UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

DEVELOPMENT AND CHARACTERIZATION OF A 2D, 700–2600 MHZ RECEIVER ARRAY, FOR THE ESTIMATION OF DIRECTION OF ARRIVAL (DOA) IN THE FREQUENCY DOMAIN

TESIS PARA OPTAR AL GRADO DE MAGISTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA

DIEGO SALVADOR JOSÉ MANUEL GALLARDO VIVES

PROFESOR GUÍA:
RICARDO FINGER CAMUS

PROFESOR CO-GUÍA:
DAVID MONASTERIO LAGOS

MIEMBROS DE LA COMISIÓN:
CESAR AZURDIA MEZA
WALTER MAX-MOERBECK ASTUDILLO

Este trabajo ha sido parcialmente financiado por CONICYT-ANID a través de sus fondos Basal ACE210002, FB210003, y FONDECYT 1221662.

SANTIAGO DE CHILE
2024

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE MAGISTER EN
CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA
POR: **DIEGO SALVADOR JOSÉ MANUEL GALLARDO VIVES**
FECHA: 2024
PROF. GUÍA: RICARDO FINGER CAMUS

DESARROLLO Y CARACTERIZACIÓN DE UN ARREGLO DE ANTENAS 2D, DE 700 A 2600 MHZ, PARA LA ESTIMACIÓN DE LA DIRECCIÓN DE ARRIBO (DOA) EN EL DOMINIO DE LA FRECUENCIA

Esta tesis presenta el desarrollo de RadioVision, un sistema receptor de 4×4 antenas para la estimación en tiempo real de direcciones de arribo (DoAs) de fuentes de radiofrecuencia (RF) en el rango de 700–2600 MHz, con una implementación novedosa del algoritmo U-ESPRIT 2D en el dominio de frecuencias. RadioVision combina un *front-end* analógico con un receptor heterodino de 70 MHz de ancho de banda de frecuencia intermedia, y un *back-end* digital con 16 ADCs operando a 140 MSPS y una FPGA para el procesamiento de señales en tiempo real, incluyendo transformadas rápidas de Fourier (FFTs) y la estimación de DoAs. Este enfoque reduce la cantidad de ADCs necesarios y permite el filtrado de bandas para suprimir interferencias indeseadas.

RadioVision incorpora una cámara de video para crear una pantalla de realidad aumentada, superponiendo marcadores en las fuentes de RF detectadas, lo que proporciona una visualización en tiempo real. Las pruebas realizadas arrojaron un error cuadrático medio de DoA de 3.12° en zonas rurales, y de 4.63° y 5.76° en zonas urbanas complejas, validando la efectividad del algoritmo U-ESPRIT 2D y demostrando la viabilidad de la localización de fuentes RF en condiciones reales.

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE MAGISTER EN
CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA
POR: **DIEGO SALVADOR JOSÉ MANUEL**
GALLARDO VIVES
FECHA: 2024
PROF. GUÍA: RICARDO FINGER CAMUS

DEVELOPMENT AND CHARACTERIZATION OF A 2D, 700–2600 MHZ RECEIVER ARRAY, FOR THE ESTIMATION OF DIRECTION OF ARRIVAL (DOA) IN THE FREQUENCY DOMAIN

This thesis presents the design, construction, and testing of RadioVision, a 4 × 4 antenna array receiver for the real-time direction of arrival (DoA) estimation of radiofrequency (RF) sources within the 700–2600 MHz range. The system features a novel implementation of the 2D U-ESPRIT algorithm in the frequency domain, leveraging a heterodyne receiver with a 70 MHz intermediate frequency bandwidth and a digital back-end with 16 ADCs operating at 140 MSPS. Processing is handled by an FPGA, enabling Fast Fourier Transforms (FFTs) and DoA estimations, reducing ADC requirements, and allowing frequency band filtering to suppress unwanted interference.

RadioVision also integrates an augmented reality display that overlays markers on detected RF sources using a video camera, giving real-time visual feedback. Tested in both rural and urban environments, the system achieved a mean squared DoA error of 3.12° at 80–180 meters in rural settings, and 4.63° and 5.76° at 10–60 meters in urban areas. These results validate the effectiveness of the frequency-domain U-ESPRIT 2D algorithm and demonstrate the feasibility of real-time RF source localization in complex environments.

Table of Content

Acronyms

- **ADC:** Analog-to-Digital Converter
- **CASPER:** Collaboration for Astronomy Signal Processing and Electronics Research
- **CW:** Continuous Wave
- **DC:** Direct Current
- **DDS:** Direct Digital Synthesizer
- **DL:** Down Link
- **DoA:** Direction of Arrival
- **DSB:** Double Side Band
- **DSP:** Digital Signal Processing
- **ESPRIT:** Estimation of Signal Parameters via Rotational Invariance Techniques
- **FDMA:** Frequency Division Multiple Access
- **FFT:** Fast Fourier Transform
- **FIR:** Finite Impulse Response
- **FPGA:** Field-Programmable Gate Array
- **IF:** Intermediate Frequency

✠ **LNA:** Low Noise Amplifier

✠ **LO:** Local Oscillator

✠ **LTE:** Long Term Evolution

✠ **MUSIC:** Multiple Emitter Location and Signal Parameter Estimation

✠ **MWL:** Millimeter Wave Lab

✠ **PFB:** Polyphase Filter Bank

✠ **RF:** Radio Frequency

✠ **RFI:** Radio Frequency Interference

✠ **RMSE:** Root Mean Square Error

✠ **ROACH:** Reconfigurable Open Architecture Computing Hardware

✠ **SNR:** Signal-to-Noise Ratio

✠ **U-ESPRIT:** Unitary-ESPRIT

✠ **UL:** Up Link

✠ **ULA:** Uniform Linear Array

✠ **UMTS:** Universal Mobile Telecommunications System

✠ **URA:** Uniform Rectangular Array

Chapter 1

Introduction

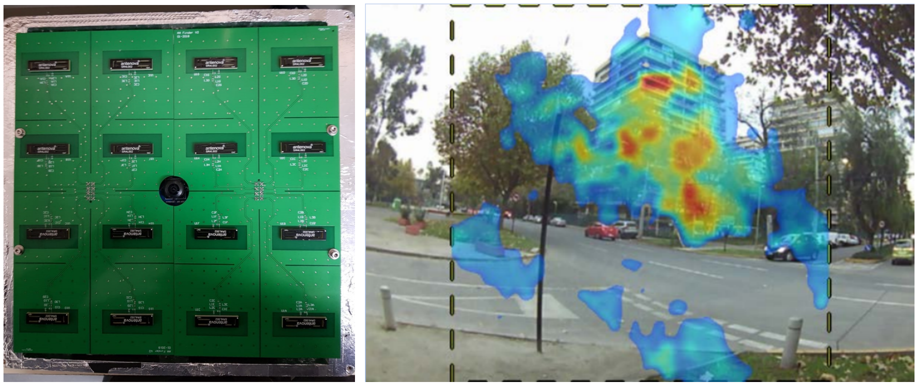
1.1 Motivation

Mobile phones are among the most commonly used objects in daily life. According to a report by eMarketer, people spend an average of approximately four hours a day on their phones [1]. Additionally, GSMA statistics indicate that 67% of the global population owns a mobile phone [2]. Given these figures, there is a high probability that if we approached a random person, they would have a mobile phone with them.

RadioVision is a project being developed at the Millimeter Wave Laboratory (MWL) of the Universidad de Chile, aimed at locating mobile phones. This technology could be invaluable for locating individuals during emergencies, such as earthquakes or avalanches. It could also be used to confiscate phones in areas where their use is prohibited, such as prisons.

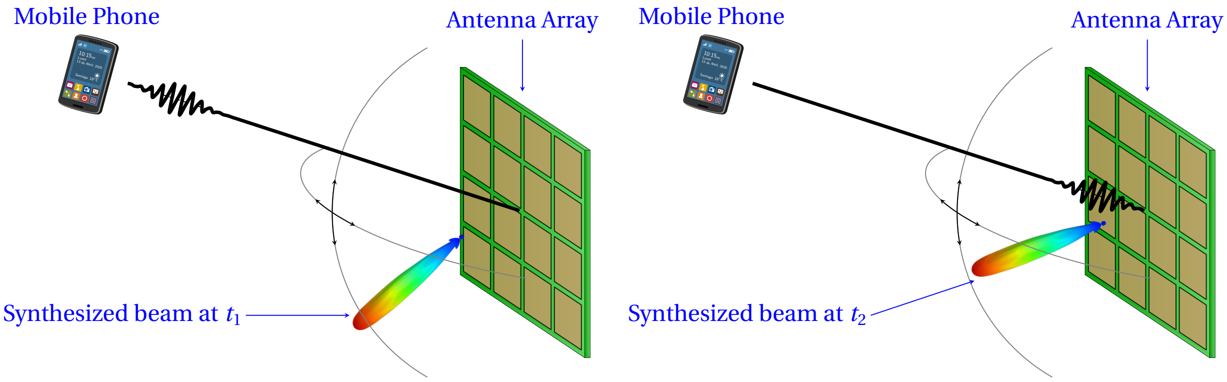
In its early stages around 2018, RadioVision operated with a 4×4 antenna array that functioned as a phased array, scanning the field of view for mobile phones, as shown in Figure 1.1. The resulting radiation map generated from this scan was then superimposed onto an image captured by an optical camera, integrated at the center of the array, creating an augmented reality view, as illustrated in panel (b) of the figure.

The system depicted in Figure 1.1 had two main limitations. First, due to the inherent nature of phased arrays, the resolution for locating phones was limited to around 20° at best, regardless of the system's signal-to-noise ratio (SNR), as the beam's size determines the resolution. Second, with the setup used at that time, which consisted of programmable phase shifters and a Raspberry Pi, it took over a second to form the radiation map since the beam had to scan the entire space, as illustrated in Figure 1.2. This delay made the system incapable of detecting phone beacons, which last around 1 ms and are crucial for applications like locating people in disasters. Due to these significant shortcomings, we decided to investigate the implementation of Direction of Arrival (DoA) estimation techniques and their integration into faster digital systems, such as Field Programmable Gate Arrays (FPGAs).



(a) (b)

Figure 1.1: (a) Preliminary version of RadioVision's antenna array integrated with an optical camera. (b) An electromagnetic image is superimposed on the corresponding optical image.



(a) (b)

Figure 1.2: Illustration of a phased array scanning the environment with its synthesized beam, highlighting its limitations in detecting a cellphone beacon. (a) Position of the beam at t_1. (b) Position of the beam at a later time t_2, emphasizing the delay that may cause the system to miss the brief signal emitted by the cellphone beacon.

1.2 DoA Estimation methods

DoA estimation using antenna arrays is used in many applications, such as mobile telecommunications, radar, and object tracking [3, 4, 5]. While there are several algorithms for DoA estimation, such as MUSIC [6], ESPRIT [7], U-ESPRIT [8], and U-ESPRIT 2D [9], most have only been validated through simulations, and may not perform well in real-life scenarios. In these situations, DoA estimation algorithms must address several challenges. These include the non-zero bandwidth of signals and potential imbalances in phase and amplitude caused by imperfections in antennas and receivers. Furthermore, the algorithms must achieve high computing speeds for real-time applications and address external propagation issues, such as interference from other sources -especially base stations- and multipath effects resulting from signal reflections and scattering.

The initial method utilized for source localization using antenna arrays was the Bartlett beamformer, a technique dating back to the Second World War. This method is used in the

2018 version of RadioVision and involves an electronic sweep across the entire field of view to identify the direction of maximum power reception, which indicates the likely location of the source [10, 11]. Over time, the Bartlett beamformer has evolved into what is now recognized as the adaptive beamformer [12], offering improved functionality. While Bartlett beamformer is widely applicable and easy to understand, it has significant limitations. Notably, its precision is contingent on the size of the antenna array, irrespective of factors such as the SNR or the temporal data collection process [4]. Due to these shortcomings, subspace-based methods emerged, such as MUSIC, ESPRIT, and U-ESPRIT 2D. One of the most significant advantages of these methods is that the array size does not limit the resolution, provided the data collection and the SNR are sufficiently large.

Sub-space methods have demonstrated remarkable performance over the years, achieving simulated source localization accuracy of less than one degree, even with signal-to-noise ratios less than 0 dB and antenna array sizes as small as 4λ [9]. However, much of the research on these methods has focused on numerical simulations that exclude significant effects from reality. For example, the impact of floor reflections is typically excluded, which can be problematic in any application where the antenna array is not located on a high-rise tower. Moreover, ground reflections can be particularly troublesome in algorithms that estimate the number of sources [13, 14, 15, 16]. Another problem that must be solved is the presence of multiple unwanted sources, such as mobile telephone base stations. Indeed, most algorithms limit the number of sources for which the direction of arrival can be determined. For instance, if we have a one-dimensional array of N antennas, MUSIC and ESPRIT can determine only $N-1$ sources. If d interferers are also considered, this number decreases to $N-1-d$. In this sense, a base station with multiple incoherent emitters ($d \geq N-1$, for example) can completely ruin the detection capability of sub-space methods. One last challenge arises from phase and amplitude imbalances presented in the non-zero bandwidth of the receiving system. When implementing sub-space algorithms in the time domain, as commonly done in the literature [6, 7, 8, 9, 17, 18, 19, 20, 21], calibrating these errors may necessitate the use of finite impulse response filters, adding complexity to both model design and resource utilization.

Although there are studies that have sought to experimentally verify subspace algorithms [18, 19, 20, 21], these have been conducted in controlled laboratory environments and do not explore realistic situations involving large distances in the open field, obstacles such as trees and buildings, and the presence of interference. Furthermore, the studies cited in [18, 19, 20, 21] only validate the subspace algorithms on the horizontal axis, failing to address the vertical axis, which presents challenges due to ground reflections.

Given the challenges and limitations of existing methods, we implemented subspace methods in the frequency domain. Although the theoretical step from time to frequency domain is known, it has not been considered until now due to the higher computational cost and slowness in the Fast Fourier Transform (FFT) calculation. However, technological advances have made possible to calculate the FFT in real time using FPGAs. This implementation offers several advantages, including the elimination of half of the ADCs necessary to form I-Q inputs (or the elimination of Direct Digital Synthesizers (DDS) otherwise), the calibration of amplitude and phase imbalances across the entire frequency range, the identification and flagging of unwanted frequency bands, and, if the modulation type allows it, the precise cal-

Table 1.1: Cell phone bands tendered in Chile. UL: up-link, DL: down-link, BW: bandwidth.

Band	**UL (MHz)**	**DL (MHz)**	**UL BW (MHz)**
28 APT (LTE)	703–748	758–803	45
5 CLR (UMTS)	824–849	869–894	25
8 E-GSM (UMTS)	902–912	947–957	10
10 AWS (LTE)	1710–1770	2110–2170	60
2 PCS (LTE)	1850-1910	1930–1990	60
7 IMT-E (LTE)	2505–2565	2625–2685	60

culation of the carrier frequency, which could lead to a better estimation of the DoAs. This proposed solution is a significant step forward in the field of DoA estimation, addressing fundamental challenges and opening up new possibilities for real-time applications.

1.3 System Description

The primary objective of this thesis is to design, implement, and characterize the performance of RadioVision: a system that implements the frequency domain version of U-ESPRIT 2D and is capable of locating multiple narrowband ($\leq$ 70 MHz) radio sources distributed across a wide range of frequencies (700–2600 MHz). To demonstrate the effectiveness and practicality of the proposed system, we conducted experiments using cell phones. The UMTS and LTE frequency bands we considered to validate RadioVision are those tendered in Chile, as shown in Table 1.1.

While the proposed system has applications in various fields, we are particularly interested in locating cell phones in diverse scenarios -stationary or moving, open fields or behind walls- and in various operation modes -texting, calling, or stand-by-. To achieve this, RadioVision performs rapid DoA estimation calculations with sub-millisecond response times, aligning with the beacon duration of cell phones. Additionally, since cell phones frequently switch between operation bands, the system must be able to switch between different bands quickly. This section provides a brief overview of the components that constitute RadioVision and are essential for rapid cell phone detection in such situations. For detailed descriptions of RadioVision's hardware and software, please refer to Chapters 4 and 5.

A simplified schematic of RadioVision is shown in Fig. 1.3. The system features a 4×4 antenna array designed to receive phone signals, ensuring coverage of all the up-links (UL) detailed in Table 1.1. The signals are then amplified and down-converted to baseband using Double SideBand (DSB) mixers and anti-aliasing filters. We can select sub-bands from Table 1.1 by varying the local oscillator (LO). To calculate the phones' DoAs in real time, the instrument incorporates a ROACH-2 platform (Reconfigurable Open Architecture Computing Hardware) developed by the CASPER group (Collaboration for Astronomy Signal Processing and Electronics Research). The board's key components include the Xilinx Virtex-6 XC6VSX475T FPGA and 16 HMCAD1511 ADCs with a sampling rate of 140 MSPS. The samples collected by the ADCs pass through Polyphase Filter Banks (PFBs) and FFT calculators implemented in the FPGA, allowing real-time spectra calculation.

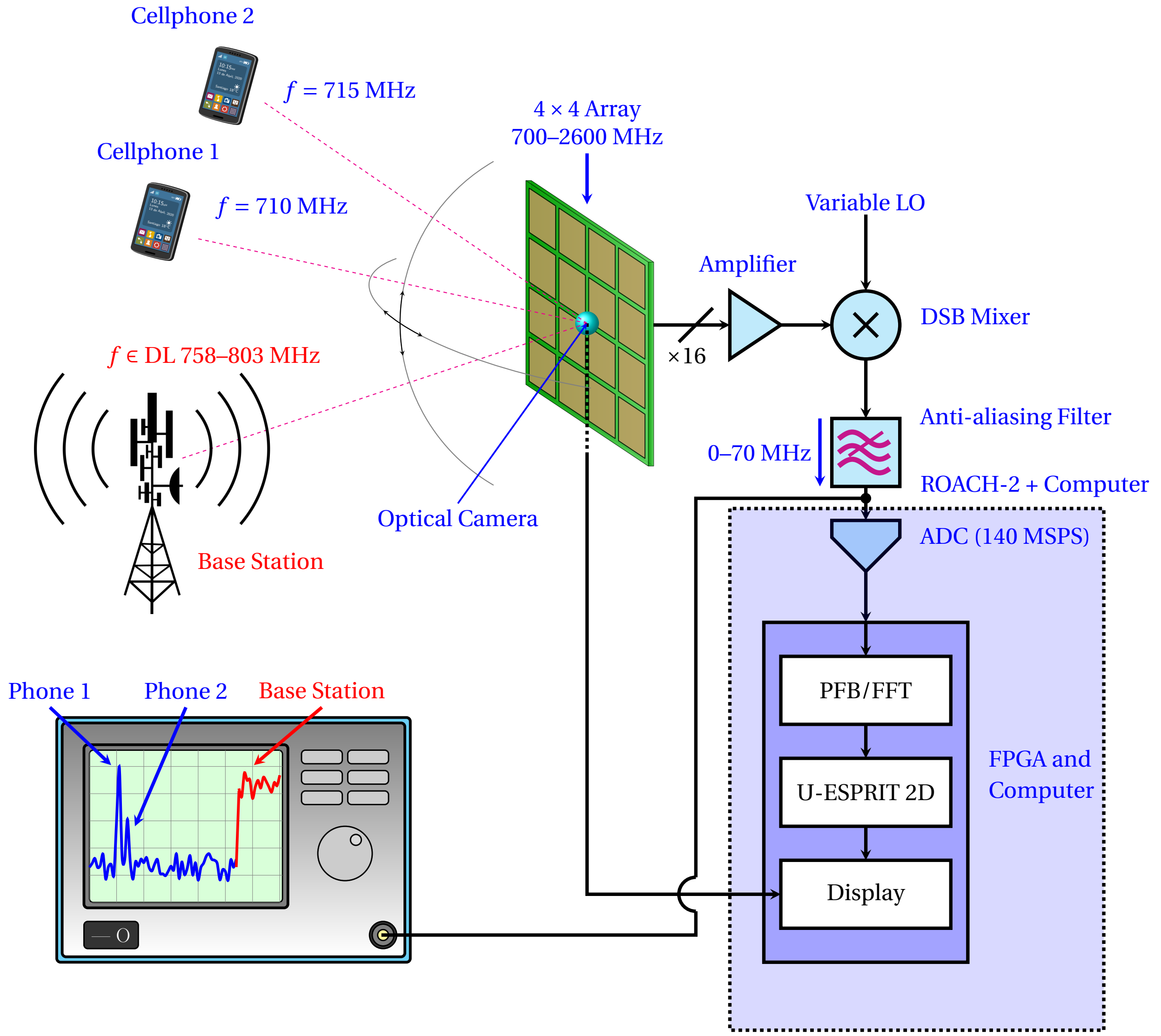


Figure 1.3: Simplified schematic of RadioVision. For example, we consider two cell phones in the UL range of 703–748 MHz and a base station in the DL range of 758–803 MHz. The variable LO allows us to select the frequency bands from Table 1.1. In this example, we set the LO at 700 MHz.

