Design and Analysis of Antenna Selection in Massive MIMO

GUILLERMO JIMÉNEZ ALONSO
MASTER´S THESIS
DEPARTMENT OF ELECTRICAL AND INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING | LTH | LUND UNIVERSITY

Design and Analysis of Antenna Selection in Massive MIMO

Guillermo Jiménez Alonso
gu2507ji-s@student.lu.se

Supervisor: Harish Venkatraman Bhat, Ericsson
Xuesong Cai, EIT (LTH)
Miguel A. Salas Natera, ETSIT (UPM)
Examiner: Buon Kiong Lau, EIT (LTH)

LTH
FACULTY OF ENGINEERING

ERICSSON

Department of Electrical and Information Technology (EIT)
Lund University
Lund, Sweden 2024

Printed by Sweden
Tryckeriet i E-huset, Lund

Abstract

Mobile communications have evolved with the commercial deployment of 5G and initial studies of 6G. New solutions and innovative technologies are being incorporated to reach the target requirements. One of the key technologies in mobile networks is massive multiple-input-multiple-output (MIMO). Increasing the number of antennas offers significant benefits in coverage, reliability, and spectral efficiency. However, the high cost, power consumption, and increased complexity of deploying large antenna arrays at the base stations (BS) could become a burden in real scenarios. Antenna selection (AS) could be a potential solution to alleviate the large number of radio frequency (RF) chains. This master thesis addresses the AS problem where a high-quality subset of the available transmit antennas at the BS are active whereas the rest of the antennas are switched off. The AS criterion is to select the submatrix with the largest Minimum Singular Value (MSV) maximizing the Signal to Noise Ratio (SNR) so the Bit Error Rate (BER) is decreased. Fully flexible full-array switching (FAS) and sub-array switching (SAS) networks are considered. Exhaustive search becomes unfeasible even as the number of antennas grows therefore sub-optimal and optimal alternatives are needed. Greedy and Branch-and-Bound (BAB) searching algorithms for both switching networks are introduced and compared in terms of complexity and relative error for obtaining the largest MSV. Simulations carried out for scenarios with different antenna configurations show that all algorithms improve the BER compared to random antenna selection. However, the gain reduces as the percentage of selected antennas to the total available antennas increases. Regarding capacity, most of the algorithms obtained better ergodic capacity than random antenna selection. In addition, there is no significant difference in the BER performance between algorithms so when there are many antennas at the BS, sub-optimal approaches might be preferred to save computation time.

Keywords: 5G, Massive MIMO, Antenna Selection, BER, Branch-and-Bound, Full-Array Switching, Sub-Array Switching, Minimum Singular Value.

Acknowledgements

First and foremost, I would like to thank Harish Venkatraman, my supervisor, and David Lundberg for allowing me to undertake this master thesis at Ericsson and for their help, guidance, and confidence in me through this process. I express my gratitude to Xuesong Cai at Lund University for his valuable comments and help at any time. I would also like to thank Juan Guirado too, for his advice and occasional coffee breaks.

I would also like to mention all the friends I have made along the way. I cannot help but look back and reflect on the wonderful experiences I have had over two years here. More than an academic achievement, it has been a journey of personal growth. For me, Lund has meant going with all of you by my side. Thank you all!

Finally, I want to dedicate this work to my old friends from many years ago and, of course, to my family. To all my grandparents who, unfortunately, I don't see them so often now, and especially to my parents, Ana María and Juan Carlos, and my sister Elena. They are my greatest source of inspiration and reason for being the way I am. Thank you for your strength, support, and unconditional love because if not, I would not be where I am nor would I have achieved all this.

Guillermo Jiménez Alonso

Popular Science Summary

The fifth generation (5G) mobile communication systems have already been deployed with advanced features to meet more demanding requirements and use cases. These new use cases and applications are made possible by the emergence of smartphones and more powerful devices that connect to the network and request resources, so traffic demand has increased as well. It is even expected that new applications and equipment will arrive creating new requirements not only related to traffic but also to delay, reliability, and energy efficiency.

One relevant technical solution to handle current and future challenges is to have a large number of antennas at the base station, which is called massive multiple-input multiple-output (MIMO). Massive MIMO is being used in 5G base stations with a moderate number of antennas (e.g., 32 transmitting and receiving antennas). It is envisioned that in the future, base stations will have an array of hundreds or even thousands of antennas to support new requirements and serve several users with the same resources. However, this technology also comes with some problems that need to be addressed to successfully deploy large antenna arrays. Those problems are related to hardware cost, energy consumption, and computational demands or complexity. This master thesis introduces and proposes the concept of antenna selection (AS) as a solution to reduce the difficulty in implementing large antenna arrays.

The idea behind AS is simple. It consists of choosing and using a subset of antennas from all the antennas at the base station. In this way, the rest of the antennas can be switched off in certain situations when they are not necessary while taking advantage of the abundance of antennas in a flexible way. In addition, if we have an excess of antennas, the hardware components and its circuitry can be reduced, and thus the complexity as well as the energy consumption and cost while retaining most of the performance offered by the unreduced massive MIMO system. The subset is intelligently selected and represents the *best* antennas that in this case will be used to transmit. Thus, antenna selection needs a criteria or objective function based on the signal quality or wireless channel experienced by each antenna. The AS problem is even more challenging in massive MIMO systems because exhaustive search is infeasible for a large number of antennas. In this master thesis, some faster sub-optimal and optimal AS algorithms are described and their performances in terms of bit error rate (BER) for different antenna configurations are analyzed.

The results show that AS can help to optimize performance and reduce complexity when not all the antennas at the base station are required to be active. By selecting the best, or even a good enough subset of antennas, we can still harness to a large extent the full potential of massive MIMO technology within 5G networks by balancing power consumption, complexity, and performance.

List of Acronyms and Abbreviations

Below is the list of acronyms listed in alphabetical order:

3GPP	3rd Generation Partnership Project
5G	5th Generation of Mobile Broadband Communications
ADC	Analog-to-Digital Converter
AS	Antenna Selection
AWGN	Additive White Gaussian Noise
BAB	Branch-and-Bound
BER	Bit Error Rate
BF	Beamforming
BS	Base Station
CSI	Channel State Information
DAC	Digital-to-Analog Converter
DL	Downlink
FAS	Full-Array Switching
FDD	Frequency-Division Duplex
LS	Least Squares
LTE	Long-Term Evolution
LoS	Line-of-Sight
MF	Matched Filter
MIMO	Multiple-Input-Multiple-Output
MISO	Multiple-Input-Single-Output
MMSE	Minimum Mean-Square Error
MRT	Maximum-Ratio Transmission
MSV	Minimum Singular Value
NR	New Radio
OFDM	Orthogonal Frequency-Division Multiplexing
RF	Radio-Frequency
Rx	Receiver
RZF	Regularized Zero-Forcing
SAS	Sub-Array Switching
SIMO	Single-Input-Multiple-Output
SISO	Single-Input-Single-Output
SNR	Signal-to-Noise Ratio
TDD	Time-Division Duplex
Tx	Transmitter
UE	User Equipment
ULA	Uniform Linear Array
UL	Uplink
UMi	Urban Microcell
ZF	Zero-Forcing

Notation

Below is the mathematical notation that have been used throughout this master thesis.

$*$	Convolution operator
a	Scalar
$\|a\|$	Magnitude of the scalar a
$\hat{a}$	Estimate of a
$\mathbf{a}$	Column vector
$\mathbf{A}$	Matrix
$\mathbf{A}^H$	Hermitian transpose or conjugate transpose of $\mathbf{A}$
$\text{diag}(\mathbf{a})$	Diagonal matrix with the elements of $\mathbf{a}$ in its main diagonal
$\|\mathbf{A}\|_F^2$	Squared Frobenius norm of $\mathbf{A}$
$\det(\mathbf{A})$	Determinant of $\mathbf{A}$
$\text{vec}(\mathbf{A})$	Transforms a matrix to a column vector, stacking its elements column-wise
$\mathbf{A}_{i,j}$	i,j-th element of matrix $\mathbf{A}$
$\mathbf{I}_M$	$M \times M$ identity matrix
$\arg$	Argument
$\max, \min$	Maximum and minimum
$\delta(x)$	Dirac delta function
$\mathbb{E}\{\cdot\}$	Expectation operator
$\mathcal{S}$	Set of elements
$\|\mathcal{S}\|$	Cardinality of $\mathcal{S}$
$\mathbb{C}$	Set of complex numbers

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The current world is digital, and it is becoming more connected than ever before. Conventional devices such as smartphones or computers are our daily tools while new ones are slowly entering the ecosystem. The progressive implementation of IoT (Internet of Things), robots, autonomous vehicles, and even Artificial Intelligence (AI) as well as more use cases that are being conceived, will increase the number of connections to wireless networks. This growth in devices and the development of more demanding applications with specific requirements make it clear that the existing and future telecommunication infrastructures should be able to support very high throughput and low latency to comply with new services.

1.1 Background and Motivation

Mobile communications have evolved considerably since the first generation, employing different multiple access techniques such as time-division, code-division, and orthogonal frequency-division multiple access (TDMA, CDMA, and OFDMA respectively). In LTE, orthogonal frequency-division multiplexing (OFDM) and multiple-input-multiple-output (MIMO) were introduced as promising technologies to enhance throughput and meet other requirements.

The fifth generation (5G) New Radio (NR) was introduced in 3GPP Release 15 in 2018, and it was fully specified in 2019 [1]. In that one and successive releases, although some useful previous technologies were adopted and improved, such as MIMO and OFDM, new technical specifications were included, e.g., frequency bands, waveform designs, and network architecture. The first 5G networks are already deployed and commercial devices are supporting NR. 5G NR is currently reaching maturity with 5G subscriptions growing worldwide. The original capabilities of 5G in the areas of enhanced Mobile Broadband (eMBB), ultra-reliable low latency communication (URLLC), and massive Machine Type Communication (mMTC) are currently being improved in 5G Advanced specified in the latest 3GPP Release 18. This and future releases will contain the technology baselines for the sixth generation (6G) standardization effort expected to begin this year.

Multi-antenna or MIMO technologies, where both the transmitter (Tx) and receiver (Rx) are equipped with more than one antenna, have been studied for decades. MIMO and beamforming (BF) have been used commercially in LTE deployments and their evolution, massive MIMO, was adopted in 5G/NR since the start. Massive MIMO employs a very large number of antennas at the base station (BS) and serves multiple antennas or several user equipment (UEs) simultaneously, commonly in the context of Multi-User MIMO (MU-MIMO). In massive MIMO, it is considered that the number of antennas at the base station is in the order of hundreds, and the number of users or their total number of antennas is lower by at least one order of magnitude.

Massive MIMO has the potential to multiplex data streams exploiting the spatial domain and increasing the capacity. Hence, this solution provides more spatial dimensions and better resolution relying on spatial multiplexing [2]. More degrees of freedom are available due to the extra antennas, and the propagation channel can be further exploited to compensate for negative effects such as propagation losses, hardware constraints, or the use of linear filters. Massive MIMO improves the spectral efficiency and energy efficiency compared to conventional MIMO systems.

However, there is a price to pay for massive MIMO, and there exist several challenges that must be overcome. The cost and complexity of the hardware increase as a result of manufacturing and implementing a large number of RF chains, up/down converters, and analog-to-digital converters (ADC) or digital-to-analog converters (DAC). Besides that, the power consumption from power amplifiers and circuits should also be considered as well as the signal processing computational complexity. To alleviate these limitations, antenna selection (AS) is an attractive solution consisting of selecting a high-quality subset of antennas while the rest are switched off. AS algorithms can be applied at both sides, Tx and/or Rx. There is what is known as MIMO Sleep which based on the traffic loads, some antenna elements could be turned off if not needed so the resources being used are optimized [3]. In that sense, the required number of RF chains is substantially decreased, and thus the complexity and energy, while still maintaining a large portion of the advantages of massive MIMO.

1.2 Objectives, Research Contributions and Limitations

This master thesis aims to study AS in massive MIMO systems. For this purpose, massive MIMO scenarios with multiple antenna configurations at both the BS and the UEs have been created, AS algorithms have been formulated and implemented at the BS, and the performance obtained regarding the Bit Error Rate (BER) is studied and analyzed. A MATLAB® 5G/NR simulator that follows the 3GPP specifications for NR has been used. The simulator was developed in previous master theses, but new features and changes are introduced to achieve this thesis' particular purposes.

The research contributions of this master thesis are the study of AS algorithms under certain conditions related to the size of the channel matrix, the analysis of the performance of those algorithms, and the comparison with various configurations and scenarios when applying AS at the BS in a 5G/NR massive MIMO system. Specifically, the research questions to be answered are (1) how to obtain high-quality or optimal subsets of the total available antennas under a certain criterion, (2) what is the performance of using those subsets compared to a system without AS, and (3) what are the possible trade-offs of the implementation. The limitations encountered throughout the master thesis are the following: Firstly, an optimal algorithm for the particular case of AS implementation at the Tx has not been found in other investigations nor derived here. Then, scenarios with more than one layer, to achieve a more realistic AS operation, have not been enabled in the simulator. Finally, the computation time of the algorithms can be unreasonably long in some cases, especially those that are optimal and in scenarios with many antennas. This last limitation, compounded by the computation time of functions in the simulator, has been the limiting factor in determining the number of channel realizations to use in each simulation to maintain a fair compromise between computation time and the reliability of results.

1.3 Thesis Outline

The report is structured as follows: Chapter 2 provides a brief overview of the main concepts, aspects, and techniques to understand the basic foundations of MIMO technology on which this thesis is based. This includes a description of the system model with the wireless channel and the processing techniques used. In addition, the chapter also provides an introduction to the main technical features of the 5G NR physical layer and an explanation of the simulator operation. In Chapter 3, AS, the topic of the master thesis, is presented. Different approaches to solve this problem are shown and a comparison in terms of complexity and performance is carried out. Moving on to Chapter 4, the AS algorithms are implemented into the simulator and some scenarios are simulated and analyzed. The simulation results compare the approaches defined in the previous chapter with each other, but also with various BS and UE configurations. Finally, Chapter 5 draws the conclusions after completing this work and some possible future lines of research on the topic are suggested.

Chapter 2

Theoretical Framework

In this chapter, massive MIMO will be introduced along with the system model definition and the main features of NR/5G which are implemented into the simulator itself. This includes a description of the wireless channel adopted in the simulator and the processing techniques applied.

2.1 Massive MIMO Brief Overview

Multiple-input multiple-output (MIMO) has been studied for the past several decades, and it consists of having more than one antenna at both the transmitter Tx and the receiver Rx. If the Tx only has one antenna, we have single-input multiple-output (SIMO), and vice versa, if the Rx is the one with one antenna only, multiple-input single-output (MISO) would be the case. Under environments with sufficiently rich multipath propagation and large enough antenna spacing, the signals arrive through different paths with diverse phases and amplitudes. These various propagation paths can be exploited and, as a result, MIMO provides four main categories of advantages [2], [4]:

- Array gain, also known as beamforming gain. It refers to the increase in signal power or signal-to-noise ratio (SNR) at the receiver through the use of antenna arrays with an increased directivity. The coherent combination effect of the different signals received from different antennas enhances the resultant signal. The average signal power increase is proportional to the number of receive antennas and gives a logarithmic gain in capacity.
- Spatial diversity gain. Each receive antenna sees an independent faded version of the same signal with different amplitudes over statistically independent paths due to the rich multipath environment. Therefore, the receiver can combine these signals to reduce amplitude variations and raise reliability. This results in an exponential decline in bit error rate (BER) as the SNR grows. If the channel fades independently between each transmit-receive antenna pair, the diversity order equals the number of transmit antennas times the number of receive antennas.
- Spatial multiplexing (SM) gain. It offers a linear capacity gain equal to the minimum number of antennas at Tx or Rx, with no bandwidth or power increase. The bit stream is divided into simultaneous transmission of several sub-streams of data over different spatial modes. The receiver, with channel knowledge, can distinguish the sub-streams.
- Interference reduction. Beamforming can also be used to reduce interference to other users or co-channel signals. This can be due to narrower beams as the power is directed only to the intended receiver as well as algorithms intentionally generating null(s) towards the other user(s) and hence the Signal-to-Interference-plus-Noise Ratio (SINR) improves.

It may not be possible to have all the advantages simultaneously maximized because of the limited spatial degrees of freedom (or number of antennas). Sometimes spatial multiplexing is maximized and sometimes diversity gain. Thus, it is important to denote that there is a trade-off between how much gain from both types can be extracted [5].

The early research and studies on MIMO were mostly based on the conventional point-to-point MIMO links with a small number of antennas. In recent years, due to the increasing demand

for mobile broadband communications, MU-MIMO scenarios, where the BS communicates with a set of UEs using the same time and frequency resource, are becoming more interesting. At the beginning of the last decade, a new concept of MIMO was conceived, consisting of having a large number of antennas at the BS, to enable a practical MU-MIMO system [6]. Massive MIMO systems bring several benefits since the extra number of antennas offers more degrees of freedom. Most importantly, increasing the number of antennas significantly at the BS side facilitates nearly orthogonal channel vectors between different users, and channel hardening also occurs, meaning that the channel becomes more deterministic and the small-scale fading is averaged out.

2.2 System Model

2.2.1 MIMO Scenario

The scenario followed as a reference in this master thesis is represented in Figure 2.1. There is a BS equipped with M antennas and a set of P UEs each with K antennas. Both antenna arrays are linear, with the elements placed in a straight line, and the antenna spacing is assumed to be half a wavelength. Two links are considered: uplink (UL), from UEs to the BS and downlink (DL), from the BS to the UEs.

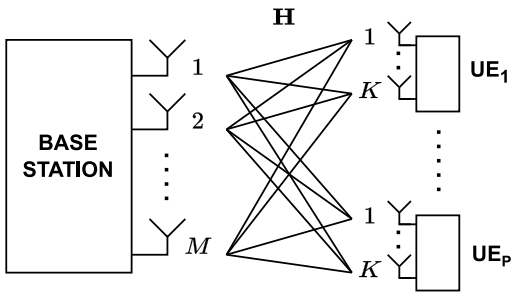


Figure 2.1: General MU-MIMO scenario. **H** corresponds to the channel matrix in frequency domain of one of the links, UL or DL. Since channel reciprocity is assumed, the channel matrix transpose corresponds to the channel of the other link.

Some general assumptions are described below.

1. Time-division duplex (TDD) operation is assumed. UL and DL transmissions are multiplexed in time and occur over different time slots but in the same frequency.
2. The channel is reciprocal. This denotes that the impulse response between any two antennas does not change for each direction, for the same time-instant and frequency range of communications. The DL channel state information (CSI) can be obtained due to channel reciprocity from the UL channel through calibration and the transposition of the uplink channel matrix. The pilots for channel estimation in the BS are sent during UL and are orthogonal.
3. During transmission, the channel is considered to remain constant, i.e., the channel coherence time is much greater than the symbol rate.
4. The equalization and precoding are fully digital and linear.

2.2.2 Wireless Channel Model

The medium through which the electromagnetic waves propagate over the air from the Tx to the Rx is known as the wireless channel. Wireless channel models are an essential part of emulating as closely as possible the physical radio channel to be able to design wireless systems and evaluate their performance. This topic is widely studied and more parameters are being added as the complexity of communication systems increases. Those models rely on mathematical representations of the channel impulse response (CIR) or the channel frequency response, containing all the necessary information to analyze and determine the transmission behavior. According to the system bandwidth, channel models can be classified as [7]:

- Wideband or frequency selective channel models. The channel is frequency selective, meaning that the response for each sub-carrier is different due to the multipath characteristics of the propagation channels.
- Narrowband or flat fading channel models. In this case, the model is not frequency selective and thus the response is the same over the system bandwidth.

Channels models can also be understood as time-variant or time-invariant. However, it is very common to assume that the channel remains approximately constant over the coherence block which is greater than the total transmission time (assumption 3), so the channel is time-invariant. When constructing channel models, there are two broad main categories:

- Deterministic channel models are based on precise descriptions and electromagnetic equations so the effect of the environment on the system is explicitly and accurately characterized at a higher computational cost, i.e., Ray Tracing techniques.
- Statistical channel models which also involve stochastic models. Stochastic models are less accurate for a particular region but they are simpler to formulate for different scenarios. They are based on the analysis of the probability density function of the channel impulse response over a large area considering the geometry of the scenario and the model is verified with field measurements data. Some examples of stochastic channel models are Rayleigh, Ricean fading, or log-normal shadowing model because they represent the channel as a random process.

In addition, standardized channel models have also been developed, most of them belonging to the geometry-based stochastic channel model (GSCMs or GBSMs) family. They aim to provide a common unified framework for the development of wireless communication systems like NR, and their evaluation and comparison. They proposed channel models for different frequency bands, or indoor and outdoor scenarios with their particularities, even including Line-Of-Sight (LOS) or Non-Line-Of-Sight (NLOS) environments. Examples of this type of channel model are the Wireless World Initiative New Radio II (WINNER II) and the Quasi Deterministic Radio Channel Generator (QuaDRiGa) [8], [9].

