Lens Antenna Arrays: An efficient framework for sparse-aware large-MIMO communications

Overview and potential applications in noncoherent schemes

A Master's Thesis submitted to:
Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona

In partial fulfillment of the requirements for the
Master in Telecommunications Engineering

Author: ***Marc Vilà Insa***
Advisor: ***Jaume Riba Sagarra***

TSC Department - Universitat Politècnica de Catalunya
Barcelona, June 2021

To my parents and friends.

Abstract

The recent increase in the demand for higher data transmission rates in wireless communications has entailed many implementation issues that can only be resolved by going through a full paradigm shift. Making use of the millimetric spectrum bands is a very attractive solution to the shortage of radio resources but, to garner all their potential, new techniques must be developed. Most of them are contained in the Massive Multiple Input Multiple Output (M-MIMO) framework: the idea of using very large antenna arrays for cellular communications. In this thesis, we propose the usage of Lens Antenna Arrays (LAA) to avoid the unbearable power and infrastructure costs posed by traditional M-MIMO architectures. This novel communication system exploits the angular-dependent power focusing capabilities of an electromagnetic lens to discern between waves with different angles of arrival and departure, without explicit signal processing. The work presented in this document motivates the use of LAAs in mmWave communications, studies some of their mathematical properties and proposes their application in noncoherent schemes. Numerical results validate the performance of this novel kind of systems and confirm their strengths in both multi-user and block fading settings. LAAs that use noncoherent methods appear to be very suitable for vehicular communications and densely populated cellular networks.

Resum

En els darrers temps, l'increment en la demanda de major velocitat de transmissió de dades en xarxes de comunicació inalàmbriques ha comportat diversos problemes d'implementació que tan sols es podran resoldre a través d'un canvi total de paradigma. Utilitzar les bandes mil·limètriques de l'espectre és una solució molt atractiva a l'escassetat de recursos de ràdio però, per tal d'extreure'n tot el seu potencial, és necessari desenvolupar noves tècniques. La majoria d'aquestes passa per la infraestructura *Massive Multiple Input Multiple Output (M-MIMO)*: la idea d'utilitzar matrius d'antenes molt grans per a comunicacions cel·lulars. En aquesta tesi, proposem l'ús de matrius d'antenes amb lent, o *Lens Antenna Arrays (LAA)*, per tal d'evitar els inassumibles costos energètics i d'instal·lació propis d'arquitectures M-MIMO tradicionals. Aquest innovador sistema de comunicacions explota les capacitats de concentració d'energia amb dependència angular de les lents electromagnètiques per tal de distingir entre ones amb diferents direccions d'arribada i de sortida, sense processament de senyal explícit. El treball presentat en aquest document motiva l'ús dels LAAs per comunicacions en bandes mil·limètriques (mmWave), n'estudia diverses propietats matemàtiques i proposa la seva aplicació en esquemes no coherents. Resultats numèrics en validen l'execució i confirmen les seves fortaleses en entorns multi-usuari i amb esvaïment en bloc. Els LAAs que utilitzen mètodes no coherents semblen ser idonis per a comunicacions vehiculars i per a xarxes cel·lulars altament poblades.

Resumen

En los últimos tiempos, el incremento en la demanda de mayor velocidad de transmisión de datos en redes de comunicación inalámbricas ha conllevado varios problemas de implementación que solo se podrán resolver a través de un cambio total de paradigma. Utilizar bandas milimétricas del espectro es una solución muy atractiva a la escasez de recursos de radio pero, para poder extraer todo su potencial, es necesario desarrollar nuevas técnicas. La mayor parte de éstas pasa por la infraestructura *Massive Multiple Input Multiple Output (M-MIMO)*: la idea de usar matrices de antenas muy grandes para comunicaciones celulares. En esta tesis, proponemos el uso de matrices de antenas con lente, o *Lens Antenna Arrays (LAA)*, para evitar los inasumibles costes energéticos y de instalación propios de las arquitecturas M-MIMO tradicionales. Este novedoso sistema de comunicaciones explota las capacidades de concentración de energía con dependencia angular de las lentes electromagnéticas para distinguir entre ondas con distintas direcciones de llegada y de salida, sin procesado de la señal explícito. El trabajo presentado en este documento motiva el uso de los LAAs en comunicaciones en bandas milimétricas (mmWave), estudia varias propiedades matemáticas y propone su aplicación en esquemas no coherentes. Resultados numéricos validan su ejecución y confirman sus fortalezas en entornos multiusuario y con desvanecimiento en bloque. Los LAAs que utilizan métodos no coherentes parecen ser idóneos para comunicaciones vehiculares y para redes celulares altamente pobladas.

Acknowledgements

Writing this thesis has been the effort of four intense months. It has required copious amounts of dedication and persistence. This would not have been possible without the support of a number of people. In this section, I would like to thank my close friends and my parents, who have given me moral support during this period of time.

Special thanks, as well, to my adviser Jaume Riba, who has presented me valuable guidelines that have contributed to enhance the quality of the project.

Contents

List of Figures

List of Tables

Notation and symbols

The next list describes several symbols that will be later used within the body of the document:

$x \triangleq y$ x is defined as y.

$x \equiv y$ x is equivalent to y.

$\{x\}$ The set of x.

$x \in \mathcal{S}$ x is an element of S.

$x[k]$ Signal that takes the value $x[k]$ at time k.

$X \cap Y$ Intersection of X and Y

$>>$ Much larger than.

$<<$ Much smaller than.

j Imaginary unit.

$\mathrm{Re}\{\cdot\}$ Real part.

$\mathrm{Im}\{\cdot\}$ Imaginary part.

$\mathrm{mod}_s 2\pi$ Symmetrical modulo operation in the interval $(-\pi, +\pi]$.

$\mathrm{sinc}(x)$ $\frac{\sin(\pi x)}{\pi x}$.

$\delta(x-a)$ Kronecker delta function centered at $x = a$.

$\lfloor \cdot \rfloor$ Floor function.

$\lceil \cdot \rfloor$ Rounding function.

$\arg\min_{x \in S} f(x)$ Value of x that minimizes $f(x)$ in the set S.

$\arg\max_{x \in S} f(x)$ Value of x that maximizes $f(x)$ in the set S.

$O\left(f(n)\right)$ Big-O of $f(n)$.

$\oplus$ Logical OR operator.

$\mathbf{a}$ Vector named a.

$\mathbf{A}$ Matrix named A.

$[\mathbf{a}_1, \dots \mathbf{a}_N]$ Matrix made with columns $a_1, \dots, a_N$.

$[\mathbf{A}]_{m,n}$ Element (m, n) of $\mathbf{A}$.

$\mathbf{I}_N$ Identity matrix of size $N \times N$.

$\mathbf{0}_{M \times N}$ Matrix of zeroes of size $M \times N$.

$\mathbf{1}_{N \times M}$ Matrix of ones of size $M \times N$.

$\cdot^*$ Complex conjugate.

$\cdot^T$ Transpose.

$\cdot^H$ Conjugate transpose.

$\mathbf{A}^{-1}$ Inverse of matrix A.

$\mathrm{Tr}\left[\cdot\right]$ Trace.

$|\mathbf{A}|$, $\det(\mathbf{A})$ Determinant of matrix A.

$|\cdot|$ Absolute value.

$\|\cdot\|$ Euclidean norm.

$\|\cdot\|_F$ Frobenius norm.

$\otimes$ Kronecker product.

$\odot$ Hadamard product.

$\mathrm{vec}(\cdot)$ Vectorized form.

$\mathrm{diag}(a_1, \ldots, a_N)$ Diagonal matrix with elements $a_1, \ldots, a_N$.

$\mathrm{diag}(\mathbf{A})$ Diagonal elements of matrix $\mathbf{A}$.

$\mathbb{E}[\cdot]$ Expected value.

$x \sim \mathcal{N}(\mu, \sigma^2)$ Real Gaussian variable x with mean μ and variance σ^2.

$x \sim \mathcal{CN}(\mu, \sigma^2)$ Complex Gaussian variable x with mean μ and variance σ^2.

$\mathbf{X} \sim \mathcal{MCN}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ Matrix complex normal distribution, such that $\mathrm{vec}(\mathbf{X}) \sim \mathcal{CN}(\mathrm{vec}(\mathbf{M}), \mathbf{V} \otimes \mathbf{V})$.

$x \sim \mathcal{U}(a, b)$ Uniformly distributed variable between a and b.

$x \sim \mathrm{Exp}(\alpha)$ Exponentially distributed variable with mean $\frac{1}{\alpha}$.

$\mathrm{erf}(x)$ Gaussian Error Function defined as $\mathrm{erf}\left(x\right) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-t^2\right) dx$.

$|\mathcal{S}|$ Size of set S.

Acronyms and abbreviations

AoA Angle Of Arrival.

AoD Angle Of Departure.

AS Antenna Selection.

AWGN Additive White Gaussian Noise.

BLAST Bell Laboratories Layered Space-Time.

BS Base Station.

CDD Conventional Differential Detection.

CSCG Circularly Symmetric Complex Gaussian.

CSI Channel State Information.

CSIR Channel State Information at the Receiver.

CSIT Channel State Information at the Transmitter.

D-OSTBC Differential Orthogonal Space-Time Block Coding.

DFDD Decision-Feedback Differential Detection.

DFE Decision-Feedback Equalization.

DFT Discrete Fourier Transform.

DL Downlink.

DM Differential Modulation.

DoF Degrees of Freedom.

DPSK Differential Phase Shift Keying.

DSTM Differential Space-Time Modulation.

EM Electromagnetic.

FDD Frequency Division Duplexing.

IPI Inter-Path Interference.

IR-UWB Impulse Radio Ultra Wideband.

ISI Inter-Symbol Interference.

IUI Inter-User Interference.

LAA Lens Antenna Array.

LoS Line-Of-Sight.

M-MIMO Massive Multiple Input Multiple Output.

M-PSK M-ary Phase Shift Keying.

MIMO Multiple Input Multiple Output.

MISO Multiple Input Single Output.

ML Maximum Likelihood.

MRC Maximum Ratio Combining.

MSDD Multiple-Symbol Differential Detection.

MSDSD Multiple-Symbol Differential Sphere Decoding.

MT Mobile Terminal.

MU-MIMO Multi-User Multiple Input Multiple Output.

nDFE Noncoherent Decision-Feedback Equalization.

NLoS Non-Line-Of-Sight.

OPDM Orthogonal Path Division Multiplexing.

OSTBC Orthogonal Space-Time Block Coding.

P2P Point-To-Point.

PDF Probability Density Function.

PDM Path Division Multiplexing.

PDMA Path Division Multiple Access.

PSP Power Space Profile.

RF Radio Frequency.

Rx Receiver.

SER Symbol Error Rate.

SIC Successive Interference Cancellation.

SIMO Single Input Multiple Output.

SINR Signal to Noise plus Interference Ratio.

SISO Single Input Single Output.

SNR Signal to Noise Ratio.

STBC Space Time Block Coding.

SVD Singular Value Decomposition.

TDD Time Division Duplexing.

Tx Transmitter.

UL Uplink.

ULA Uniform Linear Array.

V-BLAST Vertical Bell Laboratories Layered Space-Time.

Work Plan

The work required to complete this thesis has been spread throughout *4 months*. The planning we have followed can be summarized as:

- ***(1.5 months)*** **Literature research:** Gathering information and redirecting the main topics of the thesis was the first executed task.
- ***(1 month)*** **Characterization, design and independent work:** Once enough confidence in the field was obtained, we started working on new ideas and research proposals.
- ***(2 weeks)*** **Simulations and tests:** In order to assess the developed schemes and ideas, we devoted some time to prepare and conduct numerical tests on them.
- ***(1 month)*** **Writing of the thesis:** The last part of the development of this thesis was to sort all the work that had been done and present it in a self-contained, formal manner.

Chapter 1

Introduction

In recent times, wireless communications have been experiencing a very important paradigm shift that will become more and more apparent over the next decade. Many new applications are emerging, such as vehicular connectivity and pervasive health monitoring, and the usage of already well-known ones, like high definition video streaming and Internet of Things (IoT), is dramatically increasing. All these applications require the connection of a massive amount of new devices to exchange more data. The global mobile traffic is expected to grow by 30% annually up until 2024 [1]. This nonstop demand for data traffic is starting to become unbearable for the already scarce radio resources at sub-6GHz spectral bands. While a lot of research is being conducted to better exploit those frequencies by densifying the network, these efforts seem to lead to a dead-end road [2].

Fifth generation (5G) wireless communications promise higher data rates and connectivity, and reduced latency, financial cost and power consumption. Achieving these goals with the current available resources and technology is virtually unattainable. The most promising and direct solution to this dilemma is the exploitation of additional frequency bands. In particular, the millimeter-wave (mmWave) spectrum, which comprises between 30GHz - 300GHz[1], is really attractive, as it has traditionally been practically unused for wireless communications. The size of these bands will allow for an enormous increase in network capacity, a much needed enhancement in view of the 5G (and beyond) demands.

It is unquestionable that the deployment of mmWave cellular wireless networks will be key in enabling the compliance of 5G promises. The attractiveness of new bands raises the question on why have they not been explored before, and the answer is very clear: mmWave propagation poses a set of challenges that implementable technology has not been able to overcome until recently. The main drawback against the usage of millimetric wavelengths is the high free-space path-loss they suffer [3]. According to *Friis' Law*, omnidirectional waves experiment an attenuation proportional to the square of their wavelength, causing mmWaves to be especially dwindled by this phenomenon, compared to sub-6GHz ones. Atmospheric absorption is also of important concern at high frequencies. Low directivity antennas would require very high transmission power to be useful in regular outdoor cells, making them noncompliant with wireless communication regulations. They would only have a place in close range scenarios, such as indoor Wireless Personal Area Networks (WPAN) and picocells, but for general-purpose mobile communications, a more nuanced approach is required.

In mid-range outdoor mmWave applications, the most suitable systems are those that employ high-directional antennas. In a mobile environment, terminals must be able to steer narrow beams, resulting in high directional gains only in the wanted communication directions[2]. Antenna arrays with signal processing techniques, such as beamforming, can provide enough gain as to fully mitigate the highly limiting free-space loss of high frequencies, making mmWave bands viable for conventional cellular mobile communications. The amount of elements required in a mmWave array is usually rather large, hence Massive Multiple Input Multiple Output (M-MIMO) systems are of particular interest in 5G applications.

For previously exploited spectrum bands, mobile terminals (MT) and base stations (BS) were usually limited to few antennas per array, due to size constraints. This is not the case for mmWave.

[1] This roughly translates into wavelengths of 10mm to 1mm, hence its name.

[2] A derivative of communicating through narrower beams is higher robustness against eavesdropping and interference [1].

The small working wavelengths facilitate the use of very compact antennas, allowing the incorporation of reasonably sized arrays in mobile terminals. Under these circumstances, BSs show virtually no spatial limitation for massive antenna arrays. Nonetheless, there is another implementation aspect that must be pointed out which forbids direct application of traditional transceiver architectures. In conventional sub-6GHz fully-digital systems, each antenna is allocated a Radio Frequency (RF) chain in order to perform baseband processing. Translating this approach to M-MIMO arrays would be prohibitive in terms of infrastructure cost. Besides, having all the RF circuits active for full array processing would entail an unbearable power consumption, both from a financial and an ecological perspective. The environmental impact of 5G networks cannot be neglected, in particular because the Internet and Communication Technology (ICT) sector is already an immense contributor toward the global carbon footprint [4]. In this regard, energy efficiency is, more than ever, a very important design criterion for 5G systems.

While millimetric bands entail propagation issues, they can also provide desirable and potentially exploitable characteristics [5]. The small wavelength size relative to most elements in the communication environment produces a severe penetration loss. This implies that most power will either be: transmitted through a Line-Of-Sight (LoS) path, reflected onto a few single bounce Non-Line-Of-Sight (NLoS) components, or blocked. This sparse nature of the mmWave scattering opens the door to a reduced channel representation: the *beamspace channel* [6]. In a setting without rich scattering, the amount of dominant beams is very low compared to the number of antennas utilized. It then seems reasonable that only a set of RF chains of the order of the independent signal paths is required to garner the full channel potential for communication. This idea of working in the beamspace domain is one of the key concepts to enable the deployment of functional mmWave systems with acceptable infrastructure and energy costs. Alternatively, many other proposals have been developed over the last few years, such as *analog beamforming*, *hybrid analog/digital precoding* and *few-bit Analog to Digital Converters (ADC)*, but they all present their fair share of performance issues [1]. Certainly, working within the beamspace framework seems like the most promising approach.

How is it possible to transform a signal from the spatial domain to the beamspace one? The most straightforward way to achieve it is by means of pure digital signal processing over an antenna array. This approach, however, does not solve the power constraint problem at all, since the full set of array elements must be processed to transform a signal between domains. In order to overcome this limitation, we still need to introduce an additional component on the antenna array: an electromagnetic (EM) lens. This is a passive device analogous to its optical counterpart that has radiation focusing properties. For small angles, the effect of an EM lens can approximate a Discrete Fourier Transform (DFT), which is the processing we would have to explicitly perform in regular Multiple Input Multiple Output (MIMO) systems to transform the spatial domain into the beamspace domain. After this analog approximation of a Fourier transform, the dimension of the received signal is remarkably reduced. Thanks to the power focusing capability of the lens, we can apply Antenna Selection (AS) on a few sets of elements of the array and allocate very few RF chains to process them without eroding the system performance. This is in sharp contrast to antenna selection in regular MIMO, which usually compromises the communication [7].

The use of an EM lens in conjunction with an antenna array has been coined *Lens Antenna Array (LAA)*. It is a very powerful idea with a lot of potential implications for 5G. In this thesis, we have intended to study LAA-based systems in-depth in the context of mmWave wireless communications: from the properties of their mathematical model to particular practical applications. The unique characteristics of these new kinds of transceivers make them remarkably attractive for BS architectures, in which space and power constraints are loose enough to deploy extremely large arrays that can allocate plenty of antenna elements per user connected.

LAAs are a very broad topic of interest and substantial amounts of research are actively being conducted. A good deal of literature has been published proposing all sorts of uses for this exciting new technology: from regular communication schemes, to applications in user localization and radar. Notwithstanding, very few works have established the potential link between LAAs and *noncoherent communications* [8–10]. Systems of this type, which bypass the explicit acquisition of full Channel State Information (CSI) to communicate, have traditionally been overshadowed by coherent schemes. While this paradigm entails some issues that must be overcome for a successful implementation, the amount of desirable properties it delivers make it worthwhile. Not having to rely on the small-scale effects knowledge about the channel greatly simplifies the transmission of data in high mobility environments, such as vehicular communications, and within architectures that employ very large arrays, as in the case of M-MIMO. Furthermore, avoiding the use of orthogonal pilot sequences enables BSs to provide reliable service to a higher number of users simultaneously, as well as reducing their training overhead and thus increasing the achievable network transmission rate.

The noncoherent schemes proposed in the literature have mostly been conceived within a very agnostic approach on *statistical CSI*. One of the main motivations to study noncoherent communications under the LAA framework is the clear structure of the resulting channel [11]. The angle-dependent power focusing capabilities of the EM lens seem like a very natural complement for noncoherent communications. Angles of arrival (AoA) and angles of departure (AoD) in the beamspace domain are large scale effects (*i.e.* they vary much slower than the coherence time of the channel), which make them a relatively easy to acquire piece of CSI. We believe that carefully incorporating this knowledge in partially noncoherent schemes specifically tailored for LAAs can greatly improve their performance in front of fully noncoherent ones.

With this thesis, we want to broaden the fields of study of LAAs and noncoherent communications. By combining ideas taken from both, we want to display the full range of possibilities they can offer in unison.

1.1 State of the art

EM lenses are not a new topic of research, and a proof of that is the well established literature, such as [12]. Nonetheless, their joint application with antenna arrays of similar size is rather novel, especially for mmWave communications, although they are rapidly gaining attention due to the opportunities they offer [1, 3]. In [6] and [13], the authors study the *Continuous Aperture MIMO (CAP-MIMO)*, built upon the framework of beamspace MIMO. They depart from traditional hybrid analog-digital architectures by proposing the use of a *Discrete Lens Array (DLA)* to remarkably reduce the transceiver complexity and power consumption. The modeling they use for the EM lens, however, is that of a spatial DFT, which is a useful enough approximation, but far from a real system implementation.

The line of research conducted in [7] and [11] introduces a change in perspective in the analysis of LAA systems. Instead of systematically using the DFT model, as other works on the topic have done [14–16], the authors introduce an alternative study framework that treats the LAA as an integrated component. By establishing its physical properties, they derive an array response vector that is faithful to real implementations without sacrificing mathematical simplicity and usefulness. This new perspective has lead to a plethora of results by other authors, like [17–19].

A notable work that stemmed from this new line of research is [20]. In it, the authors go a step further by realizing that, at millimetric bands and with very large arrays, the plane wave models for EM signals are not, in general, accurate approximations. Instead, they follow an analysis similar to [7] but on the basis of a more general spherical wave model. The new array response vector derived allows for potential localization applications.

As for the noncoherent communications content, we have been inspired by the line of research from [21, 22]. The authors have developed a very convenient framework for differential detection in Multi-User Multiple Input Multiple Output (MU-MIMO) Uplink (UL). A variety of works have followed this approach and studied aspects and modifications of the basic scheme [23–25]. Of special interest, are [8] and [9], in which the introduction of LAAs in noncoherent UL systems is contemplated, although more physics-focused models of the EM lenses are employed. The noncoherent Downlink (DL) has also been researched, and a good example of relevant literature on the topic is [26]. In that article, the authors introduce a method to incorporate second order statistical CSI on a noncoherent Space Time Block Coding (STBC) scheme. Nevertheless, to the best of our knowledge, no works have been published that incorporate a LAA in a noncoherent DL setting.

1.2 Objectives and contributions

With this thesis we have determined to fulfill the following concrete objectives:

- Precisely motivate the use for LAAs in mmWave M-MIMO communications from the beamspace channel framework.
- Study an analytic mathematical model for the array response vector of the LAA and justify its use in practical applications.
- Examine the *power leakage problem* inherent to LAA systems.
- Analyze the symmetric LAA-MIMO channel.
- Characterize the capacity of the LAA-MIMO channel and understand its degradation factors.
- Compile a variety of coherent Point-To-Point (P2P) communication schemes, represent them in a consistent framework and particularize them for LAAs.
- Research two noncoherent P2P communications schemes presented in previous works, one for the UL and one for the DL, and generalize them to better leverage the strengths of LAAs.
- Study the multi-user extensions of the previous noncoherent schemes and propose new modifications.
- Design a full UL-DL multi-user communication system based on the previously discussed schemes. Provide solutions for AS, statistical CSI acquisition and other nontrivial aspects of the communication setting.
- Test the designed system on various simulated scenarios and compare it with well-known coherent alternatives.

1.3 Thesis outline

The content of this thesis has been organized in 6 chapters that progressively build upon each other. The common thread that serves as the backbone of the whole work can be split in three clear main subjects:

- **Preliminaries and motivation:** The first two chapters are focused on providing the mathematical foundations of the thesis and motivate the need for LAAs in mmWave communications. **Chapter 2** introduces the minimum mathematical tools required to understand the rest of the contents. We summarize some ideas on channel estimation, antenna selection and differential codification that will be revisited afterwards, so that individuals with different academic backgrounds

can easily follow all the other topics based on them. After this introductory chapter, **Chapter 3** presents the concept of *beamspace* and its applications in modeling wireless communications over millimetric bands. This framework evinces the need for a reduced channel model that can be obtained by means of array processing. LAAs come into play to avoid this demanding task by performing it analogically, effectively reducing the power consumption and infrastructure cost of the system.

- **General study of the LAA channel:** The next two chapters deal with various physical aspects of a LAA system. In **Chapter 4**, we comment on the physical implementation of an EM lens an derive a closed form solution for its array response vector. We start from a general spherical wave model and rigorously obtain an expression that encapsulates many real world scenarios. Afterwards, we particularize a simplified form that will serve as the basis for the signal processing and communications applications we are concerned with in this thesis. We also investigate the power leakage problem. With all these results, in **Chapter 5** we obtain the LAA channel matrix and study its main features. We then characterize its capacity and the factors that affect it.
- **Communication schemes over the LAA channel:** The last two content chapters of the thesis are devoted to proposing and studying data transmission schemes based on LAAs. **Chapter 6** is centered around the P2P communication problem between a BS and a MT. Firstly, some conventional coherent schemes are presented to illustrate the range of possibilities offered by LAAs. After that, two noncoherent schemes are explored, which will lay the foundations for **Chapter 7**. This final chapter recovers all the concepts and tools that have been developed throughout the thesis and builds a complete communication scheme upon them: from its general data transmission aspects to more nuanced details that play an important role in its implementation. Lastly, we subject our proposal to various simulated tests and compare its performance to other coherent schemes.

The following diagram (Figure 1.1) contains the previous outline in a synthesized form. We have indicated our explicit contributions in green. The ones which are marked with a red asterisk are relatively novel and have a clear potential for publication.

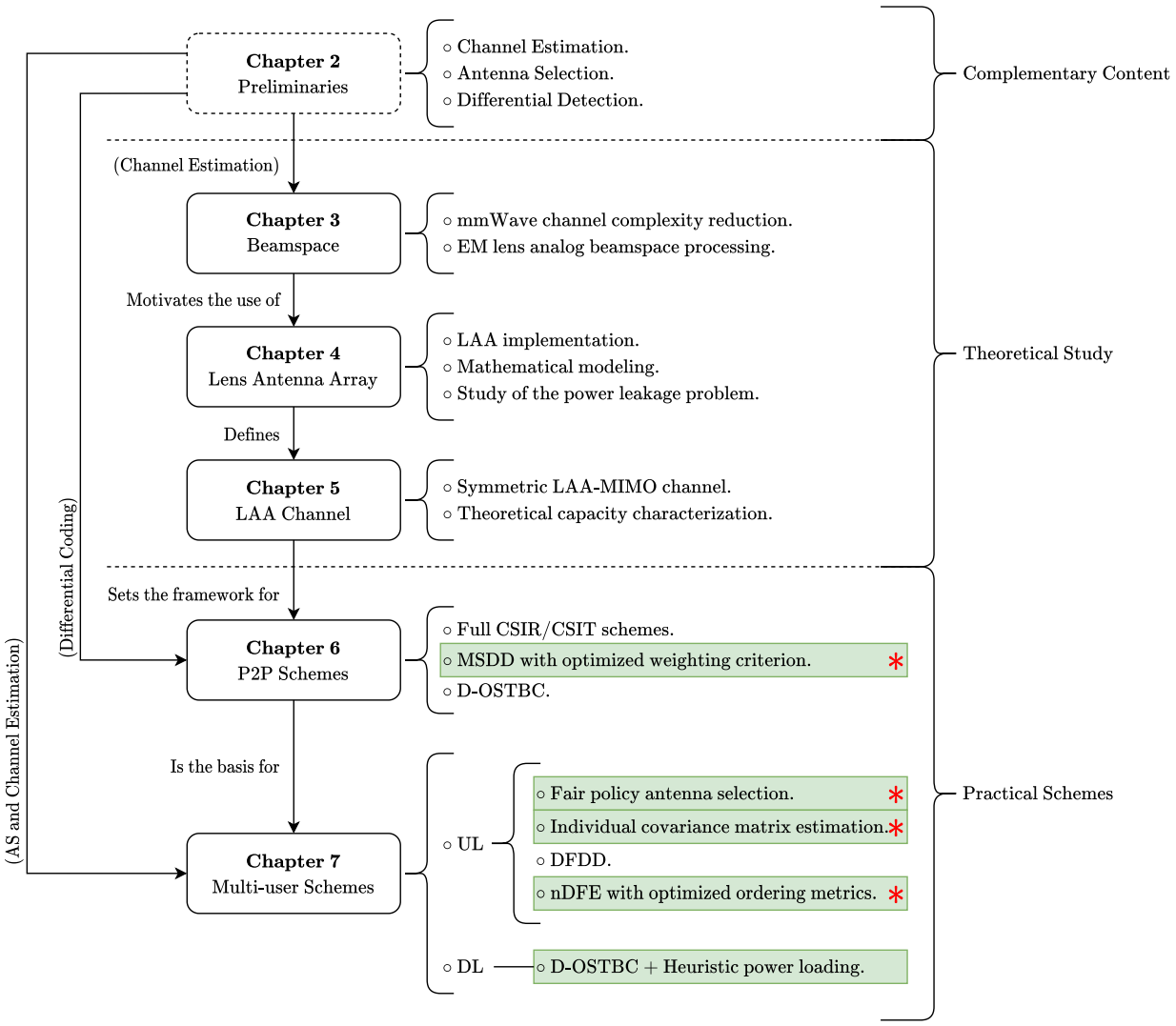


Figure 1.1: Outline of the thesis. The green boxes contain our explicit contributions.

Chapter 2

Signal processing and communications preliminaries

Before dealing with the main ideas of the thesis, we present in this chapter some preliminary results and methods that will be useful to develop various concepts afterwards. The purpose of having these sections in the body of the thesis is to make this document as self-contained as possible and allow readers with varied backgrounds to follow its line of reasoning. We have selected and summarized the topics we believe will be the most useful for such an intent. Additionally, we have presented them in a notation that will be consistent throughout the thesis.

Section 2.1 deals with the problem of channel estimation and the issues it entails in very large MIMO and multi-user environments. **Section 2.2** presents the technique of antenna selection, a crucial aspect of any system that employs antenna arrays with many elements. Both these topics will be retaken in **Chapter 7** to motivate the need for noncoherent schemes and illustrate the suitability of LAA in their implementation. Finally, **Section 2.3** introduces the concepts of differential coding, upon which we will generalize various ideas in **Chapters 6** and **7**.

2.1 Channel estimation

Channel estimation is an essential aspect of coherent wireless communications systems, since it is the main source of CSI at the transmitter and/or the receiver. It is a broad and classical topic in signal processing and a plethora of techniques have been developed over the years [1]. They have been specifically tailored for various scenarios and can incorporate many desirable characteristics (adaptive estimation, low complexity, etc.). In this introductory section, we will only provide some general notions on the topic.

To illustrate channel estimation we will employ it in a MIMO P2P context. Both the Transmitter (Tx) and the Receiver (Rx) are equipped with a multiantenna array, containing N_T and N_R elements, respectively. The Tx sends a stream of complex data symbols $\mathbf{x}[k] \in \mathbb{C}^{N_T \times 1}$ for a duration of K time instants. The model of the signal at the Rx is then:

$$\mathbf{y}[k] = \mathbf{H}\mathbf{x}[k] + \mathbf{z}[k] \in \mathbb{C}^{N_R \times 1} \quad , \quad k = 0, \dots, K-1. \tag{2.1}$$

$\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ is the *channel coefficients matrix*. Each one of its entries $[\mathbf{H}]_{m,n}$ characterizes the propagation between an element m of the Rx and an element n of the Tx. For simplicity, we assume slow *Rayleigh fading*, *i.e.* $[\mathbf{H}]_{m,n} \sim \mathcal{CN}(0,1)$ and they remain constant for the whole duration K. $\mathbf{z}[k] \in \mathbb{C}^{N_R \times 1}$ contains any noise and interference that may add uncertainty to the estimation of $\mathbf{H}$.

Broadly speaking, traditional channel estimation techniques can be divided into 2 general categories:

- **Training-based channel estimation:** Methods in this group are widely used in practical communication systems. They rely on the usage of *pilot sequence* signals known to the Rx that the Tx emits before starting the transmission of data. They serve as a probe to the Rx to estimate the channel coefficients. We denote the transmitted training block of size K_t,

$\mathbf{X}_t \triangleq [\mathbf{x}_t[0], \dots \mathbf{x}_t[K_t-1]]$. Then, the received training block takes the form

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{H}\mathbf{X}_t + \mathbf{Z} \in \mathbb{C}^{N_R \times K_t} \\ [\mathbf{y}_t[0], \dots, \mathbf{y}_t[K_t-1]] &= \mathbf{H}\left[\mathbf{x}_t[0], \dots, \mathbf{x}_t[K_t-1]\right] + [\mathbf{z}[0], \dots, \mathbf{z}[K_t-1]]\,. \end{aligned} \tag{2.2}$$

Notice that the rows of $\mathbf{X}_t$ and $\mathbf{Y}_t$ represent different Tx and Rx antennas, respectively, and their columns correspond to different time instants. Hence, this notation is referred to as *space-time*.

The *Maximum Likelihood (ML) estimation* of the channel consists in maximizing the likelihood function of the received block with respect to the channel matrix and for a known training sequence, *i.e.*

$$\widehat{\mathbf{H}}_{ML} = \arg\max_{\mathbf{H}} f_{\mathbf{Y}_t}\left(\mathbf{Y}_t|\mathbf{X}_t\right). \tag{2.3}$$

Under the Additive White Gaussian Noise (AWGN) assumption ($\mathbf{z}[k] \sim \mathcal{CN}\left(\mathbf{0}_{N_R\times 1}, P_z\mathbf{I}_{N_R}\right)$), this maximization reduces to

$$\begin{aligned} \widehat{\mathbf{H}}_{ML} &= \arg\min_{\mathbf{H}} \|\mathbf{Y}_t - \mathbf{H}\mathbf{X}_t\|_F^2 \\ &= \mathbf{Y}_t\mathbf{X}_t^+, \end{aligned} \tag{2.4}$$

where $\mathbf{X}_t^+ \triangleq \mathbf{X}_t^H\left(\mathbf{X}_t\mathbf{X}_t^H\right)^{-1}$ is the *right Moore-Penrose Pseudoinverse* of $\mathbf{X}_t$, and has the property that $\mathbf{X}_t\mathbf{X}_t^+ \equiv \mathbf{I}_{N_R}$. Then, the estimated channel matrix has the form

$$\widehat{\mathbf{H}}_{ML} = \mathbf{Y}_t\mathbf{X}_t^H\left(\mathbf{X}_t\mathbf{X}_t^H\right)^{-1} = \mathbf{H}\mathbf{X}_t\mathbf{X}_t^H\left(\mathbf{X}_t\mathbf{X}_t^H\right)^{-1} + \mathbf{Z}\mathbf{X}_t^+ = \mathbf{H} + \mathbf{Z}\mathbf{X}_t^+ = \mathbf{H} + \tilde{\mathbf{Z}}. \tag{2.5}$$

The resulting estimation is the real channel matrix plus some residual uncertainty $\tilde{\mathbf{Z}}$.

To perform this type of channel estimation, the *Gramian matrix* $\mathbf{X}_t\mathbf{X}_t^H$ must be invertible, and thus, full rank. This requires the length of the training sequence $\mathbf{X}_t$ to be $K_t \geq N_T$ and each pilot vector $\mathbf{x}_t[k]$ to be orthogonal to the rest.

- **Blind channel estimation:** The usage of training-based channel estimation techniques is widespread, since it provides very accurate estimates. Nonetheless, it presents a fair amount of drawbacks [27]. For instance, as the number of Tx antennas increases, the length of the training sequences must grow proportionally to it. This creates a large overhead before data transmission is possible, reducing the achievable rate of the system. Furthermore, the number of orthogonal pilot sequences for a given training period and bandwidth is limited, which severely restricts the amount of Txs that can communicate with the same Rx at once. Using non-orthogonal pilots for different estimations will produce less accurate results, since the obtained coefficients will be correlated with the channel matrices of other Txs[1]. This effect is known as *pilot contamination* [2, 27, 28].

  Blind methods are a remarkable alternative to explicit training-based channel estimation, although their use has not been popular in practical schemes [29]. They require no pilot sequences, hence the adjective *blind*. Instead, they are based on second-order statistics of the channel taken directly from the received data signals. They operate in the following manner [30]. Consider a transmitted signal block $\mathbf{X} \in \mathbb{C}^{N_R\times K}$, such that $\mathbf{X}\mathbf{X}^H \equiv \mathbf{I}_{N_R}$. For a received block $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}$ under AWGN, its *covariance matrix* can be obtained as

$$\begin{aligned} \mathbf{C}_{\mathbf{Y}} &= \mathbb{E}\left[\mathbf{Y}\mathbf{Y}^H\right] = \mathbb{E}\left[\left(\mathbf{H}\mathbf{X} + \mathbf{Z}\right)\left(\mathbf{H}\mathbf{X} + \mathbf{Z}\right)^H\right] \\ &= \mathbb{E}\left[\mathbf{H}\mathbf{X}\mathbf{X}^H\mathbf{H}^H\right] + \mathbb{E}\left[\mathbf{Z}\mathbf{Z}^H\right] = \mathbb{E}\left[\mathbf{H}\mathbf{H}^H\right] + P_z\mathbf{I}_{N_R} = \mathbf{C}_{\mathbf{H}} + P_z\mathbf{I}_{N_R}, \end{aligned} \tag{2.6}$$

[1] These Txs may not necessarily be trying to communicate with the same Rx.

assuming that $\mathbb{E}\left[\mathbf{H}\mathbf{X}\mathbf{Z}^H\right] \equiv \mathbf{0}_{N_R \times N_R}$ (the noise is uncorrelated with the signals). We denote the channel covariance matrix as $\mathbf{C_H} \triangleq \mathbb{E}\left[\mathbf{H}\mathbf{H}^H\right]$. In the zero-mean channel and noise case, an estimate of $\mathbf{C_H}$ from a single transmitted block[2] can be

$$\widehat{\mathbf{C}}_{\mathbf{H}} \triangleq \frac{1}{K}\mathbf{Y}\mathbf{Y}^H - P_z\mathbf{I}_{N_R}. \tag{2.7}$$

We can represent the channel matrix as $\mathbf{H} \triangleq \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, its *Singular Value Decomposition (SVD)*. $\mathbf{U} \in \mathbb{C}^{N_R \times N_R}$ and $\mathbf{V} \in \mathbb{C}^{N_T \times N_T}$ are *unitary matrices*[3], and $\mathbf{\Lambda} \in \mathbb{R}^{N_R \times N_T}$ is a rectangular diagonal matrix. Its diagonal elements are known as the *singular values* of $\mathbf{H}$, which are non-negative and ordered as $\lambda_1 \geq \cdots \geq \lambda_{n_{min}} > 0$. n_{min} is the rank of $\mathbf{H}$, which, in general, is the minimum between N_T and N_R.

With the SVD of the channel in mind, $\widehat{\mathbf{C}}_{\mathbf{H}}$ can be *eigendecomposed* as

$$\widehat{\mathbf{C}}_{\mathbf{H}} \triangleq \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}^2\widehat{\mathbf{U}}^H, \tag{2.8}$$

such that $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{\Lambda}}$ are estimates of $\mathbf{U}$ and $\mathbf{\Lambda}$, respectively. Then, the matrix $\tilde{\mathbf{H}} \triangleq \widehat{\mathbf{U}}\sqrt{\widehat{\mathbf{\Lambda}}^2}$ will approximately span the same subspace as $\mathbf{H}$. This is the reason why this approach is also known as *subspace partitioning* [27]. The phase ambiguity introduced by the blind estimation is caused by the loss of the information in $\mathbf{V}$, which is a common trade-off in methods based in second order statistics.

While blind channel estimation avoids the pilot contamination problem and requires no training overhead, it comes with its fair share of issues. For instance, if the channel phase ambiguity is not resolved with further training, only special types of modulations and codifications can make use of it, such as *differential coding*, explored in **Section 2.3**. The rigid structure of the transmitted signal blocks, so that $\mathbf{X}\mathbf{X}^H \equiv \mathbf{I}_{N_T}$ (*semi-unitary matrix*), is also a limiting factor in the system designs.

Besides these two main groups, there are channel estimation methods that rely on a mix of the two philosophies, referred to as *semi-blind methods* [29, 31]. They bring to the table an adjustable trade-off between estimation accuracy and complexity/overhead. By appending few training blocks next to or onto the data blocks, they help improve the results obtained with simple blind methods at a reasonable added cost.

2.2 Antenna Selection (AS)

It is a well-known fact in modern day wireless communications that equipping transmitters and receivers of MIMO systems with more antennas results in an increase of Degrees of Freedom (DoF) in the propagation channel, thus enabling higher data rates and reliability [28]. Even more notable benefits of this tendency can be found in multi-user environments, in which having more antennas at the BS allows it to provide service to a larger number of MTs at once.

