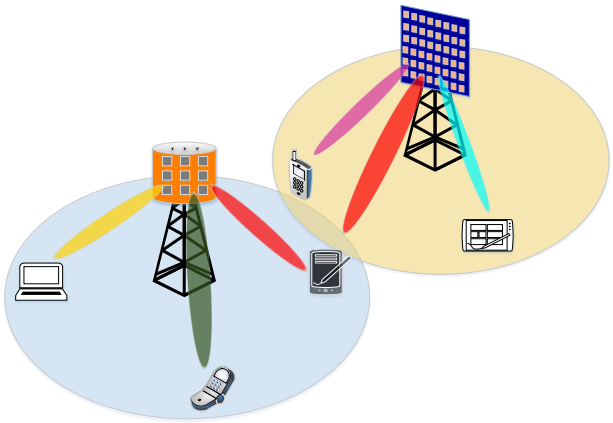
TÉCNICO LISBOA



Massive MIMO Processing for 5G Systems

Efficient Detection Algorithms and Applications

Francisco Miguel Teixeira do Rosário

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. José Eduardo Charters Ribeiro da Cunha Sanguino
Supervisor: Prof. António José Castelo Branco Rodrigues
Member of the Committee: Prof. Mário Alexandre Teles de Figueiredo

July 2015

To my family

Acknowledgments

I would like to express my special appreciation and thanks to my supervisors Professor António Rodrigues and Professor Francisco Monteiro for providing the opportunity to write this thesis. I also want to thank my friend and colleague João Lemos for his assistance in this project. Further, I would like to express my gratitude towards Professor Juan Acebrón for promptly providing the computer *Tsunami* to perform computationally intensive simulations. Additionally, I wish to express my sincere thanks to Professor João Xavier for the insightful discussions, without which part of this work would not have been possible. Last but not least, I can't help feeling a sense of gratitude towards Ivo Sousa for his time and support.

Resumo

A quinta geração de comunicações móveis tem como objectivo cumprir os requisitos em termos de capacidade, cujo crescimento na próxima década é previsto ser exponencial. Uma das propostas para satisfazer tal aumento é a extensão dos sistemas convencionais *multiple-input multiple-output* (MIMO) para um número massivo (largas dezenas a centenas) de antenas. Embora promissora, esta tecnologia traz alguns desafios que terão primeiro de ser superados. Um desses problemas é o aumento da complexidade em termos de processamento de sinal, que pode tornar-se incomportável quando técnicas tradicionais são utilizadas. É portanto necessário encontrar métodos que tenham uma performance tão perto do óptimo quanto possível e que, ao mesmo tempo, tenham baixa complexidade. Nesta tese, é dada uma especial ênfase a algoritmos para efectuar detecção e precodificação para sistemas MIMO de grandes dimensões. Em particular, quando o número de antenas na estação base é muito maior que o número de utilizadores a serem servidos, um método de processamento linear baseado na *Neumann series* e no lema de inversão de matrizes é proposto. Adicionalmente, algoritmos aleatórios para efectuar detecção em sistemas de MIMO massivo são exaustivamente estudados. Nomeadamente, uma variante de *Gibbs sampling* optimizada é sugerida, mostrando-se a sua superioridade em termos de performance face aos métodos existentes no estado da arte, incluindo uma técnica de *lattice reduction*. Finalmente, explorando as vantagens providenciadas pelo número elevado de antenas, um sistema repetidor *full-duplex* com MIMO massivo é avaliado. Este estudo inclui o desenho dos filtros de transmissão e recepção que, usando um esquema de alocação de potências óptimo, é capaz de atingir eficiências energéticas elevadas, satisfazendo em simultâneo os requisitos individuais em termos de taxa de transferência. Em todos os algoritmos aqui propostos, tanto a implementação em *hardware* como a adequabilidade para computação em paralelo foram tidas em consideração. Os resultados apresentados incluem exemplos ilustrativos e resultados numéricos elucidativos.

Palavras-chave: 5G, MIMO massivo, algoritmos de deteccção MIMO, *lattice reduction*, *Gibbs sampling*, *Neumann series*, *full-duplex*, alocação de potência óptima

Abstract

Fifth generation wireless networks aim to meet the requirements imposed by the ever increasing demand in capacity. One of the proposals under consideration is the extension of traditional multiple-input multiple-output (MIMO) systems to the use of a massive number (a few dozens to hundreds) of antennas. Even though promising, this technology brings some challenges that must first be surpassed. One of such problems is the increase in signal processing complexity, which might become unbearable when conventional techniques are used. Hence, one is to find methods that perform as close to optimal as possible, while yielding low complexity. In this thesis, a special emphasis on suitable algorithms to perform detection and precoding in large MIMO systems has been given. Particularly, for the case where the number of antennas at the base station is much larger than the number of served users, effective linear processing methods, based on the Neumann series and on the matrix inversion lemma, are proposed. Further, randomised algorithms to perform detection in symmetric massive MIMO systems are thoroughly studied. Namely, an optimised Gibbs sampling variant method is suggested and shown to outperform the existing ones in the literature, including state-of-the-art lattice reduction methods. Finally, capitalising on the advantages provided by large-scale arrays, a massive MIMO enabled full-duplex relay setup is evaluated. This includes the design of transmit and receive linear filters that, using a link dependent optimal power allocation scheme, is able to attain high energy efficiencies, whilst satisfying individual throughput requirements. In all of the proposed algorithms, both hardware feasibility and suitability for parallel computation have been taken into consideration. The results herein presented include illustrative examples and insightful numerical simulations.

Keywords: 5G, massive MIMO, MIMO detection algorithms, lattice reduction, Gibbs sampling, Neumann series, full-duplex, optimal power allocation

Contents

List of Figures

List of Tables

Nomenclature

Matrices and Vectors

$\mathbf{a}$	Column vector $\mathbf{a}$.
$\mathbf{A}$	Matrix $\mathbf{A}$.
$a_{i,j}$	Entry in the ith line and jth column of matrix $\mathbf{A}$.
$\mathbf{a}^{(k)}$	kth row of matrix $\mathbf{A}$.
$\mathbf{a}_k$	kth column of matrix $\mathbf{A}$.
$\det(\mathbf{A})$	Determinant of matrix $\mathbf{A}$.
$\lambda_i(\mathbf{A})$	ith eigenvalue of matrix $\mathbf{A}$.
$\|\|\mathbf{a}\|\|$	Norm of vector $\mathbf{a}$.
$\|\|\mathbf{A}\|\|_F$	Frobenius norm of matrix $\mathbf{A}$.
$od(\mathbf{A})$	Orthogonality deficiency of matrix $\mathbf{A}$.
$tr\{\mathbf{A}\}$	Trace of matrix $\mathbf{A}$.
$\mathbf{C}$	Covariance matrix.
$\mathbf{D}$	Diagonal matrix.
$\mathbf{e}_i$	ith column of identity matrix.
$\mathbf{I}$	Identity matrix.
$\mathbf{L}$	Lower triangular matrix.
$\mathbf{T}$	Unimodular matrix.
$\mathbf{Z}$	Gram matrix.

Operators and Sets

$\mathcal{CN}(\mu, \sigma^2)$	Complex normal distribution with mean value μ and variance σ^2.
$\delta(x)$	Dirac delta function.
$\mathbb{E}\{\cdot\}$	Expectation operator.
$\Im(x)$	Imaginary part of x.
Λ	Primal lattice.
Λ_D	Dual lattice.
$\log(x)$	base-e logarithm of x.
$\log_2(x)$	base-2 logarithm of x.
$\log_{10}(x)$	base-10 logarithm of x.

$\max(a, b)$ Maximum between a and b.

$\min(a, b)$ Minimum between a and b.

$\mathcal{N}(\mu, \sigma^2)$ Normal distribution with mean value μ and variance σ^2.

$\mathcal{O}(\cdot)$ Complexity order, big O notation.

$P(x)$ Probability of x.

$p(x)$ Probability distribution function of x.

$\mathcal{Q}(\cdot)$ Quantiser to the nearest neighbour.

$Q(\cdot)$ Q-function.

$\Re(x)$ Real part of x.

$\mathbb{C}$ Set of complex numbers.

$\mathbb{R}$ Set of real numbers.

$\mathbb{Z}$ Set of integer numbers.

$U[a, b]$ Uniform distribution between a and b.

$\mathbb{V}\text{ar}\{\cdot\}$ Variance operator.

$\lceil \cdot \rfloor$ Rounding to the closest integer.

$\lceil \cdot \rceil$ Smallest following integer.

Superscripts

$(\cdot)^\star$ Complex conjugate.

$(\cdot)^H$ Hermitian transpose.

$(a)^+$ Maximum between 0 and a.

$(\cdot)^\dagger$ Moore-Penrose pseudoinverse.

$(\cdot)^T$ Transpose.

MIMO Channel

$\mathcal{A}$, $\mathcal{B}$ Constellation alphabet.

C Capacity.

C_e Ergodic Capacity.

G_d Diversity Gain.

G_m Multiplexing Gain.

$\mathbf{H}$ Channel matrix.

$\mathbf{n}$ Additive white Gaussian noise.

N_R Number of receive antennas.

N_T Number of single-antenna users.

σ_n^2 Noise variance.

σ_x^2 Average power of transmitted symbols.

SNR Signal-to-noise ratio.

SNR_a Signal-to-noise ratio per transmit antenna.

$\mathbf{x}$ Transmitted symbol vector.

$\hat{\mathbf{x}}$ Estimated symbol vector.

$\mathbf{y}$ Received vector.

Chapter 4

β Ratio between N_T and N_R.

Δ_K Error in the K-term Neumann series inverse approximation.

$\mathbf{E}$ Hollow matrix.

f_{add} Number of FLOPS to perform an addition or multiplication operation.

f_{div} Number of FLOPS to perform a division operation.

f_{sqrt} Number of FLOPS to perform a square root operation.

γ Ratio between N_R and N_T.

K Number of terms in the Neumann series.

N_f Number of FLOPS.

$\mathbf{P}$ Partition of a matrix.

$\mathbf{r}$ Received vector after a linear filter.

θ Inverse of $(N_T + N_R)$.

U Number of updates to be performed.

$\tilde{\mathbf{Z}}_K^{-1}$ Approximation of $\mathbf{Z}^{-1}$ by the K-term Neumann series.

Chapter 5

α "Temperature" parameter.

$\mathbf{d}^{(t)}$ Difference in the maximum likelihood cost between consecutive iterations.

$f(\mathbf{x})$ Maximum likelihood cost of $\mathbf{x}$.

$\phi(\mathbf{x})$ Normalised maximum likelihood cost of $\mathbf{x}$.

$\overline{\mathbf{H}}$ Channel matrix including average gain.

$\tilde{\mathbf{H}}$ Reduced channel matrix after performing lattice reduction.

N Number of transmit and receive antennas.

n Number of real dimensions.

P Number of required restarts.

$q_1, q_{\text{opt.}}, q_\infty$ Mixing ratios for the "temperature" parameter.

R Number of restarts.

S Space state.

$\mathbf{x}^{(t)}$ Candidate vector at iteration t.

Chapter 6

α_{zf} Normalisation factor for the precoding filter.

$\mathbf{A}_{\text{zf}}$ Zero-forcing precoding filter.

$\beta_{\text{RD},k}$ Large-scale fading component of the kth receiving user.

$\beta_{\text{SR},k}$ Large-scale fading component of the kth transmitting user.

$\mathbf{D}_{\text{RD}}$ Large-scale fading matrix from the relay to the destinations.

$\mathbf{D}_{\text{SR}}$ Large-scale fading matrix from the sources to the relay.

$\mathbf{D}_{p_S}$ Diagonal matrix containing sources' power allocation.

$\epsilon_{\mathbf{H}}$ Variance in the error of the estimated matrix $\mathbf{H}$.

$\epsilon_{\mathbf{t}}$ Variance in the error of the transmitted vector $\mathbf{t}$.

$\mathcal{E}_{\mathbf{H}}$ Error associated with the estimation of channel matrix $\mathbf{H}$.

$\mathcal{E}_{\mathbf{t}}$ Error associated with the transmission of vector $\mathbf{t}$.

$\mathbf{F}_{\text{rx}}$ Loopback interference detection filter.

$\mathbf{F}_{\text{tx}}$ Loopback interference precoding filter.

$\mathbf{G}_{\text{RD}}$ Channel matrix from the relay to the destinations.

$\mathbf{G}_{\text{SR}}$ Channel matrix from the sources to the relay.

$\mathbf{H}_{\text{LI}}$ Loopback interference channel.

$\mathbf{H}_{\text{RD}}$ Small-scale fading matrix from the relay to the destinations.

$\mathbf{H}_{\text{SR}}$ Small-scale fading matrix from the sources to relay.

$\tilde{\mathbf{H}}$ Estimated channel matrix $\mathbf{H}$.

K Number of pairs being served by the relay.

$\mathbf{n}_d$ Noise vector at the destinations.

$\mathbf{n}_r$ Noise vector at the relay.

$\sigma^2_{\mathbf{n}_d}$ Average noise power at the destinations.

N_{rx} Number of receive antennas at the relay.

N_{tx} Number of transmit antennas at the relay.

$p_{\text{R},0}$ Peak transmit power at the relay.

p_{R} Power allocated at the relay.

$p_{\text{S},k}$ Power allocated at the kth source.

$p_{\text{S}_{0,k}}$ Peak transmit power at the kth source.

$\mathbf{R}_{\mathbf{n}_r}$ Relay noise covariance matrix.

R_k End-to-end achievable rate by pair k.

$R_{0,k}$ Required rate by the kth pair.

$R_{\text{RD},k}$ Achievable rate in the downlink channel by pair k.

$R_{\text{SR},k}$ Achievable rate in the uplink channel by pair k.

$\mathbf{r}$ Received vector at the relay.

$\hat{\mathbf{r}}$ Received vector at the relay after the loopback interference filter.

$\sigma^2_{\mathbf{n}_r}$ Average noise power at the relay.

σ^2_{LI} Residual loopback interference average power.

$\mathbf{t}$ Transmitted vector from the relay to the destinations.

$\hat{\mathbf{t}}$ Vector to be transmitted after the precoder.

$\tilde{\mathbf{t}}$ Vector to be transmitted from the relay to the destinations.

$\mathbf{W}_{\text{zf}}$ Zero-forcing detection filter.

$\mathbf{y}_d$ Received vector at the destinations.

$\mathbf{y}_{\mathrm{r}}$ Received vector at the relay after both detection filters.

Glossary

5G Fifth Generation of Mobile Networks

A-D Analog-to-Digital

AF Amplify-and-Forward

AWGN Additive White Gaussian Noise

BER Bit Error Rate

BS Base Station

CDF Cumulative Distribution Function

CDMA Code Division Multiple Access

CSI-R Channel State Information at the Receiver

CSI-T Channel State Information at the Transmitter

D-A Digital-to-Analog

DF Decode-and-Forward

e2e End-to-End

EE Energy Efficiency

ELR Element-Based Lattice Reduction

FBMC Filter Bank Multicarrier

FDD Frequency Division Duplexing

FDMA Frequency Division Multiple Access

FD Full-Duplex

FLOPS Floating-point Operations per Second

GFDM Generalised Frequency Division Multiplexing

GS Gibbs Sampling

I/Q In-phase/Quadrature

IBFD In-Band Full-Duplex

IoT Internet of Things

LAS Likelihood Ascent Search

LD Linear Detector

LI Loopback Interference

LLL Lenstra–Lenstra–Lovász

LOS Line-of-Sight

LRA Lattice Reduction-Aided

LR Lattice Reduction

LTE-A Long Term Evolution-Advanced

M2M Machine-to-Machine

MCMC Markov Chain Monte Carlo

MGS Mixed Gibbs Sampling

MIMO Multiple-Input Multiple-Output

ML Maximum Likelihood

MMSE Minimum Mean-Square Error

MR Multiple Restarts

MU Multiuser

NI Natural Isolation

NOMA Non-Orthogonal Multiple Access

OFDMA Orthogonal Frequency-Division Multiple Access

OGS Optimised Gibbs Sampling

OPA Optimal Power Allocation

OSIC Ordered Successive Interference Cancellation

OUPA Optimal Uniform Power Allocation

PDF Probability Density Function

RAT Radio Access Technology

RF Radiofrequency

RS Randomised Search

SDR Software Defined Radio

SD Sphere Decoder

SISO Single-Input Single-Output

SNR Signal-to-Noise Ratio

SNR$_a$ Signal-to-Noise Ratio per Transmit Antenna

SU Single-user

SV Sorted Variance

T-MGS Triple Mixed Gibbs Sampling

TDD Time Division Duplexing

TDMA Time Division Multiple Access

USRP Universal Software Radio Peripheral

V-BLAST Vertical-Bell Laboratories Layered Space-Time

Chapter 1

Introduction

The motivations behind this thesis will be detailed in this initial chapter. An overall structure of the document is provided in the last section, as well as the main contributions of this work.

1.1 Motivation

Over the past few decades we have been experiencing a rapid growth in mobile communications traffic. Everyday an increasing number of users demand more speed and reliability, anytime and anywhere. This fact is stimulating the industry to think ahead and plan future networks, which must be able to deliver an increased capacity and performance at a diminished cost. Recent studies forecast a total of 50 billion connected devices by 2020 [1], resulting in a need of a thousandfold gain in capacity per square kilometre.

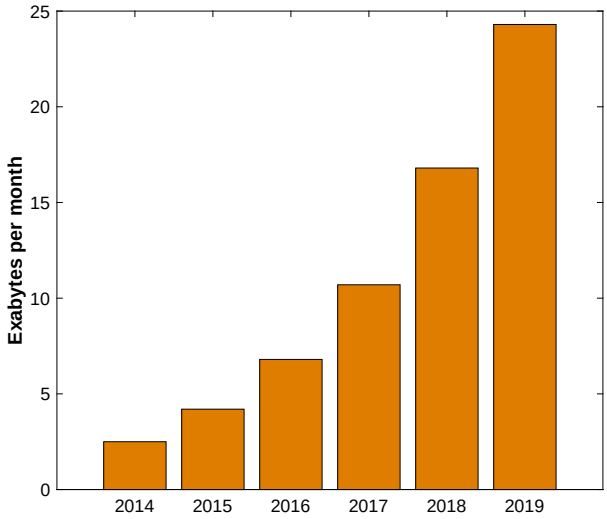
The deployment of the next and fifth generation of mobile networks (5G) is expected to emerge between 2020 and 2030. 5G will involve a combination of existing and evolving technologies, such as LTE-A[1] and Wi-Fi, and will have to support a large diversity of applications with a large variety of requirements, rendering an 'all-in-one' solution an inefficient one. Recent research has been focusing on how to deliver higher data rates, reduced latency and enhanced indoor coverage for an increased number of devices [2]. In order to meet these challenges, researchers in both academia and industry have been studying and developing promising techniques that will be present in future networks. Some of those novel approaches as well as existing ones will be discussed in the next section.

1.1.1 5G Requirements

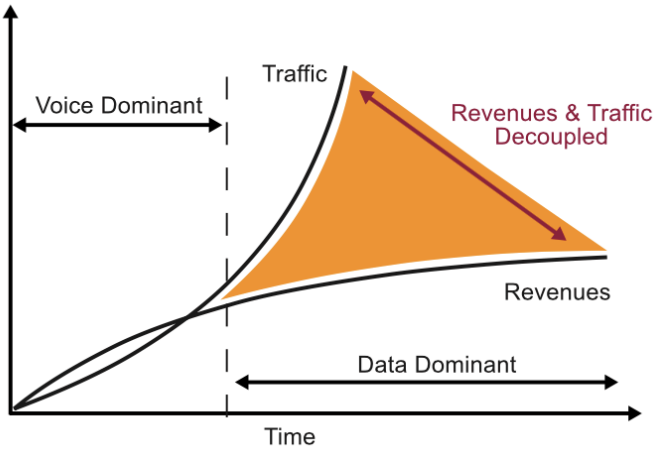
Thus far, mobile operators and the industry have been trying to follow the increasing trend of traffic with the successive release of new features and functionalities. The latest of those releases is LTE-A (Release 13) [3], whose deployments are growing globally, and expects to provide a higher capacity with the implementation of carrier aggregation, an improved spatial multiplexing scheme and support

[1] Long Term Evolution-Advanced; also known as the fourth generation of mobile networks.

for relay nodes. However, and despite the advances made so far, new market trends are imposing requirements which are forcing the development of a fifth generation of mobile networks. In addition to the current functionalities, this new system should be able to support innovative services, which so far were not possible due to capacity and latency constraints. Some examples are the Internet of Things (IoT) connectivity, enhanced e-health or even machine-to-machine (M2M) communications - the idea is to redefine human and machine networks in such a way that everything is connected at a zero-distance [4].



(a) Cisco forecasts an approximate increase of ten times in overall data traffic between 2014 to 2019. (based on data from [5]).



(b) Whilst mobile traffic volume is expected to increase exponentially, network operators' revenues have been following a logarithmic trend (reproduced from [6]).

Figure 1.1: Trends of traffic data volume in the next decade.

As a result, an exponential growth in traffic volume is expected to occur until the year of 2020, while the revenues are bound to remain nearly constant, as is depicted in figure 1.1. These facts raise an additional challenge, which is how to keep mobile networks profitable in the long term. Considering this

economic aspect and in order to fulfil the objectives described above, there are five high priority aspects that must be taken into account [1, 7, 8]:

Capacity 5G systems must be able to cope with traffic volumes which are various orders of magnitude larger than today's networks. With the most recent forecasts [7] and in order to give the impression of "infinite capacity", a thousand-fold gains per square kilometre will be needed when compared to the current generation of LTE-A.

Data Rates In general, the user experienced data rates must be higher than the current ones. Obviously, the requirements depend on the assumed scenario - office usage has different requirements from a pedestrian way situation, for instance. The idea is to deliver a better quality of service with responses to the requests perceived as instantaneous. To support that and as an illustrative example, end users should experience data rates of at least 1 Gb/s and 5 Gb/s in 95 and 20 percent of office locations, respectively, and during 99 percent of the busy period [7]. Locations with bad coverage, such as rural areas, should also be provided with rates of at least 100 Mb/s in the downlink and 20 Mb/s in uplink and latencies below 100 ms [7].

Connectivity 5G systems will have to deal with a large number of connected devices at the same time, of different types and with different requirements. Scenarios, such as very crowded places (e.g., stadiums or open air festivals) and regions with massive deployment of sensors and actuators, must be addressed in terms of connectivity. The target is to deploy a 100-fold gains in terms of possible number of simultaneously connected devices and still be able to provide rates of upto 20 Mb/s in the worst case scenario.

RAN Latency In order to give the zero-distance sensation to the end user, latency over the radio access network (RAN) should be less than 1 ms (5 times less when compared to LTE-A). This would enable the possibility for applications such as augmented reality, tactile Internet or even real-time control of machine systems.

Cost With the existing technologies, energy consumptions are following the same trend as the data traffic - an exponential growth. The vision for 5G systems should be more energy-oriented at the device, site and network levels, requiring the usage of efficient components, advanced modulation techniques and near optimal resource allocation algorithms.

As seen in the points described above, future generations will require drastic gains over current solutions. A viable option to tackle this problem is the combined use of more spectrum, higher spectral efficiency and network densification, leading to an overall reduced cost per bit. Improvements in these three directions could lead to the fulfilment of the desired goals. For the sake of demonstration, a simple example shall be looked at. Consider a given area where the network throughput is to be increased by a

thousand times. Making use of the three degrees of freedom described before, one has [9]:

$$\underbrace{\text{Throughput}}_{\text{bit/s/Area}} = \underbrace{\text{Available spectrum}}_{\text{in Hz}} \cdot \underbrace{\text{Cell density}}_{\text{Cell/Area}} \cdot \underbrace{\text{Spectral efficiency}}_{\text{bit/s/Hz/Cell}} . \tag{1.1}$$

An increase of ten times in each direction, for instance, would meet the required performance. An illustrative example of this can be seen in figure 1.2, which includes some of the techniques that will be discussed in the next section.

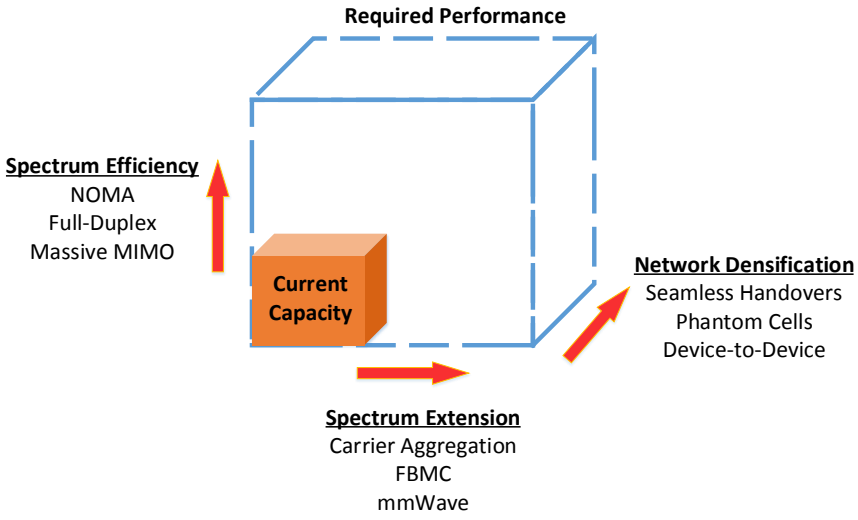


Figure 1.2: Cube representing the possible paths to attain the required capacity and some technologies currently being studied.

1.1.2 5G Design Aspects

As stated in the previous section, some of the challenges and requirements for the next generation mobile networks are already established. The techniques proposed by the industry which are potentially capable of satisfying those demands are going to be described here. Bear in mind that 5G is still a concept and hence the key technologies that will effectively be implemented are yet to be defined; every path is open. However, there is the need for an heterogenous network (HetNet), which means 5G will have to ”play nicely” with previous generations. Even though successive optimisations over the current network might cover some of the defined requirements, they will not suffice to accomplish the desired goals.

A crucial point in future deployments will be the usage of frequency bands above 3 GHz [10, 11], as below this value the spectrum usage is already saturated. The use of higher frequency carriers, denoted by millimetre waves (mmWaves), means more bandwidth is available, hence it becomes easier to attain the so desired higher data rates. So far, these bands were not used for mobile communications due to their large attenuation. However, if network density is increased, then distance between users and serving cells is reduced, enabling the use of these frequencies. Notwithstanding this fact, ultra-densification may

lead to unbearable levels of interference and, ultimately, zero throughput [12]; hence, a careful network planning is mandatory. Additionally, current modulation and detection schemes may not be optimum in these high frequency bands and, thus, new and more efficient waveform designs will be required.

The possible solutions described in the last paragraph will most likely require a new radio access technology (RAT), which has to have the capability of interworking with current systems. Therefore and in the following decades, cognitive radio networks will be of paramount importance in the maximisation of both spectrum and energy efficiencies. Some of the key features that could be implemented into the 5G radio access are [8]:

Phantom Cell The idea, proposed by DOCOMO, is to add a capacity layer to the coverage layer via the introduction of low power nodes in high traffic areas. These small cells would use higher frequencies for increased performance and throughputs. The control signalling and handover management would still be responsibility of a macro cell using lower frequencies.

Flexible Duplex One of the advantages of a cognitive radio network is its flexibility in terms of utilised carrier frequency and duplex scheme. The use of a smart combination of time division duplex (TDD) and frequency division duplex (FDD) schemes, along with an opportunistic carrier selection will play a key role in the 5G RAT.

Waveform Design 5G systems will use different spectrum bands, having different propagation conditions, and, thus, different requirements modulation-wise. While much of the processing could be optimisations of LTE-based schemes (e.g., OFDM[2]), some alternatives such as single carrier waveforms or other advanced multi-carrier waveform (e.g., filter bank multicarrier (FBMC) and generalised frequency division multiplexing (GFDM) [13, 14]) can also be utilised to prioritise coverage and support wider frequency bands as well as new scenarios.

Massive MIMO Multiple-input multiple-output (MIMO) systems with multiple antennas at the transmitter and receiver are already used to achieve higher capacity and link reliability by the exploitation of both spatial multiplexing and diversity. In order to deliver the required data rates, the use of antenna arrays with a very large number (tens to hundreds) of elements might become a reality. This technology will allow a much higher spectral efficiency with decreased energy levels, whilst serving a large number of users at the same time [15].

JSDM Another degree of freedom to be exploited is the correlation of the channels of the different users. Joint spatial division and multiplexing (JSDM) [16] to be used in multiuser MIMO downlink applies a clever pre-beamforming scheme that partitions users with approximately the same channel into groups. Hence, the transmitted signal may be designed in such a way that intergroup interference is minimised, leading to even higher spectral efficiencies.

[2]Orthogonal frequency-division multiplexing.

NOMA Hitherto an effort to use orthogonal signalling to differentiate users has been made (e.g., TDMA[3], CDMA[4] and OFDMA[5]). Non-orthogonal multiple access (NOMA) [17] methods are being studied in order to meet the various requirements of next generation networks. The idea is to transmit data from different users in the same slot (time and/or frequency) and by exploiting the power-domain (different path losses), advanced successive interference cancellation schemes can be used to successfully decode the transmitted data.

It is worth mentioning that the aim of any of these solutions is to either increase spectral efficiency or cell density, leading to higher throughputs for a larger number of devices. Moreover, they can be implemented together and be integrated with current technologies. An example of this type of combined solution is proposed in [18], where the benefits of a full-duplex relay station equipped with massive MIMO are presented. In the paper it is shown that, in theory, it is possible to efficiently serve a large number of users, while using very simple linear processing techniques and power allocation schemes with close to optimal performance. In the next few years, further technical studies and field experiments will have to be performed in order to bring the 5G concept into the real world.

1.2 Goals and Structure of this Document

In a succinct manner, the main objective of this thesis is the study and consequent implementation of algorithms that are suitable for large MIMO systems. Among other fields, this work includes the study of MIMO fundamentals, the benefits of scaling up MIMO, the understanding and implementation of existing state-of-the-art algorithms and a consequent comparative study. New methods and slight variations of the analysed algorithms are going to be proposed here. The development of this research project will involve knowledge from a large variety of signal processing techniques, such as lattice reduction, Markov chain Monte Carlo (MCMC) methods, and advanced matrix theory. When the author finds the topic to be out of the scope of the thesis, the reader will be referred to the proper bibliography. Nonetheless, the basic definitions and appropriate notation will be present when needed. Furthermore, and when the author finds it suitable, concepts and analytical results will be supported by graphics, tables and numerical outputs from simulations.

The organisation of this document in terms of chapters and respective contents will be as follows:

Chapter 2 - MIMO Overview: In chapter 2 a brief overview of MIMO systems will be given, including a review of fundamental concepts and an outline of the most utilised models in MIMO. The advantages of conventional MIMO will then be presented, along with the benefits and challenges that come up when one tries to scale up these systems to larger dimensions. The chapter ends with a concise section on existing cutting edge massive MIMO prototypes and testbeds.

[3]Time division multiple access.
[4]Code division multiple access.
[5]Orthogonal frequency-division multiple access.

Chapter 3 - MIMO Detection: The outcomes of literature review on both conventional and state-of-the-art MIMO detection will figure in chapter 3. This includes the review of some fundamental concepts in MIMO, followed by a brief introduction to the most used linear and non-linear detection techniques, including lattice reduction and randomised algorithms.

Chapter 4 - On the Massive MIMO Effect The advantages of increasing the number of antennas from a theoretical point of view are going to be described in this chapter. In particular, the channel hardening effect is described and exploited to construct near optimal, low complexity algorithms to perform both detection and precoding.

Chapter 5 - Symmetric Large MIMO Detection In this chapter, two state-of-the-art approaches to optimally solve the detection problem for symmetric large MIMO systems are considered. Firstly, a suitable lattice reduction method is reviewed and compared against conventional receivers. Secondly, a randomised algorithm based on the Gibbs sampling technique is studied, optimised and set side by side with the former.

Chapter 6 - Application of Massive MIMO A massive MIMO enabled full-duplex relay is considered in chapter 6. Filters to perform interference suppression, detection and precoding are jointly designed. The advantages of such setup are evaluated in terms of spectral efficiency, bit error rates and achievable throughputs. Further, an optimal power scheme yielding low complexity to satisfy individual pair rates is proposed.

Chapter 7 - Conclusions In the last chapter of this thesis, an overview of the most important results is conducted. In addition, some suggestions for possible follow-up research are left.

1.3 Contributions

The original contributions of this research work are presented in the final sections of chapters 4, 5 and 6. They can be summarised as follows:

Fast Matrix Inversion Updates By exploiting the channel hardening effect, Neumann series can be efficiently used to approximate the inverse matrices required in linear detection techniques. To further reduce complexity, two algorithms based on the matrix inversion lemma for a fast recomputation of the inverse are proposed in chapter 4, and for the first time in the context of massive MIMO. In particular, the impact of the updates after the approximation in the performance is numerically evaluated. This part of the work resulted in the following submission:

- F. Rosário, F. Monteiro, and A. Rodrigues, "Fast Matrix Inversion Updates for Massive MIMO Detection and Precoding," submitted to a scientific journal, May 2015.

Optimised MCMC techniques based on Gibbs sampling Randomised algorithms, namely MCMC techniques, can be effectively used to find close to optimal solutions in a variety of problems.

In chapter 5, a thorough evaluation of these methods to perform MIMO detection using a Gibbs sampler is conducted. Specifically, the stalling problem is identified and techniques to circumvent it are reviewed. Further, it is concluded that convergence rates are highly dependent on the "temperature" parameter and, based on the obtained results, a variant that proves to be faster is suggested. This part of the work resulted in the following submission:

- F. Rosário, F. Monteiro, and A. Rodrigues, "Triple Mixed Lattice Gaussian Sampling Method for Large MIMO Detection," submitted to a scientific journal, June 2015.

Filter design for an end-to-end transmission supported by a full-duplex relay Full-duplex relays can in theory double spectral efficiencies over conventional wireless systems. However and without compromising neither direct or reverse links, loopback interference must be mitigated. In chapter 6, this problem is mathematically described and jointly designed filters to suppress interference and forward information are proposed. In addition, a low-complexity linear optimal power allocation scheme based on the individual throughput requirements is contemplated. This part of the work resulted in the following submission (accepted, presented, and published):

- J. S. Lemos, F. Rosário, F. Monteiro, J. Xavier, and A. Rodrigues, "Massive MIMO Full-Duplex Relaying with Optimal Power Allocation for Independent Multipairs," in *IEEE Proc. 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2015, pp. 306–310.

Chapter 2

MIMO Overview

In this chapter both fundamentals and theoretical advantages of the MIMO channel will be provided. In addition, an introduction to the benefits and arising issues when the number of antennas is increased will be provided.

2.1 Introduction

In wireless communications the traditional structure of a link is composed of a single antenna at both transmitter and receiver sides. This system, denoted as single-input single-output (SISO), was widely used in the past few decades for both mobile and fixed communications. However, due to the increasing required data rates and capacity for the same available bandwidth, approaches with increased spectral efficiencies were required. In 1996, Foschini demonstrated that this was possible from a practical point of view with the widely known transceiver architecture denoted as V-BLAST[1] [19]. From there on, a sheer amount of MIMO related contributions were made, addressing different aspects such as channel coding and estimation, capacity computation, multiple-access schemes and detection/precoding algorithms.

Today, MIMO technology is implemented in a variety of wireless applications, such as Wi-Fi (802.11ac), LTE-A and WiMAX[2], supporting much higher data rates than the conventional SISO. Yet, there are a set of constraints that are limiting conventional MIMO systems to be limited to eight antennas on both transmit and receiving sides. The idea for future wireless communication networks is the usage of a large to a massive number of antennas at the serving base station, which will, in theory, fulfil part of the requirements defined in section 1.1.1.

2.2 Channel Model

In a wireless link, signals are transmitted with the help of a carrier wave which has a defined centre frequency and average power. Even though the choice of the carrier affects the system performance, from

[1] Vertical-Bell Laboratories layered space-time.
[2] Worldwide interoperability for microwave access.

a signal processing perspective, one is only interested in the discrete samples of the complex baseband equivalent signal. For the characterisation of the wireless medium, an uplink scenario will be considered, where N_T single-antenna users transmit N_T complex symbols $\mathbf{x} \in \mathbb{C}^{N_T \times 1}$ from a given constellation. The transmitted signals will pass through the wireless medium, denoted as channel and defined by the matrix $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$, with entries h_{ij} corresponding to the gain between jth transmit antenna and ith receive element. Furthermore, the received sampled vector $\mathbf{y} \in \mathbb{C}^{N_R \times 1}$ will be affected by additive noise, denoted as $\mathbf{n} \in \mathbb{C}^{N_R \times 1}$. Hence, the received vector can be written as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}. \tag{2.1}$$

This is the channel that is mostly used in the literature and accepted as a good enough approximation to reality, as long as $\mathbf{H}$ and $\mathbf{n}$ statistics are modelled accordingly. Moreover, the following simplifications and assumptions are often made:

- The transmitted signal is narrowband enough so that the channel can be considered frequency-nonselective.
- The channel is modelled as slow-fading which means that $\mathbf{H}$ is constant for the duration of a coding block (tens to thousands of channel symbols).
- Antenna elements are spaced sufficiently far apart such that the entries of matrix can be modelled as independent and identically distributed (i.i.d.) Gaussian random variables with zero-mean and unit variance, that is $h_{ij} \sim \mathcal{CN}(0, 1)$ (this model is also denoted in the literature as Rayleigh fading).
- The additive noise is considered independent in each radiofrequency (RF) chain and modelled as circularly symmetric complex Gaussian vector for every frequency component, such that $\mathbb{E}\{\mathbf{n}\} = 0$ and $\mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2\mathbf{I}$. In the literature, a perturbation with these characteristics is often denoted as additive white Gaussian noise (AWGN).
- The total transmitted power is constrained to P, that is $tr\{\mathbf{Q}\} \leq P$ with $\mathbf{Q} = \mathbb{E}\{\mathbf{x}\mathbf{x}^H\}$. For independent input data the covariance comes as $\mathbf{Q} = \sigma_x^2\mathbf{I}$, where σ_x^2 represents the average symbol power.
- Perfect information of the channel matrix $\mathbf{H}$, denoted as channel state information (CSI), is often considered to be known at the receiver (CSI-R). Due to the inherent complexity in estimating CSI at the transmitter side (CSI-T), it is generally not considered to be available. Furthermore, perfect CSI is impossible due to the existence of noise, interference and hardware impairments. When imperfect CSI is considered, $\mathbf{H}$ entries can be decomposed in $h_{ij} = \hat{h}_{ij} + \tilde{h}_{ij}$, where $\hat{h}_{ij}$ is the estimated gain and $\tilde{h}_{ij}$ the associated error.