On the one hand, implementing PFBs helps to avoid spectral leakage and scalloping. Figure 1.3 illustrates a situation in which it could be advantageous to prevent leakage by implementing PFBs, where two phones with FDMA modulation have very close frequencies (710 and 715 MHz). If phone #1 is significantly more powerful than phone #2 (which could happen if one phone is much closer to the array than another or if signals from one phone must pass through obstacles), the leakage of phone #1 can drown out phone #2, thereby altering its estimated location, or even wholly discarding its consideration in the source number estimation. By implementing PFBs, this leakage is prevented. On the other hand, implementing the FFT offers all the advantages mentioned in Section 1.2, i.e., amplitude and phase imbalance calibration, flagging, and frequency calculation (if necessary). For example, Fig. 1.3 shows that placing the LO at 700 MHz causes the base station from the 758–803 MHz down-link (DL) to appear in the intermediate frequency (IF). Additionally, interference in the lower sideband below 700 MHz prevents us from lowering the LO further. Since base stations have multiple incoherent emitters, their presence could completely ruin the performance of U-ESPRIT 2D in the time domain. However, by implementing the FFT, we can flag (filter) those channels to prevent them from disturbing the algorithm.

After calculating the DoAs, we display the positions as markers on an optical image taken with a camera located at the array's center. This approach makes RadioVision easier to operate and debug. Furthermore, with this approach, we can conduct tests with moving sources and compare the estimated DoAs with the actual positions after obtaining the latter with the camera recording. In this manner, the integration with the camera allows us to evaluate RadioVision's performance in multiple source positions, in contrast to what is done in [18, 19, 20, 21], which keeps the sources in fixed positions.

1.4 Objectives

1.4.1 General Objective

The primary objective of this thesis is to design, implement, and characterize the performance of a system intended to locate mobile phones.

1.4.2 Specific Objectives

1) Integrate a 4 × 4 antenna array that operates without phase ambiguities across the 700–2600 MHz range, a 16-channel receiver with an RF bandwidth of 700–2600 MHz and an IF bandwidth of 0–70 MHz, and the ROACH-2 platform into a portable device.
2) Implement the U-ESPRIT 2D algorithm in the frequency domain on the ROACH-2 platform.
3) Measure the amplitude and phase response of the receiver across all bands listed in Table 1.1 for calibration in the digital system.
4) Develop a Python software program to manage all processes necessary for operating RadioVision, which includes:
    - ✠ Configuring the FPGA, monitoring its status, and loading calibration registers.
    - ✠ Receiving data from the FPGA and computing DoAs.
    - ✠ Applying post-processing techniques to the DoAs.
    - ✠ Forming the augmented reality image.
5) Validate RadioVision's performance with mobile phones, aiming for an accuracy of under 5° in phone localization in two angular dimensions.

1.5 Hypothesis

DoA estimation methods can achieve localization accuracies of approximately 5 degrees, regardless of the array size, as long as the SNR and data accumulation are sufficiently high. Moreover, implementing these methods in the frequency domain offers several advantages, such as flagging base stations, reducing hardware requirements and computational costs, and facilitating the receiver's calibration of amplitude and phase imbalances. Finally, deploying these techniques on an FPGA can result in computation times of less than 1 ms, which is fast enough to localize cellphone beacons accurately.

In conclusion, incorporating a frequency-domain DoA estimation method, such as U-ESPRIT 2D, on an FPGA can significantly enhance the performance of RadioVision, making it suitable for practical, real-world applications beyond laboratory settings.

1.6 Thesis Outline

The rest of the document is organized as follows.

- **Chapter 2: Theoretical Framework** This chapter reviews the theoretical background necessary for completing this thesis. It focuses on the principles behind each component of RadioVision and discusses various DoA estimation methods.
- **Chapter 3: Signal Model and Application of the FFT in the Time Domain** We present the mathematical model and foundation of the U-ESPRIT 2D algorithm applied to the frequency domain.
- **Chapter 4: Hardware Design** This chapter details the hardware design, implementation, and measurement processes. It also includes several practical considerations that influenced the implementation.
- **Chapter 5: Software Design** The software design is presented in this chapter.
- **Chapter 6: Field Measurement Tests** This chapter presents the RadioVision measurements. Experiments were conducted using continuous wave sources and cell phones in call mode. Additionally, the experiments were carried out under various environmental conditions.
- **Chapter 7: Conclusions** In this chapter, we summarize the main characteristics of RadioVision. Specifically, we compare the results achieved with the proposed objectives. We also highlight aspects not covered in this work and suggest them for future research.

Chapter 2

Theoretical Framework

This work applies and extends the Direction of Arrival estimation methods referenced in [6, 7, 8, 9]. Implementing these methods in an analog-digital system requires antenna arrays and microwave systems expertise. This chapter outlines the fundamental concepts necessary to understand these fields, extensively utilized throughout the thesis.

We start in Section 2.1 by presenting the notation used throughout the thesis. In Section 2.2, we review antenna arrays and microwave receivers, which form the analog component of RadioVision. Finally, Section 2.3 introduces the DoA estimation methods from [6, 7], which serve as the foundation for the method developed in this thesis. A solid understanding of linear algebra is essential for grasping DoA estimation methods. For those interested, a detailed overview of linear algebra concepts can be found in Appendix A.

2.1 Notation

First, we introduce the notation used throughout the thesis to make the presentation as straightforward as possible. Uppercase boldface characters refer to matrices, while lowercase boldface characters refer to vectors. We use $(\mathbf{A})_{mn}$ to denote the (m, n)-th entry of the matrix $\mathbf{A}$, $(\mathbf{a})_m$ to denote the m-th entry of the vector $\mathbf{a}$, and we start the numbering with $m = 0$ and $n = 0$ (i.e., $(\mathbf{A})_{00}$ is the element in the first row and the first column of the matrix $\mathbf{A}$, while $(\mathbf{a})_0$ is the first element of the vector $\mathbf{a}$). The notation $\mathbf{A}^T$, $\mathbf{A}^*$ and $\mathbf{A}^H$ denote, respectively, the transpose, the conjugate, and the transpose conjugate of $\mathbf{A}$. If we have the $M \times 1$ vectors $\mathbf{a}_0 = [a_{00}\ a_{10}\ ...\ a_{M-1,0}]^T$, $\mathbf{a}_1 = [a_{01}\ a_{11}\ ...\ a_{M-1,1}]^T$, ..., $\mathbf{a}_{N-1} = [a_{0,N-1}\ a_{1,N-1}\ ...\ a_{M-1,N-1}]^T$, and the $N \times 1$ vectors $\mathbf{x}_0 = [a_{00}\ a_{01}\ ...\ a_{0,N-1}]^T$, $\mathbf{x}_1 = [a_{10}\ a_{11}\ ...\ a_{1,N-1}]^T$, ..., $\mathbf{x}_M = [a_{M-1,0}\ a_{M-1,1}\ ...\ a_{M-1,N-1}]^T$, we can express an $M \times N$ matrix $\mathbf{A}$ with elements $(\mathbf{A})_{mn} = a_{mn}$ as

$$\mathbf{A} = \begin{bmatrix} a_{00} & a_{01} & ... & a_{0,N-1} \\ a_{10} & a_{11} & ... & a_{1,N-1} \\ \vdots & \vdots & ... & \vdots \\ a_{M-1,0} & a_{M-1,1} & ... & a_{M-1,N-1} \end{bmatrix}, \tag{2.1}$$

or

$$\mathbf{A} = \left[\begin{array}{cccc} \mathbf{a}_0 & \mathbf{a}_1 & ... & \mathbf{a}_{N-1} \end{array} \right], \tag{2.2}$$

or

$$\mathbf{A} = \left[\begin{array}{c} \mathbf{x}_0^T \\ \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_{M-1}^T \end{array} \right]. \tag{2.3}$$

We use $j = \sqrt{-1}$ to represent imaginary numbers, $\mathbb{R}\{x\}$ and $\mathbb{I}\{x\}$ to denote the real and imaginary parts of the complex number x, and $\otimes$ to denote the Kronecker product. Finally, some valuable matrices used throughout the thesis are $\mathbf{I}_N$, which represents the $N \times N$ identity matrix, $\mathbf{\Pi}_N$, which is the $N \times N$ exchange matrix with ones in its antidiagonal and zeros elsewhere, $\mathbf{0} = [0\ 0\ ...\ 0]^T$, which is the zero-vector whose dimension can be deduced from context, and $\mathbf{O}$, which represents the matrix filled with 0s whose dimensions can be deduced from context.

2.2 Antenna Arrays and Microwave Receivers

2.2.1 Antenna Arrays

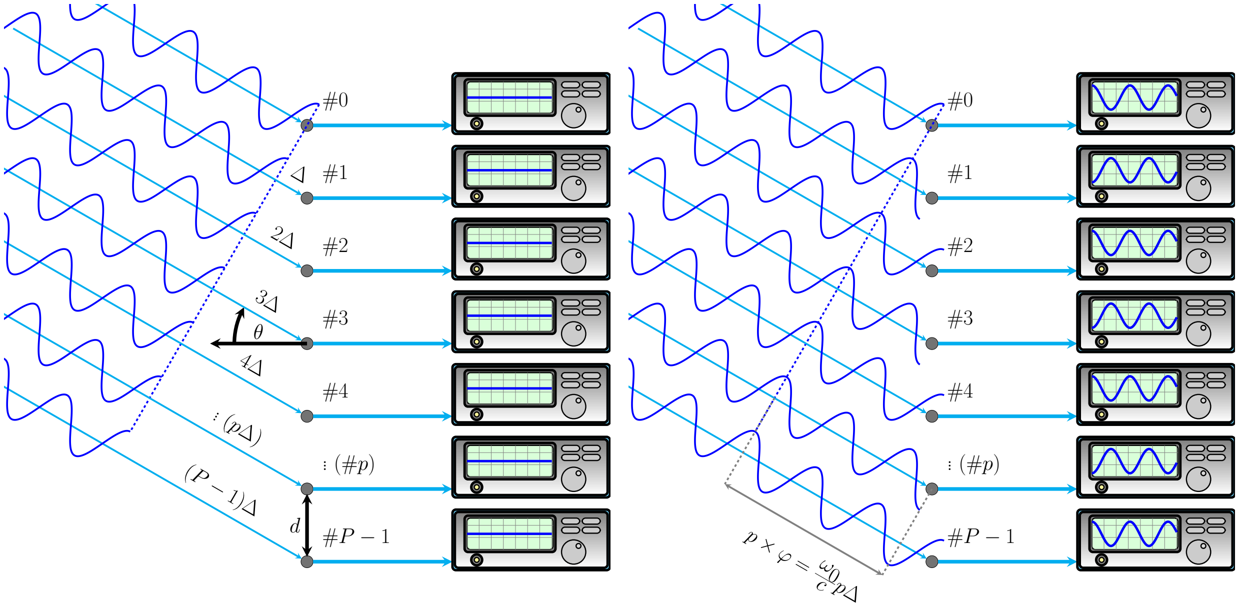
An antenna array consists of two or more antennas positioned at specific points in space. Technically, a group of antennas qualifies as an array only when their radiation patterns overlap in the far field. Although different antennas can be used, the typical approach is to use identical antennas arranged in a planar grid with fixed and equal spacing between them. In this thesis, we focus solely on arrays of identical antennas that are uniformly excited and placed at equal intervals.

This section analyzes the one-dimensional (1D) case, known as Uniform Linear Arrays (ULA). The extension from 1D to two-dimensional (2D) arrays is straightforward, and readers interested in this topic can refer to the works in [10, 11].

It is important to emphasize that while our primary focus is on signal reception -specifically within the context of RadioVision- the analysis followed is equally applicable to signal transmission. This equivalence arises from the reciprocity theorem, which ensures the symmetrical behavior of electromagnetic waves during transmission and reception in passive, linear, and time-invariant systems [10, 11].

2.2.1.1 Uniform and Linear Antenna Arrays (ULA)

The impressive capabilities of antenna arrays arise from the fact that incoming waves reach each antenna with different phases, as shown in Figure 2.1.



(a) (b)

Figure 2.1: Illustration of a ULA consisting of $P = 7$ antennas with spacing $d = \lambda$. (a) Transient state, where the incoming wavefront has only reached antenna #0. (b) Steady-state scenario, where all antennas are continuously receiving the signal.

Consider an array of P antennas, labeled from #0 to #$P-1$, separated by a distance d (as depicted in Figure 2.1). Suppose a quasi-monochromatic wave with frequency ω_0 impinges on the array at an angle θ relative to the array's normal. As panel (a) illustrates, the wavefront first reaches antenna #0. After traveling a distance Δ, it subsequently reaches antenna #1. It continues this pattern, reaching antenna #2 after traveling 2Δ, and so on, until it arrives at antenna #$P-1$ after traveling a total distance of $(P-1)\Delta$.

In the steady state scenario shown in panel (b), the signal received by antenna #p experiences a phase shift of $p \times \varphi$ relative to the reference antenna #0, where φ is given by

$$\varphi = \frac{\omega_0 \Delta}{c} = \frac{\omega_0 d \sin(\theta)}{c} = \frac{2\pi d \sin(\theta)}{\lambda}, \tag{2.4}$$

being c the speed of light, λ the wavelength associated with ω_0, and we have derived $\Delta = d \sin\theta$ from the geometric configuration shown in the figure. Note that the phase difference between every pair of consecutive antennas is φ.

The phase shifts between antennas allow antenna arrays to achieve remarkable capabilities. For example, if the array is designed to avoid phase ambiguities ($d < \lambda/2$) [22], each direction of arrival θ can be uniquely identified by the phase shift between consecutive antennas, $\varphi = \omega_0 d \sin\theta / c$. Therefore, by measuring the phase shift φ, the direction of arrival θ can be determined through the following expression:

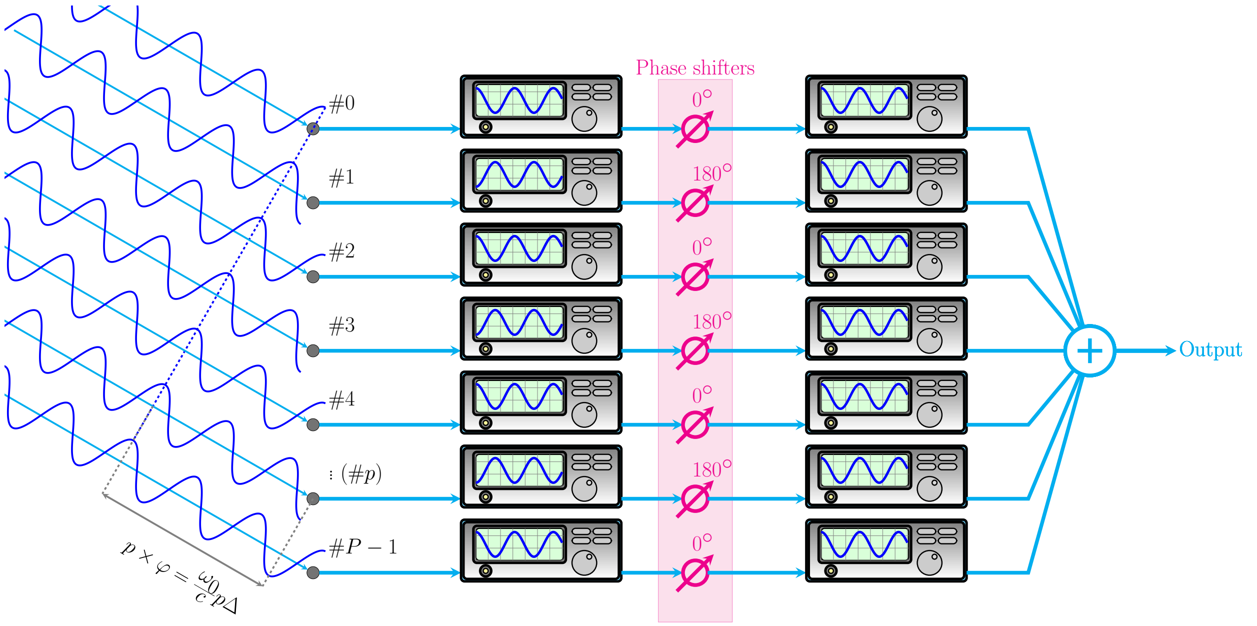


Figure 2.2: Beamformer example where the phase shift introduced by the p-th phase shifter is given by $\zeta_p = (1 - (-1)^p) \times 90°$.

$$\theta = \sin^{-1}\left(\frac{c\varphi}{\omega_0 d}\right). \tag{2.5}$$

As another example, we can connect phase shifters to the output of each receiver from Figure 2.1 and then sum the signals from all outputs, forming the system shown in Figure 2.2. In this system, the phase shifter connected to the p-th antenna introduces a phase shift denoted as ζ_p. This setup, known as a beamformer, produces a single output, enabling the entire array to function as a single-output antenna with a unique radiation pattern.

Given a fixed set of phase shifts $\{\zeta_p\}_{p=0}^{P-1}$, we can measure the total power at the single output for each angle of incidence $\theta \in [-\pi, \pi]$ in order to create the radiation pattern. However, as can be seen from Figure 2.3, the phase difference between every pair of antennas is the same for θ and $180° - \theta$. This indicates the presence of phase ambiguities for each pair of angles θ and $180° - \theta$ when measured in relation to any direction perpendicular to the line of the array's formation. This phenomenon is also evident in 2D antenna arrays, where phase ambiguity between two angles occurs with respect to the direction perpendicular to the plane in which the array is formed.

To address this issue, it is common practice to place a reflecting plane on one side of the antenna array. This approach offers two advantages: first, it helps eliminate phase ambiguity when the array is used in localization applications; second, it allows us to position electronics -such as amplification, mixing, and filtering- behind the reflecting plane without affecting the radiative characteristics of the antenna array.

To obtain a mathematical description, we must recognize that beamformers involve wave interference. Since we are adding P waves of the form $\cos(\omega_0 t + p\varphi + \zeta_p)$, with $p = 0, 1, ...,$

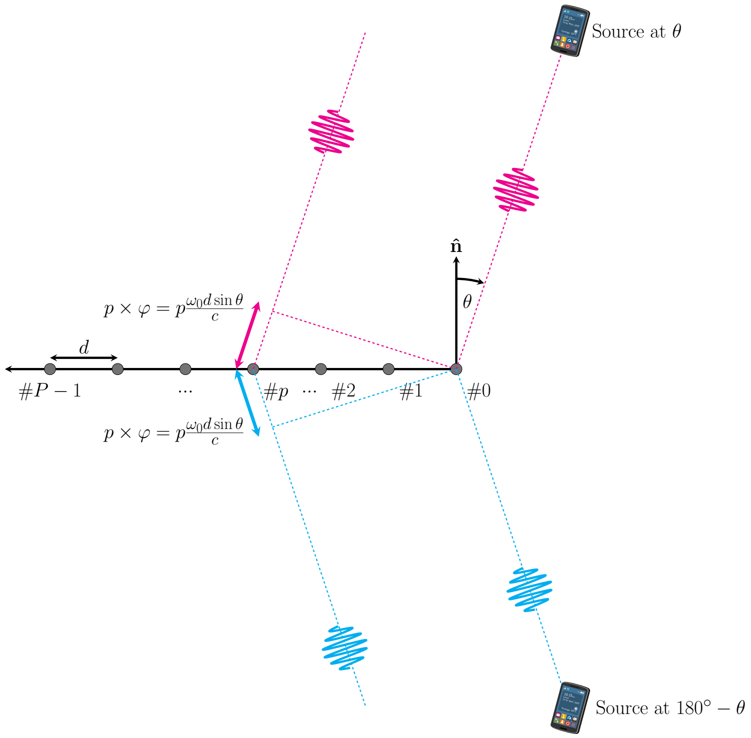


Figure 2.3: Example of a one-dimensional array that shows the phase ambiguity between each pair of angles θ and $180° - \theta$. As can be seen, the phase difference between every pair of antennas is the same for θ and $180° - \theta$.

