Terminal 3 saves the received signal in order to process it later. Resuming after the first time-slot of the downlink phase terminal 1 have $\hat{x}_2$ and terminal 2 have $\hat{x}_1$ while terminal 3 have the information in (4.21).

In the third time-slot, time-slot 2 of downlink (see Figure 4.13), the relay transmits the messages x_1 and x_3 on antenna 1 and 3, respectively, and the signals received by the terminals are

$$y_1[3] = h_{11}x_1 + h_{13}x_3 + n\,, \tag{4.26}$$

$$y_2[3] = h_{21}x_1 + h_{23}x_3 + n\,, \tag{4.27}$$

$$y_3[3] = h_{31}x_1 + h_{33}x_3 + n\,. \tag{4.28}$$

By now terminal 1 and 3 know that they are receiving their uplink original messages whereas terminal 2 can use its estimated $\hat{x}_1$. Therefore next step amounts to canceling each one of these contribution from the received signal at each terminal, resulting in

$$y_1'[3] = y_1[3] - h_{11}x_1\,, \tag{4.29}$$

$$y_2'[3] = y_2[3] - h_{21}\hat{x}_1\,, \tag{4.30}$$

$$y_3'[3] = y_3[3] - h_{31}x_1 \,. \tag{4.31}$$

and again using a simple matched filter both terminal 1 and 2 estimate $\hat{x}_3$ from (4.32) and (4.33), respectively, and terminal 3 estimates $\hat{x}_1$ from (4.34).

$$\hat{x}_3 = Q_C\left[h_{13}^{-1}\, y'_1[3]\right] \tag{4.32}$$

$$\hat{x}_3 = Q_C\left[h_{23}^{-1}\, y'_2[3]\right] \tag{4.33}$$

$$\hat{x}_1 = Q_C\left[h_{31}^{-1}\, y'_3[3]\right] \tag{4.34}$$

Terminal 1 and 2 have now all the messages. Finally, using the information in (4.21), terminal 3 can compute (4.35) and apply a simple matched filter in order to estimate $\hat{x}_2$ from (4.36).

$$y_3'[2] = y_3[2] - h_{31}\hat{x}_1 \tag{4.35}$$

$$\hat{x}_2 = Q_C\left[h_{32}^{-1}\, y_3'[2]\right] \tag{4.36}$$

This strategy was also simulated for 4 and 16-QAM modulation and the results are presented in Figures 4.14 and 4.15. One can conclude that the performance of such strategy accomplishes better results than the approach using just two time-slots described in subsection 4.3.1. Furthermore is clearly less complex than the previous one since only simple computation processes and simple matched filters are applied. However this strategy loses 50% of throughput improvement when compared with the two time-slot strategy but it still achieves a throughput improvement of 100% when compared with the traditional scheme, reducing the communication from 6 time-slots to 3 time-slots.