- Transmitted symbols $\mathbf{x}$ are often picked from an M-ary quadrature amplitude modulation (M-QAM). To guarantee symmetry in the constellation, M is chosen as an even power of two, that is $M = 2^{2n}$, with n being an integer.

These assumptions offer a very simplified characterisation of the channel, whose representation is presented in figure 2.1. Despite not entirely realistic, its main advantage is its ease of implementation; hence, it is widely used for performance comparisons in the literature (e.g., when comparing channel estimation or detection algorithms). What is more, the signal-to-noise ratio (SNR) per transmit antenna, denoted as SNR_a, in this model is straightforward to compute [20]:

$$\text{SNR}_a = \frac{\sigma_x^2}{\sigma_n^2}, \tag{2.2}$$

which is interpreted as the ratio between the average power of one transmitted symbol and the variance of the noise.

Obviously, other MIMO channel models trying to approximate reality with more precision exist. For instance, these models may consider scenarios where a line-of-sight (LOS) path is available, channel matrix entries are correlated or where the received signals resulted from a different number of scatterers [21].

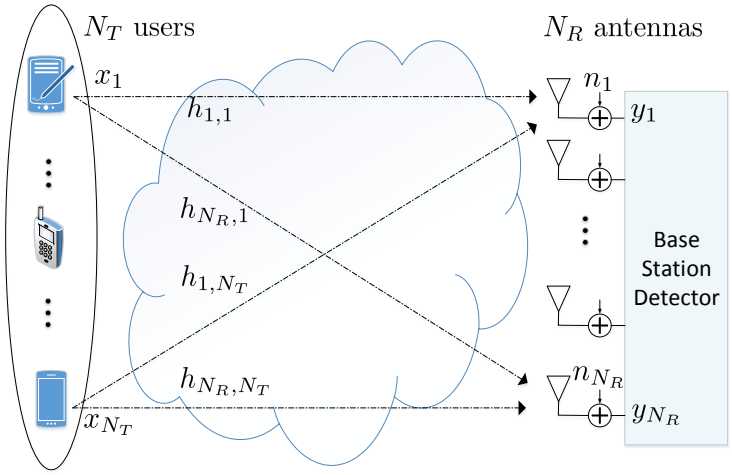


Figure 2.1: Multiuser MIMO uplink channel with N_T transmit and N_R receive antennas.

2.3 MIMO Channel Capacity

When the number of transmit and receive antennas is increased, one hopes to improve the achievable rates of the system, denoted as capacity and measured in bits per second per Hertz (bps/Hz). In order to do the necessary computations assume a single-user MIMO (SU-MIMO) scenario, where the channel statistics are as in the previous section (extension to multiuser (MU-MIMO) is similar). Furthermore, assume that the input vector $\mathbf{x}$ is i.i.d. complex Gaussian. From information theory results and under the previously stated assumptions, it is widely accepted that the expression for MIMO open-loop instantaneous capacity

comes as [22]:

$$C = \log_2 \det(\mathbf{I} + \text{SNR}_a \mathbf{H}\mathbf{H}^H). \tag{2.3}$$

Note that if CSI-T is available (closed-loop), optimal power distribution over transmit antennas methods, also known as waterfilling algorithms, can be performed, resulting in higher achievable capacities [23]. However, CSI-T is usually hard to obtain and, hence, it will not be considered in this work, unless otherwise stated. In order to gain insight into how capacity scales with the number of antennas, two distinct scenarios will be considered in the following subsections.

2.3.1 Capacity in the Low-SNR Regime

In the first analysed situation we consider a low-SNR regime at the receiver side with $\text{SNR}_a \to 0$. This can also be seen as the case where LOS propagation conditions impose that the transmit array cannot resolve individual receive elements (and vice-versa), resulting in a matrix $\mathbf{H}$ with low rank, which is denoted as r. From (2.3) we have that [21, Chapter 2]:

$$C = \log_2 \det(\mathbf{I} + \text{SNR}_a \mathbf{H}\mathbf{H}^H) = \log_2 \prod_{i=1}^{r} \left(1 + \text{SNR}_a \lambda_i^2(\mathbf{H})\right) \tag{2.4}$$

$$\approx \log_2 \left(1 + \sum_{i=1}^{r} \text{SNR}_a \lambda_i^2(\mathbf{H})\right) = \log_2 \left(1 + \text{SNR}_a tr\{\mathbf{H}\mathbf{H}^H\}\right) \approx \quad \text{SNR}_a tr\{\mathbf{H}\mathbf{H}^H\} \log_2 e, \tag{2.5}$$

where $\lambda_i(\mathbf{H})$ represents the ith eigenvalue of matrix $\mathbf{H}$. The dominance of linear terms and the fact that $\log_2(1+x) \approx x \log_2 e$ for small x were used in the first and second approximations, respectively. Computing now the ergodic capacity C_e, defined as the average behaviour over time of (2.3), one is led to:

$$C_e = \mathbb{E}\{C\} = \text{SNR}_a \mathbb{E}\{tr\{\mathbf{H}\mathbf{H}^H\}\} \log_2 e \tag{2.6}$$

$$= N_R N_T \text{SNR}_a \log_2 e, \tag{2.7}$$

where the fact that $\mathbb{E}\{tr\{\mathbf{H}\mathbf{H}^H\}\} = \mathbb{E}\{\sum_{i=1}^{N_R} \sum_{j=1}^{N_T} |h_{ij}|^2\} = N_R N_T$ was used. This equation shows that on average, and for low-SNR scenarios, capacity scales with N_R.

2.3.2 Capacity in the High-SNR Regime

When high-SNRs are available or analogously propagation conditions are favorable (similar eigenvalues $\lambda_i(\mathbf{H})$), equation 2.3 evens out to [21, Chapter 2]:

$$C = \mathbb{E}\left\{ \sum_{i=1}^{\min(N_T, N_R)} \log_2(1 + \text{SNR}_a \lambda_i^2(\mathbf{H})) \right\} \tag{2.8}$$

$$\approx \min(N_T, N_R) \log_2(\text{SNR}_a) + \sum_{i=1}^{\min(N_T, N_R)} \mathbb{E}\{\log_2 \lambda_i^2(\mathbf{H})\}, \tag{2.9}$$

where the approximation $\log_2(1+x) \approx \log_2(x)$ for large x was used. Under these conditions, one reaches the conclusion that capacity scales with $\min(N_T, N_R)$, which is a good result but that requires antennas at both transmitter and receiver sides. Despite the made simplifications, the results herein presented are promising and already exploited by many wireless communication systems. As will be seen in the following sections and chapters, the unbounded increase in the number of antennas brings new challenges that are still being identified and under intensive research.

2.4 Large MIMO Systems

Thus far it has been seen that with conventional MIMO we can, using the same amount of power as in the SISO counterpart, have a more reliable link and increased sum-rates. Today's systems employ a number of antennas between 2 and 8, achieving spectral efficiencies of less than about 30 bps/Hz [24]. In the future, when a large number of spatial dimensions (tens to hundreds of antennas) is utilised, both opportunities and challenges will arise. Some of these features will be reviewed in this section.

Large MIMO systems typical application scenarios will be regions with a dense deployment of terminals that need to be connected simultaneously. Possible examples of these are multibeam satellite and cellular networks beyond LTE-A, which may include coverage for both mobile terminals and local area networks (LAN) [25]. As seen in the previous section, more antennas means more degrees of freedom, which means small-scale fading, noise and hardware impairments can all be averaged out [26]. What was random tends to become more deterministic as the number of elements increases. Furthermore, under perfect CSI-R the total transmitted power by each user can decrease by a factor of N_R [27] ($\sqrt{N_R}$ in the case of imperfect CSI-R) and, as long as $N_R >> N_T$, linear receivers may achieve near-optimal performance whilst yielding low complexity [20, Chapter 10]. However, considerable technical challenges need to be addressed in realising such large MIMO systems. Some of these limiting factors are [28]:

Independent spatial dimensions To take full advantage of a large number of antennas, received signals at different elements should take different paths in the wireless medium. However, this effect denoted as rich scattering, does not always occur. Scenarios where an LOS is available or the pin-hole[3] effect is present will, therefore, degrade the capacity of the MIMO channel. Due to the diversity of physical environments (e.g., indoor, urban, rural) channel sounding measurements will be of extreme importance in order to characterise practical propagation environments. Some of the possible configurations for large MIMO systems can be seen in figure 2.2.

Antenna placement and RF chains The idea for large MIMO systems is to put small, inexpensive hardware in BSs and possibly terminals. This, in turn, will raise issues such as spatial correlation and reduced radiation efficiency due to the small inter-element spacing (a spacing of more than $\lambda/2$ is usually recommended). Moreover, tens to hundreds of RF chains, up/down converters,

[3]Pinhole channel is interpreted as the one where scattering energy travels through a very thin air pipe, preventing channel rank from building up [29].

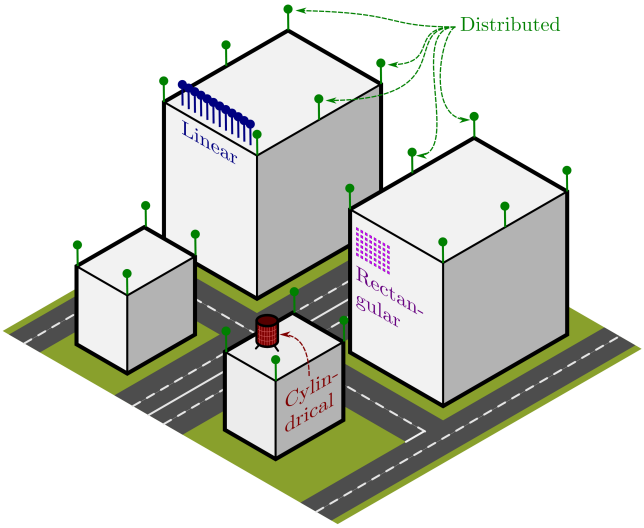


Figure 2.2: Possible antenna arrangements for base stations using large MIMO (reproduced from [30]).

A-D/D-A[4] converters will be needed to feed the antenna elements [30]. Low-cost hardware will result in more likely imperfections leading to higher phase-noise, I/Q[5] imbalance and levels of quantisation noise. Possible approaches to mitigate these issues are, for example, higher carrier frequencies (decreased wavelength) [28] and spatial modulation [31], which is able to maintain high spectral efficiencies while reducing the number of needed RF chains.

Signal Processing Complexity Even though mobile processing power has increased over the past few years, it will be important to keep algorithms' complexity low so that experienced delay is minimised. In a transmission system there are algorithms for channel coding, estimation, precoding, synchronisation, detection and waveform modulation. Hence, it is necessary to research new ways to achieve the objectives with implementable and efficient hardware.

Multicell operation and CSI-T Due to the expected aggressive reuse distance in 5G systems, it is necessary to manage resource allocation, intercell interference and BS cooperation ingeniously. For example, in a multicell MU scenario non-orthogonal pilots must be utilised [32], since orthogonal signalling would be infeasible for a large number of cells, leading to pilot contamination [33]. This will culminate into imperfect CSI, hindering the overall system capacity.

These are some of the challenges and they are interdependent in the sense that maximising one objective may result in the degradation of another (e.g., increasing spectral efficiency implies increasing power that for instance reduces energy efficiency). Consequently, subjective trade-offs between the objectives must come into play, since there is no global optimum but rather local optima [34]. What is more, different scenarios have different subjective optima that systems must be able to identify and adapt on-the-fly (e.g., in low-demand periods energy efficiency should be prioritised over spectral efficiency).

[4]Analog-to-digital and digital-to-analog.
[5]In-phase/quadrature.

2.5 Prototypes and Testbeds

In order to validate theory, researchers need to move from analytical work into the development of testbeds and prototypes. Only in this way, it is possible characterise real-world propagation conditions and determine the feasibility and commercial viability of large MIMO systems [35]. These prototypes often require a huge amount of investment in terms of money, time, resources and knowledge from a wide range of fields of expertise. For this reason, only top-tier, well funded research universities and some companies are able to develop their own frameworks, which must be versatile, flexible and scalable enough [35] to actually become cost-efficient in the long run.

One of the latest, cutting edge prototypes the author is aware of (announced October 2014) is the result of a partnership between Lund University (Sweden) and National Instruments (NI) [36]. The highly adaptable 100-antenna prototype incorporates simultaneously the award-winning LabVIEW software [37] and the state-of-the-art NI USRP RIO SDRs[6]. Even though performance results are not available as of yet, the system parameters include an RF center frequency that may oscillate between 1.2 and 6 GHz and an impressive sampling rate of 30.72 MS/s [35]. A picture of the system, including both array and USRPs, is presented in figure 2.3(a).

In July 2014, Ericsson (in association with NTT DOCOMO) made a live, over-the-air demonstration in the company labs located in Kista (Sweden) where they were able to achieve 5 Gbps in the 15 GHz frequency band. Despite the lack of further technical details, their report mentioned that the ”demonstration employs an innovative new radio interface concept and advanced MIMO technology to achieve the record-breaking results” [38].

On a smaller scale but equally important, channel measurement campaigns for different antenna configurations have been conducted in recent years and the results are very promising. Most of the studies state that performance in a large variety of scenarios is close to theoretical results and that orthogonality improves with increasing number of antennas [39–42]. Moreover, in addition to the planar array configuration, cylindrical and compact arrays (such as cubes) are also being experimented with coherent results [41, 43–45]. Some of these prototypes are depicted in figure 2.3.

2.6 Concluding Remarks

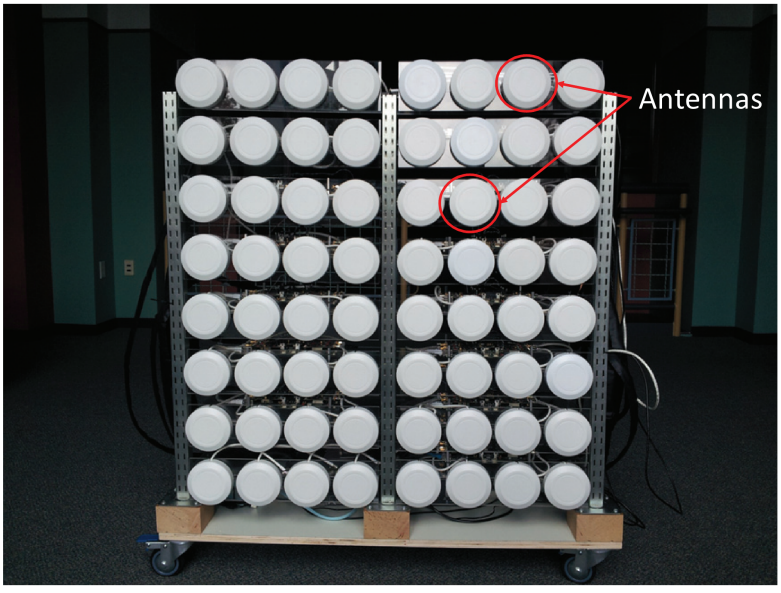
In this chapter, a brief introduction to the MIMO channel has been provided, and expressions for the underlying increase in capacity have been outlined. Further, both advantages and limiting obstacles yet to be faced in large MIMO systems have been presented. In the remaining chapters of this thesis, some of these problems will be detailed, whilst the gains provided by massive MIMO will be exploited in order to construct low complexity, near optimal methods.

[6]National Instruments universal software radio peripheral RIO software defined radio.

(a) Custom cross-polarised patch 100-antenna array developed in Lund University, in association with National Instruments (reproduced from [35]).

(b) 25-antenna array presented as part of METIS project optimised to function in the 2.3 GHz band (reproduced from [46]).

(c) 128-cyllindrical array used to measure propagation characteristics for very-large MIMO at 2.6 GHz (reproduced from [41]).



(d) Argos: a base station architecture employing 64 antennas (reproduced from [43]).

Figure 2.3: Massive MIMO array prototypes.

Chapter 3

MIMO Detection

In this chapter a brief overview of the fundamentals of MIMO detection will be first conducted. Secondly, an introduction to the various existing families of detectors will be provided, including both conventional and state-of-the-art receivers.

3.1 Introduction

A preponderant aspect of any communication link is the effectiveness of its detection algorithm. Provided some knowledge of the channel is available at the receiver side, one is to design low complexity algorithms to retrieve information from the transmitted symbols. A lower complexity means a reduced number of arithmetic operations, which leads to reduced latency and increased power efficiency. However, when reducing complexity, the designer might be compromising on the performance of the system (e.g., decreasing the diversity gain), which might be undesirable. Since optimal detection is prohibitive for large MIMO systems due to its exponential complexity, sub-optimal approaches must be considered, studied and compared. An overview of the different possible approaches will be conducted in this chapter.

Another motivation for the exhaustive research in detection algorithms is the fact that communication standards do not impose how network providers design their receivers. This fact opens up the opportunity for the implementation of adaptive equalisers, able to adjust themselves according to the required objectives at any given time. Thus, depending on the propagation conditions and loading factors, the use of hybrid detectors may become a very efficient alternative in the long run [47].

In addition to the complexity-performance trade-off, there are a set of other constraints in the design of detection algorithms. For instance, these methods should be hardware-friendly, scalable in terms of production and robust to finite-numerical resolution (fixed-point). All these aspects should, whenever possible, be taken into account, be incorporated in the simulation environment and carefully evaluated.

3.2 Detection Fundamentals

In the following analysis, the channel model described in section 2.2 will be used and, unless otherwise stated, perfect CSI-R will be considered. The purpose of a MIMO detector is to decode the N_T streams that were transmitted and mixed by the channel. Under the assumed scenario, performance will be very dependent on the conditioning of matrix $\mathbf{H}$ and noise average power σ_n^2. In this section the essential definitions to assess the behaviour of the various detectors will be presented.

One of the essential parameters to characterise the expected performance of a detector is the orthogonality deficiency of the channel matrix $\mathbf{H}$.

Definition 1 (Orthogonality Deficiency [47]). *For an $N_R \times N_T$ matrix $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_{N_T}]$, with $\mathbf{h}_n$ being the nth column of $\mathbf{H}$, its orthogonality deficiency ($od(\mathbf{H})$) is defined as*

$$od(\mathbf{H}) = 1 - \frac{\det(\mathbf{H}\mathbf{H}^H)}{\prod_{n=1}^{N_T} \|\mathbf{h}_n\|^2}, \tag{3.1}$$

where $\|\mathbf{h}_n\|$ is the norm of the nth column of $\mathbf{H}$.

The metric in (3.1) is a measure of the orthogonality of the column vectors in matrix $\mathbf{H}$, it is bounded by $0 \leq od(\mathbf{H}) \leq 1$ and takes minimum and maximum value when $\mathbf{H}$ is orthogonal and singular, respectively.

The parameter that measures how much the transmitted symbol power is above the additive noise at the receiver is the SNR.

Definition 2 (Signal-to-Noise Ratio). *For a transmitted vector $\mathbf{x} \in \mathbb{C}^{N_T \times 1}$ with covariance matrix $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \sigma_x^2 \mathbf{I}_{N_T}$ and an additive noise vector $\mathbf{n} \in \mathbb{C}^{N_R \times 1}$ distributed as $\mathcal{CN}(0, \sigma_n^2 \mathbf{I})$, the average input SNR is defined as:*

$$SNR = N_T \frac{\sigma_x^2}{\sigma_n^2}. \tag{3.2}$$

SNR is usually defined in logarithmic scale and widely used to study bit error rate (BER) performance in communication systems. For that purpose, the evolution characteristics of BER as a function of the SNR are often depicted and are used in the literature to evaluate the diversity order gain of the technique, denoted by G_d.

Definition 3 (Diversity Order Gain). *Designate the average error probability, which is function of the SNR, as $P_e(SNR)$. The diversity order gain G_d is then defined as [47]*

$$G_d = \lim_{SNR \to \infty} -\frac{\log P_e(SNR)}{\log SNR}. \tag{3.3}$$

Graphically, G_d is interpreted as the minus asymptotic shape of the BER *versus* SNR and the maximum value it can take is $G_{d,max} = N_T N_R$, in the case of uncorrelated fading [20, Chapter 2], as will be

verified in the next section. Another useful metric is the multiplexing gain, which represents the increase of the spectral efficiency with the SNR, for a given BER.

Definition 4 (Spatial Multiplexing Gain)**.** *Let $\{C(SNR)\}$ be a family of codes of block length l one at each SNR level, and $R(SNR)$ be the rate of the corresponding code. Then, the achievable spatial multiplexing gain G_m of the scheme is defined as [48]*

$$G_m = \lim_{SNR \to \infty} \frac{R(SNR)}{\log(SNR)}. \tag{3.4}$$

This parameter shows how capacity increases with SNR and is associated with the maximum number of independent subchannels (or layers), being upper bounded by $G_m \leq \min(N_T, N_R)$. For a fixed BER value, the multiplexing gain G_m measures how fast spectral efficiency can improve with the increase of the SNR while keeping the same error rate [20]. In the next section, further insight into the last two defined terms will be provided.

3.3 Diversity and Multiplexing - A Trade-off

Even though not entirely obvious, it can be shown that a multiple-antenna system has two interdependent gains [48]. For the sake of demonstration, consider the slow Rayleigh fading wireless link with N_T transmit and N_R receive antennas. If the same signal is transmitted in the N_T antennas, multiple independent faded replicas of the same symbol are received, resulting in a more reliable reception. This improvement in reliability is called diversity and for the considered scenario this advantage is expressed in the average BER which, in the high SNR regime, is made to decay with SNR^{-N_T} (better than the SISO counterpart SNR^{-1}).

Another possible approach is to send different data streams through different antennas and if individual transmit/receive pairs fade independently, the data rate can be increased by a given amount G_m. This technique is called spatial multiplexing and, for a fixed high SNR, capacity was shown to scale with $\min(N_T, N_R)$ (as verified in section 2.3.2).

Succinctly, MIMO systems provide two types of gains: diversity and spatial multiplexing. These are correlated in such a way that maximising one may not necessarily maximise the other. However, the trade-off between them and its interpretation can be expressed in a simple manner for most system parameters of interest. Consider the above scenarios and that the block length satisfies $l \geq N_T + N_R - 1$, for a multiplexing gain G_m the maximum achievable diversity gain is precisely $(N_T - G_m)(N_R - G_m)$. This result can be interpreted as G_m antennas provide multiplexing gain, whereas the remaining are used for diversity. Therefore, there is a fundamental trade-off between the two possible gains and that researchers have been trying to maximise for the past few decades. In this work and otherwise stated, no coding schemes will be employed and therefore, the study of the performance of the detectors will be solely based on their diversity gain.

3.4 Optimal Detection

As stated previously the objective of a receiver is, from the given data $\mathbf{y}$ and channel $\mathbf{H}$, to obtain an estimate of the transmitted message $\mathbf{x}$. Assuming $\mathbf{x}$ is chosen uniformly from a constellation (or alphabet) $\mathcal{A}$, the detector that minimises the error probability $P(\mathbf{x} \neq \hat{\mathbf{x}}|\mathbf{y}, \mathbf{H})$ is called maximum likelihood (ML) and is given by

$$\hat{\mathbf{x}}_{ML} = \arg \min_{\hat{\mathbf{x}} \in \mathcal{A}^{N_T}} \|\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\|, \tag{3.5}$$

where $\hat{\mathbf{x}}$ is the estimated vector. Geometrically, ML chooses the vector $\hat{\mathbf{x}}$ that yields the smallest Euclidian distance between received vector and conjectured message $\mathbf{H}\hat{\mathbf{x}}$. This is equivalent to a discrete optimisation problem over $|\mathcal{A}|^{N_T}$ candidate vectors, hence resulting in an exponential complexity with the number of transmit antennas [49, 50] (NP-hard problem). Even though ML attains optimal performance (maximum diversity), its complexity is impractical for larger dimensions and constellations. Therefore, suboptimal approaches will be introduced and briefly compared in the next sections.

3.5 Linear Detection

In order to separate the various N_T substreams, linear equalisation schemes can be employed. This is achieved by multiplying the received vector $\mathbf{y}$ with the equalisation matrix $\mathbf{A}^H$ followed by a slicer[1] $\mathcal{Q}(\cdot)$ in the following way:

$$\hat{\mathbf{x}} = \mathcal{Q}(\mathbf{A}^H\mathbf{y}). \tag{3.6}$$

This family of detectors have the advantage of being easy to implement and yielding low complexity at the expense of not collecting the same diversity as ML. The most conventional linear detectors (LD) will be presented in the next subsections.

3.5.1 Matched Filtering

The simplest form of receiver is the matched filter (MF) as it treats interference from other substreams as purely noise by making $\mathbf{A} = \mathbf{H}$. The MF estimates are, hence, given by (apart from a normalisation factor):

$$\hat{\mathbf{x}}_{MF} = \mathcal{Q}(\mathbf{H}^H\mathbf{y}). \tag{3.7}$$

Computing the above expression yields $\mathcal{O}(N_T N_R)$ operations, which is rather attractive. However, it only works properly for scenarios where $N_T << N_R$, as performance is severely degraded by the contributions of uncanceled interference when load increases.

[1] The slicer returns the nearest neighbour in the constellation being used.

3.5.2 Zero-Forcing Detection

Another LD family is the zero-forcing (ZF) methods, which impose inter-user interference to zero at the expense of an increase in noise (denoted as noise enhancement). This is achieved by designing the equalisation matrix as [51]:

$$\mathbf{A}_{ZF}^{H} = (\mathbf{H}^{H}\mathbf{H})^{-1}\mathbf{H}^{H} = \mathbf{H}^{\dagger}, \tag{3.8}$$

where $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudoinverse of a matrix [52]. After equalisation, the noise level is changed, resulting in an output SNR for the nth symbol given by [53]:

$$\text{SNR}_{ZF,n} = \frac{\text{SNR}}{[(\mathbf{H}^{H}\mathbf{H})^{-1}]_{nn}}, \tag{3.9}$$

where $[\cdot]_{nn}$ denotes the nth diagonal element. Generally, ZF performs better than the MF, is only optimal when $od(\mathbf{H}) = 0$ and it requires $\mathcal{O}(N_T^3)+\mathcal{O}(N_T^2 N_R)+\mathcal{O}(N_T N_R^2)$ multiplications and the same amount of additions [47] to equalise the received vector. In terms of diversity gain it is shown that it can achieve $N_R - N_T + 1$, which is still $N_T - 1$ times less than what ML can attain [54].

3.5.3 Minimum Mean-Square Error Detection

The linear detector whose transformation matrix objective is to minimise the mean-square error (MSE) between the transmitted and transformed received vector is called minimum mean-square error (MMSE). It can be formulated as:

$$\mathbf{A}_{MMSE}^{H} = \arg\min_{\mathbf{U}} \mathbb{E}\{\|\mathbf{x} - \mathbf{U}\mathbf{H}^{H}\mathbf{y}\|^2\}, \tag{3.10}$$

with closed form solution given by [55]:

$$\mathbf{A}_{MMSE}^{H} = (\mathbf{H}^{H}\mathbf{H} + N_T \text{SNR}^{-1}\mathbf{I})^{-1}\mathbf{H}^{H}. \tag{3.11}$$

As can be inferred from (3.11), this detector requires knowledge of the SNR but will perform better than ZF in the medium to high-SNR regimes. This is due to the reduced noise enhancement given by [53]:

$$\text{SNR}_{MMSE,n} = \frac{\text{SNR}}{[\mathbf{H}^{H}\mathbf{H} + N_T \text{SNR}^{-1}\mathbf{I}]_{nn}} - 1 = \text{SNR}_{ZF,n} + \eta_{\text{SNR},n}, \tag{3.12}$$

where $\eta_{\text{SNR},n}$ is a non-vanishing random variable, statistically independent of $\text{SNR}_{ZF,n}$ and that converges with probability one to a scaled Fisher–Snedecor distribution [56, Chapter 20](for more information on the distribution properties, see [53]).

In spite of the non-vanishing SNR gap between ZF and MMSE filters, ultimately, when SNR $\to \infty$

[53], the average ϵ-outage capacities[2] converge to the same value [57], that is:

$$\lim_{\text{SNR}\to\infty} C_{MMSE}(\epsilon) - C_{ZF}(\epsilon) \to 0. \tag{3.13}$$

Furthermore, MMSE will not improve on either the diversity gain (which is still $N_R - N_T + 1$) or on the complexity and will, hence, suffer the same problems as ZF in larger dimensions.

3.6 Non-linear Detection

As seen in the previous section, LDs are not able to collect diversity in symmetric large-scale systems ($N_T = N_R$) due to the loss of mutual information. As a consequence, non-linear methods to cancel interference across different layers have been widely studied in the literature. Popular techniques include successive and parallel interference cancellation (SIC and PIC, respectively). Henceforth, due to its simplicity, only SIC detectors will be considered.

3.6.1 Successive Interference Cancellation

The purpose of the SIC is to successively decode the different layers (using either ZF or MMSE) and cancel the effect of the estimated symbol from the received signal, assuming it was correctly detected. In order to further improve their performance, one possible approach is to always decode the symbol with the highest SNR, resulting in a need to find the optimal detection order. This method is often called Ordered-SIC (OSIC) but in the literature it is mostly known as V-BLAST[3][19].

Performance-wise, the diversity gain attained by the OSIC in the ith layer is now $N_R - N_T + i$ [59], leading to an improved diversity gain in the low to moderate SNR. However, in the high SNR regime the outage probability is dominated by the first detected layer and diversity is still $N_R - N_T + 1$. Nonetheless, there is still a considerable gain in the SNR, since at each iteration the detection is subject to one less interferer [53] (assuming previous estimations were correct). The maximum expected gain in the SNR when compared to the conventional linear detectors is shown to be around $10\log_{10} N_R$ dB [53].

3.6.2 Lattice Reduction-Aided Algorithms

So far and despite their low complexity, none of the studied detectors (apart from ML) was able to achieve full diversity. This is often in part due to the bad conditioning of the channel matrix $\mathbf{H}$ (high $od(\mathbf{H})$). The aim of lattice reduction-aided (LRA) methods is precisely to transform $\mathbf{H}$ into a new, more orthogonal basis $\tilde{\mathbf{H}} = \mathbf{HT}$, with $\mathbf{T}$ being a unimodular matrix [60]. Hence and interpreting the columns of $\mathbf{H}$ as the basis of a lattice, the purpose of the LR algorithm is to find the vectors of shortest length, which will

[2]Here, the ϵ-outage capacity is defined as the maximum supportable rate under the restriction that the outage probability is no greater than ϵ.

[3]This architecture was available for the first time in 1996 at Bell Laboratories in New Jersey in the United States, achieving spectral efficiencies ranging from 20–40 bit/s/Hz [58].

result in a lower orthogonality deficiency and as a consequence better performance when linear detection is applied.

Before proceeding and because it will be important in the subsequent chapters, some real-lattice theory shall now be presented.

Definition 5 (Primal Lattice)**.** *Let* $\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_{N_T}$ *be a linearly independent set of vectors in* $\mathbb{R}^{N_R}$ *(with* $N_T \leq N_R$*) [61]. The set of points*

$$\Lambda = \left\{ \mathbf{y} \in \mathbb{R}^{N_R} : \mathbf{y} = \sum_{i=1}^{N_T} \mathbf{h}_i x_i, x_i \in \mathbb{Z} \right\} \tag{3.14}$$

is called a lattice of dimension N_R *and whose vectors* $\mathbf{h}_1, ..., \mathbf{h}_{N_T}$ *are its basis.*

Basically, a set of discrete points in $\mathbb{R}^{N_R}$ form a lattice (also known as the primal lattice). Geometrically, it can be interpreted as an integral linear combination of the vectors $\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_{N_T}$. Given a primal basis, its dual can also be defined [20, Chapter 2].

Definition 6 (Dual Lattice)**.** *Given a primal lattice* Λ *with a basis* $\mathbf{H}$*, the dual lattice is unique and defined as*

$$\Lambda_D = \left\{ \mathbf{z} \in \mathbb{R}^{N_T} : \langle \mathbf{z}, \mathbf{x} \rangle \in \mathbb{Z}, \forall \mathbf{x} \in \Lambda \right\}. \tag{3.15}$$

In words, the dual of Λ is the set of all points whose inner product with any of the points in Λ is integer [62].

Known LR algorithms include Minkowski reduction, Hermite–Korkin-Zolotarev (HKZ) reduction [63], Lenstra-Lenstra-Lovász (LLL) and its complex counterparts [64], Seysen's algorithm (SA) [65] and element-based lattice reduction (ELR) [20, Chapter 10]. In this section only the algorithms that have a relevant contribution to MIMO detection (at the present moment) will be considered.

By far, the most used method in the literature is the LLL, mainly due to its polynomial complexity $\mathcal{O}(N_R^4)$ and for the fact that its $od(\tilde{\mathbf{H}})$ is bounded [20, Chapter 10]. Additionally, the SIC LR-based detection with LLL can attain full receive diversity order N_R with average complexity $O(N_T^3 N_R \log N_R)$ [66], which is still feasible for medium sized systems. An important improvement to the LLL was the complex-LLL (CLLL) which enabled the use of the method directly in the complex domain and, further, it is proven to only require half of the arithmetic computations needed by the real counterpart [67]. However, LLL based algorithms may not perform well for larger N_R values due to their increasing asymptotic error [68].

Another important family of LR-based methods are the very recently proposed ELR algorithms. The underlying idea is to generate an $\tilde{\mathbf{H}}$ to reduce the asymptotic pairwise error probability of conventional detectors. In MIMO detection, this translates to minimising the diagonal elements of the noise covariance matrix after lattice-reduced is applied [69]. The ELR approach is proven to yield low-complexity while attaining solid performance even for larger systems and constellation sizes [68] and will be evaluated in chapter 5.

3.6.3 Sphere Decoder

In the last sections near-optimal algorithms yielding acceptable complexity were discussed. An optimal (ML-like) detector shall now be introduced.

A MIMO detection problem is equivalent to finding the closest lattice point in the corresponding constellation. The concept of the sphere decoder (SD) is to search only through the points that are found inside a sphere with given radius $\sqrt{R}$ [61]. In this way, only the lattice points within the squared distance R are considered in the metric optimisation and as long as R is properly chosen, it is possible to achieve ML performance with lower average complexity.

SD can be equivalently seen as a search in a tree, where nodes expected values are dependent of the channel $\mathbf{H}$, noise variance and estimated transmitted message. Therefore, a large number of branches might have to be visited before reaching the solution, resulting in a non-fixed complexity. Even though for many practical problems the complexity may be upper-bounded by a polynomial function, there will be situations (channel realisations) where SD complexity becomes exponential [70], making it useless for real-time detection in large MIMO systems. Nevertheless, their performance can be used as a standard for comparison.

3.7 Detection Based on Local Search

MIMO detection involves a minimisation of the ML cost as in (3.5) and, in larger dimensions, that computation is unbearable. A possible approach, also utilised in other fields [71], is the local search that makes use of heuristic techniques. Even though there is no initial guarantee on the quality of the solution, these methods have demonstrated robust performance in a reasonable amount of time and will therefore be discussed in this section.

The idea behind detection based on local search is, given an initial point, search continuously in its neighbourhood for better solutions. This can be seen as reducing the space of search to a smaller one and find the point with minimum cost function in that subset (local optimum). The neighbourhood of a given point consists of all solutions which differ from the given solution in k coordinates (also referred as the k-coordinate away neighbourhood). For the sake of looking into different regions of space, the search procedure can be repeated several times by performing escape strategies and choosing different initial points. Attached to these random methods there must be stopping criteria (to limit complexity), that when satisfied, the point yielding the highest likelihood in the explored space is returned as the solution.

Although it is hard to analytically evaluate the performance of these algorithms, they are found to converge quickly with an output that is close to the optimal solution and, hence, they continue to be studied and compared by many researchers [28]. In the next paragraphs, some of the local search based algorithms used for detection in large MIMO systems will be presented. Henceforth in this chapter, an equivalent real-valued MIMO system will be considered[4].

[4] A complex-valued model can be transformed in a completely equivalent real-valued one (as will be seen in chapter 5).

3.7.1 Likelihood Ascent Search

In the K-stage likelihood ascent search algorithm (K-LAS) [72], the initial point is preferably given by the output solution of a low complexity detector, such as MMSE. The search consists of several substages. During the first substage, which may have more than one iteration, the algorithm updates one symbol per iteration (1-coordinate away neighbourhood) such that the cost function at each iteration decreases monotonically. It stops when a local optimum is reached, moving then to the second substage. At this point a two symbol update iteration is applied (2-coordinate away neighbourhood). If it succeeds in increasing the likelihood, the algorithm returns to the one symbol update stage. Otherwise, it moves on to a three symbol update and so on until a K-coordinate away neighbourhood fails to increase the likelihood. In order to limit complexity, symbol updates at a substage higher than the first one may only have one iteration. The cost function after the kth iteration at a given search stage is given by [28]:

$$C^{(k)} = \mathbf{x}^{(k)^T}\mathbf{H}^T\mathbf{H}\mathbf{x}^{(k)} - 2\mathbf{y}^T\mathbf{H}\mathbf{x}^{(k)}, \tag{3.16}$$

where $\mathbf{x}^{(k)}$ is the candidate vector at the kth iteration. The overall average per symbol complexity is $\mathcal{O}(N_T^2)$ which is acceptable for larger dimensions. The main disadvantage of the conventional LAS algorithm is that to achieve optimal BER performance, a very large number of receive antennas is needed, especially for larger constellations. In order to overcome this situation, alternative LAS approaches using multiple initial vectors and multiple search candidate sets have been proposed in [73].