Despite all the benefits that come with very large MIMO, there is an important price to pay for using very large antenna arrays: increased hardware complexity (RF chains), increased signal processing complexity (acquiring CSI, beamforming, etc.) and higher energy consumption. Under space and cost constraints, directly deploying Massive MIMO systems becomes unfeasible.

[2] It can be improved by averaging over various blocks, provided that the channel remains constant.

[3] Square matrices such that $\mathbf{U}^H\mathbf{U} = \mathbf{U}\mathbf{U}^H = \mathbf{I}$. They act as *rotation matrices*.

A promising technique to alleviate these costs without sacrificing the advantages of MIMO is AS [32]. At every time instant, each end of the MIMO system only employs the antennas with the associated channel coefficients that are the most beneficial for communication. Due to the sparse nature of the mmWave channel [1], the number of active antennas at every time can be much lower than the total size of the array, translating in a reduced number of active RF chains. The addition of a switching network (or alternative hardware) and the small overhead to account for an antenna selection phase in the communication are paid off by the amount of energy and resources spared.

We provide an example of AS in a P2P MIMO UL scenario based on the energy received by each antenna of the BS [11]. Consider the BS has N_{BS} antennas and $N_{RF} < N_{BS}$ available RF chains. The MT has N_{MT} antennas. Before data transmission, the system completes an antenna selection phase. Assuming no Channel State Information at the Transmitter (CSIT), the MT isotropically transmits identical training symbols s_{sel} for a duration K_{sel} and with power P_{sel}:

$$\mathbf{x}_{sel}[k] \triangleq \sqrt{\frac{P_{sel}}{N_{MT}}} s_{sel} \mathbf{1}_{N_{MT}} \quad , \quad k = 0, \dots, K_{sel} - 1. \tag{2.9}$$

Then, the signal received at the BS is

$$\mathbf{y}_{sel}[k] = \mathbf{H}\mathbf{x}_{sel}[k] + \mathbf{z}[k] \quad , \quad k = 0, \dots, K_{sel} - 1, \tag{2.10}$$

where $\mathbf{H}$ is the discrete-time channel (assumed constant during this training phase) and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_{N_{BS}\times 1}, P_z \mathbf{I}_{N_{BS}})$ is AWGN. Since the BS has only N_{RF} RF chains available, not every signal $\{\mathbf{y}_{sel}[k]\}_{k=0}^{K_{sel}-1}$ will be received by its antennas. Instead, the RF chains will need to sequentially scan over all the array elements. As a result, each one of them will only be connected to an RF chain for an effective time

$$T_{sel} \triangleq \left\lfloor \frac{K_{sel} N_{RF}}{N_{BS}} \right\rfloor. \tag{2.11}$$

After the AS training pilots have been transmitted, the BS combines the corresponding T_{sel} received symbols into a single vector

$$\begin{aligned} \bar{\mathbf{y}}_{sel} &= \frac{1}{\sqrt{T_{sel}}} \sum_{k'=0}^{T_{sel}-1} \mathbf{y}_{sel}[k'] = \frac{1}{\sqrt{T_{sel}}} \sum_{k'=0}^{T_{sel}-1} \mathbf{H}\mathbf{x}_{sel}[k'] + \frac{1}{\sqrt{T_{sel}}} \sum_{k'=0}^{T_{sel}-1} \mathbf{z}[k'] \\ &= \sqrt{\frac{T_{sel} P_{sel}}{N_{MT}}} s_{sel} \mathbf{H}\mathbf{1}_{N_{MT}} + \frac{1}{\sqrt{T_{sel}}} \sum_{k'=0}^{T_{sel}-1} \mathbf{z}[k'] \\ &= \sqrt{\frac{T_{sel} P_{sel}}{N_{MT}}} s_{sel} \sum_{n=1}^{N_{MT}} \mathbf{h}_n + \frac{1}{\sqrt{T_{sel}}} \sum_{k'=0}^{T_{sel}-1} \mathbf{z}[k'], \end{aligned} \tag{2.12}$$

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

The components of $\bar{\mathbf{y}}_{sel}$, $\bar{y}_{sel,m}$, are then sorted in descending order in terms of their associated energy as

$$\left|\bar{y}_{sel,[1]}\right|^2 \geq \cdots \geq \left|\bar{y}_{sel,[N_{RF}]}\right|^2 \geq \cdots \geq \left|\bar{y}_{sel,[N_{BS}]}\right|^2, \tag{2.13}$$

where the subindex $[\cdot]$ represents the permutation that fulfills such ordering. Let Δ_{sel} be an energy threshold below which the energy received at an antenna can be regarded as noise and N_Δ be the largest index for which $\left|\bar{y}_{sel,[N_\Delta]}\right|^2 \geq \Delta_{sel}$, *i.e.* the number of antennas with non-negligible

received power. The indexes of the selected antennas form the set $\mathcal{M}_{sel} = \{[1], \ldots, [N_{sel}]\}$, where $|\mathcal{M}_{sel}| = N_{sel} \triangleq \min\{N_\Delta, N_{RF}\}$.

The results of the AS are usually gathered in a *selection matrix* $\mathbf{\Sigma}$ such that

$$[\mathbf{\Sigma}]_{m,m} = \begin{cases} 1 & \text{if } m \in \mathcal{M}_{sel} \\ 0 & \text{otherwise} \end{cases}, \tag{2.14}$$

which is appended to received/transmitted signal vectors to mathematically model the fact that only a subset of antennas at the BS are active.

All in all, antenna selection is one of the most important enabling techniques for M-MIMO systems, which will greatly contribute to a more eco-friendly implementation [4]. As it will be made clear in the next chapters, while the application of AS on traditional M-MIMO transceivers always results in some sort of performance trade-off, this is not the case in LAAs. The rate degradation produced by employing just a fraction of the full array elements becomes almost negligible.

2.3 Conventional Differential Detection (CDD)

In most wireless communications systems, the Rx or the Tx (or both) utilizes some kind of CSI in order to compensate for the effect of the channel and obtain a reliable detection of the transmitted data. For slow fading channels, this task is simple to carry out using training sequences and similar techniques (as commented in **Section 2.1**). However, there are scenarios in which this is not possible; in M-MIMO communications, very large multi-user environments or smaller systems with a fast fading channel, the CSI becomes very hard to estimate accurately. Furthermore, the transmission of pilot symbols effectively reduces the achievable data rate.

A well-known solution to this issue is *Differential Modulation (DM)*, which can be applied in all sorts of single antenna and multiantenna systems. To illustrate its main concepts, we will present a Differential Phase Shift Keying (DPSK) scheme. Consider a typical Single Input Single Output (SISO) system

$$y = hx + z, \tag{2.15}$$

where x is the transmitted scalar symbol taken from a complex alphabet $\mathcal{S}$, $h \in \mathbb{C}$ is the channel and $z \sim \mathcal{CN}(0, P_z)$ is AWGN. Under these circumstances, if the channel is known at the Rx, the ML detection of x given y is reduced to

$$\hat{x}_{ML} \triangleq \arg\min_{x \in \mathcal{S}} |y - hx|^2, \tag{2.16}$$

which is a trivial scalar detection problem [31]. However, if h is unknown, *differential encoding and detection* can be used to effectively transmit data. The core idea of DM is to encode information in the difference in phase between two consecutive symbols, $s[k]$ and $s[k-1]$, and apply detection on them as a block. $s[k]$ are drawn from a unitary M-ary Phase Shift Keying (M-PSK) constellation $\mathcal{S} \triangleq \left\{\exp\left(j\frac{2\pi i}{M}\right) | i = 0, \ldots, M-1\right\}$, such that $|s[k]| = 1 \quad \forall s[k] \in \mathcal{S}$. Then, instead of transmitting $s[k]$ directly, we define a set of symbols $\{x[k]\}$ as:

$$\begin{aligned} &x[k] = x[k-1]s[k] \quad , \quad k > 0 \\ &x[0] \equiv 1. \end{aligned} \tag{2.17}$$

This codification ensures that $|x[k]| = 1 \quad \forall k$. The first sent symbol is set to 1 at both ends of the communication. We will refer to it as *reference symbol*.

At Rx, the received symbols are:

$$\begin{aligned} y[k] &= hx[k] + w[k] \\ &= hx[k-1]s[k] + w[k] \\ y[k-1] &= hx[k-1] + w[k-1], \end{aligned} \tag{2.18}$$

given that h remains constant for two channel uses. In this scenario, the ML detection of $s[k]$ given $y[k]$ and $y[k-1]$ reduces to [31]

$$\begin{aligned} \hat{s}_{ML} &= \underset{s[k]\in\mathcal{S}}{\arg\min} \left|y[k-1] - hx[k-1]\right|^2 + \left|y[k] - hx[k]\right|^2 \\ &= \underset{s[k]\in\mathcal{S}}{\arg\min} \left|y[k-1] - hx[k-1]\right|^2 + \left|y[k] - hx[k-1]s[k]\right|^2 \\ &= \underset{s[k]\in\mathcal{S}}{\arg\min} \left|y[k-1] - hx[k-1]\right|^2 + \left|y[k]s^*[k] - hx[k-1]s[k]s^*[k]\right|^2 \\ &= \underset{s[k]\in\mathcal{S}}{\arg\min} \left|y[k-1] - hx[k-1]\right|^2 + \left|y[k]s^*[k] - hx[k-1]\right|^2. \end{aligned} \tag{2.19}$$

Since neither h nor $x[k-1]$ are known, we can treat them as a single entity $q \triangleq hx[k-1]$. The value of q that minimizes (2.19) is

$$q_{opt} \triangleq \frac{y[k-1] + y[k]s^*[k]}{2}. \tag{2.20}$$

Then, the minimization simplifies to

$$\begin{aligned} \hat{s}_{ML} &= \underset{s[k]\in\mathcal{S}}{\arg\min} \left|y[k-1] - y[k]s^*[k]\right|^2 \qquad (2.21) \\ &= \underset{s[k]\in\mathcal{S}}{\arg\min} \left(-y^*[k-1]y[k]s^*[k] - y[k-1]y^*[k]s[k] + \left|y[k-1]\right|^2 + \left|y[k]\right|^2 \left|s[k]\right|^2\right) \\ &= \underset{s[k]\in\mathcal{S}}{\arg\max} \left(y^*[k-1]y[k]s^*[k] + y[k-1]y^*[k]s[k]\right) \\ &= \boxed{\underset{s[k]\in\mathcal{S}}{\arg\max} \operatorname{Re}\left\{y[k-1]y^*[k]s[k]\right\}}, \qquad (2.22) \end{aligned}$$

which is a trivial scalar detection problem.

With differential detection we have effectively solved the issue of requiring CSI for detection, as it utilizes the reference of $y[k-1]$ instead of h. It is an insightful observation on the core principle of Conventional Differential Detection (CDD) to view $y[k-1]$ as an effective channel estimate during the detection of $s[k]$ from $y[k]$. However, the price we must pay to accomplish it is a 3dB loss in performance. In the coherent case, the Signal to Noise Ratio (SNR) is $\text{SNR} = \frac{|h|^2}{P_z}$, for noise power P_z, while in the noncoherent one it is half of that value. This can be seen from (2.22), since this detector has to deal with twice the noise power than (2.16) [31].

DPSK has been extended to multiantenna systems in a variety of forms of DM [33]. The fundamental 3dB performance penalty has been extensively studied, and many schemes capable of arbitrarily reducing it (in slow fading channels) have been proposed [33]. Most of them are based on jointly detecting a sequence of symbols rather than individual ones.

Chapter 3

Overview on Beamspace MIMO

Before diving into LAA systems, we will motivate their application by studying a very closely related scenario presented in [6]: *Beamspace MIMO*.

Consider a MIMO wireless link with a 1D Uniform Linear Array (ULA) of length L_T at the transmitter and one of length L_R at the receiver. The two terminals are separated by a link distance of $R >> L_T, L_R$ and operate at carrier frequency f_c (wavelength λ_c). The antennas of each array are spaced equally by following the *critical spacing*[1] criterion ($d = \lambda_c/2$), thus the number of elements on the transmitter and the receiver are, respectively,

$$N_T \triangleq \left\lfloor \frac{L_T}{\lambda_c/2} \right\rfloor \quad , \quad N_R \triangleq \left\lfloor \frac{L_R}{\lambda_c/2} \right\rfloor . \tag{3.1}$$

In this setting we represent the link in the *aperture domain* with the expression introduced in **Section 2.1**:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \tag{3.2}$$

where $\mathbf{x} \triangleq [x_1, \dots, x_{N_T}]^T$ is the transmitted signal vector, $\mathbf{y} \triangleq [y_1, \dots, y_{N_R}]^T$ is the received signal vector, $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ is the aperture domain channel matrix, which contains the propagation channel between transmitter and receiver antennas, and $\mathbf{z} \in \mathbb{C}^{N_R \times 1}$ contains any noise and interference.

It is well-known [34] that a signal vector of dimension N_T can equivalently be expressed in terms of N_T orthogonal basis waveforms. In the previous MIMO scheme, orthogonal spatial beams form an optimal basis for the spatial dimension [6]. Let $\mathbf{U}_{b,T} \in \mathbb{C}^{N_T \times N_T}$ be the *transmit beamforming matrix*, that contains the transmit orthogonal beams and let $\mathbf{U}_{b,R} \in \mathbb{C}^{N_R \times N_R}$ be the *receive beamforming matrix*, with the receiver beams. We may define $\mathbf{x} \triangleq \mathbf{U}_{b,T}\mathbf{x}_b$ and $\mathbf{y}_b \triangleq \mathbf{U}_{b,R}^H \mathbf{y}$, where $\mathbf{x}_b = [x_{b,1}, \dots, x_{b,N_T}]^T$ and $\mathbf{y}_b = [r_{b,1}, \dots, r_{b,N_R}]^T$ are the transmitted and received signal vectors in the beamspace, respectively. Then we can equivalently represent (3.2) in the beamspace domain as

$$\mathbf{y}_b = \mathbf{U}_{b,R}^H \mathbf{y} = \mathbf{U}_{b,R}^H (\mathbf{H}\mathbf{x} + \mathbf{z}) = \underbrace{\mathbf{U}_{b,R}^H \mathbf{H} \mathbf{U}_{b,T}}_{\mathbf{H}_b} \mathbf{x}_b + \underbrace{\mathbf{U}_{b,R}^H \mathbf{z}}_{\mathbf{z}_b} = \mathbf{H}_b \mathbf{x}_b + \mathbf{z}_b. \tag{3.3}$$

$\mathbf{H}_b \in \mathbb{C}^{N_R \times N_T}$ is the *beamspace channel matrix* that contains the coupling between spatial beams at the receiver and transmitter. $\mathbf{x}$ can be rewritten in terms of the N_T orthogonal beams that form $\mathbf{U}_{b,T}$ as $\mathbf{x} = \mathbf{U}_{b,T}\mathbf{x}_b \triangleq \sum_{n=1}^{N_T} \mathbf{u}_T(\theta_{T,n})\, x_{b,n}$. This means that each transmitted beamspace signal, $x_{b,n}$, is mapped onto its corresponding orthogonal beam $\mathbf{u}_T(\theta_{T,n})$, which is a column of $\mathbf{U}_{b,T}$. The set of these N_T beams covers the full (one-sided) angular domain $\phi_T \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Each $\mathbf{u}_T(\theta_{T,n})$ is a steering vector that represents an all-phase complex spatial sinusoid with frequency

$$\theta_{T,n} \triangleq \frac{d}{\lambda_c} \sin(\phi_{T,n}) \overset{d=\frac{\lambda_c}{2}}{=} \frac{1}{2} \sin(\phi_{T,n}) \in \left[-\frac{1}{2}, \frac{1}{2}\right]. \tag{3.4}$$

[1] According to [34], the optimal spacing of elements in a linear array of length L is $\lambda_c/2$, which is usually referred to as critical spacing. It allows to capture the full angular resolution of $1/L$ with the least amount of antennas. Using a sparser spacing would produce aliasing (*grating lobes*) and using a denser one would not improve the resolution beyond $\frac{1}{L}$.

The elements of $\mathbf{u}_T(\theta_{T,n})$ are given by $u_{T,m}(\theta_{T,n}) \triangleq e^{-j2\pi\theta_{T,n}m}$ for

$$m \in \mathcal{M}(N_T) \triangleq \left\{ m = -\frac{N_T-1}{2}, \ldots, \frac{N_T-1}{2} \right\}. \tag{3.5}$$

Critical spacing implies there is a one-to-one relationship between $\theta_{T,n} \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ and $\phi_{T,n} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Since the spatial frequencies for the N_T beams, $\theta_{T,n}$, are uniformly spaced with spacing

$$\Delta\theta_{o,T} = \frac{1}{N_T} = \frac{\lambda_c}{2L} \Leftrightarrow \Delta\phi_{o,T} \approx \frac{\lambda_c}{L}, \tag{3.6}$$

then the basis beams $\mathbf{u}_T(\theta_{T,n})$ are orthogonal to each other [6]. $\Delta\theta_{o,T}$ and $\Delta\phi_{o,T}$ are measures of the spatial resolution (*beamwidth*) of the transmit array. The same reasoning up until this point can be done with $\mathbf{U}_{b,R}$ and $\mathbf{y}_b$ for the receiver array, by accordingly substituting the subindexes $T \leftrightarrow R$.

Notice that the construction of the orthogonal columns of $\mathbf{U}_{b,T}$ and $\mathbf{U}_{b,R}$ (we will refer to either of them as simply $\mathbf{U}_b \in \mathbb{C}^{N\times N}$) gives place to two unitary 2D DFT matrices:

$$\mathbf{U}_b \triangleq \mathbf{U}_{DFT} = \frac{1}{\sqrt{N}} \left[\mathbf{u}(\theta_n)\right]_{n\in\mathcal{M}(N)}. \tag{3.7}$$

The physical angles $\phi_n \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ are the spatial angles that cover the entire spatial horizon and correspond to $\theta_n = n\Delta\theta_o = \frac{n}{N}$. The DFT matrix relates the antenna space and the beamspace unambiguously.

In a P2P LoS link, the channel matrix in the aperture domain $\mathbf{H}$ is deterministic [35]. The aperture domain signal on every antenna element of the Rx can be related to the signal from every element of the Tx array with the steering vectors

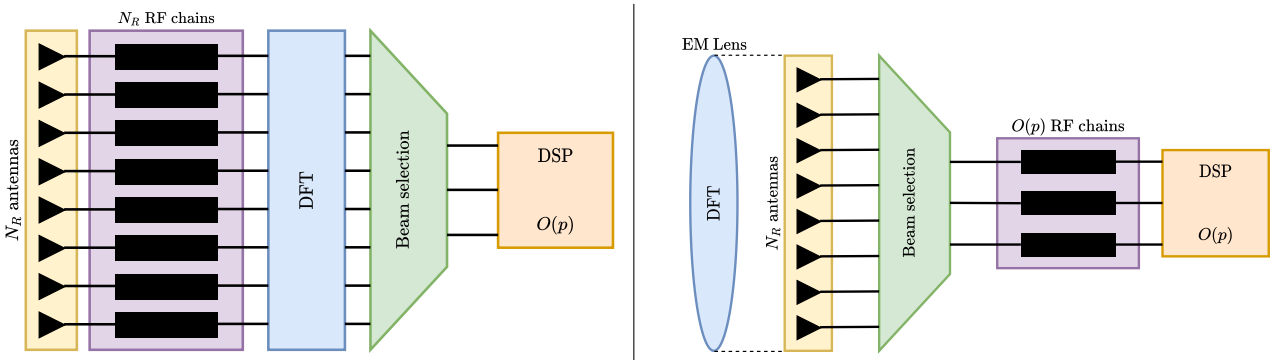
$$\mathbf{H}_{LoS} = \left[\mathbf{u}_R(\theta_{ch,n})\right]_{n\in\mathcal{M}(N_T)}, \tag{3.8}$$

where $\theta_{ch,n}$ is the spatial frequency corresponding to the angle of the nth transmit sample point with respect to the broadside. Each one of them is defined as $\theta_{ch,n} = n\Delta\theta_{ch}$. The spacing between channel frequencies, $\Delta\theta_{ch} \triangleq \frac{\lambda_c}{4R}$, is inversely proportional to the distance between Rx and Tx, and thus much smaller than the orthogonal spacing $\Delta\theta_o = \frac{1}{N} = \frac{\lambda_c}{2L}$, since $L << R$. This implies that the channel vectors will be correlated and the channel matrix low rank. The rank can be approximated by the number of beams that couple Tx and Rx elements strongly. Considering the spatial bandwidth of the receive aperture is $2\theta_{max} = 2 \times \frac{1}{2}\sin(\phi_{max}) \approx 2 \times \frac{L_R}{4R}$, the number of strong orthogonal beams is

$$p_{LoS} = \frac{2\theta_{max}}{\Delta\theta_o} = \frac{\sin(\phi_{max})}{\lambda_c/(2L_T)} \approx \frac{L_R/(2R)}{\lambda_c/(2L_T)} = \frac{L_R L_T}{\lambda_c R}. \tag{3.9}$$

This is a fundamental quantity known as the *Fresnel number* in optics [6], which is typically much smaller than N_R or N_T, for $R >> L_R, L_T$. p_{LoS} also represents the amount of orthogonal transmit beams that can be packed within the Rx aperture, each one occupying an approximately disjoint set of sample points. The Rx is then capable of distinguishing between different data streams modulated onto the orthogonal beams (*i.e.* spatial multiplexing for $p_{LoS} \geq 2$).

Wireless communications at mmWave frequencies are characterized by a very sparse nature (both in the spatial and frequency domains) and directional propagation, as expound in **Section 5**. The LoS propagation is expected to be the dominant mode, with few single-bounce multipath components. This sparsity is reflected in the low rank of the beamspace channel matrix $\mathbf{H}_b$. Beamspace representation of MIMO mmWave systems then seems like a very suitable choice. Moreover, $\mathbf{H}_b$ is nearly diagonal,



(a) Conventional beamspace MIMO Rx.

(b) Beamspace MIMO Rx with an EM lens.

Figure 3.1: Comparison between conventional and lens-aided beamspace MIMO systems.

which means that the orthogonal Fourier spatial basis vectors serve as a good approximation for the *eigenfunctions* of the LoS channel.

Conventional MIMO systems rely on processing over all the antenna elements of the ULA in order to capture the full spatial dimension of the channel. Then, they transform the aperture domain into the beamspace domain through a digital DFT [6]. The main issue hindering the practical application of this approach is that every antenna element requires an individual RF chain (see Figure 3.1a). The infrastructure cost and energy consumption become prohibitive as the arrays grow larger (M-MIMO). Solving this problem without further physical equipment is exceedingly challenging.

A promising alternative to circumvent this setback relies on replacing the digital DFT by a physical device that approximates its analog counterpart. It is a well known fact in optics [36] that converging lenses approximate the Fourier transform on waves with small incident angles (with respect to the normal vector of the lens). Instead of explicitly having to process all the elements of the array, an EM converging lens can be placed in front of the antennas to achieve a similar effect. It acts as a virtual phase shifter which focuses beams from different directions onto few different sets of antennas, producing an effective analog Fourier transform.

The main motivation behind this approach is the opportunity of reducing the complexity, infrastructure and energy consumption of the beamspace MIMO system. By placing a selection network between the array elements and the RF processing components (Figure 3.1b), the required amount of RF chains can be significantly reduced: only an amount of the order of the number of beams (*i.e.* the rank of the channel) is needed. This will allow the operation of much larger arrays, which can fully harness all the benefits of mmWave MIMO communications.

Chapter 4

The Lens Antenna Array System

A *Lens Antenna Array (LAA)* is a proposed multiantenna system that uses a convergent RF lens to focus impinging waves onto a focal region. In general, it is composed of two main elements:

- **An EM lens:** It is the mmWave analogous of an optical lens. It acts as a passive transmissive device able to alter the propagation direction and phase of EM waves. More precisely, we are interested in convergent lenses that can focus the energy of EM signals and collimate beams. In general, EM lenses can be categorized in three groups depending on the techniques with which they are implemented [17]:
  - *Dielectric lenses*: They are made of dielectric materials. Their properties come from their specifically designed surfaces.
  - *Planar lenses*: They consist of arrays of antennas connected by transmission lines with variable lengths. The design and distribution of the transmission lines is what defines their properties.
  - *Compact planar lenses*: They employ sub-wavelength spatial phase shifters.

  While we will consider the dielectric design, a fundamental principle is common among the three: to provide variable phase shifting of EM waves at different points on the lens aperture so as to achieve *angle-dependent energy focusing.* This is the core property that allows to separate signals coming from different directions and superpose them constructively onto different sections of the antenna array.
- **An array of antennas:** The elements of the matching antenna array are placed behind the EM lens. In the literature, various array configurations and geometries have been studied for the lens antenna, but the most prominent ones are the planar configuration and the semispherical one. While the former has been moderately studied [37], we will focus on the latter. The focal region behind a circular convex lens is found at a constant distance (*i.e. focal length*) from its center, hence it traces a semispherical surface. The most natural choice for array geometry is thus the one that matches it, since equal impinging waves coming from different directions will be equally focused on the antenna elements. Indeed, it is established in [19] and [20] that working with the semispehrical configuration results in clearer and more intuitive expressions.

As stated in **Chapter 3**, a converging EM lens approximates an effective DFT for small incident angles. Many works in the literature of LAAs [3, 6, 14, 15, 38] have adopted the small angle approximation to systematically model the effect of the lens as a spatial DFT filter. Even though this approach is useful to simplify signal processing solutions, it is far from the actual performance of a real LAA. A more physics-based method can be found in [39, 40], where the authors derive an array response vector from a Fourier optics perspective. The issue with this approach, however, is that the obtained expressions are not easily tractable in communication applications.

In the following sections, we will derive the array response of a LAA by following a procedure similar to the one presented in [20]. This technique is a generalization of the influential line of research started with [7]. The appeal of this approach is its clear association with a physical model of a LAA while providing a closed-form, easy to manipulate response vector expression that clearly emphasizes the potential of these multiantenna systems in mmWave communications.

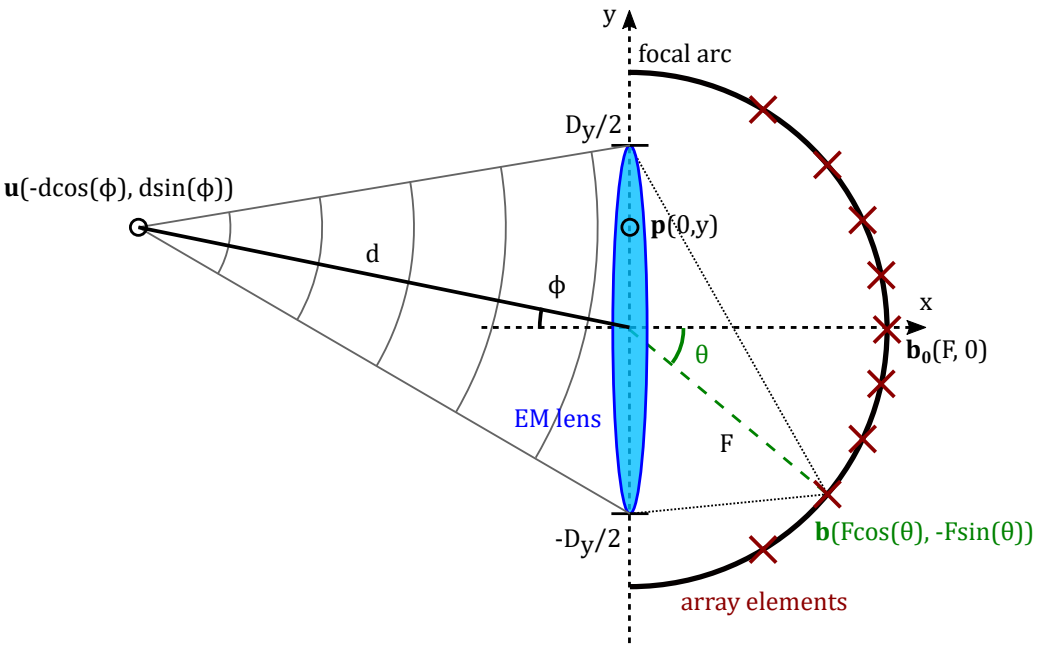


Figure 4.1: 2D section schematic of a LAA.

4.1 Array response derivation

Instead of dealing with a full-fledged 3D LAA, we will study its 2D section (*i.e. linear array*). This simplification allows for clearer notation and provides all the insights required to understand the working principles of a LAA system. The 3D model is investigated in [1, 17, 18, 41].

Consider a wireless system with a LAA acting as a receiver (see Figure 4.1). Its EM lens has a physical length of D_y and negligible width. It is centered at the origin and oriented along the y-axis. It is equipped with N_R antenna elements placed on the *focal arc* of the lens, which is shaped as a semicircle of radius F, the *focal distance* of the lens. There is a signal source located at a distance d from the center of the EM lens and at an angle[1] $\phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ relative to the x-axis, at point $\mathbf{u} \triangleq [-d\cos(\phi) \quad d\sin(\phi)]^T$. For generality, we assume the transmitter is a point source and uses an omni-directional antenna. The signal reaches the EM lens at any point $\mathbf{p} \triangleq [0 \quad y]^T$, for $y \in \left(-\frac{D_y}{2}, \frac{D_y}{2}\right)$, and takes the form

$$s(\mathbf{u}, \mathbf{p}) \triangleq \eta(\mathbf{u}, \mathbf{p}) e^{-j\kappa\|\mathbf{u}-\mathbf{p}\|}, \tag{4.1}$$

where $\kappa \triangleq \frac{2\pi}{\lambda}$ is the *wave number* corresponding to the wavelength λ, and $\eta(\mathbf{a}_1, \mathbf{a}_2) \triangleq \frac{\lambda}{4\pi\|\mathbf{a}_1-\mathbf{a}_2\|}$ is the *free-space path loss* between two points $\mathbf{a}_1$ and $\mathbf{a}_2$. The signal received by the array, r, is the contribution of every point on the lens that receives radiation from the source and focuses it on the focal arc:

$$r(\theta, d, \phi) \triangleq \int_{-D_y/2}^{D_y/2} s(\mathbf{u}, \mathbf{p}) \eta(\mathbf{p}, \mathbf{b}) e^{-j\varphi(\mathbf{p},\mathbf{b})} dy. \tag{4.2}$$

$\mathbf{b} \triangleq [F\cos(\theta) \quad -F\sin(\theta)]^T$ is any point on the focal arc and $\theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ is the angle between the center of the lens and that point (defined as positive below the x-axis and negative above it). Hence, r

[1] We can realistically assume that the signal source is in front of the LAA as BSs typically apply sectorization of the full angular range; each array usually serves $\leq 120°$.

is the integral of the received signal over all the points $\mathbf{p}$ on the lens and displaced onto a point $\mathbf{b}$ on the array with $\eta(\mathbf{p},\mathbf{b})$. $\varphi(\mathbf{p},\mathbf{b}) \triangleq \psi(\mathbf{p}) + \kappa\|\mathbf{p}-\mathbf{b}\|$ is the total phase shift of the signal. The first term is the contribution of the EM lens and the second term is the propagation delay between lens and focal arc. Expanding the previous expression we obtain

$$\begin{aligned} r(\theta,d,\phi) &= \int_{-D_y/2}^{D_y/2} \eta\left(\mathbf{u}(d,\phi),\mathbf{p}(y)\right) e^{-j\kappa\|\mathbf{u}(d,\phi)-\mathbf{p}(y)\|} \eta\left(\mathbf{p}(y),\mathbf{b}(\theta)\right) e^{-j\varphi(\mathbf{p}(y),\mathbf{b}(\theta))} dy \\ &= \int_{-D_y/2}^{D_y/2} \frac{\lambda^2 \exp\Big(-j\Big(\kappa\left\|\mathbf{u}(d,\phi)-\mathbf{p}(y)\right\| + \varphi\left(\mathbf{p}(y),\mathbf{b}(\theta)\right)\Big)\Big)}{16\pi^2\left\|\mathbf{u}(d,\phi)-\mathbf{p}(y)\right\|\left\|\mathbf{p}(y)-\mathbf{b}(\theta)\right\|} dy. \end{aligned} \tag{4.3}$$

Convex EM lenses use similar principles as their optical counterparts to achieve energy focusing of wave radiation. By carefully designing $\psi(\mathbf{p})$, we can obtain the desired focal points. In general, an incident spherical wave-front generated at point $\mathbf{c}_0 \triangleq [F_0 \quad 0]^T$ should converge at the focal point $\mathbf{b}_0 \triangleq [F \quad 0]^T$. The signal phase at that point is $\phi_0 = \kappa\|\mathbf{c}_0-\mathbf{p}\| + \psi(\mathbf{p}) + \kappa\|\mathbf{p}-\mathbf{b}_0\|$. The fixed phase shift by the lens design is then

$$\psi(\mathbf{p}) \equiv \phi_0 - \kappa\left(\|\mathbf{p}-\mathbf{b}_0\| + \|\mathbf{c}_0-\mathbf{p}\|\right), \tag{4.4}$$

so the total phase shift between points $\mathbf{p}$ and $\mathbf{b}$ is given by

$$\varphi(\mathbf{p},\mathbf{b}) = \phi_0 + \kappa\left(\|\mathbf{p}-\mathbf{b}\| - \|\mathbf{p}-\mathbf{b}_0\| - \|\mathbf{c}_0-\mathbf{p}\|\right). \tag{4.5}$$

With the previous result, we may now develop the phase shift in (4.3) and expand its L_2 norms:

$$r(\theta,d,\phi) = \int_{-D_y/2}^{D_y/2} \frac{\lambda^2 e^{-j\left(\phi_0+\kappa\left(\sqrt{d^2+y^2-2dy\sin(\phi)}+\sqrt{F^2+y^2+2yF\sin(\theta)}-\sqrt{F^2+y^2}-\sqrt{F_0^2+y^2}\right)\right)}}{16\pi^2\sqrt{\left(d^2+y^2-2dy\sin(\phi)\right)\left(F^2+y^2+2yF\sin(\theta)\right)}} dy. \tag{4.6}$$

In order to obtain a tractable closed form of the previous expression, we will assume that the angle of the source relative to the x-axis is small:

1. $d >> y$:
   - $\frac{1}{\sqrt{d^2+y^2-2dy\sin(\phi)}} \approx \frac{1}{d}$
   - 2nd Order Taylor Approximation:

$$\sqrt{d^2+y^2-2dy\sin(\phi)} \approx d - y\sin(\phi) + \frac{y^2\cos^2(\phi)}{2d} \tag{4.7}$$

2. $F >> y$:
   - $\frac{1}{\sqrt{F^2+y^2+2yF\sin(\theta)}} \approx \frac{1}{F}$
   - 2nd Order Taylor Approximation:

$$\begin{aligned} \sqrt{F^2+y^2+2yF\sin(\theta)} - \sqrt{F^2+y^2} &\approx \sqrt{F^2+y^2}\left(\frac{yF\sin(\theta)}{F^2+y^2} - \frac{1}{2}\left(\frac{yF\sin(\theta)}{F^2+y^2}\right)^2\right) \\ &\approx y\sin(\theta) - \frac{y^2\sin^2(\theta)}{2F} \end{aligned} \tag{4.8}$$

3. $F_0 >> y$:

- 2nd Order Taylor Approximation:

$$\sqrt{F_0^2 + y^2} \approx F_0 + \frac{y^2}{2F_0} \tag{4.9}$$

Applying the previous approximations on (4.6), we obtain

$$\begin{aligned} r(\theta, d, \phi) &\approx \int_{-D_y/2}^{D_y/2} \frac{\lambda^2 e^{-j\left(\phi_0 + \kappa\left(d + y(\sin(\theta) - \sin(\phi)) + y^2\left(\frac{\cos^2(\phi)}{2d} - \frac{\sin^2(\theta)}{2F} - \frac{1}{2F_0}\right) - F_0\right)\right)}}{16\pi^2 dF} dy \\ &= \frac{\lambda^2 e^{-j(\kappa d + \phi_0 - \kappa F_0)}}{16\pi^2 dF} \int_{-D_y/2}^{D_y/2} e^{-j\left(\kappa\left(y(\sin(\theta) - \sin(\phi)) + y^2\left(\frac{\cos^2(\phi)}{2d} - \frac{\sin^2(\theta)}{2F} - \frac{1}{2F_0}\right)\right)\right)} dy \end{aligned} \tag{4.10}$$

Without loss of generality, we assume $\phi_0 - \kappa F_0 \equiv 2\pi$, since ϕ_0 is common for all antenna elements and the phase term before the integral can be removed. For notation purposes, we denote $\alpha \triangleq \frac{\pi \sin^2(\theta)}{\lambda F} - \frac{\pi \cos^2(\phi)}{\lambda d} + \frac{\pi}{\lambda F_0}$ and $\beta \triangleq \frac{\sin(\theta) - \sin(\phi)}{\lambda}$. We can finally rewrite (4.10) and find a closed form solution:

$$\begin{aligned} r(\theta, d, \phi) &\approx \frac{\lambda^2 e^{-j\kappa d}}{16\pi^2 dF} \int_{-D_y/2}^{D_y/2} \exp\left(j\left(\alpha y^2 - 2\pi\beta y\right)\right) dy \\ &= \frac{\lambda^2 e^{-j\kappa d}}{16\pi^2 dF} \left[\frac{\sqrt{\pi}}{2\sqrt{\alpha}} e^{-j\left(\frac{\pi^2\beta^2}{\alpha} - \frac{5\pi}{4}\right)} \left(\operatorname{erf}\left(\frac{\alpha D_y + 2\pi\beta}{2\sqrt{\alpha}} e^{j\frac{3\pi}{4}}\right) + \operatorname{erf}\left(\frac{\alpha D_y - 2\pi\beta}{2\sqrt{\alpha}} e^{j\frac{3\pi}{4}}\right)\right)\right]. \end{aligned} \tag{4.11}$$

We normalize the previous expression and define the effective LAA response on any point $\mathbf{b}$ on the focal arc:

$$\begin{aligned} a(\theta, d, \phi) &\triangleq \frac{16\pi^2 dF}{\lambda^2 e^{-j\kappa d}} r(\theta, d, \phi) \\ &\approx \boxed{\frac{\sqrt{\pi}}{2\sqrt{\alpha}} e^{-j\left(\frac{\pi^2\beta^2}{\alpha} - \frac{5\pi}{4}\right)} \left(\operatorname{erf}\left(\frac{\alpha D_y + 2\pi\beta}{2\sqrt{\alpha}} e^{j\frac{3\pi}{4}}\right) + \operatorname{erf}\left(\frac{\alpha D_y - 2\pi\beta}{2\sqrt{\alpha}} e^{j\frac{3\pi}{4}}\right)\right)} \end{aligned} \tag{4.12}$$

Recall that θ is a continuous variable that represents the angle of a point behind the lens with respect to the x-axis, whose value must belong in the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. However, we are only interested in the positions of antenna elements. We want them to sample the y-axis uniformly, *i.e.* if their associated angles are denoted θ_n, then $\tilde{\theta}_n \triangleq \sin(\theta_n)$ must be equally spaced in the interval $[-1, 1]$:

$$\tilde{\theta}_n \triangleq \frac{n}{\tilde{D}}, \quad n \in \mathcal{M} \triangleq \left\{-\frac{N_R - 1}{2}, \ldots, 0, \ldots, \frac{N_R - 1}{2}\right\},^2 \tag{4.13}$$

where $\tilde{D} \triangleq \frac{D_y}{\lambda}$ denotes the normalized aperture of the lens. In a similar manner to the MIMO transceivers from **Section 3**, we will employ critical spacing along the y-axis. Therefore, the number of antennas the LAA should have is $N_R \triangleq 1 + \lfloor 2\tilde{D} \rfloor$, which means that larger lenses require more antennas to be deployed but can provide a higher resolution. Since the antennas are not deployed in a straight line but rather in an arc, notice they will be more densely packed at the center of the array than on its edges.

[2] We assume N_R to be odd for notational convenience.