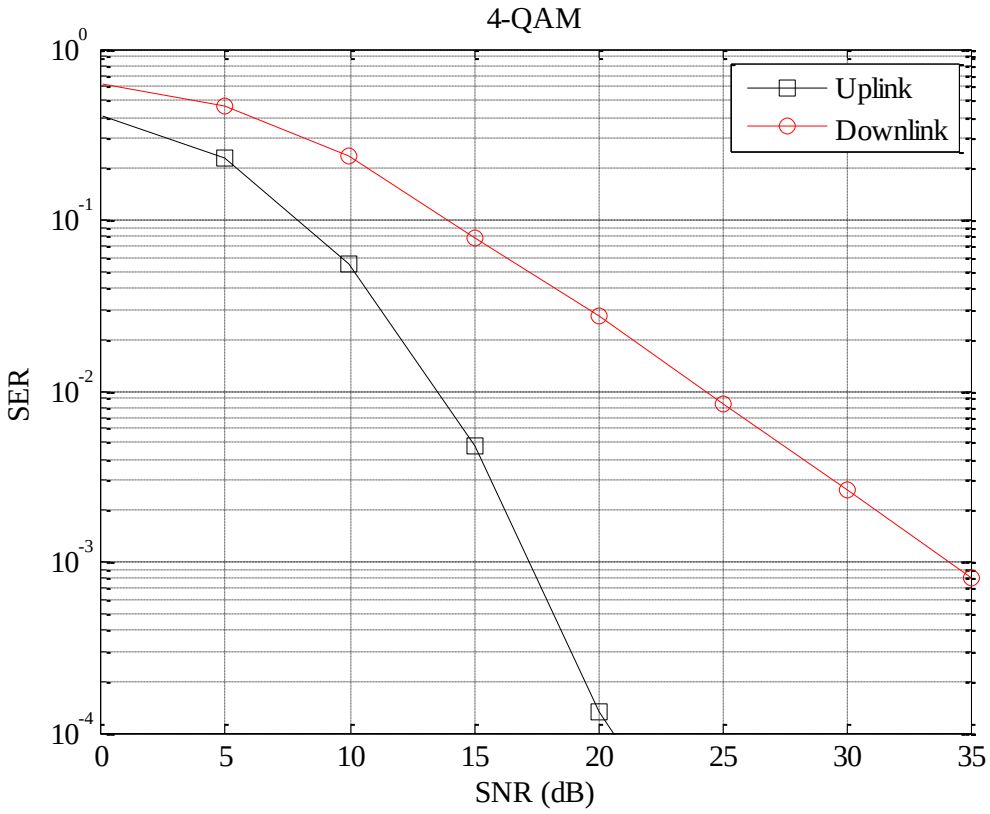

Figure 4.14. Detection three time-slots strategy with 3 terminals using 4-QAM.

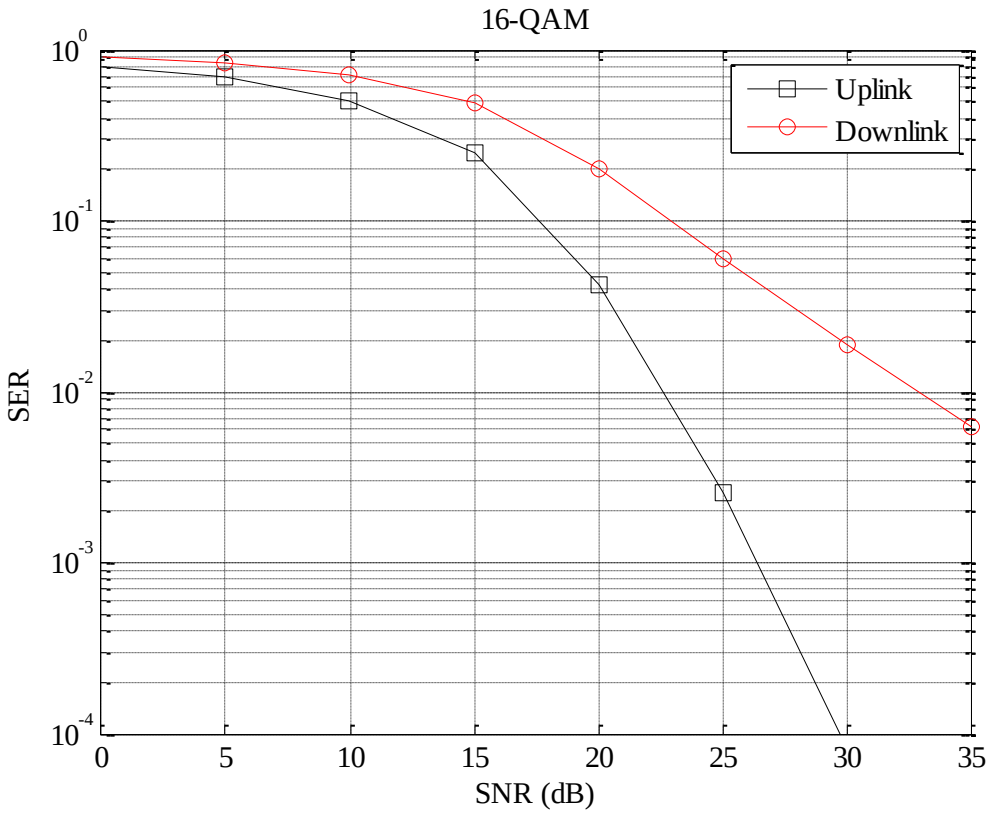

Figure 4.15 Detection three time-slot strategy with 3 terminals using 16-QAM.