WINNER II, which was the previous channel model used in the simulator, is a cluster model with some good features as it uses the clustered delay line (CDL) model. It provides link level and system level simulations for different scenarios such as local, metropolitan, and wide area wireless communications, frequencies, and variable bandwidth. However, it was released in 2007 and newer channel models have appeared. QuaDRiGa, presented in 2014, is a 3D stochastic channel model that extends WINNER II. It can represent complex propagation scenarios with massive MIMO systems and the channel parameters are determined stochastically from channel measurements. QuaDRiGa fulfills all the essential 3rd Generation Partnership Project (3GPP) requirements and thus has been widely used as a 3GPP-3D compatible implementation.

3GPP is the group of organizations in charge of the standardization and technical specifications for mobile telecommunications including 5G NR. 3GPP has also developed its spatial channel model (SCM), with the latest Technical Report (TR) 38.901 Release 17 presented in 2022. The adopted channel model for the simulator is precisely the previous version, TR 38.901 Release 16

(V16.1.0) [10] from 2020, with the steps describing the channel model detailed in [11]. This paper analyzes the 3GPP 3D channel model for 5G OFDM MIMO systems and its practical implementation in MATLAB®. The reasons for changing from WINNER II to this new channel model are that the initial results with the first channel model were not satisfactory and the implementation of WINNER II in MATLAB® has some limitations. On the other hand, QuaDRiGa is a quite complete channel model but its application within the simulator could have been complex and time-consuming. The MATLAB®functions to generate the channel coefficients following [11] were applied.

This channel model uses the CDL model that groups the scatterers into a superposition of clusters. The clusters are distributed over the area and each one contains multipath components (MPC) with comparable delays and similar values of angles of departure (AoD) and arrival (AoA). A mathematical description of the channel model is given below, however, a rigorous and detailed explanation of the channel model is out of the scope of this master thesis. Therefore, interested readers are referred to [10] and [11] for further details.

2.2.3 System Model

The signal model is considered as a superposition of specular MPCs with delays (τ), Doppler frequencies (ν), azimuth (φ_T) and elevation (θ_T) AoDs, azimuth (ϕ_R) and elevation (θ_R) AoAs, and complex path gains or coefficients [12]. Considering a MIMO system with N_T static antennas at the Tx and N_R static antennas at the Rx, and N_{cl} clusters in between, the channel matrix can be expressed as

$$\mathbf{B}(t,\tau) = \sum_{l=1}^{N_{cl}} g_l(t)\, \mathbf{a}_T(\varphi_{T,l}, \phi_{T,l})\, \mathbf{a}_R^T(\varphi_{R,l}, \phi_{R,l})\, \delta(\tau - \tau_l(t)), \tag{2.1}$$

where $g_l(t)$ contains the field patterns of the receive and transmit antenna elements and the cluster powers, and $\mathbf{a}_T(\varphi, \phi)$ and $\mathbf{a}_R(\varphi, \phi)$ are the array response vectors or steering vectors of the transmit antenna array and receive antenna array respectively [11], [13]. The impulse response between the j-th transmit antenna and i-th receive antenna is expressed in (2.2), where $\alpha_{i,j}(t,l)$ is the channel gain or coefficient between the i,j antenna pair at the instant t with the delay τ. This interpretation could be translated and used in both uplink and downlink by changing the nomenclature of the number of antennas in transmission or reception as in Figure 2.1.

$$h_{i,j}(t,\tau) = \sum_{l=1}^{N_{cl}} \alpha_{i,j}(t,l)\delta(\tau - \tau_l(t)). \tag{2.2}$$

The channel gain for the particular instant t and the cluster l, can be split into two terms, $\alpha_{i,j} = \gamma_{i,j}\sqrt{\beta_p}$. There is the small-scale fading, or fast fading coefficient $\gamma_{i,j}$ which is complex, and β_p is the large-scale coefficient comprising path loss and shadow fading to a particular UE, so this term is user-dependent. The system is considered to be wideband so the delays τ_l of the N_{cl} clusters must be taken into account. As described in [2], encapsulating all the impulse responses, the matrix in (2.1) can be understood as the $N_R \times N_T$ matrix $\mathbf{B}(t,\tau)$ (or tensor if more properly defined) described in (2.3) as

$$\mathbf{B}(t,\tau) = \begin{bmatrix} h_{1,1}(t,\tau) & h_{1,2}(t,\tau) & \dots & h_{1,N_T}(t,\tau) \\ h_{2,1}(t,\tau) & h_{2,2}(t,\tau) & \dots & h_{2,N_T}(t,\tau) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R,1}(t,\tau) & h_{N_R,2}(t,\tau) & \dots & h_{N_R,N_T}(t,\tau) \end{bmatrix}. \tag{2.3}$$

Given the signal $x_j(t)$ transmitted from the j-th transmit antenna, the received signal $y_i(t)$ at the i-th received antenna is given by (2.4). The channel includes all the effects of the antenna configuration and frequency filtering. The input-output relation can also be expressed in matrix notation as in (2.5).

$$\begin{aligned} y_i(t) &= \sum_{j=i}^{N_T} h_{i,j}(t,\tau) * x_j(t) + n_i(t) \\ &= \sum_{j=1}^{N_T} \int_o^{\tau_{max}} h_{i,j}(t,\tau) x_j(t-\tau) d\tau + n_i(t), \quad i = 1, 2, \ldots, N_R \end{aligned} \tag{2.4}$$

$$\mathbf{y}(t) = \mathbf{B}(t,\tau) * \mathbf{x}(t) + \mathbf{n}(t), \tag{2.5}$$

where $\mathbf{x}(t) = [x_1(t)\, x_2(t) \ldots x_{N_T}(t)]^T \in \mathbb{C}^{N_T \times 1}$ is the transmitted signal vector which could contain pilots or data that is assumed to be random and mutually independent. The i.i.d. additive white Gaussian noise (AWGN) is represented by $\mathbf{n}(t) = [n_1(t)\ n_2(t) \ldots n_{N_R}(t)]^T \in \mathbb{C}^{N_R \times 1}$, following $\mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{N_T})$ with the variance σ_n^2. The received signal vector is $\mathbf{y}(t) = [y_1(t)\ y_2(t) \ldots y_{N_R}(t)]^T \in \mathbb{C}^{N_R \times 1}$.

It is common to make some assumptions before this step to simplify the system model. In an OFDM-like modulation with enough cyclic prefix length, each subcarrier can be processed independently, so one can focus on a single narrowband subcarrier and work with a system model in complex baseband. Or, if the channel is frequency flat, there is only one single delay, which gives the same simplification. This can also be further simplified by assuming that the channel is time-invariant. As a result, the system model would be defined by a matrix multiplication expression. As previously said, flat or frequency selective fading does not depend only on the channel but on the relationship between the system and channel characteristics, i.e., bandwidth in this case. Thus, when the system bandwidth is much larger than the coherence bandwidth ($W \gg W_c$), the channel is frequency selective and its impulse response has multiple delays. So, in this case, since the system bandwidth is 20 MHz the simplifications cannot be made.

Now, assuming that the environment in one realization (at a fixed t) is stationary (assumption 3), the attenuation coefficients and propagation delays do not depend on time t, so we have a time-invariant channel with the following impulse response as elements of $\mathbf{B}(\tau)$:

$$h_{i,j}(\tau) = \sum_{l=1}^{N_{cl}} \alpha_{i,j}(l) \delta(\tau - \tau_l). \tag{2.6}$$

Then, (2.5) reduces to (2.7) if looking at a specific frequency point, i.e., only spatial dimension being considered. The massive MIMO system model is depicted in Figure 2.2 with the precoding and equalization filters included, which will be described in Sections 2.2.5 and 2.2.6.

$$\mathbf{y} = \mathbf{B} * \mathbf{x} + \mathbf{n}. \tag{2.7}$$

When the continuous time channel is converted to a discrete time channel, the input-output relationship can be described as in (2.8). Being the length of the transmitted signal vector L_x, the receiver vector y_i will have a length of $L_x + N_{\text{taps}} - 1$ with N_{taps} being the total number of taps necessary to represent the cluster delays. The delayed replicas of the signal $x_j[k]$ due to the N_{taps} branches, also provide frequency diversity if the receiver is able to resolve the multipaths [7]. Another possible interpretation of the discrete time-invariant frequency-selective MIMO channel, in which convolution is explicitly stated, is shown in (2.9) and (2.10).

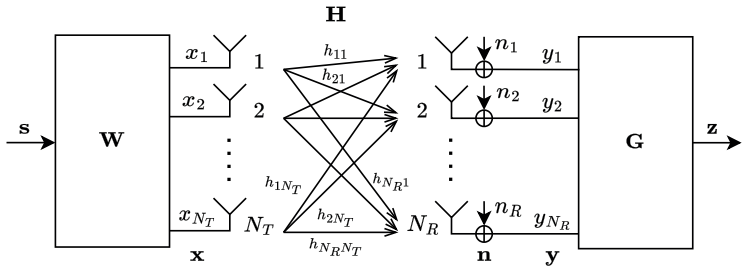


Figure 2.2: Massive MIMO system model block diagram. **W** and **G** are the precoding and equalization filters respectively.

$$y_i[k] = \sum_{j=1}^{N_T} h_{i,j}[k] * x_j[k] + n[k] = \sum_{j=1}^{N_T} \sum_{l=0}^{N_{\text{taps}}-1} h_{i,j}[l] x_j[k-l] + n_i[k] = \\ = \sum_{j=1}^{N_T} \sum_{l=0}^{N_{\text{taps}}-1} \alpha_{i,j}[l] x_j[k-l] + n_i[k], \tag{2.8}$$

$$\mathbf{y}[k] = \sum_{l=0}^{N_{\text{taps}}-1} \mathbf{B}_l \mathbf{x}[k-l] + \mathbf{n}[k], \tag{2.9}$$

$$\mathbf{B}_l = \begin{bmatrix} h_{1,1}[l] & h_{1,2}[l] & \dots & h_{1,Nt}[l] \\ h_{2,1}[l] & h_{2,2}[l] & \dots & h_{2,Nt}[l] \\ \vdots & \vdots & \ddots & \vdots \\ h_{Nr,1}[l] & h_{Nr,2}[l] & \dots & h_{Nr,Nt}[l] \end{bmatrix}. \tag{2.10}$$

2.2.4 Channel Estimation

Transmitted signals pass through the multipath channel and are distorted when reaching the receiver by many adverse effects such as multipath propagation, scattering, and interference due to the sharing of radio resources [14]. Thus, channel estimation is a critical process, to compensate for those effects and retrieve the transmitted signals and data symbols at the Rx. It is common to use pilot symbols known by both the Tx and Rx whose time-frequency resource location is also known. Using those pilots, the channel response over a coherence interval is interpolated. Each transmitting antenna has a unique pilot symbol that is orthogonal to that of another transmitting antenna.

Channel estimation is a broad and well-studied research topic with high interest in massive MIMO systems exploiting channel reciprocity to estimate the channel based on uplink pilots symbols [15]. TDD operation in that sense is also preferable compared to FDD for massive MIMO, since the number of pilots does not depend on the number of BS antennas, and the feedback from the UEs is not needed. There is even research to estimate the CSI with new Machine Learning approaches [16], [17], to tackle some challenges of channel estimation such as pilot contamination or the high computational complexity when the number of antennas is large. These new techniques are being compared to more conventional signal processing. Traditional channel estimation methods include least squares (LS) and linear minimum mean square error (LMMSE). LS does not require prior channel information, being a low computational complexity estimator. LMMSE, on the other hand, minimizes the estimation errors on average and considers noise but relies on the first-order and second-order statistics of the channel, i.e., mean and covariance matrices, so it has high computational complexity and assumes that the propagation channel can be modeled as a linear system [14].

LS is the estimator used in the simulator to obtain the channel estimates on the pilot symbols, and then linear interpolation is applied to the rest of the resource grid. The channel in this case is represented as a time-frequency resource grid, i.e., a matrix $\mathbf{H}(t, f)$ with rows as subcarriers and columns as OFDM symbols. This is the so-called time-invariant transfer function which is the Fourier transform of the previous $\mathbf{B}(t, \tau)$ [18]. The LS channel estimation that happens at the BS for a particular antenna is expressed in (2.11) below. The process at the UE would be analogous.

$$\text{vec}(\hat{\mathbf{H}}_{LS,P}) = \text{diag}(\text{vec}(\mathbf{X}_P))^{-1}\text{vec}(\mathbf{Y}_P), \tag{2.11}$$

where $\hat{\mathbf{H}}_{LS,P} \in \mathbb{C}^{Q\times N_P}$ is the LS estimate of the matrix $\mathbf{H}_P$ containing only the pilot symbols with Q total subcarriers and N_P pilot symbols. $\mathbf{X}_P \in \mathbb{C}^{Q\times N_P}$ is the matrix of transmitted pilot symbols and $\mathbf{Y}_P \in \mathbb{C}^{Q\times N_P}$ is the matrix of received pilot symbols. Reorganizing the pilot symbols of the vector $\text{vec}(\hat{\mathbf{H}}_{LS,P})$ to their known location within the time-frequency resource grid, a matrix $\hat{\mathbf{H}}_{LS} \in \mathbb{C}^{Q\times N_{\text{symb}}}$ can be built with N_{symb} the total number of OFDM symbols of the resource grid. The next step is to linearly interpolate the channel response for all symbols and subcarriers in $\hat{\mathbf{H}}_{LS}$ with the channel coefficients obtained from $\hat{\mathbf{H}}_{LS,P}$.

It is important to mention that the channel obtained here is an estimate and has errors because, among other things, it does not consider noise. Thus, perfect CSI estimates are not assumed in this master thesis. This would affect the equalization and beamforming filters, as well as the antenna selection process.

2.2.5 Linear Equalization

Once the channel is estimated, we can proceed to decode the data symbols. The optimal way to decode the transmitted vector is to use the Maximum-Likelihood decoder, assuming no channel coding or prior information on the data. To carry out the optimization, an exhaustive search is required on all possible transmitted symbol vectors. With a finite constellation of size A, the search goes through A^{M_T} vector symbols. Thus, the decoding complexity scales exponentially with the number of antennas making it very restrictive for massive MIMO specifically. However, there are some fast algorithms and common sub-optimal approaches to reduce the required computations [2].

To reduce the complexity of the ML receiver, linear filters are used to equalize the received vector, as in Figure 2.2. The post-processed vector results from and linear relationship as shown in (2.12)

$$\mathbf{z} = \mathbf{G}\mathbf{y}, \tag{2.12}$$

where $\mathbf{G} \in \mathbb{C}^{T\times N_R}$ is the matrix corresponding to the linear equalizer. Now, the process is simplified to a matrix multiplication and each data stream is decoded independently. The decoding complexity grows linearly with the size of the matrices, i.e., the number of antennas. The vector $\mathbf{z}$ is an estimate of the transmitted vector, so, in that case, the linear filter has the restriction $T = N_T$ [4]. The most common linear filters are Maximum Ratio Combining (MRC) or Match Filter (MF), Zero-Forcing (ZF), and Minimum Mean Square Error (MMSE). Those linear filters are also used for beamforming so they will be described in the next section.

2.2.6 Beamforming

Another important process is beamforming, also known as precoding. Beamforming is almost equivalent to the equalization performed at the Rx, but this time, at the Tx. The goal is to emit the signal from the BS with an appropriate weighting so that the signal is steered in the direction of the UE. The linear precoder matrix may require normalization to keep the total power of all the streams constant. As stated before, Maximum Ratio Transmission (MRT), ZF, and MMSE are the most common techniques [2], [19]. The equation (2.13) describes the matrix implementation,

$$\mathbf{x} = \sqrt{P_{Tx}}\mathbf{W}\mathbf{s}, \tag{2.13}$$

where $\mathbf{x}$ is the transmitted vector as in (2.7), P_{Tx} is the total transmit power, $\mathbf{s}$ is the signal to be transmitted with unit power containing the data symbols, and $\mathbf{W}$ is the beamforming or precoder matrix which elements are called weights. The precoding matrix is normalized to not add any power and to keep the total power among all the streams constant. Note here once again, that $\mathbf{W}$ will depend on the channel matrix which in our case is not perfect but an estimate. In fact, the channel matrix is $\hat{\mathbf{H}}_{LS}$ described before, and the beamforming is applied in the frequency-time resource grid for each transmit-receive antenna pair. However, without loss of generality, the following filters would have the common notation of $\mathbf{H}$.

Maximum Ratio Transmission or Matched Filter

MRT is the simplest linear precoding technique, which aims to maximize the output SNR. However, MRT does not take into consideration user interference, so it is mainly useful in a noise-limited scenario in low SNR or highly orthogonal channels. The MRT is the Hermitian of the channel matrix, given by,

$$\mathbf{W}_{MRC} = \mathbf{H}^H. \tag{2.14}$$

Zero-Forcing

ZF precoding is also called null-steering, and it is performed by selecting the right Moore–Penrose inverse or pseudo-inverse of the channel matrix. It is one of the solutions to the LS problem minimizing the norm $\|\mathbf{H}^H\mathbf{x} - \mathbf{s}\|$. ZF is given by the expression,

$$\mathbf{W}_{ZF} = \mathbf{H}^{\dagger} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}. \tag{2.15}$$

Interference from other UEs is canceled out, however, it does that by enhancing the noise which potentially can become correlated across the streams. It is most suitable for interference-limited scenarios.

Regularized Zero-Forcing

A more sophisticated approach is to apply Regularized Zero-Forcing (RZF) to reduce the noise enhancement while mitigating the multistream interference. The RZF is given by the following expression,

$$\mathbf{W}_{RZF} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \xi\mathbf{I})^{-1}. \tag{2.16}$$

MMSE precoding is a particular case of RZF. MMSE minimizes the MSE when $\xi = M_R\sigma_n^2/P_{TX}$. When $\xi = 0$, i.e., at very high SNR regimes, becomes ZF, and when $\xi \to \infty$, i.e., at very low SNR regimes, the filter converges to MF.

2.3 NR 5G Simulator

2.3.1 NR Overview

Before describing the simulator, some NR physical layer characteristics will be introduced since they are included in it. NR is an evolution of LTE as it is based on this previous generation architecture and reuses many of its features, however, NR provides more benefits compared to LTE. It is a new radio access technology so it has also been redesigned allowing more flexibility and forward compatibility for future use cases, higher-frequency bands exploitation with larger bandwidth transmission and high data rates, ultra-lean principle, network efficiency enhancement, interference and latency reduction, and has an extensive usage of beamforming with a larger number of antennas elements.

OFDM was also chosen for NR as it is a suitable waveform for time dispersion robustness and ease to multiplex and exploit both time and frequency domains [1]. One relevant feature of OFDM is the numerology which determines the subcarrier spacing and cycle prefix length, and it has to be balanced to resist frequency errors, phase noise, and Doppler spread. In LTE, the subcarrier

spacing and cyclic prefix were fixed at 15 kHz and 4.7 µs (there is also an extended cyclic prefix but is not used in practical LTE deployments), which offered a good balance for the envisioned scenarios. NR, on the other hand, supports scalable numerology with a flexible subcarrier spacing increasing in powers of 2 from 15 kHz, so coexistence with LTE is allowed, to 240 kHz. There are a total of 5 different numerologies as shown in Table 2.1.

Numerology (μ)	**Subcarrier Spacing (SCS)** [kHz]	**Useful Symbol Time** (T_U) [µs]	**Cyclic Prefix** (T_{CP}) [µs]
0	15	66.7	4.7
1	30	33.3	2.3
2	60	16.7	1.2
3	120	8.33	0.59
4	240	4.17	0.29

Table 2.1: Numerologies and Subcarrier Spacing in NR [1].

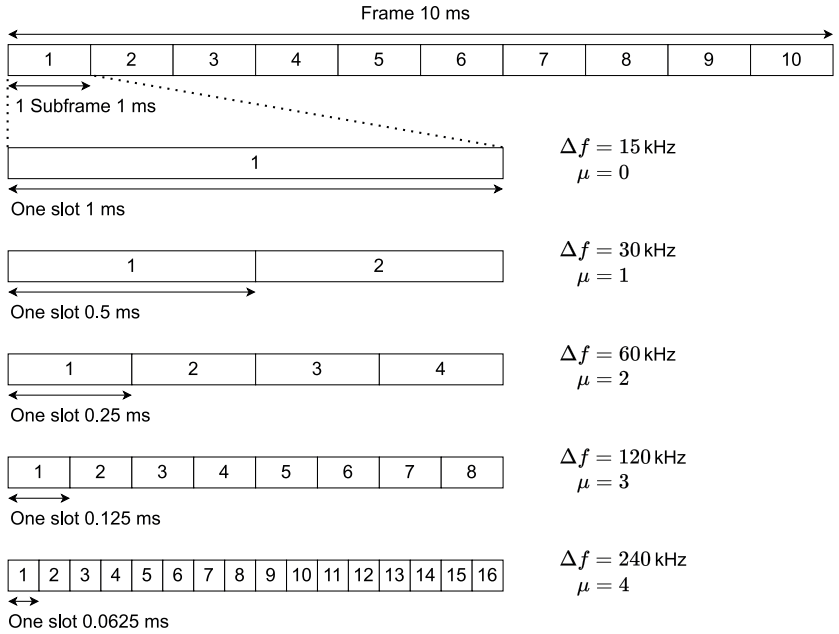


Figure 2.3: Frame and slot structure in NR [1].