3.7.2 Randomised Search

In the randomised search (RS) algorithm as in [74], the set of nodes to be tested in a local neighbourhood is selected randomly, hence the name. In order to improve search efficiency, the method keeps track of already visited nodes. Given an initial point, the first step in the algorithm is to find its neighbourhood, represented by all feasible vectors which are one symbol away from $\mathbf{x}^{(k)}$ and that have not been visited yet. Then, a randomly chosen position, denoted by m, is picked and all possible vectors differing from $\mathbf{x}^{(t)}$ in that index m, denoted by $\mathbf{d}(j)$, are computed as:

$$\mathbf{d}(j) = \mathbf{x}^{(t)} + \lambda_m\mathbf{e}_m, j = 1, ..., |\mathcal{A}| - 1, \tag{3.17}$$

where $\mathbf{e}_m$ is a vector of zeros except for the mth entry which is 1 and λ_m is a constant such that $\mathbf{d}(j) \in \mathcal{A}$, for all j. Consequently, the cost function $\beta(j)$ is computed as [28]:

$$\beta(j) = 2\lambda_m z_m - 2\lambda_m\mathbf{e}_m^T\mathbf{G}\mathbf{x}^{(\mathbf{t})} - \lambda_m^2\mathbf{G}_{m,m}, \tag{3.18}$$

with $\mathbf{G} = \mathbf{H}^T\mathbf{H}$, $\mathbf{z} = \mathbf{H}^T\mathbf{y}$ and where z_m denotes the mth element of $\mathbf{z}$. Then, the candidate vector with the highest likelihood β_{max} is either considered the new solution vector (if $\beta_{max} > 0$) or discarded (if $\beta_{max} < 0$). The algorithm stops only when all m positions have been visited. Due to its low

average complexity $\mathcal{O}(N_T^{1.4})$ and in the pursuance of better solution vectors (explore other parts of the space), the algorithm can be restarted multiple times and the final solution will be given by the vector $\mathbf{x}$ yielding the lowest ML cost. Once again and similar to LAS, RS algorithms attains close to optimal performance in large dimensions but only for small constellations [28]. Nevertheless, its performance is highly dependent on the number of required restarts, which is also a function of the SNR, number of antennas and constellation size.

3.8 Concluding Remarks

In this chapter, two distinct families of receivers, suitable for large MIMO systems, have been outlined. They are the LRA methods and the techniques based on local search. On the one hand, LR methods combine the traditional LD and SIC techniques in order to retrieve maximum diversity. On the other hand, randomised algorithms return the best solution after performing an "ingenious" exploration of space regions where the optimal solution is very likely to be. Descriptions of both theory and implementation aspects of the corresponding state-of-the-art algorithms will be detailed in chapter 5, and where an improved variant is also proposed.

Chapter 4

On the Massive MIMO Effect

In this chapter, the main benefits of increasing the number of dimensions will be outlined and exploited to construct algorithms that, without compromising the performance, yield lower complexity than conventional methods.

4.1 Introduction

In chapter 2, when the advantages of increasing the number of antennas in terms of capacity were discussed, a point-to-point system was considered. However, in this work one is mainly interested in the case where different N_T users are served simultaneously, since in that scenario the chances of uncorrelated fading are much higher. This leads to a channel matrix $\mathbf{H}$ with higher rank, which is much more desirable. In this chapter, the multiuser case will then be considered, and the consequences of increasing the number of receiving antennas without bounds will be discussed.

Whilst assuming $N_T << N_R$, it will be indeed verified that linear processing techniques perform close to optimal and, capitalising on that fact, algorithms for both fast computation and updates of inverse matrices will be evaluated.

4.2 Channel Hardening

In a communication system with a given channel and available bandwidth, one is interested in maximising the amount of transmitted information, measured in bits per channel use, and which is denoted as mutual information. Using the MIMO model in (2.1), the SNR definition in (3.2), and considering the input $\mathbf{x}$ to be Gaussian distributed [75], instantaneous capacity can be expressed as

$$C = \log_2 \det \left(\mathbf{I} + \frac{\text{SNR}}{N_T} \mathbf{H}^H \mathbf{H} \right). \tag{4.1}$$

However, because $\mathbf{H}$ is a random matrix, one is interested in the average behaviour of (4.1), that is $C_e = \mathbb{E}\{C\}$, denoted as ergodic capacity, and which can be formulated as [76]:

$$C_e = N_R \mathbb{E}\{\log_2(1 + \text{SNR}\lambda)\}, \tag{4.2}$$

where the operator $\mathbb{E}\{\cdot\}$ is over the eigenvalues of the normalised Wishart matrix $\mathbf{H}^H\mathbf{H}/N_T$ [77], denoted here by λ.

One of the solutions to increase throughputs, as was reviewed in previous chapters, is to use multiple antenna communications. The gain in capacity, from a theoretical point of view, when the number of antennas is increased without bounds will be briefly evaluated in this section. Particularly, two situations will be considered: $N_R >> N_T$ and $N_T \approx N_R$.

For the first case, using the law of large numbers, analysis is straightforward. Note that the entries of $\mathbf{H}$ are Gaussian i.i.d. random variables, hence when $\mathbf{H}$ is tall ($N_R >> N_T$), the product $\mathbf{H}^H\mathbf{H}/N_T = \mathbf{H}^H\mathbf{H}/N_R \cdot (N_R/N_T)$ tends to a scaled identity matrix, that is $(N_R/N_T)\mathbf{I}$. Therefore, when $N_R \to \infty$ the eigenvalues in (4.2) tend deterministically to (N_R/N_T), resulting in:

$$C_e^{N_R>>N_T} \sim N_R \log_2\left(1 + \text{SNR}\frac{N_R}{N_T}\right), \tag{4.3}$$

which clearly shows the advantages of increasing the number of receive antennas N_R.

On the other hand, when both number of served users (N_T) and number of BS antennas (N_R) are large, with ratio $\beta = N_T/N_R$, the eigenvalues in (4.2) do not converge deterministically any more, but rather statistically. Further, their distribution is known and given by the Marchenko-Pastur law, as provided in the following theorem [78].

Theorem 1. *Consider a matrix $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ with i.i.d. entries satisfying $\mathcal{CN}(0,1)$. Then, when $N_R, N_T \to \infty$ with $N_T/N_R \to \beta$, then the distribution $p(\lambda)$ of the eigenvalues $\lambda(\mathbf{H}^H\mathbf{H}/N_R))$ satisfies*

$$p(x) = (1 - \beta^{-1})^+\delta(x) + \frac{1}{2\pi\beta x}\sqrt{(x-a)^+(x-b)^+}, \tag{4.4}$$

where $a = (1-\sqrt{\beta})^2$, $b = (1+\sqrt{\beta})^2$, $(z)^+ = \max(0,z)$ and $\delta(x)$ represents the Dirac delta function.

A graphical representation of the eigenvalue distribution in (4.4) is depicted in figure 4.1 for different values of β (the mass function at $x = 0$ was disregarded). What can be inferred is that when the ratio β decreases, a convergence in the distribution is verified, meaning that the eigenvalue of a given channel realisation becomes more and more deterministic as the number of antennas increases.

This diagonalisation effect verified in $\mathbf{H}^H\mathbf{H}$ is called in the literature channel orthogonalisation or channel hardening, and will be exploited in the subsequent sections to construct low complexity near optimal detectors and precoders.

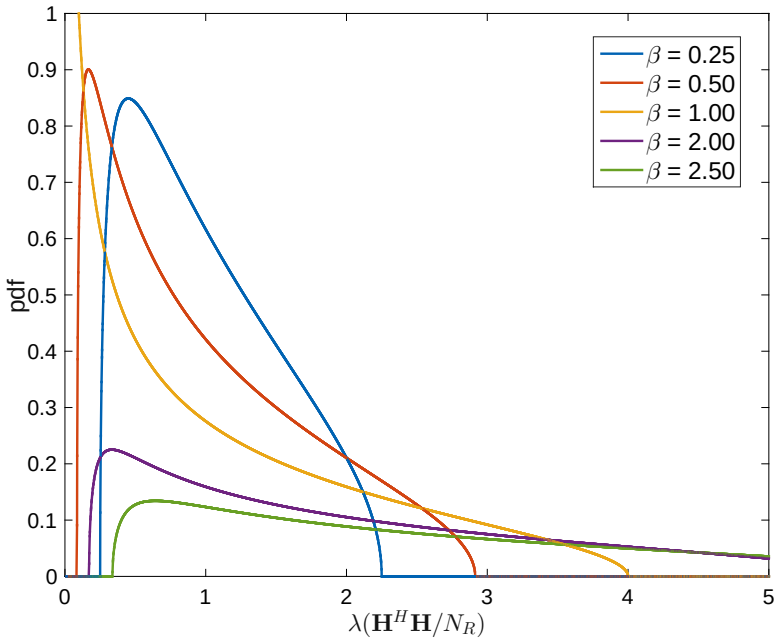


Figure 4.1: Probability density function (pdf) of the eigenvalues of $\mathbf{H}^H\mathbf{H}/N_R$ for different ratios β.

4.3 On the Optimality of Linear Processing

The diagonalisation of the channel, verified in the previous section when $1 << N_T << N_R$, can be effectively exploited to apply linear processing techniques (for either detection or precoding purposes). Specifically, these low complexity yielding filters attain near optimal performance (diversity is $N_R - N_T + 1 \approx N_R$), hence their interest in massive MIMO systems. In this work, only the detection case will be considered, though its extension to precoding is straightforward.

Denote the general, linear detector matrix by $\mathbf{A} \in \mathbb{C}^{N_R \times N_T}$. Then, by multiplying the received signal by its conjugate transpose $\mathbf{A}^H$, followed by the application of the quantiser operator $\mathcal{Q}(\cdot)$, one can estimate the transmitted symbols as:

$$\hat{\mathbf{x}} = \mathcal{Q}(\mathbf{A}^H\mathbf{y}), \tag{4.5}$$

where $\hat{\mathbf{x}}$ are the hard decisions of $\mathbf{x}$. The three conventional detectors MF, ZF and MMSE, introduced in chapter 3 will be considered, i.e.,

$$\mathbf{A} = \begin{cases} \mathbf{H}, & \text{for MF;} \\ \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}, & \text{for ZF;} \\ \mathbf{H}(\mathbf{H}^H\mathbf{H} + N_T\text{SNR}^{-1}\mathbf{I})^{-1}, & \text{for MMSE.} \end{cases} \tag{4.6}$$

Applying any of these filters to the received vector in (2.1) leads to

$$\mathbf{r} = \mathbf{A}^H \mathbf{y} = \mathbf{A}^H \mathbf{H}\mathbf{x} + \mathbf{A}^H \mathbf{n}, \tag{4.7}$$

where $\mathbf{r} \in \mathbb{C}^{N_T \times 1}$ denotes the vector containing the symbols before quantisation. Its kth element can be decomposed as follows

$$r_k = \underbrace{\mathbf{a}_k^H \mathbf{h}_k x_k}_{\text{useful signal}} + \underbrace{\sum_{i \neq k}^{N_T} \mathbf{a}_k^H \mathbf{h}_i x_i}_{\text{interference}} + \underbrace{\mathbf{a}_k^H \mathbf{n}}_{\text{enhanced noise}} . \tag{4.8}$$

Now using the MF in (4.6), taking into account the statistical distributions of $\mathbf{h}_k \sim \mathcal{CN}(0,1)$ and $\mathbf{n} \sim \mathcal{CN}(0, \sigma_n^2)$, and using the law of large numbers (as $N_R \to \infty$) [32]:

$$\frac{1}{N_R} \mathbf{h}_k^H \mathbf{h}_i \xrightarrow{\text{a.s.}} 0 \text{ (interference vanishes);} \tag{4.9}$$

$$\frac{1}{N_R} \mathbf{h}_k \mathbf{n} \xrightarrow{\text{a.s.}} 0 \text{ (noise vanishes),} \tag{4.10}$$

for $k \neq i$, $k = 1, ..., N_T$ and where $\xrightarrow{\text{a.s.}}$ denotes the almost sure convergence. Therefore, from (4.8), the signal of interest satisfies

$$\frac{1}{N_R} \mathbf{h}_k^H \mathbf{r} \xrightarrow{\text{a.s.}} \mathbb{E}\{h_{kk}^\star h_{kk}\} x_k = x_k, \tag{4.11}$$

which is the same as stating that the simplest form of detection (MF) performs optimally for any SNR, as long as $N_R \to \infty$. Obviously, this last condition is rather unrealistic, nevertheless (4.11) gives a useful insight into the benefits of scaling up the number of antennas at the receiving end.

Unlike the MF, the ZF detector considers other users' channels and forces inter-user interference to zero. This mitigation, however, comes at the cost of noise enhancement, which in the low SNR regime might lead to a drastic reduction in performance. Further, computing the ZF filter requires an inverse, which may yield a non-negligible complexity, when a large number of users are being served (a more rigorous study on this will be conducted in subsequent sections). Thus, when inter-user interference is considerable (large N_T), better performance for the MF is to be expected for the low SNR, whilst requiring a lower number of arithmetic operations. On the other hand, the MMSE filter takes into consideration both inter-user interference and noise variance in order to perform detection. Assuming that such information is available at the receiver (knowledge of both SNR and CSI), then, from the three considered detectors, MMSE is expected to perform best.

For a numerical demonstration of these facts, a simulation with a realistic number of receiving antennas, $N_R = 128$, is performed [41]. Using the detectors in (4.6) and for an increasing number of transmit antennas ($N_T = 8, 64, 128$), BER performance is depicted in figure 4.2. First thing to be noticed, is the high diversity attained by both ZF and MMSE filters when $N_T = 8 << N_R$, corroborating the statement previously made. Further and as expected, the MF filter outperformed the ZF in the low SNR for the symmetric case ($N_R = N_T$), thereby in this regime the best strategy is to treat interference also

as noise. Nevertheless, linear detection performance in this last, symmetric case is subpar, hence the need for large MIMO suitable detectors (as will be studied in chapter 5). Still, the MMSE filter, yielding about the same complexity as ZF, obtained the best performance of the three (for this reason, when linear detection-dependent methods are applied, such as lattice reduction, MMSE filters should be chosen to guarantee maximum effectiveness).

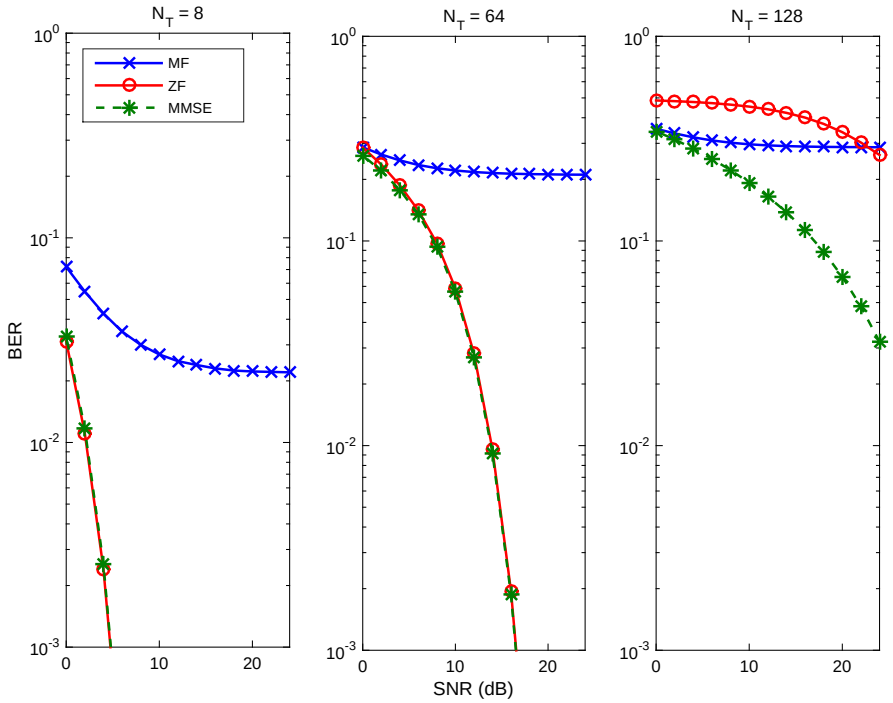


Figure 4.2: BER performance of traditional linear detectors when $N_R = 128$ and for different values of N_T.

4.4 Efficient Inversion Techniques

As was verified in previous sections, when the number of receive antennas N_R is much larger than the number of served users N_T in the uplink, channel hardening effects become preponderant. This is more noticeable especially when the ratio β is low. Mathematically, this can be seen as a diagonalisation of the entries in the Gram matrix $\mathbf{Z} = \mathbf{H}^H\mathbf{H}$, where the off-diagonal entries tend to zero, whereas the diagonal terms become closer and closer to N_R.

Under the assumption that $N_R >> N_T$, linear detection and precoding methods were confirmed to perform close to optimal. Even though these techniques yield low complexity, they often require the inversion of matrix $\mathbf{Z} \in \mathbb{C}^{N_T \times N_T}$, an operation that is not very hardware-friendly when conventional methods are utilised. For convenience purposes, the ZF and MMSE detection matrices are replicated in

this section and denoted by $\mathbf{A}_{zf}$ and $\mathbf{A}_{mmse}$. The corresponding expressions are as follows:

$$\mathbf{A}_{zf} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H = \mathbf{Z}_{zf}^{-1}\mathbf{H}^H; \tag{4.12}$$

$$\mathbf{A}_{mmse} = (\mathbf{H}^H\mathbf{H} + N_T\text{SNR}^{-1}\mathbf{I})^{-1}\mathbf{H}^H = \mathbf{Z}_{mmse}^{-1}\mathbf{H}^H. \tag{4.13}$$

Both $\mathbf{Z}_{zf}$ and $\mathbf{Z}_{mmse}$ are, for large $\gamma = N_R/N_T$, an "almost diagonal" matrix $\mathbf{D}$. For simplification purposes and without loss of generality, the ZF Gram matrix $\mathbf{Z} = \mathbf{Z}_{zf}$ will be considered in the analysis henceforth.

The computation of $\mathbf{Z}^{-1}$ is performed in the central processing unit. For large N_T, a high throughput matrix inversion can become expensive both in terms of area and power [79]. Therefore, exploiting the fact that $\mathbf{Z} \approx \mathbf{D}$, one may look for low complexity suboptimal approaches to find good approximations of $\mathbf{Z}^{-1}$. A first brute-force approach would be to approximate $\mathbf{Z}^{-1}$ by $\mathbf{D}^{-1}$, which would lead to a considerable decrease in complexity. But is this in fact a good estimate of $\mathbf{Z}^{-1}$? Unfortunately, and as will be confirmed in the following section, this is a crude approximation, severely deteriorating the performance of the system [80]. Hence, the objective is to find a hardware-realisable method that finds $\mathbf{Z}^{-1}$ with low complexity and without compromising the performance of the system. From the previous sentence, two pertinent questions arise:

1. What is the complexity of conventional stable inversion methods?
2. How acceptable must the performance of the proposed method be?

In order to answer the last question, the immediate reply is that it depends on the considered application and namely on the error correcting capabilities of the channel coding scheme being employed. For instance, LTE has a maximum 10% block error rate requirement [81]. An answer to question 1 will be given in section 4.4.3.

In this work, the study will be focused on the trade-off between performance and complexity. As a starting point Neumann series expansion will be considered, followed by a brief detail of the complexity of an exact inverse method based on the Cholesky decomposition.

4.4.1 Neumann Series Expansion

Eigenvalue Analysis

Using the fact that the eigenvalues of $\mathbf{Z}$ follow a deterministic Marchenko-Pastur distribution (as seen in theorem 1) and the analysis conducted in [80], one may infer that the largest and smallest eigenvalues of the scaled matrix $\frac{1}{N_T+N_R}\mathbf{Z} = \theta\mathbf{Z}$ converge to

$$\begin{aligned}\lambda_{max}(\theta\mathbf{Z}) &\to 1 + \frac{2\sqrt{\gamma}}{1+\sqrt{\gamma}};\\ \lambda_{min}(\theta\mathbf{Z}) &\to 1 - \frac{2\sqrt{\gamma}}{1+\sqrt{\gamma}},\end{aligned} \tag{4.14}$$

respectively, and when both N_T and N_R grow without bounds. Therefore, from (4.14), the eigenvalues of $(\mathbf{I}-\theta\mathbf{Z})$ are approximately bounded by $[-2\sqrt{\gamma}/(1+\gamma), 2\sqrt{\gamma}/(1+\gamma)]$ and when γ is large, $2\sqrt{\gamma}/(1+\gamma) \to 0$. Thus, for increasing values of γ, the following condition is satisfied:

$$\lim_{n\to\infty} (\mathbf{I} - \theta\mathbf{Z})^n \approx \mathbf{0}. \tag{4.15}$$

From [82], it is known that if $\mathbf{Z}$ satisfies (4.15), then its inverse can be expressed by a Neumann series as

$$\mathbf{Z}^{-1} \approx \sum_{n=0}^{K-1} (\mathbf{I} - \delta\theta\mathbf{Z})^n \theta, \tag{4.16}$$

with equality when the number of terms in the series K tends to infinity and where $\delta < 1$ is an attenuation factor such that the limits in (4.14) hold true. Equation (4.16) can be regarded as the summation of powers of a matrix, providing an iterative method to approximate $\mathbf{Z}^{-1}$ without directly computing any inverse. However, for a feasible and efficient implementation of (4.16), the number of considered terms (iterations) K should be finite and small. Noting that a matrix-by-matrix multiplication is order $\mathcal{O}(N_T^3)$, then the complexity in (4.16) is $\mathcal{O}((K-2)N_T^3)$, for $K \geq 3$. Even though the complexity order of an exact inverse is also $\mathcal{O}(N_T^3)$, from a hardware implementation perspective, matrix multiplications are preferable, making Neumann series worthwhile when advanced architectures are not at one's disposal. A comparison of the complexity for different K and an exact inverse method will be conducted in section 4.4.3.

4.4.2 Convergence of the Neumann Series

For realistic values of γ (finite N_T and N_R) the bounds in (4.14) may not be valid. If any of the moduli of the eigenvalues of $(\mathbf{I} - \delta\theta\mathbf{Z})$, denoted by $|\lambda(\mathbf{I} - \delta\theta\mathbf{Z})|$, is greater than 1, then convergence is not guaranteed [83]. To circumvent this problem and as was done in [80], a small modification shall now be introduced. Denote the scalar multiplication in (4.16) with a diagonal matrix $\mathbf{D}$, such that

$$\mathbf{D}^{-1} = \frac{\delta}{N_R + N_T}\mathbf{I}. \tag{4.17}$$

Substituting in (4.16), the approximate inverse can be written as

$$\mathbf{Z}^{-1} \approx \sum_{n=0}^{K-1} (\mathbf{I} - \mathbf{D}^{-1}\mathbf{Z})^n \mathbf{D}^{-1}. \tag{4.18}$$

For a fixed K, the convergence speed of (4.18) depends on the magnitude of $\rho = \lambda(\mathbf{I}-\mathbf{D}^{-1}\mathbf{Z})$. Provided $|\rho_{max} \leq 1|$, then the smaller the magnitude of $|\rho|$, the faster the convergence. Thus, the optimal choice for δ, such that $\lambda(\mathbf{I} - \mathbf{D}^{-1}\mathbf{Z})$ is as small as possible, is a non-trivial task. On the one hand, if δ is chosen too small, the convergence speed deteriorates and a large K would be required, increasing complexity.

On the other hand, if δ is too large, convergence may not be guaranteed. In order to solve this and as was proposed in [80], one may make use of Gershgorins circle theorem [84] and assume that $\mathbf{Z}$ is a diagonally dominant matrix [85] satisfying

$$|z_{ii}| > \sum_{j,i\neq j} |z_{ij}|, \quad i,j = 1,...,N_T. \tag{4.19}$$

Under this assumption, choosing the terms $\mathbf{D}^{-1} = \text{diag}(1/z_{11},...,1/z_{N_TN_T})$ guarantees that $\max_i |\rho_i| < 1$, hence assuring the convergence of the Neumman series [80].

Using the latter result, the Gram matrix $\mathbf{Z}$ can be decomposed into its diagonal $\mathbf{D}$ and its hollow $\mathbf{E}$ in the following way:

$$\mathbf{Z} = \mathbf{D} + \mathbf{E}. \tag{4.20}$$

Thus, (4.18) can be rewritten as

$$\tilde{\mathbf{Z}}_K^{-1} = \sum_{n=0}^{K-1} (-\mathbf{D}^{-1}\mathbf{E})^n \mathbf{D}^{-1}, \tag{4.21}$$

where $\tilde{\mathbf{Z}}_K^{-1}$ is the K-term approximation of $\mathbf{Z}^{-1}$. For $K = 1$, $\tilde{\mathbf{Z}}_1^{-1} = \mathbf{D}^{-1}$ corresponds to a scaled version of the MF detector. For $K = 2$, the approximation $\tilde{\mathbf{Z}}_2^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{E}\mathbf{D}^{-1}$ requires $\mathcal{O}(N_T^2)$ operations, which is still lower than the exact inverse $\mathcal{O}(N_T^3)$. For $K = 3$, the approximate inverse is given by:

$$\tilde{\mathbf{Z}}_3^{-1} = \mathbf{D}^{-1} - (\mathbf{D}^{-1}\mathbf{E})\mathbf{D}^{-1} + (\mathbf{D}^{-1}\mathbf{E})(\mathbf{D}^{-1}\mathbf{E}\mathbf{D}^{-1}). \tag{4.22}$$

Even though the complexity to compute (4.22) scales with $\mathcal{O}(N_T^3)$, only one matrix-by-matrix multiplication is required, resulting in a lower number of operations than the one of exact inverse methods. Finally, for $K \geq 4$, the expected complexity is higher than that of an exact inverse. Nonetheless, and as was stated before, a hardware implementation of an iterative method is much more preferable than when compared with exact inverse methods.

4.4.3 Complexity Analysis

In order to be able to effectively measure the trade-off between performance and complexity, an estimate of the needed number of operations is essential. As a standard for comparison, the exact inverse method based on Cholesky factorisation will be briefly detailed (note that other methods could have been chosen, such as QR and LU decompositions or Gauss-Jordan method).

The computational complexity is characterised by the sum of real-valued addition, multiplication and division operations. Compiler related complexity reduction techniques, such as strength reduction [86], will not be considered.

Cholesky Decomposition Exact Inversion

Exploiting the fact that $\mathbf{Z}$ is a positive definite symmetric matrix, Cholesky decomposition is applicable and factorises it into a lower triangular matrix $\mathbf{L}$ and its Hermitian transpose $\mathbf{L}^H$ as $\mathbf{Z} = \mathbf{L}\mathbf{L}^H$. A detailed description of the algorithm to compute $\mathbf{L}$ can be found in [84, 87] and its complexity, given in total number of FLOPS[1] N_f^{chol}, is:

$$N_f^{chol} = \frac{(N_T^3 - N_T)}{3} f_{add} + \frac{(N_T^2 - N_T)}{2} f_{div} + N_T f_{sqrt}, \tag{4.23}$$

where f_{add}, f_{div} and f_{sqrt} denote the number of necessary FLOPS to apply the addition (or multiplication), division and square root operators, respectively.

After $\mathbf{L}$ is obtained, computing $\mathbf{Z}^{-1}$ becomes less computational demanding. For that purpose, define $\mathbf{X} = \mathbf{Z}^{-1}$ and $\mathbf{M} = \mathbf{L}^{-1}$, such that [88, 89]

$$\begin{cases} \mathbf{L}\mathbf{m}_i = \mathbf{e}_i, \\ \mathbf{L}^H \mathbf{x}_i = \mathbf{m}_i, \end{cases} \tag{4.24}$$

where $\mathbf{e}_i$ is the ith column of $\mathbf{I}$. Hence, finding the inverse $\mathbf{Z}^{-1}$ from $\mathbf{L}$ requires solving two triangular left division systems, whose total number of FLOPS is given by [90]:

$$N_f^{syst} = 2N_T(N_T^2 f_{add} + N_T(f_{div} - 1)). \tag{4.25}$$

Finally, the total the number of FLOPS to compute the inverse via Cholesky decomposition is given by:

$$N_f^{total} = N_f^{chol} + N_f^{syst}, \tag{4.26}$$

corresponding to one Cholesky decomposition, followed by solving two triangular left divisions.

A graphical representation of the total number of FLOPS as a function of N_T for the computation of K terms of the Neumann series and the exact inverse via Cholesky decomposition is given in figure 4.3. The used parameters for the number of necessary FLOPS were $f_{add} = 1$, $f_{div} = 8$, $f_{sqrt} = 8$ [90]. For the Neumann series, both preprocessing matrices computation ($\mathbf{D}^{-1}$ and $\mathbf{D}^{-1}\mathbf{E}$) and matrix power operations in the series were added to the total number of FLOPS. It can be observed that when $K \leq 2$, the complexity is rather low even for a large number of users. This is justified with the fact that all multiplications involve diagonal matrices, leading to a rather low number of operations. From (4.22), for $K = 3$, one matrix-by-matrix multiplication is required, hence the depicted increased complexity was to be expected. Nonetheless, its complexity is still below the one when Cholesky factorisation is applied to compute the exact inverse; hence, the advantages for $K \leq 3$ are two-fold: not only are they easier to implement in hardware but also fewer computations are required. Lastly and even though the complexity

[1] Floating-point operations per second.

order for $K \geq 4$ is still $\mathcal{O}(N_T^3)$, the high number of required matrix-by-matrix multiplications leads to a larger number of FLOPS than when compared with the exact inverse via Cholesky decomposition. Nevertheless and depending on the impact in the performance (studied next), Neumann series is an iterative method, thus the case when $K \geq 4$ might still have a practical interest for hardware implementation.

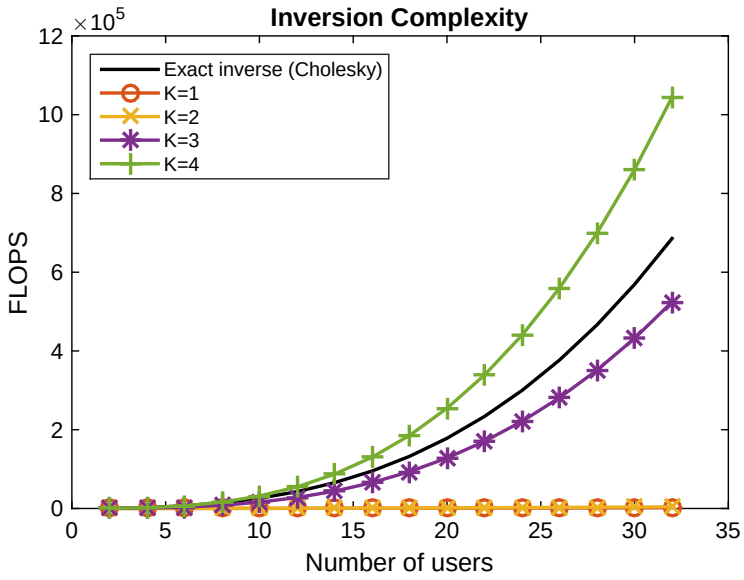


Figure 4.3: Number of required FLOPS to compute inverse of a matrix using the Cholesky decomposition and the Neumann series employing K terms.

4.4.4 Performance Impact of the Neumann Series

The reduction in complexity for $K \leq 3$ provided by the Neumann series comes at the cost of an approximation error. In this section, both the magnitude of the error and its impact in performance will be evaluated. Firstly, denote the approximation error for the inverse $\mathbf{Z}^{-1}$ as $\Delta_K = \mathbf{Z}^{-1} - \tilde{\mathbf{Z}}_K^{-1}$ [89]. Expanding it leads to:

$$\Delta_K = \sum_{k=K}^{\infty} (-\mathbf{D}^{-1}\mathbf{E})^k \mathbf{D}^{-1} \tag{4.27}$$

$$= (-\mathbf{D}^{-1}\mathbf{E})^K \sum_{k=0}^{\infty} (-\mathbf{D}^{-1}\mathbf{E})^k \mathbf{D}^{-1} \tag{4.28}$$

$$= (-\mathbf{D}^{-1}\mathbf{E})^K \mathbf{Z}^{-1}. \tag{4.29}$$

The purpose of the approximation is to equalise the received signal $\mathbf{y}$. Making use of either ZF or MMSE reviewed in (4.12) and (4.13), the symbols after equalisation with the approximation are given by [89]

$$\hat{\mathbf{r}}_K = \tilde{\mathbf{Z}}_K^{-1} \mathbf{H}^H \mathbf{y} = \mathbf{Z}^{-1} \mathbf{H}^H \mathbf{y} - \Delta_K \mathbf{H}^H \mathbf{y}. \tag{4.30}$$

The l_2-norm of the residual approximation error $\Delta_K \mathbf{H}^H \mathbf{y}$ can be bounded by

$$||\Delta_K \mathbf{H}^H \mathbf{y}||_2 = ||(-\mathbf{D}^{-1}\mathbf{E})^K \mathbf{Z}^{-1}\mathbf{H}^H \mathbf{y}||_2, \tag{4.31}$$

$$\leq ||(-\mathbf{D}^{-1}\mathbf{E})^K||_F ||\mathbf{Z}^{-1}\mathbf{H}^H \mathbf{y}||_2 \tag{4.32}$$

$$\leq || - \mathbf{D}^{-1}\mathbf{E}||_F^K ||\mathbf{Z}^{-1}\mathbf{H}^H \mathbf{y}||_2 \tag{4.33}$$

From the latter result, it is verified that if the condition $||\mathbf{D}^{-1}\mathbf{E}|| < 1$ is satisfied, then convergence is guaranteed and the error approaches zero exponentially fast as $K \to \infty$. Is this condition satisfied for large-scale MIMO systems? The authors in [89] partially answered this question with the following theorem.

Theorem 2. *Let $N_R > 4$ and assume the entries of $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ are drawn from a $\mathcal{CN}(0,1)$ distribution. Then,*

$$P(||\mathbf{D}^{-1}\mathbf{E}||_F^K < \alpha) \geq 1 - \frac{(N_T^2 - N_T)}{\alpha^{\frac{2}{K}}} \sqrt{\frac{2N_R(N_R+1)}{(N_R-1)(N_R-2)(N_R-3)(N_R-4)}}. \tag{4.34}$$

Theorem 2 provides good insight into the probability of convergence of the Neumann series. For $K = 1$ and $\alpha = 1$, it can be inferred that an increase in the ratio $\frac{N_R}{N_T}$ leads to a higher probability of convergence. In addition, when $\alpha \leq 1$ an increase in K also increases the probability that the error in the K-term approximation is smaller than α. Ultimately, when $N_R \to \infty$ with N_T constant and $\alpha \in (0, 1]$, the limit $P(||\mathbf{D}^{-1}\mathbf{E}||_F^K < \alpha) \to 1$ holds, a similar result to the one verified during the MF analysis.

4.4.5 Numerical Results

The performance of the Neumann series in massive MIMO systems will be here evaluated in terms of BER curves and using the MMSE filter. Firstly, define the quality of the channel as SNR $= N_T \frac{\sigma_x^2}{\sigma_n^2}$ and the constellation size to be 16-QAM. Secondly, fix the number of receive antennas $N_R = 128$ and vary the number of transmit antennas $N_T = 8, 16, 32$.