4.3.3 Two Time-slots Strategy with MIMO Terminals

The high SER obtained on the strategy discussed in subsection 4.3.1 is explained mostly by the lost of dimensions in the downlink since ML detection is made using the projections of the signals in just one dimension as it was explained. Besides the use of brute force ML detection increases complexity at the terminals so in the present sub-section a new version of that strategy is proposed. The goal is to maintain the throughput improvement of 150% over the traditional scheme, gain performance and reduce complexity in the detection process taking advantage from a new dimension. The scenario considered for the uplink stays the one specified for section 4.3.

In order to attain this new dimension the two time-slots strategy with MIMO users assumes that terminals are now equipped with two antennas instead of a single one. The uplink remains as it was considered (see Figure 4.8) for all the strategies discussed in this section, so the users transmit their messages at the same time-slot using a single antenna as previously stated.

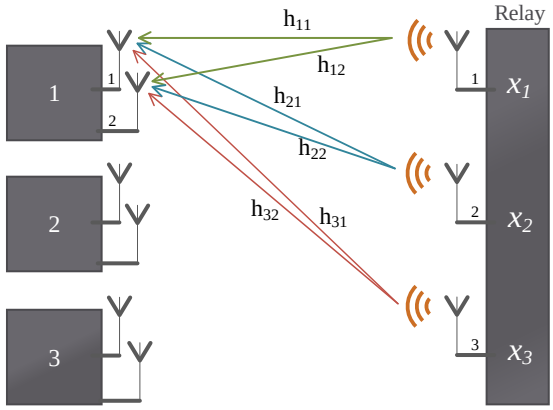

Figure 4.16. Downlink phase on two time-slot strategy with 3 MIMO terminals.

The reception on the downlink phase makes use of the two antennas of the terminals. Since the process of the detection will be the same for the three terminals it will only be reproduced the procedure example concerning reception on terminal 1 (see Figure 4.16).

MIMO relay decodes the uplink messages x_1, x_2 and x_3 then each antenna of the relay transmits one of the received messages, i.e., antenna i^{th} transmit x_i, all at the same time in the second time-slot. The transmitting configuration is assumed to be known in advance by the terminals and therefore the sum of the received signal at the two antennas of each terminal, in this example just terminal 1, is given by

$$y_{11} = h_{11}x_1 + h_{21}x_2 + h_{31}x_3 + n\,, \tag{4.37}$$

$$y_{12} = h_{12}x_1 + h_{22}x_2 + h_{32}x_3 + n\,. \tag{4.38}$$

The coefficients h are the ones represented in Figure 4.16 and y_{ij} denote the received signal at the i^{th} terminal and its j^{th} antenna. Since the terminal possesses CSIR and knows what was its own sent uplink message the first step of the decoding process amounts to canceling its own contribution from the received signal, resulting in

$$y'_{11} = y_{11} - h_{11}x_1, \tag{4.39}$$

$$y'_{12} = y_{21} - h_{12}x_1. \tag{4.40}$$

After calculate this self interference cancelation the system can be rewritten as (4.41) or equivalently (4.42).

$$\begin{bmatrix} y'_{11} \\ y'_{12} \end{bmatrix} = \begin{bmatrix} h_{21} & h_{31} \\ h_{22} & h_{32} \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} + [\mathbf{n}], \tag{4.41}$$

$$\mathbf{y}' = \mathbf{H}'\mathbf{x}' + \mathbf{n}. \tag{4.42}$$

Therefore accordingly with the model described in section 4.1 this system is equivalent to the one described in equation (3.1), once that the system model of Chapter 3 is also valid for the present chapter which means that the system of (4.42) can be interpreted as a MIMO communications of $N_T = N_R = 2$ antennas. Consequently all the detections strategies considered along Chapter 3 can be applied.

Summarizing in the two time-slots strategy with MIMO terminals each one of them exploits the linear codification induced by the wireless medium to reduce the detection problem to a 2×2MIMO equivalent system and then apply a detection strategy decoding the messages from the other two terminals. This strategy is simulated for 4 and 16-QAM modulation. As expected the performance of this scheme brings down the SER curve especially for the LRA decoders. Depending on the computation capacity of the terminals it can be chosen an appropriate decoding strategy requiring more or less computation and consequently more or less energy consumption. Within this perspective it will be presented the performances for OSIC-MMSE (see Figures 4.17 and 4.18) and LRA OSIC MMSE (see Figures 4.19 and 4.20) decoders once they are the ones that achieve a better performance and have a substantial difference of computation. Concluding PLNC combined with a MIMO relay and MIMO users guarantees the best performance of the scenario described in section 4.3, ensuring a polynomial computational time and having a throughput improvement of 150% competing with the traditional scheme. This strategy can be generalized for scenarios with N terminals equipped with $N-1$ antennas and a MIMO relay equipped with N antennas. For instance in a scenario with four terminals the throughput improvement raises up to 200% competing with the traditional scheme, reducing the communication from 8 time-slots to 2 time-slots.