NR frames have a duration of 10 ms in the time domain. Each frame is divided into 10 equal subframes of 1 ms. In turn, a subframe is divided into slots whose total number depends on the numerology, as shown in Figure 2.3. A slot commonly contains 14, or if the extended cyclic prefix is used, 12 OFDM symbols.

All devices in LTE are capable of supporting a carrier bandwidth of 20 MHz. But, in NR, devices should be able to support very wide bandwidths of up to 400 MHz for a single carrier, which is prohibitive in terms of cost. Therefore, it is important to say that NR devices can use only a portion of the carrier and not be centered on the carrier frequency, which brings some implications for handling the DC subcarrier. In the frequency domain, the smallest physical resource in NR is a resource element (RE) which consists of one subcarrier during one OFDM symbol. A resource

block (RB) is composed of 12 consecutive subcarriers in the frequency domain only, and here there is a difference compared to LTE where a resource block is a two-dimensional structure, with 12 subcarriers in frequency and one LTE slot in time. Since NR supports different numerologies, a resource block can have different bandwidths. Nevertheless, the resource block boundaries are aligned across numerologies, so one resource block with a particular subcarrier spacing occupies the same as two resource blocks with half of that subcarrier spacing.

This master thesis is mainly focused on the physical layer which is responsible for coding, modulation, and multi-antenna mapping among others. It provides services to the second layer, i.e., Medium-Access Control (MAC) and Radio-Link Control protocols. There are three types of channels in NR: logical, transport, and physical channels. The physical channels are the only ones considered and implemented in the simulator, and they map a particular transport channel to be transmitted in a set of time-frequency resources. There are a total of six Physical Channels. The Physical Downlink Shared Channel (PDSCH), the Physical Broadcast Channel (PBCH), the Physical Downlink Control Channel (PDCCH), the Physical Uplink Shared Channel (PUSCH), the Physical Uplink Control Channel (PUCCH), and the Physical Random-Access Channel (PRACH). Of these channels, the simulator specifically uses only the PDSCH and PUSCH, which are used to transmit data in downlink and uplink, respectively.

One relevant step described before was channel estimation, which may be performed in different ways. Channel knowledge is essential to implement most of the complex processing techniques in current mobile communications, even more critical when having a large number of antennas and high bandwidth. To be able to estimate the channel, pilots are sent, known as reference signals, which are predefined symbols at specific locations in the time-frequency resource grid, known by both receiver and transmitter.

In uplink, the UE sends pilots called Sounding Reference Signals (SRS) in NR. SRS transmissions are the equivalent of the downlink CSI-RS measurements [1]. CSI-RS have the objective to estimate the CSI and report back to the BS, which in NR is called gNB, and then select the precoder matrices. Since we are working in TDD and assuming reciprocity, those signals are not used in the simulator. Demodulation Reference Signals (DMRS) will be used in downlink for channel estimation at the receiver.

SRS are based on the *Zadoff-Chu* sequences which have some good properties such as a constant amplitude making them efficient in terms of power amplification. A SRS spans over two or more OFDM symbols at the end of a slot. In the frequency domain, SRS for more than one user are multiplexed so the SRS for one user are transmitted every second or fourth subcarrier. Each user is assigned a different *Zadoff-Chu* sequence with a phase rotation which makes them orthogonal. Similarly, if there is more than one antenna port SRS sequences must be orthogonal between them. In this thesis, the number of antenna ports is the same as the number of antennas used or active at that moment. Now, once the channel is estimated, linear equalizers are employed to estimate the transmitted symbols.

In downlink, as it was previously mentioned, reciprocity is assumed, so DL CSI can be obtained from the UL channel information estimates. Knowing the channel at the BS, downlink precoding or beamforming can be performed [1]. As explained in Section 2.2.6, beamforming consists of focusing the transmitted power in a certain direction or particular location in space using the CSI available. This increased directivity helps to achieve high data rates and reduce interference with other links.

2.3.2 Simulator

The 5G NR simulator was developed in previous master theses [20] and [21]. It consists of a set of modules as functions and methods in MATLAB® which implement NR specifications. The 5G NR Toolbox is not used because of license issues. On top of that, a Graphical User Interface (GUI) was also developed with tabs to separate the settings from the simulations to make the

simulator more dynamic, responsive, and intuitive for users. The modules use the configuration entered by the user and the data and variables are organized in *structs* which are structure arrays of MATLAB®.

Originally, it implemented the WINNER II channel model since MATLAB® has a toolbox with functions to configure and generate the channel coefficients. However, as stated before, WINNER II was changed for a new channel model which follows the specifications of the TR 38.901 V16.1 [10]. The simulator consists of the blocks shown in Figure 2.4.

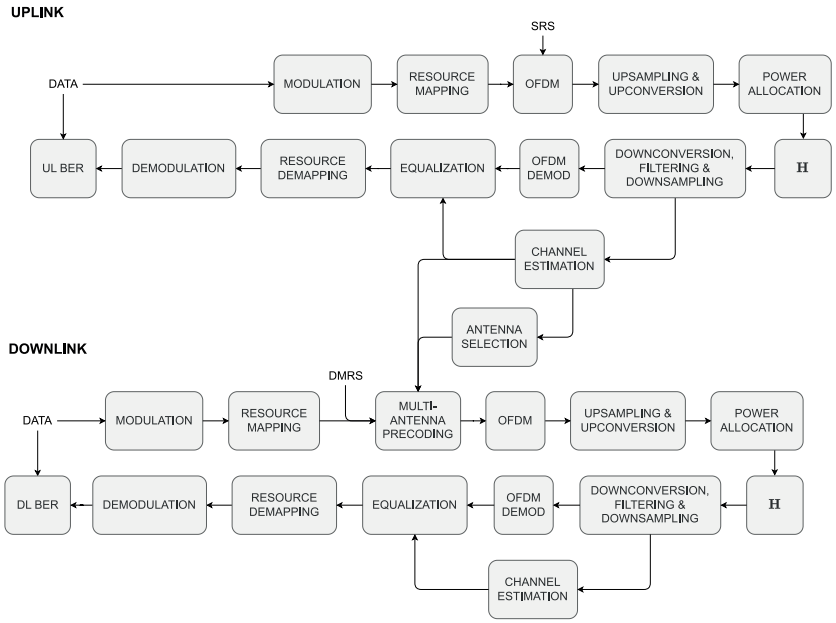


Figure 2.4: Simulator block diagram.

There are two main blockchains: uplink and downlink. The uplink transmission consists of random data generation, modulation, resource mapping, OFDM modulation, upsampling, upconversion and power allocation. The signal goes through the channel one subframe at a time. Since the channel has multiple delays, the resulting signal at the receiver would be the sum of those delayed versions. Then, at the BS in this case, the process is reversed. The steps are downconversion, filtering and downsampling, OFDM demodulation, channel equalization, demapping, and demodulation. The downlink chain is analogous but there is multi-antenna precoding. In that step, transmission layers are mapped to a set of antenna ports using a precoder matrix. This is called beamforming which in our case is implemented through digital linear filters and is only performed in the downlink. Scrambling and channel coding steps are not included in either of the two links since they do not affect the results.

The simulation flow is the following. The scenario is set up with the BS, i.e., gNB and the UEs location and link characteristics as frequency, bandwidth, modulation, etc. Then the channel coefficients can be computed. The uplink frames are transmitted with the PUSCH data and SRS symbols are placed in the resource grid. After passing through the channel, the receiver, in this case, the BS, demodulates the signal and estimates the channel from the SRS symbols. Once the received signal is decoded, the uplink BER can be calculated. After that, from the channel estimates, the gNB performs the beamforming. Then, the DMRS symbols are added and downlink frames are sent. The signal goes through the same channel again and the UEs receive the signal. The channel is equalized and the data symbols are decoded. Finally, the downlink BER can be obtained.

In the next chapter, the AS problem is explained and different approaches are addressed that will be implemented later on in the simulator.

Chapter 3

Antenna Selection

This chapter presents the antenna selection problem as well as the implementation of selection algorithms that will select a subset of antennas according to a specific criterion or objective function.

3.1 Antenna Selection

Massive MIMO can help to handle the new challenges arising in current and future wireless communications with more devices connected to the network and the development of new applications and services requiring higher data rates and spectral efficiency. However, it comes at a cost. The economic cost, power consumption, and the increasing complexity of the hardware and RF chains, containing power amplifiers, phase shifters, and ADCs/DACs, could become a real burden to the actual development and implementation of BS with a large number of antennas. To overcome those challenges, some solutions such as hybrid analog-digital beamforming or algorithm optimization have been proposed [22].

Antenna selection (AS) is a feasible alternative for reducing the complexity in both small and large-scale MIMO systems [23]. The reason is that antennas are usually inexpensive to manufacture and easy to deploy, whereas the RF chains are the critical components cost wise [19]. AS has been studied and analyzed on both Tx and Rx sides. In reception, the UEs transmit to the BS, which chooses the group of antennas to be used for receiving the signal. However, in this master thesis, we will focus on the BS while transmitting mainly, i.e., downlink. Whether at the Rx or the Tx side, the main idea is to be able to select a high-quality subset of the available antennas while the rest are switched off. In this way, only a reduced number of RF transceivers are connected while retaining to a high extent the performance of the original, full complexity massive MIMO systems [24], [25]. AS methods have been analyzed in conventional MIMO systems in the past decade [26], most of them focused on the Rx side trying to optimize the capacity.

Switching off RF chains can simplify massive MIMO circuitry and its hardware, but the performance is degraded. To obtain the minimum possible losses in performance, the channel is exploited, and based on its properties the *best*, i.e., high-quality, subset of antennas is selected out of the total available. Therefore, CSI is required at the BS and the final performance of the AS depends on the channel properties. In a typical i.i.d. Rayleigh channel, no specific antenna outperforms the others, however, in real massive MIMO channels, some antennas could contribute more than others, the latter of which can be switched off without incurring a great performance loss [27].

3.2 Antenna Selection Criterion

We consider the previous general MIMO scenario in Figure 2.2 with N_T transmit antennas and N_R receive antennas ($N_T > N_R$). In the Tx, i.e., the BS, the AS processing stage selects $L \leq N_T$ antennas out of the N_T antennas. The received signal vector is expressed in the following equation,

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

This expression is inferred from (2.7) but here $\mathbf{H}_{\mathcal{S}}$ represents the channel submatrix of $\mathbf{H}$ by selecting the columns of the subset $\mathcal{S} \subseteq \{1, \ldots, N_T\}$ with cardinality $|\mathcal{S}| = L$, so $\mathbf{H}_{\mathcal{S}} \in \mathbb{C}^{N_R \times L}$. In fact, this submatrix has yet another dimension considering the cluster delays, this will be explained in the next chapter. Let us assume that the channel matrix here has only the receiving and transmitting antenna dimensions. Thus, AS can be translated to a submatrix selection problem. This submatrix selection is performed based on an objective function that tries to maximize a certain feature. In Figure 3.1 below, it can be observed how AS could be included in the system.

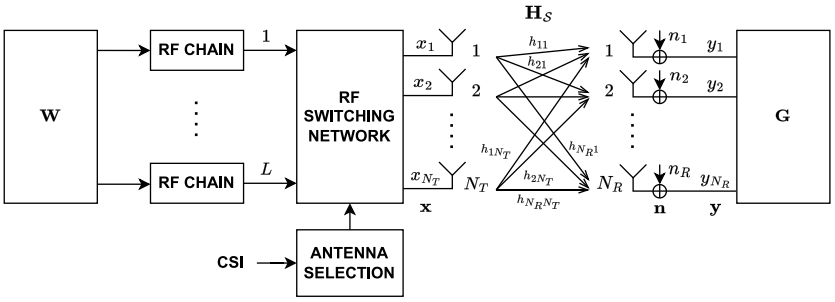


Figure 3.1: Massive MIMO block diagram with antenna selection.

3.2.1 Minimum Singular Value

In [28] it is stated that the spatial multiplexing performance with linear receivers depends on the $\text{SNR}_{\min}$ which corresponds to the lowest SNR of all streams. From (2.12) the transmitted signal can be estimated through a linear equalizer. A particular SNR_k from the k-th row of $\mathbf{H}$ is maximized as follows [24], [28]

$$\begin{aligned} \text{SNR}_{k,\min} &\propto \frac{1}{\max\limits_{k} [\mathbf{H}_{\mathcal{S}}^H \mathbf{H}_{\mathcal{S}}]_{kk}^{-1}} = \frac{1}{\max\limits_{\mathbf{e}_k} \mathbf{e}_k^H [\mathbf{H}_{\mathcal{S}}^H \mathbf{H}_{\mathcal{S}}]^{-1} \mathbf{e}_k} \\ &\geq \frac{1}{\max\limits_{\|\mathbf{a}\|^2=1} \mathbf{a}^H [\mathbf{H}_{\mathcal{S}}^H \mathbf{H}_{\mathcal{S}}]^{-1} \mathbf{a}} = \frac{1}{\sigma_{\max}([\mathbf{H}_{\mathcal{S}}^H \mathbf{H}_{\mathcal{S}}]^{-1})} = \sigma_{\min}^2(\mathbf{H}_{\mathcal{S}}), \end{aligned} \tag{3.2}$$

where the expression $[\cdot]_{kk}$ represents the k-th diagonal element, $\mathbf{e}_k$ is the k-th column of $\mathbf{I}_L$, the vector $\mathbf{a}$ is the eigenvector of the matrix $[\mathbf{H}_{\mathcal{S}}^H \mathbf{H}_{\mathcal{S}}]^{-1}$, and $\sigma_i(\mathbf{H}_{\mathcal{S}})$ denotes the i-th singular value of the matrix $\mathbf{H}_{\mathcal{S}}$. The expression derived from (3.2) implies that the minimum SNR among the receive antennas, i.e., UEs, is lower bounded by the minimum singular value (MSV) of the channel submatrix. Consequently, the selection criterion is to find the submatrix $\mathbf{H}_{\mathcal{S}}$ with the largest MSV. Authors in [28] demonstrate that this selection criterion guarantees a large minimum SNR and it is optimal for linear receivers in terms of diversity order, i.e., BER, which is the main feature to optimize in our case. Note that TDD operation is assumed so the submatrix selection is performed over the estimated uplink channel at the BS. Hence, $\mathbf{H}_{\mathcal{S}}$ would be derived from the channel estimate described in Section 2.2.4 in the simulator, but the previous analysis remains valid.

Other selection criteria can also be implemented such as maximizing the norm of the submatrix [29], with lower complexity but lower expected performance. If channel capacity is to be maximized, the optimization problem is

$$\mathbf{H}_{\mathcal{S}} = \arg\max_{\mathbf{H}_{\mathcal{S}}} \log_2 \det\big(\mathbf{I}_L + \overline{\rho} \mathbf{H}_{\mathcal{S}}^H \mathbf{P} \mathbf{H}_{\mathcal{S}}\big), \tag{3.3}$$

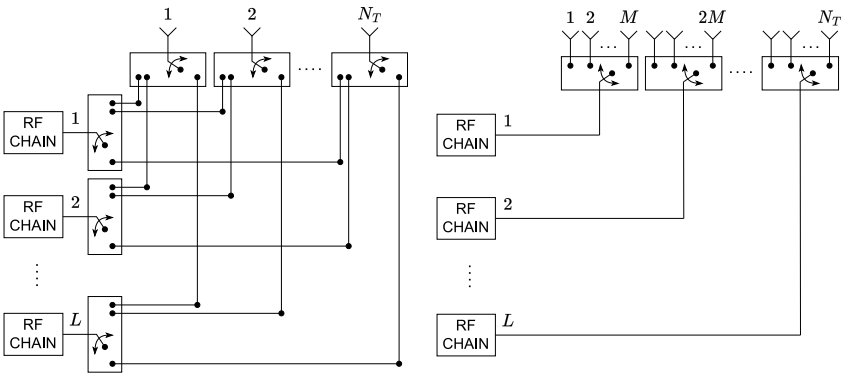
with $\overline{\rho}$ as the average SNR and $\mathbf{P}$ is a diagonal power allocation matrix [30]. In this case, the solution is to find a submatrix that gives the largest determinant above, so a different objective function could be adopted [30], [31]. It has been demonstrated that simple selection schemes related to the norm or the channel gain perform closely to the optimal solution [19]. In any case, choosing a reasonable selection criterion is expected to give better results than a random selection both for BER and capacity.

3.3 Antenna Selection Algorithm

Once the criterion is defined, the question that arises is how to find this optimal submatrix $\mathbf{H}_{\mathcal{S}}$ of all the possible submatrix combinations. The AS problem, as a search problem, is highly computationally expensive and even infeasible in massive MIMO when a large number of antennas are available to be selected. Therefore, the selection algorithm, i.e., the search algorithm, is a critical part of AS. But before introducing the algorithms, the RF switching network will be addressed.

3.3.1 RF Switching Network

Together with the RF chains, RF switches (as shown in Figure 3.1) are also one of the most decisive components. A fully flexible full-array switching (FAS) network allows the selection of an arbitrary subset of L antennas, however, designing it with low loss and high isolation can be difficult when the number of antennas is large [25]. In massive MIMO, this complexity of the RF switching could become a performance factor too. Sub-array switching (SAS) topologies have also been introduced where the full-array is divided into L sub-arrays of size $M = N_T/L$, and only one antenna is chosen in each sub-array, thus reducing the complexity of the RF switching network. The way of grouping the antennas could be done in smart ways such as an interleaving configuration when antenna correlation is considered [30]. Binary switching networks with $M = 2$ are also commonly studied structures. In Figure 3.2, FAS and SAS networks can be observed.



(a) Full-array switching (FAS) network. **(b)** Sub-array switching (SAS) network.

Figure 3.2: RF switching networks [25].

If an exhaustive search is performed, the optimal solution to this NP problem has an exponential order of complexity of $\mathcal{O}\left(\binom{N_T}{L}\right)$ and $\mathcal{O}\left(\left(\frac{N_T}{L}\right)^L\right)$ (or $\mathcal{O}\left(M^L\right)$) for the full-array and sub-array AS respectively. The complexity of sub-array selection is lower than the one of the full-array switching, but the number of possible combinations is also much smaller, which could affect the final performance of the AS. However, a few papers have shown that sub-array AS performs close both in capacity and diversity order to the full-array AS, especially when $N_T \gg N_R$ [25], [32].

3.3.2 Greedy Search

We will start by introducing a sub-optimal but faster algorithm before describing the Branch-and-Bound (BAB) algorithm. In such search problems, a greedy searching algorithm can be applied. The heuristic of a greedy algorithm is that the choice is the optimal one at each stage, i.e., the local optimal solution. Therefore, all the MSVs of the candidate subsets are examined and the maximum one is selected in the current step. The procedure continues including antennas in a growing matrix, i.e., extending the column size, until a subset of L antennas is selected. The greedy algorithm could work with a full-array switching network, or with sub-array switching. In any case, the greedy algorithm will find a sub-optimal solution but significantly reduces the search space of the exhaustive search and thus the complexity by not calculating the possible combinations with lower MSVs at a given level. The greedy algorithm is presented in Alg. 1 and $f(\mathbf{H}_{\mathcal{S}})$ represents the objective function which in our case is the MSV, i.e., $\sigma_{\min}(\mathbf{H}_{\mathcal{S}})$.

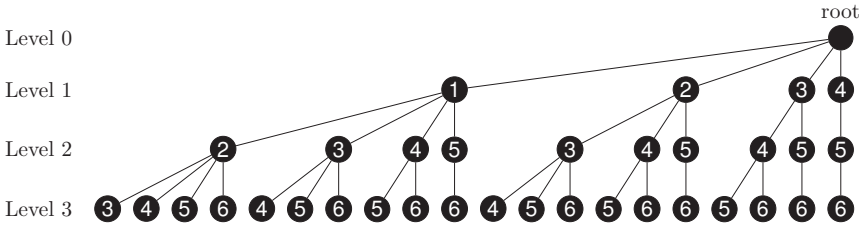
Algorithm 1 Greedy-search antenna selection.