The n-th element of the LAA is located at point $\mathbf{b}_n = [F\cos(\theta_n) \quad -F\sin(\theta_n)]$ and its array response is expressed as $a(\theta_n, d, \phi) \triangleq a_n(d, \phi)$. Hence, the full array response vector is

$$\mathbf{a}(d,\phi) \triangleq \begin{bmatrix} a_{-\frac{N_R-1}{2}}(d,\phi) \\ \vdots \\ a_0(d,\phi) \\ \vdots \\ a_{\frac{N_R-1}{2}}(d,\phi) \end{bmatrix} \in \mathbb{C}^{N_R \times 1}. \tag{4.14}$$

As stated in [20], (4.12) assumes spherical wave-fronts and is a well-suited approximation in near-field scenarios, *i.e.* $d \leq R \triangleq \frac{2D_y^2}{\lambda}$ (*Rayleigh distance*). It is thus a generalization of the previous results presented in [7]. It contains information regarding the distance between the LAA and the transmitter, in the form of a *windowing effect* [20], which can be exploited to estimate its location. Working at higher frequencies with larger lenses means the Rayleigh distance increases and far-field models become less accurate. The generality and complexity of (4.12) cannot be omitted in real world applications, even though few literature takes it into account [20].

Having said that, and for the purposes of this thesis, we will center its results around communications with LAAs and leave localization for future research. When using the plane wave-front approximation, while we lose information[3]about d, we will still keep the angular position of the transmitters. We can thus sacrifice the precision provided by (4.12) in favor of simplicity.

The far-field approximation is achieved when d and F_0 increase to infinity, making the spherical wave-fronts reduce to planar ones:

$$\begin{aligned} \lim_{d,F_0\to\infty} a_n(d,\phi) &= \lim_{d,F_0\to\infty} \frac{\sqrt{\pi}}{2\sqrt{\alpha}} e^{-j\left(\frac{\pi^2\beta^2}{\alpha}-\frac{5\pi}{4}\right)} \left(\operatorname{erf}\left(\frac{\alpha D_y + 2\pi\beta}{2\sqrt{\alpha}} e^{j\frac{3\pi}{4}}\right) + \operatorname{erf}\left(\frac{\alpha D_y - 2\pi\beta}{2\sqrt{\alpha}} e^{j\frac{3\pi}{4}}\right)\right) \\ &= \lim_{d,F_0\to\infty} \frac{-e^{j\frac{\pi}{4}} \left(\int_0^{\frac{\alpha D_y+2\pi\beta}{2\sqrt{\alpha}} e^{j\frac{3\pi}{4}}} e^{-t^2}dt + \int_0^{\frac{\alpha D_y-2\pi\beta}{2\sqrt{\alpha}} e^{j\frac{3\pi}{4}}} e^{-t^2}dt\right)}{\sqrt{\alpha} e^{j\frac{\pi^2\beta^2}{\alpha}}}. \end{aligned} \tag{4.15}$$

Recall that $\alpha = \frac{\pi\sin^2(\theta)}{\lambda F} - \frac{\pi\cos^2(\phi)}{\lambda d} + \frac{\pi}{\lambda F_0}$ and thus

$$\lim_{d,F_0\to\infty} \alpha = \frac{\pi\sin^2(\theta)}{\lambda F}. \tag{4.16}$$

Since we previously set $y << F$, we can assume $\alpha \to 0$ in the far-field. By letting $z \triangleq \sqrt{\alpha}$ we have

$$\lim_{d,F_0\to\infty} a_n(d,\phi) = \lim_{z\to 0} \frac{-e^{j\frac{\pi}{4}} \left(\int_0^{\left(\frac{zD_y}{2}+\frac{\pi\beta}{z}\right)e^{j\frac{3\pi}{4}}} e^{-t^2}dt + \int_0^{\left(\frac{zD_y}{2}-\frac{\pi\beta}{z}\right)e^{j\frac{3\pi}{4}}} e^{-t^2}dt\right)}{z e^{j\frac{\pi^2\beta^2}{z^2}}}. \tag{4.17}$$

[3]BSs located at a certain height that employ 2D and 3D LAAs can rely on the planar wave approximation without losing information about d, since it is encoded in the elevation angle.

Then, by *L'Hopital's rule*,

$$\begin{aligned}
\lim_{d,F_0\to\infty} a_n(d,\phi) &= \lim_{z\to 0} \frac{-e^{j\frac{\pi}{4}}\left(e^{-\left(\frac{zD_y}{2}+\frac{\pi\beta}{z}\right)^2 e^{j\frac{3\pi}{2}}} e^{j\frac{3\pi}{4}}\left(\frac{D_y}{2}-\frac{\pi\beta}{z^2}\right) + e^{-\left(\frac{zD_y}{2}-\frac{\pi\beta}{z}\right)^2 e^{j\frac{3\pi}{2}}} e^{j\frac{3\pi}{4}}\left(\frac{D_y}{2}+\frac{\pi\beta}{z^2}\right)\right)}{e^{j\frac{\pi^2\beta^2}{z^2}} - ze^{j\frac{\pi^2\beta^2}{z^2}} j2\frac{\pi^2\beta^2}{z^3}} \\
&= \lim_{z\to 0} \frac{-e^{j\pi}\left(e^{j\left(\frac{z^2D_y^2}{4}+\frac{\pi^2\beta^2}{z^2}+D_y\pi\beta\right)}\left(\frac{D_y}{2}-\frac{\pi\beta}{z^2}\right) + e^{j\left(\frac{z^2D_y^2}{2}+\frac{\pi^2\beta^2}{z^2}-D_y\pi\beta\right)}\left(\frac{D_y}{2}+\frac{\pi\beta}{z^2}\right)\right)}{e^{j\frac{\pi^2\beta^2}{z^2}}\left(1-j2\frac{\pi^2\beta^2}{z^2}\right)} \\
&= \lim_{z\to 0} \frac{-e^{j\left(\pi+\frac{z^2D_y^2}{4}\right)}\left(e^{jD_y\pi\beta}\left(\frac{D_yz^2}{2}-\pi\beta\right)+e^{-jD_y\pi\beta}\left(\frac{D_yz^2}{2}+\pi\beta\right)\right)}{z^2-j2\pi^2\beta^2} \\
&= \frac{e^{jD_y\pi\beta}-e^{-jD_y\pi\beta}}{j2\pi\beta} = \frac{\sin(D_y\pi\beta)}{\pi\beta} = \frac{D_y}{D_y}\frac{\sin(D_y\pi\beta)}{\pi\beta} = D_y\,\mathrm{sinc}(D_y\beta)\,.
\end{aligned} \tag{4.18}$$

Finally, we simply expand β:

$$a_n(\phi) = D_y\,\mathrm{sinc}\left(D_y\frac{\sin(\theta_n)-\sin(\phi)}{\lambda}\right) = D_y\,\mathrm{sinc}\left(\tilde{D}\tilde{\theta}_n-\tilde{D}\sin(\phi)\right). \tag{4.19}$$

By defining $\tilde{\phi}\triangleq\sin(\phi)\in[-1,1]$ (usually referred to as *spatial frequency*) and applying (4.13), we finally obtain the array response for each element:

$$a_n(\phi) = D_y\,\mathrm{sinc}\left(n-\tilde{D}\tilde{\phi}\right) \tag{4.20}$$

This expression is almost identical to the one presented in [7], which is already well-known throughout the literature [11, 17–19, 42]. The only difference is the factor $\tilde{D}$ instead of D_y. In the following sections we will refer to this normalized, dimensionless version, since it is a more natural way of expressing an array response vector:

$$\boxed{a_n(\phi) = \tilde{D}\,\mathrm{sinc}\left(n-\tilde{D}\tilde{\phi}\right)}. \tag{4.21}$$

The interest of employing a LAA rather than a traditional ULA system can be clearly seen in its response vector: the *sinc* function provides the AoA-dependent energy-focusing capability. In practical terms, any received signal with a given AoA ϕ_i will be magnified $\tilde{D}$ times on a very small set of antennas located close to the focal point, $\tilde{D}\tilde{\phi}_i$, while being almost negligible[4] for those ones far away from it ($|n-\tilde{D}\tilde{\phi}_i| >> 1$). The implications this has are very remarkable. Considering two different signals with AoAs ϕ_1 and ϕ_2 received simultaneously at the LAA , if their angular separation is sufficient ($|\phi_1-\phi_2| > \frac{1}{\tilde{D}}$, the *spatial frequency resolution* or half the width of the main lobe of the *sinc*), they will interfere one another quite lightly by simply selecting different sets of antennas.

[4]Specifically, the side lobes of a *sinc* function decrease with the distance to its center n_0 as $\frac{1}{|n-n_0|}$.

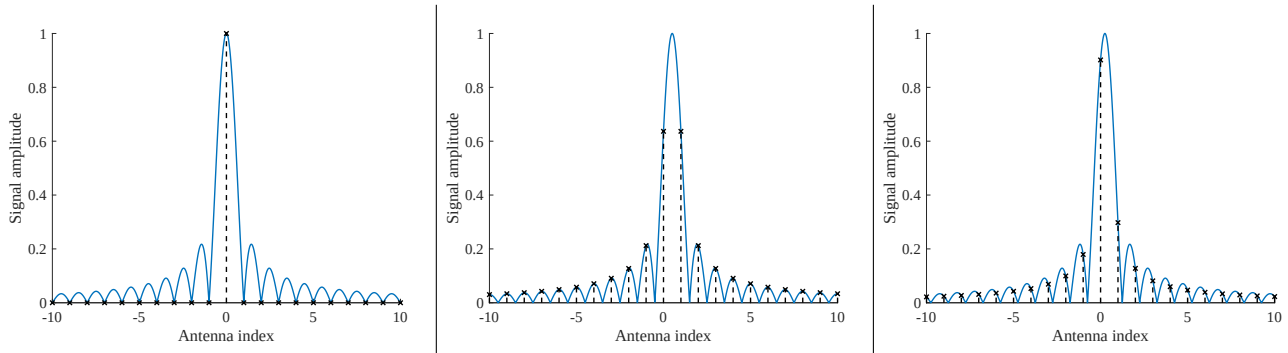
Throughout this thesis we will use the following notation

$$\mathbf{a}(\phi) \triangleq \begin{bmatrix} a_{-\frac{N_R-1}{2}}(\phi) \\ \vdots \\ a_0(\phi) \\ \vdots \\ a_{\frac{N_R-1}{2}}(\phi) \end{bmatrix} = \tilde{D} \begin{bmatrix} \operatorname{sinc}\left(-\frac{N_R-1}{2} - \tilde{D}\tilde{\phi}\right) \\ \vdots \\ \operatorname{sinc}\left(-\tilde{D}\tilde{\phi}\right) \\ \vdots \\ \operatorname{sinc}\left(\frac{N_R-1}{2} - \tilde{D}\tilde{\phi}\right) \end{bmatrix} \triangleq \boxed{\tilde{D}\operatorname{sinc}\left(\mathbf{n} - \tilde{D}\tilde{\phi}\mathbf{1}\right)}, \tag{4.22}$$

where $\mathbf{n} \triangleq \left[-\frac{N_R-1}{2} \ \ldots \ 0 \ \ldots \ \frac{N_R-1}{2}\right]^T$ contains the indices of the array elements.

Since the EM lens is a passive device, it presents reciprocity between incoming and outgoing waves, making the response vector (4.22) equivalent for both Tx and Rx systems [7].

4.2 Energy focusing and the power leakage issue



(a) Perfect alignment. **(b)** Worst case. **(c)** General leakage.

Figure 4.2: Various power leakage scenarios.

Given that $\tilde{\phi} \in [-1, 1]$, we may express it in terms of the array indices and the spatial frequency resolution:

$$\tilde{\phi} = \frac{\check{n} + \epsilon}{\tilde{D}} \quad , \quad \check{n} \triangleq \left\lceil \tilde{D}\tilde{\phi} \right\rfloor, \tag{4.23}$$

where $\check{n}$ is the antenna index nearest to the focusing point and $\epsilon \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ represents the misalignment between the antenna and that exact point. With (4.23), we can rewrite (4.22) as

$$\mathbf{a}(\phi) = \tilde{D}\operatorname{sinc}\left(\mathbf{n} - (\check{n} + \epsilon)\mathbf{1}\right). \tag{4.24}$$

Seeing that the number of antennas in the LAA is finite, it is usually not possible to perfectly sample impinging waves with random AoAs, *i.e.* to obtain $\epsilon = 0$, making their power spread across the entire array. Therefore, the *power leakage* problem is unavoidable. We may find 3 leakage situations (Figure 4.2):

(**a**) When an impinging wave is perfectly aligned with an array element all its associated power is captured by a single antenna (no power leakage). This situation is optimal in multipath scattering scenarios and multiple access schemes, since there is no interference between received signals.

(**b**) When a signal wave impinges right between two antenna elements (worst misalignment case) the power leakage is maximum. No antenna can be assigned as the center of the received wave and the interference between users or paths is the highest.

(**c**) In general, we may not find situation 4.2a nor 4.2b. Some amount of power will be lost unless all antennas are employed for a single received signal, but a single antenna will clearly be the nearest to the real center of the wave.

In [43], the authors propose a *beam aligning precoding* method to deal with the power leakage problem in the beamspace domain. In the LAA system we are considering in this thesis, we may exploit the power focusing capabilities of the EM lens to counter the power leakage in a simplified manner, by using the antenna selection ideas introduced in **Section 2.2**.

For a given signal reaching a LAA with a random AoA, indexed l, the power leaking through the array decays with the square of the distance from its focal point[5]. We can assume that the antennas located at a distance from the focal point greater than a threshold $\Delta > 0$ receive negligible signal power. We then define the support set of indices

$$\mathcal{M}_l \triangleq \left\{ n \in \mathcal{M} : |n - \tilde{D}_R \tilde{\phi}_l| < \Delta \right\} \tag{4.25}$$

assigned to its corresponding path. The antennas that are not contained in this set are considered to receive no signal power and can then be omitted from the processing. In particular, we may use an energy-based criterion (as in **Section 2.2**) to define the placement of such sets, so that the selected antennas will be the nearest one to the real AoA of the impinging wave (maximum received energy) plus a small support around it. For sufficiently spread out AoAs, the subsets of antennas assigned to different paths will be disjoint, reducing remarkably the task of (*interpath*) interference cancellation.

To illustrate and quantify the power leakage problem in a LAA multipath environment, we will define a metric derived from the one presented in [43]. The ratio between the leaked (due to misalignment and antenna selection) and total power received from a beam by the LAA is

$$\eta_{leak}^{(l)} \triangleq 1 - \frac{\sum\limits_{n \in \mathcal{M}_l} \mathbf{a}^H(\phi_n)\, \mathbf{a}(\phi_l)}{\sum\limits_{i \in \mathcal{M}} \mathbf{a}^H(\phi_i)\, \mathbf{a}(\phi_l)}, \tag{4.26}$$

where ϕ_n, ϕ_i are the AoAs corresponding to the antennas of the array and ϕ_l is the real AoA of the lth path. Alternatively, (4.26) can be expressed as

$$\eta_{leak}^{(l)} = 1 - \frac{\sum\limits_{n \in \mathcal{M}_l} \delta\left(n - \tilde{D}\tilde{\phi}_n\right) \operatorname{sinc}^2\left(n - \tilde{D}\tilde{\phi}_l\right)}{\sum\limits_{i \in \mathcal{M}} \delta\left(i - \tilde{D}\tilde{\phi}_i\right) \operatorname{sinc}^2\left(i - \tilde{D}\tilde{\phi}_l\right)} = 1 - \frac{\sum\limits_{n \in \mathcal{M}_l} \operatorname{sinc}^2\left(n - \tilde{D}\tilde{\phi}_l\right)}{\sum\limits_{i \in \mathcal{M}} \operatorname{sinc}^2\left(i - \tilde{D}\tilde{\phi}_l\right)}. \tag{4.27}$$

To give some numerical perspective, consider a LAA with $\tilde{D} = 100$ ($N_R = 201$) and an impinging planar wave such that $\tilde{D}\tilde{\phi} = 0.5$ (it impinges between antennas 0 and 1). As we have seen, this is the worst scenario in terms of power leakage since the misalignment with respect to the array elements is maximum. In that situation, selecting a support of just 4 antennas centered around the focal point results in $\eta \approx 9.7548\%$. In other words, we are able to capture more than 90% of the total power reaching the array. The conclusion we can draw from this example is that, although the power leakage issue must be thoroughly studied and considered (especially in multi-user environments), the power focusing capabilities of the LAA notably reduce its impact if the signals reaching the lens are sufficiently spread out along the angular dimension.

[5] This is caused by the decay of the sidelobes of $|\operatorname{sinc}(x)|^2$.

Chapter 5

The Lens Antenna Array Channel

Wireless mmWave communications present various particularities due to the working wavelengths, which are orders of magnitude smaller than most elements in the environment. For instance, *Friis' Law* [3] implies that simple non-directional antennas become useless at millimetric bands, since the isotropic path loss increases with the inverse of the square of the wavelength ($\propto \frac{1}{\lambda^2}$). This effect, however, can be mitigated with the introduction of directional gains, which also scale as $\propto \lambda^{-2}$, as more antenna elements can fit into the same physical area. This way of compensating free-space path loss makes the usage of very large (ideally massive) multi-antenna systems a must for mmWave communications.

Beyond free-space loss, general path loss is heavily influenced by the environment. There is a plethora [44] of statistical models to describe the channel characteristics in different situations. The one considered in this thesis is taken from [5]. The average path loss (large-scale effects) takes the form

$$\mathrm{PL}\left(d\right)\left[\mathrm{dB}\right] = A + 10B\log_{10}\left(d\right) + \xi, \quad \xi \sim \mathcal{N}\left(0, \sigma^2\right), \tag{5.1}$$

where d is the distance, A and B are linear parameters and ξ models the (lognormal) shadowing. For $B = 2$, this expression generalizes Friis' formula. It is worth mentioning that the free-space path loss of a mmWave link is no worse than at regular frequencies once it has been compensated with beamforming gain [3], making mmWave bands suitable for cellular systems.

Mathematically, the mmWave channel can be accurately described with well-known standard multipath models. In particular, we will adhere to the *Saleh-Valenzuela model* [1]. Consider a MIMO system in which the Tx has an array of N_T antennas and the Rx has N_R[1]. Their response vectors are $\mathbf{a}_T\left(\phi_T\right) \in \mathbb{C}^{N_T \times 1}$ and $\mathbf{a}_R\left(\phi_R\right) \in \mathbb{C}^{N_R \times 1}$, respectively, which depend on the angular direction of arriving and departing waves $\phi_R, \phi_T \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Under the general multipath environment, the channel impulse response can be modeled as

$$\mathbf{H}\left(t\right) \triangleq \sum_{l=1}^{L} \alpha_l \mathbf{a}_R\left(\phi_{R,l}\right) \mathbf{a}_T^H\left(\phi_{T,l}\right) \delta\left(t - \tau_l\right) \in \mathbb{C}^{N_R \times N_T}, \tag{5.2}$$

which is a matrix whose elements $\left[\mathbf{H}(t)\right]_{m,q} \triangleq h_{m,q}\left(t\right)$ denote the channel impulse response from transmitting antenna $q \in \mathcal{Q}$ to receiving antenna $m \in \mathcal{M}$, with $\mathcal{Q}$ and $\mathcal{M}$ being the sets of Tx and Rx antenna indexes, respectively. L is the number of significant channel paths or beams, which is much lower than N_R and N_T thanks to the multipath sparsity of mmWave communications [3]. Each path l is determined by 4 parameters: its AoD $\phi_{T,l}$, its AoA $\phi_{R,l}$, its associated time delay τ_l and the complex gain α_l, which may be modeled [7] as

$$\alpha_l \triangleq \sqrt{\beta\gamma_l} e^{j\omega_l} \quad , \quad \omega_l \sim \mathcal{U}\left[0, 2\pi\right). \tag{5.3}$$

β is the linear form of (5.1), *i.e.* the large-scale attenuation including distance-dependent path-loss and shadowing. κ_l represents the power fractional ratio of the lth path, such that $\sum_{l=1}^{L} \gamma_l \equiv 1$. Finally, ω_l denotes the phase shift of the lth path, which is a small-scale effect.

(5.2) is a general formulation[2] for *wideband* systems. When the channel bandwidth BW is sufficiently small, $\max_{l \neq l'} \left|\tau_l - \tau_{l'}\right| << \frac{1}{BW}$, *i.e.* $\tau_l \approx \tau_{l'} \triangleq \tau$. If the system can compensate the delays

[1] Both arrays have *critical spacing*.

[2] It may also incorporate Doppler shifts for rapidly varying channels [3].

perfectly (perfect time synchronization), we can omit them from the previous formulation and obtain its *narrowband* form

$$\boxed{\mathbf{H} \triangleq \sum_{l=1}^{L} \alpha_l \mathbf{a}_R\left(\phi_{R,l}\right) \mathbf{a}_T^H\left(\phi_{T,l}\right)}. \tag{5.4}$$

This is the version we will refer to throughout the thesis.

5.1 Symmetric channel: LAA-MIMO

Even though in later sections of the thesis we will focus on the *single-sided* LAA channels, characterizing the *symmetric* case provides the richest insights. Indeed, most LAA channels are special cases of the more general symmetric LAA-MIMO channel. In this model, both Tx and Rx are equipped with a LAA, with N_T and N_R elements available, and normalized apertures $\tilde{D}_T$ and $\tilde{D}_R$, respectively. Their antenna indices are gathered in vectors

$$\mathbf{q} \triangleq \begin{bmatrix} -\frac{N_T-1}{2} \\ \vdots \\ \frac{N_T-1}{2} \end{bmatrix} \quad , \quad \mathbf{m} \triangleq \begin{bmatrix} -\frac{N_R-1}{2} \\ \vdots \\ \frac{N_R-1}{2} \end{bmatrix}, \tag{5.5}$$

and thus according to (4.22), we may express the transmit and receive array response vectors as

$$\mathbf{a}_T\left(\phi_{T,l}\right) = \tilde{D}_T \operatorname{sinc}\left(\mathbf{q} - \tilde{D}_T \tilde{\phi}_{T,l} \mathbf{1}\right) \tag{5.6}$$

$$\mathbf{a}_R\left(\phi_{R,l}\right) = \tilde{D}_R \operatorname{sinc}\left(\mathbf{m} - \tilde{D}_R \tilde{\phi}_{R,l} \mathbf{1}\right). \tag{5.7}$$

The narrowband channel matrix (5.4) then takes the form

$$\begin{aligned} \mathbf{H} &= \sum_{l=1}^{L} \alpha_l \mathbf{a}_R\left(\phi_{R,l}\right) \mathbf{a}_T^H\left(\phi_{T,l}\right) \\ &= \underbrace{\tilde{D}_R \tilde{D}_T}_{\tilde{D}^2} \sum_{l=1}^{L} \alpha_l \operatorname{sinc}\left(\mathbf{m} - \tilde{D}_R \tilde{\phi}_{R,l} \mathbf{1}\right) \operatorname{sinc}\left(\mathbf{q}^H - \tilde{D}_T \tilde{\phi}_{T,l} \mathbf{1}^H\right), \end{aligned} \tag{5.8}$$

or, equivalently,

$$\mathbf{H} = \tilde{D}^2 \sum_{l=1}^{L} \alpha_l \operatorname{sinc}\left(\mathbf{m} - \left(m_l + \epsilon_{R,l} \mathbf{1}\right)\right) \operatorname{sinc}\left(\mathbf{q}^H - \left(q_l + \epsilon_{T,l} \mathbf{1}^H\right)\right), \tag{5.9}$$

in terms of the antennas nearest to the focus point of each lth path beam.

This channel model (and derived ones) allows for parallel data streams to be transmitted over different propagation paths with independent processing. This interesting property is presented in [7] and is named Path Division Multiplexing (PDM). It differs from conventional spatial multiplexing since it can achieve higher spatial resolution without the need for complex array processing, just by exploiting the path-sparsity and angular properties of the channel. Although general PDM cannot avoid inter-path interference without more advanced techniques (due to the power leakage problem mentioned in **Section 4.2**), the authors in [7] point out an ideal case that does not present this issue. When the AoAs are perfectly aligned with the Rx antennas, the *sinc* pattern reduces to a *Kronecker delta* and suppresses the inter-path interference completely (assuming that different antennas are selected for different paths). This scheme is termed as Orthogonal Path Division Multiplexing (OPDM).

5.1.1 Capacity characterization

To study the capacity of the previous channel model, consider its optimal case. Every signal beam l would align perfectly with its corresponding Tx and Rx antennas m_l and q_l, *i.e.* $\epsilon_{R,l}, \epsilon_{T,l} = 0 \quad \forall l = 1, \dots, L$. Additionally, all L signal paths would have distinguishable AoDs and AoAs, such that $m_l \neq m_{l'}$ and $q_l \neq q_{l'}$ for all $l \neq l'$. Thus, (5.9) reduces to

$$\begin{aligned} \mathbf{H}_{opt}(\phi_{R,l}, \phi_{T,l}) &= \tilde{D}^2 \sum_{l=1}^{L} \alpha_l \operatorname{sinc}\left(\mathbf{m} - m_l \mathbf{1}\right) \operatorname{sinc}\left(\mathbf{q}^H - q_l \mathbf{1}^H\right) \\ &= \tilde{D}^2 \sum_{l=1}^{L} \alpha_l \delta\left(\mathbf{m} - m_l \mathbf{1}\right) \delta\left(\mathbf{q}^H - q_l \mathbf{1}^H\right) \end{aligned} \tag{5.10}$$

$$[\mathbf{H}_{opt}]_{m,q} = \begin{cases} \tilde{D}^2 \alpha_l, & \text{if } m = m_l \text{ and } q = q_l, \\ 0, & \text{otherwise.} \end{cases} \tag{5.11}$$

Under these assumptions, the different signal paths can be perfectly resolved by L different antenna elements (in general, $L << N_R, N_T$) without interference. This situation allows for a simplified capacity analysis and a multiplexing scheme with reduced complexity and infrastructure that achieves it: the OPDM commented before.

According to [34], the capacity of a narrowband time-invariant wireless MIMO channel (like (5.10), for the moment) can be computed by means of the SVD of the channel matrix:

$$\mathbf{H}_{opt} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^H. \tag{5.12}$$

This representation decomposes the channel into a set of parallel, decoupled scalar Gaussian sub-channels[3], which are rotated by the unitary matrices $\mathbf{U} \in \mathbb{C}^{N_R \times N_R}$ and $\mathbf{V} \in \mathbb{C}^{N_T \times N_T}$ and scaled by the rectangular diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{N_R \times N_T}$. We assume the singular values in $\mathbf{\Lambda}$ are sorted in decreasing order. As stated in **Section 2.1**, the number of singular values n_{min} is the rank of $\mathbf{H}_{opt}$, which in general MIMO channels, is $\min(N_T, N_R)$. However, in our ideal LAA case, $\mathbf{H}_{opt}$ is sparse, and its rank is the number of different independent signal paths, *i.e.* $n_{min} \triangleq L$ (it is usually called *rank-deficient*). Then, the columns of $\mathbf{U}$ and $\mathbf{V}$ with indexes greater than L are set to $\mathbf{0}$. This low-rank aspect of $\mathbf{H}_{opt}$ can be better appreciated by rewriting (5.12) as

$$\mathbf{H}_{opt} \triangleq \sum_{l=1}^{L} \lambda_l \mathbf{u}_l \mathbf{v}_l^H, \tag{5.13}$$

given that $\mathbf{u}_l$ and $\mathbf{v}_l$ are the columns of $\mathbf{U}$ and $\mathbf{V}$, respectively.

Remember that, according to the general mmWave model (5.3), $\alpha_l = \sqrt{\beta \gamma_l} e^{j\omega_l}$. Notice that (5.10) and (5.12) have the same formulation if we set

$$\mathbf{u}_l \triangleq \delta(\mathbf{m} - m_l \mathbf{1}) e^{j\frac{\omega_l}{2}} \Rightarrow \mathbf{U} \triangleq \left[\delta(\mathbf{m} - m_1 \mathbf{1}) e^{j\frac{\omega_1}{2}}, \dots, \delta(\mathbf{m} - m_L \mathbf{1}) e^{j\frac{\omega_L}{2}}\right] \tag{5.14}$$

$$\mathbf{v}_l \triangleq \delta(\mathbf{q} - q_l \mathbf{1}) e^{-j\frac{\omega_l}{2}} \Rightarrow \mathbf{V} \triangleq \left[\delta(\mathbf{q} - q_1 \mathbf{1}) e^{-j\frac{\omega_1}{2}}, \dots, \delta(\mathbf{q} - q_L \mathbf{1}) e^{-j\frac{\omega_L}{2}}\right] \tag{5.15}$$

$$\lambda_l \triangleq \tilde{D}^2 \sqrt{\beta \gamma_l} \Rightarrow \mathbf{\Lambda} \triangleq \tilde{D}^2 \sqrt{\beta} \operatorname{diag}(\sqrt{\gamma_1}, \dots, \sqrt{\gamma_L}). \tag{5.16}$$

(5.13) is the sum of L rank-one matrices $\lambda_l \mathbf{u}_l \mathbf{v}_l^H$. Each one of them defines a scalar *eigenchannel* with an associated capacity that can support a data stream. The total capacity of the MIMO channel is the sum of the ones from each independent sub-channel. Following the procedure in [34], we allocate an

[3] Under the typical MIMO signal model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$ with AWGN $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, P_z \mathbf{I})$.

optimal signal power $P_{opt,l}$ on the lth sub-channel with the *waterfilling* method, under the constraint $P_{tot} = \sum_{l=1}^{L} P_{opt,l}$. Then, for a noise power P_z, we can easily compute the total channel capacity:

$$\boxed{C = \sum_{l=1}^{L} \log\left(1 + \frac{P_{opt,l}\,|\lambda_l|^2}{P_z}\right)}. \tag{5.17}$$

$|\lambda_l|^2 = \tilde{D}^4 \beta \gamma_l$ are the eigenvalues of

$$\mathbf{H}\mathbf{H}^H = \mathbf{U}\mathbf{\Lambda}\overbrace{\mathbf{V}^H\mathbf{V}}^{\mathbf{I}_{N_T}}\mathbf{\Lambda}^H\mathbf{U}^H = \mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}\mathbf{U}^H = \mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^H. \tag{5.18}$$

It is usually helpful to study (5.17) under the high and low SNR regimes separately. Their analysis will allow us to obtain important insights on the characteristics of the LAA channel. When the SNR is very high ($\mathrm{SNR} \triangleq P_{tot}/P_z \to \infty$), the optimal waterfilling policy asymptotically allocates the same amount of power ($P_{high,l} \triangleq P_{tot}/L \triangleq P_{high}$) to each eigenchannel. Hence

$$\begin{aligned} C_{high} &\triangleq \lim_{\mathrm{SNR}\to\infty} C = \lim_{\mathrm{SNR}\to\infty} \sum_{l=1}^{L} \log\left(1 + \frac{P_{high}\,|\lambda_l|^2}{P_z}\right) = \lim_{\mathrm{SNR}\to\infty} \sum_{l=1}^{L} \log\left(\frac{P_{tot}\,|\lambda_l|^2}{LP_z}\right) \\ &= \sum_{l=1}^{L}\left(\log\left(\frac{P_{tot}}{P_z}\right) + \log\left(\frac{|\lambda_l|^2}{L}\right)\right) = L\log\mathrm{SNR} + \sum_{l=1}^{L}\log\left(\frac{|\lambda_l|^2}{L}\right). \end{aligned} \tag{5.19}$$

Also, by *Jensen's inequality*

$$\frac{1}{L}\sum_{l=1}^{L}\log\left(1 + \frac{P_{high}|\lambda_l|^2}{LP_z}\right) \leq \log\left(1 + \frac{P_{tot}}{LP_z}\left(\frac{1}{L}\sum_{l=1}^{L}|\lambda_l|^2\right)\right) \tag{5.20}$$

$$C_{high} \leq L\log\left(1 + \frac{P_{tot}}{LP_z}\left(\frac{1}{L}\sum_{l=1}^{L}|\lambda_l|^2\right)\right) = L\log\left(1 + \frac{P_{tot}}{LP_z}\left(\frac{1}{L}\,\mathrm{Tr}\left[\mathbf{H}\mathbf{H}^H\right]\right)\right), \tag{5.21}$$

since

$$\mathrm{Tr}\left[\mathbf{H}\mathbf{H}^H\right] = \mathrm{Tr}\left[\mathbf{U}\mathbf{\Lambda}^2\mathbf{U}^H\right] = \mathrm{Tr}\left[\mathbf{\Lambda}^2\overbrace{\mathbf{U}^H\mathbf{U}}^{\mathbf{I}_{N_R}}\right] = \mathrm{Tr}\left[\mathbf{\Lambda}^2\right] = \sum_{l=1}^{L}|\lambda_l|^2. \tag{5.22}$$

It is clear that, in this regime, the path-sparsity characteristic of mmWave communications is the main limiting factor of the multiplexing capability of the channel. The rank of $\mathbf{H}_{opt}$ (*i.e.* the number of *spatial DoF*) is dictated by the amount of reliable signal paths between Tx and Rx, L, which is usually much lower than N_T or N_R. It is a known result [34] that, at high SNR, the MIMO channel with the highest capacity is the one which has uniform singular values (*i.e. well-conditioned*). This is not the case for the LAA-MIMO channel: under ideal path alignment, most singular values are 0, and even when the power leakage is maximum, the condition number of $\mathbf{H}_{opt}$, defined as

$$c\left(\mathbf{H}_{opt}\right) \triangleq \frac{\max_l \lambda_l}{\min_l \lambda_l}, \tag{5.23}$$

is still very high.

Regarding the low SNR regime, the waterfilling policy will asymptotically allocate all the power to the strongest eigenchannel (λ_{MAX}) as $\mathrm{SNR} \to 0$. Then, the capacity in that regime will be

$$C_{low} \triangleq \lim_{\mathrm{SNR}\to 0} C = \lim_{\mathrm{SNR}\to 0} \log\left(1 + \frac{P_{tot}\,|\lambda_{MAX}|^2}{P_z}\right) = \frac{P_{tot}\,|\lambda_{MAX}|^2}{P_z}\log_2 e. \tag{5.24}$$

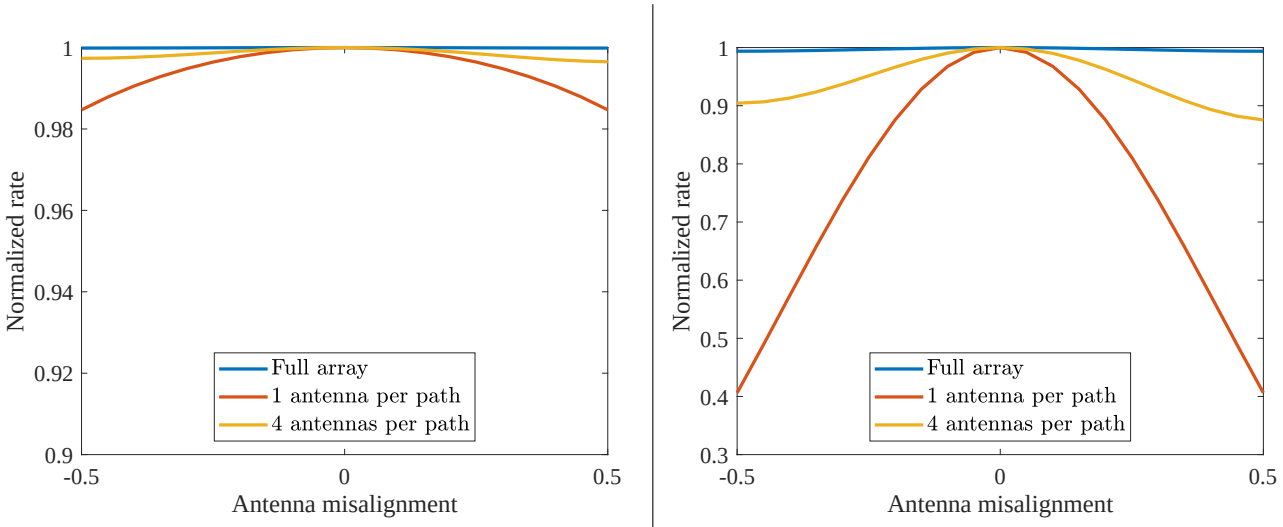
In this regime the condition number of $\mathbf{H}_{opt}$ becomes irrelevant, since the main limiting factor is how much energy can be transferred from Tx to Rx through the strongest eigenchannel. The MIMO channel provides a power gain of $|\lambda_{MAX}|^2 = \tilde{D}^4\beta\gamma_{MAX}$. The LAA proves to be well-suited for low SNR environments, as it remarkably increases the power gain thanks to the energy focusing capacity of the EM lens.

5.1.2 Capacity degradation factors

After the previous theoretical analysis of the symmetric LAA channel capacity, we will now present some numerical examples. We want to experimentally characterize how different parameters of the channel and transceivers affect its achievable rate. In particular, we will study the effect of antenna selection and path overlap.

The experimental setting studied consists in a symmetric LAA link in which the Tx and the Rx are equipped with an EM lens of aperture $\tilde{D} = 32$ ($N_T = N_R = 65$). We will consider both high-SNR (30dB) and low-SNR (-10dB) cases. For the first one, there are three paths between the transceivers which can allocate reliable transmission of data ($\alpha_l = 1, \quad l = 1, 2, 3$), while for the second one, only a single path is available ($\alpha_1 = 1$).

In the first test, we will observe the impact of antenna selection on the (normalized) maximum achievable rate in both regimes. The AoA of one of the paths will move around its assigned antenna to see the effects of various misalignments, *i.e.* ϵ_1 will vary between -0.5 and 0.5.



(a) High-SNR regime.

(b) Low-SNR regime.

Figure 5.1: Achievable rate for various antenna selection constraints.

It is clear from Figures 5.1a and 5.1b that path misalignment affects the achievable rate negligibly if the full array can be processed, since it is able to capture all the energy. A noticeable drop is experienced when only a single antenna can be allocated per path. In the high-SNR regime, the rate reduction is $< 2\%$, since there are still two remaining paths whose energy is perfectly captured by the LAA. However, the degradation is very prominent ($\sim 60\%$) in the low-SNR regime, as the full rate depends on a single path and its energy is only received by a single antenna. Nonetheless, this issue can be mitigated by allocating more antennas per beam. With just 4 antennas dedicated on each path, the rate reduction can decrease to less than 10% in the low-SNR scenario and become insignificant in the high-SNR one.

The second test is similar to the first one but only the high-SNR regime will be considered. In this case, the AoAs of paths 1 and 2 will lie on two consecutive array elements. The misalignment of one of them will increase until both paths fully overlap. As commented in the previous section, channels with similar singular values allow for larger achievable rates in the high-SNR regime. The condition number of $\mathbf{H}$ (considering only the 3 non-negligible singular values) is a good indicator of the multiplexing capabilities of the channel. When $c(\mathbf{H})$ is close to 1, the channel is said to be *well-conditioned* and it is well suited for multiple stream transmission in high-SNR. On the contrary, when its value is very high, it implies that some of the virtual independent subchannels have become too mixed for reliable multiplexing. This metric provides a more nuanced view of the channel than its rank.

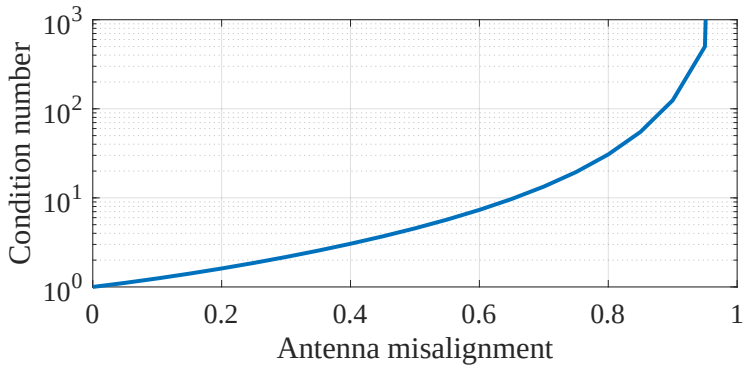


Figure 5.2: Condition number of the channel matrix at high-SNR regime.