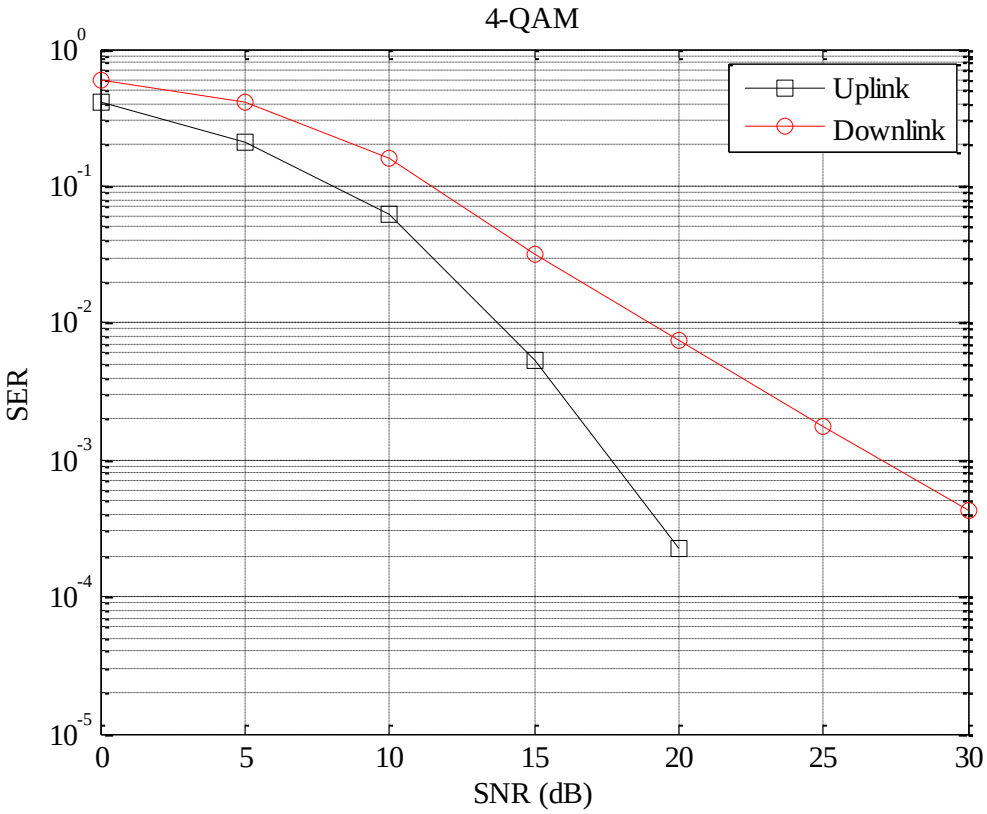

Figure 4.17. OSIC-MMSE using 4-QAM in two time-slot strategy with MIMO terminals.