Numerical simulations results are depicted in figure 4.4 and, on the whole, it can be stated that Neumann series provides an efficient way to invert $\mathbf{Z}$, especially when $K \geq 3$ and when γ is large. In particular, when $N_T = 8$ and $K \geq 3$, Neumann series and exact inversion attain the same performance, whilst for $K = 1$, BER is stalled at around 5×10^{-2}. This error floor is explained with the inaccuracy resulting from the approximation, and a similar effect is noticed when γ decreases ($N_T = 16, 32$). The depicted results were to be anticipated, since a higher K means a better approximation, leading to a better performance. It is worth mentioning, however, that when the number of antennas was set to $N_T = 32$, $K = 3$ attained lower BER than $K = 4$, a result that is counter-intuitive. This can be explained with the eigenvalues analysis conducted before; when $N_T = 32$ the condition $|\lambda_{max}(\mathbf{I} - \mathbf{D}^{-1}\mathbf{Z}) \leq 1|$ is not necessarily satisfied (this fact will be verified next), thus adding an extra term in the series is prejudicial to the approximation. As previously stated, the performance degradation verified in figure 4.4 is closely

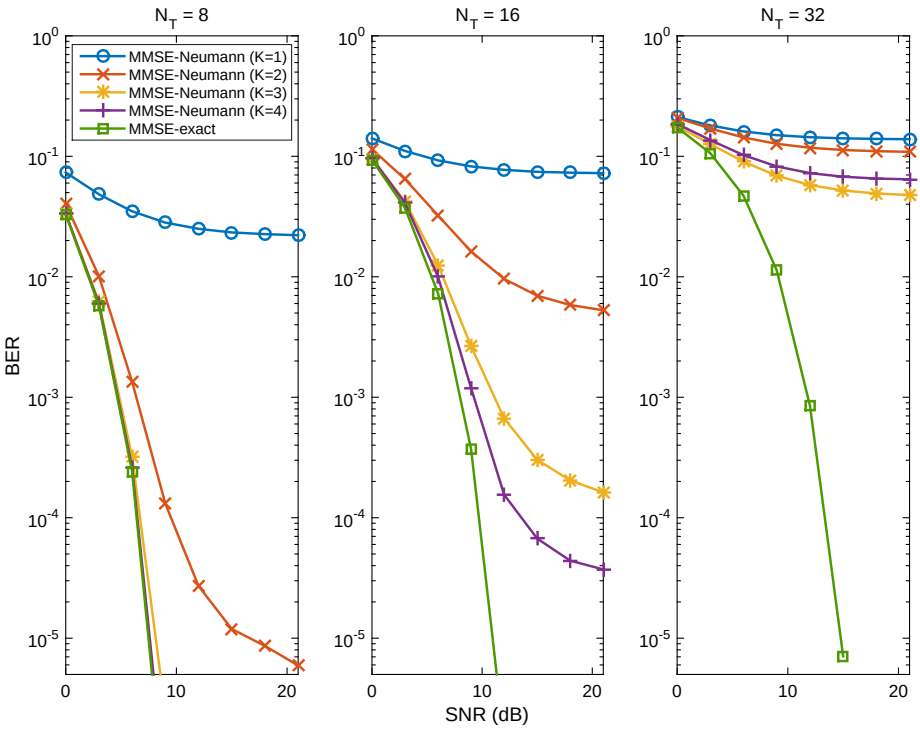


Figure 4.4: BER performance comparison between exact inverse and Neumann series with different number of terms K. Number of receive antennas was fixed to $N_R = 128$ and used constellation size was 16-QAM.

related to the modulus of the eigenvalues of $(\mathbf{I} - \mathbf{D}^{-1}\mathbf{Z})$. In order to accelerate the convergence of the series without drastically increasing complexity, one may consider off-diagonal elements for the initial matrix $\mathbf{D}$ [79]. For that purpose, define the new initial matrix as $\mathbf{D}_o$, which can be expressed as:

$$\mathbf{D}_o = \text{diag}_0(\mathbf{Z}) + \sum_{i=1}^{N_{off}} \mathbf{D}_l, \tag{4.35}$$

where N_{off} is the number of off-diagonals to be considered and $\mathbf{D}_l \in \mathbb{C}^{N_T \times N_T}$ is defined as

$$[\mathbf{D}_l]_{i,j} = \begin{cases} z_{i,j}, & |i-j| = l \\ 0, & \text{otherwise.} \end{cases} \tag{4.36}$$

Considering one off-diagonal ($N_{off} = 1$, tridiagonal matrix) and $K = 3$, the improvement in cumulative distribution function (CDF) of $|\lambda_{max}(\mathbf{I} - \mathbf{D}_o^{-1}\mathbf{Z})|$ can be verified in figure 4.5. As can be inferred, there is a non negligible gain when diagonal or tridiagonal matrices are being used. Moreover, when $N_T = 32$, the maximum eigenvalue surpasses 1 in more than half of the instances, hence justifying the counter-intuitive result in figure 4.4. The better eigenvalue distribution provided by the tridiagonal matrix leads necessarily to improved BER results, since a faster convergence in the series leads to lower error values in the approximation. This fact can be confirmed in figure 4.6, where the 3-term approximation performance using both Neumann series and tridiagonal matrix techniques is depicted.

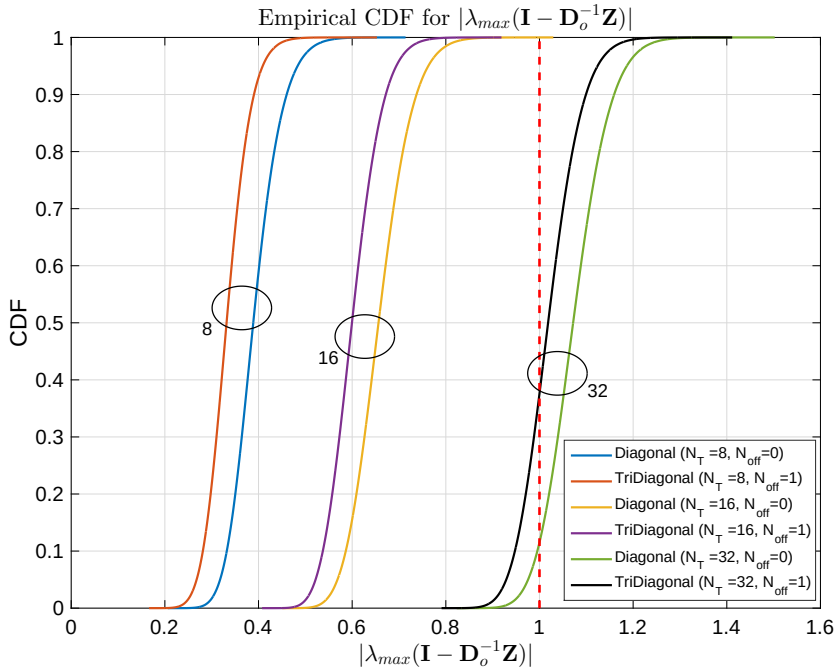


Figure 4.5: Cumulative distribution function of maximum eigenvalues of $\mathbf{I}-\mathbf{D}_o^{-1}\mathbf{Z}$ for $N_{off}=0$ and $N_{off}=1$, when $N_R=128$. Number of transmit antennas $N_T=8,16$ and 32 are considered.

4.5 Updating the Inverse of a Matrix

Until now, a method based on the Neumann series to efficiently invert matrices has been studied. Obviously, the performance of such method is highly conditioned on the ratio γ and the number of terms K. For instance, when the ratio γ decreases, performance degrades. It has been proven, however, that the inclusion of off-diagonal elements in $\mathbf{D}$ can help mitigating this deterioration, at the cost of slightly increased complexity.

Capitalising on the results obtained thus far, methods to efficiently update the inverse of a matrix after a small perturbation are proposed here for the first time in the context of massive MIMO. The reason behind this study is simple: when a user is added or removed from the system, one is interested in recomputing the new inverse $\mathbf{Z}^{-1}$ in the least time possible in order to maximise data throughput and minimise power consumption. Having this in mind, it is possible, using the matrix inversion lemma, to decrease the number of computations from $\mathcal{O}(N_T^3)$ to $\mathcal{O}(N_T^2)$, hence increasing speed. Moreover, in future wireless networks, it might make sense to update $\mathbf{Z}^{-1}$ when a new channel estimation from a single user is obtained (which corresponds to changing one column in $\mathbf{H}$ and altering a column and a row in $\mathbf{Z}$). Without recomputing the entire inverse, this can be achieved using the Sherman-Morrison formula (a special case of the matrix inversion lemma).

The algorithms and corresponding impact in the performance when a user is added, removed or updated are described in the following sections. Results here provided are done for MIMO detection but they can easily be adapted for linear precoding schemes. Moreover, the provided methods can be applied using either an initial approximation $\tilde{\mathbf{Z}}_K^{-1}$ or the exact inverse $\mathbf{Z}^{-1}$ (which coincides with $\tilde{\mathbf{Z}}_\infty^{-1}$).

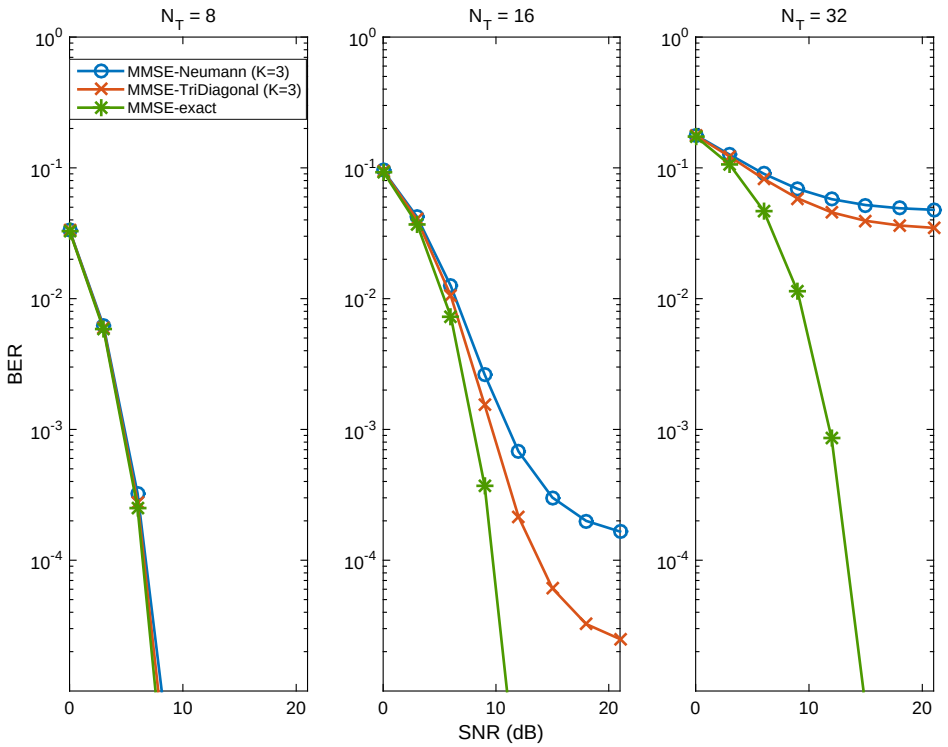


Figure 4.6: BER performance comparison between exact inverse and Neumann series with different initial "diagonal" matrices. Number of receive antennas was fixed to $N_R = 128$ and used constellation size was 16-QAM.

4.5.1 Adding and Removing a User

Assume that at a given time instant, the inverse $\mathbf{Z}^{-1} = (\mathbf{H}^H\mathbf{H})^{-1}$ is already computed (via exact inversion or Neumann series) and that a user is added to the system with estimated channel denoted by the column vector $\mathbf{h}_n$. Define the new extended matrix as $\mathbf{H}_e = [\mathbf{H}\ \mathbf{h}_n]$ (the user is added in the last column but in fact it could have been added in any position). Hence, the new Gram matrix denoted by $\mathbf{Z}_e$ is given by:

$$\mathbf{Z}_e = \begin{bmatrix} \mathbf{H}^H \\ \mathbf{h}_n^H \end{bmatrix} \begin{bmatrix} \mathbf{H} & \mathbf{h}_n \end{bmatrix} = \begin{bmatrix} \mathbf{H}^H\mathbf{H} & \mathbf{H}^H\mathbf{h}_n \\ \mathbf{h}_n^H\mathbf{H} & \mathbf{h}_n^H\mathbf{h}_n \end{bmatrix} \tag{4.37}$$

In order to find $\mathbf{Z}_e^{-1}$, one may use the general result provided by the inverse of a partitioned matrix [91], which is written as follows (see [92, Appendix B]):

$$\begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{F}_{11}^{-1} & -\mathbf{F}_{11}^{-1}\mathbf{P}_{12}\mathbf{P}_{22}^{-1} \\ -\mathbf{P}_{22}^{-1}\mathbf{P}_{21}\mathbf{F}_{11}^{-1} & \mathbf{F}_{22}^{-1} \end{bmatrix}, \tag{4.38}$$

where

$$\mathbf{F}_{11} = \mathbf{P}_{11} - \mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21}; \tag{4.39}$$

$$\mathbf{F}_{22} = \mathbf{P}_{22} - \mathbf{P}_{21}\mathbf{P}_{11}^{-1}\mathbf{P}_{12}. \tag{4.40}$$

Using the result provided above, it is possible to rewrite $\mathbf{Z}_e^{-1}$ as

$$\mathbf{Z}_e^{-1} = \begin{bmatrix} \mathbf{H}^H\mathbf{H} & \mathbf{H}^H\mathbf{h}_n \\ \mathbf{h}_n^H\mathbf{H} & \mathbf{h}_n^H\mathbf{h}_n \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{F}_{11}^{-1} & -d\mathbf{Z}^{-1}\mathbf{H}^H\mathbf{h}_n \\ -d\mathbf{h}_n^H\mathbf{H}\mathbf{Z}^{-H} & d \end{bmatrix}, \tag{4.41}$$

where

$$d = \frac{1}{\mathbf{h}_n^H\mathbf{h}_n - \mathbf{h}_n\mathbf{H}\mathbf{Z}^{-1}\mathbf{H}^H\mathbf{h}_n}; \tag{4.42}$$

$$\mathbf{F}_{11}^{-1} = \mathbf{Z}^{-1} + d\mathbf{Z}^{-1}\mathbf{H}^H\mathbf{h}_n\mathbf{h}_n^H\mathbf{H}\mathbf{Z}^{-H}. \tag{4.43}$$

Therefore, it is possible to update the inverse of a matrix $\mathbf{Z} = \mathbf{H}^H\mathbf{H}$ when a column is added to $\mathbf{H}$ at position p without explicitly recomputing $\mathbf{Z}_e^{-1}$ or $\tilde{\mathbf{Z}}_{e,K}^{-1}$. The method is summarised in algorithm 1 and, for comparison purposes, the corresponding number of required real-valued multiplications and divisions is shown on the right hand side (computation of both $\mathbf{H}^H\mathbf{h}_p$ and $\mathbf{h}_p^H\mathbf{h}_p$ were not considered since they are also required in the recomputation of $\mathbf{D}$ and $\mathbf{E}$ in the Neumann series). If the initial $\mathbf{Z}^{-1}$ is the exact inverse, then $\mathbf{Z}_e^{-1}$ is also exact. On the other hand, if an approximation $\tilde{\mathbf{Z}}_K^{-1}$ is used, then a propagation of errors is to be expected. Intuitively, the greater the error Δ_K, the greater the degradation in performance after the update.

Algorithm 1 Update $\mathbf{Z}^{-1}$, when a user $\mathbf{h}_p$ is added to $\mathbf{H}$ at position p

Input: $\mathbf{Z}^{-1}, \mathbf{H}, \mathbf{h}_p$

$\mathbf{t}_1 \leftarrow \mathbf{H}^H\mathbf{h}_p$

$\mathbf{t}_2 \leftarrow \mathbf{Z}^{-1}\mathbf{t}_1$ ▷ $4M^2$

$d \leftarrow 1/(\mathbf{h}_p^H\mathbf{h}_p - \mathbf{t}_1^H\mathbf{t}_2)$ ▷ $4M + 1$

$\mathbf{t}_3 \leftarrow d\mathbf{t}_2$ ▷ $2M$

$\mathbf{F}_{11}^{-1} \leftarrow \mathbf{Z}^{-1} + d\mathbf{t}_2\mathbf{t}_2^H$ ▷ $3M^2 + 3M$

$\mathbf{Z}_e^{-1} \leftarrow \begin{bmatrix} \mathbf{F}_{11}^{-1} & -\mathbf{t}_3 \\ -\mathbf{t}_3^H & d \end{bmatrix}$

Change last column and last row of $\mathbf{Z}_e^{-1}$ to column p and row p

Output: $\mathbf{Z}_e^{-1}$

Denote the number of updates by U, corresponding to the number of new added users after the last full inverse ($\mathbf{Z}^{-1}$ or $\tilde{\mathbf{Z}}_K^{-1}$) was computed. The extended inverse matrix after U sequential updates is designated by $\mathbf{Z}_{e,U}^{-1} \in \mathbb{C}^{(N_T+U)\times(N_T+U)}$. For instance, $U = 2$ means that algorithm 1 was run twice and its input matrices were $\tilde{\mathbf{Z}}_K^{-1}$ and $\mathbf{Z}_{e,1}^{-1}$ for the first and second repetitions, respectively.

The result using the ZF approximation $\tilde{\mathbf{Z}}_3^{-1}$ as the initial vector and performing 3 successive updates is exposed in figure 4.7. Initial number of antennas was defined as $N_T = 8$ and $N_R = 80$ and constellation size set to 64-QAM. For comparison purposes, performance using exact inverse (with dimensions $N_T + U$) and Neumann series approximation for both N_T and $N_T + U$ cases are also exhibited. The result shows that the degradation in performance when a user is added via matrix inversion lemma is neg-

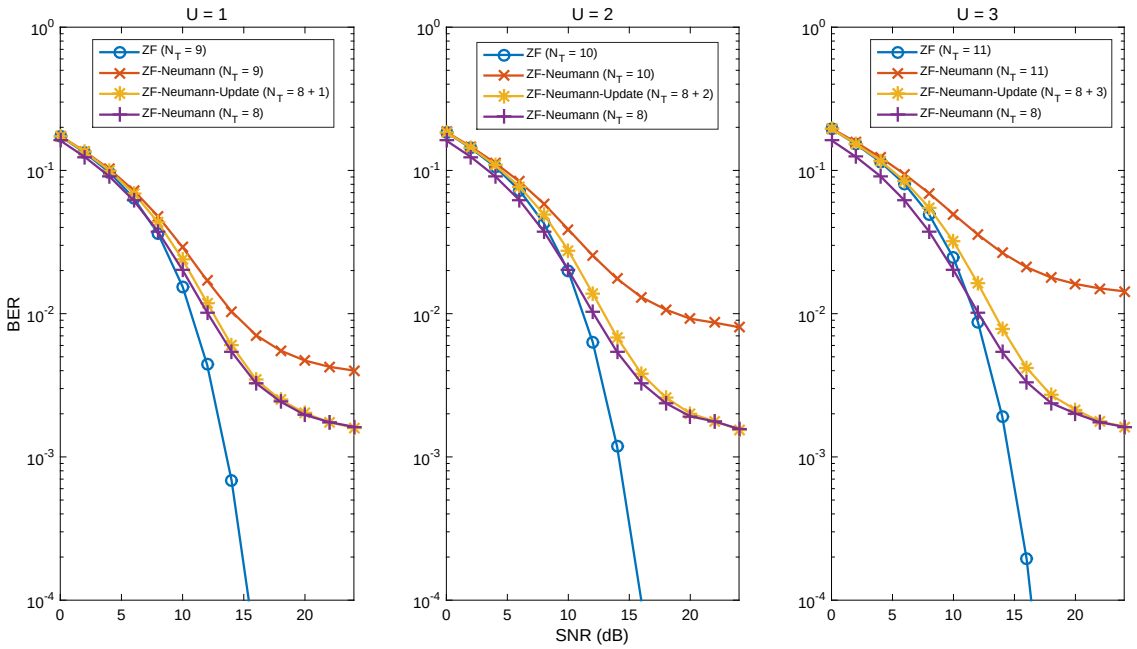


Figure 4.7: BER performance comparison between exact inverse, Neumann series approximations and update via matrix inversion lemma when U users are added. Number of receive antennas was fixed to $N_R = 80$ and used constellation size was 64-QAM.

ligible, even outperforming the case where the entire recalculation of $\tilde{\mathbf{Z}}_{e,K}^{-1}$ is done using the Neumann series. Further, the number of required operations to perform the update using algorithm 1 is lower than calculating the inverse via Neumann series ($K \geq 3$). Therefore, this result might be of interest when users need to be added on-the-fly and in cases where high matrix inversion-throughputs are required.

Based on the previous results, a similar approach can be conducted when a column from $\mathbf{H}$ is removed. Decompose the original matrix as $\mathbf{H} = [\mathbf{H}_r \ \mathbf{h}_n]$, where $\mathbf{H}_r$ is the reduced matrix from which we want to compute the inverse $\mathbf{Z}_r^{-1} = (\mathbf{H}_r^H \mathbf{H}_r)^{-1}$ and $\mathbf{h}_n$ the column corresponding to user n and that is to be removed. The initial, already computed, inverse matrix $\mathbf{Z}^{-1}$ is given by

$$\mathbf{Z}^{-1} = \begin{bmatrix} \mathbf{H}_r^H \mathbf{H}_r & \mathbf{H}_r^H \mathbf{h}_n \\ \mathbf{h}_n^H \mathbf{H}_r & \mathbf{h}_n^H \mathbf{h}_n \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{F}_{11}^{-1} & -d\mathbf{u} \\ -d\mathbf{u}^H & d \end{bmatrix}, \tag{4.44}$$

where $\mathbf{u} = \mathbf{Z}_r^{-1} \mathbf{H}_r^H \mathbf{h}_n$. Noting that $\mathbf{F}_{11}^{-1} = \mathbf{Z}_r^{-1} + d\mathbf{u}\mathbf{u}^H$ (from (4.43)), it is straightforward that

$$\mathbf{Z}_r^{-1} = \mathbf{F}_{11}^{-1} - d\mathbf{u}\mathbf{u}^H. \tag{4.45}$$

The step-by-step set of operations to compute (4.45) is outlined in algorithm 2. It is worth mentioning that the number of required computations is inferior to the one in algorithm 1. Numerical results using an analogous setup to the previous case are shown in figure 4.8. This time, BER performance using the Neumann series approximation is better if a recalculation from scratch is performed (for $N_T - U$). This is related with the smaller error propagation associated with lower-dimensional matrices. Nonetheless,

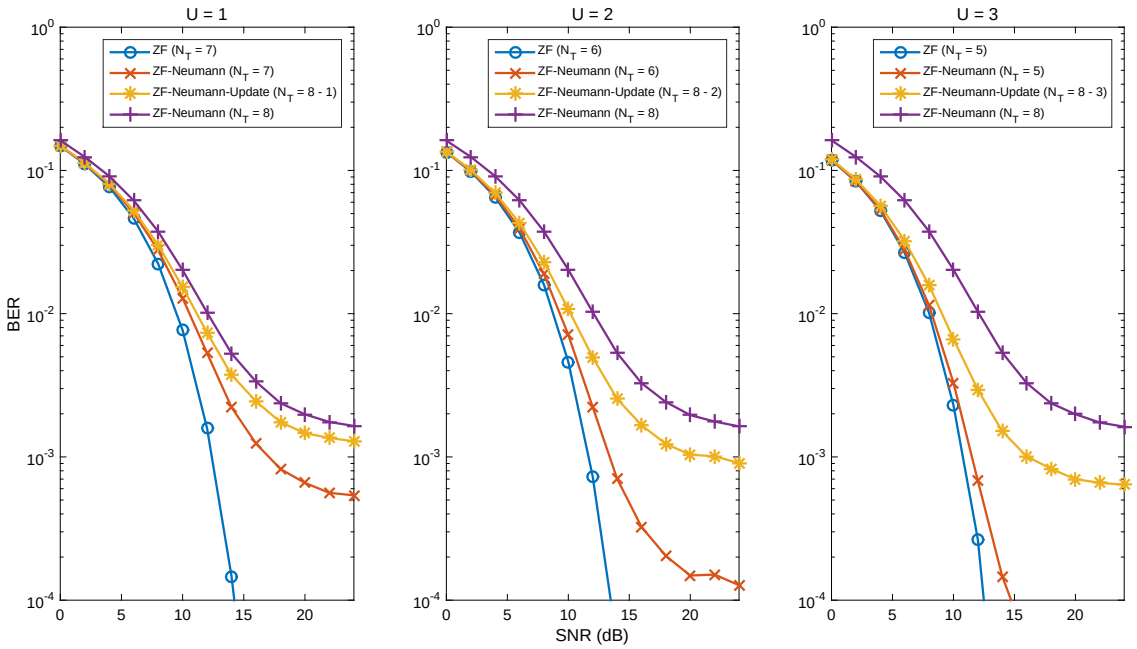


Figure 4.8: BER performance comparison between exact inverse, Neumann series approximations and update via matrix inversion lemma when U users are removed. Number of receive antennas was fixed to $N_R = 80$ and used constellation size was 64-QAM.

the gain in complexity provided by algorithm 2 might be exploited in systems where BER requirements are not too strict. In addition, this degradation could have been mitigated if a better initial approximation ($K > 3$) of $\mathbf{Z}^{-1}$ had been used.

Algorithm 2 Update $\mathbf{Z}^{-1}$, when a user $\mathbf{h}_p$ is removed from $\mathbf{H}$ at position p

Input: $\mathbf{Z}^{-1}, p$

Change column p and row p of $\mathbf{Z}^{-1}$ to last column and last row

$\mathbf{F}_{11}^{-1} \leftarrow \mathbf{Z}^{-1}(1:(N_T-1), 1:(N_T-1))$

$d \leftarrow \mathbf{Z}^{-1}(N_T, N_T)$

$\mathbf{t} \leftarrow -\mathbf{Z}^{-1}(1:(N_T-1), N_T)$

$\mathbf{Z}_r^{-1} \leftarrow \mathbf{F}_{11}^{-1} - \mathbf{t}\mathbf{t}^H/d$ ▷ $3M^2 + 3M$

Output: $\mathbf{Z}_r^{-1}$

4.5.2 Updating an Inverse when a Column is Changed

Formerly, the situations where U users are added or removed from the system were considered. Likewise, an update to the inverse when a new channel estimation for user p is obtained can be studied. This operation corresponds to replacing column p of $\mathbf{H}$ with the new estimation $\mathbf{h}_p$. After $\mathbf{H}$ is updated with the new estimation $\mathbf{h}_p$, expressed as $\mathbf{H}_u$ the new product $\mathbf{Z}_u = \mathbf{H}_u^H \mathbf{H}_u$ corresponds to altering column p and row p in the original $\mathbf{Z}$. This rank-2 perturbation can, however, be decomposed into two rank-1 perturbations. Defining a rank-1 perturbation by $\mathbf{a}\mathbf{b}^H$, one can use Sherman-Morrison formula to

recompute the inverse after the update, as follows [93]:

$$(\mathbf{Z}+\mathbf{a}\mathbf{b}^H)^{-1}=\mathbf{Z}^{-1}-\frac{(\mathbf{Z}^{-1}\mathbf{a})(\mathbf{b}^H\mathbf{Z}^{-1})}{(1+\mathbf{b}^H\mathbf{Z}^{-1}\mathbf{a})}. \tag{4.46}$$

Taking into account (4.46) and without recomputing the inverse, calculating $\mathbf{Z}_u$ from the previous $\mathbf{Z}$ is feasible. For that purpose, define $\mathbf{z}_{u,p}=\mathbf{H}_u^H\mathbf{h}_p$ and compute $\mathbf{a}_1=\mathbf{z}_{u,p}-\mathbf{z}_p$, where $\mathbf{z}_p$ is the pth column of $\mathbf{Z}$. Further, make $\mathbf{b}_2=\mathbf{a}_1^{(0,p)}$, where $\mathbf{a}_1^{(0,p)}$ is the same as $\mathbf{a}_1$ but with the pth entry zeroed. Then, computing $\mathbf{Z}_u^{-1}$ is straightforward:

$$\mathbf{Z}_t^{-1}=(\mathbf{Z}+\mathbf{a}_1\mathbf{e}_p^T)^{-1}=\mathbf{Z}^{-1}-\frac{(\mathbf{Z}^{-1}\mathbf{a}_1)(\mathbf{e}_p^T\mathbf{Z}^{-1})}{(1+\mathbf{e}_p^T\mathbf{Z}^{-1}\mathbf{a}_1)}; \tag{4.47}$$

$$\mathbf{Z}_u^{-1}=(\mathbf{Z}_t+\mathbf{e}_p\mathbf{b}_2^H)^{-1}=\mathbf{Z}_t^{-1}-\frac{(\mathbf{Z}_t^{-1}\mathbf{e}_p)(\mathbf{b}_2^H\mathbf{Z}_t^{-1})}{(1+\mathbf{b}_2^H\mathbf{Z}_t^{-1}\mathbf{e}_p)}, \tag{4.48}$$

where $\mathbf{e}_p$ is the pth column of the identity matrix $\mathbf{I}$. Equations (4.47) and (4.48) require $24M^2$ scalar multiplications and 4 divisions; thus, when M is large, updating the inverse via this method might be advantageous over exact inversion in terms of complexity. Note that updating the inverse when only one column in $\mathbf{H}$ is changed can also be regarded as removing and adding a column sequentially (algorithms 2 and 1).

Figure 4.9 depicts the situation after U different columns in $\mathbf{H}$ are changed (the dimensions of the inverse matrices after the update remain the same). As can be inferred, the sequential operations in (4.47) and (4.48) using an approximation as initial inverse led to a propagation of errors, resulting in increased BER values. However, it is interesting to note that performing a deflation followed by an inflation (running algorithm 2 and 1 sequentially) gets better results than computing a 3-term Neumann series from start. This is explained with the vanished error contribution from "removed" columns in $\mathbf{H}$, since algorithms 1 and 2 return the exact inverse if the initial inverse matrix is also exact. Ultimately, if $U=M$ updates were to be performed, then exact inverse would be attained regardless of the initial matrix (though, its cost would be similar to one of an exact inverse).

4.6 Concluding Remarks

In this chapter, the advantages of augmenting the number of antennas have been identified. In particular, it has been shown that, when using massive MIMO systems, the spectral capacity can be drastically increased. Moreover, under the assumption that $N_T << N_R$, linear processing techniques for detection and precoding were shown to perform close to optimal. By exploiting the diagonalisation of the channel, efficient inverse computation methods based on the Neumann series and matrix inversion lemma have been evaluated. It is worth emphasising, however, that perfect knowledge of the channel at the receiver was always assumed. Hence, the study of suitable channel estimation algorithms and consequent impact in performances is left for future work.

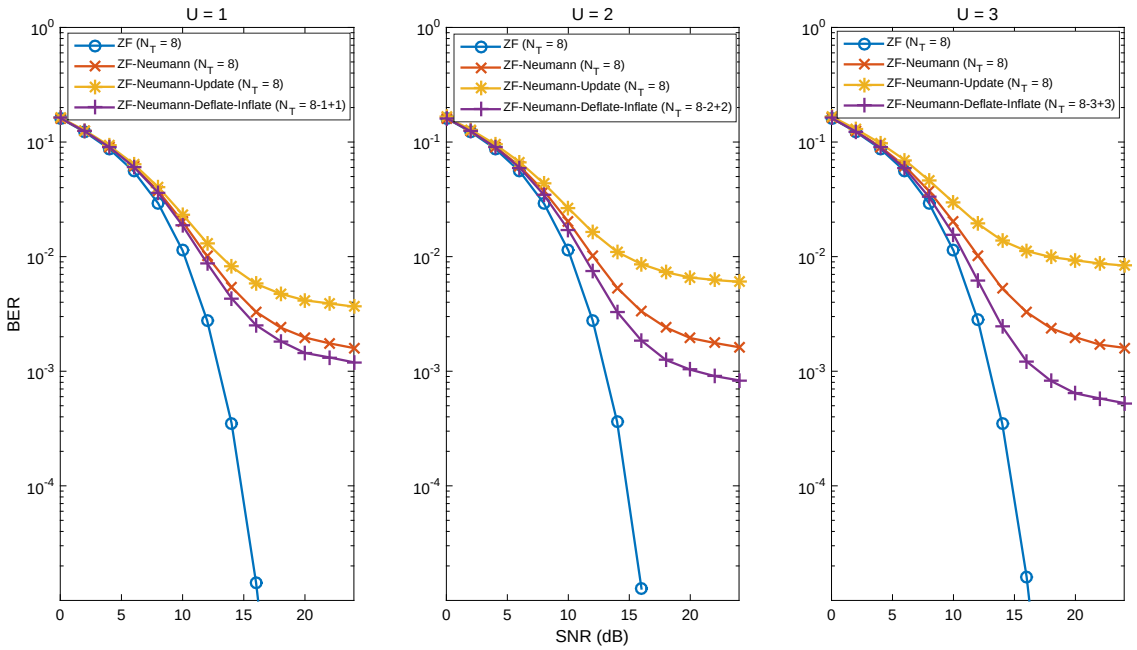


Figure 4.9: BER performance comparison between exact inverse, Neumann series approximation, inverse update via Sherman-Morrison formula and a sequential run of algorithms 2 and 1 when U users' channels are updated. Number of receive antennas was fixed to $N_R = 80$ and used constellation size was 64-QAM.

Chapter 5

Symmetric Large MIMO Detection

In this chapter a lattice reduction algorithm and several variants of the Gibbs sampling technique to perform detection will be detailed, optimised and compared. A study of the underlying complexities is also conducted.

5.1 Introduction

Large multiuser MIMO wireless systems, where the base station (BS) has tens to hundred of antennas and the users have one or more antennas has gained much attention over the past few years. This is in part due to the increasing demand from mobile operators clients for better quality of service and seamless experience. Even though most of the data traffic comes from the downlink, uplink speeds and reliability must not be neglected in the design of future wireless networks.

In the reverse link, the simultaneously transmitted signals from the users to the BS are separated by their spatial signatures, also denoted in the literature as the channel or return channel. Once received, these signals must be processed such that information can be retrieved.

It has been demonstrated in chapter 4 that when the number of served users N_T is much smaller than the number of antennas at the BS, denoted by N_R, linear processing techniques attain very good performance, while yielding low complexity. However, when $N_T \rightarrow N_R$ this is not the case anymore. Mathematically, this is explained by the achieved diversity of linear receivers, which is given by $N_R - N_T + 1$, for $N_R \geq N_T$. In the limit, when $N_R = N_T$, attained diversity is 1, which is far from optimal. In order to guarantee maximum diversity, one would have to perform maximum likelihood (ML) detection, whose complexity is exponential with the number of transmit antennas, thus making it impossible to perform detection in real time for large MIMO. The objective of this chapter is to study and evaluate near optimal, low complexity detectors that are suitable for the symmetric case when $N_R = N_T >> 1$. In particular, a state-of-the-art lattice reduction procedure and an optimised Markov chain Monte Carlo method will be described and compared.

5.1.1 System Model

Consider a MIMO system on the uplink with N_R receive antennas at the BS and $N = N_T = N_R$ served users (with one antenna each). In this work $N \leq 10$ will be considered to be conventional MIMO, $10 < N \leq 20$ to be large MIMO and $N > 20$ to be massive MIMO.

All users transmit symbols from an M-QAM alphabet $\mathcal{B}$, such that

$$\mathcal{B} = \mathcal{A} + j\mathcal{A}, \tag{5.1}$$

where $\mathcal{A} = \{-\log_2 M + 1, ..., -1, 1, \log_2 M - 1\}$. Let $x_k \in \mathcal{B}$ denote the transmitted symbol from user k and $\mathbf{x} = [x_1, ..., x_{N_T}]^T$ the vector containing the N_T symbols. The received signal $\mathbf{y} \in \mathbb{C}^{N_R \times 1}$ can be expressed as[1]

$$\mathbf{y} = \sqrt{\frac{\text{SNR}}{N_T}}\mathbf{H}\mathbf{x} + \mathbf{n}, \tag{5.2}$$

where $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ is the channel gain matrix and $\mathbf{n}$ is the noise vector, both having i.i.d. Gaussian entries $\mathcal{CN}(0, 1)$. The factor $\sqrt{\frac{\text{SNR}}{N_T}}$ is the ratio of the total transmit energy per channel use divided by the per-component noise variance. Namely, the SNR can be expressed as

$$\text{SNR} = \frac{\mathbb{E}\{||\mathbf{y} - \mathbf{n}||^2\}}{\mathbb{E}\{||\mathbf{n}||^2\}}. \tag{5.3}$$

Under the considered model, the vector that minimises the error probability (ML) is

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{B}^{N_T}} ||\mathbf{y} - \sqrt{\frac{\text{SNR}}{N_T}}\mathbf{H}\mathbf{x}|| = \arg \min_{\mathbf{x} \in \mathcal{B}^{N_T}} f(\mathbf{x}), \tag{5.4}$$

where $f(\mathbf{x})$ is the ML cost function. Equation (5.4) is an NP-hard integer least squares problem, whose aim is to find the integer lattice point (belonging to the alphabet) closest to the received signal. When the problem size is moderate, sphere decoder (SD) algorithms can be efficiently used to solve (5.4). However, as the number of dimensions increases, SD average complexity becomes exponential, hence other methods must be looked at [70].

In this work, only hard-decision outputs, where the inference about $\mathbf{x}$ is made solely based on $\mathbf{y}$, will be considered. Conversely, there is the possibility to generate soft-decision values which are preferred as inputs to the channel decoder. Nevertheless, as coded transmissions are not contemplated in this work, performance with soft-decisions will not be evaluated.

5.2 Lattice Reduction-Aided

A large MIMO suitable lattice reduction detector will now be presented, including an overview of its design and subsequent numerical results. The algorithm that is studied here belongs to the element-based

[1]Note that this channel model is identical to the ones presented in previous chapters. However, for both convenience and implementation purposes, this model is adapted in this chapter.

lattice reduction (ELR) family and it is formulated to solve the "shortest longest basis" (SLB) of the dual space [68]. For this reason, this method is denominated as Dual-Element-Based Lattice Reduction-Shortest Longest Basis (D-ELR-SLB) and its formulation will be reviewed in the next section. The outcomes of this evaluation will then be used for comparison purposes. In particular, its performance will be confronted with the ones of randomised algorithms [94], which are the main focus of this work. For analysis purposes, the gain term $\sqrt{\frac{\text{SNR}}{N_T}}$ will be dropped in the remaining parts of section 5.2.

5.2.1 D-ELR-SLB Algorithm

Element-based lattice reduction algorithms were first proposed in the context of single carrier frequency division multiple access (SC-FDMA) in [69]. Afterward, D-ELR-SLB algorithms were adapted to MIMO detection and because of their inherent low complexity, the extension to the large and massive MIMO cases was direct and conducted in [68].

Most LR algorithms aim to maximise the orthogonality of the basis $\mathbf{H}$. Reducing the basis $\mathbf{H}$ is equivalent to finding $\tilde{\mathbf{H}} = \mathbf{HT}$, with $\mathbf{T}$ being a complex unimodular matrix[2]. One of the most cited LR algorithms in the literature is the LLL and its extension to the dual (D-LLL [95]) and complex cases (CLLL [67]). However, the performance of LLL methods degrade when large MIMO systems are considered and, as a consequence, the iterative SA is often used as the metric for these systems [68]. Despite the good performance attained by SAs in large MIMO, their high computational cost for each basis update is far from ideal [96]; hence, the motivation for the design of efficient lattice reduction algorithms that are suitable for large MIMO.