Input: $\mathbf{H}$, L
Output: $\mathcal{S}_g$, B_g, N_{vg}
1: Initialize $N_T \leftarrow$ columns of $\mathbf{H}$, $\mathcal{S}_g \leftarrow \emptyset$, $l \leftarrow 0$ and $N_{vg} \leftarrow 0$
2: $l \leftarrow l + 1$
3: $M \leftarrow \lfloor N_T/L \rfloor$ ▷ for sub-array switching
4: $B_g \leftarrow 0$
5: $\mathcal{S} \leftarrow \mathcal{S}_g$
6: **if** $N = L$ **then**
7: | $N_{vg} = 1$
8: | $\mathcal{S}_g \leftarrow \{1 : N\}$
9: | $B_g \leftarrow f(\mathbf{H}_{\mathcal{S}_g})$
10: | **return** $\mathcal{S}_g$, B_g, N_{vg}
11: **end if**
12: **for** i in $N \setminus \mathcal{S}$ or $1 : M$ **do** ▷ for full-array or sub-array switching respectively
13: | $N_{vg} \leftarrow N_{vg} + 1$
14: | $\mathcal{S}_l \leftarrow \mathcal{S} \cup \{i\}$ ▷ for full-array switching
15: | $\mathcal{S}_l \leftarrow \mathcal{S} \cup \{i + (l-1)M\}$ ▷ for sub-array switching
16: | $B \leftarrow f(\mathbf{H}_{\mathcal{S}_l})$
17: | **if** $B > B_g$ **then**
18: | | $B_g \leftarrow B$
19: | | $\mathcal{S}_g \leftarrow \mathcal{S}_l$
20: | **end if**
21: **end for**
22: **if** $l = L$ **then**
23: | **return** $\mathcal{S}_g$, B_g, N_{vg}
24: **else**
25: | go to line 2
26: **end if**

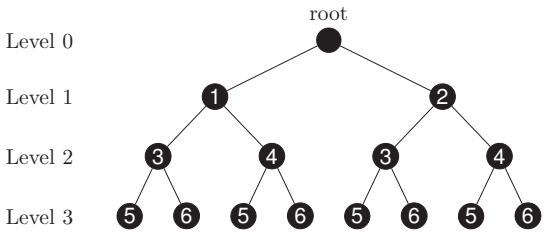
3.3.3 Branch-and-Bound Search

To explain the BAB algorithm, search trees will be introduced first. Search trees can be constructed to ease the understanding of the switching network and how the algorithms work. An example of selecting $L = 3$ out of $N_T = 6$ is shown in Figure 3.3. Figure 3.3(a) represents the full-array containing the possible 20 combinations, whereas in 3.3(b) the sub-array structure only considers 8 combinations reducing the search space. However, as was mentioned, the results with the sub-array structure may not be optimal. Let us assume that the nodes denote the index of the column to be added, i.e., the antenna included in a growing matrix. Note that the exhaustive search complexity here can be seen as the sum of all the paths from the root to a leaf node.

The greedy algorithm is sub-optimal, therefore fast algorithms that can guarantee the global optimal solution under certain conditions are needed. One of them is BAB, a popular searching algorithm with much higher searching efficiency than exhaustive search. Structuring again the search tree as in Figure 3.3, the BAB algorithm prunes branches without calculating the associated subnodes and can reach the optimal solution when some constraint is fulfilled. This condition is that there must be a monotonic relationship between the subset size and the objective function or criterion. BAB searching procedure occurs iteratively so all the leaf nodes are either evaluated or pruned obtaining the optimal solution for a particular tree structure [25].



(a) Full-array antenna selection.



(b) Sub-array antenna selection $M = 2$.

Figure 3.3: Example of search trees with $N_T = 6$ and $L = 3$.

Considering the MSV as the objective function from the results derived in (3.2), the monotonicity of MSV is proven in [33]. Given a matrix $\mathbf{A} \in \mathbb{C}^{R \times C}$ and a column vector $\mathbf{b} \in \mathbb{C}^{R \times 1}$, then

$$\sigma_{N+1}\left(\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix}\right) \leq \sigma_N\left(\mathbf{A}\right) \text{ if } R > C, \tag{3.4}$$

$$\sigma_P\left(\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix}\right) \geq \sigma_P\left(\mathbf{A}\right) \text{ if } R \leq C. \tag{3.5}$$

Here we have two cases, the first when there are more rows than columns, $R > C$, or lower or equal rows than columns, $R \leq C$. This would have implications for the efficiency and solution that the algorithms will obtain, whether is it optimal or not. Expression (3.4) indicates that adding an extra column in a tall matrix ($R > C$) does not increase the MSV. In other words, considering two subsets such that $\mathcal{S}_1 \subseteq \mathcal{S}_2$ and $|\mathcal{S}_2| \leq L$, then $\sigma_{\min}(\mathbf{H}_{S_1}) \geq \sigma_{\min}(\mathbf{H}_{S_2})$, but this only happens for tall matrices, i.e., $N_R > L$. This denotes that if we go through the search tree adding columns, the decreasing monotony is only satisfied when the total number of antennas of the UEs is larger than the L antennas chosen at the BS. Although, this may be a possible case in real life, a scenario where the opposite condition $N_R \leq L$ occurs should be also addressed.

The relationship in expression (3.5) requires a different interpretation of search trees. In Figure 3.3, both trees depict the nodes as the columns being added to a growing matrix. But it can also be understood as the columns being removed from the complete channel matrix $\mathbf{H}$, having a decreasing subset size at each level. In that sense, with the same example as in Figure 3.3(a) with $N_T = 6$ and $L = 3$, when we move to node 1, the subset to examine is $\{2, 3, 4, 5, 6\}$. This specific example is a special case because the search tree built has the same shape for both interpretations, but in a general case the number of levels would be $N_T - L$ for a decreasing matrix size.

With this in mind, (3.5) states that removing a column of the matrix decreases the MSV. Therefore, in an analogous way to the previous example, considering two subsets such that $\mathcal{S}_1 \subseteq \mathcal{S}_2$ and $|\mathcal{S}_1| \geq L$, then $\sigma_{\min}(\mathbf{H}_{S_1}) \leq \sigma_{\min}(\mathbf{H}_{S_2})$ when both matrices are fat, i.e., $N_R \leq L$. In this case, this means that the number of UEs and their total number of antennas is lower than the L antennas selected at the BS. This case would be the one assumed during the simulations.

Two different BAB-based search algorithms have been developed. An upward approach would be followed for the first case (3.4), with a gradually increasing subset size; and a downward approach for the latter case (3.5), removing elements from the subset. The efficiency of the BAB algorithm as mentioned before, lies in the fact that if the MSV of the current subset, which is the upper bound for its branches, is lower than the global bound B, those branches and subsets can be pruned without being computed. This applies always when the respective condition is satisfied so that the final solution is optimal. In Algo. 2 and 3, upward BAB and downward BAB are described. Note that the upward BAB algorithm can be executed in a sub-array switching network easily but the tree construction for the downward BAB could be more complicated. Thus, some simplifications have been made for the downward BAB SAS search, such as that the sub-array size is set to $M = 2$ and if the subset chosen at the leaf node still has more elements than L, the first L elements are taken.

Algorithm 2 BAB-upwards-search antenna selection.

Input: $\mathbf{H}$, L
Output: $\mathcal{S}_{bab-up}$, B_{bab-up}, $N_{vbab-up}$

Initialize $N_T \leftarrow$ columns of $\mathbf{H}$, $\mathcal{S}_l \leftarrow \emptyset$, $B_{bab-up} \leftarrow 0$, $l \leftarrow 0$ and $N_{vbab-up} \leftarrow 0$
$l \leftarrow l + 1$
$M = N_T - (L - l)$ ▷ for full-array switching
$M = N_T / L$ ▷ for sub-array switching
$\mathcal{S} \leftarrow \mathcal{S}_l$
if $N = L$ **then**
  $Nv = 1$
  $\mathcal{S}_{bab-up} \leftarrow \{1 : N\}$
  $B_{bab-up} \leftarrow f(\mathbf{H})$
  return $\mathcal{S}_{bab-up}$, B_{bab-up}, $N_{vbab-up}$
end if
if $\mathcal{S}$ is empty **then** ▷ for full-array switching
  $i_0 \leftarrow 1$
else
  $i_0 \leftarrow \mathcal{S}_{\text{end}} + 1$
end if
$i_0 \leftarrow 1$ ▷ for sub-array switching
if $l = L$ **then**
  for $i = \mathcal{S}_{\text{end}} + 1 : M$ **do**
    $N_{vbab-up} \leftarrow N_{vbab-up} + 1$
    $\mathcal{S}_l \leftarrow \mathcal{S} \cup \{i\}$ ▷ for full-array switching
    $\mathcal{S}_l \leftarrow \mathcal{S} \cup \{i + (l-1)M\}$ ▷ for sub-array switching
    $B \leftarrow f(\mathbf{H}_{\mathcal{S}_l})$
    if $B > B_{bab-up}$ **then**
      $B_{bab-up} \leftarrow B$
      $\mathcal{S}_{bab-up} \leftarrow \mathcal{S}_l$
    end if
  end for
else
  for $i = i_0 : M$ **do**
    $N_{vbab-up} \leftarrow N_{vbab-up} + 1$
    $\mathcal{S}_l \leftarrow \mathcal{S} \cup \{i\}$ ▷ for full-array switching
    $\mathcal{S}_l \leftarrow \mathcal{S} \cup \{i + (l-1)M\}$ ▷ for sub-array switching

```
        B ← f(H_{S_l})
        if B > B_{bab−up} then
            go to line 2
        end if
    end for
end if
```

Algorithm 3 BAB-downwards-search antenna selection.

Input: $\mathbf{H}$, L
Output: $\mathcal{S}_{bab-down}$, $B_{bab-down}$, $N_{vbab-down}$

```
Initialize N_T ← columns of H, ind ← 0, B_{bab−down} ← 0, l ← 0 and N_{vbab−down} ← 0
l ← l + 1
M = L + l                                          ▷ for full-array switching
M = 2                                              ▷ for sub-array switching
if N_T = L then
    N_{vbab−down} ← N_{vbab−down} + 1
    S_{bab−down} ← {1 : N_T}
    B_{bab−down} ← f(H)
    return S_{bab−down}, B_{bab−down}, N_{vbab−down}
end if
if ind = 0 then
    i_0 ← 1
    S ← {1 : N}
else
    i_0 ← ind + 1
end if
i_0 ← 1                                            ▷ for sub-array switching
if l = N − L then
    for i = i_0 : M do
        N_{vbab−down} ← N_{vbab−down} + 1
        S_l ← S \ {i}                              ▷ for full-array switching
        S_l ← S \ {i + (l − 1)M}                   ▷ for sub-array switching
        S_l ← S_l(1 : L)                           ▷ for sub-array switching
        B ← f(H_{S_l})
        if B > B_{bab−down} then
            B_{bab−down} ← B
            S_{bab−down} ← S_l
        end if
        if i + (l − 1)M = N_T then                 ▷ for sub-array switching
            break
        end if
    end for
else
    for i = i_0 : M do
        N_{vbab−down} ← N_{vbab−down} + 1
        S_l ← S \ {i}                              ▷ for full-array switching
        S_l ← S \ {i + (l − 1)M}                   ▷ for sub-array switching
        B ← f(H_{S_l})
        if B > B_{bab−down} then
            S ← S_l
            ind ← i
            go to line 2
        end if
    end for
end if
```

3.3.4 Algorithm Comparison

To understand and compare the actual performance of each algorithm, four simulations with N_T from 4 to 32 antennas at the BS have been conducted. Table 3.1 lists the cases particularities.

1. Same number of receive antennas as the number of available transmit antennas, with the variable number of selected antennas growing with the total number of antennas: $L = \lfloor N_T/2 \rfloor$, $N_R = N_T$. $M = \lceil N_T/L \rceil$, which is equal to 2 for even N_T.
2. Same number of selected antennas as the number of receive antennas: $L = N_R = 2$.
3. Fixed number of selected antennas, which is higher than the number of receive antennas: $L = 4 > N_R = 2$.
4. Variable number of selected antennas growing with the total number of antennas: $L = \lfloor N_T/2 \rfloor$, $N_R = 2$. Thus, $M = \lceil N_T/L \rceil$, which is equal to 2 for even N_T.

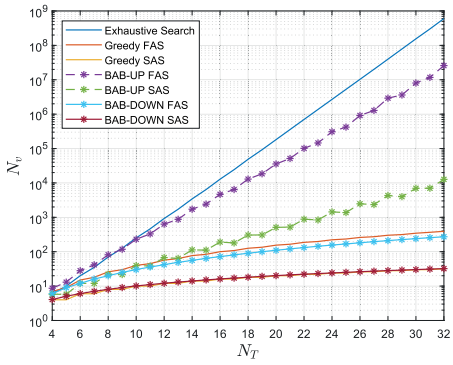
Case	N_T	N_R	L	M	**Condition**
1	$\{4, \ldots, 32\}$	N_T	$\lfloor N_T/2 \rfloor$	$\lceil N_T/L \rceil$	$N_R > L$ (3.4)
2	$\{4, \ldots, 32\}$	2	2	$\lceil N_T/2 \rceil$	$N_R = L$ (3.5)
3	$\{4, \ldots, 32\}$	2	4	$\lceil N_T/4 \rceil$	$N_R < L$ (3.5)
4	$\{4, \ldots, 32\}$	2	$\lfloor N_T/2 \rfloor$	$\lceil N_T/L \rceil$	$N_R < L$ (3.5)

Table 3.1: Algorithm comparison cases.

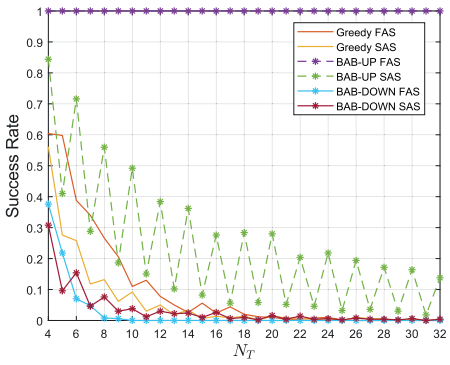
To have smoother curves and a realistic **H** channel matrix, 500 channel matrices have been generated according to TR 38.901 Release 16, i.e., the one used in the simulator and with similar characteristics. The first case represents the condition in (3.4) when $N_R > L$. The upward BAB FAS algorithm is assumed to be optimal for this case. The last three cases fulfilled the condition in (3.5). Case 2 refers to when the equality $N_R = L$ is met. In case 3, L is also fixed but greater than N_R, and in case 4, L is variable and becomes large, so the size of the tree and the number of combinations are much larger. We would expect that in all of these cases, the downward BAB FAS algorithm could find an optimal solution. The question is whether it requires visiting a significant number of nodes and calculating their MSV. The visited and calculated nodes metric is what is called N_v in the pseudocode of the algorithms. This number also equals the number of evaluations done, since when a node is visited is included (or discarded) in the current subset of columns and then the objective function is calculated. In addition, a comparison both in visiting nodes and accuracy performance would be interesting to have for all of the algorithms. Therefore, the success rate is introduced and gives the rate at which the algorithm manages to return the optimal solution.

As was previously mentioned, the exhaustive search for a full-array switching network has an exponential complexity which makes the number of visited nodes $\binom{N_T}{L}$ and this number grows with both the total number of antennas and the antennas selected since all leaf nodes should be evaluated, and it does not have pruning methods. Exhaustive search for FAS has been included as a benchmark whereas exhaustive search for sub-array switching has been omitted. In the case of greedy search for both types of switching networks the expected of visited nodes should be lower than BAB and much lower than exhaustive search since it does not deepen so much in the tree and its complexity does not depend on the channel characteristics but on the size of the tree which is directly related to N_T and L. On the other hand, BAB complexity depends not only on the structure and size of the tree but also on the channel characteristics so it can prune branches faster than exhaustive search.

For each case, the computational complexity in terms of the average of the visited nodes appears on the left, and the average success rate, on the right. In Figure 3.4, case 1 is depicted. Upward BAB FAS achieves the optimal solution since the expression (3.4) is fulfilled. Upward BAB FAS has about 1 order of magnitude reduction in N_v compared to exhaustive search. For the SAS network, this algorithm performs better in terms of success rate for an even number of antennas since the search tree that is built is more balanced. Both FAS and SAS greedy algorithms outperform the downward BAB algorithms in terms of success rate since the latter are designed for the other condition and do not solve the problem effectively.

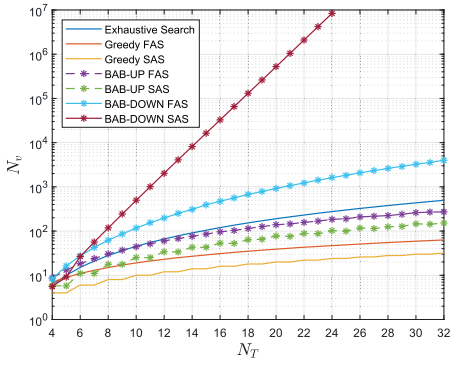


(a) Visited nodes case 1.

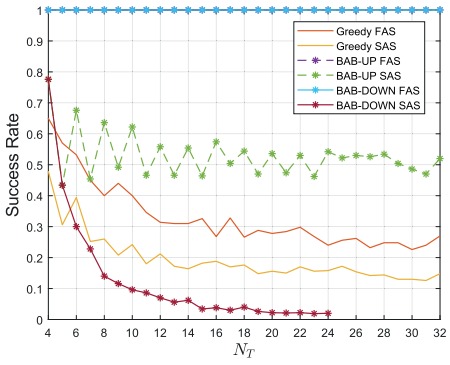


(b) Average success rate case 1.

Figure 3.4: Comparison of algorithms for case 1, $L = \lfloor N_T/2 \rfloor < N_R = N_T$.



(a) Visited nodes case 2.

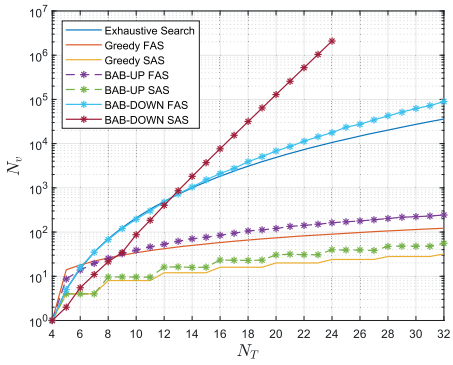


(b) Average success rate case 2.

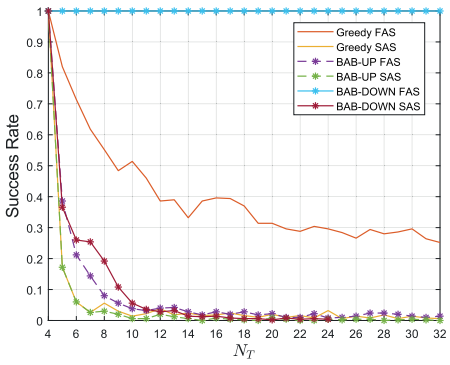
Figure 3.5: Comparison of algorithms for case 2, $L = N_R = 2$.

Case 2 can be observed in Figure 3.5. We should mention first that for both cases 2 and 3 because downward BAB SAS has a much higher complexity than even the exhaustive search, N_T has been limited to 24 antennas for that particular algorithm. Due to the fact that L is small compared to N_T, downward BAB performs the worst and even surpasses the exhaustive search since more columns have to be removed from the original matrix. In this case, downward BAB gives no advantage. Upward BAB has a complexity in-between greedy and exhaustive search which was also expected. However, with this example, no significant benefit can be obtained from the upward BAB FAS relative to the exhaustive search, although with even higher N_T, this difference could increase. Looking at the success rate, both upward BAB and downward BAB for FAS configuration reach the optimal solution since the equality $N_T = L$ is a special case. The rest of the algorithm success rates decrease with N_T and are proportional to their visited nodes, i.e, the higher the complexity, the better the success rate, except for downward BAB SAS. Note that the upward BAB SAS curve seems to stabilize around 50%, although further simulations with higher N_T should be addressed to confirm this observation.

In Figure 3.6, we have case 3. As for case 2, although not as pronounced, downward BAB gives no benefits since it has more complexity than exhaustive search for high N_T. We can also see that the larger N_T is, the more beneficial the upward BAB search becomes, reducing the complexity by more than 2 orders of magnitudes for $N_T = 32$. Regarding the success rate, the only algorithm

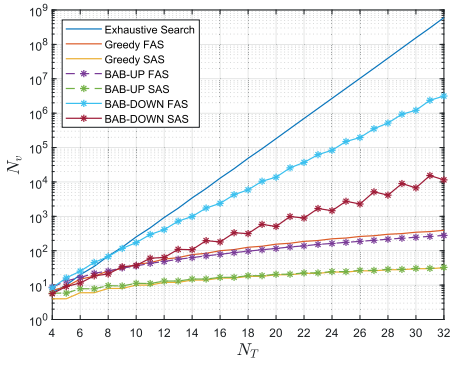


(a) Visited nodes case 3.