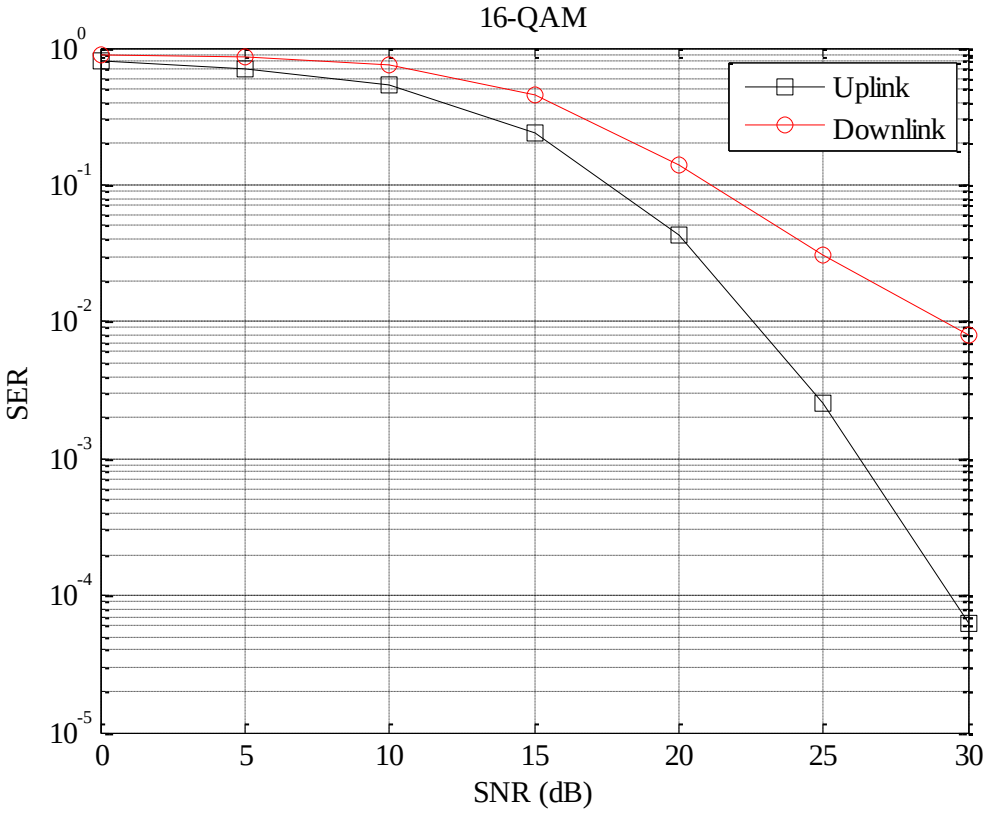

Figure 4.18. OSIC-MMSE using 16-QAM in two time-slot strategy with MIMO terminals.

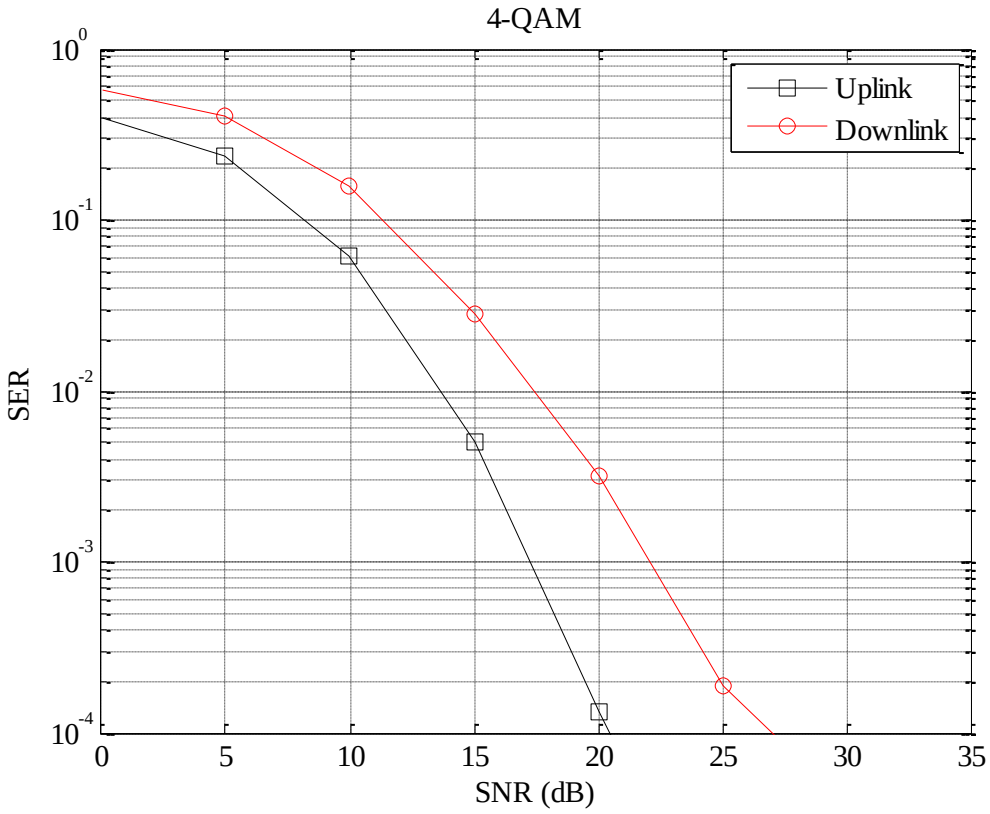

Figure 4.19. LRA OSIC-MMSE using 4-QAM in two time-slot strategy with MIMO terminals.