In Figure 5.2, we have plotted the condition number of the LAA channel matrix as two paths overlap. In general, when signal beams are well separated in the angular domain, the channel is well-conditioned. Even for the maximum misalignment considered in the model (4.24), $\epsilon = 0.5$, the condition number does not surpass 5, which is a relatively low value. However, when beams are separated by less than half the LAA resolution, the multiplexing capability of the channel for different streams of data will dwindle, fully losing a spatial DoF when completely overlapping.

Chapter 6

Point to Point Communications with Lens Antenna Arrays

In the previous chapter, we have studied the symmetric LAA-MIMO channel, for P2P mmWave communications. In that scenario, both the Tx and the Rx employ a multiantenna system with an EM lens. Real world implementations of such systems could find application in *vehicular communications*, in which space and energy constraints are reasonably soft and allow the deployment of moderately sized equipment.

One of the goals of this thesis is to formulate transmission schemes based on LAA for regular cellular mobile communications. The architecture of cellular networks and the devices they encompass present requirements that may not be met under the symmetric LAA-MIMO paradigm. While BSs are supported by static infrastructure and can allocate large equipment, MTs have very strict limitations. On the one hand, MTs are usually battery-based, so they cannot afford arrays with many elements, since the energy cost of activating and processing them is prohibitive. On the other hand, MTs must meet mobility and encapsulation standards, discarding LAAs completely, especially the ones with semispherical geometry.

The most suitable incorporation of LAAs in cellular mobile communications is the *asymmetric LAA-MIMO* paradigm. Instead of deploying EM lenses on both sides of the link, only the BS will have such equipment, while the MTs will rely on regular antenna arrays or single antennas. With this philosophy, most signal processing will be carried out at the BS, removing the computational and power load from the MTs.

The next sections are organized as follows. Firstly, we propose a variety of coherent data transmission schemes for LAA systems. We have classified them in two groups: the ones that rely on full CSIT and the ones that rely on full Channel State Information at the Receiver (CSIR). Since coherent schemes are not the focus of this thesis, these two sections do not provide an in-depth study; they simply serve to illustrate the range of possibilities that asymmetric LAA systems can offer. Afterwards, we introduce two noncoherent frameworks: one for an UL scenario and one for a DL one. The general notions derived from them will be the foundations for the multi-user schemes we will develop in **Chapter 7**.

For simplicity, all the schemes presented will consider single antenna MTs, which means that the UL will be Single Input Multiple Output (SIMO) and the DL will be Multiple Input Single Output (MISO).

6.1 Full CSIT

Let us consider a link between a LAA BS with N_{BS} elements and a single antenna MT ($N_{MT} = 1$). Since CSIT can only be fully exploited by Txs with multiple antennas, we will just consider the DL scenario (MISO). The signal received by the MT at time instant k takes the form:

$$y[k] = \mathbf{h}^H \mathbf{x}[k] + z[k]. \tag{6.1}$$

$\mathbf{h}$ is the asymmetric LAA-MISO channel vector, whose expression is

$$\mathbf{h} \triangleq \tilde{D} \sum_{l=1}^{L} \alpha_l \operatorname{sinc}\left(\mathbf{n} - \tilde{D}\tilde{\phi}_l \mathbf{1}\right) \in \mathbb{C}^{N_{BS} \times 1}. \tag{6.2}$$

As expected, it is very similar to the symmetric case (see (5.8)) but with the LAA response vector on a single side. $z[k] \sim \mathcal{CN}(0, P_z)$ is AWGN and $\mathbf{x}[k]$ is the transmitted signal, constructed as $\mathbf{x}[k] \triangleq \mathbf{\Sigma} \mathbf{w} s[k]$. $s[k]$ are *unit-energy* symbols and $\mathbf{w}$ is the *beamforming* vector, whose norm is constrained to the square root of the total transmit power: $\|\mathbf{w}\| \equiv \sqrt{P_{BS}}$. $\mathbf{\Sigma}$ is the antenna selection matrix, as defined in **Section 2.2**.

It is well-known [34] that the optimal strategy (in terms of SNR) in a MISO channel with full CSIT is aligning the data symbols along the direction of the channel vector, known as *transmit beamforming*. This technique compensates the phases of each channel component, allocates more power to antennas with better gain and converts the MISO channel into a scalar AWGN one. The beamforming vector must then be

$$\mathbf{w}_B \triangleq \sqrt{P_{BS}} \frac{\tilde{\mathbf{h}}}{\left\|\tilde{\mathbf{h}}\right\|}. \tag{6.3}$$

$\tilde{\mathbf{h}} \triangleq \mathbf{\Sigma} \mathbf{h}$ is the equivalent channel vector under antenna selection. With this approach, the received signal is

$$\begin{aligned} y[k] &= \mathbf{h}^H \mathbf{\Sigma} \mathbf{w}_B s[k] + z[k] \\ &= \sqrt{P_{BS}} \frac{\mathbf{h}^H \mathbf{\Sigma} \tilde{\mathbf{h}}}{\left\|\tilde{\mathbf{h}}\right\|} s[k] + z[k] \\ &= \sqrt{P_{BS}} \frac{\tilde{\mathbf{h}}^H \tilde{\mathbf{h}}}{\left\|\tilde{\mathbf{h}}\right\|} s[k] + z[k] = \sqrt{P_{BS}} \left\|\tilde{\mathbf{h}}\right\| s[k] + z[k]. \end{aligned} \tag{6.4}$$

With transmit beamforming we have obtained a power gain of $\left\|\tilde{\mathbf{h}}\right\|^2$. This gain is maximum when the channel beams are perfectly aligned with the Rx (*i.e.* no power leakage) and the effect of antenna selection is null:

$$\begin{aligned} \max \|\mathbf{h}\|^2 &= \tilde{D}^2 \left\| \sum_{l=1}^{L} \alpha_l \operatorname{sinc}\left(\mathbf{n} - n_l \mathbf{1}\right) \right\|^2 \\ &= \tilde{D}^2 \left\| \sum_{l=1}^{L} \alpha_l \delta\left(\mathbf{n} - n_l \mathbf{1}\right) \right\|^2 = \tilde{D}^2 \sum_{l=1}^{L} |\alpha_l|^2 . \end{aligned} \tag{6.5}$$

Then, the maximum achievable rate for this scheme is

$$\boxed{R \leq \log\left(1 + \overbrace{\frac{P_{BS} \tilde{D}^2}{P_z} \sum_{l=1}^{L} |\alpha_l|^2}^{\mathrm{SNR}_{max}}\right)}. \tag{6.6}$$

This strategy is easily translated in the scenario in which the MT is multi-antenna ($N_{MT} > 1$). The advantage of transmit beamforming in the asymmetric MIMO channel compared to the MISO one is that it possesses extra DoF ($\mathrm{DoF} = \min\left(N_{BS}, N_{MT}, L\right)$), which can be employed as power gain under low-SNR regime or as multiplexing gain under high-SNR regime [34].

6.2 Full CSIR

Since the UL channel is the reciprocal of the DL one, it is reasonable that the beamforming strategy presented before can be inverted and applied to a full CSIR scenario. Indeed, in the UL channel, the MT transmits the signal $x[k] \triangleq \sqrt{P_{MT}}s[k]$ to the BS. $s[k]$ are unit energy symbols as in the previous case and P_{MT} is the MT transmit power. The signal received at the BS is then

$$\begin{aligned}\mathbf{y}[k] &= \mathbf{\Sigma}\mathbf{h}x[k] + \mathbf{z}[k] \in \mathbb{C}^{N_{BS}\times 1} \\ &= \tilde{\mathbf{h}}x[k] + \mathbf{z}[k],\end{aligned} \tag{6.7}$$

which has a similar structure to (6.1). $\mathbf{z}[k]$ is now $\mathcal{CN}(\mathbf{0}, P_z\mathbf{I})$. The Rx then applies the linear combination (*i.e. receive beamforming*) that maximizes the SNR [45], the Maximum Ratio Combining (MRC):

$$\tilde{y}[k] = \mathbf{w}_M^H \mathbf{y}[k]. \tag{6.8}$$

The weighting vector $\mathbf{w}_M \triangleq \frac{\tilde{\mathbf{h}}}{\|\tilde{\mathbf{h}}\|}$ is defined the same way as $\mathbf{w}_B$ in (6.3) because its working principle is its natural reciprocal: to project the received signal onto $\tilde{\mathbf{h}}$. The resulting received signal is

$$\tilde{y}[k] = \mathbf{w}_M^H\left(\tilde{\mathbf{h}}x[k] + \mathbf{z}[k]\right) = \sqrt{P_{MT}}\left\|\tilde{\mathbf{h}}\right\| s[k] + \overbrace{\mathbf{w}_M^H\mathbf{z}[k]}^{\tilde{z}[k]}. \tag{6.9}$$

Once again, we have converted the SIMO channel into a scalar AWGN one. The SNR and capacity analysis is the same as in the CSIT case, since the Rx beamforming vector has unit norm and does not affect the noise power.

Regarding the full CSIR DL scenario, that is, the MT has full information of the channel state, a variety of schemes can be designed depending on the level of knowledge available at the BS. In general, when CSIR is present and the Tx employs multiple antennas, STBC can bring significant diversity gain: by transmitting special codes isotropically [34], the system can successfully garner the full DoFs of the channel. This approach has been classically antagonistic to transmit beamforming. However, by cleverly utilizing the CSI available in both ends of the link, hybrid schemes have been proposed [46, 47] that take advantage of the strengths of both philosophies.

For simplicity, we will study these methods in a channel with $L = 2$ reliable paths between BS and MT, which will allow us to use the commonly called *Alamouti scheme* [34]. Nonetheless, the same ideas can be applied to more general channels[1]. The model of the signal received at the MT follows the already presented asymmetric LAA-MISO channel (6.1):

$$\begin{aligned}y[k] &= \mathbf{h}^H\mathbf{x}[k] + z[k] \\ &= \tilde{D}\sum_{l=1}^{2}\alpha_l \operatorname{sinc}\left(\mathbf{n}^H - (n_l+\epsilon_l)\,\mathbf{1}^H\right)\mathbf{x}[k] + z[k].\end{aligned} \tag{6.10}$$

The transmitted signal is constructed as

$$\mathbf{x}[2k'] \triangleq \mathbf{\Sigma}\left(\frac{1}{\sqrt{2}}\mathbf{w}_{A,1}s_1[2k'] + \frac{1}{\sqrt{2}}\mathbf{w}_{A,2}s_2[2k']\right) \tag{6.11}$$

$$\mathbf{x}[2k'+1] \triangleq \mathbf{\Sigma}\left(\frac{1}{\sqrt{2}}\mathbf{w}_{A,1}s_1[2k'+1] + \frac{1}{\sqrt{2}}\mathbf{w}_{A,2}s_2[2k'+1]\right), \tag{6.12}$$

where $\mathbf{w}_{A,1}$, $\mathbf{w}_{A,2}$ are two distinct beamforming vectors and $s_1[k]$, $s_2[k]$ are two instances taken from a sequence of unit energy symbols, $\{s[k]\}$. $\mathbf{\Sigma}$ is the antenna selection matrix. The encoding of $s_1[k]$

	$2k'$	$2k'+1$
s_1	$s[2k']$	$-s^*[2k'+1]$
s_2	$s[2k'+1]$	$s^*[2k']$

Table 6.1: Alamouti's code.

and $s_2[k]$ is specified in Table 6.1. By taking this codification into (6.10) and assuming the channel remains constant over two symbol times, we have

$$y[2k'] = \tilde{\mathbf{h}}^H \left(\frac{1}{\sqrt{2}} \mathbf{w}_{A,1} s[2k'] + \frac{1}{\sqrt{2}} \mathbf{w}_{A,2} s[2k'+1] \right) + z[2k'] \tag{6.13}$$

$$y[2k'+1] = \tilde{\mathbf{h}}^H \left(\frac{-1}{\sqrt{2}} \mathbf{w}_{A,1} s^*[2k'+1] + \frac{1}{\sqrt{2}} \mathbf{w}_{A,2} s^*[2k'] \right) + z[2k'+1]. \tag{6.14}$$

We can write the previous two expressions in space-time matrix form[2]:

$$\begin{aligned} \begin{bmatrix} y[2k'] \\ y^*[2k'+1] \end{bmatrix} &= \frac{1}{\sqrt{2}} \begin{bmatrix} \tilde{\mathbf{h}}^H \mathbf{w}_{A,1} & \tilde{\mathbf{h}}^H \mathbf{w}_{A,2} \\ \mathbf{w}_{A,2}^H \tilde{\mathbf{h}} & -\mathbf{w}_{A,1}^H \tilde{\mathbf{h}} \end{bmatrix} \begin{bmatrix} s[2k'] \\ s[2k'+1] \end{bmatrix} + \begin{bmatrix} z[2k'] \\ z^*[2k'+1] \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} g_1 & g_2 \\ g_2^* & -g_1^* \end{bmatrix} \begin{bmatrix} s[2k'] \\ s[2k'+1] \end{bmatrix} + \begin{bmatrix} z[2k'] \\ z^*[2k'+1] \end{bmatrix} \\ \check{\mathbf{y}} &= \frac{1}{\sqrt{2}} \mathbf{G}_A \mathbf{s}[2k'] + \check{\mathbf{z}}[2k']. \end{aligned} \tag{6.15}$$

We have defined $g_1 \triangleq \tilde{\mathbf{h}}^H \mathbf{w}_{A,1}$ and $g_1 \triangleq \tilde{\mathbf{h}}^H \mathbf{w}_{A,2}$. Since the MT has full knowledge of the channel, it can obtain these two values, as well, and compute the receive weights matrix

$$\mathbf{W}_A \triangleq \frac{\mathbf{G}_A^H}{\|\mathbf{G}_A\|_F^2} = \frac{1}{|g_1|^2 + |g_2|^2} \mathbf{G}_A^H. \tag{6.16}$$

Weighting the received signal accordingly, we finally obtain

$$\begin{aligned} \tilde{\mathbf{y}} &= \mathbf{W}_A \left(\frac{1}{\sqrt{2}} \mathbf{G}_A \mathbf{s}[2k'] + \check{\mathbf{z}}[2k'] \right) \\ &= \frac{1}{\sqrt{2}} \frac{1}{|g_1|^2 + |g_2|^2} \begin{bmatrix} g_1^* & g_2 \\ g_2^* & -g_1 \end{bmatrix} \begin{bmatrix} g_1 & g_2 \\ g_2^* & -g_1^* \end{bmatrix} \begin{bmatrix} s[2k'] \\ s[2k'+1] \end{bmatrix} + \mathbf{W}_A \check{\mathbf{z}}[2k'] \\ &= \frac{1}{\sqrt{2}} \frac{1}{|g_1|^2 + |g_2|^2} \begin{bmatrix} |g_1|^2 + |g_2|^2 & 0 \\ 0 & |g_1|^2 + |g_2|^2 \end{bmatrix} \begin{bmatrix} s[2k'] \\ s[2k'+1] \end{bmatrix} + \tilde{\mathbf{z}}[2k] \\ &= \frac{1}{\sqrt{2}} \mathbf{s}[2k'] + \tilde{\mathbf{z}}[2k']. \end{aligned} \tag{6.17}$$

This scheme allows the transmission of 2 data symbols in 2 channel uses (*full rate*). It successfully transforms our original communication problem into 2 parallel AWGN subchannels. Each one of them

[1] If $L > 2$, two different approaches can be carried out: generalize the schemes for larger STBCs or arrange the paths in two disjoint groups and treat them as individual beams.

[2] Contrary to other instances of space-time notation throughout this work, the convention used in this section is rows for time.

carries a symbol with power $\frac{1}{2}$. Both noise components have power

$$\begin{aligned}\mathbb{E}\left[\left|\tilde{z}_1[2k']\right|^2\right] = \mathbb{E}\left[\left|\tilde{z}_2[2k']\right|^2\right] &= \frac{1}{\left(|g_1|^2+|g_2|^2\right)^2}\mathbb{E}\left[\left|g_1^* z[2k'] + g_2 z^*[2k'+1]\right|^2\right] \\ &= \frac{|g_1|^2 P_z + |g_2|^2 P_z + g_1^* g_2^* \mathbb{E}\left[z[2k']z[2k'+1]\right] + g_1 g_2 \mathbb{E}\left[z^*[2k']z^*[2k'+1]\right]}{\left(|g_1|^2+|g_2|^2\right)^2} \\ &= \frac{\left(|g_1|^2+|g_2|^2\right)P_z}{\left(|g_1|^2+|g_2|^2\right)^2} = \frac{P_z}{|g_1|^2+|g_2|^2}, \end{aligned} \tag{6.18}$$

so the effective SNR is

$$\boxed{\text{SNR} = \frac{|g_1|^2+|g_2|^2}{2P_z}}. \tag{6.19}$$

The design of $\mathbf{w}_{A,1}$ and $\mathbf{w}_{A,2}$ and, as a consequence, the performance of the scheme, will be conditioned by the amount of channel knowledge available at the Tx. In general terms, we can consider three degrees of CSIT:

- **AoDs:** If the BS knows the directions to reliably communicate with the MT, it can cleverly allocate its power to maximize the SNR. This is the default case in LAA systems that employ antenna selection. In the ideal case (perfect beam alignment), the beamforming vectors and their corresponding gains are:

$$\mathbf{w}_{A,l} = \sqrt{\frac{P_{BS}}{2}}\,\text{sinc}\left(\mathbf{n}-n_l\mathbf{1}\right) = \sqrt{\frac{P_{BS}}{2}}\delta\left(\mathbf{n}-n_l\mathbf{1}\right) \tag{6.20}$$

$$g_l = \mathbf{h}^H\mathbf{w}_{A,l} = \tilde{D}\alpha_l\sqrt{\frac{P_{BS}}{2}}\delta\left(\mathbf{n}^H - n_l\mathbf{1}^H\right)\delta\left(\mathbf{n}-n_l\mathbf{1}\right) = \tilde{D}\alpha_l\sqrt{\frac{P_{BS}}{2}} \quad , \quad l=1,2. \tag{6.21}$$

  This results in a SNR of

$$\text{SNR} = \frac{\tilde{D}^2 P_{BS}\left(|\alpha_1|^2+|\alpha_2|^2\right)}{4P_z}, \tag{6.22}$$

  6dB below the one in (6.6).

- **AoD with the highest gain:** Beyond knowing the AoDs, the Tx may also be aware of which one presents a higher gain. This information is especially useful at low-SNR regimes, since all the available power can be allocated to a single beam. The maximum resulting SNR in this case is

$$\text{SNR} = \frac{\tilde{D}^2 P_{BS}\left|\alpha_{max}\right|^2}{2P_z}. \tag{6.23}$$

- **AoDs + Path gains:** Finally, knowing the gain profile of each path allows the BS to optimally allocate its available power proportionally. The power assigned to each beam is

$$P_{BS,l} \triangleq \sqrt{\frac{P_{BS}|\alpha_l|^2}{|\alpha_1|^2+|\alpha_2|^2}} \quad , \quad l=1,2. \tag{6.24}$$

Hence, the resulting SNR is

$$g_l = \tilde{D}\alpha_l \sqrt{\frac{P_{BS}\left|\alpha_l\right|^2}{\left|\alpha_1\right|^2 + \left|\alpha_2\right|^2}} \Rightarrow \text{SNR} = \frac{\tilde{D}^2 P_{BS}\left(\left|\alpha_1\right|^2 + \left|\alpha_2\right|^2\right)}{2P_z} \tag{6.25}$$

This approach can yield much better performance than when only AoDs are known, only 3dB below MRC.

6.3 Noncoherent schemes

The transmission schemes we have presented so far rely on perfect CSI either on the Tx or Rx end. As explained in **Section 2.1**, this knowledge usually relies on training-based channel estimation techniques, which can imply various issues that may hinder the communication, especially in mmWave mobile environments. Given the special characteristics of the LAA channel, we strongly advocate for the development of noncoherent and blind schemes that thoroughly exploit its strengths. It must be stated that so called *truly noncoherent* methods [22] are not the topic of interest in this thesis. Indeed, in a channel with a structure as clearly defined as the LAA, discarding the statistical knowledge about it would entail a remarkable loss in performance.

In this section, we present two noncoherent schemes: Multiple-Symbol Differential Detection (MSDD) for the UL and Differential Orthogonal Space-Time Block Coding (D-OSTBC) for the DL. They are both built upon the foundations of CDD, generalizing and improving it. Beside presenting them in a generic form, we specifically characterize them for LAA-SIMO and LAA-MISO. The intuitions behind their operation will be of great usefulness in **Chapter 7**, where we will modify and extend them to the multi-user scenario.

6.3.1 Multiple Symbol Differential Detection (MSDD)

As stated in **Section 2.3**, one of the main drawbacks of CDD is the fundamental 3dB penalty in performance against coherent schemes. It is well-known that this limit can be improved by estimating a sequence of symbols jointly. In fact, the performance loss can be arbitrarily reduced as the length of the sequence increases[3]. The scheme that takes advantage of this idea is known as MSDD.

Consider the UL SIMO scenario from previous sections, with a BS equipped with a LAA of $N_{BS} >> 1$ antennas as the Rx and a single-antenna MT ($N_{MT} = 1$) as the Tx. The received signal can be represented in the compact space-time block notation as

$$\begin{gathered} \mathbf{Y} = \mathbf{h}\mathbf{x}^T + \mathbf{Z} \in \mathbb{C}^{N_{BS}\times K} \\ [\mathbf{y}[0], \ldots, \mathbf{y}[K-1]] = \mathbf{h}\,[x[0], \ldots, x[K-1]] + [\mathbf{z}[0], \ldots, \mathbf{z}[K-1]] \end{gathered} \tag{6.26}$$

in which each row is a different BS antenna (space $\downarrow$) and each column is a different time instant (time $\rightarrow$). The signal block[4] $\mathbf{x}^T$ contains a sequence of K differentially encoded symbols, $x[k] = x[k-1]s[k]$, transmitted by the MT. $s[k]$ are taken from a unitary constellation $\mathcal{S}$ of size M (M-DPSK). $\mathbf{h}$ is the channel vector, which takes the form

$$\mathbf{h} \triangleq \tilde{D}\sum_{l=1}^{L} \alpha_l \,\text{sinc}\left(\mathbf{n} - \tilde{D}\tilde{\phi}_{BS,l}\mathbf{1}\right) \in \mathbb{C}^{N_{BS}\times 1}. \tag{6.27}$$

[3] Asymptotically, performing this joint differential detection on an infinite sequence of symbols would yield the same performance as a coherent scheme. However, this is not achievable in practice, and the performance is limited by the maximum window of time in which the channel coefficients remain constant.

[4] In contrast to other well-established literature dealing with this SIMO model [8, 21–24, 48] we will keep the transpose symbol $\cdot^T$ throughout this chapter for clarity of notation when computing inner and outer products.

For simplicity, we assume *Rayleigh fading* [34] and the path gains α_l will be distributed as Circularly Symmetric Complex Gaussian (CSCG) variables $\mathcal{CN}(0, P_l)$ (which is consistent with the multi-path channel model presented in **Chapter 5**). $\mathbf{h}$ is therefore the sum of LAA response vectors scaled by univariate complex normal coefficients, *i.e.* $\mathbf{h}$ is a complex normal multidimensional variable with statistics $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C_h})$:

$$\begin{aligned}\mathbb{E}[\mathbf{h}] &= \mathbb{E}\left[\sum_{l=1}^{L} \alpha_l \mathbf{a}_{BS}(\phi_{BS,l})\right] = \sum_{l=1}^{L} \mathbb{E}[\alpha_l \mathbf{a}_{BS}(\phi_{BS,l})] \\ &= \sum_{l=1}^{L} \mathbb{E}[\alpha_l]\, \mathbf{a}_{BS}(\phi_{BS,l}) = \mathbf{0} \qquad (6.28)\end{aligned}$$

$$\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right] \triangleq \mathbf{C_h}.^{5} \qquad (6.29)$$

Finally, $\mathbf{Z} \sim \mathcal{MCN}(\mathbf{0}, P_z\mathbf{I}_{N_R}, \mathbf{I}_K)$ is a matrix that contains AWGN.

As the authors in [21] point out, this system model is closely related to Impulse Radio Ultra Wideband (IR-UWB). In that scenario, it is common to employ binary DPSK, so users transmit information in the sign of a sequence of short pulses. As a consequence of multipath propagation, the receiver perceives a number of echoes for each transmitted pulse. To avoid Inter-Symbol Interference (ISI), the symbol duration is chosen large enough so that all echoes have been received before the next symbol starts. We can draw a clear analogy to the mmWave scenario we are dealing with in the thesis: while in IR-UWB the echoes are spread along the temporal dimension, in LAA-SIMO the echoes are spread along the spatial dimension (multiple antennas of the Rx). Furthermore, the IR-UWB problem can be easily expressed using the same formulation as in the transmission block model (6.26). It is then natural to assume that LAA-SIMO shares the same advantages and disadvantages as IR-UWB systems. In particular, they both present numerous practical incentives to avoid explicit channel estimation and instead favor noncoherent approaches. For all these reasons, we will employ and extend the tools developed in [22] for IR-UWB, applied to LAA-MIMO in the line of research from [21, 23].

Let us consider the ML estimation for the transmit symbols at the Rx. The Probability Density Function (PDF) of the received signal block conditioned on a hypothesis for the transmitted sequence $\mathbf{x}$ and a channel realization $\mathbf{h}$ is

$$f_{\mathbf{Y}}(\mathbf{Y}|\mathbf{x}, \mathbf{h}) = \frac{1}{(\pi P_z)^{N_{BS}K}} \exp\left(\frac{-1}{P_z}\left\|\mathbf{Y} - \mathbf{h}\mathbf{x}^T\right\|_F^2\right), \qquad (6.30)$$

which is the PDF of the noise $\mathbf{Z}$ displaced by $\mathbf{h}\mathbf{x}^T$. If $\mathbf{h}$ was known to the receiver (full CSIR), the optimum ML coherent detector would be

$$\widehat{\mathbf{x}}_{coh}^{ML} = \underset{\mathbf{x}\in\mathcal{S}^K,\, x[0]\equiv 1}{\arg\max}\; f_{\mathbf{Y}}(\mathbf{Y}|\mathbf{x}, \mathbf{h}), \qquad (6.31)$$

with $\mathcal{S}^K$ being the set of all possible sequences of symbols from $\mathcal{S}$. Since this is not the case, we have to derive the ML noncoherent detector. To do so, we must remove the uncertainty of $\mathbf{h}$ from the detection, taking into account its statistics. The PDF of $\mathbf{h}$ is

$$f_{\mathbf{h}}(\mathbf{h}) = \frac{1}{\pi^{N_{BS}}\left|\mathbf{C_h}\right|} \exp\left(-\mathbf{h}^H {\mathbf{C_h}}^{-1}\mathbf{h}\right). \qquad (6.32)$$

We average (6.30) over it to obtain the ML marginal distribution

$$f_{\mathbf{Y}}^{ML}(\mathbf{Y}|\mathbf{x}) = \mathbb{E}_{\mathbf{h}}\left[f_{\mathbf{Y}}(\mathbf{Y}|\mathbf{x}, \mathbf{h})\right] = \int_{\mathbb{C}^{N_{BS}}} f_{\mathbf{Y}}(\mathbf{Y}|\mathbf{x}, \mathbf{h})\, f_{\mathbf{h}}(\mathbf{h})\, d\mathbf{h} d\mathbf{h}^*. \qquad (6.33)$$

[5] See **Section 7.1** for a complete derivation of $\mathbf{C_h}$ in a multiuser environment.

Then, based on this distribution, the ML detection problem is reduced to

$$\widehat{\mathbf{x}}^{MSDD} = \underset{\mathbf{x}\in\mathcal{S}^K,\, x[0]\equiv 1}{\arg\max}\; f_{\mathbf{Y}}^{ML}\left(\mathbf{Y}|\mathbf{x}\right). \tag{6.34}$$

In [22], the author derives a so-called *truly noncoherent* ML receiver from (6.34) in the context of IR-UWB. By assuming uncorrelated elements in $\mathbf{h}$, *i.e.*

$$\mathbf{C_h} = \begin{bmatrix} P_{h,1} & & 0 \\ & \ddots & \\ 0 & & P_{h,N_{BS}} \end{bmatrix}, \tag{6.35}$$

a simplified detector is obtained, which works on a best guess when no information about $\mathbf{h}$ is known. In our thesis, however, we are dealing with a LAA environment, whose associated channel has a clearly defined structure. Hence, we want to characterize (6.34) for our working conditions.

In the following derivation, we will generalize the results from [22] to any channel vector that is distributed as $\mathbf{h}\sim\mathcal{CN}\left(\mathbf{0},\mathbf{C_h}\right)$, for any arbitrary covariance matrix. Then, in further applications of this scheme (see **Section 7**), the covariance matrix can be set as required by the problem.

The integral in (6.33) can be expanded as

$$\begin{aligned} f_{\mathbf{Y}}^{ML}\left(\mathbf{Y}|\mathbf{x}\right) &= \overbrace{\frac{1}{\pi^{N_{BS}(K+1)}\left|\mathbf{C_h}\right|P_z^{N_{BS}K}}}^{C} \int_{\mathbb{C}^{N_{BS}}} \exp\left(-\mathbf{h}^H{\mathbf{C_h}}^{-1}\mathbf{h}\right)\exp\left(\frac{-1}{P_z}\left\|\mathbf{Y}-\mathbf{h}\mathbf{x}^T\right\|_F^2\right)d\mathbf{h}d\mathbf{h}^* \\ &= C\int_{\mathbb{C}^{N_{BS}}} \exp\left(-\mathbf{h}^H{\mathbf{C_h}}^{-1}\mathbf{h}-\frac{1}{P_z}\left\|\mathbf{Y}-\mathbf{h}\mathbf{x}^T\right\|_F^2\right)d\mathbf{h}d\mathbf{h}^*. \end{aligned} \tag{6.36}$$

We have gathered all the multiplicative factors that do not affect the detection under constant C. We can now develop the exponent inside the integral into a quadratic form of $\mathbf{h}$:

$$\begin{aligned} f_{\mathbf{Y}}^{ML}\left(\mathbf{Y}|\mathbf{x}\right) &= C\int_{\mathbb{C}^{N_{BS}}} \exp\left(-\mathbf{h}^H{\mathbf{C_h}}^{-1}\mathbf{h}-\frac{\left\|\mathbf{Y}\right\|_F^2-\mathbf{x}^T\mathbf{Y}^H\mathbf{h}-\mathbf{h}^H\mathbf{Y}\mathbf{x}^*+\left|\mathbf{x}\right|^2\left|\mathbf{h}\right|^2}{P_z}\right)d\mathbf{h}d\mathbf{h}^* \\ &= \overbrace{Ce^{-\frac{\left\|\mathbf{Y}\right\|_F^2}{P_z}}}^{C'(\mathbf{Y})}\int_{\mathbb{C}^{N_{BS}}} \exp\left(-\mathbf{h}^H{\mathbf{C_h}}^{-1}\mathbf{h}-\frac{K\left|\mathbf{h}\right|^2-\mathbf{x}^T\mathbf{Y}^H\mathbf{h}-\mathbf{h}^H\mathbf{Y}\mathbf{x}^*}{P_z}\right)d\mathbf{h}d\mathbf{h}^* \\ &= C'(\mathbf{Y})\int_{\mathbb{C}^{N_{BS}}} \exp\left(-\mathbf{h}^H\left({\mathbf{C_h}}^{-1}+\frac{K}{P_z}\mathbf{I}_{N_{BS}}\right)\mathbf{h}+\frac{1}{P_z}\mathbf{x}^T\mathbf{Y}^H\mathbf{h}+\frac{1}{P_z}\mathbf{h}^H\mathbf{Y}\mathbf{x}^*\right)d\mathbf{h}d\mathbf{h}^*, \end{aligned} \tag{6.37}$$

where we have used the fact that the elements of $\mathbf{x}$ are unitary (and thus $\left|\mathbf{x}\right|^2=K$) and the equivalence $\left|\mathbf{h}\right|^2\equiv\mathbf{h}^H\mathbf{I}_{N_{BS}}\mathbf{h}$.

We can apply the change of variable $d\mathbf{h}d\mathbf{h}^*\equiv 2^{N_{BS}}d\,\mathrm{Re}\left\{\mathbf{h}\right\}d\,\mathrm{Im}\left\{\mathbf{h}\right\}$ [49] and reformulate the previous expression as

$$f_{\mathbf{Y}}^{ML}\left(\mathbf{Y}|\mathbf{x}\right) = 2^{N_{BS}}C'\int_{\mathbb{C}^{N_{BS}}} \exp\left(-\mathbf{h}^H\left({\mathbf{C_h}}^{-1}+\frac{K}{P_z}\mathbf{I}_{N_{BS}}\right)\mathbf{h}+\frac{\mathbf{x}^T\mathbf{Y}^H\mathbf{h}}{P_z}+\frac{\mathbf{h}^H\mathbf{Y}\mathbf{x}^*}{P_z}\right)d\,\mathrm{Re}\left\{\mathbf{h}\right\}d\,\mathrm{Im}\left\{\mathbf{h}\right\}. \tag{6.38}$$

This integral has a known closed form solution that is widely used in *quantum field theory* (see Equation (2.139) in [49]):

$$\boxed{f_{\mathbf{Y}}^{ML}\left(\mathbf{Y}|\mathbf{x}\right) = \frac{\left(2\pi\right)^{N_{BS}}C'(\mathbf{Y})}{\det\left({\mathbf{C_h}}^{-1}+\frac{K}{P_z}\mathbf{I}_{N_{BS}}\right)}\exp\left(\frac{1}{P_z^2}\mathbf{x}^T\mathbf{Y}^H\left({\mathbf{C_h}}^{-1}+\frac{K}{P_z}\mathbf{I}_{N_{BS}}\right)^{-1}\mathbf{Y}\mathbf{x}^*\right)}. \tag{6.39}$$

The original ML detection problem reduces to the maximization

$$\begin{aligned}\widehat{\mathbf{x}}^{MSDD} &= \underset{\mathbf{x}\in\mathcal{S}^K,\, x[0]\equiv 1}{\arg\max}\; f_{\mathbf{Y}}^{ML}(\mathbf{Y}|\mathbf{x}) = \underset{\mathbf{x}\in\mathcal{S}^K,\, x[0]\equiv 1}{\arg\max}\; \log\left(f_{\mathbf{Y}}^{ML}(\mathbf{Y}|\mathbf{x})\right) \\ &= \underset{\mathbf{x}\in\mathcal{S}^K,\, x[0]\equiv 1}{\arg\max}\; \mathbf{x}^T\mathbf{Y}^H \underbrace{\left(\mathbf{C_h}^{-1} + \frac{K}{P_z}\mathbf{I}_{N_{BS}}\right)^{-1}}_{\mathbf{W}_{opt}} \mathbf{Y}\mathbf{x}^* \\ &= \boxed{\underset{\mathbf{x}\in\mathcal{S}^K,\, x[0]\equiv 1}{\arg\max}\; \mathbf{x}^T\mathbf{Y}^H\mathbf{W}_{opt}\mathbf{Y}\mathbf{x}^*}. \end{aligned} \tag{6.40}$$

It is noteworthy that the generalized noncoherent detector we have obtained with the optimized weighting matrix $\mathbf{W}_{opt}$ displays a structure very reminiscent of the classical *estimator-correlator* [50]. Unsurprisingly, it is also very similar to the one presented in [22]: the only difference appears to be in the weights matrix $\mathbf{W}_{opt}$. In the mentioned thesis, $\mathbf{W}_d$ is a diagonal matrix with elements

$$[\mathbf{W}_d]_{n,n} = \frac{1}{K + \frac{P_z}{P_{h,n}}}, \tag{6.41}$$

stemmed from the assumed uncorrelatedness of $\mathbf{h}$. On the contrary, our detector allows for any kind of covariance matrices, with higher complexity and structure. This implies that it can better exploit the statistical CSI of second order. Indeed, our detector is a generalization of the one presented in [22], which can clearly be seen from the fact that $\mathbf{W}_{opt}$ becomes $\mathbf{W}_d$ under the diagonal covariance matrix assumption:

$$\begin{aligned}\mathbf{W}_{opt}\left(\operatorname{diag}\left(P_{h,1},\dots,P_{h,N_{BS}}\right)\right) &= \left(\operatorname{diag}\left(P_{h,1},\dots,P_{h,N_{BS}}\right)^{-1} + \frac{K}{P_z}\mathbf{I}_{N_{BS}}\right)^{-1} \\ &= \operatorname{diag}\left(P_{h,1}^{-1} + \frac{K}{P_z},\dots,P_{h,N_{BS}}^{-1} + \frac{K}{P_z}\right)^{-1} \\ &= \operatorname{diag}\left(\frac{1}{\frac{K}{P_z} + \frac{1}{P_{h,1}}},\dots,\frac{1}{\frac{K}{P_z} + \frac{1}{P_{h,N_{BS}}}}\right) \\ &= \begin{bmatrix} \frac{P_z}{K+\frac{P_z}{P_{h,1}}} & & 0 \\ & \ddots & \\ 0 & & \frac{P_z}{K+\frac{P_z}{P_{h,N_{BS}}}} \end{bmatrix}. \end{aligned} \tag{6.42}$$

The factor P_z comes from the exponent in (6.39), and $P_{h,n}$ are the diagonal terms of the full channel covariance matrix, $[\mathbf{C_h}]_{n,n}$.

In general, $\mathbf{C_h}$ is not known at the BS, so the detector has to use an estimation $\widehat{\mathbf{C}}_\mathbf{h}$. Since this matrix may be badly conditioned, its inversion would entail numerical inaccuracies. Fortunately, we can apply *Woodbury's Inversion Lemma* to the computation of $\mathbf{W}_{opt}$:

$$\mathbf{W}_{opt} = \left(\widehat{\mathbf{C}}_\mathbf{h}^{-1} + \frac{K}{P_z}\mathbf{I}_{N_{BS}}\right)^{-1} = \widehat{\mathbf{C}}_\mathbf{h} - \widehat{\mathbf{C}}_\mathbf{h}\left(\frac{P_z}{K}\mathbf{I}_{N_{BS}} + \widehat{\mathbf{C}}_\mathbf{h}\right)^{-1}\widehat{\mathbf{C}}_\mathbf{h}. \tag{6.43}$$

Now the term in the parenthesis is easily invertible, thanks to the diagonal loading that acts as a sort of *Tikhonov regularization* [51].