The purpose of the studied algorithm is different from the previous; the goal is to minimise the diagonal elements $C_{i,i}$ in the noise covariance matrix after equalisation, given by $\mathbf{C} = (\mathbf{H}^H\mathbf{H})^{-1}$ [97]. For a given transmitted symbol x_k and its hard-decision $\hat{x}_k$, one is to reduce the asymptotic pairwise error probability (PEP) after linear detection given by [54, Equation 12]:

$$P(x_k \neq \hat{x}_k|\mathbf{H}) = Q\left(\sqrt{\frac{|e_x|^2}{2C_{i,i}}}\right), \tag{5.5}$$

where $e_x = x_k - \hat{x}_k$ and $Q(\cdot)$ is the tail probability of the standard normal distribution [56, Chapter 33]. When LR-aided linear equalisation is used instead, the resulting expression for the PEP is identical to (5.5), except for the fact that the diagonal elements are now $\tilde{C}_{i,i}$, with $\tilde{\mathbf{C}} = (\tilde{\mathbf{H}}^H\tilde{\mathbf{H}})^{-1} = \mathbf{T}^{-1}\mathbf{C}(\mathbf{T}^{-1})^H$ [97]. Therefore, if the diagonal elements of $\tilde{\mathbf{C}}$ are reduced, the lattice-based algorithm improves the PEP performance, resulting in a lower BER value. Having this fact in mind, one can formulate an optimisation problem to find a unimodular matrix $\mathbf{T}$ that minimises the maximum diagonal element of $\tilde{\mathbf{C}}$ and which

[2]In a complex unimodular matrix $\mathbf{T}$, all entries of $\mathbf{T}$ and $\mathbf{T}^{-1}$ are Gaussian integers, and the determinant of $\mathbf{T}$ is either ± 1 or $\pm j$.

can be expressed as follows [68]:

$$\begin{aligned} \underset{\mathbf{T}}{\text{minimise}} \quad & \max_i(\tilde{C}_{i,i}), \\ \text{subject to} \quad & \tilde{\mathbf{C}} = (\tilde{\mathbf{H}}^H\tilde{\mathbf{H}})^{-1} = \mathbf{T}^{-1}\mathbf{C}(\mathbf{T}^{-1})^H; \\ & \mathbf{T} \in GL_{N_T}(\mathbb{Z}[j]), \end{aligned} \tag{5.6}$$

where $GL_{N_T}(\mathbb{Z}[j])$ is the group of all $N_T \times N_T$ unimodular matrices. Bear in mind that as the problem is formulated, only the longest basis vector in the dual basis is minimised and its optimal solution may not be unique. Consequently, the performance can be further improved by minimising the second largest diagonal, the third one, and so on. The procedure stops when all diagonal elements have been minimised. However, this process to find the global optimum solution is computationally demanding and, thus, not feasible for large MIMO systems. Instead, an iterative method to solve the problem was developed and proposed in [68] and will be briefly summarised below.

The objective of the ELR algorithm is to find a local optimal solution to (5.6). For that purpose, the fact that $\mathbf{T}' = (\mathbf{T}^{-1})^H$ is also a unimodular matrix is exploited and as a result it can also be expressed as the product of a series of column-addition operation matrices [69]. At each iteration, the algorithm tries to reduce the diagonal elements of $\tilde{\mathbf{C}}$ and the amount of reduction in $\tilde{C}_{k,k}$ is given by (for proof, see [68]):

$$\Delta_{i,k} = -|\lambda_{i,k}|^2\tilde{C}_{i,i} - \lambda^{\star}_{i,k}\tilde{C}_{i,k} - \lambda_{i,k}\tilde{C}_{k,i} \geq 0, \tag{5.7}$$

with $\lambda_{i,k}$ computed as:

$$\lambda_{i,k} = -\left\lceil \frac{\tilde{C}_{i,k}}{\tilde{C}_{i,i}} \right\rfloor, \tag{5.8}$$

where k is the largest reducible diagonal element of $\tilde{\mathbf{C}}$, the index i is chosen as $\arg\max_{\tilde{i}=1,\tilde{i}\neq k}^{N_T} \Delta_{\tilde{i},k}$ and the operator $\lceil\cdot\rfloor$ represents the rounding function which rounds the real and imaginary parts to the closest integer values. Moreover, a diagonal element $\tilde{C}_{k,k}$ is reducible if and only if there exists $i \neq k$, such that $\lambda_{i,k} \neq 0$ ($\Delta_{i,k} \geq 0$). If there is a reducible element, then the algorithm does the following updates:

$$\begin{aligned} \mathbf{t}'_k &\leftarrow \mathbf{t}'_k + \lambda_{i,k}\mathbf{t}'_i; \\ \tilde{\mathbf{c}}_k &\leftarrow \tilde{\mathbf{c}}_k + \lambda_{i,k}\tilde{\mathbf{c}}_i; \\ \tilde{\mathbf{c}}^{(k)} &\leftarrow \tilde{\mathbf{c}}^{(k)} + \lambda^{\star}_{i,k}\tilde{\mathbf{c}}^{(i)}, \end{aligned} \tag{5.9}$$

and proceeds to the next iteration with the updated $\tilde{\mathbf{C}}$; otherwise, the algorithm terminates. In (5.9), $\mathbf{t}'_k$ denotes the kth column of $\mathbf{T}'$, $\tilde{\mathbf{c}}_k$ represents the kth column of $\tilde{\mathbf{C}}$ and $\tilde{\mathbf{c}}^{(k)}$ is the kth row of $\tilde{\mathbf{C}}$. The pseudocode for the described operations can be found in algorithm 3. In terms of complexity, algorithm 3 requires $\mathcal{O}(N_T^2)$ in order to find the index pair (i, k). Further, for each basis update, $(31N_T - 7)$ arithmetic operations are required and the number of necessary repetitions of the D-ELR algorithm is lower than when compared with SA or LLL methods [68]. Therefore, computation-wise D-ELR algorithms

Algorithm 3 D-ELR-SLB method

Input: $\mathbf{H}$
$\tilde{\mathbf{C}} = (\mathbf{H}^H\mathbf{H})^{-1}, \mathbf{T}' = \mathbf{I}_{N_T}$
while true **do**
  $\lambda_{i,k} \leftarrow -\left\lceil \frac{\tilde{C}_{i,k}}{\tilde{C}_{i,i}} \right\rfloor, \forall i \neq k$
  if (all $\lambda_{i,k} = 0, \forall i \neq k$) **then**
    Output: $\mathbf{T} = (\mathbf{T}'^{-1})^H, \tilde{\mathbf{H}} = \mathbf{HT}$
  else
    Find the largest reducible $\tilde{C}_{k,k}$
    Choose $i = \arg\max_{\tilde{i}=1,\tilde{i}\neq k}^{N_R} \Delta_{\tilde{i},k}$
    $\mathbf{t}'_k \leftarrow \mathbf{t}'_k + \lambda_{i,k}\mathbf{t}_i$
    $\tilde{\mathbf{c}}_k \leftarrow \tilde{\mathbf{c}}_k + \lambda_{i,k}\tilde{\mathbf{c}}_i$
    $\tilde{\mathbf{c}}^{(k)} \leftarrow \tilde{\mathbf{c}}^{(k)} + \lambda^{\star}_{i,k}\tilde{\mathbf{c}}^{(i)}$
  end if
end while

are better than conventional LR techniques. Finally and unlike other detection methods (randomised algorithms, for instance), LR-aided detectors complexity is almost independent from the SNR, used constellation, or channel realisations. The effect in BER performance of the herein reviewed procedure will be studied next.

Once the reduced basis $\tilde{\mathbf{H}}$ is obtained and upon doing the proper scaling and shifting, one can use the conventional receivers to decode the transmitted symbols [54]. Moreover, successive interference cancellation (SIC) techniques [98] can be used on top of LR-aided methods to further improve performance. It is worth noting, however, that by ranking the diagonal elements $\tilde{C}_{k,k}$ in the ELR procedure, a SIC-like detection order is directly obtained. Even though the sorted variance (SV) order provided by the ELR method is different from the conventional V-BLAST one [99], it has been demonstrated in [68] that performance degradation is close to negligible. In this manner, SIC sorting procedures costing $\mathcal{O}(N_T^3)$, such as QR decomposition [100], can be avoided.

In order to get the best performance out of the D-ELR algorithms, both MMSE and SV-SIC-MMSE procedures will be considered; thus, the studied receivers will henceforth be denoted by D-ELR-SLB-MMSE and D-ELR-SLB-SV-SIC-MMSE, respectively.

For comparison purposes, the best performing conventional receiver MMSE-OSIC will be confronted with both detectors mentioned above. For a large MIMO configuration ($N_T = N_R = 16$) and a 16-QAM constellation setup, the output is expressed in figure 5.1. As can be seen, the achieved diversity by both ELR algorithms is about the same and higher than the conventional MMSE-OSIC, whose diversity does not surpass 1 (as predicted in theory). Nevertheless, when no SIC scheme is used in the ELR method, MMSE-OSIC outperforms the latter in the low-SNR regime. The results here depicted for the studied detectors show that it is possible to obtain considerable performance gain by equipping additional antennas at the base station and with acceptable complexity orders.

Finally, and in order to demonstrate the behaviour of both algorithms for the massive MIMO setup

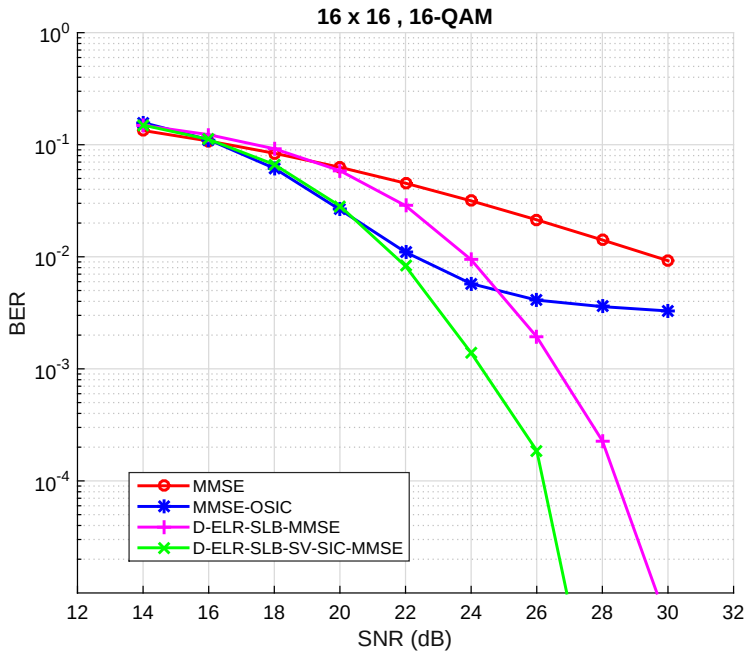


Figure 5.1: BER performance of element-based lattice reduction algorithms for $N_T = N_R = 16$ and a 16-QAM setup.

($N_T = N_R = 64$) and with a large constellation (256-QAM) configuration, a simulation was run. Its output is shown in figure 5.2. Once again, it can be inferred that a high diversity gain is achieved, a clear indication of the suitability of these low complexity detectors for massive MIMO systems.

Even though the results herein presented are as predicted in theory (i.e., LR methods attain maximum diversity), there is no indication of how far these detectors are from the optimal performance (ML). Hence, a different approach to solve the problem in (5.4) will be presented next and the ELR methods will serve a standard for comparison.

5.3 Markov Chain Monte Carlo

The use of advanced statistical methods may lead to significant gains in signal processing for wireless communications. In particular, Monte Carlo Bayesian methodologies can be employed to find a close to optimal solution in various signal reception problems, hence closing the gap in performance between state-of-the-art methods and the communication theory predicted optimal [101, Chapter 8].

Markov chain Monte Carlo (MCMC) methods refer to a class of algorithms that allow one to draw random samples from an arbitrary target probability density $p(x)$. This sampling process may be performed using a variety of methods, namely, importance sampling, rejection sampling, Metropolis–Hastings method, Gibbs sampling and slice sampling [102]. Gibbs sampling (GS), which will be considered in this work, was for the first time described in the seminal paper [103] on image processing.

Let $\mathbf{x}^{(t)}$ denote a random variable at time t and $S = \{s_j : j = 1, 2, ..., T_S\}$ to be the set of possible values that $\mathbf{x}$ can take, also called space state. A Markov chain is a sequence of random variables

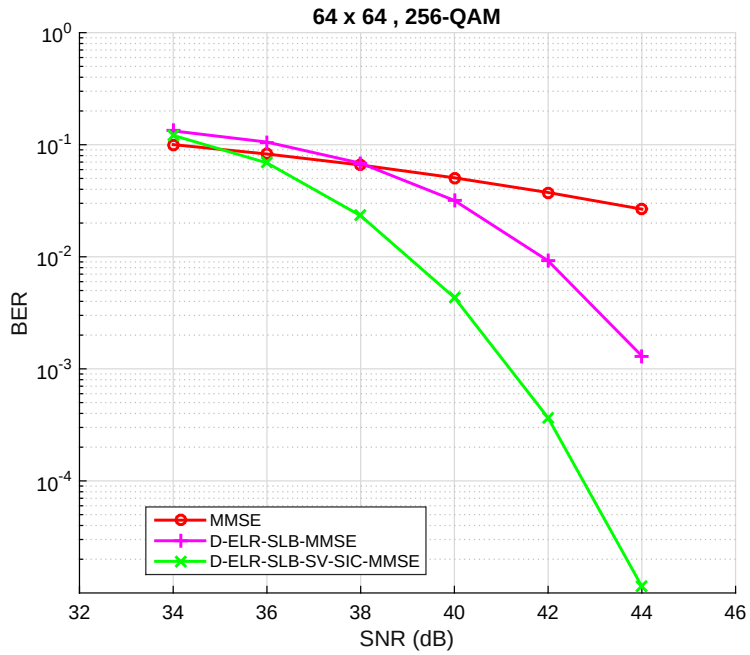


Figure 5.2: BER performance of element-based lattice reduction algorithms for $N_T = N_R = 64$ and a 256-QAM setup.

$\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, ...\mathbf{x}^{(\infty)}$, with the Markov property,

$$p(\mathbf{x}^{(t+1)} = s_j | \mathbf{x}^{(t)} = s_i, \mathbf{x}^{(t-1)} = s_i^{(t-1)}, ...) = p(\mathbf{x}^{(t+1)} = s_j | \mathbf{x}^{(t)} = s_i), \tag{5.10}$$

for every $s_j, s_i, s_i^{(t-1)}, ... \in S$ and any $t \geq 0$. In other words, the future state only depends on the present state and not on the entire history of the system. Further, a Markov chain is called reversible if the balance property is satisfied:

$$p(\mathbf{x}^{(t+1)} = s_j | \mathbf{x}^{(t)} = s_i) = p(\mathbf{x}^{(t+1)} = s_i | \mathbf{x}^{(t)} = s_j), s_i, s_j \in S. \tag{5.11}$$

This is equivalent to stating that the transition probability between s_j and s_i and between s_i and s_j are the same. The reversibility condition is necessary to ensure that the probability distribution of the chain converges to the desired one as $t \to \infty$ [102].

In the context of MIMO detection, state spaces are formed by the various possible combinations of $\mathbf{x}^{(t)}$, where each entry is taken from the alphabet $\mathcal{B}$. The desired distribution is such that ML cost is minimum; further, it is desirable that average performance improves with each iteration. A better description of the Gibbs sampling method will be provided next.

5.3.1 Conventional Gibbs Sampling

A reversible Markov chain is assumed, so that asymptotic convergence to the optimal solution is guaranteed. Thus, if the detector is run for a sufficiently long time, there is a certain positive probability that the optimal solution is visited. Both analysis and implementation will be done using the completely

equivalent real-valued MIMO channel [66, Equation 3.45]:

$$\mathbf{H}_r = \begin{bmatrix} \Re(\mathbf{H}) & -\Im(\mathbf{H}) \\ \Im(\mathbf{H}) & \Re(\mathbf{H}) \end{bmatrix}; \mathbf{y}_r = \begin{bmatrix} \Re(\mathbf{y}) \\ \Im(\mathbf{y}) \end{bmatrix}; \mathbf{x}_r = \begin{bmatrix} \Re(\mathbf{x}) \\ \Im(\mathbf{x}) \end{bmatrix}; \mathbf{n}_r = \begin{bmatrix} \Re(\mathbf{n}) \\ \Im(\mathbf{n}) \end{bmatrix}, \tag{5.12}$$

where $\mathbf{y}_r, \mathbf{n}_r \in \mathbb{R}^{2N_R \times 1}$, $\mathbf{x}_r \in \mathbb{R}^{2N_T \times 1}$ and $\mathbf{H}_r \in \mathbb{R}^{2N_R \times 2N_T}$. For notation simplification purposes, $n = 2N_T = 2N_R$ will be used and the subscript r will be dropped.

At the initial time instant $t = 0$, an n-dimensional starting candidate $\mathbf{x}^{(0)}$ (which can be randomly chosen or the solution provided by any detector; e.g., MMSE or lattice reduction method) is given to the MCMC detector. Then, the MCMC detector performs an "ingenious" random walk over the alphabet $\mathcal{A}^n$ based on the to be constructed transition rule.

Under the assumptions that the various transmitted symbols are uncorrelated and the noise is AWGN with variance σ_n^2, the integer least squares problem in (5.4) can also be regarded as a closest vector problem (CVP) in a lattice Gaussian distribution. In particular, the n-dimensional lattice formed by $\overline{\mathbf{H}} = \sqrt{\frac{\text{SNR}}{N_T}}\mathbf{H}$ (known as the lattice basis) is $\Lambda = \{\overline{\mathbf{H}}\mathbf{x} : \mathbf{x} \in \mathbb{Z}^n\}$. Then, one can define the Gaussian function centred at $\mathbf{y} \in \mathbb{R}^n$ for standard deviation $\sigma_n^2 = 1 > 0$ as

$$\rho(\overline{\mathbf{H}}\mathbf{x}) = \exp\bigg(- \frac{||\mathbf{y} - \overline{\mathbf{H}}\mathbf{x}||^2}{2}\bigg), \tag{5.13}$$

for all $\overline{\mathbf{H}}\mathbf{x} \in \mathbb{R}^n$. Hence, the previous expression suggests that the joint probability of interest satisfies

$$p(x_1, x_2, ..., x_n|\overline{\mathbf{H}}, \mathbf{y}) \ \propto \ \exp\bigg(- ||\mathbf{y} - \overline{\mathbf{H}}\mathbf{x}||^2\bigg). \tag{5.14}$$

From the above expression, the transition rule between consecutive states can be computed. Particularly, it can be regarded as a discrete Gaussian distribution over Λ; hence, sampling techniques such as Gibbs algorithm are applicable. Assume that at time index t, the current state is given by $\mathbf{x}^{(t)} \in \mathcal{A}^n$. Then, at the next time index, the method uniformly picks one random position j out of $\{1, ..., n\}$ and computes the conditional probability of transitioning to each of the possible constellation points. With the remaining $(n-1)$ positions fixed, the jth index is updated according to the following stationary distribution [104]:

$$p(\mathbf{x}_j^{(t+1)} = \omega|\theta) = \frac{\exp\bigg(- \frac{1}{2\alpha^2}||\mathbf{y} - \overline{\mathbf{H}}\mathbf{x}_{j|\omega}||^2\bigg)}{\sum_{\mathbf{x}_{j|\tilde{\omega}} \in \mathcal{A}} \exp\bigg(- \frac{1}{2\alpha^2}||\mathbf{y} - \overline{\mathbf{H}}\mathbf{x}_{j|\tilde{\omega}}||^2\bigg)}, \tag{5.15}$$

where $\mathbf{x}_{j|\omega}^T = [\mathbf{x}_{1:j-1}^{(t+1)}, \omega, \mathbf{x}_{j+1:n}^{(t)}]^T$, $\theta = \{\mathbf{x}^{(t)}, j, \mathbf{y}, \overline{\mathbf{H}}\}$ and α is a positive "temperature" parameter. So, conditioned on the jth position is chosen, the MCMC detector will with probability $p(\mathbf{x}_j^{(t+1)} = \omega|\theta)$ sample ω to position j of $\mathbf{x}^{(t+1)}$. An intuition of (5.15) suggests that the closer lattice point $\overline{\mathbf{H}}\mathbf{x}$ is to $\mathbf{y}$, the higher the probability that it is sampled. The aforementioned process is denoted in the literature by

Gibbs sampling, where one iteration involves sampling one entry at a time in the following manner [28]:

$$\begin{aligned} x_1^{(t+1)} &\sim p(x_1|x_2^{(t)},...,x_n^{(t)},\mathbf{y},\overline{\mathbf{H}}); \\ x_2^{(t+1)} &\sim p(x_2|x_1^{(t+1)},x_3^{(t)},...,x_n^{(t)},\mathbf{y},\overline{\mathbf{H}}); \\ &\vdots \\ x_n^{(t+1)} &\sim p(x_n|x_1^{(t+1)},x_2^{(t+1)},...,x_{n-1}^{(t+1)},\mathbf{y},\overline{\mathbf{H}}). \end{aligned} \tag{5.16}$$

At the end of each iteration, the ML cost in (5.4) is computed and the resulting vector is fed into the next iteration. The algorithm stops after a certain number of iterations and the algorithms' output is the vector with the best ML cost. The pseudocode related with the aforementioned MCMC detector is presented in algorithm 4.

Algorithm 4 Reversible MCMC Detector Based on Gibbs Sampling

Input: $\mathbf{y}, \mathbf{H}$, initial vector $\mathbf{x}^{(0)}$, number of iterations $t_{\max}$
Denote the decision vector by $\mathbf{z}$
Denote the ML cost function as $f(\cdot)$
for $t = 1$ to t_{max} **do**
  Pick a position index $j \in \{1, 2, ..., n\}$ from a uniform distribution
  Fix the $(n-1)$ symbols of $\mathbf{x}^{(t)}$, and transition the jth symbol of $\mathbf{x}^{(t)}$ to ω according to (5.15)
  Denote the new vector by $\mathbf{x}^{(t+1)}$
  if $f(\mathbf{x}^{(t+1)}) < f(\mathbf{z})$ **then**
    Update $\mathbf{z} = \mathbf{x}^{(t+1)}$
  end if
end for
Output: $\mathbf{z}$

Note that other randomised lattice Gaussian sampling exist in the literature, namely Klein and Gibbs-Klein algorithms [105, 106]. However, these will not be considered since the next studied method is proven to perform better [107].

5.3.2 Implementation and Complexity of Gibbs Sampling

For the sake of implementation (to limit complexity), a sequential rather than a reversible MCMC detector will be considered. The only difference resides in the way the updated position index j is chosen. In sequential detection, each iteration consists of a block iteration, where the n indices are updated sequentially, hence the name.

Denoting the difference at iteration t as $\mathbf{d}^{(t)} = \mathbf{y} - \sqrt{\frac{\text{SNR}}{N_T}}\mathbf{H}\mathbf{x}^{(t)}$ and in order to further reduce complexity, one may notice that only the jth symbol is changed when (5.15) is computed [104]. Consequently, $\mathbf{d}^{(t)}$ can be expressed as:

$$\mathbf{d}^{(t)} = \mathbf{d}^{(t-1)} - \sqrt{\frac{\text{SNR}}{N_T}}\mathbf{h}_j \Delta x_{j|\omega}, \tag{5.17}$$

where $\Delta x_{j|\omega} = x_{j|\omega}^{(t)} - x_{j|\omega}^{(t-1)}$ and $\mathbf{h}_j$ is the jth column of $\mathbf{H}$. Therefore, the complexity of computing the norms in (5.15), when the jth symbol is changed, is $2n$ complex operations (n multiplications for each of the products $(\mathbf{d}^{(t)}))^T\mathbf{d}^{(t)}$ and $\mathbf{h}_j\Delta x_{j|\omega}$). Further and assuming that the number of required iterations will be large (processing for massive MIMO systems), the complexity of an MCMC detector can be reduced even further. This can be achieved by means of a QL factorisation of the channel matrix, $\mathbf{H} = \mathbf{QL}$ and by rewriting (5.4) as [108]

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{A}^n} ||\tilde{\mathbf{y}} - \sqrt{\frac{\text{SNR}}{N_T}}\mathbf{L}\mathbf{x}||^2, \tag{5.18}$$

where $\tilde{\mathbf{y}} = \mathbf{Q}^T\mathbf{y}$. Exploiting the fact that $\mathbf{L}$ is a lower triangular matrix, the first difference $\mathbf{d}^{(0)} = \tilde{\mathbf{y}} - \mathbf{L}\mathbf{x}$ costs $n + 2\sum_{i=1}^{n} i = n(n+2)$ multiplication and addition operations (which is lower than the conventional model $2n^2$). Moreover, at each iteration the difference $\mathbf{d}^{(t-1)} - \sqrt{\frac{\text{SNR}}{N_T}}\mathbf{l}_j\Delta x_{j|\omega}$ will require $2(\frac{n+1}{2})(|\mathcal{A}| - 1)$ arithmetic operations. If the conventional model was used, the corresponding complexity would be $2n(|\mathcal{A}|-1)$. Then, and taking into account that computing $\tilde{\mathbf{y}}$ yields $\mathcal{O}(\frac{2}{3}n^3+2n^2)$, there is a gain in complexity if

$$n(n+2) + t_{max}(n+1)(|\mathcal{A}| - 1) + \frac{2}{3}n^3 + 2n^2 < 2nt_{max}(|\mathcal{A}| - 1) + 2n^2, \tag{5.19}$$

which holds true when

$$t_{max} > \frac{\frac{2}{3}n^2 + n + 2}{(|\mathcal{A}| - 1)(n-1)}n \approx \frac{\frac{2}{3}n^2 + n}{|\mathcal{A}| - 1}. \tag{5.20}$$

Another possibility to keep complexity low, especially in higher constellations (64-QAM and higher), is to only sample one-symbol away neighbours, as suggested in [107].

5.3.3 Gibbs Sampling Performance

So far, the details, including complexity, of the conventional Gibbs sampling for MIMO detection were given. The performance of the method shall now be evaluated. To this purpose, set the maximum number of block iterations $t_{\text{max}} = 500$ and the "temperature" parameter equal to the noise variance, that is $\alpha^2 = 1$. For a symmetric system $N_T = N_R = 10$ and a 4-QAM constellation, the corresponding BER plot is shown in figure 5.3. For comparison purposes, performance of MMSE and D-ELR-SLB-SV-SIC-MMSE are also depicted. As can be inferred for low SNR, the performance is very close to optimal. However, in the medium to high SNR, the BER did not improve any further. This problem is referred in the literature as the "stalling problem" [109]. This effect can be explained as follows: consider a case where for a candidate vector $\mathbf{x}^{(t)}$, the norm $||\mathbf{y} - \overline{\mathbf{H}}\mathbf{x}^{(t)}||$ is significantly smaller than $||\mathbf{y} - \overline{\mathbf{H}}\mathbf{x}_{j|\tilde{\omega}}^{(t)}||$, where $\mathbf{x}_{j|\tilde{\omega}}^{(t)}$ is the same $\mathbf{x}^{(t)}$ with jth position changed with $\tilde{\omega}$. In this situation and, especially when SNR is high, one may find that after sampling $P(\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)}|\mathbf{y}, \mathbf{H}) \approx 1$ and $P(\mathbf{x}^{(t+1)} = \mathbf{x}_{j|\tilde{\omega}}^{(t)}|\mathbf{y}, \mathbf{H}) \approx 0$ [109]. Thus, the conventional Gibbs sampler may remain stuck in the same state for a very long time and that effect consequence is depicted in figure 5.3, as the number of iterations was fixed to $t_{\text{max}} = 500$. All

in all and even though the chain is guaranteed to converge asymptotically when $t \to \infty$ [102], stalling may occur. A possible solution to alleviate this problem is studied next.

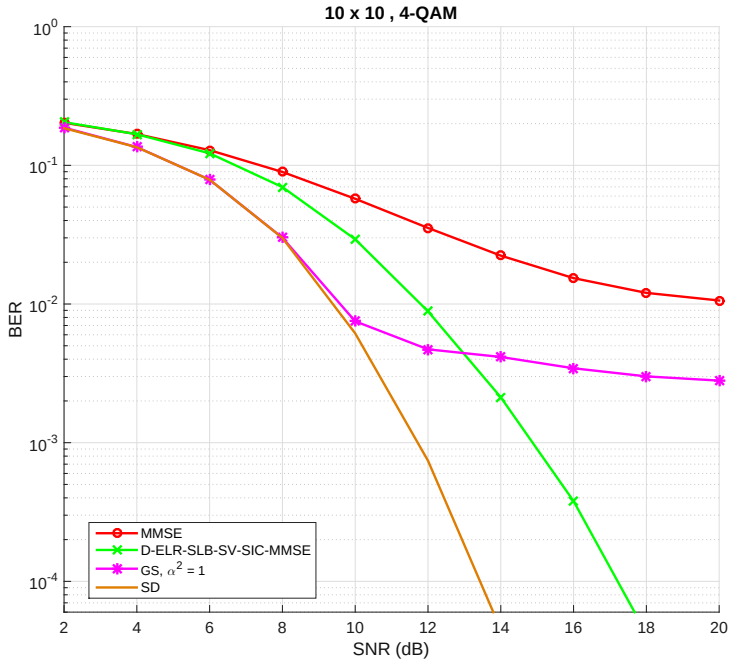


Figure 5.3: BER performance for conventional Gibbs sampling detection algorithm for $N_R = N_T = 10$ and a 4-QAM constellation.

5.3.4 Mixed Gibbs Sampling with Multiple Restarts

As stated previously, stalling occurs because the ML cost of the state vector may be trapped in a local minima for many iterations [107]. That is in part due to the norm inside the exponential in (5.14), which may cause disparate sampling probabilities. One possible approach is to try and optimise the temperature parameter α such that the probability of encountering the optimal solution is as high as possible. This evaluation was done in [104] and its performance will be shown in a subsequent section. In this part and following what was done in [107], a mixed Gibbs sampling will be studied. This consists of a slight change in the update rule in (5.16) . Instead of drawing samples from (5.15) with probability one, there is a chance $q_\infty = (1 - q_1)$ that the values are instead taken from a uniform distribution as

$$p(\mathbf{x}_j^{(t+1)} = \omega|\theta) \sim U[0, 1], \tag{5.21}$$

such that $\sum p(\mathbf{x}_j^{(t+1)} = \omega|\theta) = 1$. This corresponds to setting $\alpha = \infty$ in (5.15). Hence, the method acquires the name of mixed Gibbs sampling (MGS), and as long as q_∞ is chosen accordingly, the stalling problem is alleviated: if stuck in a local minima there is a non-zero probability q_∞ that the detector performs a random walk to move away from the local solution. Unless otherwise stated, the mixing ratio value is set to $q_\infty = \frac{1}{n}$ (as suggested in [107]), which is equivalent to performing one random walk per

each block iteration on average.

To further improve performance, a multiple restart (MR) technique, as proposed in [107] will also be contemplated. Due to the random nature of MGS, different instances of the algorithm may lead to different outputs, even when the inputs values are the same. Denote the maximum number of allowed restarts by R. When $R > 1$, the output solution is the one yielding the lowest ML cost and since each of the restarts is independent of the others, a parallel computation of the various instances can be performed to decrease overall detection time.

With the introduction of MR, the algorithm has now three distinct loop levels: R restarts, t block iterations and n sampling processes per block iteration. Thus, the method can become computationally demanding. However, when a close enough solution to optimal has already been found, additional iterations may be redundant. Hence, in order to limit complexity, suitable stopping criterion can be defined.

If the ML cost remains unchanged for two consecutive block iterations, stalling is said to have occurred [107]. Then, stalling mode is entered, where only a maximum number of iterations θ_s more is allowed. If during stalling mode, a vector $\mathbf{x}$ with lower ML cost is found, then stalling mode is left; else, the algorithm terminates after θ_s iterations. The parameter θ_s is chosen large if the stalled ML cost is high and chosen small when the margin for improvement is narrow. If the algorithm does not stop as a result of the stalling mode, then the initially set maximum number of block iterations $t_{\max}$ are performed.

Under the considered model and when $\mathbf{x}$ is error free, the ML cost is $||\mathbf{n}||^2$, whose distribution is a chi-square with $2N_R$ degrees of freedom, mean $N_R\sigma_n^2$ and variance $N_R\sigma_n^4$ [56, Chapter 11]. Thus, a standardised ML cost of solution vector $\mathbf{x}$ can be characterised as [107]

$$\phi(\mathbf{x}) = \frac{||\mathbf{y} - \sqrt{\frac{\text{SNR}}{N_T}}\mathbf{H}\mathbf{x}||^2 - N_R\sigma_n^2}{\sqrt{N_R}\sigma_n^2}. \tag{5.22}$$

Metric (5.22) can be seen as a a difference between ML cost and the mean of $||\mathbf{n}||^2$ normalised by the standard deviation of $||\mathbf{n}||^2$ (note that under the channel model in (5.2), $\sigma_n^2 = 1$). In addition to the ML cost, the stopping criteria θ_s should also take into account the size of the QAM constellation. Accordingly, [107] proposed the following:

$$\theta_s(\mathbf{x}) = \lceil \max(c_{\min}, 10\log_2 M \exp(\phi(\mathbf{x}))) \rceil, \tag{5.23}$$

where c_{min} is the minimum number of iterations following a stalling event.

Likewise, a stopping criteria for the maximum number of repetitions R can be defined. In a similar fashion to the previous case (see [107] for further insight), the algorithm stops if the number of restarts performed so far is greater than the following integer threshold:

$$P = \lfloor \max(0, 0.5\log_2 M\phi(\mathbf{x})) \rfloor + 1. \tag{5.24}$$

The usage of both MGS and the stopping criteria in (5.23) and (5.24) will provide close to optimal solutions, while yielding limited complexity, as will be verified in the next section.

5.3.5 Numerical Results

The performance of the described mixed Gibbs sampling with restarts will be evaluated here. Setting the number of antennas $N = 10$, maximum number of restarts $R = 15$, maximum number of iterations $t_{\max} = 160$ and $c_{\min} = 10$, the results are depicted in figure 5.4. The initial vector for the first restart was the output of the MMSE filter and for the remaining restarts a random initial vector was used. As can be inferred, the stalling problem was efficiently fixed with the use of MGS-MR. Further and as was expected, the attained diversity by MGS-MR is similar to the one obtained by the lattice reduction-aided method. Moreover, the gain in BER of MGS-MR over D-ELR-SLB-SV-SIC-MMSE is not negligible. In particular, the randomised algorithm requires around less 3 dB than the studied LR algorithm to achieve a BER of 10^{-3}.

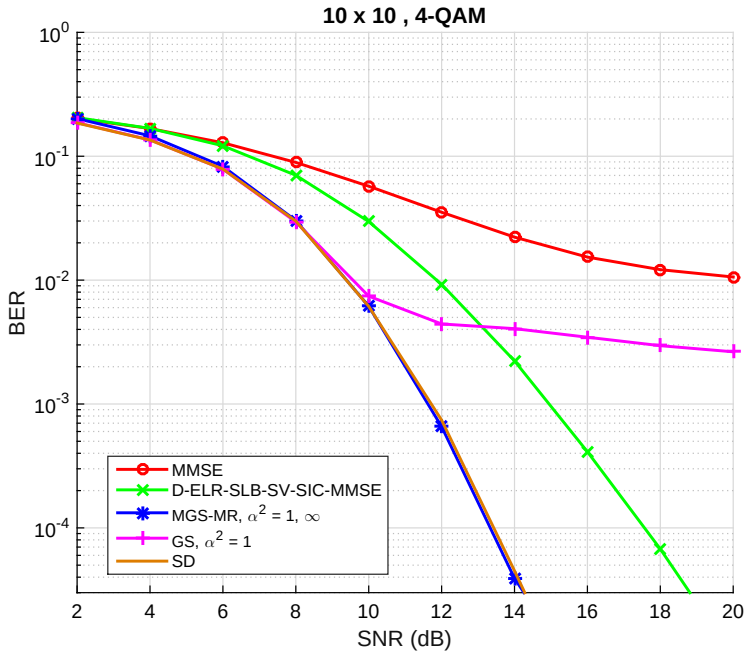


Figure 5.4: BER performance for both conventional Gibbs sampling and mixed Gibbs sampling with restarts detection algorithms for $N_R = N_T = 10$ and a 4-QAM constellation.

Large MIMO detection with MGS-MR shall now be evaluated. Increasing dimensions to $N_T = N_R = 16$, using a 16-QAM constellation and setting $t_{\max} = 512$, the results are shown in figure 5.5. As was expected, in the low SNR regime, the performance is near ML and better than the one achieved by LR method. However, for SNR $\geq$ 20 dB, the MCMC detector suffers a degradation in performance. This phenomenon is related with the rather low convergence rate (mixing time) of the underlying Markov chains, since $\alpha^2 = 1$. Particularly, this choice for α will cause the Markov chain to take a long time to reach its stationary distribution [104]. To solve this problem, three different approaches are possible:

1. Try and optimise the mixing ratio q_∞ as a function of the SNR.

2. Find the dependence of α with the value of the SNR.

3. Increase the maximum number of iterations $t_{\max}$ and use different, better initial vectors $\mathbf{x}^{(0)}$, at the cost of increased complexity.

Option 2 from the above will be evaluated next.