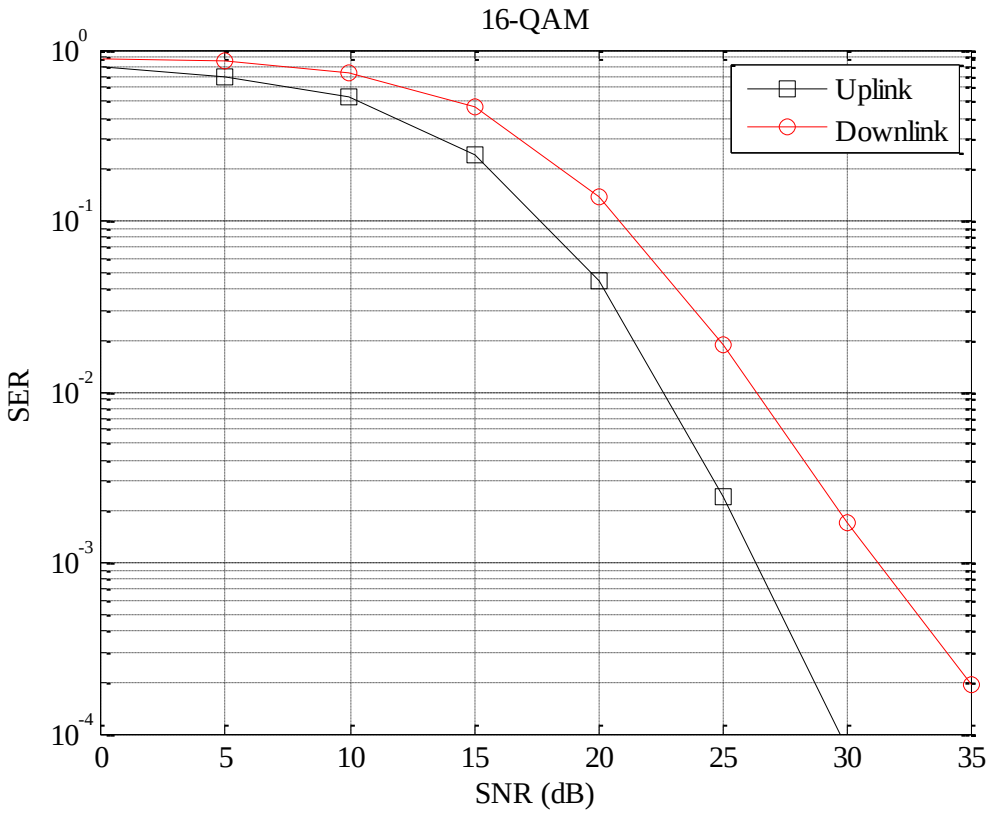

Figure 4.20. LRA OSIC-MMSE using 16-QAM in two time-slot strategy with MIMO terminals.

Chapter 5

Conclusions

The present chapter points to the conclusions of the dissertation and describes possible future directions for the work developed.

5.1 Main Conclusions

This thesis chiefly deals with the concept of interference cancellation when a receiver knows the channel that each of the interfering messages have gone through. This may be the case in "pure" MIMO, or when merging the concept of successive interference cancelation (SIC) with physical layer network coding, interpreting PLNC as a first stage of SIC.