MSDD is based on the autocorrelation of a signal block. Considering (6.40), we define matrix

$$\begin{aligned}\mathbf{R} &\triangleq \mathbf{Y}^H\mathbf{W}_{opt}\mathbf{Y} \\ r_{k,k'} &\triangleq [\mathbf{R}]_{k,k'} = \mathbf{y}^H[k]\mathbf{R}\mathbf{y}[k'].\end{aligned} \tag{6.44}$$

From this definition we can understand $r_{k,k'}$ as the weighted correlation coefficient between received signal $\mathbf{y}[k]$ and $\mathbf{y}[k']$. We can express the noncoherent detection rule in a double summation form:

$$\widehat{\mathbf{x}}^{MSDD} = \underset{\mathbf{x}\in\mathcal{S}^K,\, x[0]\equiv 1}{\arg\max}\ \mathbf{x}^T\mathbf{R}\mathbf{x}^* = \underset{\mathbf{x}\in\mathcal{S}^K,\, x[0]\equiv 1}{\arg\max} \sum_{k=0}^{K-1}\sum_{k'=0}^{K-1} r_{k,k'}x[k]x^*[k']. \tag{6.45}$$

For $k=k'$, $x[k]x^*[k] = |x[k]|^2 = 1$, so we can remove these terms from the summation. Furthermore, since $\mathbf{W}_{opt}$ is a symmetric matrix, we only need to consider the correlation terms from one of its two triangular halves:

$$\boxed{\widehat{\mathbf{x}}^{MSDD} = \underset{\mathbf{x}\in\mathcal{S}^K,\, x[0]\equiv 1}{\arg\max} \sum_{k=1}^{K-1} x^*[k] \sum_{k'=0}^{k-1} x[k']r_{k',k}} \tag{6.46}$$

Before introducing a noncoherent scheme for the DL in the following section, let us briefly comment on some numerical results regarding the use of the optimized weights matrix. Consider a LAA P2P UL scenario (model (6.26)), with $\tilde{D}=50$ ($N_{BS}=101$) and $L=3$. A MT transmits a signal block of length $K_{UL}=401$ with power $P_{MT}=1$. We assume the channel $\mathbf{h}$ remains invariant during the full transmission. The SNR perceived by the Rx can be defined as

$$\text{SNR} \triangleq \frac{\mathbb{E}\left[\|\mathbf{h}\|^2\right]}{N_{BS}P_z}. \tag{6.47}$$

To perform MSDD, the BS must estimate the covariance matrix of the channel, $\mathbf{C}_{\mathbf{h}}$, from the received signal $\mathbf{y}[k]$. Its correlation matrix is computed as

$$\begin{aligned}\mathbf{C}_{\mathbf{y}} \triangleq \mathbb{E}\left[\mathbf{y}[k]\mathbf{y}^H[k]\right] &= \mathbb{E}\left[\left(\mathbf{h}x[k]+\mathbf{z}[k]\right)\left(\mathbf{h}x[k]+\mathbf{z}[k]\right)^H\right] \\ &= \mathbb{E}\left[\mathbf{h}x[k]x^*[k]\mathbf{h}^H + \mathbf{h}x[k]\mathbf{z}^H[k] + \mathbf{z}[k]x^*[k]\mathbf{h}^H + \mathbf{z}[k]\mathbf{z}^H[k]\right] \\ &= \mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right] + \mathbb{E}\left[\mathbf{z}[k]\mathbf{z}^H[k]\right] = \mathbf{C}_{\mathbf{h}} + P_z\mathbf{I}.\end{aligned} \tag{6.48}$$

We have used the fact that the AWGN $\mathbf{z}[k] \sim \mathcal{CN}\left(\mathbf{0}, P_z\mathbf{I}\right)$ is uncorrelated with the transmitted signal and the channel, and $x[k]$ is unitary ($|x[k]|^2=1$), since it is drawn from an M-PSK constellation. From this result we conclude that $\mathbf{C}_{\mathbf{h}}$ can be obtained by estimating $\mathbf{C}_{\mathbf{y}}$ and removing the noise power contribution. An appropriate option to do such estimation is the *unbiased sample covariance* from the received block:

$$\widehat{\mathbf{C}}_{\mathbf{y}} \triangleq \frac{1}{K_{UL}}\mathbf{Y}\mathbf{Y}^H. \tag{6.49}$$

Then, we subtract the noise load to obtain an estimate of the channel covariance matrix:

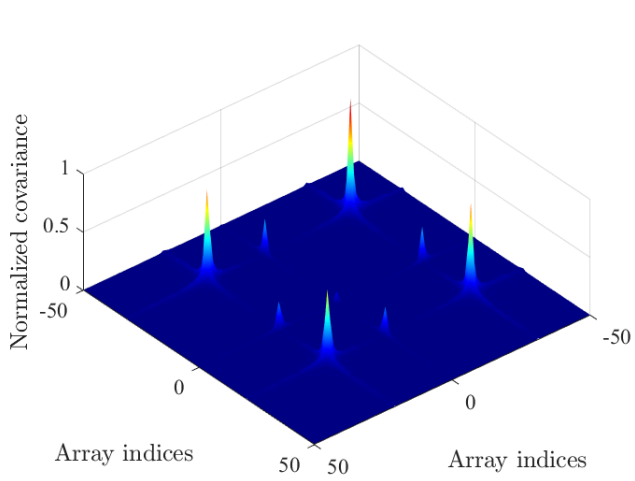
$$\widehat{\mathbf{C}}_{\mathbf{h}} \triangleq \widehat{\mathbf{C}}_{\mathbf{y}} - \hat{P}_z\mathbf{I}. \tag{6.50}$$

$\widehat{P}_z$ is an estimate of the noise power, obtained from antennas moderately separated (in the angular sense) from the signal AoAs.

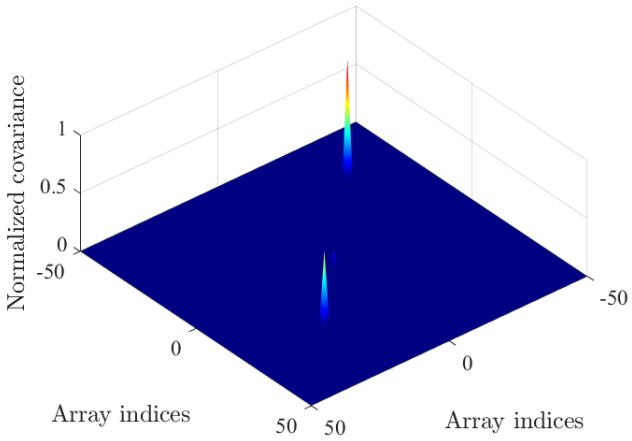
Let us make some comparisons between our generalized covariance matrix method and the truly noncoherent method from [22] and its subsequent line of research. A qualitative illustration of $\widehat{\mathbf{C}}_{\mathbf{h}}$ computed from the previous received signal block $\mathbf{Y}$ at SNR = 0dB can be seen in Figure 6.1a. Using only its diagonal elements (Figure 6.1b) to produce the weights matrix discards an important amount

of signal power and information from the crossed path terms. This loss can be observed in Figures 6.1c and 6.1d, where the latter has substantially less active elements than the former.

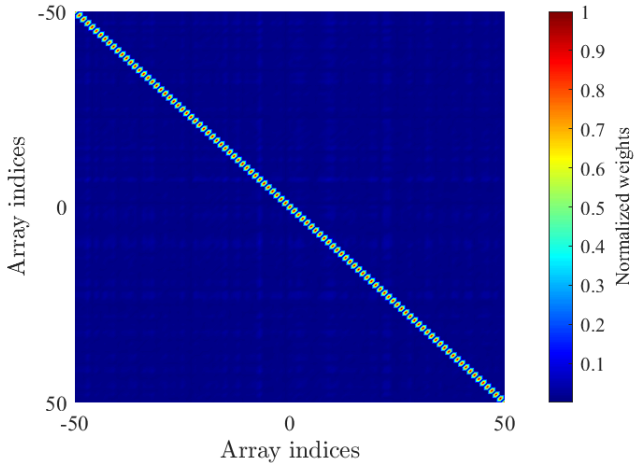
In practice, this results in the expected effect. After averaging over 3000 UL transmissions at various SNRs, we have obtained the Symbol Error Rate (SER) displayed in Figure 6.2. It is then clear that better exploiting the available statistical CSI produces a more reliable noncoherent system. With this proposal, we have successfully generalized and improved upon a well-known MSDD scheme for a P2P UL.



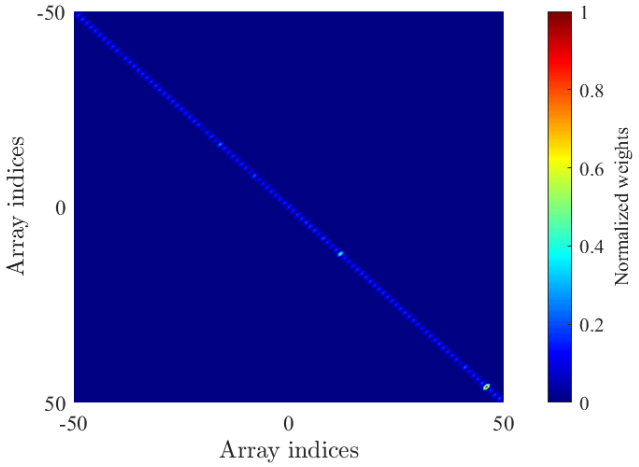
(a) Full estimated channel covariance matrix.



(b) Estimated channel covariance under the uncorrelatedness hypothesis.



(c) Weights matrix computed with the full covariance matrix.



(d) Weights matrix computed with the diagonal covariance matrix.

Figure 6.1: Comparison between MSDD weighting approaches.

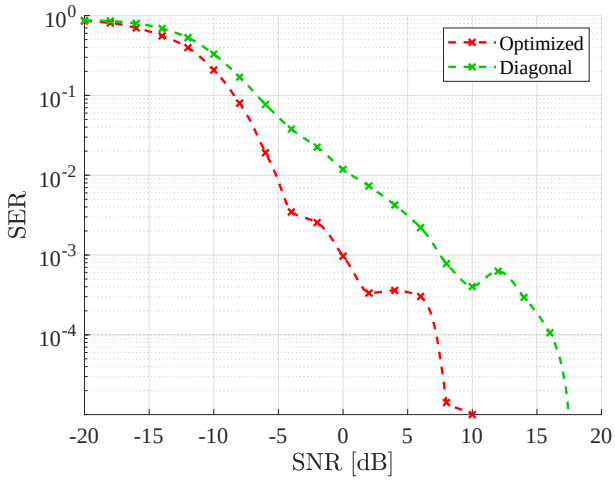


Figure 6.2: SER comparison between optimized and truly noncoherent MSDD methods for various values of SNR.

6.3.2 Differential Orthogonal Space-Time Block Coding (D-OSTBC)

Traditional space-time coding schemes remove the CSI acquisition task from Txs and rely instead on full CSIR. They are specially suitable for slowly fading channels and Rxs with enough resources (energy, computational power) to devote to signal processing, since the channel matrix coefficients must be estimated accurately. An UL would be an appropriate scenario). This approach, however may not work in fast fading channels or schemes with battery-constrained receivers. A noncoherent approach seems like the more natural alternative.

Differential Space-Time Modulation (DSTM) is the multiantenna extension of DM for SISO systems (see **Section 2.3**). Instead of dealing with scalars, we now manipulate space-time block matrices, in which rows ($\downarrow$) represent the spatial dimension (Tx antennas) and columns ($\rightarrow$) are different instances of time. Consider a MISO DL[6] setting in which a LAA BS with N_{BS} antennas transmits data to a single-antenna MT ($N_{MT} = 1$). The space-time model of the received signal is

$$\begin{aligned} \breve{\mathbf{y}}^T &= \mathbf{h}^H \breve{\mathbf{X}} + \breve{\mathbf{z}}^T \in \mathbb{C}^{1\times K} \\ [y[0], \dots, y[K-1]] &= \mathbf{h}^H \left[\mathbf{x}[0], \dots, \mathbf{x}[K-1]\right] + [z[0], \dots, z[K-1]] \\ \left[\mathbf{y}^T[0], \dots, \mathbf{y}^T[K-K_b]\right] &= \mathbf{h}^H \left[\mathbf{X}[0], \dots, \mathbf{X}[K-K_b]\right] + \left[\mathbf{z}^T[0], \dots, \mathbf{z}[K-K_b]\right], \end{aligned} \tag{6.51}$$

where $\breve{\mathbf{y}}^T$ contains the full sequence of $\{y[k]\}$ received signals, $\mathbf{h}^H \in \mathbb{C}^{1\times N_{BS}}$ is the reciprocal channel vector from (6.26), $\breve{\mathbf{X}} \in \mathbb{C}^{N_{BS}\times K}$ contains the K transmitted signals $\mathbf{x}[k] \in \mathbb{C}^{N_{BS}\times 1}$, for $k = 0, \dots, K-1$, and $\breve{\mathbf{z}}^T[k] \in \mathbb{C}^{1\times N_{BS}}$ is additive noise and interference. $\mathbf{x}[k]$ are constructed as $\mathbf{x}[k] = \mathbf{G}[k]\mathbf{c}[k]$, such that $\mathbf{G}[k] \in \mathbb{C}^{N_{BS}\times N_s}$ is a power-loading/beamforming matrix and $\mathbf{c}[k]$ are the transmitted symbols of dimension N_s, a parameter we will refer to as its *spatial size* [29]. These symbols are coded into space-time blocks of length K_b, $\mathbf{C}[i] \triangleq [\mathbf{c}[i], \mathbf{c}[i+1], \dots, \mathbf{c}[i+K_b-1]]$. The same occurs with the beamformed signals $\mathbf{X}[i] \triangleq \mathbf{G}\mathbf{C}[i] = [\mathbf{x}[i], \mathbf{x}[i+1], \dots, \mathbf{x}[i+K_b-1]]$.

A full DL transmission of duration K is composed of $N_b \triangleq \lfloor \frac{K}{K_b} \rfloor$ signal blocks[7]. The block-wise model can be expressed as

$$\mathbf{y}^T[i] = \mathbf{h}^H \mathbf{X}[i] + \mathbf{z}^T[i] \quad , \quad i = 0, K_b, 2K_b, \dots, (N_b-1)\,K_b. \tag{6.52}$$

[6] The MIMO extension of this scheme is straightforward and can be found in [31].
[7] We assume K is a multiple of K_b for convenience.

For the moment, we assume the channel remains constant over the duration of at least two blocks.

Let $\mathcal{S}_{\mathcal{U}} \triangleq \{\mathbf{S}[i]\}$ be a set of unitary matrices[8] to be transmitted. Extending the ideas of **Section 2.3**, instead of transmitting $\mathbf{S}[i]$ directly, we encode the information differentially on a new set $\{\mathbf{C}[i]\}$:

$$\begin{aligned} \mathbf{C}[i] &= \mathbf{C}[i-1]\mathbf{S}[i] \\ \mathbf{C}[0] &\equiv \mathbf{I}_{N_s}. \end{aligned} \tag{6.53}$$

Since $\{\mathbf{S}[i]\}$ are unitary, we can prove that $\mathbf{C}^H[i]\mathbf{C}[i] = \mathbf{I}_{N_s}$ by following the recursion

$$\begin{aligned} \mathbf{C}^H[i]\mathbf{C}[i] &= \mathbf{S}^H[i]\mathbf{C}^H[i-1]\mathbf{C}[i-1]\mathbf{S}[i] \\ &= \mathbf{S}^H[i]\mathbf{S}^H[i-1]\dots\mathbf{S}^H[1]\overbrace{\mathbf{C}^H[0]\mathbf{C}[0]}^{\mathbf{I}_{N_s}}\mathbf{S}[1]\dots\mathbf{S}[i] \\ &= \mathbf{S}^H[i]\dots\overbrace{\mathbf{S}^H[1]\mathbf{S}[1]}^{\mathbf{I}_{N_s}}\dots\mathbf{S}[i] \\ &= \mathbf{I}_{N_s}. \end{aligned} \tag{6.54}$$

This property guarantees that the transmit power is kept constant if the beamforming matrix is properly normalized. By transmitting $\mathbf{X}[i-1]$ and $\mathbf{X}[i]$, the received signal blocks are

$$\mathbf{y}^T[i-1] = \mathbf{h}^H\mathbf{X}[i-1] + \mathbf{z}^T[i] \tag{6.55}$$

$$\begin{aligned} \mathbf{y}^T[i] &= \mathbf{h}^H\mathbf{X}[i] + \mathbf{z}^T[i] \\ &= \mathbf{h}^H\mathbf{G}[k]\mathbf{C}[i] + \mathbf{z}^T[i] \\ &= \mathbf{h}^H\mathbf{G}[k]\mathbf{C}[i-1]\mathbf{S}[i] + \mathbf{z}^T[i] \\ &= \mathbf{h}^H\mathbf{X}[i-1]\mathbf{S}[i] + \mathbf{z}^T[i]. \end{aligned} \tag{6.56}$$

If $\mathbf{h}^H\mathbf{X}[i-1]$ was known, we could simply use the coherent ML detector

$$\widehat{\mathbf{S}}_{ML} \triangleq \underset{\mathbf{S}[i]\in\mathcal{S}_{\mathcal{U}}}{\arg\min} \left\| \mathbf{y}^T[i] - \mathbf{h}^H\mathbf{X}[i-1]\mathbf{S}[i] \right\|^2. \tag{6.57}$$

Since this is not the case, we will decode $\mathbf{S}[i]$ using two consecutive blocks $\mathbf{y}^T[i]$ and $\mathbf{y}^T[i-1]$:

$$\begin{aligned} \widehat{\mathbf{S}}_{ML}[i] &= \underset{\mathbf{S}[i]\in\mathcal{S}_{\mathcal{U}}}{\arg\min} \left\| \mathbf{y}^T[i] - \mathbf{h}^H\mathbf{X}[i-1]\mathbf{S}[i] \right\|^2 + \left\| \mathbf{y}^H[i-1] - \mathbf{h}^H\mathbf{X}[i-1] \right\|^2 \\ &= \underset{\mathbf{S}[i]\in\mathcal{S}_{\mathcal{U}}}{\arg\min} \left\| \mathbf{y}^T[i]\mathbf{S}^H[i] - \mathbf{h}^H\mathbf{X}[i-1] \right\|^2 + \left\| \mathbf{y}^H[i-1] - \mathbf{h}^H\mathbf{X}[i-1] \right\|^2. \end{aligned} \tag{6.58}$$

The vector $\mathbf{h}^H\mathbf{X}[i-1]$ is completely unknown, so we use the value that minimizes (6.58):

$$\widehat{\mathbf{h}^H\mathbf{X}}[i-1] = \frac{1}{2}\left(\mathbf{y}^T[i]\mathbf{S}^H[i] + \mathbf{y}^T[i-1]\right). \tag{6.59}$$

We replace it in (6.58) and obtain the simplified minimization problem:

$$\begin{aligned} \widehat{\mathbf{S}}_{ML}[i] &= \underset{\mathbf{S}[i]\in\mathcal{S}_{\mathcal{U}}}{\arg\min} \left\| \mathbf{y}^T[i]\mathbf{S}^H[i] - \mathbf{y}^T[i-1] \right\|^2 \\ &= \underset{\mathbf{S}[i]\in\mathcal{S}_{\mathcal{U}}}{\arg\min} \left(\|\mathbf{y}[i]\|^2 - \mathbf{y}^T[i]\mathbf{S}^H[i]\mathbf{y}^*[i-1] - \mathbf{y}^T[i-1]\mathbf{S}[i]\mathbf{y}^*[i] + \|\mathbf{y}[i-1]\|^2 \right) \\ &= \underset{\mathbf{S}[i]\in\mathcal{S}_{\mathcal{U}}}{\arg\min} -2\,\mathrm{Re}\left\{ \mathbf{y}^T[i-1]\mathbf{S}[i]\mathbf{y}^*[i] \right\} \\ &= \boxed{\underset{\mathbf{S}[i]\in\mathcal{S}_{\mathcal{U}}}{\arg\max}\, \mathrm{Re}\left\{ \mathbf{y}^T[i-1]\mathbf{S}[i]\mathbf{y}^*[i] \right\}}. \end{aligned} \tag{6.60}$$

[8] This constrains the temporal size of the block to be equal to its spatial size, *i.e.* $K_b \equiv N_s$.

Notice that both (6.57) and (6.60) ML detectors are the natural space-time extension of the coherent and noncoherent detectors introduced in **Section 2.3**.

For an arbitrary set of unitary code matrices $\{\mathbf{S}[i]\}$, the ML detector (6.60) becomes computationally expensive, since it requires a search through all possible $\mathbf{S}[i]$. Instead, using Orthogonal Space-Time Block Coding (OSTBC) rather than group codes, the complexity can be remarkably reduced and an overall better performance can be achieved [31].

Differential encoding of matrix blocks requires square and unitary matrices. However, code matrices for OSTBC are semi-unitary[9] and not square in general. The only instances in which they are square are for $K_b = 2, 4, 8$. They can be constructed using the following linear model:

$$\mathbf{S}[i] \triangleq \sqrt{\frac{1}{M_s}} \sum_{m=1}^{M_s} \left(\mathrm{Re}\left\{ s_m[i] \right\} \mathbf{A}_m + j \, \mathrm{Im}\left\{ s_m[i] \right\} \mathbf{B}_m \right), \tag{6.61}$$

where $\{s_1[i], \ldots, s_{M_s}[i]\}$ is a set of complex symbols taken from a unitary constellation $\mathcal{S}$ to be transmitted in a space-time block and $\{\mathbf{A}_m, \mathbf{B}_m\}_{m=1}^{M_s} \in \mathbb{C}^{N_s \times K_b}$ are fixed code matrices. This scheme allows us to send M_s symbols over K_b time intervals, hence its rate is defined as $R_b \triangleq \frac{M_s}{K_b}$. Symbol blocks of sizes 2, 4 and 8 have rates of 1, 3/4 and 1/2, respectively. The scheme can be extended to larger blocks but their rate becomes so low ($<1/2$) that are of little interest in practice.

For (6.61) to be *orthogonal*, $\{\mathbf{A}_m, \mathbf{B}_m\}$ must be an *amicable orthogonal design* [31], which is a special set of matrices whose structure follows some defined constraints. These ensure the following unitary property:

$$\mathbf{S}[i]\mathbf{S}^H[i] = \frac{1}{M_s} \sum_{m=1}^{M_s} |s_m|^2 \, \mathbf{I}_{K_b} = \frac{1}{M_s} \sum_{m=1}^{M_s} \mathbf{I}_{K_b} = \mathbf{I}_{K_b}. \tag{6.62}$$

By constructing space-time code blocks with these designs, OSTBC guarantees that the ML detection of different symbols in a block is decoupled. The ML detector (6.60) simplifies considerably compared to searching for general unitary code matrices. Its final expression is derived as follows:

$$\begin{aligned}
\widehat{\mathbf{S}}_{ML}[i] &= \underset{\mathbf{S}[i] \in \mathcal{S}_{\mathcal{U}}}{\arg\max} \, \mathrm{Re}\left\{ \mathbf{y}^T[i-1] \mathbf{S}[i] \mathbf{y}^*[i] \right\} \\
&= \underset{\mathbf{s}[i] \in \mathcal{S}}{\arg\max} \, \mathrm{Re}\left\{ \sqrt{\frac{1}{M_s}} \mathbf{y}^T[i-1] \left(\sum_{m=1}^{M_s} \left(\mathrm{Re}\left\{ s_m[i] \right\} \mathbf{A}_m + j \, \mathrm{Im}\left\{ s_m[i] \right\} \mathbf{B}_m \right) \right) \mathbf{y}^*[i] \right\} \\
&= \underset{\mathbf{s}[i] \in \mathcal{S}}{\arg\max} \, \mathrm{Re}\left\{ \sum_{m=1}^{M_s} \left(\mathrm{Re}\left\{ s_m[i] \right\} \mathbf{y}^T[i-1] \mathbf{A}_m \mathbf{y}^*[i] + j \, \mathrm{Im}\left\{ s_m[i] \right\} \mathbf{y}^T[i-1] \mathbf{B}_m \mathbf{y}^*[i] \right) \right\} \\
&= \boxed{\underset{\mathbf{s}[i] \in \mathcal{S}}{\arg\max} \sum_{m=1}^{M_s} \left(\mathrm{Re}\left\{ s_m[i] \right\} \mathrm{Re}\left\{ \mathbf{y}^T[i-1] \mathbf{A}_m \mathbf{y}^*[i] \right\} + \mathrm{Im}\left\{ s_m[i] \right\} \mathrm{Im}\left\{ \mathbf{y}^T[i-1] \mathbf{B}_m \mathbf{y}^*[i] \right\} \right)}.
\end{aligned} \tag{6.63}$$

The main practical issue regarding D-OSTBC scheme is the structural rigidity of the code matrices which can only be square for certain N_s. In many systems and scenarios, the conditions may not be optimal for any block size. Particularly, under the LAA-MISO channel model studied in this section, there may be L reliable signal paths between the BS, and we may want to assign a different spatial dimension on each one of them (*i.e. spatial multiplexing*). Since L is arbitrarily given by the geometry of the communication setting, it may not be suitable for any D-OSTBC design. This would

[9] A matrix $\mathbf{M}$ is semi-unitary when either $\mathbf{M}^H\mathbf{M} = \mathbf{I}$ or $\mathbf{M}\mathbf{M}^H = \mathbf{I}$.

force the scheme to employ fewer paths than the available ones to accommodate a block code, hence underutilizing the full channel diversity.

To circumvent this inconvenience, we can apply a simple modification to the basic scheme that will make it more adaptable to channel characteristics. Given a differentially encoded matrix $\mathbf{C}[i] \in \mathbb{C}^{N_s \times K_b}$, we define the *trimming matrix*

$$\mathbf{\Phi}_L \triangleq \left[\mathbf{I}_L \, \mathbf{0}_{L\times(K_b-L)}\right], \tag{6.64}$$

in which K_b is 2, 4 or 8 accordingly, and transmit $\mathbf{\Phi}_L \mathbf{C}_{K_b}[i]$ instead of $\mathbf{C}_{K_b}[i]$ (the subscript K_b indicates the size of the unitary matrix). This is equivalent to only transmitting the first L rows of $\mathbf{C}_{K_b}[i]$:

$$\mathbf{y}^T[i] = \mathbf{h}^H \mathbf{G}[k] \mathbf{\Phi}_L \mathbf{C}_{K_b}[i] + \mathbf{z}^T[i]. \tag{6.65}$$

The receiver can use the regular ML detector with $\mathbf{S}[i]$ defined for $N_s = K_b$. With this scheme, we can successfully exploit the LAA channel for $N_s = 2, \ldots, 8$. While we could easily extend it to larger spatial dimensions, mmWave channels are characterized by their multipath sparsity, hence $L > 8$ is not common in most environments [5].

D-OSTBC presents the same diversity order ($G_d = \text{rank}\left\{\mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right]\right\} \propto L$) as its coherent counterpart. The only difference between their performances is the expected 3dB loss in SNR. Indeed, for coherent detection, the SNR is

$$\text{SNR}_{coh} = \frac{K_b}{M_s N_{BS}} \frac{\|\mathbf{h}\|^2}{P_z} = \frac{\|\mathbf{h}\|^2}{R_b N_{BS} P_z}, \tag{6.66}$$

for noise power P_z, assuming total transmitted energy equal to one during each time instant and ignoring the impact of $\mathbf{G}[k]$. Notice that there is a gain inversely proportional to the block rate in SNR. Then, in the noncoherent scenario, it can be proven [31] that

$$\text{SNR}_{non} = \frac{\|\mathbf{h}\|^2}{2 R_b N_{BS} P_z} = \frac{\text{SNR}_{coh}}{2}, \tag{6.67}$$

due to the effect of using $\mathbf{y}^T[i-1]$ instead of $\mathbf{h}^H$ in the ML detector, which doubles the noise power.

Chapter 7

Multi-user Communications with Lens Antenna Arrays

In the final chapter of this thesis we are going to study the problem of LAA communications in a multi-user environment. All the tools we have presented so far will be useful in this setting, but the presence of both noise and Inter-User Interference (IUI) will require the development of new methods with a more in-depth analysis. We want to garner all the benefits that LAAs deliver for mmWave wireless communications and combine them with the strengths of noncoherent schemes.

The setting we are going to study in particular consists in a group of N_U users in the vicinity of a BS that want to establish simultaneous communication with it. Each user has a single-antenna MT and the BS is equipped with a LAA of sufficiently large size in order to give service to all N_U users at once. Therefore, we will employ a generalized multi-user version of the SIMO and MISO models discussed in the previous chapter: MU-MIMO [27]. Since MTs are powered by batteries, they are constrained in terms of energy consumption, which translates into limited computational power. For this reason, we want the BS to carry out the CSI acquisition for both the UL and DL, thus removing most of the signal processing from the users' terminals.

We have designed a full UL-DL scheme that exploits the advantages of using very large LAA with partial CSI at the BS. Not only have we considered the communication side of the system, but also CSI acquisition, antenna selection and other signal processing aspects that are crucial for the scheme to be robust and successful. We have organized the exposition in two sections: one for the UL and one for the DL. Each one of them has been divided in various subsections that contain encapsulated techniques that, once combined, form a synchronized whole. After these explanations, the performance of the resulting scheme will be tested in various numerical simulations.

7.1 Uplink

In the UL stage, N_U MT users (Tx) start a communication with the BS (Rx). This means that a set of N_U SIMO systems have to be processed simultaneously (MU-MIMO). Its associated signal model is the result of combining those N_U SIMO ones:

$$\begin{aligned}\mathbf{Y} = [\mathbf{y}[0], \dots, \mathbf{y}[K_{UL}-1]] &= \sum_{u=1}^{N_U} \mathbf{h}_u \mathbf{x}_u^T + \mathbf{Z} \in \mathbb{C}^{N_{BS} \times K_{UL}} \\ &= [\mathbf{h}_1, \dots, \mathbf{h}_{N_U}] \left[\mathbf{x}_1, \dots, \mathbf{x}_{N_U}\right]^T + \mathbf{Z} \\ &= \boxed{\mathbf{H}\mathbf{X}^T + \mathbf{Z}}. \end{aligned} \tag{7.1}$$

Each row of the signal block $\mathbf{X}^T$, $\mathbf{x}_u^T$, corresponds to a different user u that transmits K_{UL} symbols from an M-ary constellation $\mathcal{S}$. Each column of matrix $\mathbf{H}$, $\mathbf{h}_u$, is the channel associated to user u, which takes the form

$$\mathbf{h}_u \triangleq \tilde{D} \sum_{l \in \mathcal{L}_u} \alpha_l^{UL} \operatorname{sinc}\left(\mathbf{n} - \left(n_l^{(u)} + \epsilon_l^{(u)}\right)\mathbf{1}\right), \tag{7.2}$$

where $\mathcal{L}_u$ is the set of channel paths connecting Tx u with the Rx. Vector $\mathbf{n} = \left[-\frac{N_{BS}-1}{2}, \dots, \frac{N_{BS}-1}{2}\right]^T$ contains the BS antenna indices. Once again, we assume Rayleigh fading as in **Section 6.3.1** ($\alpha_l^{UL} \sim \mathcal{CN}(0, P_l)$). The path gains can then be expressed as

$$\alpha_l^{UL} \triangleq \left|\alpha_l^{UL}\right| e^{j\omega_l^{UL}} \quad \Rightarrow \quad \left|\alpha_l^{UL}\right|^2 \sim \text{Exp}\left(\frac{1}{P_l}\right) \quad , \quad \omega_l^{UL} \sim \mathcal{U}[0, 2\pi)\,. \tag{7.3}$$

Finally, $\mathbf{Z} \sim \mathcal{MCN}(\mathbf{0}, P_z\mathbf{I}_{N_R}, \mathbf{I}_{K_{UL}})$ is AWGN.

7.1.1 Antenna Selection

As explained in **Section 2.2**, very large antenna arrays cannot afford the infrastructure and power requirements to employ all their elements at once, and LAAs are no exception. For this reason, the system designed in this chapter requires the use of antenna selection as well. However, while the energy-based AS criterion discussed in **Section 2.2** is well-suited for P2P communications, multi-user scenarios require a more nuanced approach.

Straightforward energy based antenna selection does not distribute available resources in a fair manner across all the MTs establishing communication with the BS. While this technique maximizes the available sum rate, it does so by favoring users with strong channel paths and neglecting others with worse SNR conditions. To circumvent this non-ideal situation, we propose an approach that first divides the available RF chains fairly along the users and then applies energy detection to each one of them individually.

This new antenna selection procedure is described as follows. Each MT transmits training symbols $s_{sel}^{(u)}$ for a duration K_{sel} with power $P_{sel}^{(u)}$:

$$x_{sel}^{(u)}[k] \triangleq \sqrt{P_{sel}^{(u)}}\, s_{sel}^{(u)}. \tag{7.4}$$

An advantage of not having to explicitly estimate the channel coefficients is the fact that the training symbols are not required to be orthogonal. As mentioned in **Section 2.2**, since the amount of RF chains available at the BS is lower than the number of Rx antennas ($N_{RF} < N_{BS}$), it has to perform the AS phase by scanning across all its array elements, which results in an effective training time of

$$T_{sel} \triangleq \left\lfloor \frac{K_{sel} N_{BS}}{N_{RF}} \right\rfloor. \tag{7.5}$$

The BS receives signals of the form

$$\begin{aligned} \mathbf{y}_{sel}[k] &= \mathbf{H}\mathbf{x}_{sel}[k] + \mathbf{z}[k] \quad , \quad k = 0, \dots, K_{sel} - 1. \\ &= [\mathbf{h}_1, \dots, \mathbf{h}_{N_U}] \left[x_{sel}^{(1)}[k], \dots, x_{sel}^{(N_U)}[k]\right]^T + \mathbf{z}[k] \end{aligned} \tag{7.6}$$

with $\mathbf{H}$ being the same as in (7.1) and $\mathbf{z}[k]$ being complex AWGN ($\mathbf{z}[k] \sim \mathcal{CN}(\mathbf{0}, P_z\mathbf{I})$). Once all the training symbols have been received, the BS combines them into a single vector, in the same manner as in **Section 2.2**:

$$\begin{aligned} \bar{\mathbf{y}}_{sel} &= \frac{1}{\sqrt{T_{sel}}} \sum_{k'=0}^{T_{sel}-1} \mathbf{y}_{sel}[k'] = \frac{1}{\sqrt{T_{sel}}} \sum_{k'=0}^{T_{sel}-1} \left(\mathbf{H}\mathbf{x}_{sel}[k'] + \mathbf{z}[k']\right) \\ &= \frac{1}{\sqrt{T_{sel}}} \sum_{k'=0}^{T_{sel}-1} \sum_{u=1}^{N_U} \mathbf{h}_u x_{sel}^{(u)}[k'] + \overbrace{\frac{1}{\sqrt{T_{sel}}} \sum_{k'=0}^{T_{sel}-1} \mathbf{z}[k']}^{\bar{\mathbf{z}}} \\ &= \sqrt{T_{sel}} \tilde{D} \sum_{u=1}^{N_U} \sqrt{P_{sel}^{(u)}}\, s_{sel}^{(u)} \sum_{l \in \mathcal{L}_u} \alpha_l \,\text{sinc}\left(\mathbf{n} - \tilde{D}\tilde{\phi}_{BS,l}\mathbf{1}\right) + \bar{\mathbf{z}}. \end{aligned} \tag{7.7}$$

Then, to perform an energy based selection, the BS considers

$$\bar{\bar{\mathbf{y}}}_{sel} \triangleq \bar{\mathbf{y}}_{sel} \odot \bar{\mathbf{y}}^*_{sel} = \begin{bmatrix} |\bar{y}_{sel,1}|^2 \\ \vdots \\ |\bar{y}_{sel,N_{BS}}|^2 \end{bmatrix}. \tag{7.8}$$

At this point, our proposal starts to depart from the one in **Section 2.2**. If we closely analyze the structure of $\bar{\bar{\mathbf{y}}}_{sel}$, we can clearly see it presents very prominent energy peaks, thanks to the focusing properties of the LAA channel. Each one of these peaks corresponds to a different signal beam l. In other words, a local maximum in $\bar{\bar{\mathbf{y}}}_{sel}$ corresponds to the antenna closest to the angular direction of a beam. We will denote it as n_l, using a notation similar to (4.24), and $\bar{\bar{y}}_{sel,n_l}$ will be denoted p_l. As stated in **Section 2.2**, antennas whose energy level is under a certain threshold Δ_{sel} will be considered noise and be left out of the selection process.

The task of assigning each identified path l to its corresponding origin MT is a very complex problem beyond the scope of this thesis, especially considering we are dealing with a noncoherent scheme that does not employ orthogonal training pilots. Nonetheless, we propose general brush strokes on how this task could be carried out. After the antenna selection phase[1], each MT could transmit K_{id} identifying symbols. Without the need for orthogonal identifiers, we propose that some sort of signaling is encoded in each sequence autocorrelation in such a way that it uniquely identifies and differentiates its corresponding user from the rest.

Having said that and for the purposes of the following sections, we assume that the BS is fully aware of the number of users in active communication (N_U) and which signal beams correspond to each one of them ($\mathcal{L}_u, \quad \forall u = 1, \dots, N_U$). This information may have been gathered from previous UL stages or by the means we have just suggested.

For a given number of simultaneous users, the BS should distribute its available resources equally among them, by following a *fairness policy.* In particular, for N_{RF} available RF chains, each user should be allocated $N_{RF,u} \triangleq \left\lfloor \frac{N_{RF}}{N_U} \right\rfloor$ of them. Then, given the set of beams coming from user u, $\mathcal{L}_u$, each one of them is assigned an amount of RF chains M_l, proportionally to their associated $p_l^{(u)}$:

$$M_{l,u} \triangleq \left\lfloor \frac{p_l^{(u)}}{\sum_{l \in \mathcal{L}_u} p_l^{(u)}} N_{RF,u} \right\rfloor. \tag{7.9}$$

The RF chains allocated to user u that have not been assigned to any of its corresponding AoAs will be devoted to the one with highest p_l, *i.e.* $\arg\max_{l \in \mathcal{L}_u} p_l$. This approach ensures a fair distribution of the available resources while trying to maximize the received energy per user.

After the RF chains have been assigned to each beam, the BS defines a support of antennas around every estimated AoA ($n_l^{(u)}$). The way in which this task is implemented is displayed in **Appendix A** (1.). In a nutshell, the support increases on both sides of each $n_l^{(u)}$ until all its assigned RF chains have been paired with consecutive antennas or another AoA from a different user has been reached. This condition is applied to reduce the amount of IUI captured by a set of antennas: while some supports may overlap, each $n_l^{(u)}$ will only belong to a single one of them. Refer to Figure 7.1 for a simple diagram on how this algorithm operates. If there are spare RF chains after the antenna support of a beam has been defined, they are assigned to the following strongest path (per user). This procedure is iterated until all the resources have been allocated to their corresponding users.

[1]This procedure could be carried out during the selection phase by cleverly employing the second order statistics of the transmitted symbols.

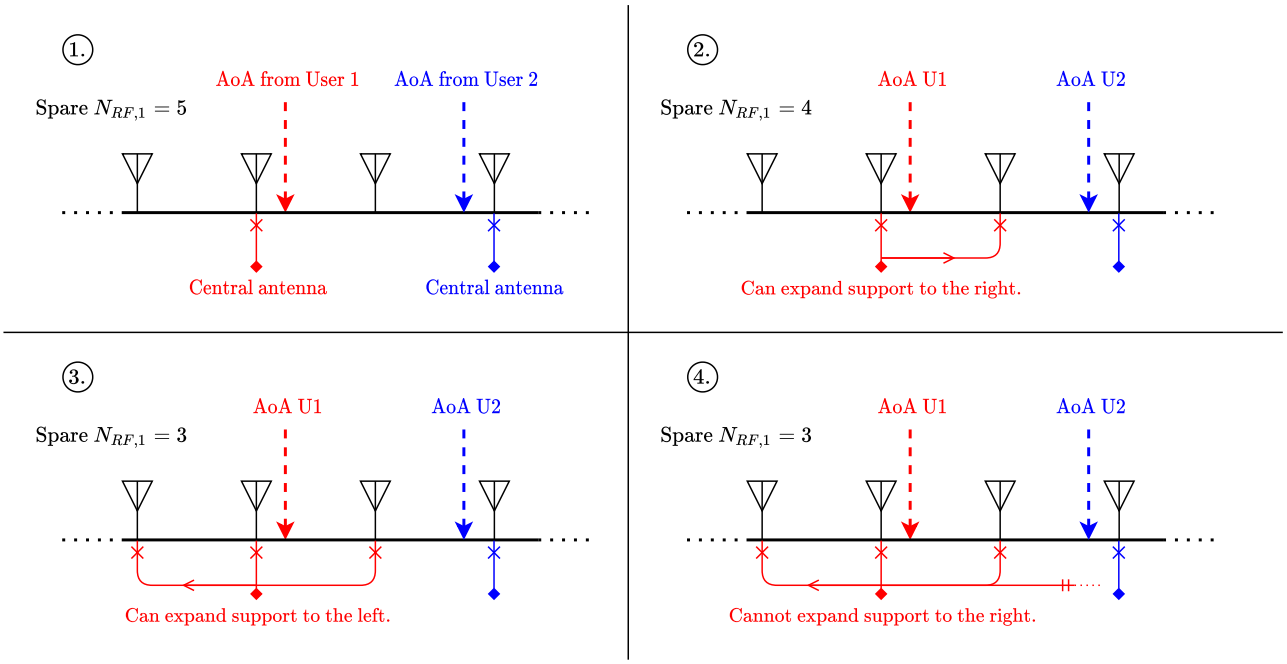


Figure 7.1: Four illustrative steps of the AS algorithm.