$P-1$, the best way to analyze their interference is through the complex sum

$$S(t) = \sum_{p=0}^{P-1} e^{j(\omega_0 t + p\varphi + \zeta_p)}, \tag{2.6}$$

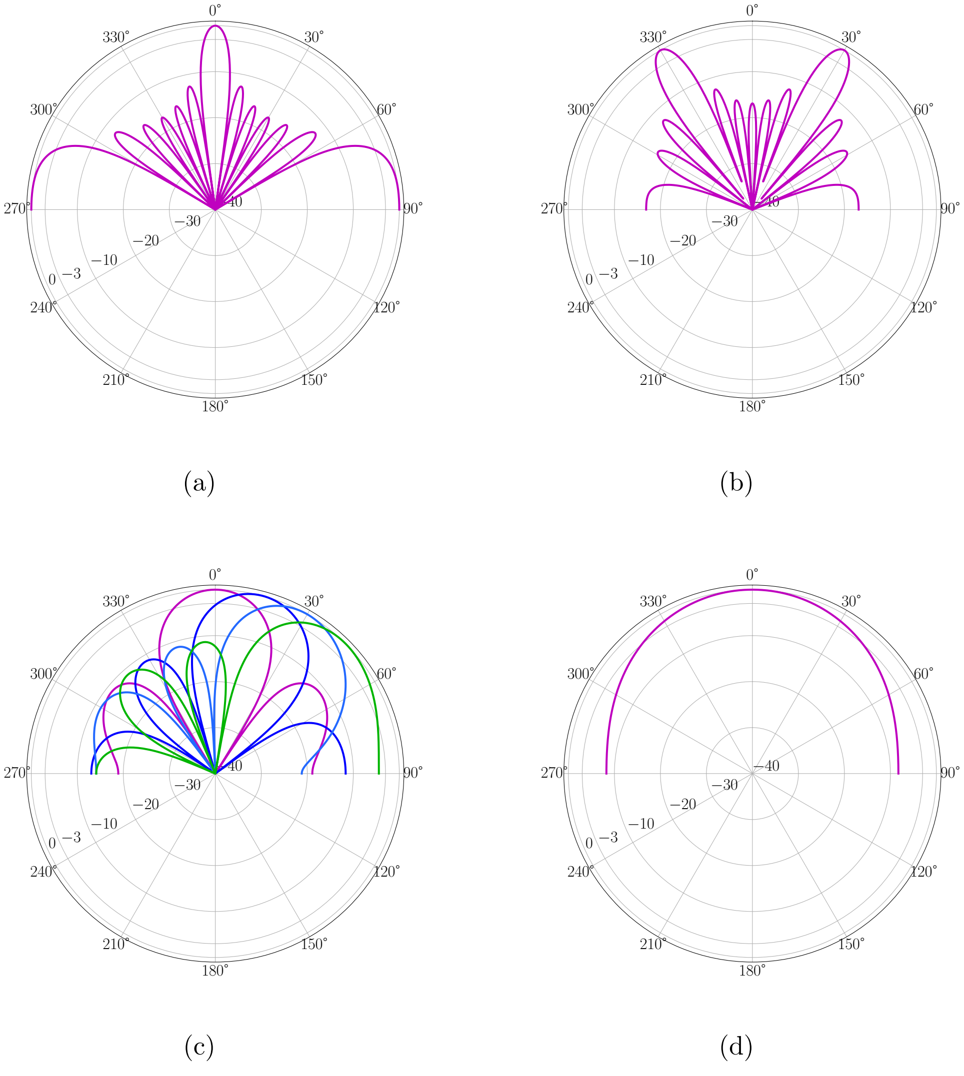
and then, the beamformer's output is given by $\mathbb{R}\{S(t)\}$.

Notice that the maximum of $S(t)$ is achieved only when all waves are in phase. If we initially ignore the phase shifters (i.e., assume they are all set to zero), this condition occurs when

$$\begin{aligned} \varphi = \pm 2\pi m \Leftrightarrow \frac{2\pi\Delta}{\lambda} = \pm 2\pi m \\ \Leftrightarrow \Delta = \pm m\lambda \Leftrightarrow \theta = \pm \sin^{-1}\left(\frac{\lambda m}{d}\right). \end{aligned} \tag{2.7}$$

Equation (2.7) indicates that a maximum of $S(t)$ occurs only when the path difference Δ for the signals impinging at an angle θ fits an integer multiple of the wavelength. For $d = \lambda$, illustrated in Figure 2.2, a full wavelength fits between the antennas along the array line (i.e., when $\theta = \pm 90°$), resulting in a maximum of $S(t)$ at $\theta = \pm 90°$. Since the phase shifters are turned off, a signal impinging normally on the array at $\theta = 0°$ also arrives in phase at all antennas, producing another maximum in $S(t)$.

The radiation pattern considering $\zeta_p = 0$ is shown in Figure 2.4(a) [1]. This pattern exhibits three main lobes, with maxima at $\theta = 0°$, $\theta = 90°$, and $\theta = -90°$. The lobe at $\theta = 0°$ depends only on the fact that $\zeta_p = 0$, while the other lobes are determined by the array spacing, as derived in (2.7). These additional lobes are called grating lobes because they depend on the array's grating.



(a)

(b)

(c)

(d)

Figure 2.4: Normalized array factors (in dB) for the beamformer example shown in Figure 2.2. (a) Array factor without any phase shifts ($\zeta_p = 0$). (b) Array factor with phase shifts $\zeta_p = (1-(-1)^p) \times 90°$. (c) Array factor for a 4×4 antenna array with a spacing of $d = 7.5$ cm, operating at a frequency of 1.85 GHz. The four colored curves represent different phase shift configurations $\{\zeta_p\}_{p=0}^{P-1}$. It can be observed that the direction of maximum received power can be adjusted by varying the values of ζ_p. (d) Array factor for the same 4×4 antenna array operating at 0.7 GHz.

---

[1] We only plot the pattern for $\theta \in [-\pi/2, \pi/2]$ since we consider the implementation of reflecting planes in practical applications (thus eliminating the radiation for $\theta \in [\pi/2, 3\pi/2]$).

Grating lobes arise because the phase differences between antennas are identical for $\theta = \pm 90^\circ$ and $\theta = 0^\circ$. If we slightly change the observed angle to $\theta = 90^\circ - \varepsilon$ (with ε small), the phase difference between antennas should be the same as that for $\theta = 0^\circ - \varepsilon$. Extending this reasoning, for $d = \lambda$, every measured phase difference is associated with two angles. This phenomenon is called phase ambiguity and presents challenges in many array applications. For example, in direction finding, phase ambiguity prevents us from determining whether a signal originated from angle θ_1 or $\theta_2 \neq \theta_1$. As a result, antenna arrays are typically designed to avoid phase ambiguity by ensuring that the spacing between antennas is less than $\lambda/2$ [22].

Now consider the scenario where ζ_p is not zero for all p. Specifically, let $\zeta_p = (1-(-1)^p) \times 90^\circ$ as illustrated in Figure 2.2. Here, with $d = \lambda$, the phase differences captured by the array at $\theta = \pm 30^\circ$ are compensated for by the phase shifters, creating a maximum in $S(t)$ at these angles. This results in the radiation pattern shown in Figure 2.4(b).

As demonstrated, adjusting the phase shifters ζ_p allows us to steer the main lobe of the antenna array, effectively scanning the space for sources, as depicted in Figure 2.4(c). This technique was utilized in the 2018 version of RadioVision, shown in Figure 1.1. The capability to resolve two sources depends on the width of the main lobe. According to the Rayleigh criterion [23], the resolution is determined by half the first-null beamwidth of the array, typically equal to the half-power beamwidth. From Figure 2.4(c), using a spacing $d = 7.5$ cm and $P = 4$ antennas for RadioVision at a frequency of 1.85 GHz, we observe that the resolution is approximately 30°. However, at a frequency of 0.7 GHz, the radiation pattern shown in Figure 2.4(d) does not exhibit nulls, indicating an inability to resolve two sources at this frequency. As a result, RadioVision would struggle to localize sources correctly. Consequently, RadioVision cannot operate effectively using the classic beamformer method, prompting us to explore the DoA estimation methods discussed in Section 2.3.

2.2.2 Microwave Receivers

A microwave receiver is a crucial component in radio communication systems, designed to capture and process microwave signals, typically ranging from 300 MHz to 300 GHz.

The typical architecture of a microwave receiver, illustrated in Figure 2.5, begins with an antenna that captures the incoming microwave signal, referred to as $s_{RF}(t)$. Once this signal is received, it is fed into a low-noise amplifier (LNA). The LNA amplifies the weak microwave signal while minimizing additional noise, critical for preserving the signal's quality.

Following amplification, the signal moves to the mixer, where the high-frequency microwave signal is downconverted to an intermediate frequency (IF), resulting in a signal $s_{IF}(t)$. The mixer achieves this by combining the microwave signal with a local oscillator (LO) signal, $V_l \cos(\omega_l t)$, producing both sum and difference frequencies.

After the frequency conversion, the intermediate frequency signal passes through a low-pass filter, which isolates the desired band and removes higher frequencies. Finally, the signal goes through an IF amplifier, which increases the system gain and prepares the signal for further processing or analysis.

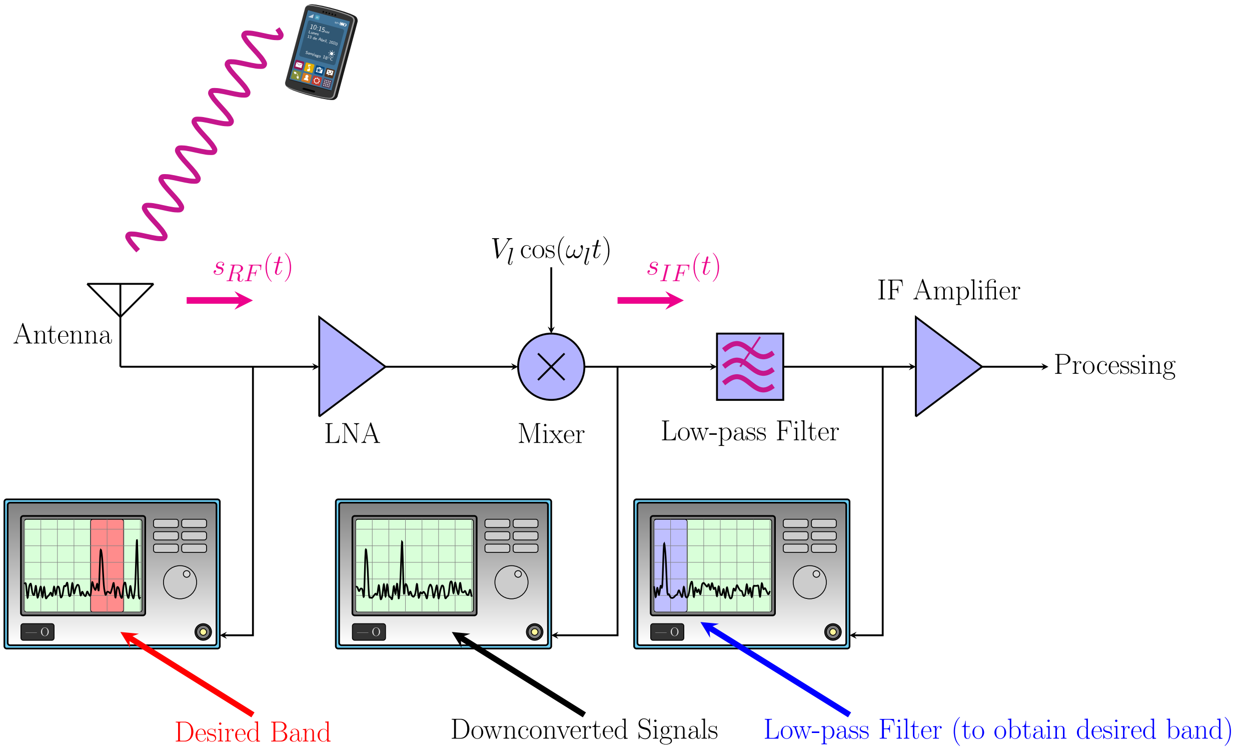


Figure 2.5: Schematic representation of a microwave receiver as used in RadioVision.

2.2.2.1 Mixers

Mixers play a crucial role in frequency conversion within a receiver. Ideally, a mixer operates by multiplying the received signal, $s_{RF}(t)$, by $\cos(\omega_l t)$, shifting the signal to a different band for further processing.

In a microwave receiver, the role of the mixer is to recover the baseband representation of the RF signal $s_{RF}(t)$, which has a bandwidth limited by $\pm\Delta\omega$, as shown in Figure 2.6 (the portion covered by $\Delta\omega$ is called the upper-side band, while the one covered by $-\Delta\omega$ is called the lower-side band). Mathematically, if $\hat{\boldsymbol{S}}_{RF}(\omega)$ represents the Fourier transform of $s_{RF}(t)$, then, a mixer should be able to recover $\hat{\boldsymbol{S}}_{RF}(\omega \pm \omega_l)$, where ω_l is such that $\hat{\boldsymbol{S}}_{RF}(\omega \pm \omega_l) \neq 0$ only for $\omega \in [-\Delta\omega, \Delta\omega]$, as shown in the figure.

For an ideal mixer, after multiplying $s_{RF}(t)$ by $\cos(\omega_l t)$, the Fourier transform of the resultant IF signal is given by

$$\begin{aligned} \hat{\boldsymbol{S}}_{IF}(\omega) &= \hat{\boldsymbol{S}}_{RF}(\omega) * \pi(\delta(\omega - \omega_l) + \delta(\omega + \omega_l)) \\ &= \pi\hat{\boldsymbol{S}}_{RF}(\omega - \omega_l) + \pi\hat{\boldsymbol{S}}_{RF}(\omega + \omega_l), \end{aligned} \tag{2.8}$$

where $*$ denotes convolution. Equation (2.8) shows that $\hat{\boldsymbol{S}}_{IF}(\omega)$ contains two copies of the original signal spectrum, shifted to $-\omega_l$ and ω_l, effectively enabling the frequency downconversion to recover the desired signal $\hat{\boldsymbol{S}}_{RF}(\omega \pm \omega_l)$ (by applying a low-pass filter).

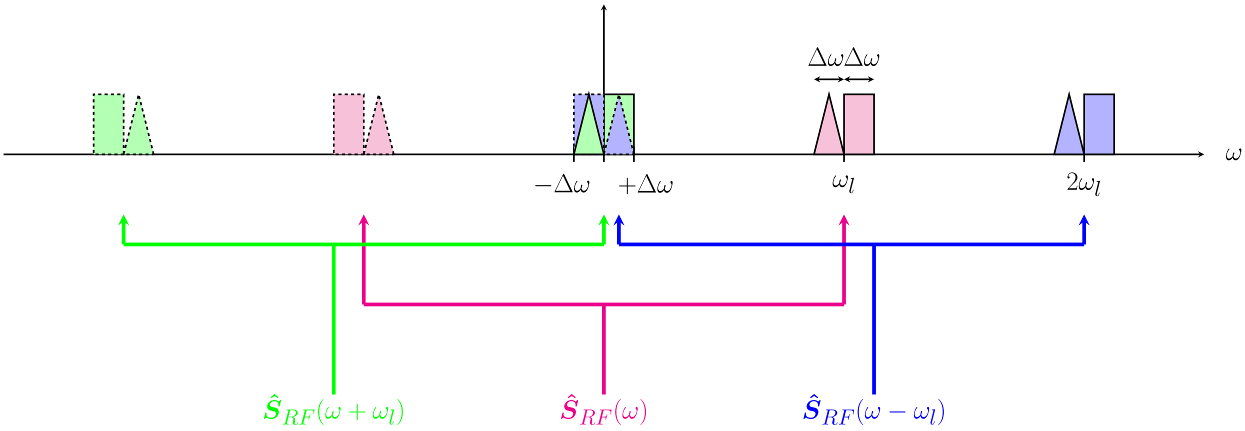


Figure 2.6: Representation of the upconversion and downconversion of an ideal mixer. $\hat{\boldsymbol{S}}_{RF}(\omega)$ represents the Fourier transform of $s_{RF}(t)$.

While this is the ideal behavior, no physical mixer can perfectly perform this multiplication. Real mixers rely on nonlinear devices -such as diodes or transistors- to approximate this multiplication. Despite their non-idealities, they still succeed in generating the desired frequency conversion by exploiting the nonlinear characteristics of these components. Indeed, let us consider a diode mixer as an example, characterized by a nonlinear current-voltage relation expressed as

$$I(V) = I_s(e^{\alpha V} - 1), \tag{2.9}$$

where I_s is the saturation current, and α is a constant related to the diode characteristics. We can express the exponential function using its Taylor series expansion:

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}. \tag{2.10}$$

When applying this to (2.9), we obtain:

$$I(V) = I_s\left(\alpha V + \frac{\alpha^2 V^2}{2!} + \frac{\alpha^3 V^3}{3!} + ...\right), \tag{2.11}$$

and if we consider $V(t) = V_l \cos\omega_l t + s_{RF}(t)$, then we can write

$$\begin{aligned} I(V) &= I_s \sum_{i=1}^{\infty} \frac{\alpha^i (V_l \cos \omega_l t + s_{RF}(t))^i}{i!} \\ &= I_s \sum_{i=1}^{\infty} \sum_{k=0}^{i} \frac{\alpha^i}{i!} \binom{i}{k} (V_l \cos \omega_l t)^{i-k} \times s_{RF}^k(t) \\ &= I_s \sum_{i=1}^{\infty} \sum_{k=0}^{i} \frac{\alpha^i}{k!(i-k)!} (V_l \cos \omega_l t)^{i-k} \times s_{RF}^k(t), \end{aligned} \tag{2.12}$$

were we used Newton's binomial theorem. Equation (2.12) captures the various contributions resulting from mixing the incoming microwave signal with the local oscillator signal. In this context, the IF signal is composed of an infinite series of terms of the form $(V_l \cos \omega_l t)^{i-k} \times s_{RF}^k(t)$, with k ranging from 0 to i. For instance, considering $i = 2$ illustrates that the current $I(V)$ is composed of the terms $V_l^2 \cos^2(\omega_l t) \propto 1 + \cos(2\omega_l t)$, the ideal mixer product $V_l \cos(\omega_l t) s_{RF}(t)$, and $s_{RF}^2(t)$. Similarly, for $i = 3$, the equation would include products of three signals, and this pattern continues indefinitely.

Although (2.12) is exact as long as the current-voltage relation (2.9) holds, this representation becomes cumbersome and impractical in many cases. Consequently, a small-signal model is often adopted, assuming that $V(t)$ remains small. In such a case, we can apply the approximation

$$e^x \approx \sum_{i=0}^{2} \frac{x^i}{i!}. \tag{2.13}$$

The rationale behind this approximation is grounded in the fact that, even after the LNA, the signals typically range around ≤ 6 dBm, which corresponds to RMS voltages of approximately $\sqrt{50\Omega \times 0.004 \text{ W}} \approx 0.4$ V, for which (2.13) holds. If power levels exceed 10 dBm, a third-order component $x^3/3!$ should be considered.