In this dissertation lattices (described in Chapter 2) have a prominent role given that their structure is closely related to the MIMO detection problem, which has been a central research topic in MIMO communication in the last decade. From a lattice perspective, MIMO detection is known as the closest vector problem (CVP), as described in Chapter 3. In this chapter, several suboptimal solutions addressing this problem have also been described. Taking into account the detection techniques implemented for MIMO, the performance results point to the conclusion that although LRA receivers capture the same diversity order of ML detection, $d = N$, the LRA OSIC-MMSE receiver is the one that achieves the best performance since it is the one that has the smaller power penalty compared to the optimal ML detection. In some configurations the power penalty in respect to ML is reduced approximately to 1 dB only. It is remarkable that LRA OSIC-MMSE achieves this performance in polynomial time instead of the exponential performed by ML (which is chiefly due to the polynomial complexity of the celebrated LLL algorithm). On the other hand, excluding ML detection, this receiver is the one that needs more computation. However, it is important to notice that most of the complexity is not associated with the detection of each symbol vector but rather associated with the reduction of each matrix representing a channel. Consequently, the lattice reduction algorithm is only needed to be computed when the channel changes, which for slow fading channels is sparsely need. Apart from LRA receivers, the OSIC-MMSE receiver has the best performance since it is the one that minimizes the power penalty in comparison to ML detection. In general, ZF, MMSE, and OSIC receivers all have the same diversity order ($d = 1$) but OSIC receivers present a larger gain than the others. The performance gain from ZF to MMSE is small and tends to decrease with a higher number of antennas and higher-order M-QAM constellations. The implementation of complex-LLL not only reduces the complexity of the lattice reduction algorithm, but also achieves that without sacrificing any performance when compared with traditional real-valued LLL. Additionally, it also helps to reduce the complexity of other parts of the MIMO receivers involving matrix computations.

Chapter 4 puts in practice the concepts of PLNC (introduced in Chapter 1) and proposes a set of new strategies combining MIMO with PLNC in scenarios that move beyond the traditional TWRC. One should notice that in all these strategies, the uplink is actually a distributed MIMO system and in the downlink we looked at three different approaches, all of which make use of some sort of self-interference cancelation. The two time-slots strategy proved to have the lowest performance, which can be explained by the fact that ML detection is implemented using the projections of the signals in

just one dimension rather than deciding for symbols in a multi-dimensional lattice. Since ML detection is used, this strategy has exponential complexity; however, it has a 150% capacity increase in respect to traditional TDMA. The three time-slots strategy exceeds the performance of the two time-slots strategy and is clearly less complex because it only involves a simple cancelation of terms and a matched filter. This latter strategy achieves a 100% throughput improvement compared with the traditional scheme, but nevertheless it loses 50% of throughput improvement when compared with the previous one. Finally the two time-slots strategy with MIMO terminals is the most attractive. It provides a 150% improvement in respect to the traditional scheme and does it using simple computation processes in a first step and then can perform any of the suboptimal solutions found on Chapter 3. The superior performance achieved by this strategy is obtained using LRA OSIC-MMSE receivers in all terminals and also at the relay. One should note that this strategy can always be generalized to scenarios with N terminals, each of which equipped with $N-1$ antennas and the MIMO relay equipped with N antennas, allowing the exchange of messages in just two time-slots.

5.2 Future Work

A natural next step for this research is to implement the two time-slots strategy with MIMO terminals in scenarios with more than three terminals and demonstrate the generalization above mentioned to N terminals. A more elaborate extension would be implementing full-duplex nodes, i.e., nodes that transmit and receive at the same time in the same frequency band, combined with PLNC and MIMO techniques. In full-duplex configurations the detection processes need not only to consider the interference coming from other special streams or from other terminals, but self-interference needs also to be taken into account.

The properties of lattices made them central in the other branch of PLNC, which is the *Compute and Forward* approach, as introduced in Chapter 1 and that should probably become the main road to be explored in PLNC for instance by forcing an isomorphism between the lattice structure and group codes with worlds associated with lattice points, meaning that the sum of any two codewords always give rise to another valid codeword.

References

[1] F. Boccardi, B. Clerckx, A. Ghosh, E. Hardouin, G. Jöngren, K. Kusume, E. N. Onggosanusi and Y. Tang, “Multiple-antenna techniques in LTE-advanced,” IEEE Commun. Magazine, vol. 50, no.2, pp. 114-121, February 2012.

[2] D. Tse and P. Viswanath, “Fundamentals of Wireless Communication,” Cambridge University Press, New York, USA, 2005.

[3] S. Zhang, S. C. Liew, P. P. Lam, “Hot Topic: Physical-layer Network Coding,” ACM MobiCom’06, pp. 358-365, September 2006.

[4] P. Popovski and H. Yomo, “The anti-packets can increase the achievable throughput of a wireless multi-hop network,” in Proc. of IEEE Int. Conf. on Commun., Istanbul, Turkey, pp. 3885–3890, June 2006.