Figure 7.2 illustrates the result of this antenna selection technique[2]. The dark blue area represents the normalized $\bar{\bar{\mathbf{y}}}_{sel}$ over the antenna indices and the black dashed lines are the real AoAs of each beam. Each user is connected to the BS by 3 reliable paths. Notice that the union of supports (colored lines) for each user has roughly the same number of elements due to the fairness policy. Each support contains and is centered around a single AoA. Due to the energy proportionality criterion, an AoA of User 5 (light blue) is not covered by any support of antennas, since its associated power is negligible compared to its other two beams.

The supports of antennas allocated to each user after the antenna selection process, $\mathcal{M}_{sel}^{(u)}$, are gathered in individual selection matrices $\mathbf{\Sigma}_u$, as in **Section 2.2**:

$$[\mathbf{\Sigma}_u]_{m,m} = \begin{cases} 1 & \text{if } m \in \mathcal{M}_{sel}^{(u)} \\ 0 & \text{otherwise} \end{cases}, \quad u = 1, \ldots, N_U \tag{7.10}$$

A global selection matrix $\mathbf{\Sigma}$ for the BS can be constructed by combining all the individual ones:

$$\mathbf{\Sigma} \triangleq \bigoplus_{u=1}^{N_U} \mathbf{\Sigma}_u \tag{7.11}$$

On a final note regarding antenna selection, realize that AoAs are a large-scale effects and hence, changes in them are slow compared to the transmission of signal blocks. This implies that, while full AS is unavoidable on the first UL, antennas selected during this phase can be reused for the following DLs and ULs, thanks to channel reciprocity. In order to keep track of AoAs[3], however, the BS may need to perform AS periodically. Fortunately, due to the slowness in these changes, previously known AoAs can be employed to only perform a partial procedure (*i.e.* on fewer array elements), which is simpler and faster.

[2]This plot has been obtained from the UL simulations presented in **Section 7.3** at SNR = 0dB.

[3]AoA tracking is a classical and well understood problem in signal processing. Works like [52] study it under the LAA paradigm.

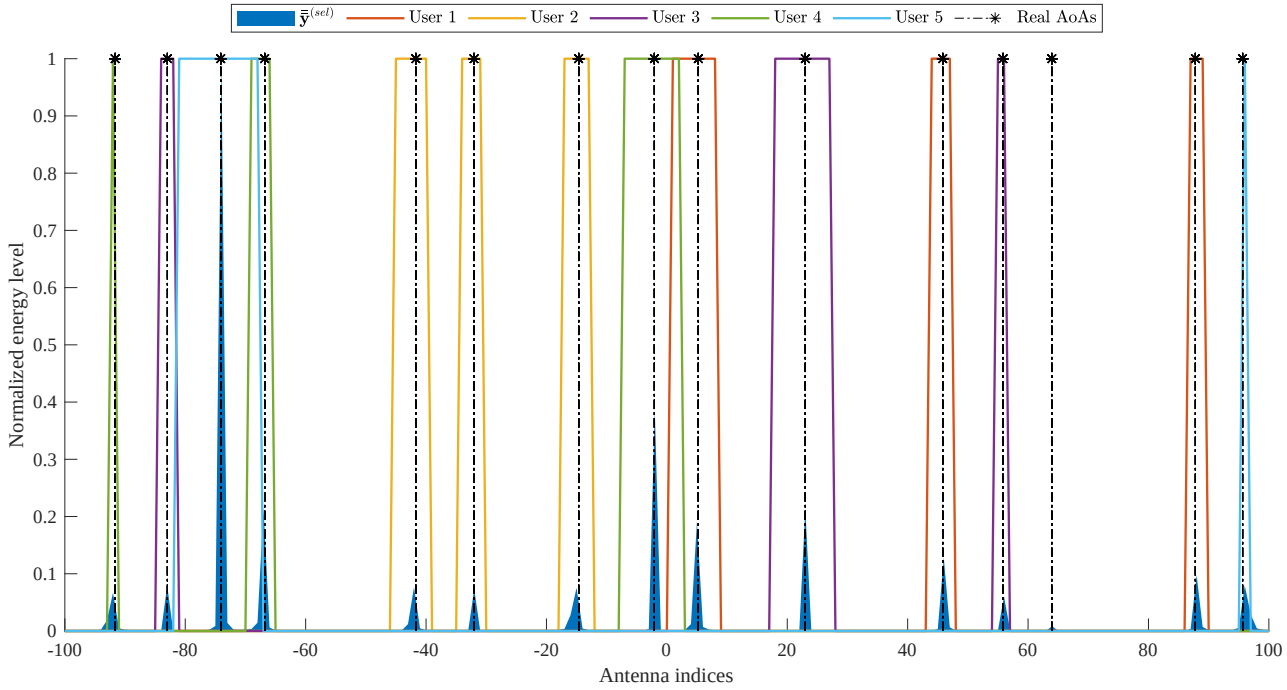


Figure 7.2: Example of antenna selection and assignment. The colored lines represent the supports for each user.

7.1.2 Channel covariance estimation

Once the AS is complete, the N_U MTs can start the transmission of relevant data to the BS, for a duration K_{UL}. We apply the selection matrix $\mathbf{\Sigma}$ onto (7.1):

$$\mathbf{Y} = \mathbf{\Sigma}\left(\mathbf{H}\mathbf{X}^T + \mathbf{Z}\right). \tag{7.12}$$

One of the design choices we have set for our multi-user communication scheme is *partial noncoherence*. Even without full CSI, we will rely on second order statistical knowledge about the channel, provided by its covariance matrix. Previously, in **Section 6.3.1**, we have presented a possible way to estimate it in the P2P case. However, since this is a multi-user setting, we will need an approach that takes ISI into consideration.

To gain some insights, let us study the structure of $\mathbf{C_h}^{(u)}$, the covariance matrix of the channel vector associated with each user u. Firstly, we recover (6.48) and generalize it for multiple users:

$$\begin{aligned}\mathbf{C_y} \triangleq \mathbb{E}\left[\mathbf{y}[k]\mathbf{y}^H[k]\right] &= \mathbb{E}\left[\left(\sum_{u=1}^{N_U}\mathbf{h}_u x_u[k] + \mathbf{z}[k]\right)\left(\sum_{u=1}^{N_U}\mathbf{h}_u x_u[k] + \mathbf{z}[k]\right)^H\right] \\ &= \mathbb{E}\left[\sum_{u=1}^{N_U}\mathbf{h}_u x_u[k]\sum_{u=1}^{N_U}x_u^*[k]\mathbf{h}_u^H\right] + P_z\mathbf{I}\end{aligned} \tag{7.13}$$

Under the assumed model, the channel vector $\mathbf{h}_u$ corresponding to each user is independent from the others, which implies $\mathbb{E}[\mathbf{h}_u\mathbf{h}_{u'}^H] = \mathbf{0}$ for $u \neq u'$. This reduces the first term of the covariance to a single

summation:

$$\mathbf{C_h} = \mathbb{E}\left[\sum_{u=1}^{N_U} \mathbf{h}_u x_u[k] x_u^*[k] \mathbf{h}_u^H\right] + P_z \mathbf{I}. \tag{7.14}$$

Furthermore, since we are employing an M-PSK modulation, every transmitted symbol is unitary, so $x_u[k]x_u^*[k] = |x_u[k]|^2 = 1$. Thus, the covariance matrix of the received signal takes the form:

$$\mathbf{C_y} = \overbrace{\mathbb{E}\left[\sum_{u=1}^{N_U} \mathbf{h}_u \mathbf{h}_u^H\right]}^{\mathbf{C_h}} + P_z \mathbf{I} = \sum_{u=1}^{N_U} \overbrace{\mathbb{E}\left[\mathbf{h}_u \mathbf{h}_u^H\right]}^{\mathbf{C_h}^{(u)}} + P_z \mathbf{I}. \tag{7.15}$$

This means that the full covariance matrix of the N_U channels is the sum of their individual covariances. We can characterize the expression of $\mathbf{C_h}$ for our LAA-MU-MIMO channel:

$$\begin{aligned}\mathbf{C_h} &= \sum_{u=1}^{N_U} \mathbb{E}\left[\sum_{l\in\mathcal{L}_u} \alpha_l^{UL} \mathbf{a}_{BS}\left(\tilde{\phi}_{BS,l}\right) \sum_{l'\in\mathcal{L}_u} \alpha_{l'}^{UL*} \mathbf{a}_{BS}^H\left(\tilde{\phi}_{BS,l'}\right)\right] \\ &= \sum_{u=1}^{N_U} \sum_{l\in\mathcal{L}_u} \sum_{l'\in\mathcal{L}_u} \mathbb{E}\left[\alpha_l^{UL} \mathbf{a}_{BS}\left(\tilde{\phi}_{BS,l}\right) \alpha_{l'}^{UL*} \mathbf{a}_{BS}^H\left(\tilde{\phi}_{BS,l'}\right)\right].\end{aligned} \tag{7.16}$$

In general, every signal path gain is independent from the rest, which translates into $\mathbb{E}\left[\alpha_l^{UL}\alpha_{l'}^{UL*}\right] = 0$ for $l \neq l'$. This assumption nulls the crossed terms in the double summation of (7.16) and leaves only the repeated ones

$$\begin{aligned}\mathbf{C_h} &= \sum_{u=1}^{N_U} \sum_{l\in\mathcal{L}_u} \mathbb{E}\left[\left|\alpha_l^{UL}\right|^2\right] \mathbf{a}_{BS}\left(\tilde{\phi}_{BS,l}\right) \mathbf{a}_{BS}^H\left(\tilde{\phi}_{BS,l}\right) \\ &= \sum_{u=1}^{N_U} \sum_{l\in\mathcal{L}_u} P_l \mathbf{a}_{BS}\left(\tilde{\phi}_{BS,l}\right) \mathbf{a}_{BS}^H\left(\tilde{\phi}_{BS,l}\right) \\ &= \boxed{\tilde{D}^2 \sum_{l\in\mathcal{L}} P_l \operatorname{sinc}\left(\mathbf{n} - (n_l - \epsilon_l)\,\mathbf{1}\right) \operatorname{sinc}\left(\mathbf{n}^H - (n_l - \epsilon_l)\,\mathbf{1}^H\right)}.\end{aligned} \tag{7.17}$$

$\mathcal{L} \triangleq \bigcup_{u=1}^{N_U} \mathcal{L}_u$ is the set of all channel paths coming from all N_U users and P_l is their associated power gain. $\mathbf{C_h}$ presents a very sparse structure. Since we have assumed inter-path independence, it is only made of the $|\mathcal{L}| << N_{BS}$ contributions. All its non-negligible terms will be gathered around the diagonal because it is a sum of 2D *sinc* functions with the same displacement in both axes, which contrasts with the richer structure seen in Figure 6.1a. This is expected as in that case, the channel coefficients stayed constant during the full transmission. However, in a real case scenario this assumption might not hold (*block fading* [34]). For this reason, only the diagonal *sinc* contributions remain. By sacrificing information from crossed terms in $\mathbf{C_h}$ we increased robustness against channel changes, as will be made clear in **Section 7.3**.

After closely analyzing the structure of $\mathbf{C_h}$, we must design a method to estimate it from the received signal block $\mathbf{Y}$ and extract the individual covariance matrices per user. We resort to the sample covariance once again. To slightly improve its accuracy, we can add more samples to its computation by simply appending the selection signals to $\mathbf{Y}$:

$$\tilde{\mathbf{Y}} \triangleq \left[\mathbf{\Sigma}\left(\mathbf{y}_{sel}[0], \ldots, \mathbf{y}_{sel}[T_{sel}-1]\right), \mathbf{Y}\right] \in \mathbb{C}^{N_{BS}\times(T_{sel}+K_{UL})} \Rightarrow \widehat{\mathbf{C}}_{\mathbf{y}} = \frac{1}{K+T_{sel}} \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^H. \tag{7.18}$$

Then, we remove the noise contribution as in (6.50): $\widehat{\mathbf{C}}_{\mathbf{h}} = \widehat{\mathbf{C}}_{\mathbf{y}} - \hat{P}_z \mathbf{I}$.

The individual channel covariances have the form

$$\mathbf{C_h}^{(u)} = \tilde{D}^2 \sum_{l \in \mathcal{L}_u} P_l \operatorname{sinc}\left(\mathbf{n} - (n_l - \epsilon_l)\,\mathbf{1}\right) \operatorname{sinc}\left(\mathbf{n}^H - (n_l - \epsilon_l)\,\mathbf{1}^H\right). \tag{7.19}$$

Notice that only P_l and ϵ_l are unknown, since $\tilde{D}$ is defined by the LAA and n_l have been obtained in the AS phase. Under the assumption that AoAs are sufficiently separated, *i.e.* small overlap between beams, relevant terms around the diagonal of $\mathbf{C_h}$ are composed by a single path contribution, with negligible leakage from the others. For this reason, by fitting a curve of the type *sinc* squared on a few selected points around $[\mathbf{C_h}]_{n_l,n_l}$, we can obtain a good approximation[4] for P_l and ϵ_l, the only degrees of freedom in the fitting process. The implementation of this procedure can be found in **Appendix A** (2.). With these results, we finally obtain the estimated channel covariance matrix per user:

$$\boxed{\widehat{\mathbf{C}}_{\mathbf{h}}^{(u)} = \tilde{D}^2 \sum_{l \in \mathcal{L}_u} \hat{P}_l \operatorname{sinc}\left(\mathbf{n} - (n_l - \hat{\epsilon}_l)\,\mathbf{1}\right) \operatorname{sinc}\left(\mathbf{n}^H - (n_l - \hat{\epsilon}_l)\,\mathbf{1}^H\right)}. \tag{7.20}$$

P_l is often referred to as Power Space Profile (PSP) in the literature [9, 23, 53].

7.1.3 Decision-Feedback Differential Detection (DFDD)

At the beginning of this chapter, we stated that one of the goals of our scheme is to be able to operate without the phase information of the channel, neither on the transmitter nor on the receiver ends (*i.e.* partially noncoherent scheme), thus avoiding the need for orthogonal pilot sequences and all the issues their use entails. Our approach will be based on differential modulation and detection.

As discussed in **Section 2.3**, CDD implies a 3dB loss in performance compared to coherent schemes. In **Section 6.3.1**, we have presented MSDD, a technique to overcome this limitation by performing joint differential detection on a sequence of consecutive symbols. The issue that hinders its implementation is the fact that the complexity of decoding a sequence of length K_{UL} scales exponentially with such length (*i.e.* $O\left(M^{K_{UL}-1}\right)$), making it computationally very demanding and not implementable in practice in most systems. Several attempts have been made in order to reduce its computational cost without sacrificing the performance gains of MSDD. Some prominent examples are based on sphere decoding algorithms (Multiple-Symbol Differential Sphere Decoding (MSDSD) [33]), and on decision feedback (Decision-Feedback Differential Detection (DFDD)). We will explore this last method in this section.

Consider the single symbol differential detection problem as an MSDD of block length $K_{UL} = 2$:

$$\begin{aligned}
\hat{\mathbf{x}}^{CDD} \equiv \hat{\mathbf{x}}^{MSDD}\Big|_{K_{UL}=2} &= \underset{\hat{\mathbf{x}} \in \mathcal{S}^2,\, \hat{x}[0]=1}{\arg\max}\ \operatorname{Re}\left\{\hat{x}^*[1] \sum_{k'=0}^{1-1} r_{k',1} \hat{x}[k']\right\} \\
&= \underset{\hat{\mathbf{x}} \in \mathcal{S}^2,\, \hat{x}[0]=1}{\arg\max}\ \operatorname{Re}\left\{\hat{x}^*[1] \hat{x}[0] r_{0,1}\right\} \\
&= \underset{\hat{\mathbf{x}} \in \mathcal{S}^2,\, \hat{x}[0]=1}{\arg\max}\ \operatorname{Re}\left\{\hat{x}[1] \hat{x}^*[0] r_{1,0}\right\}.
\end{aligned} \tag{7.21}$$

Since we are only interested in the information symbols $s[k] = x[k]x^*[k-1]$, the previous detection problem reduces to

$$\begin{aligned}
\hat{s}^{CDD}[k] &= \underset{\hat{s} \in \mathcal{S}}{\arg\max} \operatorname{Re}\left\{\hat{s}[k] \hat{x}[k-1] \hat{x}^*[k-1] r_{k,k-1}\right\} \\
&= \underset{\hat{s} \in \mathcal{S}}{\arg\max} \operatorname{Re}\left\{\hat{s}[k] r_{k,k-1}\right\} \\
&= \exp\left(j \mathcal{Q}_{PSK}\left(r_{k,k-1}\right)\right),
\end{aligned} \tag{7.22}$$

[4] The approximation becomes exact when there is no power leakage, *i.e.* $\epsilon_l = 0,\ \forall l \in \mathcal{L}$.

which is the quantization of the phase of $r_{k,k-1}$ over the M-ary constellation $\mathcal{S}$, by defining

$$\mathcal{Q}_{PSK}(x) \triangleq \frac{2\pi}{M}\left\lceil \frac{M}{2\pi}\arg(x) \right\rfloor. \tag{7.23}$$

Therefore, the correlation coefficient $r_{k,k-1}$ is the decision variable for the M-PSK information symbol.

To improve the performance over this detection scheme while avoiding full MSDD, we can extend the length of the block of symbols without having to perform a search over all possible sequences. To do so, we detect new symbols by taking into account the already decided ones, in a successive procedure; this is the core idea behind DFDD. Starting with $\hat{x}^{DFDD}[0] \equiv 1$, the following symbol estimates are

$$\begin{aligned}
\hat{x}^{DFDD}[1] &= \exp\left(j\mathcal{Q}_{PSK}\left(\hat{x}^{DFDD}[0]r_{0,1}\right)\right) \\
\hat{x}^{DFDD}[2] &= \exp\left(j\mathcal{Q}_{PSK}\left(\hat{x}^{DFDD}[0]r_{0,2} + \hat{x}^{DFDD}[1]r_{1,2}\right)\right) \\
&\vdots \\
\hat{x}^{DFDD}[k] &= \exp\left(j\mathcal{Q}_{PSK}\left(\sum_{k'=0}^{k-1}\hat{x}^{DFDD}[k']r_{k',k}\right)\right), \quad k = 1, \ldots, K_{UL} - 1.
\end{aligned} \tag{7.24}$$

Then, to recover the information symbols $\hat{s}^{DFDD}[k]$ we simply apply differential decoding:

$$\hat{s}^{DFDD}[k] = \hat{x}^{DFDD}[k]\prod_{k'=0}^{k-1}\hat{s}^{DFDD*}[k'], \quad k = 1, \ldots, K_{UL} - 1. \tag{7.25}$$

This scheme presents a complexity of the order of $O(M)$, since only a single symbol has to be detected at a time [33]. The trade-off to achieve this complexity reduction is a performance loss compared to MSDD and an increased vulnerability to error propagation.

The order in which the decisions are made can greatly influence the error rate of DFDD, and optimizing it can lead to marginal loss in performance compared to MSDD. The optimum decision order is a variation on the well-known concept of Decision-Feedback Equalization (DFE), the basis of Bell Laboratories Layered Space-Time (BLAST) systems. The metric to sort the reliability of the symbols in each step will be the phase quantization error over $\mathcal{S}$, which is computed with the function

$$\Delta\mathcal{Q}_{PSK}(x) \triangleq (\arg(x) - \mathcal{Q}_{PSK}(x)) \mod_s 2\pi. \tag{7.26}$$

It is important to highlight the conceptual difference between BLAST schemes and DFDD: in the former, sorting is performed per channel realization among users, while in the latter, it is performed per user among symbols in a block.

In the first iteration, we select the pair of symbols ($\hat{k}_0$ and $\hat{k}_1$) that are the most reliable to decide. According to (7.21), they are connected through their corresponding correlation coefficient ($r_{\hat{k}_0,\hat{k}_1}$), so we want to find the one which produces the smallest quantization error:

$$\left[\hat{k}_0, \hat{k}_1\right] = \underset{\substack{\hat{k}_0,\hat{k}_1 \in \{0,\ldots,K_{UL}-1\} \\ \hat{k}_0 < \hat{k}_1}}{\arg\min} \left|\Delta\mathcal{Q}_{PSK}\left(r_{\hat{k}_0,\hat{k}_1}\right)\right|. \tag{7.27}$$

We set $\hat{x}^{DFDD}[\hat{k}_0] \equiv 1$ and decode $\hat{x}^{DFDD}[\hat{k}_1]$ with (7.24). Next, taking these two decisions into account, we search among the remaining symbols $\hat{x}^{DFDD}[\hat{k}], \hat{k} \in \{0, \ldots, K_{UL} - 1\} / \left\{\hat{k}_0, \hat{k}_1\right\}$ which one can be decided in the next step the most reliably:

$$\hat{k}_2 = \underset{\substack{\hat{k} \in \{0,\ldots,K_{UL}-1\} \\ / \left\{\hat{k}_0,\hat{k}_1\right\}}}{\arg\min} \left|\Delta\mathcal{Q}_{PSK}\left(\hat{x}^{DFDD}[\hat{k}_0]r_{\hat{k},\hat{k}_0} + \hat{x}^{DFDD}[k_1]r_{\hat{k},\hat{k}_1}\right)\right|. \tag{7.28}$$

We continue this procedure with the optimum decision order

$$\hat{k}_n = \underset{\substack{\hat{k}\in\{0,\dots,K_{UL}-1\} \\ /\{\hat{k}_0,\dots,\hat{k}_{n-1}\}}}{\arg\min} \left| \Delta\mathcal{Q}_{PSK}\left(\sum_{m=0}^{n-1} \hat{x}^{DFDD}[\hat{k}_m] z_{\hat{k}_n,\hat{k}_m} \right) \right|, \quad n = 0,\dots,K_{UL}-1, \tag{7.29}$$

and the symbol estimates at every step

$$\hat{x}^{DFDD}[\hat{k}_n] = \exp\left(j\mathcal{Q}_{PSK}\left(\sum_{m=0}^{n-1} \hat{x}^{DFDD}[k_m] r_{\hat{k}_n,\hat{k}_m} \right) \right), \tag{7.30}$$

setting $\hat{x}^{DFDD}[\hat{k}_0] \equiv 1$. After the complete block has been detected, every symbol (with the original ordering) can be sequentially decoded. The DFDD algorithm can be performed very efficiently since the successive detection of symbols can be *vectorized*.

There is one minor detail regarding the order of detection that must be stated. The detection of a given symbol in the sequence is based on an increasing number of prior decisions. In particular, the first detected symbol, which is the one with the least phase quantization error, relies on no prior result, while the last one benefits from $K_{UL}-1$ estimates. Since it is reasonable to consider that more priors translate into more reliable estimations, this situation leads to varying conditions for the individual decisions. In [24], the authors propose a method to combat this effect. After all symbols have been detected, the DFDD algorithm can iterate once again over each one of them. This way, the detection is performed with better conditions, as always $K_{UL}-1$ prior estimations are available. Note that, after the first iteration has been completed, any additional one does not require symbol ordering, since each one of them is virtually the last.

In the original article [24], it is numerically shown that this simple modification exhibits significant gains compared to the non-iterated scheme. It is also stated that the number of additional iterations can be kept rather small: a single additional iteration provides almost the full obtainable gain.

A version of this algorithm coded in Matlab can be found in **Appendix A** (5.).

7.1.4 Noncoherent Decision-Feedback Equalization (nDFE)

A central issue in any wireless multi-user scheme is IUI. How effectively the interference is combated can be the determinant factor on whether a system is implementable in practice or not. The multi-user LAA system we are studying in this section is no exception. The PDM capability of the LAA guarantees small Inter-Path Interference (IPI) for sufficiently spread paths in the angular domain. Since different MTs will, in general, communicate to the BS through different paths, their IUI will, as well, be smaller thanks to the use of the lens. This extension of the PDM concept to a multi-user environment has been coined Path Division Multiple Access (PDMA) in [17]. Nevertheless, although the IUI effect is less significant than in other multi-user environments, it still degrades the overall performance and achievable sum rate of the system.

A typical approach to deal with IUI in coherent multi-user systems is sorted DFE over the users, also referred to as BLAST in MIMO contexts [54–56]. In a nutshell, it consists in detecting each user sequentially and using already detected ones to cancel interference (Successive Interference Cancellation (SIC)). The order in which this procedure is performed is dictated by some metric that quantifies the reliability of decoding a user sequence at each step. The interference cancellation, however, requires full channel knowledge. It would seem that this idea is not applicable in our noncoherent scheme, but this is not the case. In [23], the authors propose a variation of DFE that operates with statistical CSI: Noncoherent Decision-Feedback Equalization (nDFE).

Let us closely analyze the correlation matrix of each user,

$$\mathbf{R}_u \triangleq \mathbf{Y}^H \mathbf{W}_u \mathbf{Y}. \tag{7.31}$$

Omitting the effect of the selection matrix $\mathbf{\Sigma}_u$, $\mathbf{R}_u$ is decomposed in 5 distinguishable terms:

$$\begin{aligned}
\mathbf{R}_u = \mathbf{Y}^H \mathbf{W}_u \mathbf{Y} &= \left(\sum_{\nu=1}^{N_U} \mathbf{x}_\nu^* \mathbf{h}_\nu^H + \mathbf{Z}^H \right) \mathbf{W}_u \left(\sum_{\nu=1}^{N_U} \mathbf{h}_\nu \mathbf{x}_\nu^T + \mathbf{Z} \right) \\
&= \left(\mathbf{h}_u^H \mathbf{W}_u \mathbf{h}_u \right) \mathbf{x}_u^* \mathbf{x}_u^T && (7.32) \\
&\quad + \sum_{\nu \neq u} \left(\mathbf{h}_\nu^H \mathbf{W}_u \mathbf{h}_\nu \right) \mathbf{x}_\nu^* \mathbf{x}_\nu^T && (7.33) \\
&\quad + \sum_{\mu=1}^{N_U} \sum_{1 \leq \nu < \mu} \left(\left(\mathbf{h}_\nu^H \mathbf{W}_u \mathbf{h}_\mu \right) \mathbf{x}_\mu^* \mathbf{x}_\nu^T + \left(\mathbf{h}_\mu^H \mathbf{W}_u \mathbf{h}_\nu \right) \mathbf{x}_\nu^* \mathbf{x}_\mu^T \right) && (7.34) \\
&\quad + \sum_{\nu=1}^{N_U} \mathbf{x}_\nu^* \mathbf{h}_\nu^H \mathbf{W}_u \mathbf{Z} + \mathbf{Z}^H \mathbf{W}_u \sum_{\nu=1}^{N_U} \mathbf{h}_\nu \mathbf{x}_\nu^T && (7.35) \\
&\quad + \mathbf{Z}^H \mathbf{W}_u \mathbf{Z} && (7.36)
\end{aligned}$$

(7.32) contains the desired correlation coefficients of the sequence transmitted by user u, (7.33) is the autocorrelation contribution of all the other users, (7.34) has the cross-correlation between sequences by different users (including u), (7.35) is the correlation between signals and noise, and finally, (7.36) is the noise correlation term.

We now characterize the statistics of all the previous terms to be able to cancel their effect. In [23], the authors assume the uncorrelated channel model from [22] (commented in **Section 6.3.1**) and derive simple statistics based on normal distributions. Since we are dealing with a more structured channel model, we propose a slightly more involved derivation of the first and second order moments of the previous terms.

Let us define the correlation factors

$$\xi_{u,\nu} \triangleq \mathbf{h}_\nu^H \mathbf{W}_u \mathbf{h}_\nu = \sum_{n=1}^{N_U} \sum_{m=1}^{N_U} w_{m,n}^{(u)} h_{\nu,m}^* h_{\nu,n} \tag{7.37}$$

$$\xi_{u,\nu,\mu} \triangleq \mathbf{h}_\nu^H \mathbf{W}_u \mathbf{h}_\mu = \sum_{n=1}^{N_U} \sum_{m=1}^{N_U} w_{m,n}^{(u)} h_{\nu,m}^* h_{\mu,n}. \tag{7.38}$$

(7.37) corresponds to the autocorrelation factor of the channel from user ν weighted in the detection of user u, while (7.38) represents the cross-correlation between channels from users ν and μ weighted in that same detection. Their expected values are, respectively

$$\begin{aligned}
\eta_{u,\nu} \triangleq \mathbb{E}\left[\xi_{u,\nu}\right] &= \mathbb{E}\left[\mathbf{h}_\nu^H \mathbf{W}_u \mathbf{h}_\nu\right] = \mathbb{E}\left[\mathrm{Tr}\left[\mathbf{h}_\nu^H \mathbf{W}_u \mathbf{h}_\nu\right]\right] = \mathbb{E}\left[\mathrm{Tr}\left[\mathbf{W}_u \mathbf{h}_\nu \mathbf{h}_\nu^H\right]\right] \\
&= \mathrm{Tr}\left[\mathbb{E}\left[\mathbf{W}_u \mathbf{h}_\nu \mathbf{h}_\nu^H\right]\right] = \mathrm{Tr}\left[\mathbf{W}_u \mathbb{E}\left[\mathbf{h}_\nu \mathbf{h}_\nu^H\right]\right] = \boxed{\mathrm{Tr}\left[\mathbf{W}_u \mathbf{C}_{\mathbf{h}}{}^{(\nu)}\right]} && (7.39) \\
\eta_{u,\nu,\mu} \triangleq \mathbb{E}\left[\xi_{u,\nu,\mu}\right] &= \mathrm{Tr}\left[\mathbf{W}_u \mathbb{E}\left[\mathbf{h}_\mu \mathbf{h}_\nu^H\right]\right] = \mathrm{Tr}\left[\mathbf{W}_u \mathbf{0}\right] = 0. && (7.40)
\end{aligned}$$

In (7.39), we have used the cyclic property of the trace and the commutation of $\mathbb{E}[\cdot]$ and $\mathrm{Tr}[\cdot]$, since they are both linear operators and, in (7.40), we have used the fact that different users have uncorrelated

channel vectors. Regarding the variance of $\xi_{u,\nu}$,

$$\begin{aligned}\sigma^2_{u,\nu} &\triangleq \mathbb{E}\left[\xi_{u,\nu}\xi^*_{u,\nu}\right] - |\eta_{u,\nu}|^2 = \mathbb{E}\left[\mathbf{h}_\nu^H \mathbf{W}_u \mathbf{h}_\nu \mathbf{h}_\nu^H \mathbf{W}_u \mathbf{h}_\nu\right] - \mathrm{Tr}\left[\mathbf{W}_u \mathbf{C_h}^{(\nu)}\right]^2 \\ &= \mathbb{E}\left[\sum_{n=1}^{N_U}\sum_{m=1}^{N_U} w^{(u)}_{m,n} h^*_{\nu,m} h_{\nu,n} \sum_{n'=1}^{N_U}\sum_{m'=1}^{N_U} w^{(u)}_{m',n'} h_{\nu,m'} h^*_{\nu,n'}\right] - \mathrm{Tr}\left[\mathbf{W}_u \mathbf{C_h}^{(\nu)}\right]^2 \\ &= \sum_n\sum_m\sum_{n'}\sum_{m'} w^{(u)}_{m,n} w^{(u)}_{m',n'} \mathbb{E}\left[h^*_{\nu,m} h_{\nu,n} h_{\nu,m'} h^*_{\nu,n'}\right] - \mathrm{Tr}\left[\mathbf{W}_u \mathbf{C_h}^{(\nu)}\right]^2. \end{aligned} \tag{7.41}$$

The first term of the previous expression is a sum of fourth order moments of a complex multivariate normal distribution, since $\mathbf{h}_\nu \sim \mathcal{CN}(\mathbf{0}, \mathbf{C_h})$. By using *Isserlis' Theorem* [57], and assuming $\mathbf{h}_\nu$ is circularly symmetric (*i.e.* , $\mathbb{E}\left[\mathbf{h}_\nu \mathbf{h}_\nu^T\right] = \mathbf{0}$), we can proceed in the following manner:

$$\begin{aligned}\sigma^2_{u,\nu} &= \sum_{n,m,n',m'} w^{(u)}_{m,n} w^{(u)}_{m',n'} \left(\mathbb{E}\left[h^*_{\nu,m} h_{\nu,n}\right]\mathbb{E}\left[h_{\nu,m'} h^*_{\nu,n'}\right] + \mathbb{E}\left[h^*_{\nu,m} h_{\nu,m'}\right]\mathbb{E}\left[h_{\nu,n} h^*_{\nu,n'}\right]\right) - \mathrm{Tr}\left[\mathbf{W}_u \mathbf{C_h}^{(\nu)}\right]^2 \\ &= \sum_{n,m,n',m'} w^{(u)}_{m,n} w^{(u)}_{m',n'} c^{(\nu)}_{m,n} c^{(\nu)}_{m',n'} + \sum_{n,m,n',m'} w^{(u)}_{m,n} w^{(u)}_{m',n'} c^{(\nu)}_{m,m'} c^{(\nu)}_{n',n} - \mathrm{Tr}\left[\mathbf{W}_u \mathbf{C_h}^{(\nu)}\right]^2. \end{aligned} \tag{7.42}$$

We have denoted $\left[\mathbf{C_h}^{(\nu)}\right]_{m,n}$ as $c^{(\nu)}_{m,n}$ for visual clarity. Then, we simply express the two summations in matrix form:

$$\begin{aligned}\sigma^2_{u,\nu} &= \mathrm{Tr}\left[\mathbf{W}_u \mathbf{C_h}^{(\nu)}\right]^2 + \mathrm{Tr}\left[\mathbf{W}_u \mathbf{C_h}^{(\nu)} \mathbf{W}_u \mathbf{C_h}^{(\nu)}\right] - \mathrm{Tr}\left[\mathbf{W}_u \mathbf{C_h}^{(\nu)}\right]^2 \\ &= \boxed{\mathrm{Tr}\left[\left(\mathbf{W}_u \mathbf{C_h}^{(\nu)}\right)^2\right]}. \end{aligned} \tag{7.43}$$

Deriving the variance $\sigma^2_{u,\nu,\mu}$ is much less involved since it depends on independent variables (cross-signal interference):

$$\begin{aligned}\sigma^2_{u,\nu,\mu} &= 2\,\mathbb{E}\left[\xi_{u,\nu,\mu}\xi^*_{u,\nu,\mu}\right] = 2\,\mathbb{E}\left[\sum_{n=1}^{N_U}\sum_{m=1}^{N_U} w^{(u)}_{m,n} h^*_{\nu,m} h_{\mu,n} \sum_{n'=1}^{N_U}\sum_{m'=1}^{N_U} w^{(u)}_{m',n'} h_{\nu,m'} h^*_{\mu,n'}\right] \\ &= 2\sum_n\sum_m\sum_{n'}\sum_{m'} w^{(u)}_{m,n} w^{(u)}_{m',n'} \mathbb{E}\left[h^*_{\nu,m} h_{\nu,m'} h_{\mu,n} h^*_{\mu,n'}\right] \\ &= 2\sum_n\sum_m\sum_{n'}\sum_{m'} w^{(u)}_{m,n} w^{(u)}_{m',n'} \mathbb{E}\left[h^*_{\nu,m} h_{\nu,m'}\right]\mathbb{E}\left[h_{\mu,n} h^*_{\mu,n'}\right]. \end{aligned} \tag{7.44}$$

We have used the fact that $\mathbf{h}_\nu$ is independent from $\mathbf{h}_\mu$, so

$$\mathbb{E}\left[h^*_{\nu,m} h_{\nu,m'} h_{\mu,n} h^*_{\mu,n'}\right] = \mathbb{E}\left[h^*_{\nu,m} h_{\nu,m'}\right]\mathbb{E}\left[h_{\mu,n} h^*_{\mu,n'}\right]. \tag{7.45}$$

The factor 2 is to account for the conjugate transpose term in (7.34). $\sigma^2_{u,\nu,\mu}$ is then expressed in terms of the covariance matrices associated to users ν and μ:

$$\sigma^2_{u,\nu,\mu} = 2\sum_n\sum_m\sum_{n'}\sum_{m'} w^{(u)}_{m,n} w^{(u)}_{m',n'} c^{(\nu)}_{m,m'} c^{(\mu)}_{n,n'} = \boxed{2\,\mathrm{Tr}\left[\mathbf{W}_u \mathbf{C_h}^{(\nu)} \mathbf{W}_u \mathbf{C_h}^{(\mu)}\right]}. \tag{7.46}$$

Finally, we will characterize the cross influence of noise with signal and the noise autocorrelation (both

are zero-mean) under the same factor:

$$\begin{aligned}\sigma^2_{n,u} &= \overbrace{2\,\mathbb{E}\left[\sum_{\mu=1}^{N_U}\sum_{\nu=1}^{N_U}\mathbf{h}_\nu^H\mathbf{W}_u\mathbf{Z}\mathbf{Z}^H\mathbf{W}_u\mathbf{h}_\mu\right]}^{\text{signal}\times\text{noise}} + \overbrace{\mathbb{E}\left[\mathrm{Tr}\left[\mathbf{Z}^H\mathbf{W}_u\mathbf{Z}\mathbf{Z}^H\mathbf{W}_u\mathbf{Z}\right]\right]}^{\text{noise}\times\text{noise}} \\ &= 2\,\mathbb{E}\left[\sum_{\mu}\sum_{\nu}\mathrm{Tr}\left[\mathbf{W}_u\mathbf{Z}\mathbf{Z}^H\mathbf{W}_u\mathbf{h}_\mu\mathbf{h}_\nu^H\right]\right] + \mathrm{Tr}\left[\mathbf{W}_u\,\mathbb{E}\left[\mathbf{Z}\mathbf{Z}^H\right]\mathbf{W}_u\,\mathbb{E}\left[\mathbf{Z}\mathbf{Z}^H\right]\right] \\ &= 2\sum_{\mu}\sum_{\nu}\mathrm{Tr}\left[\mathbf{W}_u\,\mathbb{E}\left[\mathbf{Z}\mathbf{Z}^H\right]\mathbf{W}_u\,\mathbb{E}\left[\mathbf{h}_\mu\mathbf{h}_\nu^H\right]\right] + \mathrm{Tr}\left[\mathbf{W}_u P_z\mathbf{I}\mathbf{W}_u P_z\mathbf{I}\right]. \end{aligned} \tag{7.47}$$

We have used the same rationale as in (7.43) and the uncorrelatedness between the channel and the noise. In the first term, the only addends that do not null are the ones in which $\mu=\nu$. The result in matrix form is

$$\boxed{\sigma^2_{n,u} = 2P_z\,\mathrm{Tr}\left[\mathbf{W}_u^2\sum_{\nu=1}^{N_U}\mathbf{C_h}^{(\nu)}\right] + P_z^2\,\mathrm{Tr}\left[\mathbf{W}_u^2\right]} \tag{7.48}$$

After these parameters have been particularized, we are now in the position to define the nDFE principles and operation. The main idea behind DFE is the successive subtraction of interference caused by already detected users. Under the noncoherent paradigm, this subtraction can only be done statistically. Since the channel coefficients are not explicitly known, the best guess we can have is its expected value. More precisely, we can subtract the signal correlation effect of already detected users from the correlation matrix $\mathbf{R}_u$ of user u. The scale of this subtraction is dictated by the interference autocorrelation mean, $\eta_{u,\nu}$. The set of already detected users is $\nu\in\mathcal{D}$, with estimated signal sequences $\widehat{\mathbf{x}}_\nu$. The correlation matrix for detection of user u with removed interference is computed as

$$\boxed{\mathbf{R}'_u = \mathbf{R}_u - \sum_{\nu\in\mathcal{D}}\eta_{u,\nu}\widehat{\mathbf{x}}_\nu^*\widehat{\mathbf{x}}_\nu^T} \tag{7.49}$$

Analogous to DFDD (and thus, analogous to BLAST) the order in which the interference cancellation is conducted can greatly affect the overall performance of the scheme. We want to start dealing with users whose detection is the most reliable, so that removing their interference from the other received sequences propagates the minimum amount of errors. A metric that quantifies the transmission conditions of each user is their Signal to Noise plus Interference Ratio (SINR) after interference from previously detected ones has been suppressed. From the previously described statistics of the received signals, we define the SINR of user u as

$$\begin{aligned}\mathrm{SINR}_u &\triangleq \boxed{\frac{\eta_{u,u}^2+\sigma_{u,u}^2}{\left(\sum\limits_{\nu\neq u}\sigma_{u,\nu}^2 + \sum\limits_{\substack{\nu\notin\mathcal{D}\\ \nu\neq u}}\eta_{u,\nu}^2\right) + \left(\sum\limits_{\nu<\mu}\sigma_{u,\nu,\mu}^2\right) + \left(\sigma_{n,u}^2\right)}} \\ &= \frac{\text{desired signal power}}{(\text{signal}\times\text{interf.}) + (\text{interf.}\times\text{interf.}) + (\text{interf.}\times\text{noise} + \text{noise}\times\text{noise})}\end{aligned} \tag{7.50}$$

Similarly to what happened with DFDD, the last user to be detected profits the most from the cancellation of all the other mean interference. Once every user has been detected, estimates for all the sequences are available. Hence, an additional iteration of interference suppression can be applied

to improve the detection on all users, without having to sort them again by the SINR criterion. The interference-reduced correlation matrix of this second iteration is computed as

$$\mathbf{R}''_u = \mathbf{R}_u - \sum_{\forall \nu \neq u} \eta_{u,\nu} \widehat{\mathbf{x}}^*_\nu \widehat{\mathbf{x}}^T_\nu \tag{7.51}$$

This procedure could be repeated various times. However, in [24] it is shown numerically that the majority of the gain achievable with this method is already obtained by just one additional nDFE iteration.