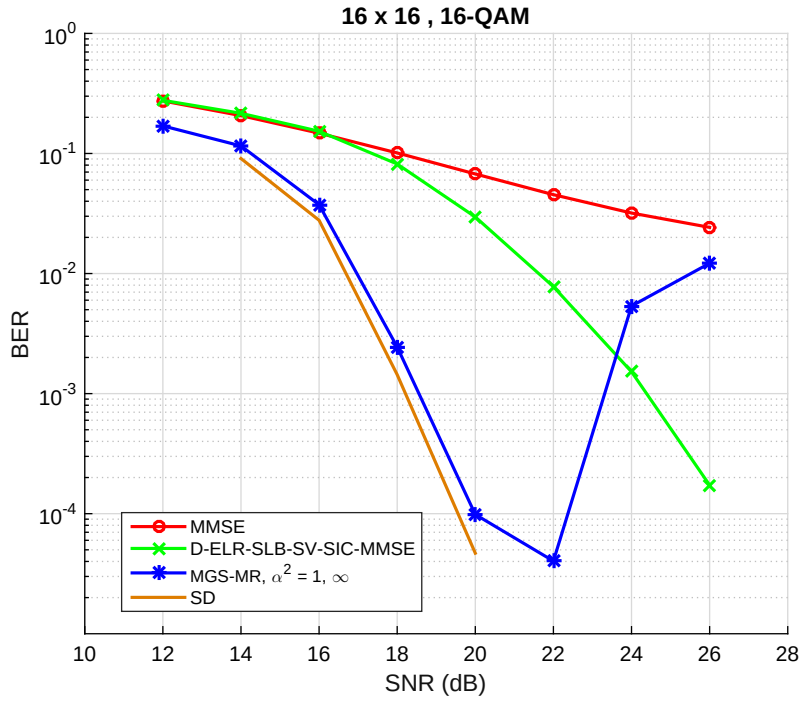


Figure 5.5: BER performance for both conventional Gibbs sampling and mixed Gibbs sampling with multiple restarts detection algorithms for $N_R = N_T = 16$ and 16-QAM constellation.

5.3.6 Gibbs Sampling with Optimised "Temperature" Parameter

The "temperature" parameter α controls the mixing time of the underlying Markov chain and its choice should be such that the probability of encountering the optimal solution is maximum [104]. On the one hand, if α is chosen too small, convergence rate becomes low, which is undesirable. On the other hand, with a large α the chances of finding the optimal solution in stationary distribution become exponentially small. Hence, there is a trade-off in the choice for α. The MGS-MR method tried to circumvent this difficulty by using an alternating sampling method between $\alpha = 1$ and $\alpha = \infty$. However, as seen previously, this choice did not prove to be effective in the high SNR regime.

The mixing time of MCMC detector is closely related to the existence of local minima and the average number of these increases with the number of antennas. Hassibi *et al.* [104] verified this fact and derived an optimal α^2 for the case when a $\{-1, 1\}$ constellation set is used. In this work, the result will be used for larger constellations without further proof. Thus, the proposed $\alpha^2_{\text{opt.}}$ in [104] will be denoted

here as the "optimal temperature" parameter, being given by:

$$\alpha^2_{\text{opt.}} = \frac{\text{SNR}}{\log(n)} + \sqrt{\left(\frac{\text{SNR}}{\log(n)}\right)^2 - 2\frac{\text{SNR}}{\log(n)}}, \tag{5.25}$$

only valid for SNR $> 2\log n$, as one is interested in. The result in (5.25) suggests that $\alpha_{\text{opt.}}$ scales with $\mathcal{O}(\sqrt{\text{SNR}})$, for fixed n, which means that faster mixing times are required in the high SNR regime.

In order to demonstrate that MCMC methods employing Gibbs sampling converge to the optimal solution and the importance of the parameter α, a study of BER evolution in terms of the number of iterations shall now be performed. For that purpose, SNR was fixed and a set of values for α^2 was chosen, including $\alpha^2 = 1$ and $\alpha^2 = \alpha^2_{\text{opt.}}$. Using the MMSE filter output as the initial vector, figures 5.6 and 5.7 show the change in BER with an increasing number of iterations. For comparison purposes, conventional detectors are also depicted. The first thing to be noticed in figure 5.6 is the fast convergence of the GS method to the ML performance when $\alpha^2 =$ "optimal" (blue line). Moreover, it is interesting to observe the rapid BER decay in the first few iterations for when $\alpha^2 = 1$ and then the stalling effect, as had been previously verified. Higher choices of α^2 (namely, $\alpha^2 = 9$ and $\alpha^2 = 18$) proved not to improve the MMSE initial solution any further.

Denote the detector using $\alpha^2 =$ "optimal", mixing ratio $q_\infty = 0$ and multiple restarts, as optimised Gibbs sampling with multiple restarts (OGS-MR). The results corresponding to OGS-MR are depicted in figure 5.8. For comparison purposes, the remaining used parameters are identical to the ones used to obtain figure 5.4. As can be seen and as expected, the problem in the high SNR regime was partially mitigated, confirming that choosing an α different from 1 is indeed beneficent. Nonetheless, the stalling problem is present once again and that is due to the existence of relative minimums, as explained in the comments regarding figure 5.3.

Having in mind the results obtained so far, a combination of both MGS and OGS methods will be proposed next. As will be seen, the suggested variant will close the gap in performance for large MIMO detection and for high SNR values.

5.3.7 Triple Mixed Gibbs Sampling

Thus far, it has been verified that a good choice of the temperature "parameter" α is preponderant in the performance of randomised algorithms based on Gibbs sampling. The choice of α may be seen as a trade-off between reliability and convergence speed; hence, a poor choice for α may lead to rather slow progress in high dimensions [102]. To circumvent the problem, two different approaches have been characterised and evaluated; though neither of them performed as one would expect, especially for high SNR.

Based on the results obtained in the previous sections and corresponding justifications, a triple mixed Gibbs sampling (T-MGS) algorithm is going to be proposed here. The idea is to, at each iteration, pick a value for α based on the set $\{1, \text{"optimal"}, \infty\}$ with a given probability distribution. At a given time instant, choosing $\alpha = 1$ ensures convergence to optimal solution, $\alpha^2 =$"optimal" assures a quick mix

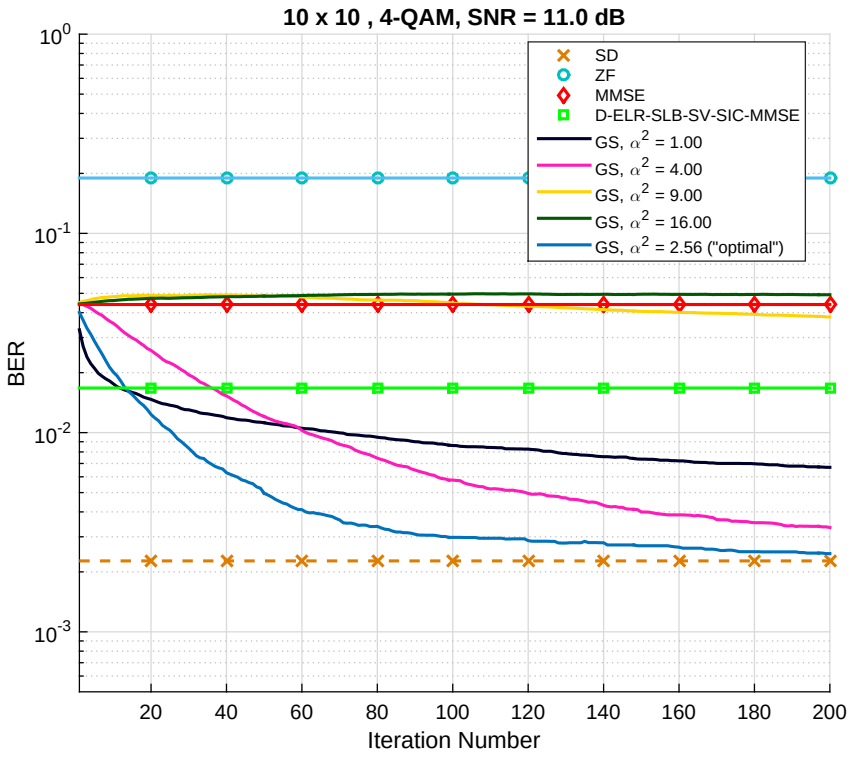


Figure 5.6: BER performance for conventional Gibbs sampling with different "temperature" parameters. $N_R = N_T = 10$ and a 4-QAM constellation were used.

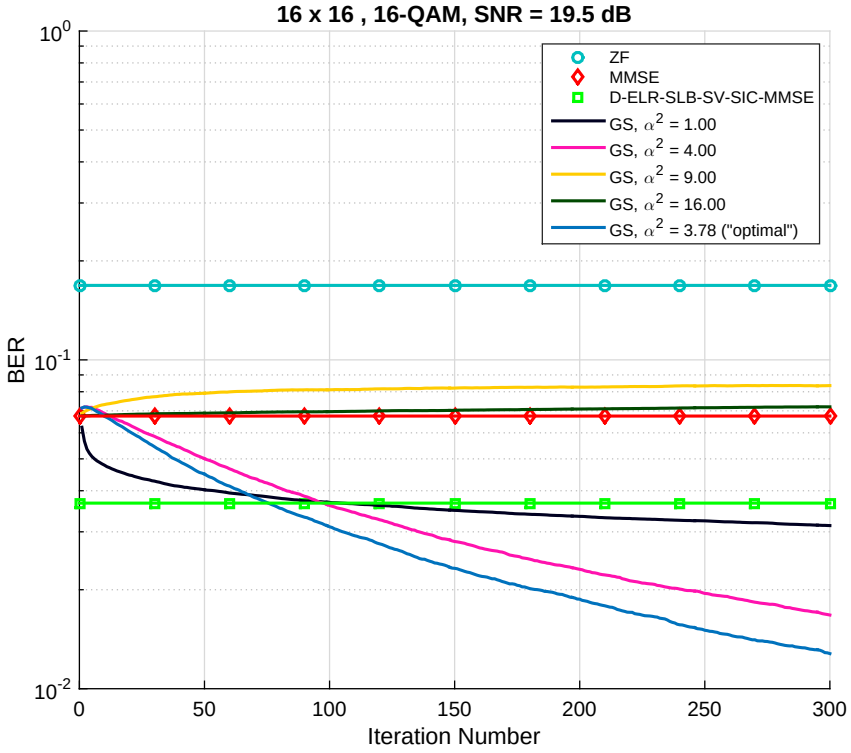


Figure 5.7: BER performance for conventional Gibbs sampling with different "temperature" parameters. $N_R = N_T = 16$ and a 16-QAM constellation were used.

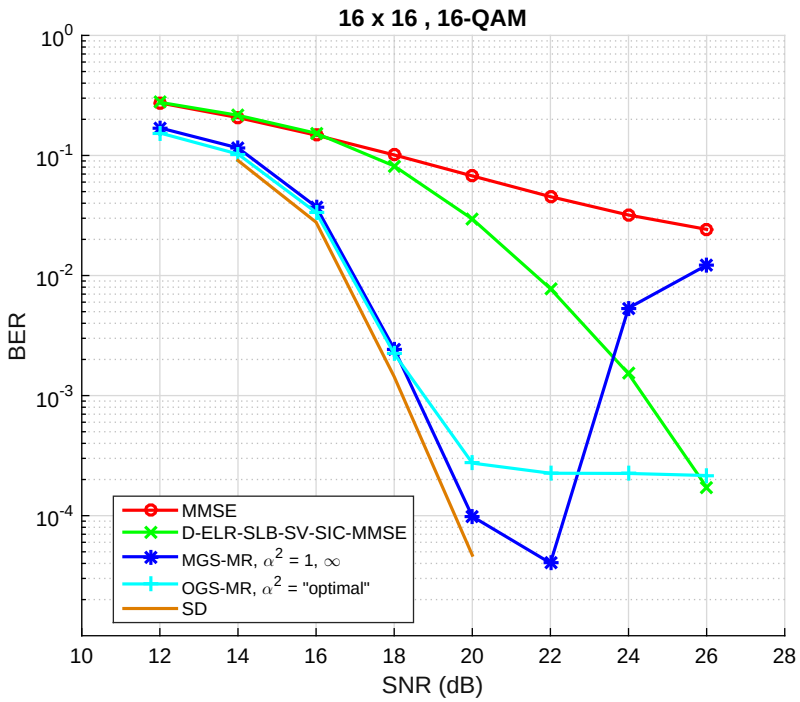


Figure 5.8: BER performance for optimised Gibbs sampling with multiple restarts detection algorithm for $N_R = N_T = 16$ and a 16-QAM constellation.

to steady state and $\alpha = \infty$ (pure random walk) is necessary to avoid local minima (which leads to the observed stalling effect). As will be verified, with a properly chosen mixing ratio one can solve the previously identified problems.

As a first approach, let q_∞ denote the probability that at each iteration a random walk is performed. If it is not executed, then the chances of picking $\alpha^2 = 1$ or $\alpha^2 =$ "optimal" are given by $(1 - q_\infty)q_1$ and $(1 - q_\infty)q_{\text{opt.}}$, respectively. When MGS-MR ($q_{\text{opt.}} = 0$) was discussed, a value of $q_\infty = 1/n$ was used [107]. However, with the addition of a third possible value for α^2, setting $q_\infty = 1/n$ is not a good choice any more as the number of stalling events is lower when $\alpha^2 > 1$.

In order to get a grasp of the optimal value for q_∞, consider $q_1 = q_{\text{opt.}} = 1/2$ (note that under the used model, the condition $q_1 + q_{\text{opt.}} = 1$ must be satisfied). Fixing the SNR at 18 and 19 dB, an evaluation of the BER performance in terms of the number of iterations, for different mixing ratios q_∞, is presented in figure 5.9. The number of antennas was $N = 16$, a constellation size of 16-QAM was used and the maximum number of iterations was set to $t_{max} = 800$. Only one restart was performed and the used initial vector was the MMSE solution. First thing to be inferred from figure 5.9 is that in both cases it is a non-zero mixing ratio ($q_\infty \neq 0$) that minimises the BER, hence proving the advantage of having a triple MGS algorithm. Additionally, the minimum value for different SNR values is attained for different values of q_∞, implying that the optimal value of q_∞ also has a dependence with the SNR. Nevertheless, the gain of such optimisation is almost negligible and should be mitigated when more than one restart is performed. Henceforth and otherwise stated, the value of $q_\infty = 1/(20N_T)$ will be adopted.

A step-by-step description of the proposed algorithm including multiple restarts can be found in

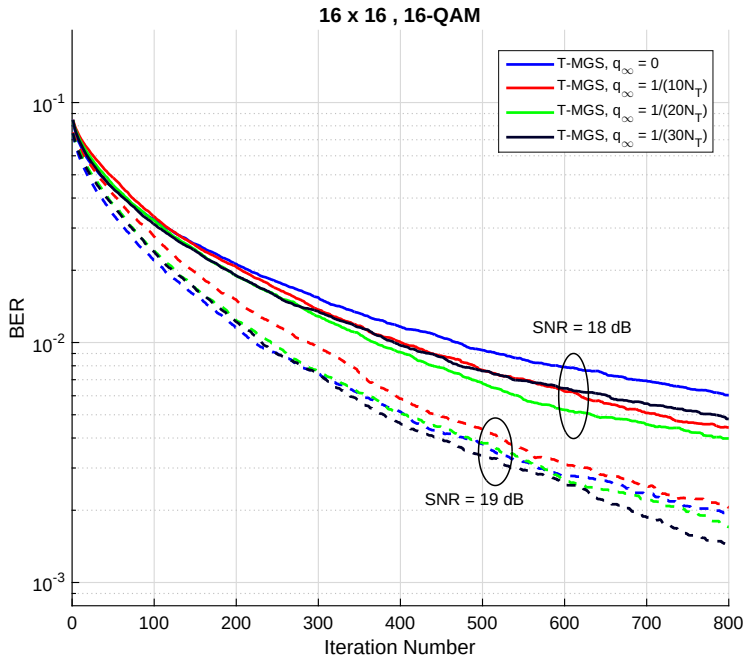


Figure 5.9: Study of the impact of average BER performance for different mixing ratios q_∞. SNR was fixed at 18 and 19 dB, the number of antennas set to $N_T = N_R = 16$ and the constellation size was 16-QAM.

algorithm 5, where pmf stands for probability mass function. It is worth mentioning that the previously mentioned MGS-MR corresponds to doing $q_\infty = 1/n$ and $q_1 = 1$, whilst OGS-MR coincides with setting $q_\infty = q_1 = 0$ in algorithm 5. The "temperature" parameters and corresponding mixing ratios are summarised in table 5.1.

Finally, a performance comparison for both convergence rates and BER values of the LR method and randomised algorithms is conducted. In order to demonstrate the superiority of the here proposed variant over the algorithms in [107] and [104], the average BER as a function of the iteration number was simulated and is depicted in figure 5.10. Number of antennas was set to $N = 16$, a 16-QAM constellation was used and the number of iterations was $t_{\max} = 1000$. Further, two SNR values were studied, namely 15 and 20 dB. For comparison purposes the sphere decoder performance (ML-like) is also shown. For each channel realisation, only one restart was performed and the MMSE filter output was used as the initial solution.

Mixing Ratio / Variant	q_∞ ($\alpha^2 = \infty$)	q_1 ($\alpha^2 = 1$)	$q_{\text{opt.}}$ ($\alpha^2 = \alpha^2_{\text{opt.}}$)
MGS	$1/2N_T$	1	0
OGS	0	0	1
T-MGS	$1/20N_T$	$1/2$	$1/2$

Table 5.1: "Temperature" parameters and corresponding mixing ratios used to compare the various GS methods, including the proposed one, for $N = 16$ and when a 16-QAM constellation size is used.

Algorithm 5 Triple Mixed Gibbs Sampling Algorithm with Multiple Restarts and Stopping Criteria

Input: $\mathbf{y}, \mathbf{H}, \mathbf{x}^{(0)}, t_{\max}, q_\infty, q_1, R$
Define:
$f(\mathbf{x}) \rightarrow$ ML cost function (eq. (5.4))
$\phi(\mathbf{x}) \rightarrow$ Normalised ML cost function (eq. (5.22))
$\theta_s(\mathbf{x}) \rightarrow$ Stalling limit function (eq. (5.23))
$p(\mathbf{x}, \alpha|\theta) \rightarrow$ Transition probability distribution (eq. (5.15))
$c_b \rightarrow$ Cost of best candidate so far; set initial c_b to ∞
Compute:
for $r = 1$ to R **do**
  Initial candidate: $\mathbf{z} = \mathbf{x}^{(0)}$; Initial cost: $\beta = f(\mathbf{z})$
  $t = 0$
  while $t < t_{\max}$ **do**
    for $i = 1$ to n **do**
      generate $r_1, r_2 \sim U[0, 1]$
      if $r_1 < q_\infty$ **then**
        generate pmf $p(x_i^{(t+1)} = \omega) \sim U[0, 1], \forall \omega \in \mathcal{A}$
        sample $x_i^{(t+1)}$ from this pmf
      else
        if $r_2 < q_1$ **then**
          $\alpha^2 = 1$
        else
          $\alpha^2 \propto$ SNR (according to eq. (5.25) and table 5.1)
        end if
        $x_i^{(t+1)} \sim p(x_i, \alpha | x_1^{(t+1)}, ..., x_{i-1}^{(t+1)}, x_{i+1}^{(t)}, ..., x_n^{(t)})$
      end if
    end for
    $\gamma = f(\mathbf{x}^{(t+1)})$
    if $\gamma \leq \beta$ **then**
      $\mathbf{z} = \mathbf{x}^{(t+1)};\ \beta = \gamma$
    end if
    $t = t + 1$; $\beta_v^{(t)} = \beta$
    if $\beta_v^{(t)} == \beta_v^{(t-1)}$ **then**
      compute $\theta_s(\mathbf{z})$
      if $\theta_s < t$ **then**
        if $\beta_v^{(t)} == \beta_v^{(t-\theta_s)}$ **then**
          break while
        end if
      end if
    end if
  end while
  if $f(\mathbf{z}) < c_b$ **then**
    $\mathbf{z}_b = \mathbf{z};\ c_b = f(\mathbf{z}_b)$
  end if
  compute $P = \lfloor \max(0, 0.5 \log_2 M\phi(\mathbf{z}_b)) \rfloor + 1$
  if $P < r$ **then**
    break for
  end if
end for
Output: $\mathbf{z}_b$

For 15 dB it can be verified in figure 5.10 that all of the algorithms converge to the SD solution; though both OGS and the proposed variants require less iterations in order to achieve it. This faster convergence rate becomes even more preponderant in the high SNR regime (20 dB) where the proposed T-MGS beats both MGS and OGS alternatives. It is worth emphasising that for 20 dB average BER performance is not optimal (SD-like); hence showing the importance of the multiple restarts feature.

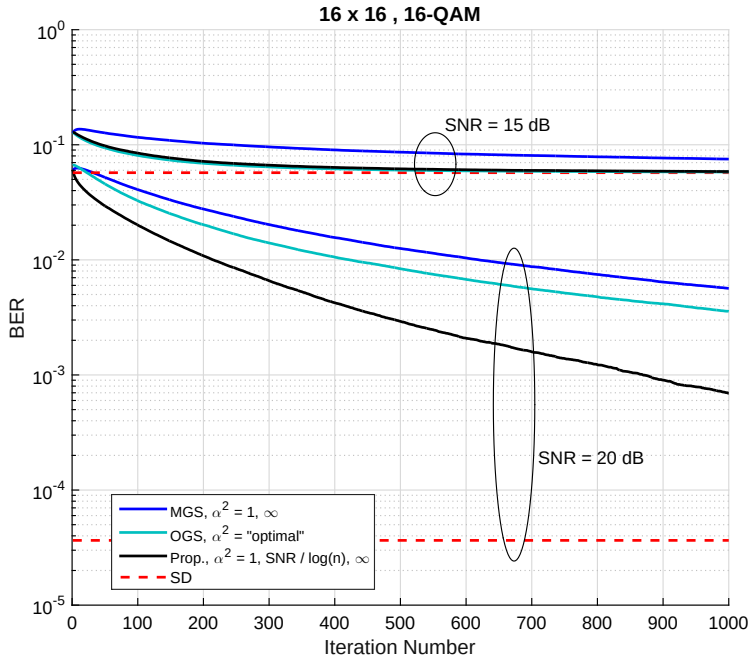


Figure 5.10: BER performance as a function of the number of iterations for the optimised, mixed and triple mixed Gibbs sampling algorithms. The number of antennas was set to $N_R = N_T = 16$ and a 16-QAM constellation was used.

Reconsidering the case where $R = 15$ and the stopping criteria as was done for figures 5.5 and 5.8, BER performance including the T-MGS detector is shown in figure 5.11. The previously identified problems have been mitigated by T-MGS and the curve shows that optimal diversity is indeed obtained for the entirety of the studied SNR range. Moreover, for a reference BER value of 10^{-3}, a gain of 5 dB is attained by the proposed algorithm over the state-of-the-art LR method. Taking into account the previously mentioned complexity reduction techniques and the attained performance, T-MGS is certainly a suitable candidate for solving the detection problem in large MIMO systems. In addition, it is worth emphasising that the algorithm is parallel architecture-friendly, which is, in this day and age, becoming an increasingly important feature.

In order to further confirm the suitability of the proposed algorithm for very large MIMO ($N = 32$), a simulation using $q_\infty = 1/(10N_T)$, $R = 15$ and $t_{max} = 1000$ was performed. The result is depicted in figure 5.12, and the previously obtained conclusions can also be applied here. Note that, further augmenting the number of antennas or the constellation size would lead to a drastic increase in complexity, even using the stopping criteria or the complexity reduction techniques suggested in section

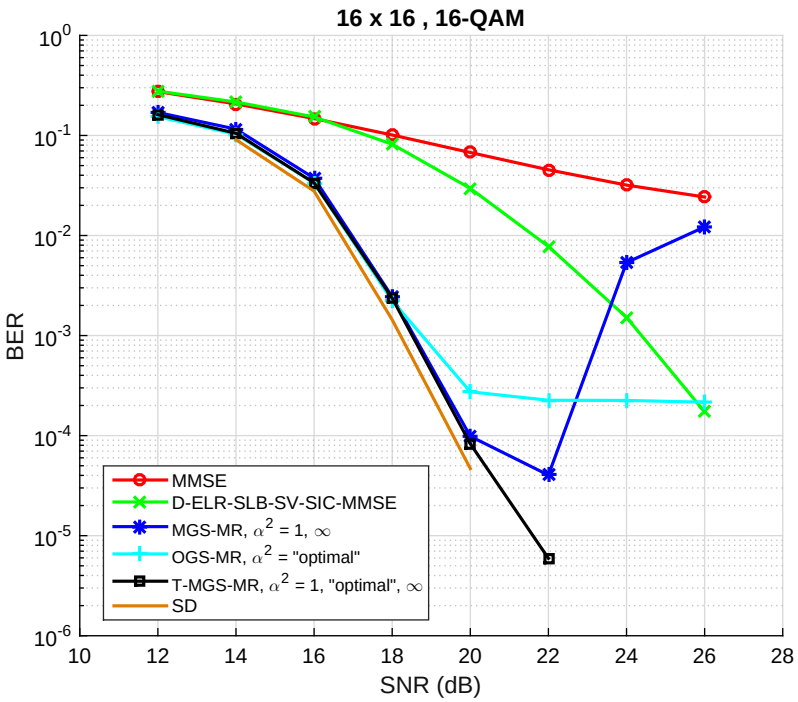


Figure 5.11: BER performance as a function of the SNR for the optimised, mixed and triple mixed Gibbs sampling algorithms with multiple restarts for $N_R = N_T = 16$ and a 16-QAM constellation.

5.3.2. Nonetheless, the algorithm can be efficiently used to obtain an estimate of SD performance for any number of dimensions, provided the initial vector is close to the initial solution (for instance, feeding the solution of the D-ELR algorithm as the initial vector in T-MGS would lead to a lower number of average number of required iterations).

5.4 Concluding Remarks

In this chapter, an optimised randomised algorithm based on Gibbs sampling has been evaluated. Particularly, a trade-off between convergence rate and the probability of finding the optimal solution has been identified. The existing approaches in the literature have been replicated here and based on the corresponding results, a faster, better performing alternative has been proposed. Specifically and yielding the same complexity, the suggested variant converges to the optimal solution faster, whilst alleviating the stalling problem in the high SNR regime. Further, it was demonstrated that the implemented algorithm outperforms a state-of-the-art lattice reduction method.

Even though some techniques to limit and reduce Gibbs sampling complexity have been reviewed in this thesis, it is left for future work the introduction of more efficient methods to perform detection in massive MIMO ($N > 32$) with large constellations (64-QAM and beyond). Furthermore, the extension to soft-decoding of the proposed algorithm and corresponding hardware implementation (making use of the GPU[3][110], for example) would also be an interesting topic for research.

[3]Graphical Processing Unit

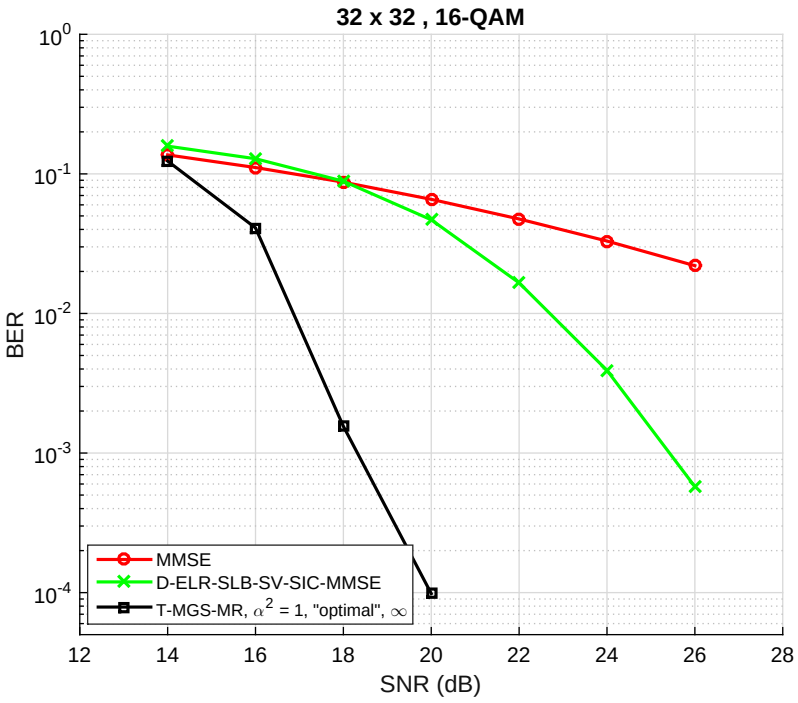


Figure 5.12: BER performance as a function of the SNR for the proposed triple mixed Gibbs sampling algorithm with multiple restarts for $N_R = N_T = 32$ and a 16-QAM constellation.

Chapter 6

Application of Massive MIMO

(this chapter results from a joint work with João Sande Lemos [111])

In this chapter the system model of an in-band full-duplex relay employing massive MIMO is detailed. The challenges as well as the benefits of such setup are studied. In addition, simple yet effective filtering and power allocation schemes are proposed and evaluated.

6.1 In-Band Full-Duplex

In a communications system using a given channel, the bidirectionality of the transmission is called duplexing. Its feasibility is guaranteed when the interference between transmission and reception is eliminated. Presently, in wireless communications, the interference is avoided using independent transmissions in either time (TDD) or frequency (FDD). These systems are denoted as half-duplex (HD), since only half of the available spectrum and/or time-slots are actually utilised for an one-way communication. Taking into account the requirements of future generation networks, the use of in-band full-duplex (IBFD), i.e., transmit and receive data in the same time-slot and frequency band, would be advantageous. The gain would be in terms of spectral efficiency, since, using the same resources, IBFD offers the possibility of doubling the transmission rates when compared to the HD counterpart.

However, the problem with IBFD radios is the self-interference component that appears on the received signal and that is caused by the system's own transmit antennas. Because the power density of an electromagnetic wave is proportional to the inverse of the square of the distance, the loopback interference (LI) power can be several times greater than the power of the signal of interest. Therefore and in order to achieve full-duplex capacity, self-interference suppression techniques must be utilised. This problem has attracted significant attention and a considerable amount of follow-up research has been conducted in the past few years.

The main techniques used to mitigate LI in the literature can be divided in three categories [112]:

Antenna Separation and Natural Isolation This passive method consists of increasing the separation distance between the transmitting and receiving antennas. A larger separation leads to better can-

cellation but also implies the increase of the size of the terminals. Electrical insulating materials to attenuate any line-of-sight or scatterer component between the transmitting and receiving sides can also be employed.

Analog-Domain Cancellation Another possibility is to use active suppression mechanisms where a cancelling signal is sent through RF chains to cancel the self-interference signal at the analog domain of the receive antenna.

Digital-Domain Cancellation The knowledge of the transmitted signal or of some of its statistics can be effectively exploited to suppress LI in the digital-domain using signal processing techniques.

The simultaneous use of all of the aforementioned techniques is possible and is often assumed in the literature. This is due to the considerable amount of self-interference that must be suppressed. As an example [113], for a typical femto base station to achieve the link SNR equal to that of a HD counterpart, it must cancel the LI component by more than 106 dB. Moreover, other hindrances, such as limited A-D converter resolution and dynamic-range, channel estimation errors and amplifier nonlinearities must also be taken into consideration in the design of the system.

In this chapter, the focus of the study will be digital signal processing techniques, while both analog-domain cancellation and antenna separation are assumed to be present.

6.1.1 In-Band Full-Duplex Applications

Widespread commercial in-band full-duplex radios may still be a few years away. This is mainly due to the technological challenges behind suppressing LI. Nonetheless, a large number of academic and research laboratories have already experimentally demonstrated the feasibility of IBFD [114]. It is believed that IBFD technology will first appear in relays serving small-scale wireless communication environments, such as wireless routers [115] or femto base stations. In these short-range systems, the employed power values are typically low, hence it is less difficult to deal with the LI [113].

In order to further improve the transmission rates, IBFD and MIMO technologies can be combined. As will be confirmed in this chapter, MIMO processing provides an effective means of suppressing LI in the spatial domain [18]; a gain that becomes more preponderant in larger dimensions, i.e., when massive MIMO is being employed.

6.2 System Model

In this section, the system model for a massive MIMO full-duplex relay architecture will be presented. Some assumptions commonly present in the literature will also be made in this work. These simplifications come from the inherent difficulty in finding a channel model that is consistent with real measurements.

Consider that K user pairs establish a wireless connection through a relay station, which operates in IBFD mode. Each pair consists of K sources and K destinations, equipped with a single antenna. The relay uses two massive MIMO arrays to simultaneously transmit and receive signals from the different users. The number of antennas used for reception and transmission purposes is denoted by N_{rx} and N_{tx}, respectively. All sources transmit in the same time slot, hence at the relay we have both self-interference and inter-user interference to be dealt with.

The nonexistence of direct links between the different users is assumed. This presumption might be unrealistic but, depending on the scenario, it might be considered acceptable due to path loss effect and severe shadowing (as will be seen later in this chapter, the allocated power at the relay is a few dB greater in magnitude than the assigned powers at the sources). Nonetheless, a direct link between the different users could be exploited in order to design a more reliable system, using both received signal from relay and corresponding pair [116].

The number of antennas at the relay may go up to a few hundred and the number of served users is in general considered to be much lower, i.e., $K << N_{tx}, N_{rx}$. The relay operates in decode-and-forward (DF) mode, which means detection is applied to the received symbols before retransmission. Alternatively, an amplify-and-forward (AF) relay could have been considered [117, 118]. A simplified representation of the studied model is presented in figure 6.1.

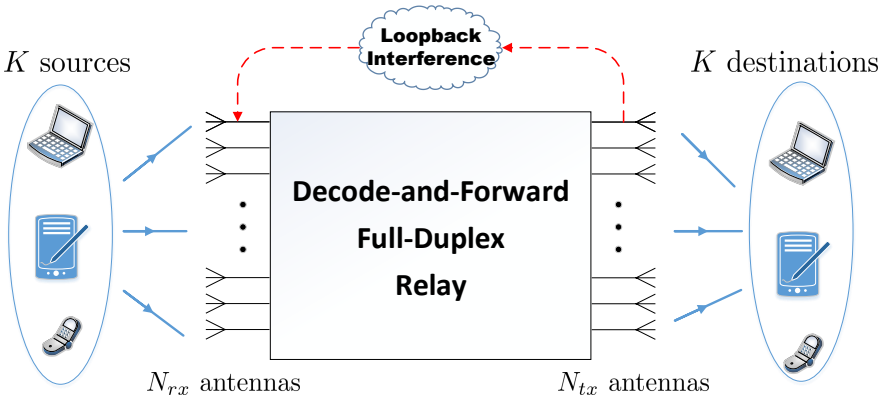


Figure 6.1: Model for full-duplex relay serving K pairs of users.

6.2.1 Channel Model

For the evaluation of the performance of the above described setup, the vectors containing the transmitted symbols from the sources to the relay and the beamformed symbols from the relay to the destinations, denoted by $\mathbf{x} \in \mathbb{C}^{K\times 1}$ and $\mathbf{t} \in \mathbb{C}^{N_{tx}\times 1}$, respectively, have to be taken into consideration. Vector $\mathbf{x}$ is taken from an M-QAM constellation and the output signals to be normalised and uncorrelated, such that $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$ and $\mathbb{E}\{\mathbf{t}^H\mathbf{t}\} = 1$. In this manner, the power transmitted by each source is independent of

K and the total transmitted power at the relay dissociated from N_{tx}. The received vectors at the relay and at the destinations are given by (6.1) and (6.2), respectively.

$$\mathbf{r} = \mathbf{G}_{\text{SR}}\mathbf{D}_{p_{\text{S}}}^{1/2}\mathbf{x} + \sqrt{p_{\text{R}}}\mathbf{H}_{\text{LI}}\mathbf{t} + \mathbf{n}_r. \tag{6.1}$$

$$\mathbf{y}_d = \sqrt{p_{\text{R}}}\mathbf{G}_{\text{RD}}\mathbf{t} + \mathbf{n}_d. \tag{6.2}$$

The uplink symbols in $\mathbf{x}$ pass through a MIMO channel from the sources to the relay denoted by $\mathbf{G}_{\text{SR}} \in \mathbb{C}^{N_{rx}\times K}$, whereas the symbols from the relay to the destinations are filtered by $\mathbf{G}_{\text{RD}} \in \mathbb{C}^{K\times N_{tx}}$. Both matrices are expressed as $\mathbf{G}_{\text{SR}} = \mathbf{H}_{\text{SR}}\mathbf{D}_{\text{SR}}^{1/2}$ and $\mathbf{G}_{\text{RD}} = \mathbf{D}_{\text{RD}}^{1/2}\mathbf{H}_{\text{RD}}$, where $\mathbf{D}_{\text{SR}}$ and $\mathbf{D}_{\text{RD}}$ are diagonal matrices with entries $\beta_{\text{SR},k}$ and $\beta_{\text{RD},k}$ taken from a lognormal distribution [56, Chapter 29], which account for large-scale fading. The fast-fading channel components are present in $\mathbf{H}_{\text{SR}}$ and $\mathbf{H}_{\text{RD}}$, both with independent entries taken from a $\mathcal{CN}(0,1)$ distribution. The LI channel is represented by $\mathbf{H}_{\text{LI}}$ with distribution $\mathcal{CN}(0,\sigma_{\text{LI}}^2)$, where σ_{LI}^2 accounts for the residual LI power after self-interference suppression imposed in the analog-domain. The vectors $\mathbf{n}_r \sim \mathcal{CN}(0,\sigma_{\mathbf{n}_r}^2)$ and $\mathbf{n}_d \sim \mathcal{CN}(0,\sigma_{\mathbf{n}_d}^2)$ account for additive white Gaussian noise at the relay and destinations, respectively. Finally, the diagonal matrix $\mathbf{D}_{p_{\text{S}}}$ with entries $p_{\text{S,k}}$ regulates each source transmit power, while p_{R} denotes the relay's average transmit power.