[5] B. Nazer and M. Gastpar, “Computing over multiple-access channels with connections to wireless network coding,” in Proc. of IEEE Int. Symp. on Inf. Theory, Seattle, USA, pp. 1354–1358, July 2006.

[6] Y. Wu, P. A. Chou, and S.-Y. Kung, “Information exchange in wireless networks with network coding and physical-layer broadcast,” Microsoft Research, Redmond, WA, Tech. Rep. MSR-TR-2004-78, August 2004.

[7] C. Feng, D. Silva and F. R. Kschischang, "An algebraic approach to physical layer network coding," in Proc. of ISIT'10 – The Inter. Symp. on Information Theory, Austin, TX, USA, pp. 1017-1021, June 2010.

[8] S. C. Liew, S. Zhang and L. Lu, “Physical-layer network coding: Tutorial, survey, and beyond,” Physical Communication, vol. 6, pp. 4–42, March 2013.

[9] B. Nazer and M. Gastpar, “Compute-and-Forward: Harnessing Interference through Structured Codes,” IEEE Trans. Info. Theory, to appear, June 2011.

[10] B. Nazer and M. Gastpar, “Reliable Physical Layer Network Coding,” Proceedings of the IEEE, vol. 99, no. 3, March 2011.

[11] B. Nazer and M. Gastpar, “Compute-and-forward: A novel strategy for cooperative networks,” 42nd Asilomar Conference on Signals, Systems and Computers, October 2008.

[12] O. Revev, lecture notes, "Lattices in Computer Science," Tel Aviv University, Fall 2009, http://www.cims.nyu.edu/~regev/teaching/lattices_fall_2009/.

[13] D. Micciancio, “Lattices Algorithms and Applications,” University of California, San Diego, Winter 2010.

[14] 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.

[15] X. Ma and W. Zhang, “Performance analysis for MIMO systems with lattice-reduction aided linear equalization,” IEEE Transactions on Commun., vol. 56, no. 2, pp. 309-318, February 2008.

[16] 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.

[17] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multiple antennas,” Bell Labs Syst. Tech. J., pages 41–59, Autumn 1996.

[18] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, “Simplified processing for high spectral efficiency wireless communications employing multi-element arrays,” IEEE J. Select. Areas Commun., 17:1841–1852, November 1999.

[19] G. J. Foschini and M. J. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” Wirel. Personal Commun., 6: 311–335, March 1998.

[20] P. Viswanath, D. Tse, and V. Anantharam, “Asymptotically optimal waterfilling in vector multiple-access channels,” IEEE Trans. Inf. Theory, vol. 47, no. 1, pp. 241–267, January 2001.

[21] A. Sibille, C. Oestges and A. Zanella, “MIMO From Theory to Implementation,” Academic Press, Elsevier, 2011.

[22] R. C. de Lamare, “Massive MIMO Systems. Signal Processing Challenges and Research Trends,” October 2013, http://arxiv.org/abs/1310.7282.

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

[24] P. Q. Phong and B. Vallée, “The LLL algorithm: survey and applications,” Heidelberg: Springer, 2010.

[25] L. Babai, “On Lovasz' lattice reduction and the nearest lattice point problem,” Combinatorica, vol. 6, no.1, pp. 1-13, 1986.

[26] H. Yao and G. W. Wornell, “Lattice-reduction-aided detectors for MIMO communication systems,” in Proc. Globecom, Taipei, China, November 17-21, 2002.

[27] C. Ling, “Towards characterizing the performance of approximate lattice decoding in MIMO communications,”4th International Sump. on Turbo Codes & Related Topics, April 2006.

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

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

[30] 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.

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

[31] A. Goldsmith, “Wireless Communications,” Cambridge Univ. Press, 2005.

[32] K. Su, “Detection and decoding of signals transmitted over linear MIMO channels,” PhD dissertation, Department of Engineering, University of Cambridge, 2005.

[33] M. Morelli, “Timing and frequency synchronization for the uplink of an OFDMA system,” IEEE Trans. Commun., vol. 52, no. 2, pp. 296-306, February. 2004.

[34] N. Xu, “Physical-layer network coding for MIMO systems,” PhD dissertation, Computer Science and Engineering, University of North Texas, May 2011.