7.1.5 Some remarks on joint DFDD + nDFE

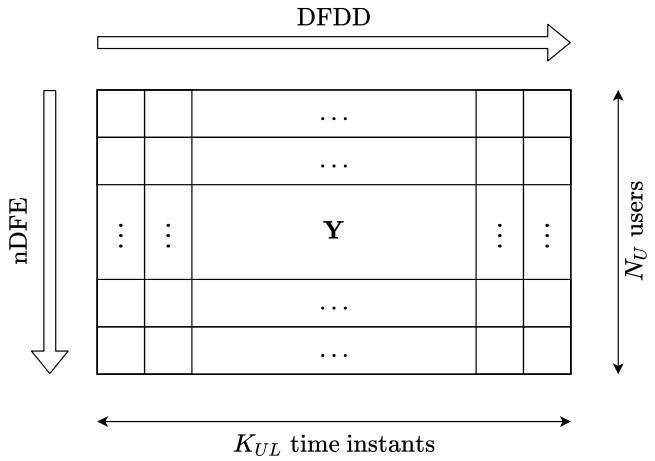


Figure 7.3: Graphical representation of DFDD+nDFE on a signal block $\mathbf{Y}$.

As we have seen, the uplink communication phase (not accounting for AS) is based on two DFE tasks along 2 orthogonal dimensions (Figure 7.3): users (nDFE) and time (DFDD). They both present significant performance gains under iterative applications. The sequential nature of nDFE and DFDD raises the question on whether their results can be improved by jointly performing the user/temporal sorting.

It turns out that detecting the received symbols over the whole user/time plane jointly requires a redesign of the weighting matrix of each user. It is numerically shown in [48] that the performance gains of this joint scheme are marginal in front of performing nDFE and DFDD separately, while its increased complexity limits its application. Hence, in terms of performance-complexity trade-off, DFDD/nDFE should be the default choice.

Another aspect that should be addressed is how to combine the various iterations of both algorithms. In [24], different configurations are tested. The simulations show that the best performance is achieved when iterative nDFE over the users is applied and, on each one, iterative DFDD is employed, as well. This is the approach we will consider in the simulations. A Matlab code version of the full DFDD/nDFE algorithm can be found in **Appendix A** (4.).

7.2 Downlink

In systems that rely on full CSI and channel reciprocity, *i.e.* $\mathbf{h}_u^{UL} \equiv \mathbf{h}_u^{DL*}$, Time Division Duplexing (TDD) can be employed for the DL by simply reusing the pilot information from the UL. This is a very strong assumption and not always realistic. In addition, Frequency Division Duplexing (FDD) is not easily implementable, since channel reciprocity is not ensured at different spectral bands. It cannot reuse the same pilots between UL and DL and would require further training time.

The scheme we are going to present in this section only expects *path reciprocity*; this means that the amount of signal paths ($L_u \triangleq |\mathcal{L}_u|, \quad \forall u = 1, \dots, N_U$) as well as their corresponding AoAs and AoDs do not change between UL and DL ($\tilde{\phi}_{BS,l} = \tilde{\phi}_{MT,l}$). This allows the use of TDD and FDD equally[5]. Since the lens is a passive component, at least in TDD it is realistic to expect the maintenance of second order statistics, *i.e.* power, between UL and DL ($\left|\alpha_l^{UL}\right|^2 = \left|\alpha_l^{DL}\right|^2$). In contrast to the channel phases ω_l, which are small-scale effects, path gains are large-scale effects and vary slowly, reasserting our channel reciprocity assumptions. This will allow the reuse of the partial CSI acquired during the UL on the DL phase, translating into performance gains in front of completely agnostic noncoherent schemes.

In this setting, the BS transmits data to the same N_U MTs from the UL simultaneously (MU-MIMO). The space-time block notation of the signal received by each user u is

$$\begin{aligned} \breve{\mathbf{y}}_u^T &= \mathbf{h}_u^H \breve{\mathbf{X}} + \breve{\mathbf{z}}^T \in \mathbb{C}^{1 \times K_{DL}} \\ [y_u[0], \dots, y_u[K_{DL}-1]] &= \mathbf{h}_u^H \left[\mathbf{x}_u[0], \dots, \mathbf{x}_u[K_{DL}-1]\right] + [z_u[0], \dots, z_u[K_{DL}-1]] \\ \left[\mathbf{y}_u^T[0], \dots, \mathbf{y}_u[K_{DL}-K_b^{(u)}]\right] &= \mathbf{h}^H \left[\mathbf{X}_u[0], \dots, \mathbf{X}_u[K_{DL}-K_b^{(u)}]\right] + \left[\mathbf{z}_u^T[0], \dots, \mathbf{z}_u^T[K_{DL}-K_b^{(u)}]\right] \end{aligned} \tag{7.52}$$

$$\mathbf{y}_u^T[i] = \mathbf{h}_u^H \mathbf{X}_u[i] + \mathbf{z}_u^T[i] \in \mathbb{C}^{1 \times K_{DL}} \quad , \quad i = 0, K_b^{(u)}, \dots, K_{DL} - K_b^{(u)}, \tag{7.53}$$

the same one employed in **Section 6.3.2**.

7.2.1 D-OSTBC with Power Loading (D-OSTBC-PL)

As explained in **Section 6.3.2**, DSTM allows to bypass the CSIR acquisition required in regular space-time schemes. This is well motivated in our DL scheme, because we want to remove the processing load from MTs. In **Section 6.2**, we have declared that the isotropic nature of basic STBC methods is not in contradiction with special kinds of Tx beamforming and power loading. As we will show next, the combination of the two philosophies will prove very useful in LAA channels. We have imported some ideas from [26], in which the authors propose a way to exploit the available statistical CSI at the Tx to improve the performance of basic D-OSTBC.

For simplicity, we will display the communication between the BS and a single MT u. Extending the scheme to the multi-user setting is just a matter of superposing the N_U parallel MISO systems. Consider an orthogonal DSTM scheme based on the linear model from (6.61):

$$\mathbf{S}_u[i] \triangleq \sqrt{\frac{1}{M_s^{(u)}}} \sum_{m=1}^{M_s^{(u)}} \left(\mathrm{Re}\left\{ s_m^{(u)}[i] \right\} \mathbf{A}_m^{(u)} + j\, \mathrm{Im}\left\{ s_m^{(u)}[i] \right\} \mathbf{B}_m^{(u)} \right), \tag{7.54}$$

[5] *Beam-squint* is the frequency dependence of the beam power distribution of a signal path component. This issue is present in many wideband multiantenna systems. Similarly, LAA systems also present a type of beam squint (called *dispersion* in optics), which is the result of different refraction indices at different wavelengths. This implies that with FDD, if the operation frequencies of the UL and DL are too far apart, the beam squint becomes non-negligible. For the purposes of this thesis, we will ignore it but, otherwise, it should be corrected with array processing techniques.

which has rate $R_b^{(u)}$ and dimensions $M_s^{(u)} \times K_b^{(u)}$. Once again, we assume that K_{DL} is a multiple of $K_b^{(u)}$ for convenience. The differentially encoded symbol block is then computed as $\check{\mathbf{C}}_u[i] = \check{\mathbf{C}}_u[i-1]\mathbf{S}_u[i]$, with $\check{\mathbf{C}}_u[0] \equiv \mathbf{I}_{M_s^{(u)}}$.

As stated in **Section 6.3.2**, the spatial dimension of the transmitted blocks and, to a extent, the amount of symbols per block, is determined by L_u, the number of reliable signal paths between the BS and MT u. Since the propagation conditions can vary across users ($L_u \neq L_{u'}$, $u \neq u'$), each one of them might receive signal blocks of different sizes and rates. To adapt the encoded blocks $\check{\mathbf{C}}_u[i]$ to the channel characteristics, we apply a suitable trimming matrix to them: $\mathbf{C}_u[i] \triangleq \mathbf{\Phi}_{L_u}\check{\mathbf{C}}_u[i]$. We will refer to this trimmed matrix as *codeword*.

The main idea of [26] is to transmit the codewords for user u along the eigenvectors of its associated channel correlation matrix, $\mathbf{C}_{\mathbf{h}}^{(u)}$, with appropriated power loading on each eigenvector. This is known in the literature as *eigen-beamforming* [3, 45]. The eigendecomposition of $\mathbf{C}_{\mathbf{h}}^{(u)}$ is

$$\mathbf{C}_{\mathbf{h}}^{(u)} = \mathbf{U}_u\mathbf{\Lambda}_u^2\mathbf{U}_u^H. \tag{7.55}$$

It presents a very similar structure to the *Gramian matrix* $\mathbf{H}\mathbf{H}^H$ from **Section 5.1.1**. $\mathbf{\Lambda}_u^2$ only contains L_u non-negligible eigenvalues. For this reason, we need to project the codewords onto the subspace spanned by their corresponding eigenvectors (or *eigenbeams*), with each component scaled by a suitable power loading coefficient d_l. The resulting transmitted block should be constructed as follows:

$$\mathbf{X}_u[i] \triangleq \sqrt{M_s^{(u)}P_{BS}^{(u)}}\overbrace{\tilde{\mathbf{U}}_u\mathbf{D}_u}^{\mathbf{G}_u}\mathbf{\Phi}_{L_u}\check{\mathbf{C}}_u[i], \tag{7.56}$$

where $P_{BS}^{(u)}$ is the transmit power per symbol and user. $\tilde{\mathbf{U}}_u$ contains the first L_u columns of $\mathbf{U}_u$ and $\mathbf{D}_u \triangleq \operatorname{diag}(d_1, \ldots, d_{L_u})$ is a diagonal matrix that contains the power loading coefficients. Together, they form the power loading matrix $\mathbf{G}_u$ commented in **Section 6.3.2**. We impose the constraint $\sum_{l=1}^{L_u} d_l^2 \equiv 1$ so that $\mathbf{X}_u[i]$ has constant power along the transmission:

$$\begin{aligned}
\operatorname{Tr}\left[\mathbf{X}_u[i]\mathbf{X}_u^H[i]\right] &= M_s^{(u)}P_{BS}^{(u)}\operatorname{Tr}\left[\mathbf{C}_u^H[i]\mathbf{\Phi}_{L_u}^H\mathbf{D}_u^H\tilde{\mathbf{U}}_u^H\tilde{\mathbf{U}}_u\mathbf{D}_u\mathbf{\Phi}_{L_u}\mathbf{C}_u[i]\right] \\
&= M_s^{(u)}P_{BS}^{(u)}\operatorname{Tr}\left[\mathbf{\Phi}_{L_u}^H\mathbf{D}_u^2\mathbf{\Phi}_{L_u}\mathbf{C}_u[i]\mathbf{C}_u^H[i]\right] = M_s^{(u)}P_{BS}^{(u)}\operatorname{Tr}\left[\mathbf{\Phi}_{L_u}\mathbf{\Phi}_{L_u}^H\mathbf{D}_u^2\right] \\
&= M_s^{(u)}P_{BS}^{(u)}\operatorname{Tr}\left[\mathbf{D}_u^2\right] = M_s^{(u)}P_{BS}^{(u)}\sum_{l=1}^{L_u} d_l^2 = M_s^{(u)}P_{BS}^{(u)}.
\end{aligned} \tag{7.57}$$

Thanks to the recursive nature of DSTM, the power loading is only required on the reference symbol, *i.e.*

$$\mathbf{X}_u[0] \triangleq \sqrt{M_s^{(u)}P_{BS}^{(u)}}\tilde{\mathbf{U}}_u\mathbf{D}_u\mathbf{\Phi}_{L_u}. \tag{7.58}$$

It is then easy to check that its effect is carried over to the rest of transmit blocks by simply applying the regular differential encoding ($\mathbf{X}[i] = \mathbf{X}[i-1]\mathbf{S}[i]$ for $0 < i < K_{DL} - K_b^{(u)}$). Without further eigenbeamforming operation, all the successive blocks are automatically transmitted along the eigenbeams. The eigendecomposition of $\mathbf{C}_{\mathbf{h}}^{(u)}$ should not be computationally complex since, under antenna selection, the covariance matrix of the channel has many null blocks. Besides, only the largest L_u eigenpairs are required, from a total of $N_{BS} >> L_u$.

The design of the power loading matrix $\mathbf{D}_u$ is one of the central aspects of this scheme, although it is not particularly sensitive, as it will be shown in the next section. In [26], the authors study the error probability of DSTM over spatially correlated P2P channels. By considering a simplified expression of

the SER, they are able to derive a closed form solution for the coefficients of $\mathbf{D}_u$ that minimize such approximate error rate:

$$\left[\mathbf{D}_{opt}^{(u)}\right]_{l,l} \triangleq \sqrt{\frac{1}{L_u} + \frac{2P_z}{P_{BS}^{(u)} \sin^2\left(\frac{\pi}{M}\right)} \left(\frac{1}{L_u} \sum_{\substack{l'=1 \\ l' \neq l}}^{L_u} \frac{1}{\lambda_{u,l'}^2} \right)}, \tag{7.59}$$

where $\lambda_{u,l}^2$ are the non-negligible eigenvalues of $\mathbf{C_h}^{(u)}$.

Alternatively, we propose a more heuristic option based on the power loading techniques from **Section 6.2**. This approach consists in loading each eigenvector with a weight proportional to its associated eigenvalue:

$$\left[\mathbf{D}_h^{(u)}\right]_{l,l} \triangleq \sqrt{\frac{\lambda_{u,l}^2}{\sum_{l'=1}^{L_u} \lambda_{u,l'}^2}} = \sqrt{\frac{\lambda_{u,l}^2}{\mathrm{Tr}\left[\mathbf{\Lambda}^2\right]}}. \tag{7.60}$$

This way, stronger eigenbeams are favored and can transmit symbols more reliably.

7.3 Numerical results

Now that we have fully defined a complete UL-DL system for LAA-MU-MIMO, it would be of interest to test its performance and compare it to other well-known coherent schemes. With that purpose in mind, we have designed a simulated communication environment that will help us assess our scheme under various circumstances. The main configuration of these tests is the following.

Consider a simultaneous full communication between $N_U = 5$ single-antenna MTs and a BS, composed of an UL transmission followed by a DL one. In order to provide service to the 5 users at once, the BS is equipped with an EM lens with normalized aperture $\tilde{D} = 100$, which translates into $N_{BS} = 201$ antenna elements. The LAA has $N_{RF} = 75$ RF chains available. Each MT can transmit with a power $P_{MT} = 1$ and the BS allocates the same amount of power $P_{BS} = 1$ per symbol and user. The constellation employed both in the UL and DL is 8-PSK ($M = 8$).

Setting parameter	Symbol	Value
Number of users	N_U	5
Number of antennas per MT	N_{MT}	1
Normalized lens aperture	$\tilde{D}$	100
LAA elements	N_{BS}	201
Available RF chains	N_{RF}	75
MT transmit power	P_{MT}	1
BS transmit power per user	P_{BS}	1
Signal constellation size	M	8

Table 7.1: Configuration of the communication equipment.

Due to the geometry of the physical setting, each user can connect to the BS through $L = 3$ paths, and channel reciprocity is expected between UL and DL ($\alpha_l^{UL} \equiv \alpha_l^{DL} \triangleq \alpha_l$ and $\tilde{\phi}_{BS,l} \equiv \tilde{\phi}_{MT,l} \triangleq \tilde{\phi}_l$). From the 3 available beams per user, one of them corresponds to a LoS link ($\alpha_1^{(u)} \sim \mathcal{CN}(0,1)$ while the other two are NLoS ($\alpha_2^{(u)}, \alpha_3^{(u)} \sim \mathcal{CN}(0,1/2)$).

Setting parameter	Symbol	Value
Number of beams per user	L	3
LoS channel gain	α_1	$\mathcal{CN}(0,1)$
NLoS channel gain	α_2, α_3	$\mathcal{CN}\left(0, \frac{1}{2}\right)$

Table 7.2: Channel characteristics.

The UL is divided in two phases: AS and data transmission. The AS training time is $T_{sel} = 100$ and the transmission time is $K_{UL} = 401$. We will compare our proposed DFDD+nDFE scheme against Vertical Bell Laboratories Layered Space-Time (V-BLAST) with both perfect and estimated CSI[6]. The channel estimation will be performed during the AS phase, with a sequence of orthogonal pilot signals per user of length $T_{train} = T_{sel} = 100$. We have chosen this coherent technique as a benchmark for our system because it is widely used in practice and clear analogies can be drawn between it and DFDD+nDFE. By default, both DFDD and nDFE will perform three independent iterations (#nDFE=#DFDD=3) on each space-time signal block. This way we will be able to observe the improvement in reliability between iterating 1, 2 or 3 times over every detected symbol/user. Regarding the weights matrix in DFDD, our main choice will be the optimized version we have derived in **Section 6.3.1**, although we will compare its performance against the diagonal version from [9] and a uniform weighting (*i.e.* $\mathbf{W}_u = \mathbf{I}$).

Setting parameter	Symbol	Value
UL transmission time	K_{UL}	401
AS/CSI acquisition time	T_{sel}	100
(Default) nDFE iterations	#nDFE	3
(Default) DFDD iterations	#DFDD	3
(Default) Weights matrix	$\mathbf{W}_u$	$\mathbf{W}_{opt}^{(u)}$

Table 7.3: UL configuration parameters.

The DL phase simply consists in a transmission of data during $K_{DL} = 400$. We are going to implement D-OSTBC with power loading adapted to each MT. Since the spatial dimension of the channel is $M_s = L = 3$, we can transmit 3 symbols per block. The duration of each signal block must be $K_b = 4$, so the transmission rate will be $R_b = \frac{3}{4}$ and $N_b = 100$ blocks will be transmitted. This means that, in practice, the BS will send $N_{sym}^{DL} = 300$ symbols per user during 400 time instants. We are going to analyze the effects of power loading by studying 3 approaches: uniform power loading ($\mathbf{D}_u = \mathbf{I}$), the proposal from [26] and our heuristic one.

Comparing this scheme with regular OSTBC is not truly fair, in the sense that in our scenario, only the BS has acquired CSI during the UL. Instead, we propose coherent beamforming (both with approximated and perfect CSI) as a benchmark candidate. It must be pointed out, however, that only 300 time instants are required to transmit 300 symbols with beamforming, since it is a full rate scheme.

[6]The perfect CSI schemes will employ the full LAA without AS. Their performance is meant to serve as a bound for the others.

Setting parameter	Symbol	Value
DL transmission time	K_{DL}	400
Symbol block size	K_b	4
Symbols per block	M_s	3
Symbol block rate	R_b	$\frac{3}{4}$
Transmitted blocks	N_b	100
Transmitted symbols	N_{sym}^{DL}	300

Table 7.4: DL configuration parameters.

In order to evaluate the reliability of each scheme, we will average the SER of each UL and DL over $N_{tst} = 3000$ simulations. In particular, on the UL, only 400 symbols will be accounted, since the first one is a reference and does not contain data. Similarly, on the DL, 297 are relevant in the transmission and the first block of 3 is the differential reference, as well. The SER will be plotted for various SNR values[7], between -20dB and 20dB. For the UL, SNR is defined as

$$\mathrm{SNR}_{UL} = \frac{1}{N_U} \sum_{u=1}^{N_U} \frac{P_{MT} \left\| \mathbf{h}_u \right\|^2}{P_z N_{BS}} = \frac{1}{N_U} \frac{\left\| \mathbf{H} \right\|_F^2}{P_z N_{BS}}, \tag{7.61}$$

since $P_{MT} = 1$. It can be seen as the average of the SNR for each user at every simulation. As for the DL, we have two different definitions. For the beamforming schemes, SNR is defined like in the UL, while for the D-OSTBC ones,

$$\mathrm{SNR}_{DL} = \frac{1}{N_U} \sum_{u=1}^{N_U} \frac{K_b P_{BS} \left\| \mathbf{h}_u \right\|^2}{M_s P_z N_{BS}} = \frac{1}{N_U} \frac{\left\| \mathbf{H} \right\|_F^2}{R_b P_z N_{BS}} \equiv \frac{\mathrm{SNR}_{UL}}{R_u} \tag{7.62}$$

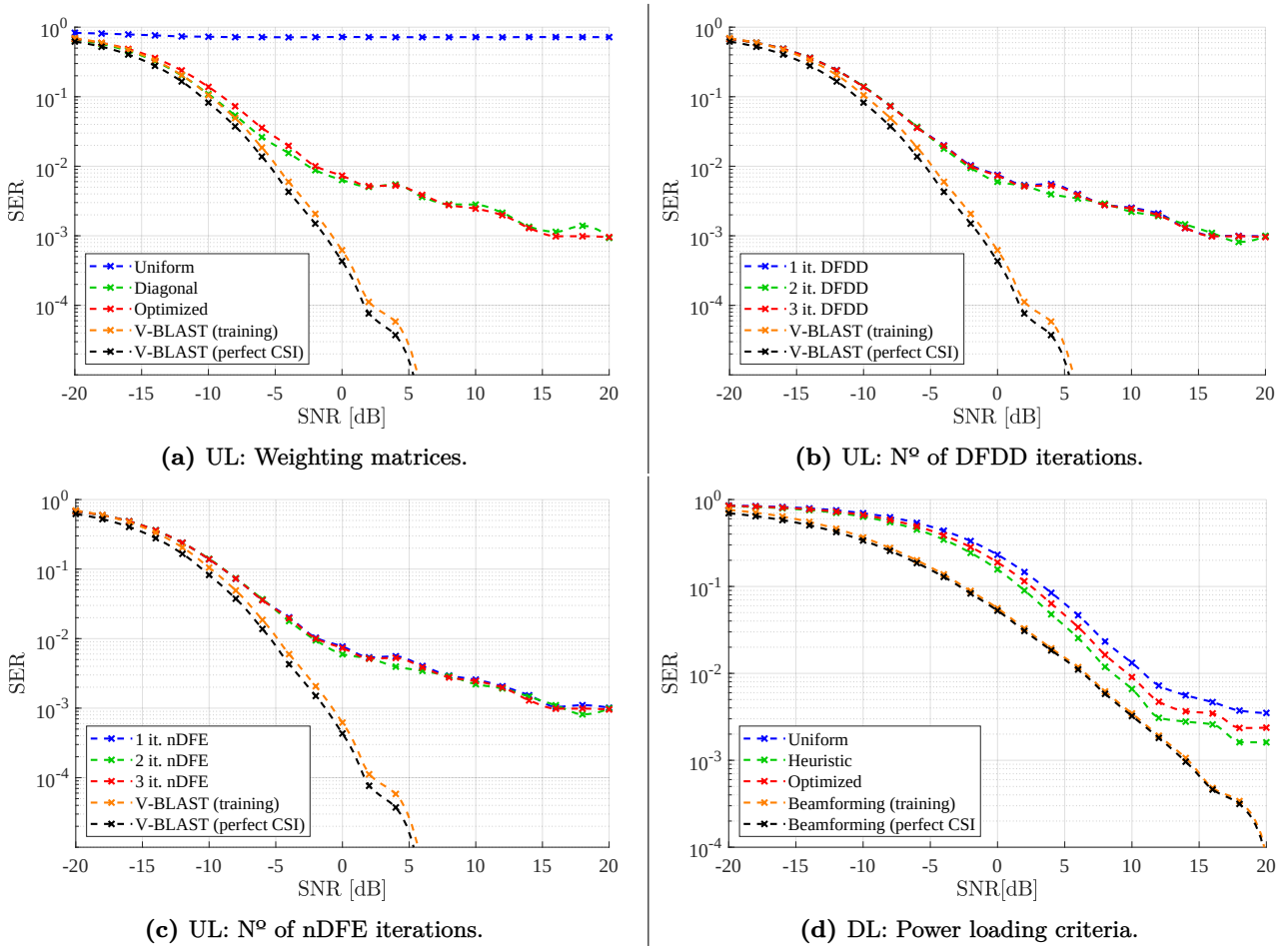
7.3.1 Ideal scenario

In the first test, we have set nominal operation conditions, which is the reason why we refer to it as *ideal*. It will serve as the baseline for the other two tests. The channel coefficients remain constant during both the UL and DL and there is full channel reciprocity. AoAs are chosen randomly across the aperture of the LAA. Since, at the moment, we are not interested in observing the effect of strong IUI, we have constrained this selection in a way so that two consecutive beams impinge the array with a minimum separation of 5 elements (with misalignment considered). As the 5 users have the same communication conditions, we have averaged their relative SERs in order to provide a cleaner graphical representation in Figures 7.4.

Let us comment on the UL results (Figures 7.4a, 7.4b and 7.4c). In general, both coherent and noncoherent schemes show similar performances at low SNR. On the contrary, as the SNR increases, the SER curves of DFDD+nDFE start flattening and a very prominent difference in performance between coherent and noncoherent schemes is present. This behavior is common in DFDD-based communications [8, 23, 24, 48].

In Figure 7.4a, three weighting approaches of the DFDD are considered. At a first glance, it is clear that uniform weighting is not a useful choice and results in unreliable transmissions at both low and high SNRs. The reasoning behind this phenomenon is that the weighted correlation matrix captures too much noise and interference and does not favor the signal contributions. In contrast with the results obtained in **Section 6.3.1**, it must be pointed out that there is no substantial difference in SER between the diagonal and optimized weightings. This discrepancy can be easily explained by

[7] SNR at the Rx.



(a) UL: Weighting matrices.

(b) UL: Nº of DFDD iterations.

(c) UL: Nº of nDFE iterations.

(d) DL: Power loading criteria.

Figure 7.4: UL and DL performance results in an ideal scenario.

the fact that, in this practical scheme, we have discarded the crossed terms in the channel covariance matrices to provide robustness against fading. Therefore, this strategy has lost the advantage it had for considering constant channel coefficients. In the low SNR region, the diagonal weighting slightly outperforms the optimized one, and it is not until the SNR is high that the off-diagonal terms start providing some gain.

Figures 7.4b and 7.4c present the performance gains of iterating more than once the DFDD and nDFE algorithms, respectively. Since IUI is low, due to the directional nature of each user's beams and the AoA spacing we have enforced, neither executing additional nDFE nor DFDD rounds provides any substantial increase in performance. Almost all the achievable reduction of SER is obtained on the first iteration for both algorithms.

Finally, Figure 7.4d displays the DL SER obtained with D-OSTBC schemes with power loading and Tx beamforming schemes. The curves corresponding to coherent methods presents a similar shape to those from the UL. There is almost no difference in performance between the estimated and perfect CSI. Regarding the D-OSTBC schemes, their curves flatten at higher SNR than those corresponding to noncoherent schemes in the UL, and start diverging from the coherent ones at around 12dB.

As expected, uniform power loading is the least effective policy, since it distributes power equally among dominant eigenbeams and does not fully exploit the available statistical CSI. The loss in performance compared to the other two approaches, however, is not significant. Surprisingly, our heuristic power loading proposal slightly outperforms the optimized method, especially at high SNRs.

This may be explained by the fact that the optimized method was conceived to minimize the SER in a P2P setting and the effect of IUI was not taken into consideration in its design.

The conclusions we can draw from this test are summarized in the following points:

1. Weighting based on statistical CSI is fundamental for the correct operation of DFDD.
2. When dealing with constant gain channels, the optimized weighting method for DFDD can provide a substantial gain against more agnostic approaches. The crossed terms of each user's covariance matrix can be estimated like the diagonal ones, with the curve fitting method proposed in **Section 7.1.2**.
3. If the crossed terms of the covariance matrix cannot be employed when constructing the optimized weighting matrix, this method brings negligible gain against diagonal weighting.
4. When users' signal beams are sufficiently spaced, a single round of nDFE+DFDD suffices to obtain most of the achievable SER reduction. Further processing should be employed if lower rates are required.
5. The heuristic power loading policy we have presented should be the method of choice in multi-user D-OSTBC environments for its improved performance. Even so, this election is not critical for the operation of the DL scheme.

7.3.2 Path overlap

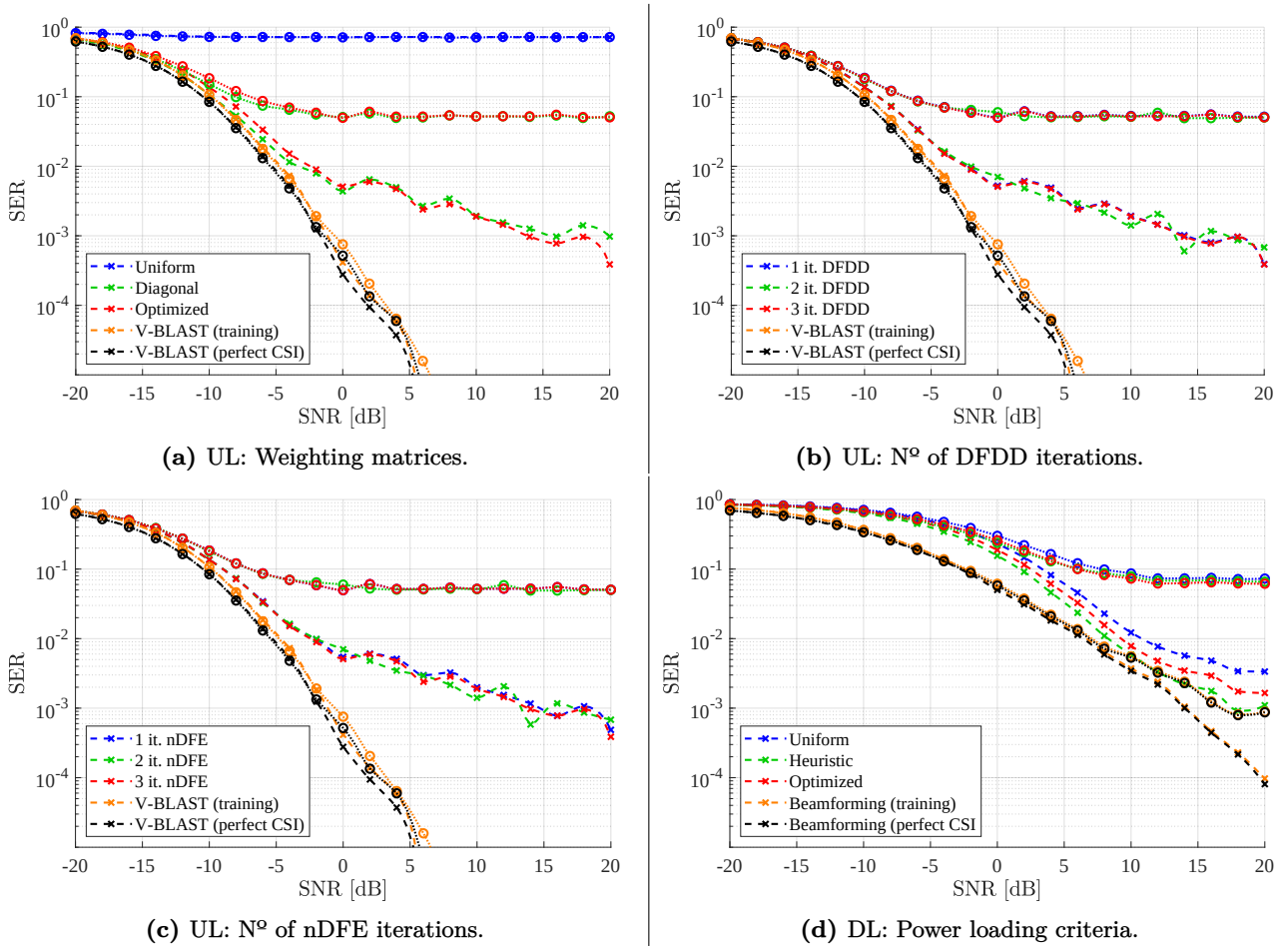
The second test we have implemented is similar to the first one but some MTs experience strong IUI. In particular, the LoS beams of User 1 and User 2 impinge the LAA on antennas separated by just between 1 and 3 indices. Despite the energy focusing capabilities of the EM lens, a considerable amount of interference corrupts both users' signals. For the sake of visual clarity in Figures 7.5, we have averaged the SERs of users in equal communication conditions; in other words, users 1 and 2 form a group (dotted lines with hollow circles) and users 3, 4 and 5 form another one (dashed lines with crosses).

Unsurprisingly, the V-BLAST schemes do not present significant issues providing reliable communications in the presence of strong IUI. Notice, however, that regular beamforming in the DL presents a flattened SER curve at high SNR because it does not employ any kind of interference suppression method.

On the contrary, neither of the noncoherent schemes can provide reliable communications at high SNR for the MTs that suffer strong IUI. At some point (around 0dB for the UL schemes, and around 10dB for the DL ones), their SER curves reach a floor and do not decrease further.

Regarding the UL phase, the three weighting alternatives behave as expected from the previous test. For the various amounts of DFDD and nDFE rounds, we encounter two distinct cases. On the one hand, for MTs with low IUI, adding more iterations does not provide better SERs, as commented for the previous test. On the other hand, for MTs with high IUI, the same result applies but for the opposite reason. The interference is so strong that nDFE cannot cancel it. It is clear from [24, 48], that nDFE rounds are effective at increasing the scheme reliability in moderate interference environments. Large mmWave LAA channels present a very different behavior, due to the high directivity of their response vectors: *sinc* squared profiles concentrate the majority of their power on a central narrow lobe. Most of the time, users encounter either very high IUI (overlapping beams) or very low IUI (well spaced beams). In neither of those cases nDFE can improve the performance.

In the DL phase, a similar phenomenon occurs, since the D-OSTBC technique has not been equipped with any measure against interference. The three power loading policies behave as in the previous test.



(a) UL: Weighting matrices.

(b) UL: Nº of DFDD iterations.

(c) UL: Nº of nDFE iterations.

(d) DL: Power loading criteria.

Figure 7.5: UL and DL performance results for a scenario in which users 1 and 2 have high IUI (dotted line, hollow circles), while users 3, 4 and 5 have low IUI (dashed line, crosses).

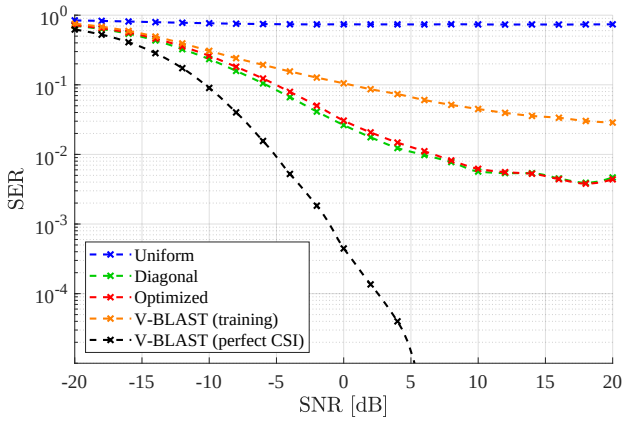
This test has allowed us to draw some remarkable conclusions about the operation of our schemes in presence of strong IUI:

1. It is clear that none of the proposed communication methods are able to provide reliable data transmission under strong interference. Further processing and more involved techniques will be required.

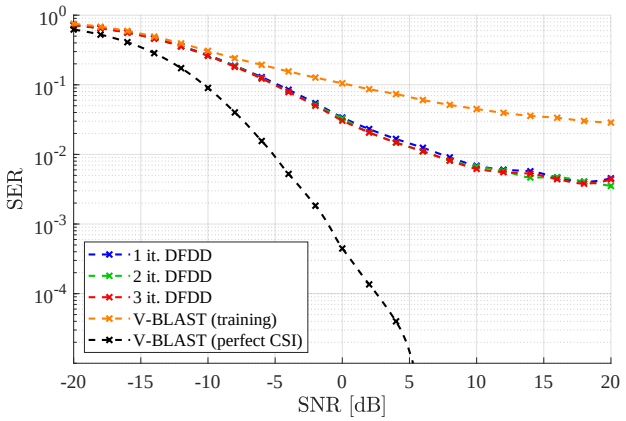
2. The performance of simple nDFE+DFDD does not improve regardless of how many times it is iterated. Therefore, a single round provides most of its achievable gain in low and high IUI scenarios.

7.3.3 Block fading

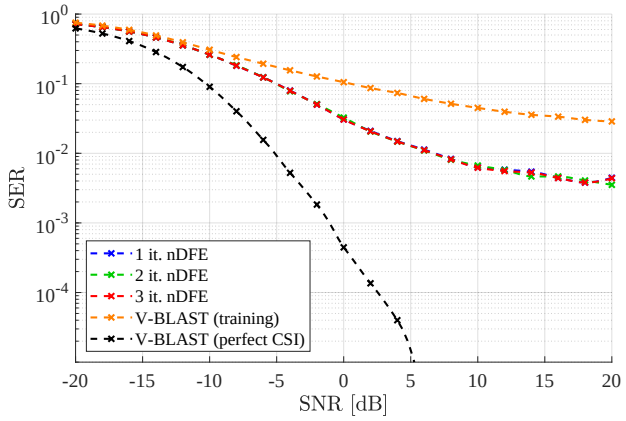
The objective of the final test is to observe how the different presented schemes behave in the presence of *block fading* [34]. Up until this point, we have considered communications in settings in which the channel coefficients remain constant during a full transmission. In practical scenarios, this might not be possible. If we relied on full CSI, we would have to limit each data transmission to fit in the so-called *coherence time* of the channel. After that, we could not guarantee its coefficients to be the same and would require CSI acquisition to reliably continue the communication. This situation is very limiting for coherent schemes, especially if the fading is fast.



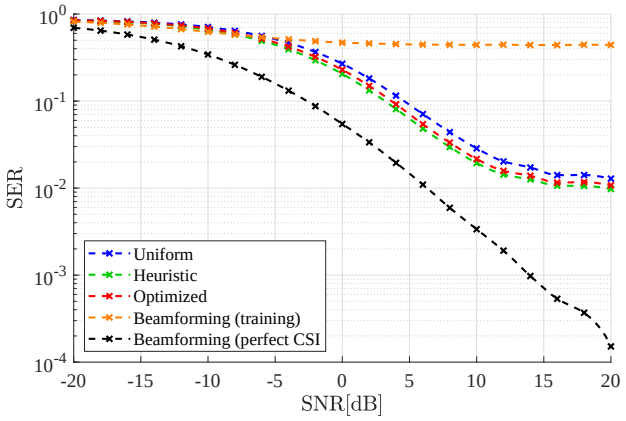
(a) UL: Different weighting matrices.



(b) UL: Different DFDD nº of iterations.



(c) UL: Different nDFE nº of iterations.



(d) DL: Different power loading criteria.

Figure 7.6: UL and DL performance results for a block fading scenario (phase and power gain change).

In our simulated test, the channel experiences block fading once in the middle of the UL and once in the middle of the DL. At those points, both power gain and phase of the channel coefficients will change suddenly, while their AoAs and AoDs will be maintained. Under this model, half of the symbols (on both UL and DL) will be transmitted through a channel that the BS will have estimated (coherently or statistically) and the other half through a completely unknown channel (except for the AoAs and AoDs). We consider low IUI once again.