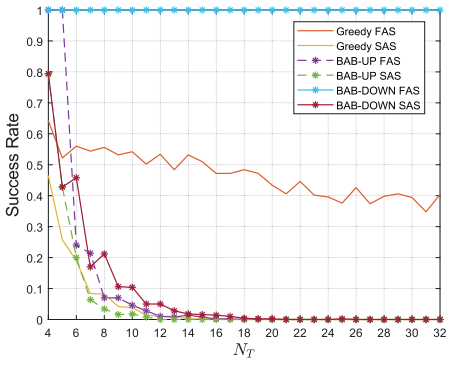


(b) Average success rate case 3.

Figure 3.6: Comparison of algorithms for case 3, $L = 4 > N_R = 2$.



(a) Visited nodes case 4.



(b) Average success rate case 4.

Figure 3.7: Comparison of algorithms for case 4, $L = \lfloor N_T/2 \rfloor > N_R = 2$.

obtaining the optimal solution is the downward BAB since (3.5) is satisfied. For the rest of the algorithms, as the possible number of combinations increases, their success rate worsens significantly, except for the greedy FAS search which also has a decreasing tendency but outperforms upward BAB FAS in success rate with fewer evaluations.

The results of case 4 are presented in Figure 3.7. This is the most interesting case since this will be the case replicated for the scenarios in the next chapter. Note that for cases 1 and 4, since L also grows, the absolute complexity also increases and the y-axis scale is larger. A behavior analogous to case 1 is observed. Here downward BAB has lower complexity than exhaustive search since L is not much smaller compared to N_T, in fact, it is half as much. For $N_T = 32$, the complexity is reduced by more than 2 orders of magnitude. Note in this case the high number of visited nodes of the figures reached almost 10^9 for 32 antennas with exhaustive search. For downward BAB more than 10^6 nodes are visited which is still a huge number, clearly differentiated from the rest of the algorithms. This could be problematic, even if the optimal solution is obtained. Upward BAB and greedy algorithms have similar complexity for each switching configuration. However, upward BAB FAS and greedy FAS appear to diverge, and upward BAB SAS and greedy SAS seem to converge with increasing N_T. The success rate is similar to case 3, with downward BAB being optimum and greedy FAS performing much better than the rest with slightly more evaluations than upward BAB FAS.

An alternative way to compare complexity is through computation time. In Tables 3.2 to 3.5 the average computation time for each case and algorithm for 4, 8, 16, and 32 available antennas is presented. This computation time is expressed in ms and the average has been performed over 500 channels. We should indicate that these execution times have been obtained from the execution of the algorithms in MATLAB® in a computer with 12th Generation Intel® Core™ i5-1245U @ 1.6GHz and 10 cores equipped with 16GB of RAM. Perhaps the computation time could be reduced on real and specific hardware where the computations could be optimized. This time would be closely related to the visited nodes but it allows a more practical interpretation of the complexity of each algorithm. For cases 1 and 4, exhaustive search has not been computed for 32 antennas since it is infeasible to carry out, and the corresponding BAB algorithms already obtain the optimal solution. For cases 2 and 3, downward BAB SAS has no value for 32 antennas as it stops at 24.

In general, greedy algorithms remain low in computation time for all cases, with longer time when L is greater, as in cases 1 and 4. Firstly, we have case 1 in Table 3.2, where upward BAB FAS is optimal with a reduction in computation time as well but still at a clearly expensive cost for high N_T. Regarding Table 3.3, case 2, it is interesting to notice that upward BAB algorithms also have decent execution times even slightly lower than the ones for exhaustive search. For downward BAB, the highest execution times are achieved. The algorithms have similar behavior in case 3, Table 3.4, as well, but the values are generally larger than in case 2. However, the reduction for sub-optimal algorithms, greedy and upward BAB, is about 3 orders in magnitude with respect to exhaustive search. Downward BAB algorithms, as previous figures have shown, obtain the highest computation times. The computation times for case 4 are shown in Table 3.5. Here the complexity increases. Downward BAB FAS achieves the optimal subset but it incurs many orders of magnitude higher the computation time compared to the rest of the sub-optimal approaches.

Average computation time (ms) - Case 1: $L = \lfloor N_T/2 \rfloor < N_R = N_T$							
N_T	**Exh. Search**	**Greedy FAS**	**Greedy SAS**	**BAB Up FAS**	**BAB Up SAS**	**BAB Do. FAS**	**BAB Do. SAS**
4	0.0726	0.2340	0.0307	0.0603	0.0299	0.2336	0.1388
8	0.4518	0.4337	0.0752	0.5461	0.1462	0.6668	0.2487
16	165.1650	1.2720	0.1915	53.9352	2.3247	3.2284	0.7330
32	-	7.3027	0.7332	1046900	354.7091	22.0179	2.9942

Table 3.2: Computation time comparison for case 1.

Average computation time (ms) - Case 2: $L = N_R = 2$.							
N_T	**Exh. Search**	**Greedy FAS**	**Greedy SAS**	**BAB Up FAS**	**BAB Up SAS**	**BAB Do. FAS**	**BAB Do. SAS**
4	0.0731	0.2336	0.0352	0.0649	0.0312	0.2739	0.1635
8	0.1523	0.2665	0.0435	0.1517	0.0802	1.8090	3.3403
16	0.4286	0.2945	0.0595	0.3875	0.2066	11.9042	777.6870
32	1.4354	0.3814	0.0887	1.0699	0.5742	98.5369	-

Table 3.3: Computation time comparison for case 2.

Average computation time (ms) - Case 3: $L = 4 > N_R = 2$							
N_T	**Exh. Search**	**Greedy FAS**	**Greedy SAS**	**BAB Up FAS**	**BAB Up SAS**	**BAB Do. FAS**	**BAB Do. SAS**
4	0.0248	0.0440	0.0141	0.0122	0.0087	0.0113	0.0089
8	0.3251	0.4493	0.0701	0.1364	0.0487	1.8636	0.5814
16	6.1310	0.5190	0.0919	0.3252	0.0959	47.9069	180.6572
32	113.4185	0.7504	0.1454	0.8080	0.2078	2070.1481	-

Table 3.4: Computation time comparison for case 3.

Average computation time (ms) - Case 4: $L = \lfloor N_T/2 \rfloor > N_R = 2$							
N_T	**Exh. Search**	**Greedy FAS**	**Greedy SAS**	**BAB Up FAS**	**BAB Up SAS**	**BAB Do. FAS**	**BAB Do. SAS**
4	0.0459	0.2085	0.0275	0.0552	0.0258	0.2460	0.1511
8	0.2973	0.4275	0.0679	0.1305	0.0467	1.8323	0.5707
16	44.4754	0.8037	0.1371	0.3337	0.0787	55.8107	4.2615
32	-	2.4417	0.2100	1.2751	0.1750	165287	280.4314

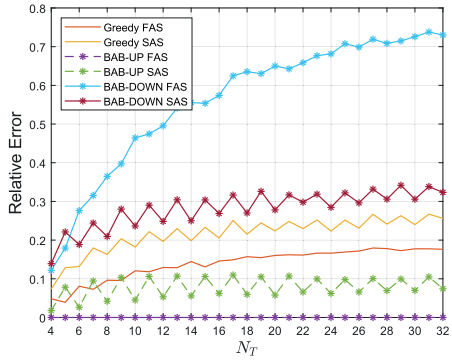
Table 3.5: Computation time comparison for case 4.

The success rate figures presented previously represent the accuracy performance of the algorithm. If they are sub-optimal, the success rate is expected to decrease as the combinations grow. However, the visited nodes and computation time and thus the complexity also increase significantly for the optimal algorithms. The next question to be answered is how different is the overall performance of these sub-optimal antenna subsets. That performance will be measured in terms of the MSV obtained, so the MSV of each algorithm will be compared to the maximum MSV corresponding to the optimal subset, obtaining the relative error. The idea is to have a more reasonable trade-off between complexity and performance depending on the case and perhaps a few orders of magnitude in complexity can be saved with only a minor performance loss.

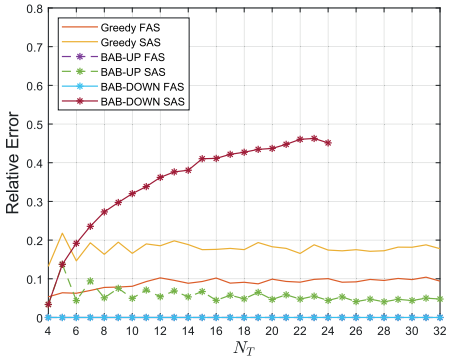
Figure 3.8 shows the relative errors of the MSVs obtained by each algorithm for the different cases. The results for case 1 are plotted in Figure 3.8(a). Here, upward BAB FAS is optimal and with a SAS topology, the error fluctuates by over 10%. Greedy algorithms show less relative error than the downward BAB algorithms with similar visited nodes but less computation time. For case 2, it can be observed from Figure 3.8(b) that, although downward BAB SAS has the greatest complexity, passing through more nodes than exhaustive search, it is the one with the largest relative error. As mentioned before, this occurs because the closer L is to N_T, the more the complexity will be reduced with downward BAB algorithms. On the other hand, if L is small compared with N_T, more columns will have to be discarded so the number of visited nodes increases. One can also see that upward BAB SAS obtains a lower relative error than both greedy algorithms with just a few more nodes evaluated than greedy SAS but in the same order of computation time as greedy FAS.

As seen in Figure 3.8(c), for case 3, the greedy FAS algorithm achieves an acceptable relative error with a reduction of more than 3 orders of magnitude in complexity for 32 antennas. Both upward BAB algorithms although having more complexity (but similar computation times) than their greedy counterparts, yield more relative error. Observe that in both cases 2 and 3, as the difference between a fixed L and variable N_T increases, the error for downward BAB SAS stabilizes.

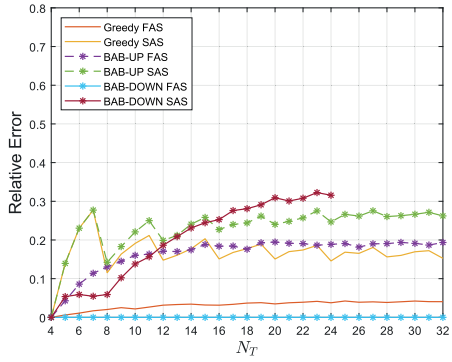
Case 4 is depicted in Figure 3.8(d). Here the greedy FAS algorithm also has a very small relative error, less than 0.1%, with a reduction of about 4 orders of magnitude in visited nodes and 5 orders of magnitude in computation time compared to downward BAB FAS for 32 antennas, being the latter optimal. The error difference between greedy SAS and downward BAB SAS is not very large with a reduction of more than 2.5 orders of magnitude in visited nodes and a significant reduction in computation time too. Both upward BAB algorithms have comparable errors with a reduction of 1 order of magnitude in complexity for the SAS network.



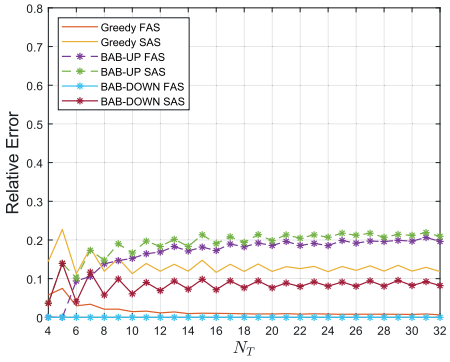
(a) Relative error case 1, $L = \lfloor N_T/2 \rfloor < N_R = N_T$.



(b) Relative error case 2, $L = N_R = 2$.



(c) Relative error case 3, $L = 4 > N_R = 2$.



(d) Relative error case 4, $L = \lfloor N_T/2 \rfloor$, $N_R = 2$.

Figure 3.8: Relative errors of algorithms for all the cases.

In the following chapter, these algorithms will be implemented into the simulator to find an optimal or a sub-optimal subset of the antennas. The goal is to analyze what performance can be achieved when AS is carried out at the BS under some scenarios.

Chapter 4

AS Simulations Performance

In the previous chapter, antenna selection (AS) has been proposed as a feasible solution to alleviate the complexity of MIMO systems. For this purpose, an AS criterion has been defined, and several search algorithms have been described to obtain a high-quality subset out of the total number of available antennas at the BS. This chapter presents the scenarios where the AS will be implemented and what results can be achieved.

4.1 System Configuration and Scenarios

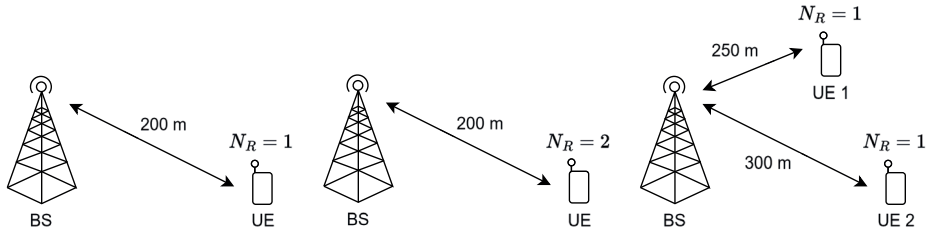
The simulator has been configured with the system configuration parameters found in Table 4.1. Although three sectors, dual-polarization, and indoor UEs are allowed, only one sector, single-polarization, and no indoor users are considered. These parameters are used throughout all the different scenarios.

Parameters	**Value**
Scenario	UMi
No. of sectors	1
Carrier frequency f_c	2.6 GHz
Sampling frequency f_s	30.72 MHz
Numerology μ	0
OFDM Symbol Time	66.67 µs
Bandwidth B	20 MHz
Subcarrier spacing (SCS) Δ	15 kHz
N_{FTT}	2048
Active Subcarriers	1332
Array geometry	ULA
Antenna spacing	$\lambda/2$
Constellation	QPSK
No. of antennas at BS	$\{16, 32, 64, 128\}$
No. OFDM symbols	14
Transmit Power BS	46 dBW
Transmit Power UE	-7 dBW
No. of UEs	$\{1, 2\}$
No. of antennas at UEs	$\{1, 2\}$
SNR range	$[-14, 15]$
LoS	No

Table 4.1: Simulation parameters.

The cases and scenarios simulated throughout this chapter are described below.

- In Section 4.3, a MISO system BER comparison is presented between $N_T = 32$ and 16 antennas at the BS when AS is applied for different numbers of L antennas selected. There is only one single antenna UE located in broadside direction 200 m from the BS. The number of channel realizations per BER curve, i.e, per antenna configuration and AS algorithm, is 200.
- In Section 4.4, a MIMO system capacity comparison is presented between $N_T = 32$ and 16 antennas at the BS when AS is applied for different numbers of L antennas selected. There is only one UE with two receive antennas located in broadside direction 200 m from the BS. 2048 subcarriers are used to obtain the Fourier transform, but then, only the 1332 active central subcarriers are considered to calculate the average capacity per subcarrier. A total of 2500 channel realizations have been simulated to compute the capacity distribution.
- In Section 4.5, MISO and MIMO systems BER comparisons are presented between $N_T = 128$ and 64 antennas at the BS when AS is applied for a fixed $L = N_T/2$ antennas selected. The number of channel realizations per BER curve is 100. The proposed scenarios can be observed in Figure 4.1. A 32 antenna configuration has also been included for benchmarking purposes in Scenarios 1 and 2:
  - Scenario 1: a MISO scenario with 1 UE placed at 200 m from the BS (along the broadside direction of the BS) equipped with 1 transmit and receive antenna.
  - Scenario 2: a MIMO scenario with 1 UE placed at 200 m from the BS (along the broadside direction of the BS) equipped with 1 transmit antenna and 2 receive antennas.
  - Scenario 3: a MU-MIMO scenario with 2 UEs placed at 250 and 300 m, both equipped with 1 transmit and receive antenna. In Figure 4.2, the layout for the scenario is presented.



(a) Scenario 1. **(b)** Scenario 2. **(c)** Scenario 3.

Figure 4.1: Scenarios with $N_T = 128$ and 64 antennas at the BS.

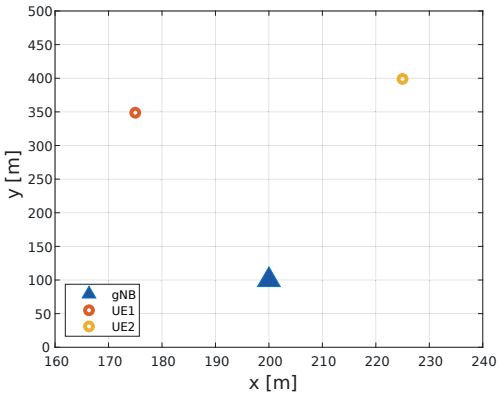


Figure 4.2: Scenario 3 layout.

4.2 Methodology

As stated in Chapter 2, the system duplexing mode is TDD. In a TDD system, the frequency used for downlink and uplink transmission is the same and the transmission in each link occurs in different time slots. The steps followed are depicted in Figure 4.3. First, the UE sends the SRS pilots symbols. The CSI is estimated at the BS and since channel reciprocity is considered, that channel information is used in the beamforming so the precoding matrix is generated and the BS, gNB, transmits the downlink data.

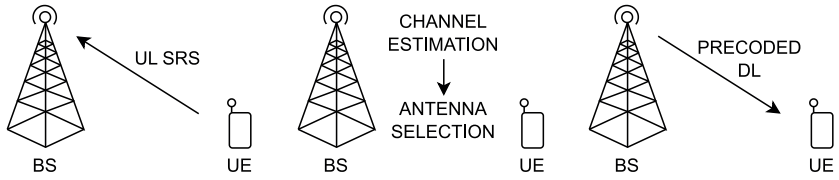


Figure 4.3: Uplink and downlink structure for beamforming in TDD.

In Figures 4.4 and 4.5, examples of two channels generated for one UE with 1 antenna to a BS with 128 antennas are shown. Applying the Fourier transform to the impulse response $B(n, \tau)$ with respect to the delay (τ) results in the transfer function $H(n, f)$ or frequency response for antennan. The channel coefficients are similar for antennas located nearby as the antenna coefficients correlation matrices show in Figures 4.6(a) and 4.6(b). However, the channel coefficients are different in the delay domain, and some channels are more favorable than others, i.e., higher or lower coefficient gain. In Appendix A, in Figures A.1(a) and A.1(b), a cross-section of the channel coefficients for a particular antenna for each channel can be observed. The distribution for both channel examples can be observed in the Appendix as well in Figures A.2(a) and A.2(b) where it can be noticed that the absolute amplitude follows a Rayleigh distribution which is a particular case of the Rician distribution with non-line-of-sight (NLOS). There is no dominant LOS component since both distributions coincide.