Using the approximation from (2.13) in the diode equation (2.9), and considering $V(t) = V_l \cos \omega_l t + s_{RF}(t)$, we derive

$$\begin{aligned} I(V) &\approx I_s \left(\alpha (V_l \cos(\omega_l t) + s_{RF}(t)) + \frac{\alpha^2}{2} (V_l^2 \cos^2(\omega_l t) + 2 V_l \cos(\omega_l t) s_{RF}(t) + s_{RF}^2(t)) \right) \\ &= I_s \left(\frac{\alpha^2 V_l^2}{4} + \alpha (V_l \cos(\omega_l t) + s_{RF}(t)) + \ldots \right. \\ &\left. \frac{\alpha^2}{2} \left(\frac{V_l^2 \cos(2\omega_l t)}{2} + 2 V_l \cos(\omega_l t) s_{RF}(t) + s_{RF}^2(t) \right) \right). \end{aligned} \tag{2.14}$$

In this expression, we can identify the frequency-shifting terms highlighted in blue. Let us denote those terms as

$$V_x(t) = \frac{V_l^2 \cos(2\omega_l t)}{2} + 2V_l \cos(\omega_l t) s_{RF}(t) + s_{RF}^2(t). \tag{2.15}$$

The term $2V_l \cos(\omega_l t) s_{RF}(t)$ corresponds to the ideal mixer output we seek, while the term $V_l^2 \cos(2\omega_l t)/2$ represents a harmonic signal with frequency $2\omega_l$. To analyse the remaining term $s_{RF}(t)$, let us consider the Fourier transform of $s_{RF}(t)$, $\hat{\boldsymbol{S}}(\omega)$. Then, we can express $s_{RF}^2(t)$ as

$$s_{RF}^2(t) = \left(\int_{-\infty}^{\infty} \hat{\boldsymbol{S}}(\omega) e^{j\omega t} d\omega\right)\left(\int_{-\infty}^{\infty} \hat{\boldsymbol{S}}(\omega) e^{j\omega t} d\omega\right). \tag{2.16}$$

This representation implies that $s_{RF}^2(t)$ corresponds to the summation of all possible products of the form $e^{j\omega_1 t} e^{j\omega_2 t}$ for each combination of $\omega_1, \omega_2 \in (-\infty, \infty)$. Colloquially, this means that $s_{RF}^2(t)$ corresponds to the mixing of all the harmonic components that make up $s_{RF}(t)$. More rigorously, we recognize that the Fourier transform of the product of two signals results in the convolution of their respective Fourier transforms. Therefore, we can express the Fourier transform of $s_{RF}^2(t)$, $\mathcal{F}\{s_{RF}^2(t)\}$, as

$$\mathcal{F}\{s_{RF}^2(t)\} = \int_{-\infty}^{\infty} \hat{\boldsymbol{S}}(\Omega) \hat{\boldsymbol{S}}(\omega - \Omega) d\Omega, \tag{2.17}$$

which reflects how each frequency component of $s_{RF}(t)$ interacts with every other frequency component. Consequently, we can write

$$s_{RF}^2(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{\boldsymbol{S}}(\Omega) \hat{\boldsymbol{S}}(\omega - \Omega) e^{j\omega t} d\Omega d\omega. \tag{2.18}$$

To illustrate the implications of (2.18), let's consider a practical example where the signal $s_{RF}(t)$ is defined as

$$s_{RF}(t) = \cos(\omega_r t) \operatorname{sinc}\left(\frac{\Delta\omega}{2\pi} t\right). \tag{2.19}$$

In this case, $s_{RF}(t)$ represents a signal with a constant amplitude spectrum bandlimited within $\Delta\omega$ and centered around the frequency ω_r, as shown in Figure 2.7(a).

Now, when we compute $s_{RF}^2(t)$, we perform a convolution of the spectrum of $s_{RF}(t)$ with itself. This graphical convolution will yield the spectrum of $s_{RF}^2(t)$ as shown in Figure 2.7(b).

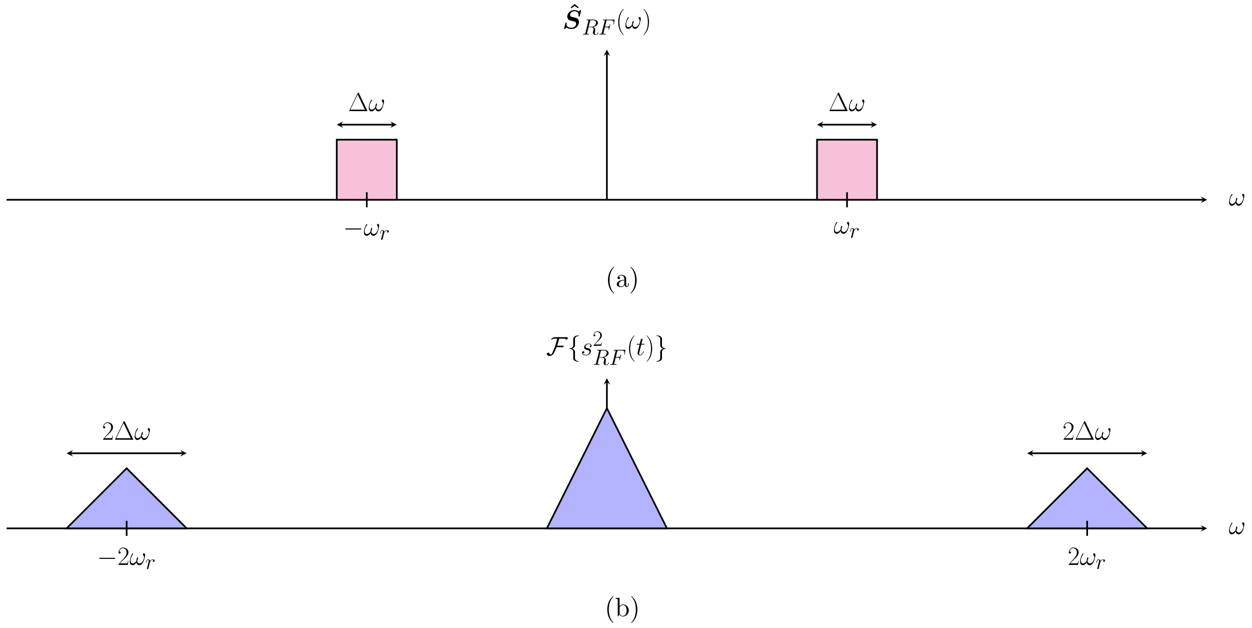


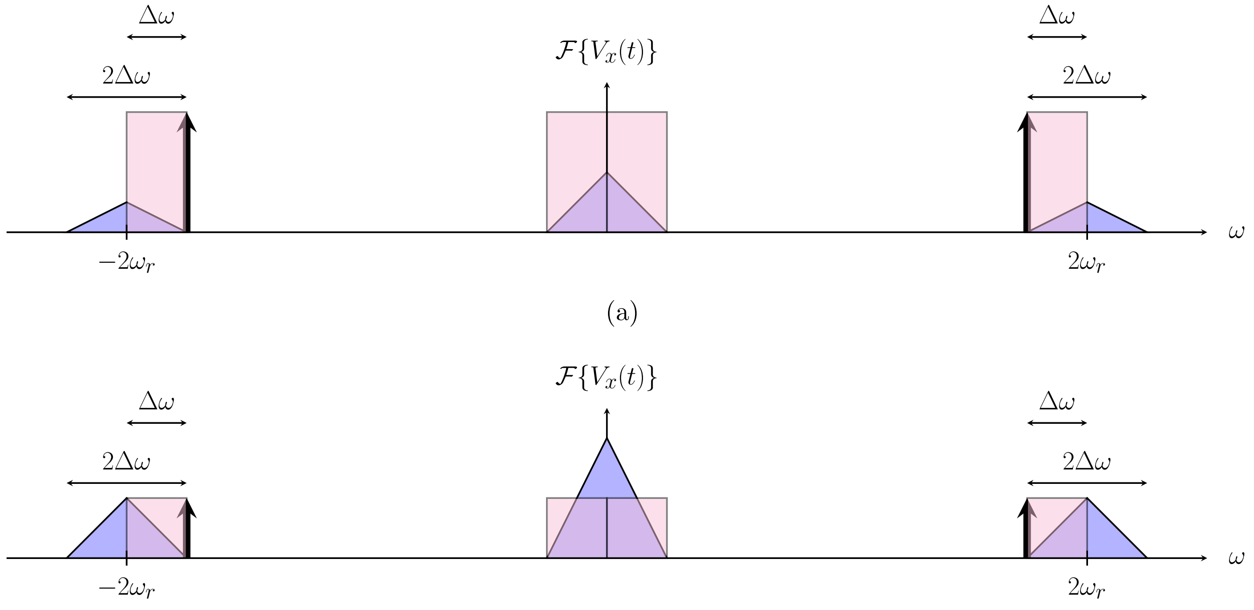
Figure 2.7: Representation of the self-mixing components within $s_{RF}(t)$. (a) Fourier transform of $s_{RF}(t)$. (b) Fourier transform of $s_{RF}^2(t)$.

The resulting spectrum contains a component with a bandwidth of $2\Delta\omega$ centered at $\omega = 0$. If $\Delta\omega$ is broad enough, the products of $s_{RF}^2(t)$ can fall into the IF (outside of just the DC component), thus increasing the noise floor and contaminating the desired downconverted signals. The derived conclusion highlights that this result does not depend on the local oscillator frequency ω_l. Instead, it emphasizes that any band with sufficient bandwidth and power can adversely affect the noise floor observed in the IF.

Finally, in Figure 2.8, we illustrate the total spectrum of $V_x(t)$ for two distinct scenarios, both with the LO frequency set to $\omega_l = \omega_r - \Delta\omega/2$. In panel (a), we show the case where the LO power is significantly higher than the RF power. In this situation, the desired frequency-shifted version of $\hat{\boldsymbol{S}}_{RF}(\omega)$, i.e., $\hat{\boldsymbol{S}}_{RF}(\omega + \omega_l)$, is clearly distinguishable. The main signal is shifted to the IF band as expected, and any noise from the $s_{RF}^2(t)$ component remains relatively small, allowing for the desired signal to be recovered with minimal interference. In panel (b), the RF power exceeds the LO power. In this case, we observe that the signal $\hat{\boldsymbol{S}}_{RF}(\omega+\omega_l)$ becomes buried in the noise resulting from the mixing process, making it difficult or even impossible to recover the signal of interest cleanly.

2.2.2.2 DSB Mixers

As shown in Figure 2.6, if the mixing process only involves multiplying $s_{RF}(t)$ with $\cos\omega_l t$, both the upper and lower sideband appear in the IF. Indeed, the lower sideband (corresponding to the signal with a triangular spectrum in the figure) appears in the IF due to the translation $(\omega - \omega_l)$. In contrast, the upper sideband (corresponding to the signal with a rectangular spectrum in the figure) appears in the IF due to the translation $(\omega + \omega_l)$. When the mixing process is such that the upper and lower sidebands appear mixed in the same IF, the mixer is called a double sideband (DSB) mixer.



(a)

(b)

Figure 2.8: Representation of the frequency-shifting terms $V_x(t)$ for two scenarios. (a) LO power $\gg$ RF power. (a) RF power $\gtrsim$ LO power.

In many cases, using DSB mixers is counterproductive, as it mixes the two sidebands, making it impossible to distinguish the original information from each sideband. However, in other situations, placing the LO in strategic positions is possible, such that only one sideband contains valuable information. In such cases, all information can be recovered in the IF. As shown in Section 4.1.2, for the RadioVision application, the LO can always be placed such that only one relevant band appears in the IF.

2.3 DoA Estimation Algorithms

This section begins with a review of the core DoA estimation methods that serve as the foundation for this work, following the progression from the MUSIC algorithm to ESPRIT.

We start by outlining the basic concepts behind the MUSIC algorithm, which uses spectral analysis to estimate the angles of arrival. ESPRIT builds upon this by simplifying the computational process by exploiting the rotational invariance property of antenna arrays. U-ESPRIT 2D then extends ESPRIT for two-dimensional arrays, making it suitable for Uniform Rectangular Arrays (URAs) such as RadioVision.

For clarity, the review of MUSIC and ESPRIT will focus on one-dimensional Uniform Linear Arrays (ULAs). In contrast, the more relevant U-ESPRIT 2D is reviewed in detail for two dimensions in Chapter 3, aligning with the system requirements of RadioVision. Although these methods can be generalized for non-uniform arrays or varied antenna types, this thesis concentrates on uniform array configurations with identical antennas.

2.3.1 MUSIC: Multiple Emitter Location and Signal Parameter Estimation

2.3.1.1 Signal Model

Let us consider a ULA with P elements, surrounded by r sources positioned at angles $\{\theta_l\}_{l=0}^{r-1}$, as shown in Figure 2.9. The array elements are uniformly spaced by a distance d, with the origin of the coordinate system located at the 0-th antenna (see the figure). All the sources are narrowband, centered around a carrier frequency ω_0, uncorrelated, and located within an isotropic and non-dispersive medium in the far field of the array. Under these conditions, the radiation from the sources can be approximated as plane waves impinging on the array.

The signal radiated by the l-th source can be written as $s_l(t) = u_l(t)e^{j(\omega_0 t + v_l(t))}$, where $u_l(t)$ and $v_l(t)$ are the amplitude and phase modulations of $s_l(t)$. Since we are working with complex signals, we assume that the receiver incorporates an IQ mixer, which enables processing these complex representations. In free space, the real component of the signal would be $u_l(t)\cos(\omega_0 t + v_l(t))$. Still, by employing an IQ mixer, we can extend this real signal to its complex form, $u_l(t)\cos(\omega_0 t + v_l(t)) + ju_l(t)\sin(\omega_0 t + v_l(t))$, allowing us to handle both in-phase and quadrature components simultaneously in the digital processing system.

Using the concepts developed in Section 2.2.1, we know that each signal $s_l(t)$ arrives at each antenna element at slightly different times due to the array's geometry. Specifically, if we consider the signal arrival time at the reference antenna (the 0-th antenna), the time delay between the p-th and 0-th antennas (Figure 2.9) is given by

$$\tau_p(\theta_l) = \frac{pd\sin\theta_l}{c} \tag{2.20}$$

Hence, the total signal impinging at the p-th antenna can be expressed as

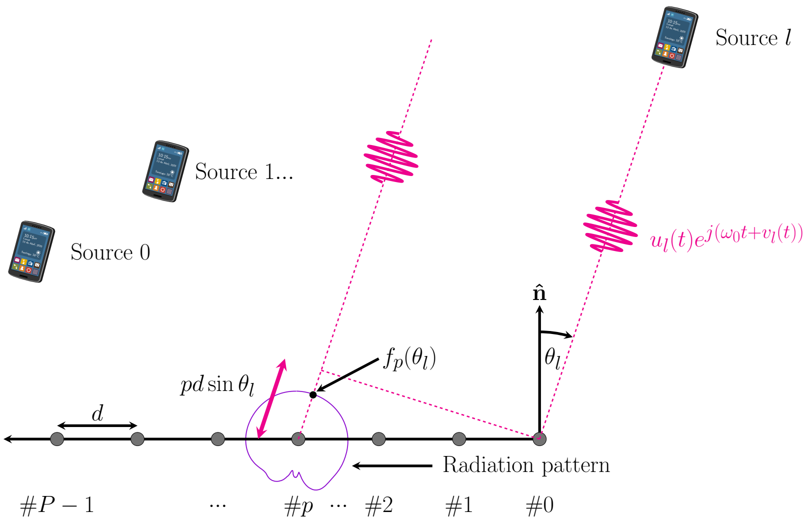


Figure 2.9: Array geometry used in the MUSIC algorithm.

$$x_p(t) = \sum_{l=0}^{r-1} f_p(\theta_l) s_l(t - \tau_p(\theta_l)) + \eta_p(t), \tag{2.21}$$

where $f_p(\theta_l)$ is the complex radiation pattern of the p-th antenna evaluated at θ_l, and $\eta_p(t)$ is the additive noise present at the p-th antenna. For simplicity, we assume that $f_p(\theta_l)$ does not vary with frequency and does not alter the time characteristics of the signal. Further details on this assumption can be found in Appendix C.

If we assume that all antennas are identical and that coupling effects between antennas are negligible, the complex radiation pattern can be considered the same for all antennas, allowing us to drop the index p from f_p and write

$$\begin{aligned} x_p(t) &= \sum_{l=0}^{r-1} f(\theta_l) s_l(t - \tau_p(\theta_l)) + \eta_p(t) \\ &= \sum_{l=0}^{r-1} s_l(t - \tau_p(\theta_l)) + \eta_p(t), \end{aligned} \tag{2.22}$$

where we have absorbed the effects of the radiation pattern $f(\theta_l)$ into the signal s_l by updating the signal as $s_l \leftarrow f(\theta_l) s_l$.

Now, since the signals are narrowband, $u_l(t)$ and $v_l(t)$ vary slowly compared to $e^{j(\omega_0 t)}$. This allows us to make the following approximation:

$$\begin{aligned} s_l(t - \tau_p(\theta_l)) &\approx u_l(t) e^{j(\omega_0(t - \tau_p(\theta_l)) + v_l(t))} \\ &= e^{-j\omega_0 \tau_p(\theta_l)} s_l(t). \end{aligned} \tag{2.23}$$

Therefore, the signal received by the p-th antenna becomes

$$x_p(t) = \sum_{l=0}^{r-1} s_l(t) e^{-j\omega_0 \tau_p(\theta_l)} + \eta_p(t), \tag{2.24}$$

or equivalently,

$$\begin{bmatrix} x_0(t) \\ x_1(t) \\ \vdots \\ x_{P-1}(t) \end{bmatrix} = \sum_{l=0}^{r-1} \begin{bmatrix} e^{-j\omega_0 \tau_0(\theta_l)} \\ e^{-j\omega_0 \tau_1(\theta_l)} \\ \vdots \\ e^{-j\omega_0 \tau_{P-1}(\theta_l)} \end{bmatrix} s_l(t) + \begin{bmatrix} \eta_0(t) \\ \eta_1(t) \\ \vdots \\ \eta_{P-1}(t) \end{bmatrix} \tag{2.25}$$

Let us define the $P \times r$ matrix $\mathbf{A}$, whose l-th column is the so-called steering vector for the direction θ_l: $\left[e^{-j\omega_0\tau_0(\theta_l)}\ e^{-j\omega_1\tau_0(\theta_l)}\ \ldots\ e^{-j\omega_0\tau_{P-1}(\theta_l)}\right]^T$. Since the summation in (2.25) is a linear combination of column vectors, we can write the signal model in matrix form as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\eta}(t), \tag{2.26}$$

where $(\mathbf{x})_p = x_p(t)$, $(\mathbf{s})_l = s_l(t)$, and $(\boldsymbol{\eta})_p = \eta_p(t)$. This signal model describes the data immediately after the antenna array. The additional effects introduced by the receiver hardware, such as amplification, downconversion, and analog-to-digital conversion, are further detailed in Appendices C and D. Still, these processes do not alter the general structure of the signal model given in (2.26).

2.3.1.2 Array Manifold and the column space of A

The array manifold $\mathcal{A}$ of an antenna array is the set of all possible steering vectors corresponding to different angles of arrival:

$$\mathcal{A} = \{\mathbf{a}(\theta) = \left[e^{-j\omega_0\tau_0(\theta_l)}\ e^{-j\omega_1\tau_0(\theta_l)}\ \ldots\ e^{-j\omega_0\tau_{P-1}(\theta_l)}\right]^T \in \mathbb{C}^P : \theta \in [-\pi, \pi]\}. \tag{2.27}$$

It is important to note that $\mathcal{A}$ is not a vector subspace. For instance, given three different angles $\theta_1, \theta_2, \theta_3$, it is not possible to express $\mathbf{a}(\theta_1) = \beta\mathbf{a}(\theta_2) + \gamma\mathbf{a}(\theta_3)$ for some constants β and γ, considering that the array does not present phase ambiguities, as it is shown in Appendix B.