6.2.2 Channel Estimation

In order to efficiently apply detection, precoding and LI mitigation techniques, channel estimates $\tilde{\mathbf{H}}_{\text{SR}}$, $\tilde{\mathbf{H}}_{\text{LI}}$ and $\tilde{\mathbf{H}}_{\text{RD}}$ of the true $\mathbf{H}_{\text{SR}}$, $\mathbf{H}_{\text{LI}}$ and $\mathbf{H}_{\text{RD}}$ are required. Applying any of the estimation algorithms proposed in the literature [119], it is assumed that any error in the channel matrices is well modelled by $\mathcal{E}_{\mathbf{H}_{\text{SR}}}, \mathcal{E}_{\mathbf{H}_{\text{RD}}}, \mathcal{E}_{\mathbf{H}_{\text{LI}}} \sim \mathcal{CN}(0,\epsilon_{\mathbf{H}}^2)$. The difference between the non-ideal estimates and the true channel values are then assumed to be given by

$$\mathbf{H}_{\text{SR}} = \tilde{\mathbf{H}}_{\text{SR}} + \mathcal{E}_{\mathbf{H}_{\text{SR}}}; \mathbf{H}_{\text{LI}} = \tilde{\mathbf{H}}_{\text{LI}} + \mathcal{E}_{\mathbf{H}_{\text{LI}}}; \mathbf{H}_{\text{RD}} = \tilde{\mathbf{H}}_{\text{RD}} + \mathcal{E}_{\mathbf{H}_{\text{RD}}}. \tag{6.3}$$

The magnitude of the errors of the different matrices is dependent of various factors, including the power used for piloting, number of time slots allocated for channel estimation or noise variance. The large-scale fading matrices, $\mathbf{D}_{\text{SR}}$ and $\mathbf{D}_{\text{RD}}$, are assumed to be perfectly known.

Another aspect that may cause errors in the estimations are hardware imperfections, which similarly to the previous case are going to be taking into consideration by modelling the transmitted signal as [120]:

$$\mathbf{t} = \tilde{\mathbf{t}} + \mathcal{E}_{\mathbf{t}},$$

where $\tilde{\mathbf{t}}$ is the vector to be transmitted after baseband filtering and where all the elements of $\mathcal{E}_{\mathbf{t}} \sim \mathcal{CN}(0,\epsilon_{\mathbf{t}}^2)$ are assumed to be uncorrelated with $\tilde{\mathbf{t}}$. It is worth mentioning that the covariance of the LI term $\sqrt{p_{\text{R}}}\mathbf{H}_{\text{LI}}\mathbf{t}$ in (6.1) is controlled by the parameters p_{R}, σ_{LI}^2, $\epsilon_{\mathbf{t}}^2$ and $\epsilon_{\mathbf{H}}^2$. A complete representation of

the designed model, including the used notation, is shown in figure 6.2.

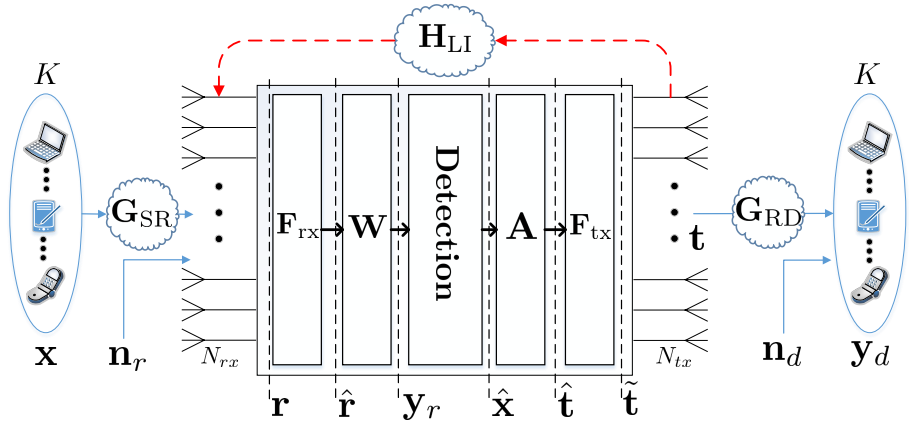


Figure 6.2: Model for full-duplex relay serving K pairs of users, including used filters and notation.

6.2.3 Detection and Precoding

The objective of a conventional bidirectional communication system is to reliably transmit and receive data. In order to achieve this goal in an IBFD system, the constraint posed by the LI must be dealt with. In this work, the filters to perform detection, beamforming and LI mitigation are going to be designed in a dissociated way. With this approach, the chance to use other detection and precoding methods in future work becomes available, as well as the possibility to design different schemes to suppress LI.

Consider, for the moment, an equivalent received signal $\hat{\mathbf{r}}$, whose LI component has been minimised. Hence, one is in the presence of an uplink channel where $K << N_{rx}$. Under this assumption, linear processing techniques are proven to perform close to optimal [18]. The usage of ZF filtering for both detection and beamforming [121] will be considered henceforth. Note that other detection and precoding methods, at the expense of more computational complexity, could be utilised. However, given the dimensions of the considered system, performance gain would be negligible. Nevertheless, none of the results given in this chapter require the condition $K << N_{rx}$, thus the study of an hybrid solution for when $K \to N_{rx}$ could be conducted.

Using the estimation of $\mathbf{G}_{\mathrm{SR}}$, denoted by $\tilde{\mathbf{G}}_{\mathrm{SR}} = \tilde{\mathbf{H}}_{\mathrm{SR}}\mathbf{D}_{\mathrm{SR}}^{1/2}$, the estimated symbols after a ZF filter $\mathbf{W}_{\mathrm{zf}}$ are given by [121]:

$$\hat{\mathbf{x}} = \mathcal{Q}(\mathbf{W}_{\mathrm{zf}}\hat{\mathbf{r}}) = \mathcal{Q}\big((\tilde{\mathbf{G}}_{\mathrm{SR}}^{H}\tilde{\mathbf{G}}_{\mathrm{SR}})^{-1}\tilde{\mathbf{G}}_{\mathrm{SR}}^{H}\hat{\mathbf{r}}\big) = \mathcal{Q}(\tilde{\mathbf{G}}_{\mathrm{SR}}^{\dagger}\hat{\mathbf{r}}), \tag{6.4}$$

where $\mathcal{Q}(\cdot)$ is a symbol-wise quantiser to the constellation set. Upon detection based on $\hat{\mathbf{x}}$, the estimated symbols are forwarded to the destinations, which are assumed to have limited processing capabilities.

Thus, a ZF precoder $\mathbf{A}_{\text{zf}}$ is used. The corresponding expression is

$$\hat{\mathbf{t}} = \mathbf{A}_{\text{zf}}\hat{\mathbf{x}} = \alpha_{\text{zf}}\tilde{\mathbf{G}}_{\text{RD}}^{H}(\tilde{\mathbf{G}}_{\text{RD}}\tilde{\mathbf{G}}_{\text{RD}}^{H})^{-1}\hat{\mathbf{x}}, \tag{6.5}$$

where $\tilde{\mathbf{G}}_{\text{RD}} = \mathbf{D}_{\text{RD}}^{1/2}\tilde{\mathbf{H}}_{\text{RD}}$ is the estimation of the true $\mathbf{G}_{\text{RD}}$ and $\alpha_{\text{zf}} = \left(\mathbb{E}(tr\{(\tilde{\mathbf{G}}_{\text{RD}}\tilde{\mathbf{G}}_{\text{RD}}^{H})^{-1}\})\right)^{-\frac{1}{2}}$ is a scalar chosen to normalise the power of $\hat{\mathbf{t}}$, i.e., $\mathbb{E}(\hat{\mathbf{t}}^{H}\hat{\mathbf{t}}) = 1$. From the definition of $\tilde{\mathbf{H}}_{\text{RD}} \sim \mathcal{CN}(0, 1 + \epsilon_{\mathbf{H}}^{2})$, α_{zf} comes as (cf. appendix A):

$$\alpha_{\text{zf}} = \sqrt{\frac{(N_{tx} - K)}{\sum_{k=1}^{K}\left(\beta_{\text{RD},k}(1 + \epsilon_{\mathbf{H}}^{2})\right)^{-1}}}. \tag{6.6}$$

6.3 Loopback Interference Mitigation

Thus far, only the detection and precoding stages were considered and were designed taking into account that LI was not present. The problem of suppressing LI using linear processing will now be evaluated.

6.3.1 Linear Filtering

Under the model given by (6.1), one is interested in minimising the LI term $\sqrt{p_{\text{R}}}\mathbf{H}_{\text{LI}}\mathbf{t}$, while preserving the signal $\mathbf{G}_{\text{SR}}\mathbf{D}_{p_{\text{S}}}^{1/2}\mathbf{x}$ and taking into account the covariance of the noise vector $\mathbf{R}_{\mathbf{n}_r}$ at the relay. Hence, the mean square error between the desired signal and the one received by the relay station is $\mathbb{E}((\mathbf{G}_{\text{SR}}\mathbf{D}_{p_{\text{S}}}^{1/2}\mathbf{x} - \hat{\mathbf{r}})(\mathbf{G}_{\text{SR}}\mathbf{D}_{p_{\text{S}}}^{1/2}\mathbf{x} - \hat{\mathbf{r}})^{H})$. Using the method proposed in [120], a linear prefilter $\mathbf{F}_{\text{tx}}$ and postfilter $\mathbf{F}_{\text{rx}}$ are considered, such that $\tilde{\mathbf{t}} = \mathbf{F}_{\text{tx}}\hat{\mathbf{t}}$ and $\hat{\mathbf{r}} = \mathbf{F}_{\text{rx}}\mathbf{r}$. The filter that finds the MMSE when $\mathbf{F}_{\text{rx}}$ is fixed satisfies (cf. appendix B)

$$\mathbf{F}_{\text{rx}}\tilde{\mathbf{H}}_{\text{LI}}\mathbf{F}_{\text{tx}} = \mathbf{0}, \tag{6.7}$$

and when $\mathbf{F}_{\text{tx}}$ is fixed becomes (cf. appendix B)

$$\mathbf{F}_{\text{rx}} = \tilde{\mathbf{G}}_{\text{SR}}\mathbf{D}_{p_{\text{S}}}\tilde{\mathbf{G}}_{\text{SR}}^{H}(\tilde{\mathbf{G}}_{\text{SR}}\mathbf{D}_{p_{\text{S}}}\tilde{\mathbf{G}}_{\text{SR}}^{H} + p_{\text{R}}\tilde{\mathbf{H}}_{\text{LI}}\mathbf{R}_{\mathbf{t}}\tilde{\mathbf{H}}_{\text{LI}}^{H} + \mathbf{R}_{\mathbf{n}_r})^{-1}, \tag{6.8}$$

where $\mathbf{R}_{\mathbf{t}} = \mathbf{F}_{\text{tx}}\mathbf{A}_{\text{zf}}\mathbf{R}_{\hat{\mathbf{x}}}\mathbf{A}_{\text{zf}}^{H}\mathbf{F}_{\text{tx}}^{H} + \epsilon_{\mathbf{t}}^{2}\mathbf{I}$. The solution provided by (6.7) corresponds to the null-space projection already presented in [120], and in general finding a reliable solution for large dimensions is hard. On the contrary, (6.8) is a closed-form expression containing the channel estimates and covariance matrices of transmitted vectors and noise, that are assumed to be known. Under the considered model and for a given set of parameters, a trade-off in the end-to-end (e2e) BER is expected. On the one hand, a lower transmitted power at the relay p_{R} reduces the LI power and diminishes any impact of the estimation errors and radiofrequency impairments. On the other hand, a higher p_{R} leads to improved signal-to-noise ratio (SNR) levels at the destinations and hence a lower BER in the forward link channel. Therefore, for a fixed source transmit power, an optimal choice for p_{R} that minimises the e2e BER can be anticipated.

Numerical results for this fact are shown and evaluated in section 6.6.2.

6.4 Achievable Rates

Other figures of merit of the system, other than BER performance, can take advantage of the inclusion of LI mitigation filters $\mathbf{F}_{\text{rx}}$ and $\mathbf{F}_{\text{tx}}$. The theoretical expressions for the achievable rates, as well as their dependence with the allocated powers, will be detailed in this section.

For a given channel realisation, it is possible to compute the achievable rate of each individual link. As was done in [122], the overall rate R_k is considered to be the minimum of both downlink and uplink rates in the following manner:

$$R_k = \min(R_{\text{SR},k}, R_{\text{RD},k}), \tag{6.9}$$

where $R_{\text{SR},k}$ and $R_{\text{RD},k}$ denote the achievable rates between the sources and the relay and between the relay and the destinations, respectively. To obtain the overall achievable rate R_k, one starts by considering the kth user received signal at the relay after filtering and before detection given by

$$\begin{aligned} y_{\text{r},k} =& \sqrt{p_{\text{S},k}}(\mathbf{W}_{\text{zf}}\mathbf{F}_{\text{rx}})_k^T \mathbf{g}_{\text{SR},k} x_k \\ &+ \sum\nolimits_{j\neq k}^{K} \sqrt{p_{\text{S},j}}(\mathbf{W}_{\text{zf}}\mathbf{F}_{\text{rx}})_k^T \mathbf{g}_{\text{SR},j} x_j \\ &+ \sqrt{p_{\text{R}}}(\mathbf{W}_{\text{zf}}\mathbf{F}_{\text{rx}})_k^T \mathbf{H}_{\text{LI}}\mathbf{t} + (\mathbf{W}_{\text{zf}}\mathbf{F}_{\text{rx}})_k^T \mathbf{n}_r, \end{aligned} \tag{6.10}$$

where $\mathbf{g}_{\text{SR},k}$ denotes the kth column of $\mathbf{G}_{\text{SR}}$. The received signal at each destination link before detection is given by

$$\begin{aligned} y_{\text{d},k} =& \sqrt{p_{\text{R}}}\mathbf{g}_{\text{RD},k}^T(\mathbf{F}_{\text{tx}}\mathbf{A}_{\text{zf}})_k \hat{x}_k \\ &+ \sqrt{p_{\text{R}}}\sum\nolimits_{j\neq k}^{K} \mathbf{g}_{\text{RD},k}^T(\mathbf{F}_{\text{tx}}\mathbf{A}_{\text{zf}})_j \hat{x}_j + n_{d,k}. \end{aligned} \tag{6.11}$$

Both expressions (6.10) and (6.11) may be seen as a known mean gain times the desired signal plus an uncorrelated effective noise term that includes channel impairment effects, interpair and loopback interference, and Gaussian noise. A valid technique commonly used in large MIMO systems [123] is to approximate the effective noise term by a Gaussian noise term under the central limit theorem, and which gives a good approximation for $R_{\text{SR},k}$ as follows [18]

$$R_{\text{SR},k} = \log_2\left(1 + \frac{p_{\text{S},k}\text{MV}_{\text{SR},k}}{p_{\text{S},k}\text{V}_{\text{SR},k} + \sum_{j\neq k}^{K} p_{\text{S},j}\text{MP}_{\text{SR},(k,j)} + p_{\text{R}}\text{LI}_{\text{SR},k} + \text{AN}_{\text{SR},k}}\right), \tag{6.12}$$

where $\text{MV}_{\text{SR},k} = |\mathbb{E}\{(\mathbf{W}_{\text{zf}}\mathbf{F}_{\text{rx}})_k^T \mathbf{g}_{\text{SR},k}\}|^2$, $\text{V}_{\text{SR},k} = \mathbb{V}\text{ar}\{(\mathbf{W}_{\text{zf}}\mathbf{F}_{\text{rx}})_k^T \mathbf{g}_{\text{SR},k}\}$, $\text{MP}_{\text{SR},(k,j)} = \mathbb{E}\{|(\mathbf{W}_{\text{zf}}\mathbf{F}_{\text{rx}})_k^T \mathbf{g}_{\text{SR},j}|^2\}$, $\text{LI}_{\text{SR},k} = \mathbb{E}\{||\mathbf{w}_{\text{zf},k}^T \mathbf{F}_{\text{rx}}\mathbf{H}_{\text{LI}}\mathbf{F}_{\text{tx}}\mathbf{A}_{\text{zf}}||^2\}$ and $\text{AN}_{\text{SR},k} = \sigma_{n_r}^2 \mathbb{E}\{||(\mathbf{W}_{\text{zf}}\mathbf{F}_{\text{rx}})_k||^2\}$. The term $\text{MV}_{\text{SR},k}$ denotes the mean gain after filtering associated with the kth uplink user, while $\text{V}_{\text{SR},k}$ stands for the associated variance in the actual gain. The term $\text{MP}_{\text{SR},(k,j)}$ corresponds to the inter-user

interference from user j in user k and $\mathrm{LI}_{\mathrm{SR},k}$ the residual interference that the filters cannot suppress. Finally, $\mathrm{AN}_{\mathrm{SR},k}$ denotes the additive enhanced noise.

Similarly, $R_{\mathrm{RD},k}$ is well approximated by [18]

$$R_{\mathrm{RD},k} = \log_2\left(1 + \frac{p_{\mathrm{R}}\mathrm{MV}_{\mathrm{RD},k}}{p_{\mathrm{R}}\mathrm{V}_{\mathrm{RD},k} + p_{\mathrm{R}}\mathrm{MP}_{\mathrm{RD},k} + \mathrm{AN}_{\mathrm{RD},k}}\right), \tag{6.13}$$

where $\mathrm{MV}_{\mathrm{RD},k} = |\mathbb{E}\{\mathbf{g}_{\mathrm{RD},k}^T(\mathbf{F}_{\mathrm{tx}}\mathbf{A}_{\mathrm{zf}})_k\}|^2$, $\mathrm{V}_{\mathrm{RD},k} = \mathbb{V}\mathrm{ar}\{\mathbf{g}_{\mathrm{RD},k}^T(\mathbf{F}_{\mathrm{tx}}\mathbf{A}_{\mathrm{zf}})_k\}$, $\mathrm{MP}_{\mathrm{RD},k} = \sum_{j\neq k}^{K}\mathbb{E}\{|\mathbf{g}_{\mathrm{RD},k}^T(\mathbf{F}_{\mathrm{tx}}\mathbf{A}_{\mathrm{zf}})_j|^2\}$ and $\mathrm{AN}_{\mathrm{RD},k} = \sigma_{n_d}^2$. Insight into each term is similar to the previous case. A graphical representation of the overall rate R_k in terms of the power at the relay p_{R}, for fixed sources' power $p_{\mathrm{S},k} = 1$[1], is represented in figure 6.3. There, one may observe the following: there is an optimal power at the relay that maximises the overall rate. For a low value of p_{R}, it is the downlink that constrains the rate, since the destinations' received power is not sufficient to overcome the noise power $\sigma_{\mathbf{n}_d}^2$. Conversely, when high values are set for p_{R}, it is the LI power that hinders the performance of the system, as the signal-to-interference ratio at the relay becomes very low.

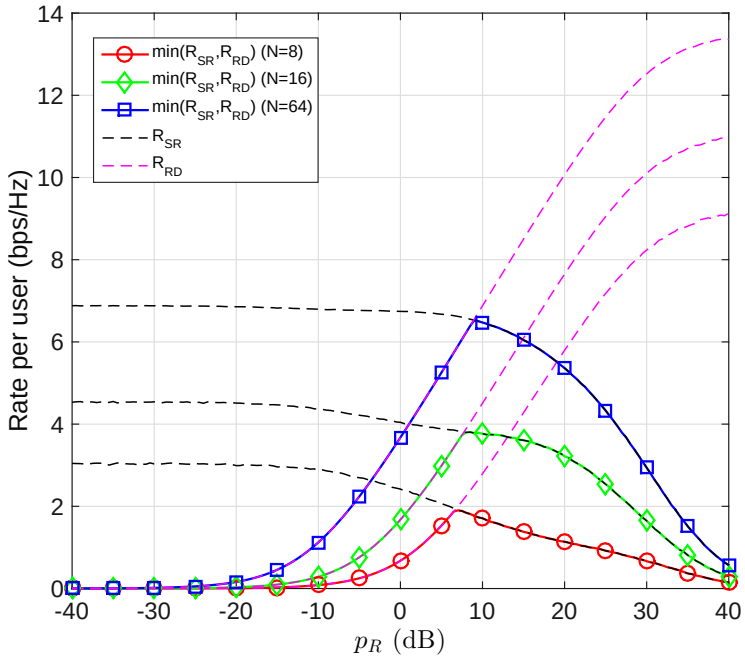


Figure 6.3: Achievable rate per user when no power allocation at the sources or large-scale fading are considered. Simulation parameters are $\sigma_{\mathbf{n}_d}^2 = 1$, $K = 5$ pairs, $\mathrm{SNR}_{\mathrm{R}} = 10$ dB and $\epsilon_{\mathbf{H}} = \epsilon_{\mathbf{t}} = 10^{-3}$.

Even though expressions (6.12) and (6.13) provide a theoretical insight into the achievable capacity of the system, there are some observations that are worth of mention. On the one hand, these expressions are only a measure of spectral efficiency (bits/s/Hz), hence do not account for possible bandwidth allocation schemes or the usage of different constellation sizes by the various pairs, which would lead to better real

[1]For this analysis $\mathbf{D}_{\mathrm{SR}}^{1/2} = \mathbf{D}_{\mathrm{RD}}^{1/2} = \mathbf{I}$ was considered. Nonetheless, if large-scale fading components were considered, one could allocate $p_{\mathrm{S},k}$, such that $\beta_{\mathrm{SR},k}p_{\mathrm{S},k} = 1$, for all k, provided power constrains are met. If this was done, conclusions would be similar to the presented ones.

rates. On the other hand, there might be situations where bandwidth allocation can not be performed (e.g., hardware constraints) or where a high noise figure does not permit the increase of the constellation size, thus the practical importance of (6.12) and (6.13) as will be seen in next section.

6.5 Optimal Power Allocation

One of the primary goals for future wireless networks is to maximise power efficiency. However, one does not want to compromise on the individual quality of service. With this objective in mind, one wants to find the system optimal power allocation (OPA), i.e., compute the required power transmitted by both sources and relay, such that the desired rate for each communication pair k is guaranteed. Moreover, peak power constraints should be satisfied and overall power consumption taken into consideration; in other words, the system energy efficiency, defined as $\text{EE} = \sum_{k=1}^{K} R_k/(p_{\text{R}} + \sum_{k=1}^{K} p_{\text{S},k})$, should be maximised. This can be described as the following optimisation problem:

$$\begin{aligned} \underset{p_{\text{R}},p_{\text{S},1},..,p_{\text{S},K}}{\text{minimise}} \quad & \sum_{k=1}^{K} p_{\text{S},k} + p_{\text{R}}, \\ \text{subject to} \quad & R_k \geq R_{0,k},\ k = 1,\ldots,K; \\ & 0 \leq p_{\text{S},k} \leq p_{\text{S}_{0,k}},\ k = 1,\ldots,K; \\ & 0 \leq p_{\text{R}} \leq p_{\text{R}_0}, \end{aligned} \tag{6.14}$$

where $R_{0,k}$ and $p_{\text{S}_{0,k}}$ are the required rate and peak power for pair k respectively, and p_{R_0} is the relay station peak power.

The solution to problem (6.14) involves the computation of the channel statistics which are the output of involved expressions due to the non-linear dependence of the optimisation variables in $\mathbf{F}_{\text{rx}}$. Note that these statistics can be either computed numerically (via simulation) or, by exploiting the knowledge of the channel statistics, good approximations for the closed-form expressions can be obtained. The latter was done in [18] for the case where no LI mitigation filters are employed, that is $\mathbf{F}_{\text{rx}} = \mathbf{F}_{\text{tx}} = \mathbf{I}$. However, when the MMSE LI suppression filters are considered, finding a closed form expression becomes a non-trivial task (if even possible). For that reason, a numerical method will be employed, consisting of a genie-aided receiver (all the information regarding the channel statistics is available). An algorithm to find the optimal power allocation that maximises the energy efficiency while meeting the individual rate requirements is proposed next.

Exploiting the fact that problem (6.14) is a linear program, one may construct a simple and computationally efficient algorithm, capable of obtaining the referred OPA. Several channel realisations are drawn to compute the corresponding channel statistics coefficients, LI mitigation filters and power vectors. The procedure is done iteratively, until the optimal power vector is reached. For this purpose, one stopping criteria is considered: a maximum number of iterations L. In addition, a threshold in the consecutive power vectors could also have been utilised. However, as will be seen in the numerical results,

the convergence of the algorithm is rather fast (few iterations are needed), hence this additional stopping criteria will not be considered. The method is summarised in algorithm 6.

Algorithm 6 - Input: Peak powers $p_{S_{0,k}}$ and p_{R_0} ; **Output:** Optimal powers: p_{S_k} and p_R

---

1. Initialisation: Set $i = 1$; initialise powers $p_{\mathrm{S},k,1} = p_{\mathrm{S}_{0,k}}$ and $p_{\mathrm{R},0} = p_{\mathrm{R}_0}$; define L as the total number of iterations and set N_{it} as the number of channel realisations per iteration.

2. Iteration i:

1) Compute channel statistics:

for $n = 1$ to N_{it} **do**

i) Generate $\mathbf{G}_{\mathrm{SR}}$, $\mathbf{G}_{\mathrm{RD}}$, $\mathbf{H}_{\mathrm{LI}}$ and compute $\mathbf{W}_{\mathrm{zf}}$ and $\mathbf{A}_{\mathrm{zf}}$

ii) Compute filter $\mathbf{F}_{\mathrm{rx}}$ with $p_{\mathrm{S},k,i}$ and $p_{\mathrm{R},i}$

iii) Compute instantaneous rate coefficients for all k pairs (as in (6.12) and (6.13))

end for

2) Average the coefficients over the realisations to obtain channel statistics.

3) Solve linear program (6.14) with the coefficients found in step 2) to obtain the new $p_{\mathrm{S},k,i}$ and $p_{\mathrm{R},i}$.

4) Set $p_{\mathrm{S},k,i+1} = p_{\mathrm{S},k,i}$ and $p_{\mathrm{R},i+1} = p_{\mathrm{R},i}$.

3. Check:

if $i = L$ **then return** $p_{\mathrm{S},k,L}$ and $p_{\mathrm{S},L}$

else set $i \leftarrow i + 1$.

end if

It is worth mentioning that the design of an iterative method such as the one presented in algorithm 6 would not be possible if it was not for the fact that (6.14) is a linear program. For instance, if the individual rate requirement was not present and a given sum spectral efficiency was to be met instead, then the optimisation problem would become a geometric programming [18], whose complexity renders iterative methods impractical.

However, since the genie receiver is being utilised in algorithm 6, the output powers p_{R} and $p_{\mathrm{S},k}$ are given *a posteriori*, that is, after the transmissions were done, which is not practical for real systems. Hence, algorithm 6 is used to demonstrate that, for a given channel realisation, there is an optimal power allocation that meets the requirements of the various users, whilst keeping energy efficiencies high. In order to obtain a similar power allocation *a priori*, a closed-form expression for the rates would have been required.

6.6 Numerical Results

In this section, the performance of the proposed filters is compared and a special emphasis on both e2e link reliability and efficiency of the power allocation scheme is provided. All results are obtained via Monte Carlo simulation, using uncoded MIMO. The optimisation problem is solved using cvx (a Matlab toolbox available at [124]).

6.6.1 System Parameters

A symmetric system with the same number of users K on both sides with one antenna each and $N = N_{tx} = N_{rx}$ antennas at the relay is assumed. Transmission and reception at the relay are done in the same time-slot and frequency band, and an arbitrary processing delay $d \geq 1$ is assumed such that at time instant i: $\hat{\mathbf{x}}[i] = f(\mathbf{x}[i-d])$, where $f(\cdot)$ denotes a generic function. Moreover, $\mathbf{F}_{\text{rx}}$ as in (6.8) and $\mathbf{F}_{\text{tx}} = \mathbf{I}$ are used. The latter choice is due to the fact that one is in the presence of a distributed MIMO system where the various K destinations are independent of each other. A filter $\mathbf{F}_{\text{tx}} \neq \mathbf{I}$ would deteriorate the downlink performance as the purpose of the beamforming filter $\mathbf{A}_{\text{zf}}$ would become distorted. However, if the considered system was a point-to-point transmission (e.g., relay to relay) then a different choice for $\mathbf{F}_{tx}$ could have been done, since a relay has the decoding and channel estimation capacities that the considered destinations are not assumed to have.

Finally and without loss of generality, $\sigma^2_{\text{LI}} = 1$ is considered. It is noteworthy that σ^2_{LI} controls the variance of the entries in the LI channel $\mathbf{H}_{\text{LI}}$ but also that the variance of the LI term $\sqrt{p_{\text{R}}}\mathbf{H}_{\text{LI}}\mathbf{t}$ is defined by both p_{R} and σ^2_{LI}. Therefore, the study will be conducted with fixed $\sigma^2_{\text{LI}} = 1$ and varying p_{R}. The SNR at the relay is defined as $\text{SNR}_{\text{R}} = \frac{\sum_{k=1}^{K} \beta_{\text{SR},k} p_{\text{S},k}}{\sigma^2_{\mathbf{n}_r}}$.

6.6.2 Allocated Power at the Relay *vs*. BER

An evaluation of the performance of the system in terms of BER at both relay and destinations, considering only small-scale fading, that is $\mathbf{D}^{1/2}_{\text{SR}} = \mathbf{D}^{1/2}_{\text{RD}} = \mathbf{I}$, will now be conducted. Firstly, for a given SNR_{R} and fixed uniform transmitted power $p_{\text{S,k}} = 1$ for all k, one is interested in studying different allocated powers at the relay p_{R} and for an increasing number of antennas N. For comparison purposes, natural isolation (NI) (when $\mathbf{F}_{\text{rx}} = \mathbf{F}_{\text{tx}} = \mathbf{I}$, which corresponds to ignoring the LI component) and the HD counterpart (when $\mathbf{t} = 0$ and $\mathbf{F}_{\text{rx}} = \mathbf{F}_{\text{tx}} = \mathbf{I}$) were implemented. Setting the variance in the errors to be $\epsilon^2_{\mathbf{t}} = \epsilon^2_{\mathbf{H}} = 10^{-3}$, curves of BER for different numbers of antennas and relay power p_{R} are depicted in figure 6.4. It can be seen that for an increasing number of antennas, the system using the MMSE filter becomes more robust to the LI, showing a BER performance closer to the HD for higher values of power p_{R} than when compared with NI. More specifically, the performance gap between NI and MMSE filters are 10 and 15 dB for $N = 16$ and $N = 64$, respectively. It would be interesting to have the BER curve corresponding to the null-space projection pair of filters $\mathbf{F}_{\text{rx}}$ and $\mathbf{F}_{\text{tx}}$ and that is left for future work.

The e2e BER performance in terms of the allocated power p_{R} and using the same setup as before is then studied. For this purpose, one sets the variance of the noise at the destinations $\sigma^2_{\mathbf{n}_d}$, quantises the received signals in (6.2), and compares them with the symbols transmitted by the sources. The results are shown in figure 6.5. It is confirmed that for a given configuration there is an optimal choice for the power at the relay that minimises the e2e BER. Moreover, it is concluded that a larger number of antennas attains the minimum BER with less power. Particularly, and as was expected, when the noise variance at the destinations increased by 10 dB, the minimum attainable BER was 10 times higher and

the required power was about 7 dB above.

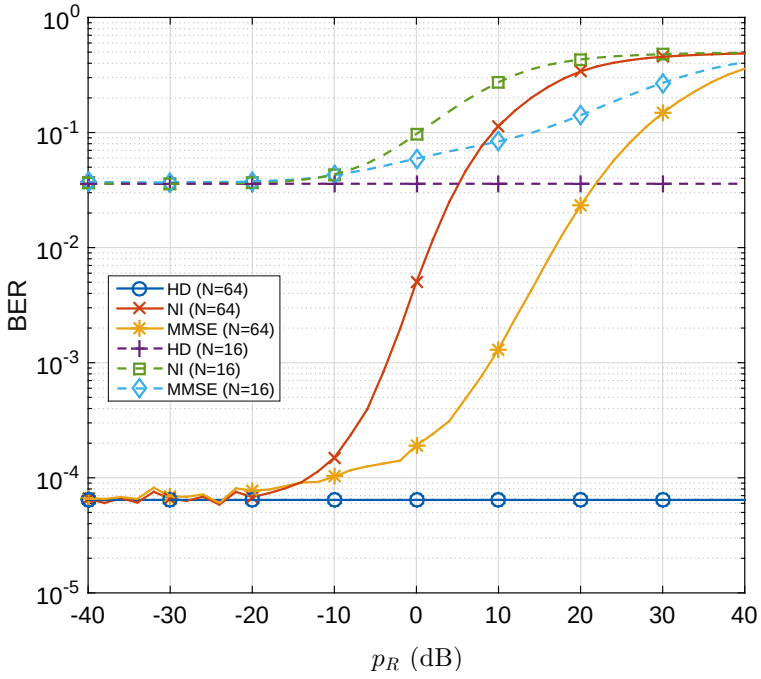


Figure 6.4: BER performance at the relay for different numbers of antennas N, $K = 5$ pairs, $\text{SNR}_\text{R} = 8$ dB and 16-QAM.

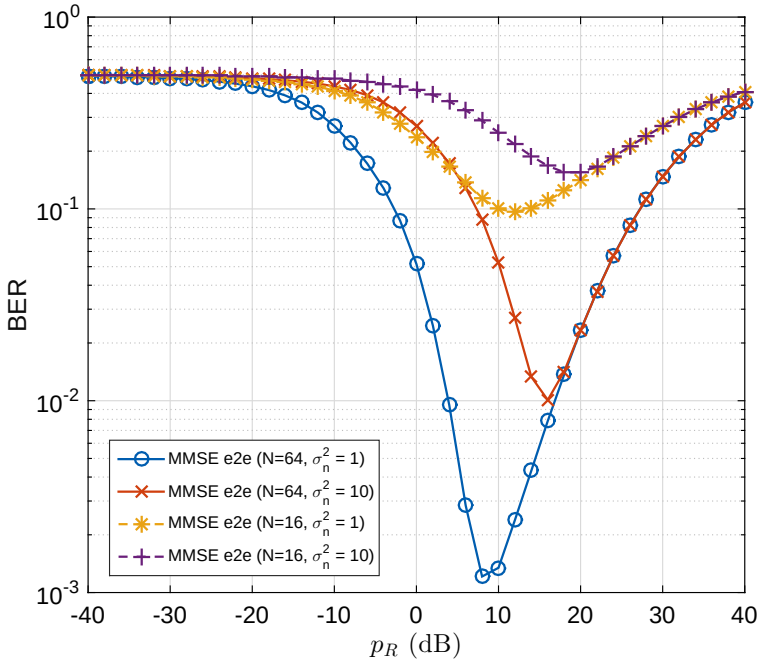


Figure 6.5: End-to-end BER performance for different N and $\sigma^2_{\mathbf{n}_d}$, $K = 5$ pairs, $\text{SNR}_\text{R} = 8$ dB and 16-QAM.

6.6.3 Optimal Power Allocation Algorithm

Finally, the results of the proposed algorithm in section 6.5, whose objective was to find the OPA that meets the rate constraints of each link, while minimising the overall power of the system, is here eval-

uated. Large-scale fading effects are now taken into consideration, more precisely, $\beta_{\text{SR},k}$ and $\beta_{\text{RD},k}$ are independent variables generated from a lognormal distribution with mean value $m = 1$ and standard deviation $\sigma = 6$ dB, that is $\beta_{\text{SR},k}, \beta_{\text{RD},k} \sim m \exp(\sigma \mathcal{N}(0,1))$. Note that, even though the optimisation problem in (6.14) is a linear program[2], there is no guarantee of convergence of algorithm 6. Specifically, convergence exists if and only if at each iteration the rate outcomes are the desired ones, i.e., $R_k \geq R_{0,k}$, and if the error in successive cost functions

$$\Delta_i = ||\big(\sum_{k=1}^{K} p_{S,k,(i-1)} + p_{R,(i-1)}\big) - \big(\sum_{k=1}^{K} p_{S,k,i} + p_{R,i}\big)||^2$$

is a decreasing function in terms of the number of iterations L. For a given channel realisation, the required rates $R_{0,k}$, channel gains $\beta_{\text{SR},k}$ and $\beta_{\text{RD},k}$, allocated powers and corresponding errors Δ_{i+1} are presented in table 6.1. As can be inferred, for $L \geq 3$, the changes in the output powers are residual, thus algorithm 6 is a reliable approach to obtain the optimal power allocation *a posteriori*.

User k	$R_{0,k}$	$\beta_{\text{SR},k}$	$\beta_{\text{RD},k}$	Iter. i	$p_{S,1,i}$	$p_{S,2,i}$	$p_{S,3,i}$	$p_{S,4,i}$	$p_{R,i}$	Δ_{i+1}
1	1.20	0.027	0.296	1	2.0000	2.0000	2.0000	2.0000	10.0000	2.87e+02
2	1.24	0.234	0.328	2	0.2386	0.0286	0.0020	0.0305	0.7375	8.74e-03
3	1.17	3.083	0.026	3	0.1644	0.0196	0.0014	0.0210	0.7373	4.79e-07
4	1.39	0.260	0.082	4	0.1635	0.0197	0.0014	0.0210	0.7374	5.38e-08

Table 6.1: Optimal power allocation algorithm output for $K = 4$ pairs with different requirements in terms of rates and affected by different large-scale fading values. The number of iterations was set to $L = 4$ and the remaining parameters were as follows: $N = 128$, $\text{SNR}_\text{R} = 10$ dB, $\epsilon_{\mathbf{H}} = \epsilon_{\mathbf{t}} = 10^{-3}$, $\sum_{k=1}^{K} R_{0,k} = 5$, $\sigma^2_{\mathbf{n}_d} = 1$. The normalised peak power of both sources and relay were set as $p_{\text{S}_{0,k}} = 3$ dB and $p_{\text{R}_0} = 10$ dB.