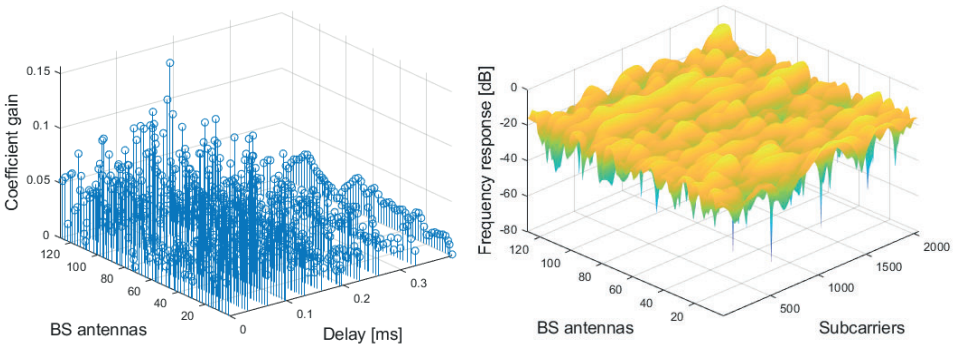
To estimate and compare properly system performance in different scenarios, e.g., different antenna configurations at both BS or UE or UEs placed in different locations, the channel power may need to be normalized to make fair comparisons. The idea is to preserve only the relative power variations over time and frequency and more importantly over the different antenna arrays so that the power is distributed evenly across all transmitting antennas. Path loss and shadowing are compensated due to noise scaling, however, small-scale coefficients should be normalized so that the expectation of the sum of all cluster delays power is 1. The normalized channel matrix containing only small-scale coefficients can be expressed as

$$\mathbf{B}_n^{\text{norm}} = \sqrt{\frac{N_R N_T N_{ch}}{\sum_{n=1}^{N_{ch}} \|\mathbf{B}_n\|_F^2}} \mathbf{B}_n, \tag{4.1}$$

where $\mathbf{B}_n$ is the $N_R \times N_T \times N_{\text{taps}}$ channel matrix corresponding to particular n-th channel realization out of a total of N_{ch} realizations. In that way, considering $h_{i,j,l,n}$ as the coefficient between the i-th receive and j-th transmit antenna respectively corresponding to the l-th tap and n-th channel realization, the following expressions are verified,

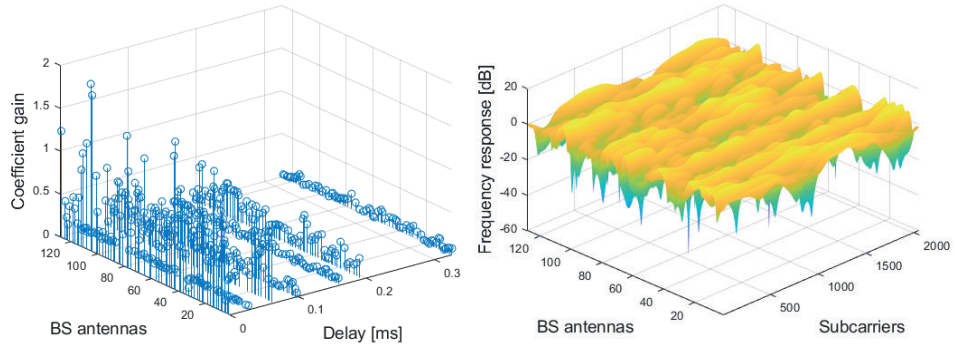
$$\mathbb{E}\left\{ \sum_{l=0}^{N_{\text{taps}}-1} \left| h_{i,j,l,n}^{\text{norm}} \right|^2 \right\} = 1, \tag{4.2}$$

$$\frac{1}{N_{ch}} \sum_{n=1}^{N_{ch}} \sum_{i=1}^{N_R} \sum_{j=1}^{N_T} \sum_{l=0}^{N_{\text{taps}}-1} \left| h_{i,j,l,n}^{\text{norm}} \right|^2 = N_R N_T. \tag{4.3}$$



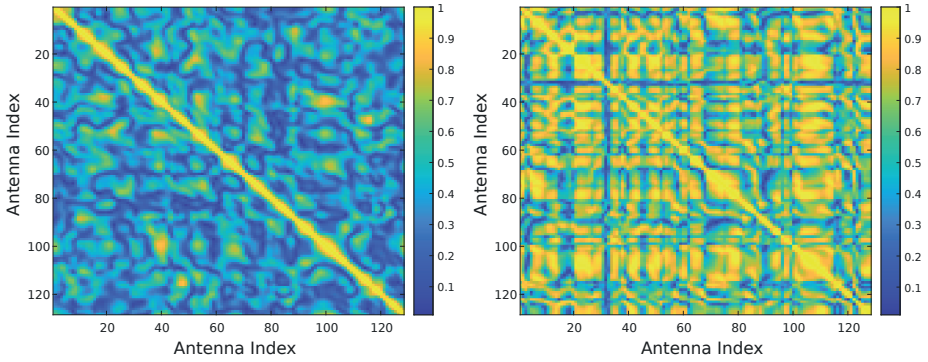
(a) Channel coefficients gain channel 1. **(b)** Frequency response channel 1.

Figure 4.4: Example of channel generation 1.



(a) Channel coefficients gain channel 2. **(b)** Frequency response channel 2.

Figure 4.5: Example of channel generation 2.



(a) Channel 1. **(b)** Channel 2.

Figure 4.6: Antenna channel coefficients correlation.

4.2.1 Antenna Selection Matrix

Before analyzing AS algorithm performances, the $\mathbf{H}$ matrix, from which the $\mathbf{H}_S$ with the L selected BS antennas will be obtained, has not yet been fully described. As said in Section 2.2.4, the channel matrix estimated is a time-frequency $Q \times N_{\text{symb}}$ matrix $\mathbf{H}(t, f)$ where Q is the number of active subcarriers, 1332, and N_{symb} the OFDM symbols in one subframe, which for the chosen numerology, $\mu = 0$, has 14 symbols. The pilot symbols have been allocated to the last four OFDM symbols, and as described in Section 2.3.1, if there is more than one user, these are multiplexed over the subcarriers. The trade-off between channel-estimation accuracy and pilot symbols overhead might be relevant since this estimation will affect the next processes, however, the number of pilot symbols will be fixed to four. The simulator generates a $Q \times N_{\text{symb}} \times N_T$ channel tensor $\mathcal{H}$ from the UL estimates. The UEs in the simulator only transmit on one antenna, even if they have more. Therefore, the AS in the BS is based only on that information. In DL, however, the UE receives from all its antennas. In other words, in UL, the rank of the channel is always considered to be 1, and in DL, it will be equal to the number of antennas of the UEs. The rank of the channel also corresponds to the number of layers.

To perform the AS, a dimension reduction has to be carried out. The resulting matrix must describe the channel between the base station antennas and all receiving antennas. The dimension reduction is done along the subcarriers and OFDM symbols for each transmit antenna such that the channel coefficients at the central frequency are chosen. We can finally build the necessary $1 \times N_T$ channel matrix $\mathbf{H}$ for each UE. More intense processes could be applied to reduce dimensions as well. For instance, AS selection could be applied for each frequency-symbol coefficient of all the transmit antennas, and then select the subset with the largest MSV. However, that would be more costly since at least Q AS subsets would have to be calculated (it is Q and not $Q \times N_{\text{symb}}$ because the channel estimated is assumed to be the same for all the symbols of a particular subcarrier). This process can be observed in Figure 4.7, where we have the $Q \times N_{\text{symb}}$ matrices for all the transmit antennas. The central frequency coefficient is taken out for each matrix to generate the $\mathbf{H}$ where the AS will be performed to obtain the final $\mathbf{H}_S$. This procedure could be extended so that in the case of the UE transmitting from more than one antenna, the BS would have another dimension so the tensor $\mathcal{H}$ would be $Q \times N_{\text{symb}} \times N_R \times N_T$ resulting in a $N_R \times N_T$ channel matrix $\mathbf{H}$.

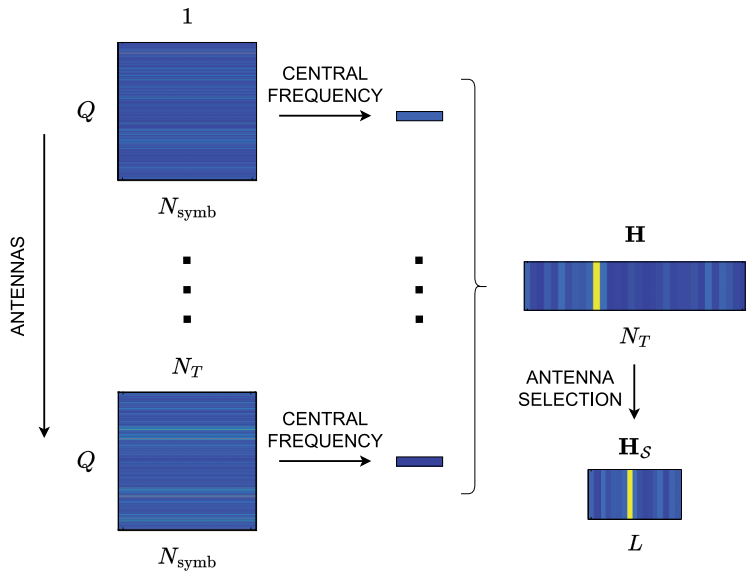


Figure 4.7: Obtaining $\mathbf{H}$ matrix for AS.

4.3 BER Comparison for 32 and 16 Antennas

In Chapter 3, a total of six algorithms have been proposed to tackle the AS problem. AS has been understood as a subset search based on the largest MSV criterion function so that the lower bound of the SNR is increased. These algorithms were conceived to reduce the computational time or complexity that it takes for the exhaustive search which becomes infeasible even for a not-so-large N_T, L pair. However, as it has been seen in the algorithm comparison, if a global optimal solution is to be obtained, although the visited nodes have been decreased in relation to the exhaustive search, the complexity that it entails, remains high.

It is important to highlight here that, in the BER calculation process, the necessary bits have been sent according to the chosen constellation and the time and frequency resources in a subframe. To ensure the reliability of the results, thresholds have been established for the minimum number of errors obtained and bits transmitted, so that the loop is not exited until the threshold is exceeded. The thresholds are 10^5 bits transmitted and the errors obtained following a decreasing function that depends on the SNR and the number of antennas, being lower for high SNR. If more than 10^5 bits have been transmitted, but there is still not a sufficient number of errors, the threshold of transmitted bits is increased to 10^7, regardless of the errors obtained. That may lead to not very reliable results for high SNR because the number of errors and bits transmitted may not be very representative. Similarly, the more channel realizations that are included, the more realistic the results will be since they will be averaged over both favorable and bad cases. However, this incurs a longer simulation time. With the selected values, for each case of L antennas selected, each simulation configuration takes the time in the order of one week to complete.

The results for the first case, where $L = \lfloor N_T/5 \rfloor$, is displayed in Figures 4.8 and 4.9 for FAS and SAS respectively. When having 16 antennas at the BS, 3 antennas are selected and when 32 antennas are available, 6 antennas are active. For 16 antennas, upward BAB FAS performs the best compared to Greedy FAS and downward BAB FAS, the latter two of which are very similar. The algorithms for AS with 32 antennas have a more expected behavior. Downward BAB FAS presents the lowest BER curve followed by Greedy FAS and lastly upward BAB FAS. Interestingly, in the SAS case, the algorithms obtain better results. Greedy SAS and upward BAB SAS show almost identical curves for both antenna configurations. Downward BAB SAS complexity and computational time increase greatly for 32 antennas since more antennas need to be removed, so it is not included. For 16 antennas, the SNR gap when AS with $L = \lfloor 0.20 N_T \rfloor$ is performed has a loss of around 10 dB and a gain of 2 dB at a BER of 10^{-4} relative to the N_T ("DL 16") and L ("DL 3") systems respectively. In general, for the 32 antenna configuration, the corresponding SNR loss is less than 10 dB for both FAS and SAS cases, relative to the N_T system ("DL 32"). On the other hand, the SNR gain is around 1 dB in the FAS cases and around 3 dB for the SAS cases at a BER of 10^{-5}, relative to the L system ("DL 6").

In Figure 4.10, $L = N_T/2$ means that 8 and 16 antennas are chosen and the results for the FAS algorithms are shown. The algorithms behave similarly for both antenna configurations. Greedy FAS seems to act as a lower bound. For the SAS case displayed in Figure 4.11, upward BAB appears to provide larger performance benefits (in terms of SNR gap) with increasing SNR, relative to the other cases. Greedy SAS and downward BAB SAS are very similar for both 16 and 32 antennas. For 16 antennas the SNR gap between AS and the N_T ("DL 16") and L ("DL 8") systems is around 4 dB at a BER of 10^{-5} for both systems and switching configurations. The SNR gain is around 1 dB and slightly more for the SAS configuration. For 32 antennas, at a BER of 10^{-5}, there is an SNR loss of approximately 3 dB compared to the N_T system ("DL 32") and an SNR gain of 0.5 dB relative to the L system ("DL 16").

Figures 4.12 and 4.13 illustrate the case for $L = \lfloor 4N_T/5 \rfloor$ where 12 and 25 antennas are selected with FAS configuration and SAS configuration, respectively. Since the number of discarded antennas is small, the downward BAB algorithms perform slightly better for both antenna and switching configurations. In this case, the performance obtained with AS is closer to the L system ("DL 12" and "DL 25") with an SNR loss of 2 dB and 1 dB approximately at a BER of 10^{-5} for

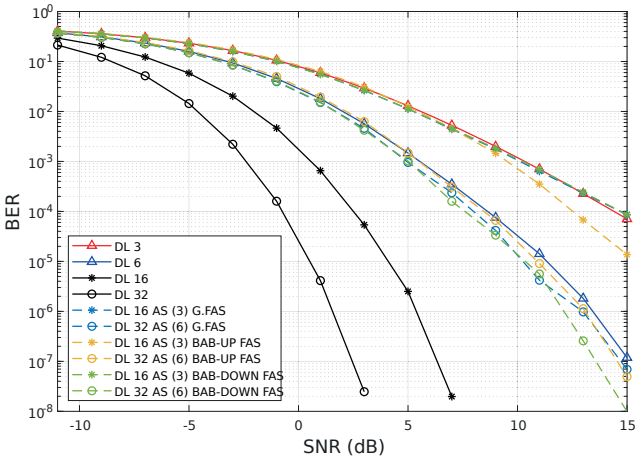


Figure 4.8: Average BER for $N_T = \{16, 32\}$ antennas with AS $L = \lfloor N_T/5 \rfloor = \{3, 6\}$ for FAS configuration.

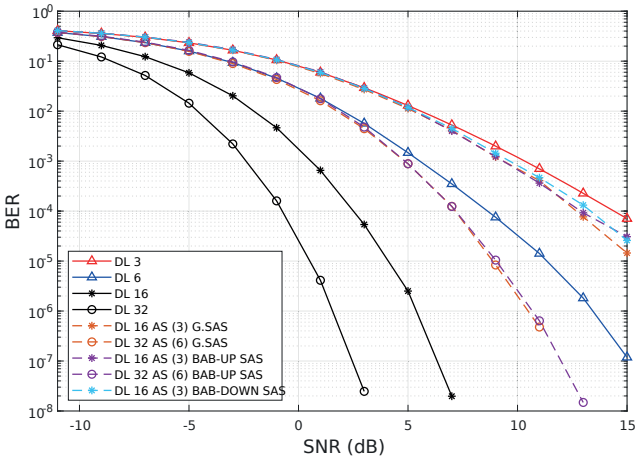


Figure 4.9: Average BER for $N_T = \{16, 32\}$ antennas with AS $L = \lfloor N_T/5 \rfloor = \{3, 6\}$ for SAS configuration.

16 and 32 antennas. In this last case, it seems that AS does not provide any substantial gain (less than 0.5 dB) except for downward BAB SAS for the 16-antenna configuration.

AS incurs some performance loss or degradation compared to full complexity systems due to the power loss and diversity gain loss from removing the $N_T - L$ antennas [24]. However, in any case, we can observe that when AS is applied, the performance is between the $N_T \times 1$ and $L \times 1$ MISO systems and this insight can be also extrapolated to MIMO systems. This was the main purpose of AS, to reduce the number of active antennas while keeping some benefits of the complete antenna array.

As for the algorithms, at least for this number of available antennas, none excels in comparison to the others and in general there are no significant differences in terms of results. However, there

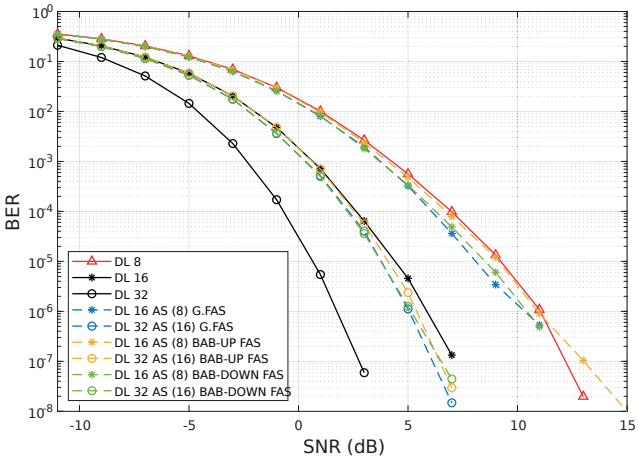


Figure 4.10: Average BER for $N_T = \{16, 32\}$ antennas with AS $L = N_T/2 = \{8, 16\}$ for FAS configuration.

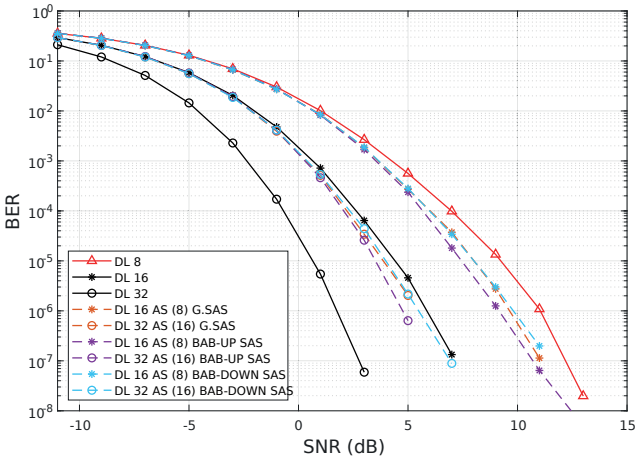


Figure 4.11: Average BER for $N_T = \{16, 32\}$ antennas with AS $L = N_T/2 = \{8, 16\}$ for SAS configuration.

are significant differences in terms of execution time and complexity as we have seen in the previous chapter. We should be reminded here that the case study in all the scenarios presented in this chapter corresponds to the condition of expression (3.5) where the number of receive antennas is less or equal to the number of selected antennas, i.e., $N_R \leq L$. Specifically, the simulations follow case 4 analyzed in the previous chapter, in which the relative errors of the algorithms are less dispersed and close to each other, as shown in Figure 3.8(d). For the other condition (3.4), the results might be different. Systems with L transmit antennas without AS can be considered as fulfilling the random AS criterion and thus they are expected to act as an upper bound for any of the proposed algorithms. In addition, the AS process is based on estimates and not perfect CSI which probably gives more realistic results, also acting as an upper bound for simulations with perfect channel knowledge.

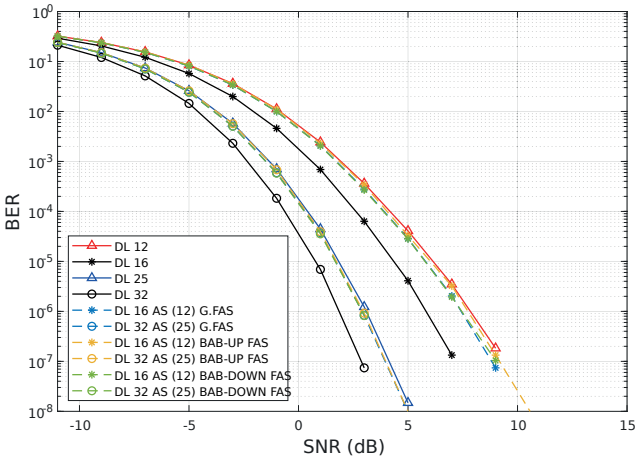


Figure 4.12: Average BER for $N_T = \{16, 32\}$ antennas with AS $L = \lfloor 4N_T/5 \rfloor = \{12, 25\}$ for FAS configuration.

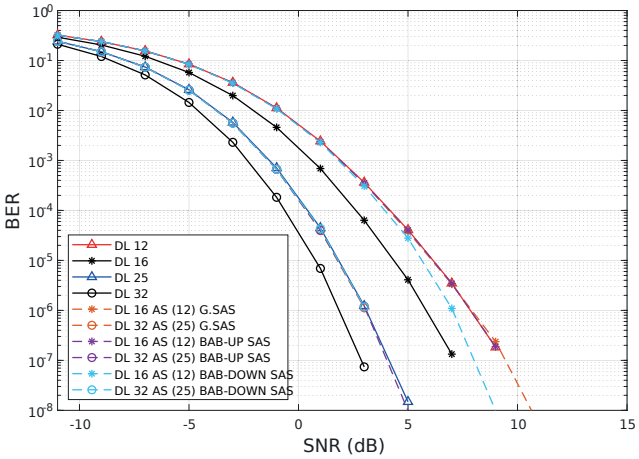
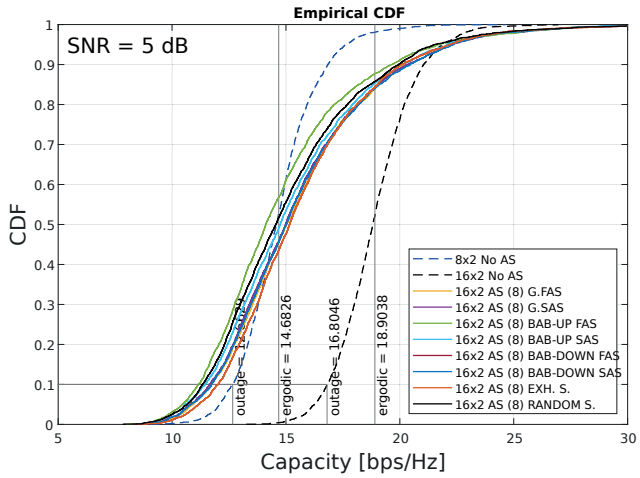


Figure 4.13: Average BER for $N_T = \{16, 32\}$ antennas with AS $L = \lfloor 4N_T/5 \rfloor = \{12, 25\}$ for SAS configuration.