This distinction highlights that the array manifold behaves more like a curved path, or as described in [7], it resembles a rope weaving through the complex space $\mathbb{C}^P$ (in this particular case that we are considering a unidimensional array).

Note that the array manifold depends solely on the structure of the antenna array, meaning it is a known, deterministic geometric entity. Each column of $\mathbf{A}$ must lie on this array manifold, as depicted in Figure 2.10. Since the array manifold is not a vector space, linear combinations of the columns of $\mathbf{A}$ will not necessarily belong to $\mathcal{A}$. Nevertheless, if we form the column space of $\mathbf{A}$, $\mathcal{C}(\mathbf{A})$ (see Appendix A), the array manifold must intersect it. Therefore, if we can accurately estimate $\mathcal{C}(\mathbf{A})$, we can determine $\{\theta_l\}_{l=0}^{r-1}$, as shown in the example from Figure 2.10.

2.3.1.3 Estimating $\mathcal{C}(\mathbf{A})$

Since the sample vectors $\mathbf{x}(t)$ are shifted by $\boldsymbol{\eta}(t)$ from the column space of $\mathbf{A}$, after taking enough samples, we expect these vectors to cluster around $\mathcal{C}(\mathbf{A})$, as shown in Figure 2.11. Mathematically, we collect $N \gg P$ samples $\{\mathbf{x}(t_n)\}_{n=0}^{N-1}$, and stack them into a matrix

$$\mathbf{X} = \left[\ \mathbf{x}(t_0)\ \ \mathbf{x}(t_1)\ \ \ldots\ \ \mathbf{x}(t_{N-1})\ \right]. \tag{2.28}$$

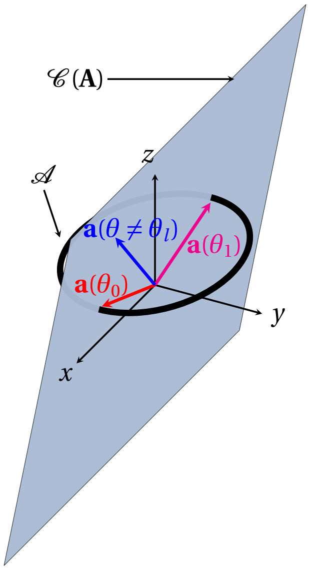


Figure 2.10: Simplified illustration of the array manifold and the column space of $\mathbf{A}$. We are considering an antenna number $P = 3$ and two sources ($r = 2$) for this illustration. Note that both $\mathcal{A}$ and $\mathcal{C}(A)$ are complex sets. Therefore, we cannot accurately illustrate the total space $\mathbb{C}^3$ (with dimension 6), nor the column space (with dimension 4). However, to illustrate the key ideas, we will consider a representation in the real space. In this case, the total space is $\mathbb{R}^P = \mathbb{R}^3$, and the column space of $\mathbf{A}$ is a plane of dimension $r = 2$.

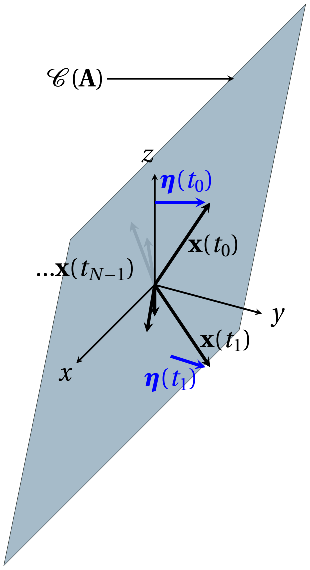


Figure 2.11: Illustration of how the sample vectors $\{\mathbf{x}(t_n)\}_{n=0}^{N-1}$ cluster around $\mathcal{C}(\mathbf{A})$.

Despite the presence of noise, which causes the sample vectors to become linearly independent and results in the column space of $\mathbf{X}$ being $\mathbb{C}^P$ rather than $\mathcal{C}(\mathbf{A})$, we can still extract a meaningful basis for $\mathcal{C}(\mathbf{A})$. By applying the singular value decomposition (SVD) [33, 34], we can optimally identify the most significant components of the column space of $\mathbf{X}$. The SVD reveals the ”most important” vectors in determining the column space, known as the principal components. Expressly, this information is provided by the left singular vectors and the singular values of $\mathbf{X}$, which correspond to the eigenvectors and the square root of the eigenvalues of $\mathbf{X}\mathbf{X}^H$.

What makes the SVD particularly useful here is that even though $\mathcal{C}(\mathbf{X}) = \mathbb{C}^P$ due to noise, it highlights the directions that are most relevant for the signal. Since the noise only perturbs $\mathbf{X}$ from its actual signal space $\mathcal{C}(\mathbf{A})$, we expect that the left singular vectors corresponding to the largest eigenvalues form a robust basis for $\mathcal{C}(\mathbf{A})$.

We first compute $\mathbf{X}\mathbf{X}^H$ as outlined in (2.29) to eventually obtain the left singular vectors and singular values. Since the signals are assumed to be zero-mean and uncorrelated, $\mathbf{X}\mathbf{X}^H$ corresponds to the estimated covariance matrix of the received signals, and is given by

$$\mathbf{X}\mathbf{X}^H = (\mathbf{A}\mathbf{S} + \mathbf{N})(\mathbf{A}\mathbf{S} + \mathbf{N})^H, \tag{2.29}$$

where

$$\begin{aligned} \mathbf{S} &= \left[\begin{array}{cccc} \mathbf{s}(t_0) & \mathbf{s}(t_1) & ... & \mathbf{s}(t_{N-1}) \end{array} \right] \\ \mathbf{N} &= \left[\begin{array}{cccc} \boldsymbol{\eta}(t_0) & \boldsymbol{\eta}(t_1) & ... & \boldsymbol{\eta}(t_{N-1}) \end{array} \right]. \end{aligned} \tag{2.30}$$

By expanding (2.29), we obtain

$$\begin{aligned} \mathbf{X}\mathbf{X}^H &= \mathbf{A}\mathbf{S}\mathbf{S}^H\mathbf{A}^H + \mathbf{A}\mathbf{S}\mathbf{N}^H + \mathbf{N}\mathbf{S}^H\mathbf{A}^H + \mathbf{N}\mathbf{N}^H \\ &= \mathbf{A}\mathbf{S}\mathbf{S}^H\mathbf{A}^H + \sigma^2\mathbf{I} \end{aligned} \tag{2.31}$$

where we have considered that the added noise is modeled as white and uncorrelated, with zero-mean and covariance matrix $\mathbf{N}\mathbf{N}^H = \sigma^2\mathbf{I}$.

Let us now analyze the eigenvectors of $\mathbf{X}\mathbf{X}^H$ (i.e., the left singular vectors of the SVD). We know that all eigenvectors of a matrix $\mathbf{A}$ must lie either in $\mathcal{C}(\mathbf{A})$ or $\mathcal{N}(\mathbf{A})$ (see Appendix A). Since $\mathbf{X}\mathbf{X}^H$ is Hermitian by construction, all eigenvectors are orthogonal, and all eigenvalues are real (see Theorem A.10). To gain insight into this decomposition, consider first the idealized case without noise, i.e., where $\mathbf{X}\mathbf{X}^H = \mathbf{A}\mathbf{S}\mathbf{S}^H\mathbf{A}^H$.

In the noise-free case, $\operatorname{rank}(\mathbf{X}\mathbf{X}^H) = r$, $\mathcal{C}(\mathbf{X}\mathbf{X}^H) = \mathcal{C}(\mathbf{A})$ and $\mathcal{N}(\mathbf{X}\mathbf{X}^H) = \mathcal{N}(\mathbf{A})$. Also, we have $\dim(\mathcal{C}(\mathbf{A})) = r$ and $\dim(\mathcal{N}(\mathbf{A})) = P - r$. Then, given that $\mathbf{X}\mathbf{X}^H$ is Hermitian, there are exactly r eigenvectors in $\mathcal{C}(\mathbf{A})$ and $P - r$ in $\mathcal{N}(\mathbf{A})$, with the latter being orthogonal to the columns of $\mathbf{A}$. Therefore, we have:

(1) For the eigenvectors in $\mathcal{N}(\mathbf{A})$, these must be orthogonal to $\mathbf{A}$, meaning they satisfy $\mathbf{A}^H\mathbf{v}_n = \mathbf{0}$.

(2) For the eigenvectors in $\mathcal{C}(\mathbf{A})$, they must satisfy $\mathbf{A}\mathbf{S}\mathbf{S}^H\mathbf{A}^H\mathbf{v}_s = \lambda\mathbf{v}_s$ with $\lambda \neq 0$.

This gives us a hint as to what to examine in the case where there is noise. For case (1) listed above, the eigenvectors $\mathbf{v}_n$ of $\mathbf{A}\mathbf{S}\mathbf{S}^H\mathbf{A}^H$ (equal to $\mathbf{X}\mathbf{X}^H$ only when there is no noise) that are orthogonal to the columns of $\mathbf{A}$ satisfy

$$\mathbf{X}\mathbf{X}^H\mathbf{v}_n = \mathbf{A}\mathbf{S}\mathbf{S}^H\mathbf{A}^H\mathbf{v}_n + \sigma^2\mathbf{I}\mathbf{v}_n = \sigma^2\mathbf{v}_n, \tag{2.32}$$

indicating that $\mathbf{v}_n$ are also eigenvectors of $\mathbf{XX}^H$ with eigenvalues σ^2. Since $\dim(\mathcal{N}(\mathbf{A})) = P - r$, there are $P - r$ such eigenvectors, denoted as $\{\mathbf{v}_{nl}\}_{l=0}^{(P-r)-1}$, each with eigenvalue σ^2. In summary, from case (1), we are concluding that the eigenvectors in $\mathcal{N}(\mathbf{A})$ do not change due to noise, as long as this is effectively white noise, uncorrelated and independent, with $\mathbf{NN}^H = \sigma^2\mathbf{I}$.

For case (2) listed above, the eigenvectors $\mathbf{v}_s$ of $\mathbf{ASS}^H\mathbf{A}^H$ satisfy

$$\mathbf{XX}^H\mathbf{v}_s = \mathbf{ASS}^H\mathbf{A}^H\mathbf{v}_s + \sigma^2\mathbf{I}\mathbf{v}_s = \lambda\mathbf{v}_n + \sigma^2\mathbf{v}_s = (\lambda + \sigma^2)\mathbf{v}_s. \tag{2.33}$$

Thus, $\mathbf{v}_s$ are also eigenvectors of $\mathbf{XX}^H$ with eigenvalues $(\lambda + \sigma^2)$. Since $\dim(\mathcal{C}(\mathbf{A})) = r$, there are r such eigenvectors, denoted as $\{\mathbf{v}_{sl}\}_{l=0}^{r-1}$, though they may have different eigenvalues.

This demonstrates that noise does not affect the eigenvectors. Consequently, if the assumptions listed below hold, the column space of $\mathbf{A}$ -determined entirely by the r eigenvectors $\{\mathbf{v}_{sl}\}_{l=0}^{r-1}$-, and therefore the DoAs $\{\theta_l\}_{l=0}^{r-1}$, can be perfectly identified, **independently of the array size**.

MUSIC Assumptions

1) All antennas are identical, and mutual coupling is negligible.
2) The sources are located in the far field of the array, with no obstacles nearby.
3) The antenna array is designed without phase ambiguities.
4) The data collection (N) is sufficiently large so that the estimated covariance matrix accurately represents the true covariance matrix.
5) The noise is white, independent and identically distributed, with zero mean.

Now, notice that even though we have the set of eigenvectors $\{\mathbf{v}_{sl}\}_{l=0}^{r-1}$, determining the vectors $\mathbf{a}(\theta_0), \mathbf{a}(\theta_1), ..., \mathbf{a}(\theta_{r-1}) \in \mathcal{A}$ that intersect with $\mathcal{C}(\mathbf{A})$ is not straightforward. For example, one might consider taking the dot product between each unknown $\mathbf{a}(\theta) \in \mathcal{A}$ and $\{\mathbf{v}_{sl}\}_{l=0}^{r-1}$, but this will not yield meaningful results, as can be inferred from Figure 2.12 (for example, the dot product between $\boldsymbol{v}_{s1}$ and $\boldsymbol{a}(\theta_1)$ is large, while the dot product between $\boldsymbol{v}_{s0}$ and $\boldsymbol{a}(\theta_0)$ is almost 0). However, if we instead compute the dot product between each $\mathbf{a}(\theta) \in \mathcal{A}$ and $\{\mathbf{v}_{nl}\}_{l=0}^{(P-r)-1}$, only the solution vectors $\mathbf{a}(\theta_0), \mathbf{a}(\theta_1), ..., \mathbf{a}(\theta_{r-1}) \in \mathcal{A}$ will yield a zero result. Hence, we could compute the normalized dot product

$$\mathbf{p}^T = \frac{\mathbf{a}(\theta)^H \overbrace{\begin{bmatrix}\mathbf{v}_{n0} \; \mathbf{v}_{n1} \; ... \; \mathbf{v}_{n(P-r-1)}\end{bmatrix}}^{\mathbf{V}_n}}{|\mathbf{a}(\theta)|}, \tag{2.34}$$

for each possible $\theta \in [-\pi, \pi]$ (scanning the array manifold). Only at the actual DoAs will this computation approach the $1 \times (P - r)$ zero vector $\mathbf{0}^T$. Moreover, if we then compute $\mathbf{p}^T\mathbf{p}$, given by

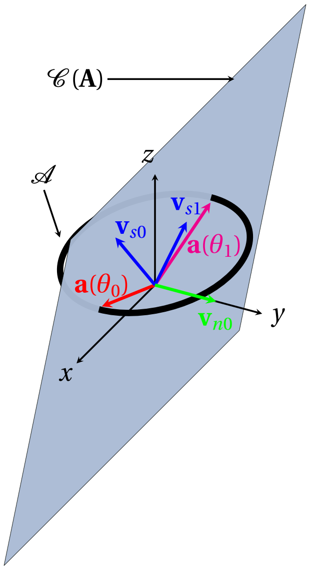


Figure 2.12: Representation of the eigenvectors in the column space $\mathcal{C}(\mathbf{A})$ and the null space $\mathcal{N}(\mathbf{A})$. In this example, with $P = 3$ antennas and $r = 2$ sources, the null space $\mathcal{N}(\mathbf{A})$ forms a one-dimensional line, as its dimension is $P - r = 1$.

$$\mathbf{p}^T\mathbf{p} = \frac{\mathbf{a}(\theta)^H\mathbf{V}_n\mathbf{V}_n^H\mathbf{a}(\theta)}{\mathbf{a}(\theta)^H\mathbf{a}(\theta)}, \tag{2.35}$$

only for the DoAs this computation will be close to 0 (scalar). This idea forms the basis of the MUSIC spectrum [6], defined as

$$P_{\text{MUSIC}}(\theta) = \frac{\mathbf{a}(\theta)^H\mathbf{a}(\theta)}{\mathbf{a}(\theta)^H\mathbf{V}_n\mathbf{V}_n^H\mathbf{a}(\theta)}. \tag{2.36}$$

The estimated DoAs are the r maxima of $P_{\text{MUSIC}}(\theta)$. With this, the MUSIC algorithm is summarized as follows:

MUSIC Algorithm

1) Collect N data samples $\mathbf{X}$ and estimate the covariance matrix $\mathbf{X}\mathbf{X}^H$.
2) Perform the eigendecomposition of $\mathbf{X}\mathbf{X}^H$: $\mathbf{X}\mathbf{X}^H = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$.
3) Estimate the number of sources r by counting the number of eigenvalues close to σ^2, which gives $P - r$, and hence r.
4) Evaluate $P_{\text{MUSIC}}(\theta) = \frac{\mathbf{a}(\theta)^H\mathbf{a}(\theta)}{\mathbf{a}(\theta)^H\mathbf{V}_n\mathbf{V}_n^H\mathbf{a}(\theta)}$ for $\theta \in [-\pi, \pi]$.
5) Identify the r largest peaks of $P_{\text{MUSIC}}(\theta)$ to estimate the DoAs.

2.3.2 ESPRIT: Estimation of Signal Parameters via Rotational Invariance Techniques

The MUSIC algorithm is based on the principle that vector samples collected by an antenna array cluster around the column space of the matrix **A**, intersecting the array manifold only at the DoAs. Importantly, MUSIC does not require specific antenna array characteristics; it can operate effectively with arbitrary antennas placed in random positions, provided there are no phase ambiguities. However, this underutilization of resources indicates that if the specific features of the array were taken advantage of, the limitations of MUSIC could be improved.

Based on MUSIC's principles, the ESPRIT algorithm enhances performance by leveraging the typical characteristics of arrays for increased accuracy. Specifically, we will focus on the one-dimensional array depicted in Figure 2.9 [2]. Thus, we adopt the model presented in equation (2.26).

Next, we introduce the array characteristic that the ESPRIT algorithm utilizes: every antenna element experiences the same time delay relative to its adjacent antenna when excited by an incoming signal. In other words, for an incoming signal $s_l(t)$ with DoA θ_l, the time delay between the 0-th and 1-th antennas, as well as between the 1-th and 2-th antenna, and so forth, up to the p-th and the $(p+1)$-th antenna, is given by $(d\sin\theta_l)/c$. To understand the implications of this in the signal model, let us first consider the case without noise, where

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) = \sum_{l=0}^{r-1} \begin{bmatrix} e^{-j\omega_0\tau_0(\theta_l)} \\ e^{-j\omega_0\tau_1(\theta_l)} \\ \vdots \\ e^{-j\omega_0\tau_{P-1}(\theta_l)} \end{bmatrix} s_l(t) = \sum_{l=0}^{r-1} \begin{bmatrix} 1 \\ e^{-j\mu(\theta_l)} \\ \vdots \\ e^{-j(P-1)\mu(\theta_l)} \end{bmatrix} s_l(t), \tag{2.37}$$

being $\mu(\theta_l) = \omega_0\tau_1(\theta_l)$. Then, notice that for all columns $\{\mathbf{a}_l\}_{l=0}^{r-1}$ of **A**, it happens that

$$e^{-j\mu(\theta_l)} \times \underbrace{\begin{bmatrix} 1 \\ e^{-j\mu(\theta_l)} \\ e^{-j2\mu(\theta_l)} \\ e^{-j3\mu(\theta_l)} \\ \vdots \\ e^{-j(P-3)\mu(\theta_l)} \\ e^{-j(P-2)\mu(\theta_l)} \end{bmatrix}}_{\text{First } P-1 \text{ elements of the } l\text{-th column of } \mathbf{A}} = \underbrace{\begin{bmatrix} e^{-j\mu(\theta_l)} \\ e^{-j2\mu(\theta_l)} \\ e^{-j3\mu(\theta_l)} \\ \vdots \\ e^{-j(P-3)\mu(\theta_l)} \\ e^{-j(P-2)\mu(\theta_l)} \\ e^{-j(P-1)\mu(\theta_l)} \end{bmatrix}}_{\text{Last } P-1 \text{ elements of the } l\text{-th column of } \mathbf{A}} . \tag{2.38}$$

---

[2] While the ESPRIT algorithm is generally more versatile, this thesis will only utilize the configuration shown in Figure 2.9.