Parameters $N_{it} = 10^3$ and $L = 5$ in algorithm 6 are, therefore, empirically defined, such that both channel statistics and output powers in algorithm 6 are good approximations of their real value. Additionally, the normalised peak power of the sources and relay are set as $p_{\text{S}_{0,k}} = 3$ dB and $p_{\text{R}_0} = 10$ dB, respectively. The performance of algorithm 6 is determined in terms of EE for the case where interference effects are disregarded (OPA-NI) and when the MMSE filter is used (OPA-MMSE). This method is also compared with the situation where an optimal uniform power allocation (OUPA) is used, which corresponds to considering $p_{\text{S},k} = p_\text{S}$, for all k, in (6.14). Figure 6.6 shows the curves of average EE for different values of desired e2e sum-rate, defined as $\sum_{k=1}^{K} R_{0,k}$, where the individual required rates $R_{0,k}$ are taken from a discrete uniform distribution (cf. appendix C).

As one may see in figure 6.6, for the same desired sum-rate, the EE of OPA-MMSE is improved significantly when compared to OUPA, while guaranteeing that no link is in outage ($R_k \geq R_{0,k}$ for all k). Furthermore, since MMSE filtering effectively reduces the LI, a lower amount of energy is spent to achieve the same sum-rate when compared with OPA-NI, an effect that becomes more preponderant in higher rates regimes. In detail, when the number of antennas increased from 64 to 128, energy efficiency

[2]If feasible, a linear program guarantees the convergence to a global minimum [125].

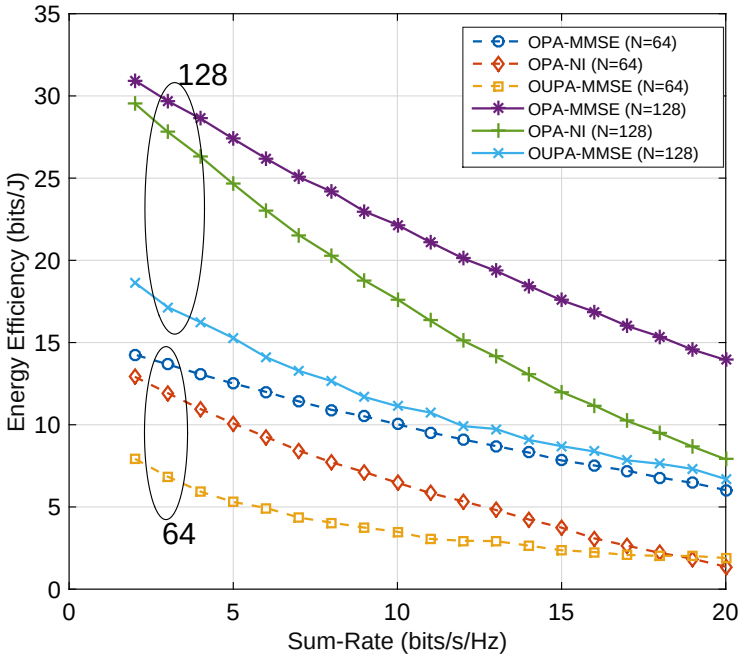


Figure 6.6: Energy efficiency for different power allocation schemes, filters and number of antennas. The remaining parameters are $\sigma^2_{\mathbf{n}_d} = 1$, $K = 10$ pairs and $\text{SNR}_\text{R} = 16$ dB.

more than doubled. This is explained by the massive MIMO effect: a better LI mitigation leads to lower required powers at both relay and destinations, thus higher energy efficiencies.

As a side note, it is important to mention that problem (6.14) is not always possible to solve. Given a set of channel statistics and peak power constraints, there might be a situation where one or more pairs' individual rate requirements become impossible to satisfy. Define outage as the situation where the achievable rate R_k is below the minimum required $R_{0,k}$, i.e., $R_k < R_{0,k}$. When one or more pairs are in outage, then the optimisation problem is said to be infeasible. In this case, the best alternative solution for power allocation is application dependent. For example, users in outage can be regarded as nonexistent until their link quality improves or, alternatively, instead of trying to meet the individual requirements, a sum-rate approach (waterfilling-like) could have been followed, as was done in [18]. For simulations purposes, whenever problem (6.14) was infeasible, the corresponding channel realisations were discarded.

6.7 Concluding Remarks

It has been demonstrated in this chapter that massive MIMO and full-duplex systems can effectively be used together in order to simultaneously serve a reasonable number of user pairs. In particular, linear filters to suppress interference and power allocation schemes to maximise the energy efficiency have been designed. The results have shown that the gains increase with the number of antennas at the relay. Note, however, that some herein made assumptions (also made in the literature) may not be coherent with experimental results yet to be performed; mainly the ones related with the loopback interference channel, whose modelling is of the uttermost importance.

Chapter 7

Conclusions

As the final chapter of this document, an overview of the statements made in the previous chapters will be carried out. In addition, some suggestions for future work will be left.

7.1 Overall Conclusions

An exponential growth in data traffic is to be expected in the next decade, hence the need for the fifth generation of wireless networks. This will incorporate a wide range of efficient solutions capable of fulfilling capacity demands. One of the most promising alternatives is the increase in the number of antennas at the transceiver sides, which promises to improve current spectral efficiencies. However, when massive MIMO systems are employed, performance of conventional detectors may become subpar and corresponding complexities unbearable; thus the need for further research in this emerging topic. The contributions of this thesis are three-fold.

Firstly, when $N_R >> N_T$, linear detection techniques (ZF and MMSE) were proven to attain close to optimal diversity, while yielding low complexity. Nevertheless, these filters require the computation of an inverse matrix, which might be cumbersome to implement. Exploiting the channel orthogonalisation present in massive MIMO systems, the performance of the hardware-friendly Neumann series was verified, resulting in acceptable bit error rates (BER), especially when the number of considered terms in the series was above two. In order to further lower complexity and improve throughputs, two algorithms to update the inverse of a matrix were proposed. They consist in matrix inflation and deflation operations, and without compromising on the performance of the system, the method is able to reduce the number of required computations from $\mathcal{O}(N_T^3)$ to $\mathcal{O}(N_T^2)$.

Secondly, and unlike the previous case, when $N_T = N_R$, linear detection methods attain diversity equal to one, which is undesirable. Since performing maximum likelihood detection is impractical, one needs to find more suitable methods that yield low complexity and perform as close to optimal as possible. Taking this fact into account, parallel architecture-friendly randomised algorithms based on Gibbs sampling have been introduced and evaluated. The stalling problem was identified, characterised and

mitigated using two distinct alternatives. However, none of the variants attained maximum diversity in the high SNR regime and, as a result, a triple mixed Gibbs sampling method was proposed and optimised. The method proved to solve the problem identified in the high SNR, whilst requiring a lower number of iterations to achieve a given BER value.

Lastly, capitalising on the benefits of large-scale arrays, a setup including both full-duplex and massive MIMO technologies has been studied. Linear filters to jointly perform detection, precoding and loopback interference mitigation were designed and the corresponding impact in performance was assessed. Particularly, a trade-off in the end-to-end BER was identified, meaning that an optimal power allocation at the relay that minimises outage probability can be found. Additionally, it has been shown using a linear program that an optimal power allocation scheme satisfying individual throughput requirements, whilst maximising energy efficiency, exists.

On the whole, it was demonstrated in this thesis that the problem arising in massive MIMO detection can be efficiently surpassed. Given the amount of existing testbeds and prototypes, it is very likely that these systems are actually brought out and, as a consequence, that future wireless networks' requirements are met. Notwithstanding the good progress made so far in the development of such systems, intense research efforts still need to be made if this technology is to see the light of the day [126].

7.2 Future Work

This section's intention is not to suggest new areas of possible research but rather to make the reader aware of possible paths that can be followed in future work. Among other things, eventual follow-up research could include:

Correlated Fading Throughout this dissertation, the widely used Rayleigh channel has been considered. Even though this assumption proves to be a good approximation of reality in massive MIMO systems, other, more suitable channel models can be applied and corresponding impact in performance assessed.

Imperfect Channel State Information Apart from chapter 6, perfect channel state information at the receiver was assumed. Similarly to the previous item, it would be important to study and implement relevant channel estimation algorithms and ensuing influence. It is worth mentioning, however, that channel acquisition methods based on Gibbs sampling exist and, hence, would serve as a suitable extension to this work.

Coded transmissions The herein studied detectors output hard-decisions. However, if coded transmissions are considered instead, soft-outputs are preferable, as channel decoders attain better performance when these are utilised. Therefore, it would be interesting to see the extension of the studied randomised algorithm to soft-outputs, as well as an evaluation of the resilience of the Neumann series approximation when error correcting codes are employed.

Bibliography

(the hyperlinks provided below have proved to be working on 12/07/2015)

[1] A. Osseiran *et al.*, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[2] V. Jungnickel *et al.*, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 44–51, 2014.

[3] D. Flore, "Initial priorities for the evolution of LTE in Release-13," 3GPP, 2015. [Online]. Available: http://www.3gpp.org/news-events/3gpp-news/1628-rel13

[4] W. Tong and Z. Peiying, "5G: A technology vision," *WinWin*, 2014. [Online]. Available: http://www.huawei.com/en/static/HW-329327.pdf

[5] Cisco, "Visual Networking Index: Global Mobile Data Traffic Forecast Update," White Paper, February 2015. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html

[6] Openet Telecom, "Closing the mobile data revenue gap," White Paper, 2010. [Online]. Available: http://ibmtelconewsletter.files.wordpress.com/2011/02/wp_closing_mobile_data_revenue_gap_a4.pdf

[7] P. Popovski *et al.*, "Scenarios, requirements and KPIs for 5G mobile and wireless system," Deliverable, Project Number ICT-317669, METIS, April 2013. [Online]. Available: https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D1.1_v1.pdf

[8] NTT DOCOMO, "Requirements, concept and technologies," White Paper, July 2014. [Online]. Available: https://www.nttdocomo.co.jp/english/corporate/technology/whitepaper_5g/

[9] M. Abolfazl and A. Fumiyuki, "Energy-efficient resource allocation for next generation wireless networks," slides presented at the *22nd European Signal Processing Conference (EUSIPCO)* Tutorial, September 2014. [Online]. Available: http://www.eusipco2014.org/program/tutorials/

[10] S. Nie, G. R. MacCartney, S. Sun, and T. S. Rappaport, "28 GHz and 73 GHz signal outage study for millimeter wave cellular and backhaul communications," in *IEEE Proc. International Conference on Communications (ICC)*, 2014, pp. 4856–4861.

[11] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[12] X. Zhang and J. G. Andrews, "Downlink cellular network analysis with multi-slope path loss models," *IEEE Transactions on Communications*, vol. 63, no. 5, pp. 1881–1894, 2015.

[13] A. Farhang, N. Marchetti, L. E. Doyle, and B. Farhang-Boroujeny, "Filter bank multicarrier for massive MIMO," in *IEEE Proc. 80th Vehicular Technology Conference (VTC)*, 2014, pp. 1–7.

[14] N. Michailow, I. Gaspar, S. Krone, M. Lentmaier, and G. Fettweis, "Generalized frequency division multiplexing: Analysis of an alternative multi-carrier technique for next generation cellular systems," in *IEEE Proc. 9th International Symposium on Wireless Communication Systems (ISWCS)*, 2012, pp. 171–175.

[15] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[16] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing - the large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[17] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *IEEE Proc. 77th Vehicular Technology Conference (VTC)*, 2013, pp. 1–5.

[18] H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 9, pp. 1721–1737, 2014.

[19] G. Golden, C. Foschini, R. A. Valenzuela, and P. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *IET Electronics Letters*, vol. 35, no. 1, pp. 14–16, 1999.

[20] M. Silva and F. A. Monteiro, *MIMO Processing for 4G and Beyond: Fundamentals and Evolution*. CRC Press, 2014.

[21] H. Huang, C. B. Papadias, and S. Venkatesan, *MIMO Communication for Cellular Networks*. Springer, 2011.

[22] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 684–702, 2003.

[23] M.-A. Khalighi, J. Brossier, G. Jourdain, and K. Raoof, "Water filling capacity of Rayleigh MIMO channels," in *IEEE Proc. 12th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2001, pp. 155–158.

[24] D. Bai *et al.*, "LTE-advanced modem design: Challenges and perspectives," *Communications Magazine*, vol. 50, no. 2, pp. 178–186, 2012.

[25] R. de Lamare, "Massive MIMO systems: Signal processing challenges and research trends," *URSI Radio Science Bulletin*, no. 347, 2013. [Online]. Available: http://arxiv.org/abs/1310.7282

[26] J. Hoydis, "Massive MIMO and HetNets: Benefits and challenges," slides presented in *Newcom Summer School on Interference Management for Tomorrow's Wireless Networks*, 2013. [Online]. Available: http://www.laneas.com/sites/default/files/attachments-11/newcom_presentation.pdf

[27] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, pp. 1436–1449, 2013.

[28] A. Chockalingam and B. S. Rajan, *Large MIMO Systems*. Cambridge University Press, 2014.

[29] P. Almers, F. Tufvesson, and A. F. Molisch, "Measurement of keyhole effect in a wireless multiple-input multiple-output (MIMO) channel," *IEEE Communications Letters*, vol. 7, no. 8, pp. 373–375, 2003.

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

[31] M. Di Renzo, H. Haas, A. Ghrayeb, S. Sugiura, and L. Hanzo, "Spatial modulation for generalized MIMO: challenges, opportunities, and implementation," *Proceedings of the IEEE*, vol. 102, no. 1, pp. 56–103, 2014.

[32] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?" in *IEEE 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2011, pp. 545–550.

[33] N. Krishnan, R. Yates, and N. B. Mandayam, "Uplink linear receivers for multi-cell multiuser MIMO with Pilot contamination: Large system analysis," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4360–4373, 2014.

[34] E. Bjornson, E. Jorswieck, M. Debbah, and B. Ottersten, "Multiobjective Signal Processing Optimization: The way to balance conflicting metrics in 5G systems," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 14–23, 2014.

[35] E. Luther, "5G Massive MIMO Testbed : From Theory to Reality," Technical Report, National Instruments, 2014. [Online]. Available: http://www.ni.com/white-paper/52382/en/#toc

[36] J. Vieira, S. Malkowsky, K. Nieman, Z. Miers, N. Kundargi, L. Liu, I. Wong, V. Owall, O. Edfors, and F. Tufvesson, "A flexible 100-antenna testbed for Massive MIMO," in *IEEE Proc. GLOBECOM Workshop on Massive MIMO: from theory to practice*, 2014, pp. 287–293.

[37] C. Elliott, V. Vijayakumar, W. Zink, and R. Hansen, "National instruments LabVIEW: a programming environment for laboratory automation and measurement," *Journal of the Association for Laboratory Automation*, vol. 12, pp. 17–24, 2007.

[38] Ericsson, "Ericsson 5G delivers 5 Gbps speeds," Press Release, 2014. [Online]. Available: http://www.ericsson.com/news/1810070

[39] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear precoding performance in measured very-large MIMO channels," in *IEEE Proc. 74th Vehicular Technology Conference (VTC Fall)*, 2011, pp. 1–5.

[40] J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, "Channel measurements for large antenna arrays," in *IEEE Proc. 9th International Symposium on Wireless Communication Systems (ISWCS)*, 2012, pp. 811–815.

[41] X. Gao, F. Tufvesson, O. Edfors, and F. Rusek, "Measured propagation characteristics for very-large MIMO at 2.6 GHz," in *IEEE Proc. 46th Asilomar Conference on Signals, Systems, and Computers*, 2012, pp. 295–299.

[42] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," submitted to *IEEE Transactions on Wireless Communcations*, 2015. [Online]. Available: http://arxiv.org/abs/1403.3376

[43] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *ACM Proc. 18th Annual International Conference on Mobile Computing and Networking (MobiCom)*, 2012, pp. 53–64.

[44] J. Nielsen, J. B. Andersen, P. C. F. Eggers, G. F. Pedersen, K. Olesen, and H. Suda, "Measurements of indoor 16×32 wideband MIMO channels at 5.8 GHz," in *IEEE Proc. 8th International Symposium on Spread Spectrum Techniques and Applications (ISSSTA)*, 2004, pp. 864–868.

[45] J. Zheng, X. Gao, Z. Zhang, and Z. Feng, "A compact eighteen-port antenna cube for MIMO systems," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 2, pp. 445–455, 2012.

[46] V. Nurmela *et al.*, "Initial channel models based on measurements," Deliverable, Project Number ICT-317669, METIS, 2014. [Online]. Available: https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D1.2_v1.pdf

[47] X. Ma and W. Zhang, "Fundamental limits of linear equalizers: diversity, capacity, and complexity," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3442–3456, 2008.

[48] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, 2003.

[49] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2201–2214, 2002.

[50] S. Verdú, "Computational complexity of optimum multiuser detection," *Algorithmica*, vol. 4, no. 1, pp. 303–312, 1989.

[51] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[52] R. Penrose, "A generalized inverse for matrices," in *Mathematical proceedings of the Cambridge philosophical society*, vol. 51, no. 03, 1955, pp. 406–413.

[53] Y. Jiang, M. K. Varanasi, and J. Li, "Performance analysis of ZF and MMSE equalizers for MIMO systems: an in-depth study of the high SNR regime," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2008–2026, 2011.

[54] X. Ma and W. Zhang, "Performance analysis for MIMO systems with lattice-reduction aided linear equalization," *IEEE Transactions on Communications*, vol. 56, no. 2, pp. 309–318, 2008.

[55] E. N. Onggosanusi, A. G. Dabak, T. Schmidl, and T. Muharemovict, "Capacity analysis of frequency-selective MIMO channels with sub-optimal detectors," in *IEEE Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, 2002, pp. 2369 –2372.

[56] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical distributions*, 4th ed. John Wiley & Sons, 2011.

[57] X. Zhang and S.-Y. Kung, "Capacity analysis for parallel and sequential MIMO equalizers," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2989–3002, 2003.

[58] P. W. Wolniansky, G. J. Foschini, G. Golden, R. Valenzuela *et al.*, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *URSI Proc. International Symposium on Signals, Systems, and Electronics (ISSSE)*, 1998, pp. 295–300.

[59] S. Loyka and F. Gagnon, "Performance analysis of the V-BLAST algorithm: an analytical approach," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1326–1337, 2004.

[60] D. Wubben, D. Seethaler, J. Jaldén, and G. Matz, "Lattice reduction," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 70–91, 2011.

[61] F. Oggier and E. Viterbo, "Algebraic number theory and code design for Rayleigh fading channels," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 3, pp. 333–415, 2004.

[62] D. Aharonov and O. Regev, "Lattice problems in NP ∩ coNP," *Journal of the ACM*, vol. 52, no. 5, pp. 749–765, 2005.

[63] B. Helfrich, "Algorithms to construct Minkowski reduced and Hermite reduced lattice bases," *Elsevier Theoretical Computer Science*, vol. 41, pp. 125–139, 1985.

[64] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "LLL reduction achieves the receive diversity in MIMO decoding," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4801–4805, 2007.

[65] D. Seethaler, G. Matz, and F. Hlawatsch, "Low-complexity MIMO data detection using Seysen's lattice reduction algorithm," in *IEEE Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, 2007, pp. 53–56.

[66] L. Bai and J. Choi, *Low Complexity MIMO Detection*. Springer, 2012.

[67] B. Gestner, W. Zhang, X. Ma, and D. V. Anderson, "Lattice reduction for MIMO detection: from theoretical analysis to hardware realization," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 4, pp. 813–826, 2011.

[68] Q. Zhou and X. Ma, "Element-based lattice reduction algorithms for large MIMO detection ," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 274–286, 2013.

[69] ——, "Designing low-complexity detectors for generalized SC-FDMA systems," in *IEEE Proc. 45th Annual Conference on Information Sciences and Systems (CISS)*, 2011, pp. 1–6.

[70] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1474–1484, 2005.

[71] T. Stützle, "Local search algorithms for combinatorial problems," Ph.D. dissertation, 1998. [Online]. Available: http://iridia.ulb.ac.be/~stuetzle/publications/Thesis.ThomasStuetzle.pdf

[72] S. K. Mohammed, A. Chockalingam, and B. Sundar Rajan, "A low-complexity near-ML performance achieving algorithm for large MIMO detection," in *IEEE Proc. International Symposium on Information Theory (ISIT)*, 2008, pp. 2012–2016.

[73] P. Li and R. D. Murch, "Multiple output selection-LAS algorithm in large MIMO systems," *IEEE Communications Letters*, vol. 14, no. 5, pp. 399–401, 2010.

[74] A. Kumar, S. Chandrasekaran, A. Chockalingam, and B. S. Rajan, "Near-optimal large-MIMO detection using randomized MCMC and randomized search algorithms," in *IEEE Proc. International Conference on Communications (ICC)*, 2011, pp. 1–5.

[75] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Wiley European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[76] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1893–1909, 2004.

[77] A. K. Gupta and D. K. Nagar, *Matrix variate distributions*. CRC Press, 1999.

[78] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge University Press, 2011.

[79] H. Prabhu, O. Edfors, J. Rodrigues, L. Liu, and F. Rusek, "Hardware efficient approximative matrix inversion for linear pre-coding in massive MIMO," in *IEEE Proc. International Symposium on Circuits and Systems (ISCAS)*, 2014, pp. 1700–1703.

[80] H. Prabhu, J. Rodrigues, O. Edfors, and F. Rusek, "Approximative matrix inverse computations for very-large MIMO and applications to linear pre-coding systems," in *IEEE Proc. Wireless Communications and Networking Conference (WCNC)*, 2013, pp. 2710–2715.

[81] J. C. Ikuno, M. Wrulich, and M. Rupp, "System level simulation of LTE networks," in *IEEE Proc. 71st Vehicular Technology Conference (VTC Spring)*, 2010, pp. 1–5.

[82] G. W. Stewart, *Matrix Algorithms, Volume 1: Basic Decompositions*. SIAM, 2001.

[83] D. Zhu, B. Li, and P. Liang, "On the matrix inversion approximation based on Neumann series in Massive MIMO systems," to appear in *IEEE Proc. International Conference on Communications (ICC)*, 2015. [Online]. Available: http://arxiv.org/abs/1503.05241

[84] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012.

[85] D. G. Feingold, R. S. Varga *et al.*, "Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem," *Pacific Journal of Mathematics*, vol. 12, no. 4, pp. 1241–1250, 1962.

[86] K. D. Cooper, L. T. Simpson, and C. A. Vick, "Operator strength reduction," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 23, no. 5, pp. 603–625, 2001.

[87] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing, Third Edition*. Cambridge University Press, 2007.

[88] A. Burian, J. Takala, and M. Ylinen, "A fixed-point implementation of matrix inversion using Cholesky decomposition," in *IEEE Proc. 46th Midwest Symposium on Circuits and Systems (MWSCAS)*, vol. 3, 2003, pp. 1431–1434.

[89] M. Wu, B. Yin, G. Wang, C. Dick, J. R. Cavallaro, and C. Studer, "Large-Scale MIMO Detection for 3GPP LTE: Algorithms and FPGA Implementations," *IEEE J. Sel. Top. Sign. Proces.*, vol. 8, pp. 916–929, 2014.

[90] T. Minka, "The lightspeed Matlab toolbox," Toolbox, 2014. [Online]. Available: http://research.microsoft.com/en-us/um/people/minka/software/lightspeed/

[91] W. W. Hager, "Updating the inverse of a matrix," *SIAM review*, vol. 31, no. 2, pp. 221–239, 1989.

[92] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, Gatsby Computational Neuroscience Unit, University College London, UK, 2003.

[93] K. S. Riedel, "A Sherman-Morrison-Woodbury identity for rank augmenting matrices with application to centering," *SIAM journal on matrix analysis and applications*, vol. 13, no. 2, pp. 659–662, 1992.

[94] S. Liu, C. Ling, and D. Stehlé, "Decoding by sampling: a randomized lattice algorithm for bounded distance decoding," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 5933–5945, 2011.

[95] C. Ling, "Approximate lattice decoding: Primal versus dual basis reduction," in *IEEE Proc. International Symposium on Information Theory (ISIT)*, 2006, pp. 1–5.

[96] W. Zhang, X. Ma, and A. Swami, "Designing low-complexity detectors based on Seysen's algorithm," *IEEE Transactions on Wireless Communications*, vol. 9, no. 10, pp. 3301–3311, 2010.

[97] Q. Zhou and X. Ma, "Improved element-based lattice reduction algorithms for wireless communications," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4414–4421, 2013.

[98] Y. Shang and X.-G. Xia, "On fast recursive algorithms for V-BLAST with optimal ordered SIC detection," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 2860–2865, 2009.

[99] B. Hassibi, "An efficient square-root algorithm for BLAST," in *IEEE Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, 2000, pp. 737–740.

[100] D. Wubben, R. Bohnke, V. Kuhn, and K.-D. Kammeyer, "MMSE extension of V-BLAST based on sorted QR decomposition," in *IEEE Proc. 58th Vehicular Technology Conference (VTC)*, vol. 1, 2003, pp. 508–512.

[101] X. Wang and H. V. Poor, *Wireless communication systems: Advanced techniques for signal reception*. Prentice Hall Professional, 2004.

[102] D. MacKay, *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, 2002.

[103] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 721–741, 1984.

[104] B. Hassibi, M. Hansen, A. G. Dimakis, H. A. J. Alshamary, and W. Xu, "Optimized Markov Chain Monte Carlo for Signal Detection in MIMO Systems: An Analysis of the Stationary Distribution and Mixing Time," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4436–4450, 2014.

[105] P. Klein, "Finding the closest lattice vector when it's unusually close," in *SIAM Proc. 11th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2000, pp. 937–941.

[106] Z. Wang, C. Ling, and G. Hanrot, "Markov chain Monte Carlo algorithms for lattice Gaussian sampling," in *IEEE Proc. International Symposium on Information Theory (ISIT)*, 2014, pp. 1489–1493.

[107] T. Datta, N. A. Kumar, A. Chockalingam, and B. S. Rajan, "A novel Monte-Carlo-sampling-based receiver for large-scale uplink multiuser MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3019–3038, 2013.

[108] C. Studer, M. Wenk, A. Burg, and H. Bolcskei, "Soft-output sphere decoding: Performance and implementation aspects," in *IEEE Proc. 40th Asilomar Conference on Signals, Systems and Computers (ACSSC)*, 2006, pp. 2071–2076.

[109] B. Farhang-Boroujeny, H. Zhu, and Z. Shi, "Markov chain Monte Carlo algorithms for CDMA and MIMO communication systems," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1896–1909, 2006.

[110] M. Wu, Y. Sun, S. Gupta, and J. R. Cavallaro, "Implementation of a high throughput soft MIMO detector on GPU," *Journal of Signal Processing Systems*, vol. 64, no. 1, pp. 123–136, 2011.

[111] J. S. Lemos, F. Rosário, F. A. Monteiro, J. Xavier, and A. Rodrigues, "Massive MIMO full-duplex relaying with optimal power allocation for independent multipairs," in *IEEE Proc. 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 306–310.

[112] A. C. Cirik, "On duality of MIMO relays and performance limits of full-duplex MIMO radios," Ph.D. dissertation, University of California, Riverside, 2014. [Online]. Available: http://www.escholarship.org/uc/item/6386f0p3

[113] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE Journal On Selected Areas In Communications*, vol. 32, no. 9, pp. 1637–1652, 2014.

[114] M. Heino *et al.*, "Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays," *IEEE Communications Magazine*, vol. 53, no. 5, pp. 91–101, 2015.

[115] J. Kim and I. Lee, "802.11 WLAN: history and new enabling MIMO techniques for next generation standards," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 134–140, 2015.

[116] C.-H. Wu, W.-H. Chung, and C.-E. Chen, "MMSE-based precoder design in nonregenerative relay systems with direct link," in *IEEE Proc. 77th Vehicular Technology Conference (VTC Spring)*, 2013, pp. 1–5.

[117] J. Joung and A. H. Sayed, "Multiuser two-way amplify-and-forward relay processing and power control methods for beamforming systems," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1833–1846, 2010.

[118] H. Liu, H. Gao, and T. Lv, "Low-complexity user scheduling for MMSE relaying with large-scale arrays," *IET Electronics Letters*, vol. 51, no. 2, pp. 147–149, 2015.

[119] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 4296–4307, 2012.

[120] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5983–5993, 2011.

[121] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, no. 31, pp. 172–179, 2013.

[122] T. Riihonen, S. Werner, and R. Wichman, "Transmit power optimization for multiantenna decode-and-forward relays with loopback self-interference from full-duplex operation," in *IEEE Proc. 45th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2011, pp. 1408–1412.

[123] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[124] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," Toolbox, 2014. [Online]. Available: http://cvxr.com/cvx

[125] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2009.

[126] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: 10 myths and one grand question," submitted to *IEEE Communications Magazine*, 2015. [Online]. Available: http://arxiv.org/abs/1503.06854

[127] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," *Technical University of Denmark*, 2008.

[128] A. M. Tulino and S. Verdú, *Random matrix theory and wireless communications*. Now Publishers, 2004.

[129] J. Seo and J. Paik, "A full-duplex MIMO relay with an MMSE filter and a modified Alamouti encoder against spatial loop interference in wireless multihop networks," *SERSC International Journal of Software Engineering and Its Applications*, vol. 6, no. 3, pp. 17–26, 2012.

[130] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2740–2746, 2007.

Appendix

Appendix A - Normalisation Factor $\alpha_{\mathbf{zf}}$

The objective is to design α_{zf} such that $\mathbb{E}\{\mathbf{t}^H\mathbf{t}\} = 1$. Using its definition $\mathbf{t} = \alpha_{\text{zf}}\tilde{\mathbf{G}}_{\text{RD}}^H(\tilde{\mathbf{G}}_{\text{RD}}\tilde{\mathbf{G}}_{\text{RD}}^H)^{-1}\hat{\mathbf{x}} = \alpha_{\text{zf}}\mathbf{P}\hat{\mathbf{x}}$, and the fact that $\hat{\mathbf{x}}$ is zero-mean with covariance matrix $\mathbf{C} = \mathbb{E}\{\hat{\mathbf{x}}\hat{\mathbf{x}}^H\} = \mathbf{I}$, it follows that (from [127])

$$\mathbb{E}\{\mathbf{t}^H\mathbf{t}\} = \alpha_{\text{zf}}^2\mathbb{E}\{(\mathbf{P}\hat{\mathbf{x}})^H(\mathbf{P}\hat{\mathbf{x}})\} = \alpha_{\text{zf}}^2 tr\{\mathbf{P}\mathbf{C}\mathbf{P}^H\} = 1. \tag{1}$$

Exploiting $\mathbf{C} = \mathbf{I}$, it follows that $tr\{\mathbf{P}\mathbf{C}\mathbf{P}^H\}$ is given by

$$tr\{(\tilde{\mathbf{G}}_{\text{RD}}\tilde{\mathbf{G}}_{\text{RD}}^H)^{-1}\} = tr\{(\mathbf{D}_{\text{RD}}\tilde{\mathbf{H}}_{\text{RD}}\tilde{\mathbf{H}}_{\text{RD}}^H)^{-1}\}. \tag{2}$$

Now noting that $\mathbf{W} = \tilde{\mathbf{H}}_{\text{RD}}\tilde{\mathbf{H}}_{RD}^H$ is a central Wishart matrix, where the columns of $\tilde{\mathbf{H}}_{\text{RD}} \in \mathbb{C}^{K\times N_{tx}}$ are zero-mean complex Gaussian vectors with covariance matrix $(1+\epsilon_{\mathbf{H}}^2)\mathbf{I}$, it follows from [128, Lemma 2.10] that

$$\mathbb{E}\{(tr\{\mathbf{D}_{\text{RD}}\mathbf{W}\})^{-1}\} = \frac{\sum_{k=1}^{K}\left(\beta_{\text{RD,k}}(1+\epsilon_{\mathbf{H}}^2)\right)^{-1}}{N_{tx}-K}. \tag{3}$$

Merging the results in (1), (2) and (3), the normalisation factor α_{zf} in (6.6) follows.

Appendix B - $\mathbf{F_{tx}}$ and $\mathbf{F_{rx}}$ Expressions

According to [129], the error covariance matrix $\mathbf{Q}$ of the relay input signal is given by $\mathbb{E}((\mathbf{G}_{\text{SR}}\mathbf{x} - \hat{\mathbf{r}})(\mathbf{G}_{\text{SR}}\mathbf{x} - \hat{\mathbf{r}})^H)$, yielding

$$\begin{aligned}\mathbf{Q} =& (\mathbf{I} - \mathbf{F}_{\text{rx}})\tilde{\mathbf{G}}_{\text{SR}}\mathbf{D}_{p_{\text{S}}}\tilde{\mathbf{G}}_{\text{SR}}^H(\mathbf{I} - \mathbf{F}_{\text{rx}})^H \\ &+ p_{\text{R}}\mathbf{F}_{\text{rx}}\tilde{\mathbf{H}}_{\text{LI}}\mathbf{R}_t\tilde{\mathbf{H}}_{\text{LI}}^H\mathbf{F}_{\text{rx}}^H + \mathbf{F}_{\text{rx}}\mathbf{R}_{\mathbf{n}_r}\mathbf{F}_{\text{rx}}^H,\end{aligned} \tag{4}$$

where $\mathbf{R}_{\mathbf{t}} = \mathbf{R}_{\tilde{\mathbf{t}}} + \mathbf{R}_{\mathcal{E}_{\mathbf{t}}} = \mathbb{E}\{\tilde{\mathbf{t}}\tilde{\mathbf{t}}^H\} + \mathbb{E}\{\mathcal{E}_{\mathbf{t}}\mathcal{E}_{\mathbf{t}}^H\} = \mathbf{F}_{\text{tx}}\mathbf{A}_{\text{zf}}\mathbf{R}_{\hat{\mathbf{x}}}\mathbf{A}_{\text{zf}}^H\mathbf{F}_{\text{tx}}^H + \epsilon_{\mathbf{t}}^2\mathbf{I}$. Using the fact that the function $f = tr\{\mathbf{Z}\mathbf{A}_0\mathbf{Z}^H\mathbf{A}_1\}$ has derivative given by $\frac{d}{d\mathbf{Z}^*}f = \mathbf{A}_1\mathbf{Z}\mathbf{A}_0$ [130, Table 4], one may easily

find $\frac{\partial}{\partial \mathbf{F}_{\text{tx}}^*} tr\{\mathbf{Q}\} = \mathbf{0}$ and $\frac{\partial}{\partial \mathbf{F}_{\text{rx}}^*} tr\{\mathbf{Q}\} = \mathbf{0}$. Fixing $\mathbf{F}_{\text{tx}}$ comes

$$\begin{aligned}\frac{\partial}{\partial \mathbf{F}_{\text{rx}}^*} tr\{\mathbf{Q}\} =& \mathbf{F}_{\text{rx}}(p_{\text{R}}\tilde{\mathbf{H}}_{\text{LI}}\mathbf{R}_t\tilde{\mathbf{H}}_{\text{LI}}^H + \mathbf{R}_{\mathbf{n}_r}) \\ &- (\mathbf{I} - \mathbf{F}_{\text{rx}})\tilde{\mathbf{G}}_{\text{SR}}\mathbf{D}_{p_{\text{S}}}\tilde{\mathbf{G}}_{\text{SR}}^H = \mathbf{0},\end{aligned} \tag{5}$$

which when solved with respect to $\mathbf{F}_{\text{rx}}$ gives (6.8). Fixing $\mathbf{F}_{\text{rx}}$ leads to

$$\begin{aligned}\frac{\partial}{\partial \mathbf{F}_{\text{tx}}^*} tr\{\mathbf{Q}\} =& \mathbf{D}_{p_{\text{S}}}\mathbf{F}_{\text{rx}}\tilde{\mathbf{H}}_{\text{LI}}(p_{\text{R}}\mathbf{F}_{\text{rx}}^H\tilde{\mathbf{H}}_{\text{LI}}^H\mathbf{F}_{\text{tx}}\mathbf{A}_{zf}\mathbf{R}_{\hat{\mathbf{x}}}\mathbf{A}_{\text{zf}}^H) \\ =& \mathbf{D}_{p_{\text{S}}}\mathbf{F}_{\text{rx}}\tilde{\mathbf{H}}_{\text{LI}}(\mathbf{F}_{\text{rx}}\tilde{\mathbf{H}}_{\text{LI}}\mathbf{F}_{\text{tx}})\mathbf{A}_{\text{zf}}\mathbf{R}_{\hat{\mathbf{x}}}\mathbf{A}_{\text{zf}}^H = \mathbf{0},\end{aligned} \tag{6}$$

whose solution is given by (6.7).

Appendix C - Uniform Discrete Distribution

The objective is to find different values for $R_{0,k}$, for all k, such that a desired sum-rate S is satisfied, i.e., $S = \sum_{k=1}^{K} R_{0,k}$. In this work, a very simple model for $R_{0,k}$ will be considered. In future work, more advanced and realistic values and model can be integrated. Denote a given mean value for the rates $M = S/K$ and set a maximum deviation p, i.e., how far are the individual rates $R_{0,k}$ allowed to be from M. The individual rates are computed as follows

$$\tilde{R}_{0,k} = M(s_k + 1), \tag{7}$$

where $s_k \sim \lceil pU[0,1] \rceil$. The ceiling function is used to discretise the rates. Finally, in order to satisfy the condition $S = \sum_{k=1}^{K} R_{0,k}$, the following normalisation is applied

$$R_{0,k} = \frac{S}{\sum_{k=1}^{K} \tilde{R}_{0,k}} \tilde{R}_{0,k}. \tag{8}$$

The results for $R_{0,k}$ presented in table 6.1 were obtained using $S = 5$ and $p = 10\%$.