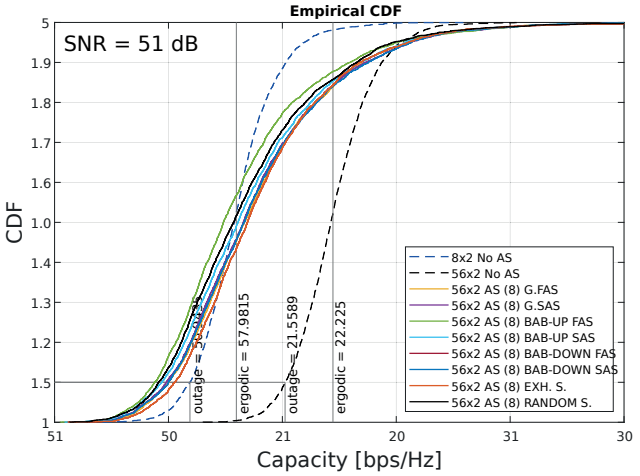
Regarding the percentage of antennas selected, if this number is high with only a few antennas discarded, any antenna subset that the algorithms obtain little differs from a random antenna subset and thus will obtain comparable results. This is precisely what occurs in the last case with $L = \lfloor 4N_T/5 \rfloor$ (refer again to Figures 4.12 and 4.13). Also, when the percentage of antennas selected is high, BAB downward can obtain the optimal solution at a low computation cost. However, even for situations where less than half the antennas are chosen, the complexity and number of evaluations for this algorithm increase greatly, which do not justify the marginal improvements from optimal AS, if any. Even with a greedy approach that is sub-optimal, similar or better results could be obtained. For the next simulations with a larger number of antennas, the Greedy FAS and upward BAB FAS algorithms will be used among the rest to have a deeper understanding of these two algorithms.

4.4 Capacity Comparison for 32 and 16 Antennas

The AS problem has been proposed in the literature mainly to optimize capacity following (3.3). Although the algorithms presented in this master thesis have not been conceived for optimizing capacity, they can be compared in that regard. A total of 2500 channels have been generated for $N_T = 16$ and $N_T = 32$ antennas with $L = 8$ and $L = 25$ ($L = N_T/2$ and $L = \lfloor 4N_T/5 \rfloor$), respectively. The channel is assumed to be perfectly known by the transmitter, so the waterfilling algorithm can be applied.



(a) SNR = 5dB.

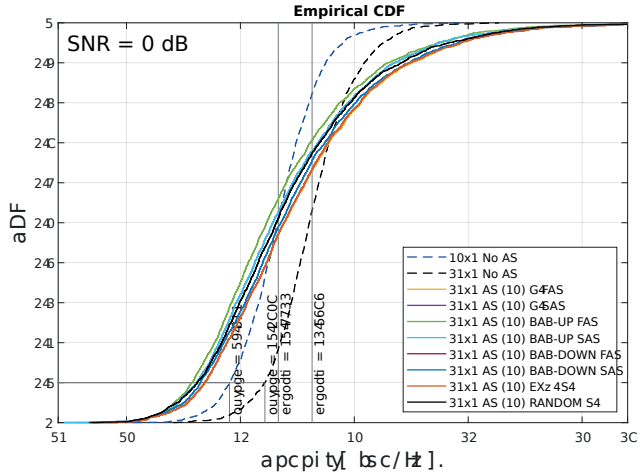


(b) SNR = 10dB

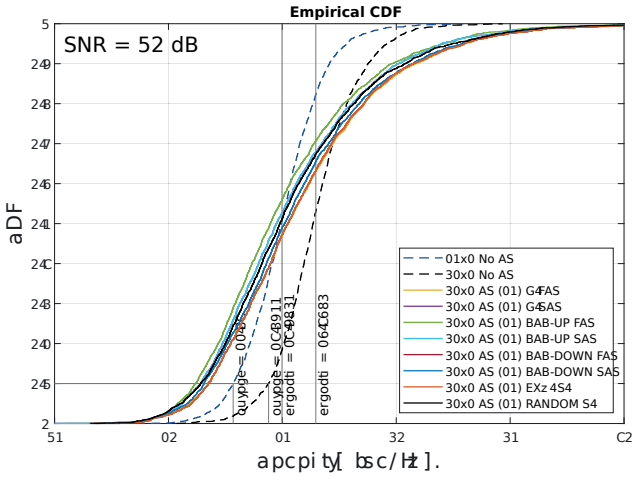
Figure 4.14: CDF of the capacity for 16×2 MIMO system and AS with $L = N_T/2 = 8$.

In Figures 4.14 and 4.15 the cumulative distribution functions (CDFs) of the channel capacity per subcarrier are presented for 16 and 32 antennas, respectively. The behavior is very similar in both cases. Exhaustive and random searches have been included as references or benchmarks

to compare the algorithm performances. Note that here, downward BAB is still optimal for the MSV criterion, but not upward BAB. For 16 antennas in Figures 4.14(a) and 4.14(b) upward BAB performs the worst of all the algorithms, and upward BAB FAS is even worse than the random search. Greedy SAS and downward BAB SAS have similar curves whereas Greedy FAS is slightly better than them. Since downward BAB is optimal, the result is the same as the exhaustive search that overlaps the former, obtaining the highest capacity of all algorithms for both antenna configurations.



(a) SNR = 5dB.

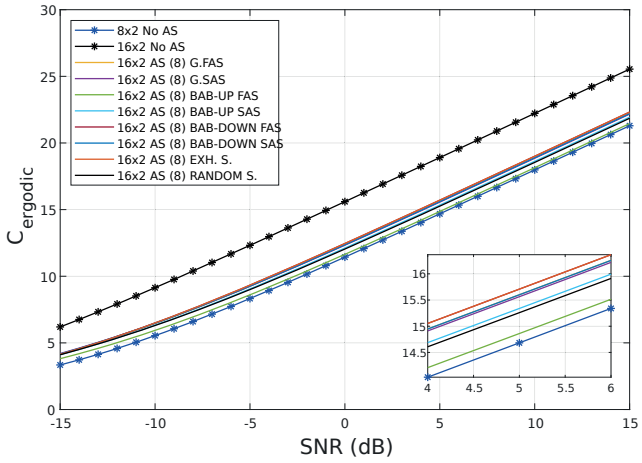


(b) SNR = 10dB

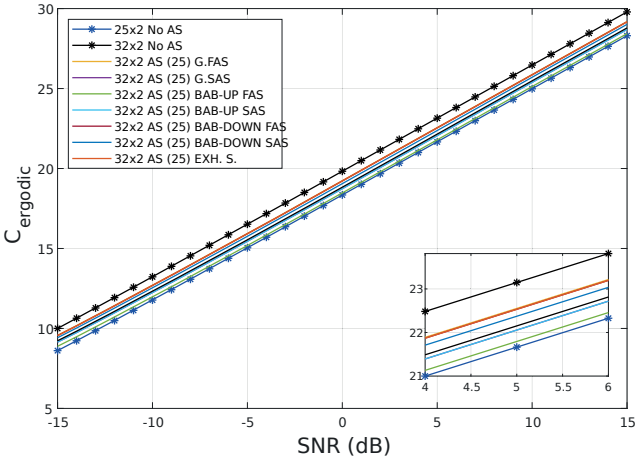
Figure 4.15: CDF of the capacity for 32×2 MIMO system and AS with $L = \lfloor 4N_T/5 = 25 \rfloor$.

For 32 antennas in Figures 4.15(a) and 4.15(b), the differences are that now upward BAB SAS also performs worse than the random search but is still better than upward BAB FAS. Greedy SAS has the same CDF curve as downward BAB SAS. Greedy FAS and downward BAB FAS are likewise fairly comparable in this case.

It is interesting to notice the outage capacities. The outage capacity quantifies the capacity that is guaranteed with a certain percentage or reliability of the channel realizations. For a 10% outage capacity, i.e., $P(C \leq C_{out,10}) = 10\%$, the no AS system performs better. That is because of the lack of diversity as the *selected* antennas are always the same for all the channels. AS algorithms designed to maximize capacity should consider this outage capacity so that it does not worsen in relation to the case of no AS.



(a) 16×2 system.



(b) 32×2 system

Figure 4.16: Ergodic capacity versus SNR from -15 to 15 dB.

Another important value is the ergodic capacity which has been defined as the expectation of the capacity over all the channel realizations, i.e., $C_{ergodic} = \overline{C} = \mathbb{E}\{C\}$, which is also shown in Figures 4.14 and 4.15. Note that the ergodic capacity is not the same as the 50th percentile. The former represents the mean, whereas the latter is the median. The steeper the CDF curve is, the more similar the mean and the median will be.

The median capacity of the random selection is the same as that of no AS, and this is to be expected since performance should be the same for half of the channel realizations due to the lack of any selection criterion. Except for the upward BAB, the rest of the algorithms have better median capacities than the random selection. The ergodic capacities versus SNR for 16 and 32, respectively, are shown in Figures 4.16(a) and 4.16(b). On average all algorithms maximized the channel capacity per subcarrier.

4.5 Scenarios Simulation Results

Monte Carlo simulations have been conducted with a total of 100 channel realizations for the three scenarios. In addition, some thresholds for the number of bits transmitted and errors obtained have been determined to obtain more reliable results. All the channels are generated and normalized first to be used in each run. The number of antennas selected has been fixed to be half of the total available antennas at the BS, i.e., $L = N_T/2$. Although the downward BAB algorithm obtains the optimal subsets, due to its high computation time and previous results for this selected antenna percentage, it has been discarded.

The number of selected antennas L will depend on N_T in the simulation and analysis of the scenarios, but the performance of different antenna configurations can be compared when L is fixed for all of them. Some gain is expected to be obtained when $L = 32$ antennas are selected out of 128 total available antennas compared to when there are 64 due to the increase in diversity. Figure 4.17 illustrates the BER for scenario 1 with $L = 32$. Although there seems to be some gain, it is marginal, and more appropriate simulations should be conducted to confirm whether it is beneficial to increase the number of available antennas to obtain better performance, because an increase in N_T even with a fixed L also leads to a higher computational cost as discussed in the previous chapter in cases 2 and 3.

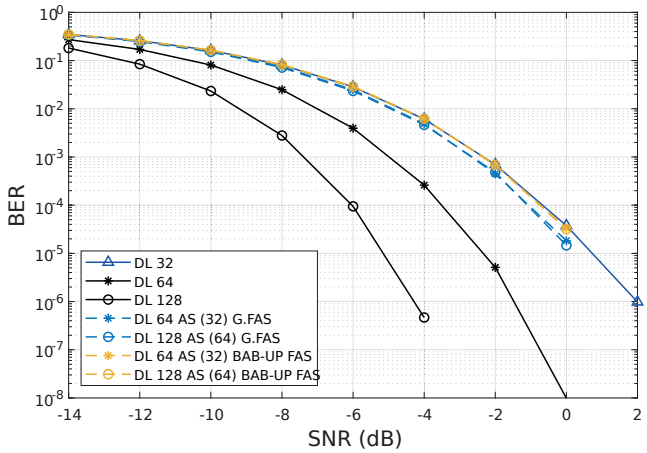


Figure 4.17: Average BER for Scenario 1 with 1 UE with $N_R = 1$ and $N_T = \{128, 64\}$ antennas with fixed AS $L = 32$ for FAS configuration.

Moving on to the scenarios, Figure 4.18 illustrates Scenario 1's average BER. In Scenario 1, the BS has 128 or 64 available antennas and the antennas selected are $L = N_T/2$, i.e., 64 and 32 respectively with a single antenna UE. Greedy FAS performs better than upward BAB which has a very similar performance as the no AS systems. However, the gain of Greedy FAS is barely around 0.25 dB at a BER of 10^{-5} for both antenna configurations. Scenario 2 is shown in Figure 4.19. Here although the AS is performed based on the channel estimation of the transmit antenna, the

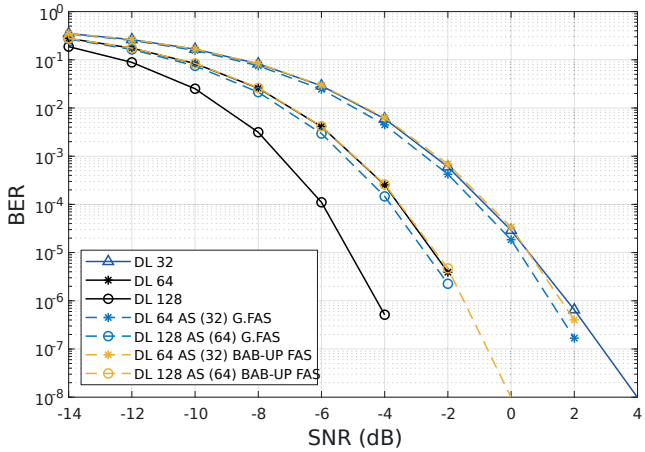


Figure 4.18: Average BER for Scenario 1 with 1 UE with $N_R = 1$ and $N_T = \{128, 64\}$ antennas with AS $L = N_T/2 = \{64, 32\}$ for FAS configuration.

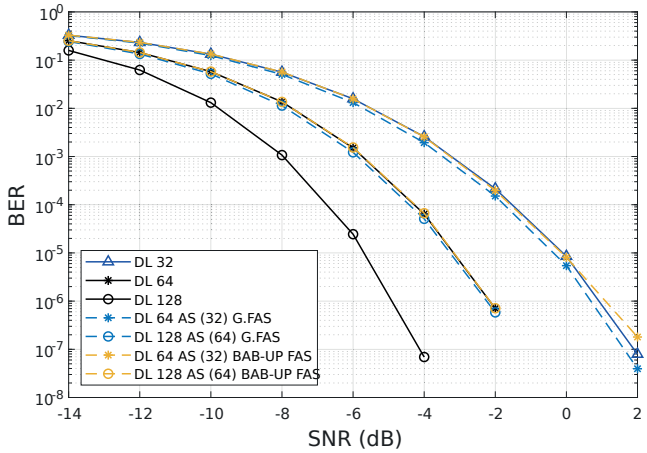


Figure 4.19: Average BER for Scenario 2 with 1 UE with $N_R = 2$ and $N_T = \{128, 64\}$ antennas with AS $L = N_T/2 = \{64, 32\}$ for FAS configuration.

UE has two receive antennas and thus some diversity gain is anticipated compared to Scenario 1. The algorithms present similar behaviours but the gain is even lower in this case with the results being very similar to the no AS systems. This is probably due to not implementing a suitable AS for both receive antennas. In Figure 4.20 both Scenarios 1 and 2 are compared. The diversity gain of having two antennas at the UE can be noticed in the increased magnitude of the slope of the BER curves.

Finally, the result for Scenario 3 is depicted in Figure 4.21. In this case, there are 2 single antenna UEs at 250 m (UE 1) and 300 m (UE 2), respectively. Here, the UEs are not located in the boresight of the BS antenna array as in the previous scenarios, but diagonally as shown in Figure 4.2. The AS is now done for both UEs so the channel tensor $\mathcal{H}$ defined in the previous has another dimension. In this way, the channel matrix $\mathbf{H}$ after the dimension reduction becomes $2 \times N_T$. The AS selects the subset with the largest MSV for both UEs together but not the optimal

for each particular one. However, an improvement in BER performance may be expected compared to the system without AS. As observed in the figure, the gain is also marginal. Both algorithms perform very similarly with slightly better results with Greedy FAS for UE 2. Remember that the transmitter does not have perfect CSI so the channel estimation contains some error. Nevertheless, the approach and methods could be improved when having several UEs with a more effective joint AS algorithm.

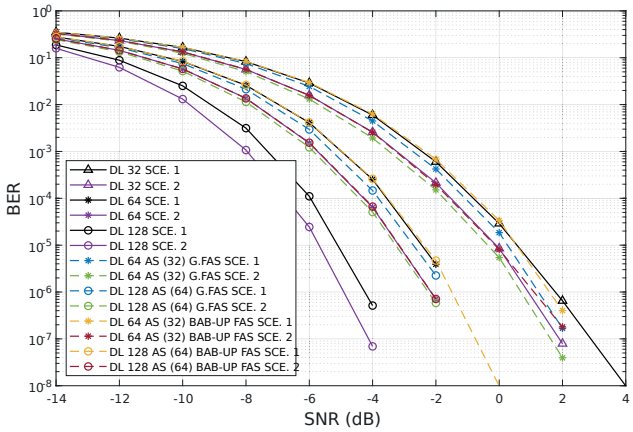


Figure 4.20: Average BER comparison for Scenario 1 and Scenario 2.

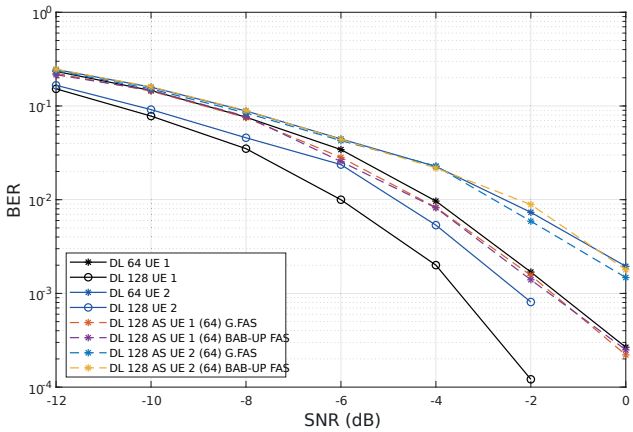


Figure 4.21: Average BER comparison for Scenario 3 with 2 single antenna UEs ($N_R = 1$) and with $N_T = \{128, 64\}$ antennas with AS $L = N_T/2 = \{64, 32\}$ for FAS configuration.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

In this master thesis the AS problem has been described as a potential solution to alleviate the difficulties and complexities in deploying massive MIMO systems with a large number of antennas at base stations, but still retaining much of the benefits of the full complexity massive MIMO. The massive MIMO scenario was defined and a new channel model following the 3GPP TR 38.901 Release 16 has been included in the MATLAB® simulator. Assuming TDD and reciprocity, the CSI obtained from uplink SRS can be exploited at the BS for transmission.

More specifically the goal was to obtain a high-quality subset of antennas from the BS to transmit in downlink. For this purpose, the largest minimum singular value (MSV) has been determined as the selection criterion in the objective function. The submatrix with the largest MSV ensures a large minimum SNR which is optimal to ensure low BER, that is with respect to the diversity order for linear receivers. Hardware components and RF switches are also critical parts of practical MIMO systems. The switching network could be a conventional fully-switching RF network (FAS) but a sub-array switching (SAS) architecture could be considered instead. The SAS configuration reduces the hardware complexity compared to the FAS network at the expense of limiting the connectivity and thus the possible antenna subsets to be selected.

A sub-optimal and optimal AS approaches for each switching configuration have been proposed based on the largest MSV criterion. A greedy search algorithm is described which is sub-optimal but low complexity. Branch-and-Bound guarantees the global optimal solution with lower computational complexity than exhaustive search if certain conditions related to the size of the channel matrix are fulfilled. If the total number of antennas of the UEs is greater than the antennas selected at the BS, i.e., $L < N_R$, an upward BAB search algorithm goes through the search tree adding columns. On the other hand, if the total number of antennas of the UEs is lower than the antennas selected, i.e., $L \geq N_R$, we have a downward BAB search algorithm removing columns from the full channel matrix. A few cases with different channel matrix sizes and antennas selected have been performed to compare the complexity and performance of the algorithms. In particular, case 4 is the most relevant with $L > N_R$ where downward BAB is optimal with fewer evaluations than exhaustive search but still at a high computational cost. Greedy search yields a low relative error even for a significant number of antennas and thus combinations.

Several simulations have been carried out with 32 and 16 antennas at the BS and percentage of antennas selected, $L = \{\lfloor N_T/5 \rfloor, \lfloor N_T/2 \rfloor, \lfloor 4N_T/5 \rfloor\}$. All algorithms have some gain compared to L antennas with no selection, but this gain diminishes as L increases as well as the computation time except for downwards BAB, which becomes faster when the ratio L/N_T approaches 1, i.e., the percentage of antennas selected is larger. However, downwards BAB does not significantly improve the BER compared to the other approaches as expected at first since it is optimal to find the largest MSV at a still high computational cost. In fact, Greedy performs very closely with

much less complexity. A brief capacity comparison is also carried out. The algorithms have as an objective function the largest MSV and not optimizing capacity but most of them outperform the ergodic capacity of a random AS.

Two scenarios with one UE with 1 and 2 receive antennas and a third scenario with two single-antenna UEs are described for Greedy FAS and upwards BAB FAS. The BS is equipped with 128 and 64 antennas. When the UE has 2 receive antennas it yields more diversity gain, however, the AS results are slightly worse in comparison probably because of not considering the second antenna for the AS process. The UEs in the third scenario are not located at the boresight direction BS. The AS in this case has been performed by taking both UEs. Therefore, the final antenna subset is not optimal for each particular UE but some gain can still be achieved, especially for the UE closer to the BS. In general terms, sub-optimal approaches such as Greedy Search and a simpler RF switching network could help to simplify the AS problem reducing the complexity and computation time while obtaining similar results to optimal results which require more costly solutions.