Thus, we can define the $(P-1) \times P$ matrix operators $\mathbf{J}_{\text{first}}$ and $\mathbf{J}_{\text{last}}$ given by

$$\begin{aligned} \mathbf{J}_{\text{first}} &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \\ \mathbf{J}_{\text{last}} &= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{2.39}$$

which, when applied to a vector, take the first and last $P-1$ components of the vector, respectively. With these operators, we can write (2.38) as

$$e^{-j\mu(\theta_l)} \times \mathbf{J}_{\text{first}}\mathbf{a}_l = \mathbf{J}_{\text{last}}\mathbf{a}_l \tag{2.40}$$

for every $l = 0, ..., r-1$. Hence, we have

$$\begin{aligned} &\begin{bmatrix} \mathbf{J}_{\text{first}}\mathbf{a}_0 & \mathbf{J}_{\text{first}}\mathbf{a}_1 & ... & \mathbf{J}_{\text{first}}\mathbf{a}_{r-1} \end{bmatrix} \begin{bmatrix} e^{-j\mu(\theta_0)} & 0 & 0 & 0 \\ 0 & e^{-j\mu(\theta_1)} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & e^{-j\mu(\theta_{r-1})} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{J}_{\text{last}}\mathbf{a}_0 & \mathbf{J}_{\text{last}}\mathbf{a}_1 & ... & \mathbf{J}_{\text{last}}\mathbf{a}_{r-1} \end{bmatrix}, \end{aligned} \tag{2.41}$$

or equivalently,

$$\mathbf{J}_{\text{first}}\mathbf{A}\boldsymbol{\Omega} = \mathbf{J}_{\text{last}}\mathbf{A}, \tag{2.42}$$

where

$$\boldsymbol{\Omega} = \begin{bmatrix} e^{-j\mu(\theta_0)} & 0 & 0 & 0 \\ 0 & e^{-j\mu(\theta_1)} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & e^{-j\mu(\theta_{r-1})} \end{bmatrix}. \tag{2.43}$$

Now, we can utilize the previously derived fact from Section 2.3.1 that the signal eigenvectors reside in $\mathcal{C}(\mathbf{A})$, regardless of the noise. Thus, we can express $\mathbf{V}_s$ as $\mathbf{V}_s = \mathbf{AT}$, and by applying (2.42), we obtain

$$\begin{aligned} &\mathbf{J}_{\text{first}}\mathbf{V}_s\mathbf{T}^{-1}\boldsymbol{\Omega} = \mathbf{J}_{\text{last}}\mathbf{V}_s\mathbf{T}^{-1} \\ \Leftrightarrow\ &\mathbf{J}_{\text{first}}\mathbf{V}_s(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T}) = \mathbf{J}_{\text{last}}\mathbf{V}_s. \end{aligned} \tag{2.44}$$

Then, recalling that $\mathbf{J}_{\text{first}}$ and $\mathbf{J}_{\text{last}}$ are operators that extract the first and last $P-1$ components of a vector (or all the row vectors in the case of a matrix), we find that

$$\mathbf{V}_{s1}(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T}) = \mathbf{V}_{s2}, \tag{2.45}$$

being $\mathbf{V}_{s1}$ the matrix formed after removing the last row of $\mathbf{V}_s$ (i.e., the matrix whose column vectors are the first $P-1$ components of the eigenvectors $\mathbf{v}_s$), and $\mathbf{V}_{s2}$ the matrix formed after removing the first row of $\mathbf{V}_s$ (i.e., the matrix whose column vectors are the last $P-1$ components of the eigenvectors $\mathbf{v}_s$).

Now, observe that $(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T})$ is an $r \times r$ full-rank matrix, while $\mathbf{V}_{s1}$ and $\mathbf{V}_{s2}$ are $(P-1)\times r$, r-rank matrices. Therefore, if $P-1 \geq r$, we can solve (2.45) in a least-square sense by applying the pseudo inverse

$$(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T}) = \mathbf{V}_{s1}^{+}\mathbf{V}_{s2}, \tag{2.46}$$

where $\mathbf{V}_{s1}^{+} = (\mathbf{V}_{s1}^{T}\mathbf{V}_{s1})^{-1}\mathbf{V}_{s1}^{T}$. Summarizing, we have obtained a matrix $(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T})$ by performing successive computations on the sample vectors $\mathbf{x}(t)$. The matrix $\boldsymbol{\Omega}$ is a diagonal matrix that contains the DoAs. However, we have obtained $(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T})$, and we need to extract $\boldsymbol{\Omega}$ somehow from this matrix. The solution is quite simple: on the one hand, since $\boldsymbol{\Omega}$ is a diagonal matrix, its eigenvalues are $\{e^{-j\mu(\theta_l)}\}_{l=0}^{r-1}$ (i.e., the eigenvalues are the DoAs that we are looking for). On the other hand, $\boldsymbol{\Omega}$ and $(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T})$ are similar matrices by definition, meaning they share the same eigenvalues. Consequently, if we compute the eigendecomposition of $(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T})$,

$$(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T}) = \boldsymbol{\Upsilon}\mathbf{L}\boldsymbol{\Upsilon}^{-1}, \tag{2.47}$$

we conclude

$$\mathbf{L} = \begin{bmatrix} \lambda_0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_{r-1} \end{bmatrix} = \begin{bmatrix} e^{-j\mu(\theta_0)} & 0 & 0 & 0 \\ 0 & e^{-j\mu(\theta_1)} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & e^{-j\mu(\theta_{r-1})} \end{bmatrix}. \tag{2.48}$$

The ESPRIT algorithm is summarized as follows:

ESPRIT Algorithm

1) Collect N data samples $\mathbf{X}$ and estimate the covariance matrix $\mathbf{X}\mathbf{X}^H$.
2) Perform the eigendecomposition of $\mathbf{X}\mathbf{X}^H$: $\mathbf{X}\mathbf{X}^H = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$.
3) Estimate the number of sources r by counting the number of eigenvalues close to σ^2, which gives $P - r$, and hence r.
4) Form the $P \times r$ matrix $\mathbf{V}_s$ whose columns are the r eigenvectors corresponding to the largest eigenvalues.
5) Compute $\mathbf{V}_{s1}$ and $\mathbf{V}_{s2}$ by eliminating the last and first rows of $\mathbf{V}_s$, respectively.
6) Solve the equation $\mathbf{V}_{s1}(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T}) = \mathbf{V}_{s2}$ in the least-square sense, obtaining $(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T}) = \mathbf{V}_{s1}^{+}\mathbf{V}_{s2}$.
7) Compute the eigenvalues $\{\lambda_l\}_{l=0}^{r-1}$ of $(\mathbf{T}^{-1}\boldsymbol{\Omega}\mathbf{T})$. The DoAs are then calculated as $\theta_l = -\sin^{-1}\left(\frac{c}{\omega_0 d}\tan^{-1}\left(\frac{\mathbb{I}\{\lambda_l\}}{\mathbb{R}\{\lambda_l\}}\right)\right)$.

Chapter 3

Signal Model and Application of the FFT in the Time Domain

In this chapter, we present the enhancements made in this thesis to the U-ESPRIT 2D algorithm by integrating Polyphase Filter Banks and Fast Fourier Transforms.

3.1 Problem Assumptions and Signal Model

In this section, we develop the mathematical model for the signals an antenna array receives in the presence of multiple sources. Since the literature has covered this development extensively [6, 7, 8, 9, 17, 24], we will emphasize only the critical aspects of our system.

To determine an expression for the signals arriving at the ADCs, let us consider a uniform rectangular array of $P \times Q$ elements (4×4 in the case of RadioVision) and r sources (cell phones in our application), as shown in Figure 3.1. We will consider the origin of the coordinate system at the center of the antenna array and use the indices $p \in \{0, 1, ..., P-1\}$ and $q \in \{0, 1, ..., Q-1\}$ so that the (p, q)-th antenna is the one centered at $\left(\left(-\frac{P-1}{2} + p\right) d_x, \left(-\frac{Q-1}{2} + q\right) d_y\right)$, where d_x and d_y are the antenna spacings in the x and y axes, respectively.

We consider the sources narrowband, non-coherent, and located within an isotropic and non-dispersive medium in the far field. This allows us to consider their radiation as plane waves when impinging on the array. Although our system can choose sub-bands by varying the LO, RadioVision receives only one narrow 70 MHz band at a time. Therefore, we will consider that the r sources share the same carrier frequency ω_0. With all this, we can express the signals $\{s_l\}_{l=0}^{r-1}$ radiated by the sources as $s_l(t) = u_l(t) \cos(\omega_0 t + v_l(t))$, where $u_l(t)$ and $v_l(t)$ are the amplitude and phase modulations of $s_l(t)$, which vary slowly compared to $\cos(\omega_0 t)$ (since $\{s_l\}_{l=0}^{r-1}$ are narrowband signals).

The position of the l-th source is determined by the angles $\phi_{xl} = \cos^{-1}(\hat{\boldsymbol{r}}_l \cdot \hat{\boldsymbol{x}})$ and $\phi_{yl} = \cos^{-1}(\hat{\boldsymbol{r}}_l \cdot \hat{\boldsymbol{y}})$, where $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are the standard unit vectors of the Cartesian coordinate system, and $\hat{\boldsymbol{r}}_l$ is the unit vector pointing towards the l-th source (see Figure 3.1). Note that when the source is at the center, we have $\phi_{xl} = \phi_{yl} = 90°$. By convention, it is best to define

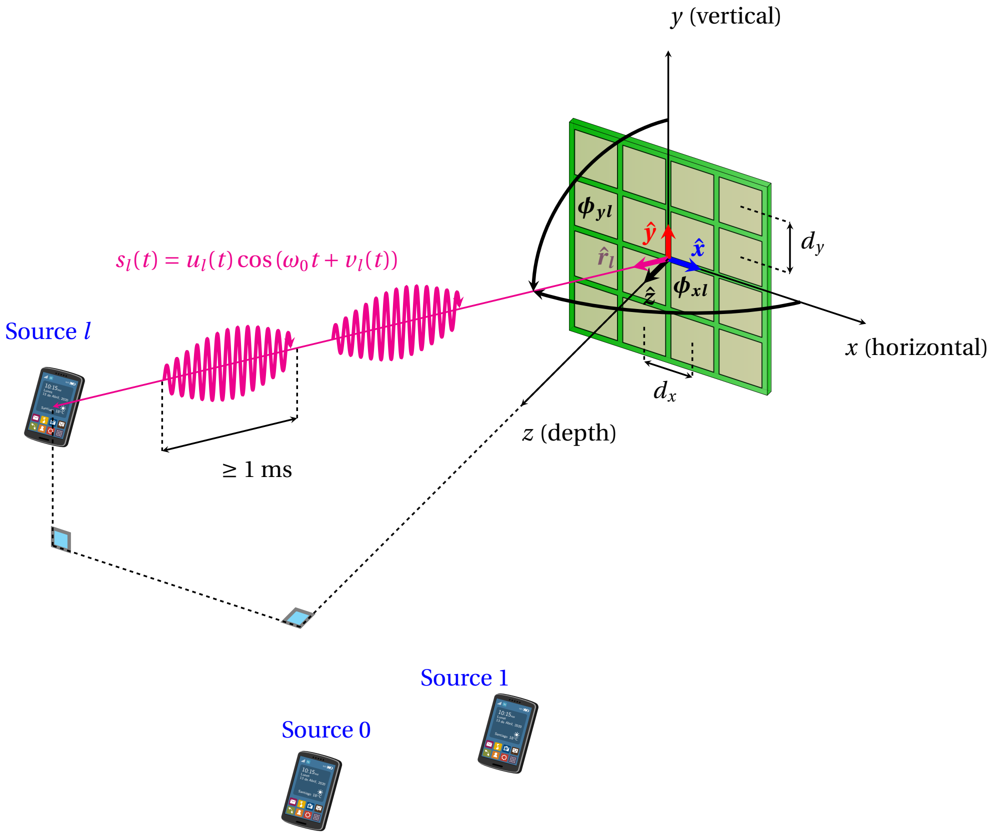


Figure 3.1: Illustration of a 4×4 ($P \times Q$) rectangular array surrounded by three (r) sources (phones). d_x and d_y are the spacings in the x and y axes, respectively. ϕ_{xl} and ϕ_{yl} are the angles of arrival of the l-th source measured relative to the x and y axes, respectively. All sources share the same carrier frequency ω_0. Each source l has its own amplitude and phase modulations $u_l(t)$ and $v_l(t)$, with $l = 0, 1, ..., r-1$. Depending on the mode of operation, phones can radiate packets with durations as short as 1 ms.

$\theta_{xl} = 90^\circ - \phi_{xl}$ and $\theta_{yl} = 90^\circ - \phi_{yl}$, such that $\theta_{xl} = \sin^{-1}(\hat{\boldsymbol{r}}_l \cdot \hat{\boldsymbol{x}})$ and $\theta_{yl} = \sin^{-1}(\hat{\boldsymbol{r}}_l \cdot \hat{\boldsymbol{y}})$. Hence, for a source at the center, $\theta_{xl} = \theta_{yl} = 0^\circ$.

Considering all the definitions above, and following the traditional delayed-propagation model shown in [6, 7, 8, 9, 17, 24], we get that the l-th signal impinging on the (p,q)-th antenna, $x_{pql}(t)$, is given by

$$x_{pql}(t) = s_l(t - \tau_{pq}(\theta_{xl}, \theta_{yl})), \tag{3.1}$$

where

$$\begin{aligned} \tau_{pq}(\theta_{xl}, \theta_{yl}) &= \tau_{pql} = \tau_{pl,x} + \tau_{ql,y} \\ &= \left(\frac{P-1}{2} - p\right)\frac{d_x \sin\theta_{xl}}{c} + \left(\frac{Q-1}{2} - q\right)\frac{d_y \sin\theta_{yl}}{c} \end{aligned} \tag{3.2}$$

is the propagation delay between the (p,q)-th antenna and the array's center, and c is the speed of light. Then, since $u_l(t)$ and $v_l(t)$ vary slowly compared to $\cos(\omega_0 t)$, we can approximate

$$\begin{aligned} x_{pql}(t) &= u_l(t-\tau_{pql})\cos\left(\omega_0(t-\tau_{pql})+v_l(t-\tau_{pql})\right) \\ &\approx u_l(t)\cos\left(\omega_0(t-\tau_{pql})+v_l(t)\right) \\ &= u_l(t)\cos\left(\omega_0 t-\mu_{pl}-\nu_{ql}+v_l(t)\right) \\ &= \frac{1}{2}\left(e^{-j(\mu_{pl}+\nu_{ql})}\xi_l(t)+e^{j(\mu_{pl}+\nu_{ql})}\xi_l^*(t)\right), \end{aligned} \tag{3.3}$$

where we have defined $\mu_{pl}=\omega_0\tau_{pl,x}$, $\nu_{ql}=\omega_0\tau_{ql,y}$, and $\xi_l(t)=u_l(t)e^{j(\omega_0 t+v_l(t))}$.

The $P\times Q$ antennas receive the sum of the r signals centered at ω_0, each one modified by the radiation pattern response at $(\theta_{xl},\theta_{yl})$. Subsequently, all the received signals pass through identical heterodyne receivers, composed of amplifiers, low-pass filters, and mixers with an LO frequency ω_{LO}, for posterior digitization by $P\times Q$ ADCs. If we assume that all receiver chains are equal, the analog signals sensed by the ADCs can be written as

$$z_{pq}(t)=\frac{1}{2}\sum_{l=0}^{r-1}\left(e^{-j(\mu_{pl}+\nu_{ql})}\chi_l(t)+e^{j(\mu_{pl}+\nu_{ql})}\chi_l^*(t)\right), \tag{3.4}$$

where $\{\chi_l(t)\}_{l=0}^{r-1}$ are baseband signals centered at $\omega_0-\omega_{\mathrm{LO}}$, and they comprise the radiation pattern and front-end modifications over the RF signals $x_{pql}(t)$.

Note that in the standard DoA literature, only terms of the form $e^{-j(\mu_{pl}+\nu_{ql})}\chi_l(t)$ are considered. This simplification arises because phasors are used, and the need for both in-phase (I) and quadrature (Q) components is omitted in the mathematical formulation of the algorithms and simulations. However, in the implementation of a real system, both terms $e^{-j(\mu_{pl}+\nu_{ql})}\chi_l(t)$ and $e^{j(\mu_{pl}+\nu_{ql})}\chi_l^*(t)$ appear in the received signals, as shown in (3.4). This discrepancy invalidates the standard procedures followed in the literature [6, 7, 8, 9], which rely on received signals expressed solely in terms of $e^{-j(\mu_{pl}+\nu_{ql})}\chi_l(t)$.

To be able to apply the standard procedures, three methods allow us to obtain only terms of the form $e^{-j(\mu_{pl}+\nu_{ql})}\chi_l(t)$ from the real signals stated in (3.4). The first method is analog and uses I-Q mixers. In this case, more complex mixers are needed, and the post-intermediate frequency (IF) hardware and the number of ADCs are also doubled. The second method is digital and consists of using digital down-converters [25, 26]. For this, DDSs must be implemented to form cosines and sines, in addition to multiple digital filters, which implies an increase in the use of digital resources. As can be seen, both methods require increasing the system's complexity and resource usage. In this sense, the third method, which consists of using the FFT, is optimal since the benefits that come with the use of the FFT are multiple, as mentioned in Chapter 1.

Now that we have the model of the signals that enter the ADCs, we proceed in Section 3.2 with the mathematical development of the incorporation of PFBs and the FFT in the 2D U-ESPRIT algorithm.

3.2 2D U-ESPRIT with PFB-FFT applied to the time domain

Let us take the expression obtained in (3.4) and consider the digitization of the signals by the ADCs, where we will write the discrete-time dependence with (n) instead of (t). Given this, we can express the $P \times Q$ signals entering the digital system as

$$\mathbf{Z}(n) = \frac{1}{2}\sum_{l=0}^{r-1} \boldsymbol{a}_{Pl}\boldsymbol{a}_{Ql}^{T}\chi_l(n) + \frac{1}{2}\sum_{l=0}^{r-1} \boldsymbol{a}_{Pl}^{*}\boldsymbol{a}_{Ql}^{H}\chi_l^{*}(n), \tag{3.5}$$

where $\mathbf{Z}(n)$ is a $P \times Q$ matrix whose (p,q)-th entry is $z_{pq}(n)$, and $\{\boldsymbol{a}_{Pl}\}_{l=0}^{r-1}$ and $\{\boldsymbol{a}_{Ql}\}_{l=0}^{r-1}$ are $P \times 1$ and $Q \times 1$ vectors given by

$$\boldsymbol{a}_{Pl} = \left[e^{-j\mu_{0l}}\ e^{-j\mu_{1l}}\ \ldots\ e^{-j\mu_{(P-1)l}}\right]^{T} \tag{3.6}$$

$$\boldsymbol{a}_{Ql} = \left[e^{-j\nu_{0l}}\ e^{-j\nu_{1l}}\ \ldots\ e^{-j\nu_{(Q-1)l}}\right]^{T}. \tag{3.7}$$

Upon entering the digital system, the received signals are transformed by applying the PFBs and the FFT. To provide a clearer understanding of these processes, we will first address the application of the FFT in Section 3.2.1, followed by an explanation of the integration of PFBs in Section 3.2.2.