As expected, V-BLAST that relies on channel estimation suffers a remarkable degradation in its performance, as seen in Figures 7.6. It can only decode reliably half of the transmitted symbols, since it does not have any information regarding the channel coefficients of the second half of the transmission. This issue is even more prominent during the DL, because regular beamforming is very vulnerable to fading.

On the contrary, by not relying on full CSI, the noncoherent schemes are much more robust against block fading. While the performance of DFDD+nDFE is moderately degraded compared to previous tests with no fading, it can still transmit data reliably as SNR increases. This loss in performance can be easily explained. The channel coefficients change in between the transmission of two symbols $x_u[k]$ and $x_u[k+1]$, making the first be affected by a channel and the second by a different one. This means that the deterministic component of the differential phase between them is lost, producing a detection error on the information encoded in $x_u[k+1]$. This first symbol transmitted within the new channel, however, virtually becomes a new reference for the subsequent ones.

As previously stated, the optimized weights matrix for DFDD is computed assuming channel fading, which is an appropriate assumption in this test. Unsurprisingly, there is no difference in performance with respect to the diagonal weighting.

Regarding the DL, the SER degradation is moderate as well. Notice in Figure 7.6d that the three power loading approaches present little difference in performance. While on the first half of the DL, the policies that effectively account for the statistical CSI are undoubtedly superior to those which do not (*i.e.* uniform loading), this advantage does not longer hold when the channel changes. In fact, uneven power loading can become detrimental. After block fading, the eigenvalues of the channel covariance matrix might have changed. Adapting the transmitted power to the eigenvalues distribution of the previous channel becomes suboptimal in the new one, in general. Indeed, the heuristic and optimized power loading methods perform almost no better than with uniform loading, in average.

To sum it up, from this final test we deduce the following remarks:

1. One of the greatest strengths of noncoherent schemes is their robustness against block fading. The more agnostic a scheme is in terms of statistical CSI, the more consistent its performance is across different communications.
2. When we can ensure a channel will be invariant during a whole data transmission, the weighting matrix of choice for DFDD should be the optimized one. In the multi-user case, the crossed terms of the covariance matrix should be estimated with an improved version of the curve fitting method presented in **Section 7.1.2**.
3. When there is risk that block fading might occur in the middle of a transmission, the weighting matrix of choice should be the diagonal one. It is computationally simpler to implement than the optimized one and provides negligible loss in performance.
4. Similarly, when choosing a power loading policy for D-OSTBC, the possibility of channel fading must be taken into account. In constant channels, approaches that employ statistical CSI should be selected, whereas in block fading channels, no eigenvalue distribution should be prioritized over the rest and uniform loading should be the default option.

Chapter 8

Conclusions and future research

The thesis presented in this document has been the result of over four months of intensive and very rewarding work. A satisfactory way of bringing it to a close would be to get the full picture of what has been achieved.

In **Chapter 3** we have presented the beamspace representation of a wireless MIMO channel. Thanks to the sparse scattering in mmWave propagation, this domain transformation through a spatial DFT allows us to convert a regular channel matrix into its much simpler beamspace domain counterpart. The clear low rank displayed by this representation hints at a simplified RF infrastructure to process signals transmitted within millimetric bands. Still, traditional transceiver architectures have no choice but to go over the full antenna array in order to garner all the benefits from the beamspace domain, which becomes unfeasible to implement in M-MIMO due to power and cost constraints. It is at this point of the common thread of the thesis in which EM lenses come into play. Their angular-dependent power focusing capabilities allow the separation of waves with different AoAs. This property effectively approximates an analog Fourier transform without the explicit signal processing and RF chains required otherwise, hence motivating the application of LAAs in mmWave communications.

In **Chapter 4**, we have discussed some aspects about the physical implementation of a LAA system. Afterwards, we have rigorously derived an expression for its array response vector from a spherical wave model. While we have acknowledged that the plane wave approximation is not an accurate representation of a real mmWave setting, we have adopted it for its simplicity and usefulness in signal processing. We have also assessed the power leakage problem inherent to LAA systems and briefly quantified its impact. Subsequently, we have obtained the expression of the symmetric LAA-MIMO channel in **Chapter 5**, based on well established propagation models. With it, we have characterized its capacity in the low and high SNR regimes and investigated how it is affected by different parameters.

After the more theory-centered sections of the thesis, we have applied the studied ideas to practical communication problems. In **Chapter 6**, we have considered the P2P setting between a BS and a MT, both ways (UL and DL). We have integrated the LAA in traditional and mixed schemes within the two general coherent scenarios (full CSIT and full CSIR), and also in noncoherent schemes, which has been the main goal of the thesis. For the UL, we have taken the basic MSDD architecture and generalized its working principles to better exploit well-structured channels. The obtained solution shows important performance gains against more agnostic approaches when no fading is present. Regarding the DL, we have adapted D-OSTBC to the LAA-MISO channel.

In **Chapter 7**, we have designed a full UL-DL system based on the presented noncoherent schemes and statistical CSI. We have addressed AS in LAA systems and proposed a procedure for multi-user settings that fairly allocates the available resources among the connected users. We have also suggested a way to approximate each user's individual channel covariance matrix from the combination of all the received signals at the BS. In the extent of the UL noncoherent detection, we have applied the results from the previous chapter to DFDD, as well as derived nDFE metrics better suited for LAA. We have used power loading together with D-OSTBC in the DL and introduced a heuristic weighting policy.

Last but not least, we have subjected our designed system to various numerical tests and compared its performance to V-BLAST and beamforming, both coherent schemes. The conclusions drawn from these simulations have been very insightful. When coherent schemes can effectively estimate the CSI (*i.e.* ideal propagation conditions), their performance is unbeatable by noncoherent methods. This

difference becomes especially vast when trying to cancel strong IUI. The advantages of noncoherent schemes, however, start to surface in the presence of block fading. Since they do not rely on full CSI, they can provide sound communications even in changing channels. On the contrary, coherent schemes become completely unreliable when they are not able to safely estimate the channel. These observations suggest the use of noncoherent LAA schemes in high mobility scenarios, such as vehicular communications, or in very populated networks in which orthogonal training sequences cannot be allocated for every connected user.

The studies conducted in this thesis have helped slightly broaden the scopes of LAA-M-MIMO and noncoherent communications, and open the door to many potential lines of research. Let us enumerate a few of them:

- In reference to the spherical wave model presented in **Section 4.1**, the complete array response vector presents a *windowing effect* that depends on the distance between Rx and Tx. In [20], the authors study potential applications of this behavior in user position sensing and regular communications. Following the philosophy of the thesis, it would be of interest to investigate whether there are some potential advantages in using this property to enhance noncoherent schemes.
- While radar technologies have not been considered in this thesis, they are a very active area of research. As seen in [58], there is room for application of LAAs and, in particular, for the spherical wave model.
- The use of LAAs with AS is very well suited for designs of M-MIMO transceivers with a special focus on energy efficiency [4].
- As commented in **Section 7.2**, general FDD schemes cannot rely on channel reciprocity between UL and DL and must estimate the propagation statistics at every stage. The training overhead required for these CSI acquisitions can become unacceptably high for M-MIMO systems. A very promising technique to overcome this limitation is called *covariance conversion* [59]: the idea of estimating the DL channel covariance matrix from the UL one. The LAA channel seems very appropriate for the implementation of such methods, thanks to its rigid structure and path reciprocity. The *beam squint* phenomenon between different frequency bands should be carefully studied, more so considering AS reuse between UL and DL (see **Section 7.1.1**).
- The user assignment method based on second order statistics, briefly stated in **Section 7.1.1**, may be properly expanded and developed. Of special interest is its joint implementation within the AS phase.
- Although the gains obtained with our optimized DFDD weighting method have been displayed, we have not fully assessed the performance of the optimized nDFE metrics nor of the heuristic D-OSTBC power loading. A rigorous analysis should be conducted on the two.
- The method to estimate the covariance matrix of each individual user can be further improved with a curve fitting technique on cross channel components. This would allow for a better performance in channels without fading during the transmission of data.
- As it has become very clear in **Section 7.3**, the implementations we have proposed cannot provide reliable communications in presence of strong IUI. Further research is needed in that regard. A very promising starting point for better interference cancellation are *subspace-based methods* [60–62].
- Finally, an alternative to orthogonal training pilots in multi-user environments is the so-called *Non-Orthogonal Multiple Access (NOMA)* [63]. Its joint application within the LAA paradigm

could greatly benefit from the special characteristics of this wireless channel.

With all this in mind, we expect that this Master's Thesis has served to slightly push forward the field of signal processing in view of the next generation of wireless communications.

Bibliography

[1] K. Hassan, M. Masarra, M. Zwingelstein, and I. Dayoub, "Channel Estimation Techniques for Millimeter-Wave Communication Systems: Achievements and Challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1336–1363, 2020.

[2] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[3] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.

[4] K. N. R. S. V. Prasad, E. Hossain, and V. K. Bhargava, "Energy Efficiency in Massive MIMO-Based 5G Networks: Opportunities and Challenges," *IEEE Wireless Communications*, vol. 24, no. 3, pp. 86–94, 2017.

[5] M. R. Akdeniz *et al.*, "Millimeter Wave Channel Modeling and Cellular Capacity Evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[6] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for Millimeter-Wave Communications: System Architecture, Modeling, Analysis, and Measurements," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 7, pp. 3814–3827, 2013.

[7] Y. Zeng and R. Zhang, "Millimeter Wave MIMO With Lens Antenna Array: A New Path Division Multiplexing Paradigm," *IEEE Transactions on Communications*, vol. 64, no. 4, pp. 1557–1571, 2016.

[8] S. Bucher, G. Yammine, R. F. H. Fischer, and C. Waldschmidt, "A Noncoherent Massive MIMO System Employing Beamspace Techniques," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 11 052–11 063, 2019.

[9] G. Yammine, S. Bucher, and R. F. H. Fischer, "Noncoherent Detection for an EM-Lens-Enabled Massive MIMO System," in *SCC 2019; 12th International ITG Conference on Systems, Communications and Coding*, 2019, pp. 1–6.

[10] S. Bucher and C. Waldschmidt, "Advanced Noncoherent Detection in Massive MIMO Systems via Digital Beamspace Preprocessing," *Telecom*, vol. 1, no. 3, pp. 211–227, 2020.

[11] L. Yang, Y. Zeng, and R. Zhang, "Channel Estimation for Millimeter-Wave MIMO Communications With Lens Antenna Arrays," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 4, pp. 3239–3251, 2018.

[12] J. Thornton and K.-C. Huang, *Modern Lens Antennas for Communications Engineering*. John Wiley and Sons, Ltd, 2013.

[13] A. Sayeed and J. Brady, "Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies," in *2013 IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 3679–3684.

[14] A. Sayeed and N. Behdad, "Continuous aperture phased MIMO: Basic theory and applications," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010, pp. 1196–1203.

[15] L. Dai, X. Gao, S. Han, I. Chih-Lin, and X. Wang, "Beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," in *2016 IEEE/CIC International Conference on Communications in China (ICCC)*, 2016, pp. 1–6.

[16] X. Gao, L. Dai, S. Zhou, A. M. Sayeed, and L. Hanzo, "Wideband Beamspace Channel Estimation for Millimeter-Wave MIMO Systems Relying on Lens Antenna Arrays," *IEEE Transactions on Signal Processing*, vol. 67, no. 18, pp. 4809–4824, 2019.

[17] Y. Zeng and R. Zhang, "Cost-Effective Millimeter-Wave Communications with Lens Antenna Array," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 81–87, 2017.

[18] Y. Zeng, L. Yang, and R. Zhang, "Multi-User Millimeter Wave MIMO With Full-Dimensional Lens Antenna Array," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2800–2814, 2018.

[19] J.-N. Shim, H. Park, C.-B. Chae, D. K. Kim, and Y. C. Eldar, "Cramér–Rao Lower Bound on AoA Estimation Using an RF Lens-Embedded Antenna Array," *IEEE Antennas and Wireless Propagation Letters*, vol. 17, no. 12, pp. 2359–2363, 2018.

[20] J. Yang, Y. Zeng, S. Jin, C.-K. Wen, and P. Xu, "Communication and Localization With Extremely Large Lens Antenna Array," *IEEE Transactions on Wireless Communications*, vol. 20, no. 5, pp. 3031–3048, 2021.

[21] A. Schenk and R. F. H. Fischer, "Noncoherent Detection in Massive MIMO Systems," in *WSA 2013; 17th International ITG Workshop on Smart Antennas*, 2013, pp. 1–8.

[22] A. Schenk, "Coding, Modulation, and Detection for Impulse-Radio Ultra-Wideband Communications," Ph.D. dissertation, Der Technischen Fakultät der Friedrich-Alexander-Universität Erlangen-Nürnberg zur Erlangung des Grades, Erlangen, 2013.

[23] R. F. H. Fischer and M. Bense, "Noncoherent Decision-Feedback Equalization in Massive MIMO Systems," in *International Zurich Seminar on Communications proceedings*, Mar. 2014, pp. 112–115.

[24] C. Stierstorfer, R. F. H. Fischer, and G. Yammine, "Iterative Decision-Feedback in Non-Coherent Multi-User Massive MIMO Systems," in *WSA 2015; 19th International ITG Workshop on Smart Antennas*, 2015, pp. 1–8.

[25] G. Yammine and R. F. Fischer, "Decision-Feedback Differential Detection with Optimum Detection Order Metric for Noncoherent Massive MIMO Systems," in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1658–1662.

[26] X. Cai and G. Giannakis, "Differential space-time modulation with eigen-beamforming for correlated MIMO fading channels," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1279–1288, 2006.

[27] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An Overview of Massive MIMO: Benefits and Challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[28] F. Rusek *et al.*, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[29] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, 1st. USA: Cambridge University Press, 2008.

[30] J. Liu and A. Host-Madsen, "Novel communication schemes with blind channel estimation in TDD MIMO system," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2004, pp. ii–33.

[31] E. G. Larsson, P. Stoica, and G. Ganesan, *Space-Time Block Coding for Wireless Communications*. USA: Cambridge University Press, 2003.

[32] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 68–73, 2004.

[33] C. Xu *et al.*, "Sixty Years of Coherent Versus Non-Coherent Tradeoffs and the Road From 5G to Wireless Futures," *IEEE Access*, vol. 7, pp. 178 246–178 299, 2019.

[34] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. USA: Cambridge University Press, 2005.

[35] G. H. Song, J. Brady, and A. Sayeed, "Beamspace MIMO transceivers for low-complexity and near-optimal communication at mm-wave frequencies," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 4394–4398.

[36] J. Goodman, *Introduction to Fourier Optics*, ser. Electrical Engineering Series. McGraw-Hill, 1996.

[37] Y. Zeng, R. Zhang, and Z. N. Chen, "Electromagnetic Lens-Focusing Antenna Enabled Massive MIMO: Performance Improvement and Cost Reduction," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1194–1206, 2014.

[38] J. H. Brady and A. M. Sayeed, "Differential beamspace MIMO for high-dimensional multiuser communication," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2015, pp. 310–314.

[39] T. Kwon, Y.-G. Lim, and C.-B. Chae, "Limited channel feedback for RF lens antenna based massive MIMO systems," in *2015 International Conference on Computing, Networking and Communications (ICNC)*, 2015, pp. 6–10.

[40] T. Kwon, Y.-G. Lim, B.-W. Min, and C.-B. Chae, "RF Lens-Embedded Massive MIMO Systems: Fabrication Issues and Codebook Design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 7, pp. 2256–2271, 2016.

[41] X. Cheng, Y. Yang, B. Xia, N. Wei, and S. Li, "Sparse Channel Estimation for Millimeter Wave Massive MIMO Systems With Lens Antenna Array," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 11 348–11 352, 2019.

[42] L. Yang, Y. Zeng, and R. Zhang, "Efficient channel estimation for millimeter wave MIMO with limited RF chains," in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.

[43] T. Xie, L. Dai, D. W. K. Ng, and C.-B. Chae, "On the Power Leakage Problem in Millimeter-Wave Massive MIMO With Lens Antenna Arrays," *IEEE Transactions on Signal Processing*, vol. 67, no. 18, pp. 4730–4744, 2019.

[44] R. Ertel, P. Cardieri, K. Sowerby, T. Rappaport, and J. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Personal Communications*, vol. 5, no. 1, pp. 10–22, 1998.

[45] M. A. Jensen and J. W. Wallace, *Space-Time Processing for MIMO Communications*. John Wiley and Sons, Ltd, 2005.

[46] J. Wang, P. Ding, M. Zoltowski, and D. Love, "Space-time coding and beamforming with partial channel state information," in *GLOBECOM '05. IEEE Global Telecommunications Conference, 2005.*, vol. 5, 2005, 5 pp.–3153.

[47] W. Meng, L. Gu, and C. Li, "The combined beamforming and space-time block coding technique for downlink transmission," in *2005 International Conference on Wireless Networks, Communications and Mobile Computing*, vol. 1, 2005, 481–486 vol.1.

[48] R. F. Fischer, M. Bense, and C. Stierstorfer, "Noncoherent Joint Decision-Feedback Detection in Multi-User Massive MIMO Systems," in *WSA 2014; 18th International ITG Workshop on Smart Antennas*, 2014, pp. 1–8.

[49] R. Rosenfelder, *Path Integrals in Quantum Physics*, 2017. arXiv: `1209.1315`.

[50] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1997.

[51] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[52] K. Wu, W. Ni, T. Su, R. P. Liu, and Y. J. Guo, "Exploiting Spatial-Wideband Effect for Fast AoA Estimation at Lens Antenna Array," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 902–917, 2019.

[53] F. Alsifiany, A. Ikhlef, M. Alageli, and J. Chambers, "Differential Downlink Transmission in Massive MU-MIMO Systems," *IEEE Access*, vol. 7, pp. 86 906–86 919, 2019.

[54] T. M. Duman and A. Ghrayeb, *Coding for MIMO Communication Systems*. 2008.

[55] M. Jankiraman, *Space-time Codes and MIMO Systems*, ser. Artech House universal personal communications series. Artech House, 2004.

[56] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *1998 URSI International*

Symposium on Signals, Systems, and Electronics. Conference Proceedings (Cat. No.98EX167), 1998, pp. 295–300.

[57] L. Isserlis, "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika*, vol. 12, no. 1/2, pp. 134–139, 1918.

[58] Z.-M. Jiang, P. Zhang, L. Huang, J. Zhang, X. He, and M. Rihan, "Lens Antenna Arrays Aided Co-Existing Radar and Communication Systems With Energy Harvesting," *IEEE Access*, vol. 8, pp. 56 160–56 169, 2020.

[59] L. Miretti, R. L. Cavalcante, and S. Stanczak, "FDD Massive MIMO Channel Spatial Covariance Conversion Using Projection Methods," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 3609–3613.

[60] R. F. H. Fischer, C. Stierstorfer, and R. R. Mueller, "Subspace Projection and Noncoherent Equalization in Multi-User Massive MIMO Systems," in *SCC 2015; 10th International ITG Conference on Systems, Communications and Coding*, 2015, pp. 1–6.

[61] G. Yammine and R. F. Fischer, "Performance/complexity comparison of SVD vs. SURV for subspace estimation in noncoherent massive MIMO," in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, 2017, pp. 234–239.

[62] ——, "Optimization of Subspace Projection in Noncoherent Massive MIMO Systems," in *SCC 2019; 12th International ITG Conference on Systems, Communications and Coding*, 2019, pp. 1–6.

[63] F. Santagiustina Lucangeli, "Multiuser random beamforming in mmWave channels," M.S. thesis, Escola Tècnica d'Enginyeria de Telecomunicació de Barcelona, Universitat Politècnica de Catalunya, Barcelona, 2019.

Appendix A

Matlab Code

In this annex, we have gathered all the Matlab code implementations we have utilized throughout the simulations, based on the theoretical foundations of the thesis. The code has been thoroughly commended inline (green color) for a clear understanding of its operation. It has been organized into functions.

1. **Perform AS and assign antennas to each user following the fairness policy:**

```
function [Wsel_u, Wsel, phi_est] = select_and_assign(Y, nRx, phi_real, Nu, ...
    Rsel, Nrf, L, Dsel, coh)
%   Inputs:
%       Y       : Training sequence (space-time block).
%       nRx     : Antenna indices.
%       phi_real: Real AoAs.
%       Nu      : Nº of users.
%       Rsel    : Effective selection time.
%       Nrf     : Nº of available RF chains.
%       L       : Nº of paths per user.
%       Dsel    : Energy threshold.
%       coh     : Coherent/noncoherent flag.
%   Outputs:
%       Wsel_u  : Selection matrix per user.
%       Wsel    : Global selection matrix.
%       phi_est : Estimated AoAs.
    Nrx = length(nRx);              % Nº of antennas.
    Nrf_u = floor(Nrf/Nu);          % Nº of RF/user.
    Npaths = Nu*L;                  % Total nº of paths.
    Wsel_u = zeros(Nrx, Nrx, Nu);   % Selection matrix/user.
    phi_est = zeros(Nu, L);         % Estimated directions indexes.
    %% Coherent detection (V-BLAST). %%
    if coh
        for u = 1:Nu    % For all users.
            % Diagonal of the selection matrix.
            w_sel = zeros(Nrx, 1);
            % Energy detection per user.
            y = abs(1/sqrt(Rsel)*sum(Y(:,u:Nu:end), 2)).^2;
            % Sort antennas by energy.
            [~, idx] = sort(y, 'descend');
            % Add selected antennas to user selection matrix.
            w_sel(idx(1:Nrf_u)) = 1;
            Wsel_u(:,:,u) = diag(w_sel);
        end
        % Global selection matrix.
        Wsel = any(Wsel_u, 3);
    %% Noncoherent detection (DFDD+nDFE). %%
```

```
    else
        % Energy detection.
        y = abs(1/sqrt(Rsel)*sum(Y, 2)).^2;
        % Detect beams: maximum value and associated antenna.
        [path_val, path_idx] = findpeaks(y, 'MinPeakHeight', Dsel);
        % Sort paths by energy.
        [~, sorted_paths] = sort(path_val, 'descend');
        if length(sorted_paths) >= Npaths
            sorted_paths = sorted_paths(1:Npaths);
        end
        path_idx = path_idx(sorted_paths);
        path_val = path_val(sorted_paths);
        % Assign paths and size of corresponding supports to each user.
        val_est = zeros(Nu, L); % Assigned energy per path.
        Nsup = zeros(Nu, L);    % Nº of supports per path per user.
        for u = 1:Nu    % For all users.
            for l = 1:L % For all paths.
                % Nearest detected AoA to real one.
                [~, idx] = min(abs(nRx(path_idx) - phi_real(u,l)));
                phi_est(u,l) = path_idx(idx);
                val_est(u,l) = path_val(idx);
            end
            % Nº of supports per path per user with energy criterion.
            Nsup(u,:) = floor(val_est(u,:)/sum(val_est(u,:))*Nrf_u);
            % RF chains left after assignment.
            extraRF = Nrf_u - sum(Nsup(u,:));
            % Assign RF chains left to strongest path.
            if extraRF
                [~, max_path] = max(val_est(u,:));
                Nsup(u,max_path) = Nsup(u,max_path) + extraRF;
            end
        end
        % Create and assign supports around AoAs.
        for u = 1:Nu    % For all users.
            for l = 1:L % For all beams.
                % Counters of the size of the support on both sides
                % of the AoA.
                Lcount = 1;     % Left support expander.
                Rcount = 1;     % Right support expander.
                % Flags that indicate if the support can keep expanding
                % to the left and to the right.
                Lkeep = true;   % Left expander flag.
                Rkeep = true;   % Right expander flag.
                % Add estimated AoAs antennas to selection matrix.
                if Nsup(u,l)
                    Wsel_u(phi_est(u,l),phi_est(u,l),u) = 1;
                    Nsup(u,l) = Nsup(u,l) - 1;
                end
                % While there are available RF chains for path l and can
```

```
                % keep expanding to the left or to the right.
                while (Nsup(u,l) && (Lkeep || Rkeep))
                    % Add support to the left.
                    idx = phi_est(u,l) - Lcount;
                    % If path l has RF chains left, can keep adding
                    % to the left, the antenna index is in bounds and
                    % there is no overlapping with a different AoA.
                    if Nsup(u,l) && Lkeep && (idx > 0) && ...
                            ~any(idx == phi_est, 'all')
                        % Add support antenna to selection matrix.
                        Wsel_u(idx,idx,u) = 1;
                        Nsup(u,l) = Nsup(u,l) - 1;
                        % Move counter to the left.
                        Lcount = Lcount + 1;
                    else
                        % Cannot keep adding support antennas to the left.
                        Lkeep = false;
                    end
                    % Add support to the right.
                    idx = phi_est(u,l) + Rcount;
                    % If path l has RF chains left, can keep adding
                    % to the right, the antenna index is in bounds and
                    % there is no overlapping with a different AoA.
                    if Nsup(u,l) && Rkeep && (idx <= Nrx) && ...
                            ~any(idx == phi_est, 'all')
                        % Add support antenna to selection matrix.
                        Wsel_u(idx,idx,u) = 1;
                        Nsup(u,l) = Nsup(u,l) - 1;
                        % Move counter to the right.
                        Rcount = Rcount + 1;
                    else
                        % Cannot keep adding support antennas to the right.
                        Rkeep = false;
                    end
                end
                % Assign spare RF to next strongest path.
                if Nsup(u,l) && l < L
                    [~, max_path] = max([NaN(1, l), ...
                        val_est(u,(l + 1):end)], [], 'omitnan');
                    Nsup(u,max_path) = Nsup(u,max_path) + Nsup(l);
                end
            end
        end
        % Global selection matrix.
        Wsel = any(Wsel_u, 3);
    end
end
```

2. **Estimate the individual covariance matrix of each user:**

```
function [Ch] = estimate_covariance(Y, phi, Pz, nD, Sig_u, Sig, glob)
%   Inputs:
%       Y:      Received signal block.
%       phi:    Estimated AoAs.
%       Pz:     Noise power.
%       nD:     Normalized lens aperture.
%       Sig:    Selection matrices.
%   Outputs:
%       Ch:     Estimated covariance matrix per user.
    [Nrx, K] = size(Y);         % Array size and temporal dimension.
    [Nu, L] = size(phi);        % Nº of users and beams per user.
    Ch = zeros(Nrx, Nrx, Nu);   % Estimated covariance matrices.
    % Diagonal elements of cov. matrix.
    P = 1/K*sum(Y.*conj(Y), 2) - Pz;
    % Curve fitting configuration.
    opts = optimset('Display','off');
    for u = 1:Nu    % For all users.
        for l = 1:L % For all paths.
            % Select interpolation points in bounds.
            if phi(u,l) > 0
                if phi(u,l) <= Nrx
                    a = phi(u,l) - 1;
                    b = phi(u,l);
                    c = phi(u,l) + 1;
                else
                    a = phi(u,l) - 2;
                    b = phi(u,l) - 1;
                    c = phi(u,l);
                end
            else
                a = phi(u,l);
                b = phi(u,l) + 1;
                c = phi(u,l) + 2;
            end
            % Group interpolation points for curve fitting.
            x = [a, b, c];
            % Power associated to those points.
            y = [P(a), P(b), P(c)];
            % Fit a squared sinc surface.
            est_par = lsqcurvefit(@(Par, x)Par(1)*(sinc(x - Par(2)).^2), ...
                [nD, a], x, y, [1, a], [nD^3, c], opts);
            % Apply antenna selection on the obtained curve.
            selected_sinc = sinc((1:Nrx)' - est_par(2));
            selected_sinc(~diag(Sig_u(:,:,u))) = 0;
            % Add path contributin to covariance matrix.
            Ch(:,:,u) = Ch(:,:,u) + est_par(1)*(selected_sinc*selected_sinc');
        end
    end
```

```
end
```

3. **Compute the DFDD weights matrix:**

```
function [W] = weight_matrix(Ch, Pz, K, type)
%   Inputs:
%       Ch:     Individiual covariance matrix of each user.
%       Pz:     Noise power.
%       K:      Space–time block length.
%       type:   Type of weighting.
%   Outputs:
%       W:  Weights matrix for each user.
    [~, Nrx, Nu] = size(Ch);    % NÂº of antennas and users.
    W = zeros(Nrx, Nrx, Nu);    % Weights matrices.
    switch type % Type of weighting.
        case 'uniform'  % Uniform weighting.
            W = repmat(eye(Nrx), 1, 1, Nu); % Identity matrices.
        case 'basic'    % Diagonal weighting.
            for u = 1:Nu
                % Only considers the diagonal elements of the covariance
                % matrices.
                Ph = diag(Ch(:,:,u));
                W(:,:,u) = diag(Ph./(K*Ph + Pz));
            end
        case 'optimized'    % Optimized weighting.
            for u = 1:Nu
                % Considers the full covariance matrix.
                W(:,:,u) = Ch(:,:,u)*(eye(Nrx) - ...
                    (Pz/K*eye(Nrx) + Ch(:,:,u))\Ch(:,:,u));
            end
    end
end
```

4. **Perform iterative nDFE over users and iterative DFDD over time:**

```
function [X] = iterative_nDFE(Y, W, Ch, Pz, Nu, M, Nit_ndfe, Nit_dfdd)
%   Inputs:
%       Y:          Full signal block.
%       W:          Weights matrix of every user.
%       Ch:         Channel covariance matrix of every user.
%       Pz:         Noise power.
%       Nu:         Nº of users.
%       M:          Size of the signal constellation.
%       Nit_ndfe:   Nº of iterations of nDFE.
%       Nit_dfdd:   Nº of iterations of DFDD.
%   Outputs:
%       X:      Coded symbols, each row is a user.
    K = size(Y, 2);         % Space–time block length.
    d = zeros(Nu, 1);       % Detected users flags.
    sinr = zeros(Nu, 1);    % SINR metric of every user.
```

```
    Ru = zeros(K, K, Nu);   % Correlation matrix of every user.
    X = zeros(Nu, K);       % Coded symbols.
    ord = zeros(Nu, 1);     % Detection order.
    eta_int = zeros(Nu);    % Interference contribution.
    o = 1;                  % Order counter.
    while ~all(d)   % While some users are not detected.
        for u = 1:Nu
            if ~d(u)
                % Compute SINR metric and mean interference per user.
                [sinr(u), eta_int(:,u)] = SINR_u(Ch, W, u, d, Pz);
            end
        end
        % Remove already detected users from SINR vector.
        sinr(d == 1) = NaN;
        % Choose user with better SINR metric.
        [~, u_sel] = max(sinr, [], 'omitnan');
        % User detection order.
        ord(o) = u_sel;
        o = o + 1;
        % Weighted correlation matrix of the selected user.
        Ru(:,:,u_sel) = Y'*W(:,:,u_sel)*Y;
        for nu = 1:Nu   % For all users.
            if d(nu)
                % Interference removal caused by detected users nu.
                Ru(:,:,u_sel) = Ru(:,:,u_sel) - ...
                    eta_int(nu,u_sel)*X(nu,:)'*X(nu,:);
            end
        end
        % User detected.
        d(u_sel) = 1;
        % Detection of user sequence.
        X(u_sel, :) = optDFDD(Ru(:,:,u_sel), M, Nit_dfdd, 0);
    end
    for i = 2:Nit_ndfe  % Additional iterations
        for u = 1:Nu
            % User to be detected.
            d(ord(u)) = 0;
            % Recompute interference power.
            [~, eta_int(:,u)] = SINR_u(Ch, W, ord(u), d, Pz);
            % Remove interference from all the other users.
            for nu = 1:Nu
                if d(nu)
                    Ru(:,:,ord(u)) = Ru(:,:,ord(u)) - ...
                        eta_int(nu,ord(u))*X(nu,:)'*X(nu,:);
                end
            end
            % User detected.
            d(ord(u)) = 1;
            % Detection of user sequence.
```

```
            X(ord(u), :) = optDFDD(Ru(:,:,ord(u)), M, Nit_dfdd, 0);
        end
    end
end
```

5. **Perform iterative DFDD:**

```
function [x] = optDFDD(Ru, M, N_it, varargin)
%   Inputs:
%       Z:          Weighted correlation matrix.
%       M:          Constellation size.
%       N_it:       Nº of iterations.
%       varargin:   1st flag and previous estimation (optional).
%   Outputs:
%       x:      Encoded symbols.
    K = size(Ru, 2);    % Signal block size.
    x = [1 zeros(1, K - 1)];        % Encoded symbols.
    phi = zeros(1, K);              % Differential phase.
    % Constellation phase differential.
    delta = 2*pi/M;
    % Symbol phases.
    ang = zeros(1, K);
    % Phase quantization error per symbol.
    Derr = zeros(1, K);
    if ~varargin{1}     % There are not prior results.
        d = [true false(1, K - 1)];    % Detected symbols.
        while ~all(d)   % While there are undetected symbols.
            % Vectorized DFDD detection.
            phi(~d) = x(d)*Ru(d, ~d);
            ang(~d) = atan2(imag(phi(~d)), real(phi(~d)));
            Derr(~d) = abs(ang(~d) - delta*round(ang(~d)/delta));
            % Remove already detected symbols from the comparison.
            Derr(d) = NaN;
            % Detect symbol with less quantization error.
            [~, k] = min(Derr, [], 'omitnan');
            % Symbol is detected.
            d(k) = true;
            % Choose symbol from M-ARY constellation.
            x(k) = exp(1j*delta*round(atan2(imag(phi(k)), ...
                real(phi(k)))/delta));
        end
    else    % There are prior results.
        x = varargin{2};
    end
    for i = 2:N_it  % Use the full detected sequence to perform DFDD.
        for k = 1:K
            phi(k) = x((1:K) ~= k)*Ru((1:K) ~= k, k);
            x(k) = exp(1j*delta*round(atan2(imag(phi(k)), ...
                real(phi(k)))/delta));
        end
```

```
    end
end
```

6. **Compute each user's SINR metrics for nDFE:**

```
function [SINR, eta_int] = SINR_u(Ch, W, u, d, Pz)
%   Inputs:
%       Ch:     Channel covariance matrix of each user.
%       W:      Signal correlation matrix of each user.
%       u:      Selected user.
%       d:      Detected users.
%       Pz:     Noise power.
%   Outputs:
%       SINR:   Signal-To-Noise-And-Interference Ratio
    Nu = size(Ch, 3);   % Nº of users.
    Nrx = size(Ch, 1);  % Nº of antennas.
    eta_int = zeros(Nu, 1); % Mean interference power per user.
    % Product of weights matrix by other users' covariance matrices.
    WCnu = zeros(Nrx, Nrx, Nu);
    WCnu(:,:,u) = W(:,:,u)*Ch(:,:,u);
    % Desired signal power.
    Pu = sum(diag(WCnu(:,:,u)))^2 + sum(sum(WCnu(:,:,u).*WCnu(:,:,u).'));
    % Signal x Interference power.
    Pint = 0;
    % Interference x Interference power.
    Pcross = 0;
    for mu = 1:Nu
        if mu ~= u
            WCnu(:,:,mu) = W(:,:,u)*Ch(:,:,mu);
            Pint = Pint + sum(sum(WCnu(:,:,mu).*WCnu(:,:,mu).'));
            if ~d(mu)
                eta_int(mu) = sum(diag(WCnu(:,:,mu)));
                Pint = Pint + eta_int(mu)^2;
            end
        end
        for nu = 1:(mu - 1)
            Pcross = Pcross + 2*sum(sum(WCnu(:,:,nu).*WCnu(:,:,mu).'));
        end
    end
    WuWu = W(:,:,u)*W(:,:,u);
    % Interference x Noise + Noise x Noise
    Pnoise = 2*Pz*sum(sum(WuWu.*sum(Ch, 3).')) + Pz^2*sum(diag(WuWu));
    % SINR metric.
    SINR = Pu/(Pint + Pcross + Pnoise);
end
```

7. **Perform Minimum Mean Squared Error V-BLAST detection over users:**

```
function [Xest] = mmse_vblast(H, Y, Pz, Nu, M)
%   Inputs:
```

```
%       H:  Estimated channel.
%       Y:  Received signal block.
%       Pz: Noise power.
%       Nu: NÂº of users.
%       M:  Signal constellation size.
%   Outputs:
%       Xest:   Detected symbols.
    Xest = zeros(Nu, size(Y, 2));
    delta = 2*pi/M;     %   Phase differential.
    d = false(1, Nu);   %   Detected users vector.
    for i = 1:Nu    % For all users.
        % Precoding matrix.
        G = H/(H'*H + Pz*eye(Nu));
        % Ignore detected users.
        G(:,d) = NaN;
        % Find user with better communication conditions.
        [~, u] = min(vecnorm(G), [], 'omitnan');
        % Detect user sequence.
        y = G(:,u)'*Y;
        Xest(u,:) = exp(1j*delta*round(atan2(imag(y), real(y))/delta));
        % User has been detected.
        d(u) = true;
        % Remove user contribution from received signal and channel.
        Y = Y - H(:,u)*Xest(u,:);
        H(:,u) = 0;
    end
end
```

8. **Compute the power loading matrix for D-OSTBC:**

```
function [Duni, Dh, Dopt, U] = power_load(Ch, L, M, Pz)
%   Inputs:
%       Ch:     Channel covariance matrix per user.
%       L:      Nº of beams per user.
%       M:      Size of the symbol constellation.
%       Pz:     Noise power.
%   Outputs:
%       Duni:   Uniform power loading matrix.
%       Dh:     Heuristic power loading matrix.
%       Dopt:   Optimized power loading matrix.
%       U:      Matrix with L most significant eigenvectors per user.
    Nu = size(Ch, 3);       % Nº of users.
    Nrx = size(Ch, 1);      % Nº of antennas.
    U = zeros(Nrx, L, Nu);  % Eigenvectors per user.
    % Power loading matrices per user.
    Dh = zeros(L, L, Nu);
    Dopt = zeros(L, L, Nu);
    % Uniform power loading.
    Duni = 1/sqrt(L)*repmat(eye(L), 1, 1, Nu);
    for u = 1:Nu    % For all users.
```

```
        % Partial eigendecomposition. L eigenpairs.
        [U(:,:,u), D] = eigs(Ch(:,:,u), L);
        % For each path.
        for i = 1:L
            % Heuristic weight.
            Dh(i,i,u) = sqrt(D(i,i)/sum(diag(D(1:L,1:L))));
            % Optimized weight.
            Dopt(i,i,u) = sqrt(1/L + 2*Pz/sin(pi/M)^2* ...
                1/L*(sum(1./diag(D(1:(i - 1),1:(i - 1)))) + ...
                sum(1./diag(D((i + 1):L,(i + 1):L)))));
        end
    end
end
```

9. **Decode differentially encoded OSTBCs:**

```
function [S] = diff_decoder(y1, y2, M, A, B)
%   Inputs:
%       y1: Received signal at time k - 1.
%       y2: Received signal at time k.
%       M:  Size of the signal constellation.
%       A:  First matrix component of the amicable design.
%       B:  Second matrix component of the amicable design.
%   Outputs:
%       S:  Decoded symbol block.
    S = zeros(1,3);
    for p = 1:3 % For all the symbols in a block.
        dec = 0;
        for i = 0:(M - 1)   % Explore all the symbols in the constellation.
            s = exp(1j*2*pi/M*i);
            % Maximize the decoding metric.
            newdec = real(y1*A(:,:,p)*y2')*real(s) - ...
                imag(y1*B(:,:,p)*y2')*imag(s);
            % If the current symbol improves the metric for the received
            % symbols, it is selected.
            if newdec > dec
                dec = newdec;
                S(p) = s;
            end
        end
    end
end
```

10. **Generate OSTBCs of rate 3/4:**

```
function [S] = OSTBC3_4(s1, s2, s3)
%   Inputs:
%       s1, s2, s3: Symbols to be encoded.
%   Outputs:
%       S:  Encoded space-time block.
    S = [   s1,     0,      s2,     -s3;
            0,      s1,     s3',    s2';
            -s2',   -s3,    s1',    0;
            s3',    -s2,    0,      s1']/sqrt(3);
end
```