5.2 Future Work

The main purpose of this master thesis is to answer the research questions proposed at the beginning of the master thesis. For this purpose, the AS problem in transmission from the BS has been formulated, a criterion to obtain a high-quality subset of antennas has been defined, different approaches and algorithms have been described to solve this problem, and finally, the implications and performance of those algorithms in various massive MIMO scenarios have been observed. This has been summarized in the conclusions before. However, there are some areas that would be interesting to study and analyze further.

- The largest MSV was chosen as an AS criterion but there are other criteria such as maximizing the Frobenius norm of the matrix. The criterion is related to the characteristic to be optimized, in this case, the BER, but other research and papers focus on maximizing downlink sum-rate capacity. Other criteria could be explored and their respective BER and capacity performance and complexity could be compared.
- Future work can be done in analyzing the impact of the CSI in the AS process. How it affects performance when having perfect CSI or different values of partial CSI knowledge, that is different number of SRS symbols to estimate the channel or strategies to allocate them in the time-frequency resource grid. In that sense, more advanced channel estimation techniques could be applied to obtain better channel estimates. In addition, the simulator could be improved by adding more layers in the uplink and enabling better exploitation of the channel with more accurate AS.
- A comparison with different channel models could be also interesting. Also, the use of a simpler flat narrowband channel model could simplify the simulations when comparing multiple AS approaches. The antenna configuration can be also further investigated with more complex configurations such as uniform planar arrays (UPA) and dual polarization instead of linear arrays (ULA) with a single polarization.
- At the moment the number of antennas selected, L, is an input variable. However, it could be optimized to maintain a certain characteristic, such as the trade-off between complexity and SNR gain, and a more sophisticated algorithm could be devised to set this value.

References

[1] Erik Dahlman, Stefan Parkvall, and Johan Skold. *5G NR: The Next Generation Wireless Access Technology*. Academic Press, 2020.

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

[3] David López-Pérez, Antonio De Domenico, Nicola Piovesan, Geng Xinli, Harvey Bao, Song Qitao, and Mérouane Debbah. A Survey on 5G Radio Access Network Energy Efficiency: Massive MIMO, Lean Carrier Design, Sleep Modes, and Machine Learning. *IEEE Communications Surveys & Tutorials*, 24(1):653–697, 2022.

[4] Juan Vidal Alegría. *Beyond Massive MIMO: Trade-offs and Opportunities with Large Multi-Antenna Systems*. PhD thesis, Electrical and Information Technology, Lund University., 2023.

[5] Lizhong Zheng and David N. C. Tse. Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels. *IEEE Transactions on Information Theory*, 49(5):1073–1096, 2003.

[6] Thomas L Marzetta. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Transactions on Wireless Communications*, 9(11):3590–3600, 2010.

[7] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005. doi: 10.1017/CBO9780511807213.

[8] Yvo de Jong Bultitude and Terhi Rautiainen. IST-4-027756 WINNER II D1. 1.2 V1. 2 WINNER II Channel Models. *EBITG, TUI, UOULU, CU/CRC, NOKIA, Tech. Rep*, 2007.

[9] Stephan Jaeckel, Leszek Raschkowski, Kai Börner, and Lars Thiele. QuaDRiGa: A 3-D Multi-Cell Channel Model With Time Evolution for Enabling Virtual Field Trials. *IEEE Transactions on Antennas and Propagation*, 62(6):3242–3256, 2014.

[10] 3GPP TR 38.901 V16.1.0. Study on Channel Model for Frequencies from 0.5 to 100 GHz. 2020. URL `https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3173`.

[11] Daniel Gaetano Riviello, Francesco Di Stasio, and Riccardo Tuninato. Performance Analysis of Multi-User MIMO Schemes under Realistic 3GPP 3-D Channel Model for 5G mmWave Cellular Networks. *Electronics*, 11(3):330, 2022.

[12] Xuesong Cai, Erik L Bengtsson, Ove Edfors, and Fredrik Tufvesson. A Switched Array Sounder for Dynamic Millimeter-Wave Channel Characterization: Design, Implementation and Measurements. *IEEE Transactions on Antennas and Propagation*, 2023.

[13] Emil Björnson and Özlem Tuğfe Demir. *Introduction to Multiple Antenna Communications and Reconfigurable Surfaces*. Boston-Delft: now publishers, 2024. doi: 10.1561/9781638283157.

[14] An Le Ha, Trinh Van Chien, Tien Hoa Nguyen, Wan Choi, and Van Duc Nguyen. Deep Learning-Aided 5G Channel Estimation. In *2021 15th International Conference on Ubiquitous*

Information Management and Communication (IMCOM), pages 1–7, 2021. doi: 10.1109/IMCOM51814.2021.9377351.

[15] Trinh Van Chien, Christopher Mollén, and Emil Björnson. Large-Scale-Fading Decoding in Cellular Massive MIMO Systems With Spatially Correlated Channels. *IEEE Transactions on Communications*, 67(4):2746–2762, 2019. doi: 10.1109/TCOMM.2018.2889090.

[16] Guoda Tian, Xuesong Cai, Tian Zhou, Weinan Wang, and Fredrik Tufvesson. Deep-Learning Based Channel Estimation for OFDM Wireless Communications. In *2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication (SPAWC)*, pages 1–5, 2022. doi: 10.1109/SPAWC51304.2022.9834008.

[17] Özlem Tuğfe Demir and Emil Björnson. Channel Estimation in Massive MIMO Under Hardware Non-Linearities: Bayesian Methods Versus Deep Learning. *IEEE Open Journal of the Communications Society*, 1:109–124, 2020. doi: 10.1109/OJCOMS.2019.2959913.

[18] Andreas F. Molisch. *Wireless Communications*, volume 34. John Wiley & Sons, 2012.

[19] Xiang Gao. *Doctoral Thesis: Massive MIMO in Real Propagation Environments*. PhD thesis, Department of Electrical and Information Technology, Lund University, 2016.

[20] Juan Guirado López-Puigcerver. Design and Optimization of Bandwidth Part Selection for Massive Beamforming. Master's thesis, Electrical and Information Technology, Lund University., 2020.

[21] Johanna Bengtsson. Adaptive Frequency Resolution for Downlink Beamforming in 5G NR. Master's thesis, Electrical and Information Technology, Lund University., 2021.

[22] Adrian Garcia-Rodriguez, Christos Masouros, and Pawel Rulikowski. Reduced Switching Connectivity for Large Scale Antenna Selection. *IEEE Transactions on Communications*, 65 (5):2250–2263, 2017. doi: 10.1109/TCOMM.2017.2669030.

[23] Andreas F. Molisch and Moe Z. Win. MIMO Systems with Antenna Selection. *IEEE Microwave Magazine*, 5(1):46–56, 2004. doi: 10.1109/MMW.2004.1284943.

[24] Yuan Gao, Wei Jiang, and Thomas Kaiser. Bidirectional Branch and Bound Based Antenna Selection in Massive MIMO Systems. In *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 563–568, 2015. doi: 10.1109/PIMRC.2015.7343363.

[25] Yuan Gao and Thomas Kaiser. Antenna Selection in Massive MIMO Systems: Full-array Selection or Subarray Selection? In *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 1–5, 2016. doi: 10.1109/SAM.2016.7569725.

[26] Andreas F. Molisch, Moe Z. Win, Yang-Seok Choi, and J.H. Winters. Capacity of MIMO Systems With Antenna Selection. *IEEE Transactions on Wireless Communications*, 4(4): 1759–1772, 2005. doi: 10.1109/TWC.2005.850307.

[27] Xiang Gao, Ove Edfors, Fredrik Tufvesson, and Erik G. Larsson. Massive MIMO in Real Propagation Environments: Do All Antennas Contribute Equally? *IEEE Transactions on Communications*, 63(11):3917–3928, 2015. doi: 10.1109/TCOMM.2015.2462350.

[28] R.W. Heath, S. Sandhu, and A. Paulraj. Antenna Selection for Spatial Multiplexing Systems with Linear Receivers. *IEEE Communications Letters*, 5(4):142–144, 2001. doi: 10.1109/4234.917094.

[29] Chongjun Ouyang, Zeliang Ou, Lu Zhang, and Hongwen Yang. Optimal Transmit Antenna Selection Algorithm in Massive MIMOME Channels. In *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2019. doi: 10.1109/WCNC.2019.8886342.

[30] Yuan Gao, Han Vinck, and Thomas Kaiser. Massive MIMO Antenna Selection: Switching Architectures, Capacity Bounds, and Optimal Antenna Selection Algorithms. *IEEE Transactions on Signal Processing*, 66(5):1346–1360, 2018. doi: 10.1109/TSP.2017.2786220.

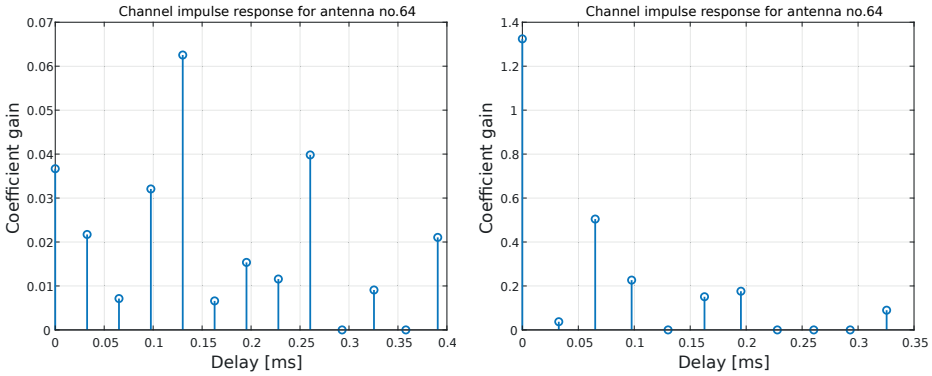
[31] Xiang Gao, Ove Edfors, Jianan Liu, and Fredrik Tufvesson. Antenna Selection in Measured Massive MIMO Channels Using Convex Optimization. In *2013 IEEE Globecom Workshops (GC Wkshps)*, pages 129–134, 2013. doi: 10.1109/GLOCOMW.2013.6824974.

[32] Xiang Gao, Ove Edfors, Fredrik Tufvesson, and Erik G. Larsson. Multi-Switch for Antenna Selection in Massive MIMO. In *2015 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2015. doi: 10.1109/GLOCOM.2015.7417765.

[33] Yi Cao and Vinay Kariwala. Bidirectional Branch and Bound for Controlled Variable Selection: Part I. Principles and Minimum Singular Value Criterion. *Computers Chemical Engineering*, 32(10):2306–2319, 2008. doi: https://doi.org/10.1016/j.compchemeng.2007.11.011.

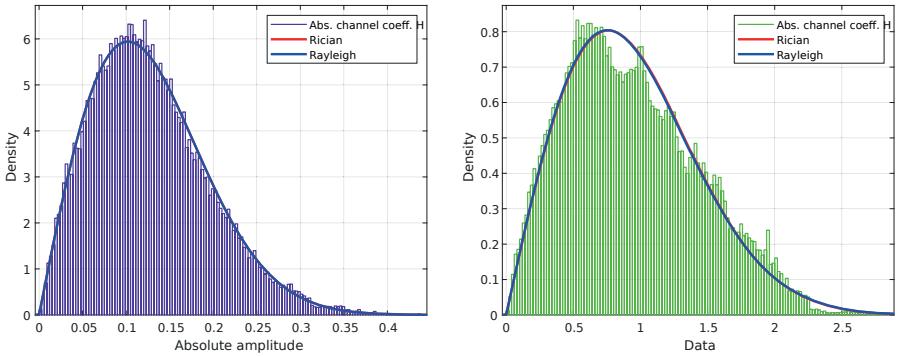
Appendix A

In this Appendix, complementary figures are shown.



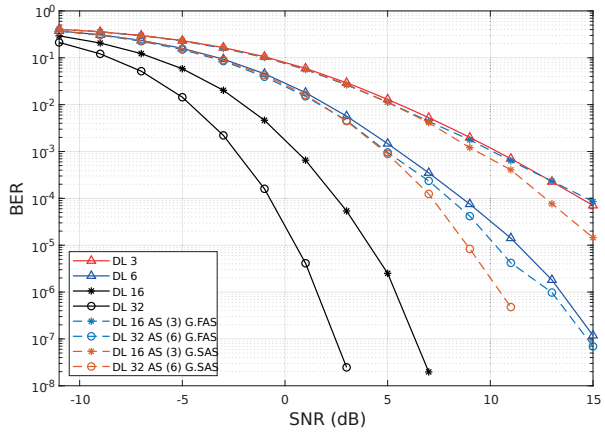
(a) Channel impulse response 1. **(b)** Channel impulse response 2.

Figure A.1: Example of channel coefficients gain for antenna no. 64 for channel 1 and 2.

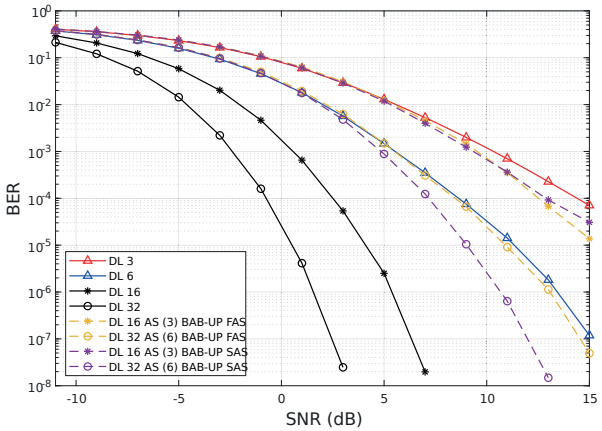


(a) Channel impulse response 1. **(b)** Channel impulse response 2.

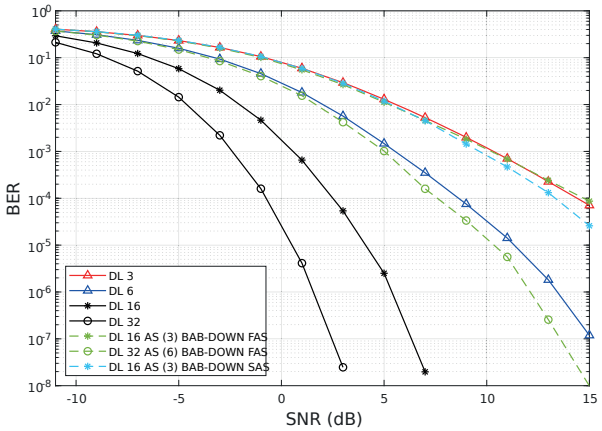
Figure A.2: Absolute amplitude distribution.



(a) Greedy FAS and Greedy SAS.

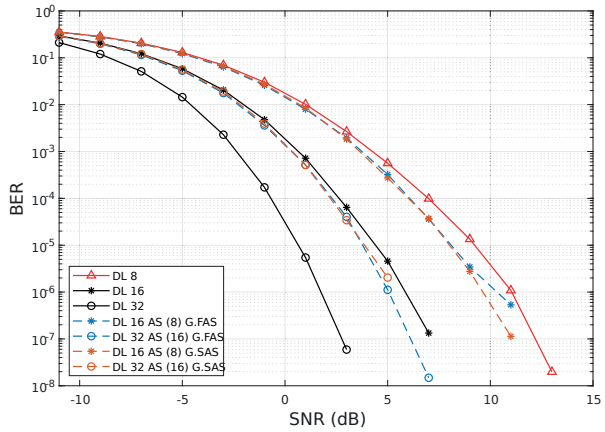


(b) BAB-UP FAS and BAB-UP SAS.

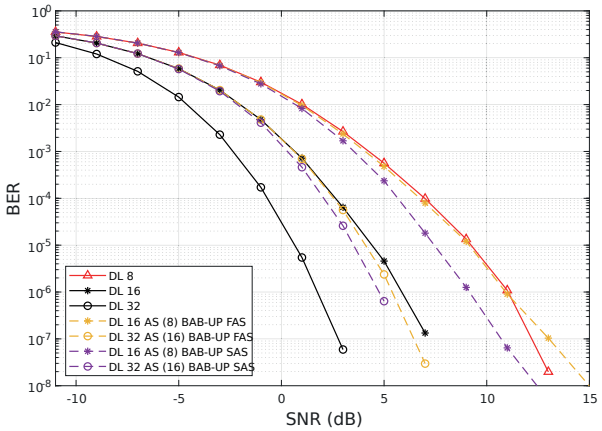


(c) BAB-DOWN FAS and BAB-DOWN SAS.

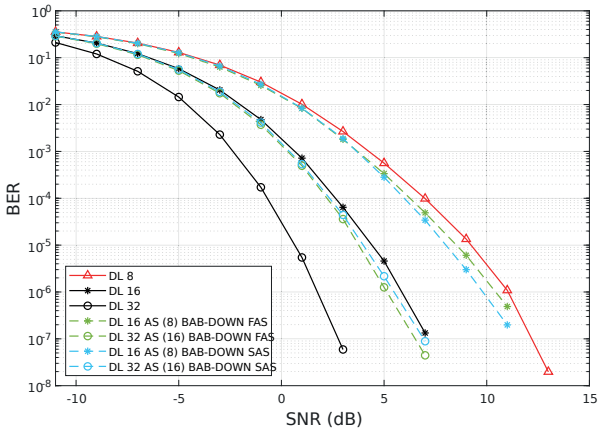
Figure A.3: Average BER for $N_T = \{16, 32\}$ antennas with AS $L = \lfloor N_T/5 \rfloor = \{3, 6\}$.



(a) Greedy FAS and Greedy SAS.

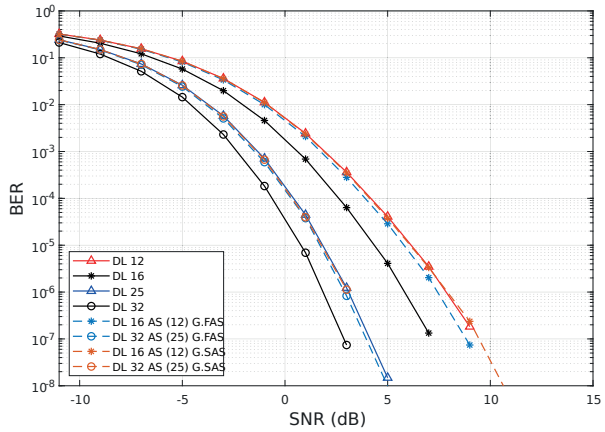


(b) BAB-UP FAS and BAB-UP SAS.

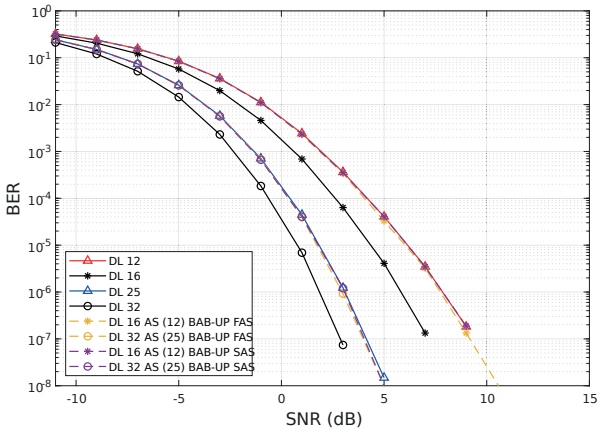


(c) BAB-DOWN FAS and BAB-DOWN SAS.

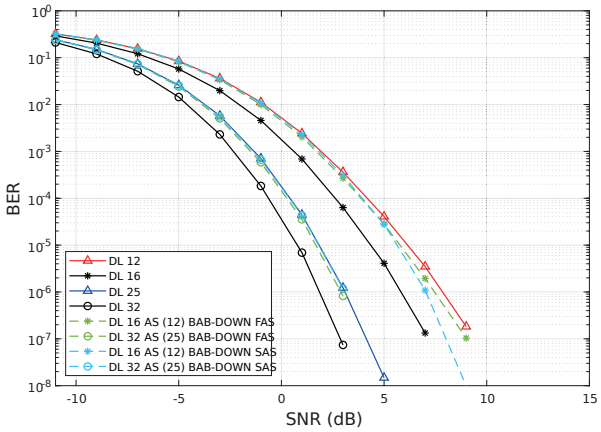
Figure A.4: Average BER for $N_T = \{16, 32\}$ antennas with AS $L = N_T/2 = \{8, 16\}$.



(a) Greedy FAS and Greedy SAS.



(b) BAB-UP FAS and BAB-UP SAS.



(c) BAB-DOWN FAS and BAB-DOWN SAS.

Figure A.5: Average BER for $N_T = \{16, 32\}$ antennas with AS $L = \lfloor 4N_T/5 \rfloor = \{12, 25\}$.

Printed by Tryckeriet i E-huset, Lund 2024

LUND
UNIVERSITY

Series of Master's theses
Department of Electrical and Information Technology
LU/LTH-EIT 2024-1013
http://www.eit.lth.se