3.2.1 Application of the FFT

After computing the N-sized FFT of each $z_{pq}(n)$, we can write

$$\begin{aligned}\hat{\mathbf{Z}}(k) &= \frac{1}{2}\sum_{l=0}^{r-1} \boldsymbol{a}_{Pl}\boldsymbol{a}_{Ql}^{T}\sum_{n=0}^{N-1}\chi_l(n)e^{-j\frac{2\pi nk}{N}} \\ &+ \frac{1}{2}\sum_{l=0}^{r-1} \boldsymbol{a}_{Pl}^{*}\boldsymbol{a}_{Ql}^{H}\sum_{n=0}^{N-1}\chi_l^{*}(n)e^{-j\frac{2\pi nk}{N}},\end{aligned} \tag{3.8}$$

being $\hat{\mathbf{Z}}(k)$ a $P \times Q$ matrix whose (p,q)-th entry is the k-th frequency bin of the FFT of the sequence $\{z_{pq}(n)\}_{n=0}^{N-1}$, that is, $\sum_{n=0}^{N-1} z_{pq}(n)e^{-j\frac{2\pi nk}{N}}$. Then, we introduce the centrosymmetric computations performed in U-ESPRIT [8, 9] to obtain

$$\begin{aligned}
\hat{\mathbf{Z}}'(k) &= \mathbf{Q}_P^H \hat{\mathbf{Z}}(k) \mathbf{Q}_Q^* \\
&= \frac{1}{2} \sum_{l=0}^{r-1} (\mathbf{Q}_P^H \boldsymbol{a}_{Pl})(\mathbf{Q}_Q^H \boldsymbol{a}_{Ql})^T \sum_{n=0}^{N-1} \chi_l(n) e^{-j\frac{2\pi nk}{N}} \\
&+ \frac{1}{2} \sum_{l=0}^{r-1} (\mathbf{Q}_P^H \boldsymbol{a}_{Pl}^*)(\mathbf{Q}_Q^H \boldsymbol{a}_{Ql}^*)^T \sum_{n=0}^{N-1} \chi_l^*(n) e^{-j\frac{2\pi nk}{N}} \\
&= \frac{1}{2} \sum_{l=0}^{r-1} \boldsymbol{\Psi}_l \hat{\chi}_{1l}(k) + \frac{1}{2} \sum_{l=0}^{r-1} \boldsymbol{\eta}_l \hat{\chi}_{2l},
\end{aligned} \tag{3.9}$$

where $\mathbf{Q}_P$ and $\mathbf{Q}_Q$ are defined by the general formulas

$$\mathbf{Q}_{2K} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_K & j\mathbf{I}_K \\ \mathbf{\Pi}_K & -j\mathbf{\Pi}_K \end{bmatrix} \tag{3.10}$$

if P or Q are even numbers, and

$$\mathbf{Q}_{2K+1} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_K & \mathbf{0} & j\mathbf{I}_K \\ \mathbf{0}^T & \sqrt{2} & \mathbf{0}^T \\ \mathbf{\Pi}_K & \mathbf{0} & -j\mathbf{\Pi}_K \end{bmatrix} \tag{3.11}$$

if P or Q are odd numbers, while $\{\hat{\chi}_{1l}(k)\}_{k=0}^{N-1}$ and $\{\hat{\chi}_{2l}(k)\}_{k=0}^{N-1}$ are the FFTs of $\{\chi_l(n)\}_{n=0}^{N-1}$ and $\{\chi_l^*(n)\}_{n=0}^{N-1}$, respectively, and $\boldsymbol{\Psi}_l$ and $\boldsymbol{\eta}_l$ are $P \times Q$ real matrices given by

$$\boldsymbol{\Psi}_l = (\mathbf{Q}_P^H \boldsymbol{a}_{Pl})(\mathbf{Q}_Q^H \boldsymbol{a}_{Ql})^T \tag{3.12}$$

$$\boldsymbol{\eta}_l = (\mathbf{Q}_P^H \boldsymbol{a}_{Pl}^*)(\mathbf{Q}_Q^H \boldsymbol{a}_{Ql}^*)^T. \tag{3.13}$$

If we consider $\chi_l(n) = f_l(n) + jg_l(n)$, we can write (3.9) as

$$\begin{aligned}
\hat{\mathbf{Z}}'(k) &= \sum_{l=0}^{r-1} \frac{\boldsymbol{\Psi}_l + \boldsymbol{\eta}_l}{2} \sum_{n=0}^{N-1} f_l(n) e^{-j\frac{2\pi nk}{N}} \\
&+ j \sum_{l=0}^{r-1} \frac{\boldsymbol{\Psi}_l - \boldsymbol{\eta}_l}{2} \sum_{n=0}^{N-1} g_l(n) e^{-j\frac{2\pi nk}{N}} \\
&= \sum_{l=0}^{r-1} \mathbf{M}_{1l} \hat{f}_l(k) + j \sum_{l=0}^{r-1} \mathbf{M}_{2l} \hat{g}_l(k),
\end{aligned} \tag{3.14}$$

where $\mathbf{M}_{1l} = (\boldsymbol{\Psi}_l + \boldsymbol{\eta}_l)/2$, $\mathbf{M}_{2l} = (\boldsymbol{\Psi}_l - \boldsymbol{\eta}_l)/2$, and $\{\hat{f}_l(k)\}_{k=0}^{N-1}$ and $\{\hat{g}_l(k)\}_{k=0}^{N-1}$ are the FFTs of $\{f_l(n)\}_{n=0}^{N-1}$ and $\{g_l(n)\}_{n=0}^{N-1}$, respectively. Then, to align with the standard ESPRIT-like mathematical developments, we must write each matrix $\hat{\mathbf{Z}}'(k)$ in vector form, which is accomplished with the $vec(\cdot)$ operation: $\hat{\mathfrak{z}}(k) = vec(\hat{\mathbf{Z}}'(k))$. In this way, the (p,q)-th entry of $\hat{\mathbf{Z}}'(k)$ is mapped into the $(P(q-1)+p)$-th position of the $PQ \times 1$ complex vector $\hat{\mathfrak{z}}(k)$. After applying the $vec(\cdot)$ operation to the expression of $\hat{\mathbf{Z}}'(k)$ given in (3.14), we get

$$\begin{aligned}
\hat{\mathfrak{z}}(k) &= \sum_{l=0}^{r-1} vec(\mathbf{M}_{1l})\hat{f}_l(k) + j\sum_{l=0}^{r-1} vec(\mathbf{M}_{2l})\hat{g}_l(k) \\
&= \left[vec(\mathbf{M}_{10})\ vec(\mathbf{M}_{11})\ ...\ vec(\mathbf{M}_{1(r-1)})\right] \hat{\boldsymbol{f}}(k) \\
&\quad + j\left[vec(\mathbf{M}_{20})\ vec(\mathbf{M}_{21})\ ...\ vec(\mathbf{M}_{2(r-1)})\right] \hat{\boldsymbol{g}}(k) \\
&= \mathbf{S}_{m1}\hat{\boldsymbol{f}}(k) + j\mathbf{S}_{m2}\hat{\boldsymbol{g}}(k),
\end{aligned} \tag{3.15}$$

where $\mathbf{S}_{m1} = \left[vec(\mathbf{M}_{10})\ vec(\mathbf{M}_{11})\ ...\ vec(\mathbf{M}_{1(r-1)})\right]$ and $\mathbf{S}_{m2} = \left[vec(\mathbf{M}_{20})\ vec(\mathbf{M}_{21})\ ...\ vec(\mathbf{M}_{2(r-1)})\right]$ are $PQ \times r$ real matrices, and $\hat{\boldsymbol{f}}(k) = \left[\hat{f}_0(k)\ ...\ \hat{f}_{r-1}(k)\right]^T$ and $\hat{\boldsymbol{g}}(k) = \left[\hat{g}_0(k)\ ...\ \hat{g}_{r-1}(k)\right]^T$ are $r \times 1$ real vectors corresponding to the k-th frequency bin of the FFTs of the vector sequences $\boldsymbol{f}(n) = [f_0(n)\ ...\ f_{r-1}(n)]^T$ and $\boldsymbol{g}(n) = [g_0(n)\ ...\ g_{r-1}(n)]^T$. Then, we can stack all vectors $\hat{\mathfrak{z}}(k)$ from $k = 0$ to $k = N/2 - 1$ (which corresponds to taking half the FFT spectrum) to form the matrix

$$\begin{aligned}
\hat{\mathfrak{z}} &= \left[\hat{\mathfrak{z}}(0)\ ...\ \hat{\mathfrak{z}}(N/2-1)\right] \\
&= \mathbf{S}_{m1}\left[\hat{\boldsymbol{f}}(0)\ ...\ \hat{\boldsymbol{f}}(N/2-1)\right] + \mathbf{S}_{m2}\left[\hat{\boldsymbol{g}}(0)\ ...\ \hat{\boldsymbol{g}}(N/2-1)\right] \\
&= \mathbf{S}_{m1}\mathbf{F}_N\mathbf{W}_H + j\mathbf{S}_{m2}\mathbf{G}_N\mathbf{W}_H,
\end{aligned} \tag{3.16}$$

being $\mathbf{F}_N = [\boldsymbol{f}(0)\ ...\ \boldsymbol{f}(N-1)]$, $\mathbf{G}_N = [\boldsymbol{g}(0)\ ...\ \boldsymbol{g}(N-1)]$ and $\mathbf{W}_H$ the $N \times N/2$ FFT matrix obtained after removing half the spectrum, given by

$$\mathbf{W}_H = \begin{bmatrix} 1 & 1 & ... & 1 \\ 1 & W_N & ... & W_N^{N/2-1} \\ \vdots & \vdots & ... & \vdots \\ 1 & W_N^{N-1} & ... & W_N^{(N/2-1)(N-1)} \end{bmatrix}, \tag{3.17}$$

where $W_N = e^{-j\frac{2\pi}{N}}$. The rationale for considering only half of the FFT spectrum lies in the fact that, for real-valued signals, the real part of the FFT is even, while the imaginary part is odd. Consequently, the phase contributions from each half of the spectrum, derived from the imaginary to the real part ratio, are opposite in sign. This would result in erroneous DoA estimates if both halves were considered.

After aligning with the standard ESPRIT-like formulations, we can compute the correlation matrix as

$$\begin{aligned}
\mathbf{R}_{zz} &= \hat{\mathfrak{z}}\hat{\mathfrak{z}}^H \\
&= (\mathbf{S}_{m1}\mathbf{F}_N\mathbf{W}_H + j\mathbf{S}_{m2}\mathbf{G}_N\mathbf{W}_H) \\
&\times (\mathbf{W}_H^H\mathbf{F}_N^T\mathbf{S}_{m1}^T - j\mathbf{W}_H^H\mathbf{G}_N^T\mathbf{S}_{m2}^T) \\
&= \mathbf{S}_{m1}\mathbf{F}_N\mathbf{S}_w\mathbf{F}_N^T\mathbf{S}_{m1}^T + \mathbf{S}_{m2}\mathbf{G}_N\mathbf{S}_w\mathbf{G}_N^T\mathbf{S}_{m2}^T \\
&+ j(\mathbf{S}_{m2}\mathbf{G}_N\mathbf{S}_w\mathbf{F}_N^T\mathbf{S}_{m1}^T - \mathbf{S}_{m1}\mathbf{F}_N\mathbf{S}_w\mathbf{G}_N^T\mathbf{S}_{m2}^T),
\end{aligned} \tag{3.18}$$

whose real part is given by

$$\begin{aligned}
\mathbb{R}\{\mathbf{R}_{zz}\} &= \frac{1}{2}(\mathbf{R}_{zz} + \mathbf{R}_{zz}^*) \\
&= \mathbf{S}_{m1}\mathbf{F}_N\mathbb{R}\{\mathbf{S}_w\}\mathbf{F}_N^T\mathbf{S}_{m1}^T + \mathbf{S}_{m1}\mathbf{F}_N\mathbb{I}\{\mathbf{S}_w\}\mathbf{G}_N^T\mathbf{S}_{m2}^T \\
&+ \mathbf{S}_{m2}\mathbf{G}_N\mathbb{R}\{\mathbf{S}_w\}\mathbf{G}_N^T\mathbf{S}_{m2}^T - \mathbf{S}_{m2}\mathbf{G}_N\mathbb{I}\{\mathbf{S}_w\}\mathbf{F}_N^T\mathbf{S}_{m1}^T \\
&= \mathbf{S}_{m1}\mathbf{F}_N(\mathbf{S}_{wr}\mathbf{F}_N^T\mathbf{S}_{m1}^T + \mathbf{S}_{wi}\mathbf{G}_N^T\mathbf{S}_{m2}^T) \\
&+ \mathbf{S}_{m2}\mathbf{G}_N(\mathbf{S}_{wr}\mathbf{G}_N^T\mathbf{S}_{m2}^T - \mathbf{S}_{wi}\mathbf{F}_N^T\mathbf{S}_{m1}^T) \\
&= \begin{bmatrix} \mathbf{S}_{m1}\mathbf{F}_N & \mathbf{S}_{m2}\mathbf{G}_N \end{bmatrix} \begin{bmatrix} \mathbf{S}_{wr} & \mathbf{S}_{wi} \\ -\mathbf{S}_{wi} & \mathbf{S}_{wr} \end{bmatrix} \begin{bmatrix} \mathbf{F}_N^T\mathbf{S}_{m1}^T \\ \mathbf{G}_N^T\mathbf{S}_{m2}^T \end{bmatrix} \\
&= \mathbf{S}_m\mathbf{\Gamma}\mathbf{S}_m^T,
\end{aligned} \tag{3.19}$$

where we have defined $\mathbf{S}_m = [\mathbf{S}_{m1}\mathbf{F}_N \; \mathbf{S}_{m2}\mathbf{G}_N]$, $\mathbf{S}_w = \mathbf{W}_H\mathbf{W}_H^H$, $\mathbf{S}_{wr} = \mathbb{R}\{\mathbf{S}_w\}$, $\mathbf{S}_{wi} = \mathbb{I}\{\mathbf{S}_w\}$, and

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{S}_{wr} & \mathbf{S}_{wi} \\ -\mathbf{S}_{wi} & \mathbf{S}_{wr} \end{bmatrix}. \tag{3.20}$$

Note that (3.19) is written as a quadratic form, and since $\mathbf{S}_{wr}^T = \mathbf{S}_{wr}$ and $\mathbf{S}_{wi}^T = -\mathbf{S}_{wi}$, $\mathbf{\Gamma}$ is symmetric.

Now, since the signals are not correlated, $\mathbf{F}_N$ and $\mathbf{G}_N$ have full rank r. Therefore, the column space of $\mathbf{S}_{m1}\mathbf{F}_N$ equals the column space of $\mathbf{S}_{m1}$, and the column space of $\mathbf{S}_{m2}\mathbf{G}_N$ equals the column space of $\mathbf{S}_{m2}$ (see Theorem A.4). Hence, the column space of $[\mathbf{S}_{m1}\mathbf{F}_N \; \mathbf{S}_{m2}\mathbf{G}_N]$ includes the column space of $\mathbf{S}_{m1}+\mathbf{S}_{m2}$ (see Theorem A.3). In consequence, there are r eigenvectors of $\mathbb{R}\{\mathbf{R}_{zz}\}$ that span the column space of $\mathbf{S}_{m1}+\mathbf{S}_{m2}$. Now, notice that $\mathbf{S}_{m1}+\mathbf{S}_{m2}$ is just

$\mathbf{\Psi} = [vec(\mathbf{\Psi}_0) \ ... \ vec(\mathbf{\Psi}_{r-1})]$. Hence, if we concatenate those aforementioned r eigenvectors as the columns of a matrix $\mathbf{V}_r$, we can write

$$\mathbf{\Psi} = \mathbf{V}_r \mathbf{T}, \tag{3.21}$$

where $\mathbf{T}$ is a non-singular $r \times r$ real matrix. From now on, we can follow the standard ESPRIT-like procedure documented in [9]. Specifically, we must consider that each $vec(\mathbf{\Psi}_l)$ follows the invariance relationships [9]

$$\begin{aligned} \tan\left(\frac{\mu_l}{2}\right) \mathbf{K}_{\mu 1} \ vec(\mathbf{\Psi}_l) &= \mathbf{K}_{\mu 2} \ vec(\mathbf{\Psi}_l) \\ \tan\left(\frac{\nu_l}{2}\right) \mathbf{K}_{\nu 1} \ vec(\mathbf{\Psi}_l) &= \mathbf{K}_{\nu 2} \ vec(\mathbf{\Psi}_l), \end{aligned} \tag{3.22}$$

where $\mu_l = (\omega_0 d_x \sin\theta_{xl})/c$, $\nu_l = (\omega_0 d_y \sin\theta_{yl})/c$, $\mathbf{K}_{\mu 1}$ and $\mathbf{K}_{\mu 2}$ are $(P-1)Q \times PQ$ matrices given by

$$\mathbf{K}_{\mu 1} = \mathbf{I}_Q \otimes \mathbb{R}\{\mathbf{Q}_{P-1}^H \mathbf{J}_P \mathbf{Q}_P\} \tag{3.23}$$

$$\mathbf{K}_{\mu 2} = \mathbf{I}_Q \otimes \mathbb{I}\{\mathbf{Q}_{P-1}^H \mathbf{J}_P \mathbf{Q}_P\}, \tag{3.24}$$

and $\mathbf{K}_{\nu 1}$ and $\mathbf{K}_{\nu 2}$ are $(Q-1)P \times PQ$ matrices given by

$$\mathbf{K}_{\nu 1} = \mathbb{R}\{\mathbf{Q}_{Q-1}^H \mathbf{J}_Q \mathbf{Q}_Q\} \otimes \mathbf{I}_P \tag{3.25}$$

$$\mathbf{K}_{\nu 2} = \mathbb{I}\{\mathbf{Q}_{Q-1}^H \mathbf{J}_Q \mathbf{Q}_Q\} \otimes \mathbf{I}_P, \tag{3.26}$$

where $\mathbf{J}_P$ and $\mathbf{J}_Q$ are $(P-1) \times P$ and $(Q-1) \times Q$ selection matrices given by

$$\mathbf{J}_P = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{P-1} \end{bmatrix}, \ \mathbf{J}_Q = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{Q-1} \end{bmatrix}. \tag{3.27}$$

After concatenating the r invariance relationships (3.22) (from $l = 0$ to $l = r-1$) into a matrix form, we obtain

$$\mathbf{K}_{\mu 1} \mathbf{\Psi} \mathbf{\Omega}_\mu = \mathbf{K}_{\mu 2} \mathbf{\Psi} \tag{3.28}$$

$$\mathbf{K}_{\nu 1} \mathbf{\Psi} \mathbf{\Omega}_\nu = \mathbf{K}_{\nu 2} \mathbf{\Psi}, \tag{3.29}$$

and after using (3.21), we get

$$\begin{aligned} \mathbf{K}_{\mu 1} \mathbf{V}_r (\mathbf{T} \mathbf{\Omega}_\mu \mathbf{T}^{-1}) &= \mathbf{K}_{\mu 2} \mathbf{V}_r \\ \mathbf{K}_{\nu 1} \mathbf{V}_r (\mathbf{T} \mathbf{\Omega}_\nu \mathbf{T}^{-1}) &= \mathbf{K}_{\nu 2} \mathbf{V}_r, \end{aligned} \tag{3.30}$$

where $\boldsymbol{\Omega}_\mu$ and $\boldsymbol{\Omega}_\nu$ are diagonal matrices such that $(\boldsymbol{\Omega}_\mu)_{ll} = \tan\frac{\mu_l}{2}$ and $(\boldsymbol{\Omega}_\nu)_{ll} = \tan\frac{\nu_l}{2}$. Finally, we can solve the two systems of equations (3.30) to obtain $(\mathbf{T}\boldsymbol{\Omega}_\mu\mathbf{T}^{-1})$ and $(\mathbf{T}\boldsymbol{\Omega}_\nu\mathbf{T}^{-1})$, which are similar matrices to $\boldsymbol{\Omega}_\mu$ and $\boldsymbol{\Omega}_\nu$, and therefore share the eigenvalues, which in turn contain the DoAs via the expressions $\tan\frac{\mu_l}{2}$ and $\tan\frac{\nu_l}{2}$.

The practical application of the development just described is summarized below.

1) Calculate the N-sized FFT of the $P \times Q$ data streams received by the antenna array and eliminate half of the spectrum.
2) Perform the centrosymmetric operations stated in (3.9), via the matrices $\mathbf{Q}_P$ and $\mathbf{Q}_Q$.
3) Calculate the real part of the correlation matrix $\mathbf{R}_{zz}$ of the antenna array using each frequency bin as a sample.
4) Considering r sources (assumed as a known number in this work), compute the eigendecomposition of $\mathbb{R}\{\mathbf{R}_{zz}\}$ and select the r eigenvectors associated with the r greatest eigenvalues. Concatenate those r eigenvectors into a matrix $\mathbf{V}_r$.
5) Solve the linear equations $\mathbf{K}_{\mu 1}\mathbf{V}_r\mathbf{X} = \mathbf{K}_{\mu 2}\boldsymbol{\Psi}$ and $\mathbf{K}_{\nu 1}\mathbf{V}_r\mathbf{Y} = \mathbf{K}_{\nu 2}\boldsymbol{\Psi}$ to find $\mathbf{X}$ and $\mathbf{Y}$.
6) Form the complex matrix $\mathbf{X} + j\mathbf{Y}$ and find its complex eigenvalues $\{\lambda_l\}_{l=0}^{r-1}$.
7) Obtain the estimated DoAs $(\theta_{xl}, \theta_{yl})$ for the l-th signal as:

$$\theta_{xl} = \sin^{-1}\left(\frac{2c}{\omega_0 d_x}\tan^{-1}(\mathbb{R}\{\lambda_l\})\right) \tag{3.31}$$

$$\theta_{yl} = \sin^{-1}\left(\frac{2c}{\omega_0 d_y}\tan^{-1}(\mathbb{I}\{\lambda_l\})\right) \tag{3.32}$$

3.2.2 Application of the PFBs

For the application of the PFB, the PFB acts as a windowing function implemented efficiently before the FFT. If we consider a mathematical development analogous to the one followed in (3.8), we can write the signals transformed by the PFBs and then by the FFT as

$$\begin{aligned}\hat{\mathbf{Z}}(k) = \frac{1}{2}\sum_{l=0}^{r-1}\boldsymbol{a}_{Pl}\boldsymbol{a}_{Ql}^T\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}h(n+mN)\chi_l(n+mN)e^{-j\frac{2\pi nk}{N}}\\ +\frac{1}{2}\sum_{l=0}^{r-1}\boldsymbol{a}_{Pl}^*\boldsymbol{a}_{Ql}^H\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}h(n+mN)\chi_l^*(n+mN)e^{-j\frac{2\pi nk}{N}},\end{aligned} \tag{3.33}$$

where M is the tap number of the polyphase structure and $h(n)$ represents the impulse response of a low-pass filter with MN coefficients [27, 28, 29, 30, 31]. Then, performing the same operations described in Section 3.2.1, we get an expression analogous to (3.16), given by

$$\hat{\mathfrak{z}} = \mathbf{S}_{m1}\mathbf{F}_N'\boldsymbol{\mathcal{H}}\mathbf{W}_H + j\mathbf{S}_{m2}\mathbf{G}_N'\boldsymbol{\mathcal{H}}\mathbf{W}_H, \tag{3.34}$$

where $\mathbf{F}'_N$ and $\mathbf{G}'_N$ are matrices analogous to $\mathbf{F}_N$ and $\mathbf{G}_N$, but with the difference that, in this case, they contain NM samples and are written as $\mathbf{F}'_N = [\boldsymbol{f}(0) \ ... \ \boldsymbol{f}(N) \ \boldsymbol{f}(N+1) \ ... \ \boldsymbol{f}(NM)]$, $\mathbf{G}'_N = [\boldsymbol{g}(0) \ ... \ \boldsymbol{g}(N) \ \boldsymbol{g}(N+1) \ ... \ \boldsymbol{g}(NM)]$ (that is to say, they are $r \times NM$ matrices), and $\boldsymbol{\mathcal{H}}$ is a $NM \times N$ real matrix given by

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_{M-1} \end{bmatrix}, \tag{3.35}$$

where each $\{\mathbf{H}_m\}_{m=0}^{M-1}$ is a diagonal $N \times N$ real matrix with $(\mathbf{H}_m)_{nn} = h(mN + n)$, with $n = 0, 1, ..., N-1$.

Given (3.34), we notice that all the results from Section 3.2.1 are valid for the PFB implementation, with the difference that now $\mathbf{\Gamma}$ is given by

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbb{R}\{\boldsymbol{\mathcal{H}}\mathbf{W}_H\mathbf{W}_H^H\boldsymbol{\mathcal{H}}^H\} & \mathbb{I}\{\boldsymbol{\mathcal{H}}\mathbf{W}_H\mathbf{W}_H^H\boldsymbol{\mathcal{H}}^H\} \\ -\mathbb{I}\{\boldsymbol{\mathcal{H}}\mathbf{W}_H\mathbf{W}_H^H\boldsymbol{\mathcal{H}}^H\} & \mathbb{R}\{\boldsymbol{\mathcal{H}}\mathbf{W}_H\mathbf{W}_H^H\boldsymbol{\mathcal{H}}^H\} \end{bmatrix}. \tag{3.36}$$

Finally, note that from (3.34), the PFB-FFT implementation can be interpreted as a complex digital mixing (multiplication by $\cos\left(\frac{2\pi nk}{N}\right) + j\sin\left(\frac{2\pi nk}{N}\right)$) followed by a low-pass filter $h(n)$, and hence, it constitutes and efficient way for implementing a digital downconverter, with the additional benefits stated in Section 1.3.

Chapter 4

Hardware Design

In this chapter, we present the design and implementation of RadioVision. The system comprises an analog and digital part, as shown in Figure 4.1. Each of these parts are described in the following subsections.

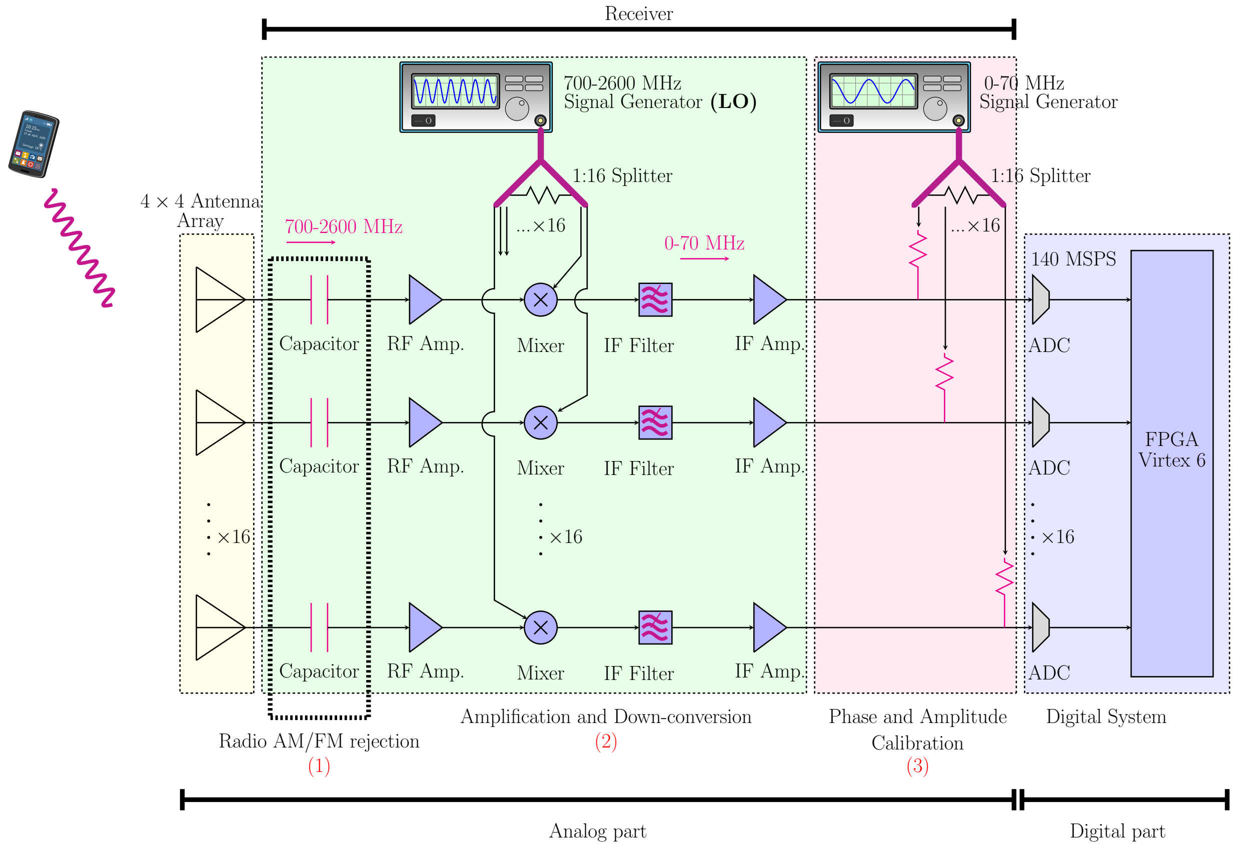


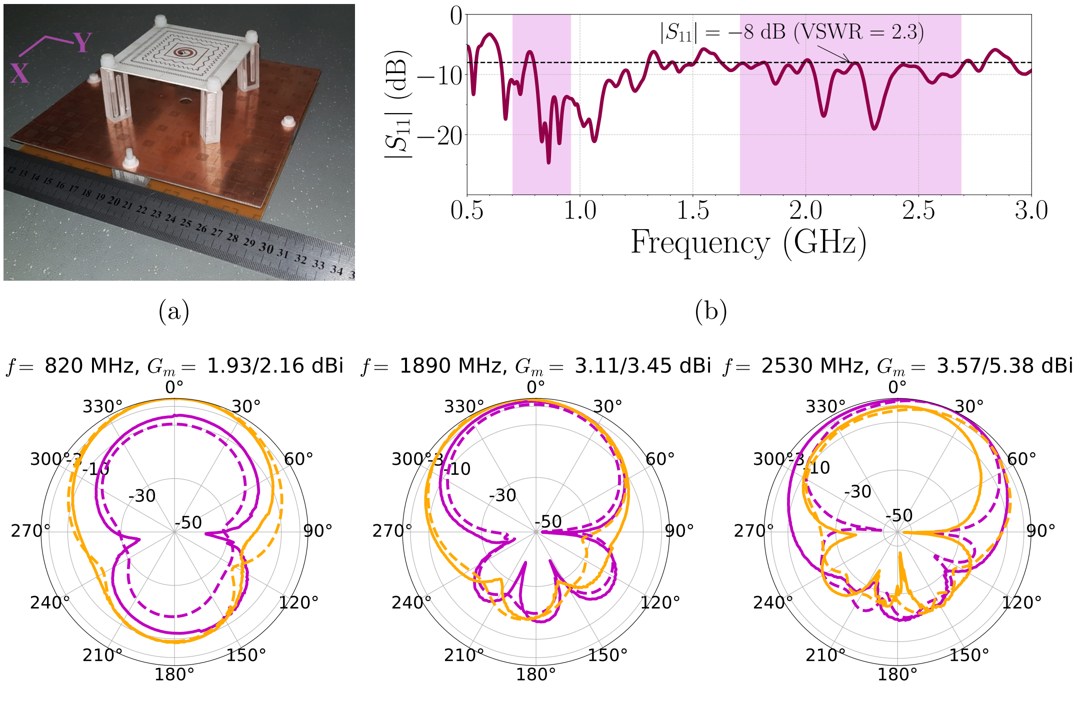
Figure 4.1: Schematic representation of RadioVision hardware.

4.1 Analog system

The analog system is composed of two parts: the antenna array and the receiver. In Sections 4.1.1 and 4.1.2, we review these two components separately, followed by the integration of both in Section 4.1.3.

4.1.1 Antenna Array

Since our system is intended for phone localization, an antenna array that works over the entire mobile phone spectrum is required. To accomplish this, we selected the antenna presented in [32], which has several advantages. First, it covers all the up-links from the UMTS and LTE bands tendered in Chile (700–2600 MHz). Second, it features a wide average half-power beamwidth of 72°, which defines the field of view of the localization system. Third, its built-in reflector ensures a half-space radiation pattern, which is necessary to avoid phase ambiguities between the front and back of the array, and it also prevents interference and coupling with the reception system. Finally, its small size (7.5 cm) allows operation without phase ambiguities (or grating lobes) across the entire bandwidth within the field of view. The key performance metrics of the antenna are summarized in Figure 4.2.



(a)

(b)

(c)

(d)

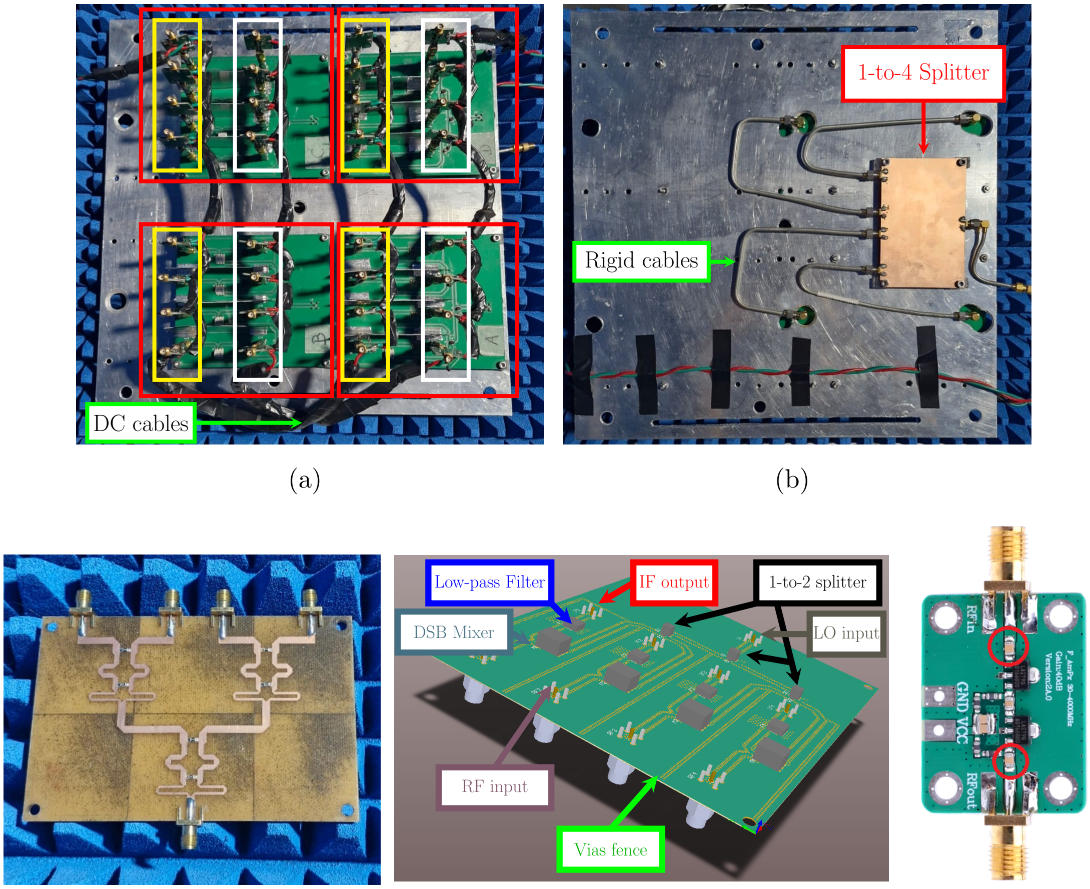
(e)

Figure 4.2: Antenna presented in [32] and its leading figures of merit. (a) Fabricated antenna. (b) Measured S_{11} parameter. The purple colored bands show the frequencies in which the antenna must operate to cover all the LTE and UMTS uplinks. (c, d, e) Measured (solid lines) and simulated (dashed lines) radiation patterns of the single-element antenna in dB at 820, 1890, and 2530 MHz. Purple lines correspond to the X-pol, while orange lines correspond to the Y-pol. Measurements are normalized with respect to their maximum (left value of G_m), and simulations are normalized with respect to their maximum (right value of G_m).

4.1.2 Receiver

The receiver can be decomposed into three parts, enumerated as (1)–(3) in Figure 4.1. The first part (1) is a series capacitor with a cut-off frequency of 100 MHz, which acts as a high-pass filter and prevents AM and FM radio from entering the system and eventually reaching the IF (0–70 MHz). The second part (2) has the typical elements expected on a heterodyne receiver, which are RF and IF amplifiers, a DSB-mixer to obtain the baseband signal (up to 70 MHz), and a low-pass filter to prevent aliasing in the ADCs. It also has a 1-to-16 splitter to distribute the LO signal. In the third part (3), a 1-to-16 splitter distributes a signal that sweeps from 0 to 70 MHz. This signal is used to calibrate phase imbalances digitally, as explained in Section 4.2.1.2.

Although Figure 4.1 presents the theoretical design of the receiver, several practical considerations shaped its implementation. We opted for a modular approach when designing the 16-chain receiver, dividing it into multiple PCBs to house different components, as shown in Figures 4.3(a) and (b). The mixers and low-pass filters are spread across four identical PCBs known as downconverter boards. The RF amplifiers and series capacitors are placed together



(a) (b)

(c) (d) (e)

Figure 4.3: RadioVision's 16-chain receiver. (a) Top layer of the receiver, housing the 4 downconverter boards (red), 16 RF amplifiers (yellow), 16 IF amplifiers (white) and all DC connections to bias the amplifiers. (b) Bottom layer of the receiver, housing the 1-to-4 LO splitter, which is facing the metal plate to reduce coupling. (c) 1-to-4 LO splitter. (d) 3D layout of the downconverter board. (e) Commercially available amplifiers.

on 16 PCBs, and the IF amplifiers are also housed on separate PCBs. Additionally, a 1-to-4 splitter, shown in Figure 4.3(c), is used to distribute the LO signal to the four downconvert boards, where the LO is further divided into the four channels of each downconverter board by the use of successive 1-to-2 SMD splitters, as shown in Figure 4.3(d).

We placed the mixers and filters on four separate boards for economic reasons, as the unit price decreases when external companies commission larger quantities of identical PCBs. The 32 amplifiers were placed on individual boards for better thermal management. This decision was crucial because the receiver consumes 24 W of power, and consolidating all components on a single PCB could lead to overheating and potential damage to the integrated circuits.

The 3D layout of the four downconverter boards can be seen in Figure 4.3(d), while Figure 4.3(e) illustrates the commercially available PCB used for both the RF and IF amplification stages. The same amplifier design was selected for both stages because it is wideband, operating effectively from 10 MHz to 4 GHz. Each amplifier board includes its own DC-block capacitors, highlighted in red in the figure. In the 16 boards dedicated to RF amplification, we replaced the DC-block capacitors with 1 nF capacitors to filter out signals below 100 MHz, such as AM and FM radio, preventing them from reaching the IF stage. The 16 boards designated for IF amplification remained unchanged.

Before continuing with the measurements of the complete receiver, we list some fundamental considerations.

- ✠ Since each amplifier provides a gain of 30–40 dB across its entire bandwidth, the complete receiving chains exhibit very high overall gain. This high gain implies we must prevent signal couplings, standing waves, and feedback loops.
- ✠ We opted for two separate DC inputs for the power supply: one for the 16 RF amplifiers and another for the 16 IF amplifiers. This design choice is based on the fact that each set of 16 amplifiers requires approximately 3 A of current, which is easily manageable by the DC sources available in our laboratory. Using a single input for all 32 amplifiers could introduce feedback loops within the IF signals and would demand around 6 A, which exceeds the capacity of some of our DC sources. While using more than two outputs is technically feasible, it would necessitate additional power sources, which we deemed undesirable for simplicity and manageability.
- ✠ Ferrites and absorbent material were placed on the DC cables to mitigate any coupling of high-frequency signals.
- ✠ Given the high gain of the system and the significant difference between the LO power and the received signals (with LO power injected into the first splitter being around 20 dBm, while signals from the antennas are below -40 dBm), it was crucial to prevent LO signal coupling into the RF tracks, which could otherwise saturate the amplifiers. To address this problem, we attached the first LO splitter to the metal plate holding the receiver, as shown in Figure 4.3(b), effectively creating a shielding effect similar to a Faraday cage. In addition, rigid cables were used for the LO connections (Figure 4.3(b)), as they offer better insulation compared to flexible cables.
- ✠ The cables entering the RF amplifiers from the antennas and those exiting the IF amplifiers to the ADCs were also chosen to be rigid for two main reasons: (1) Rigid cables provide better insulation compared to flexible alternatives, which helps to reduce

coupling, and (2) rigid cables are more stable against mechanical movements, and considering the high gain of the system, good stability in general is sought.

- ✠ As a precautionary measure, vias fences were installed on the downconverter boards to separate each receiving chain and reduce potential interference. In addition, we added small aluminum plates to enhance the physical separation further and minimize signal coupling between adjacent channels. Subsequent coupling measurements showed no appreciable difference between applying and not applying the aluminum plates. However, as it required work to remove them, and they made no difference, we decided to leave them fitted in the system.
- ✠ The order of the components in the receiver is relevant. Placing the anti-aliasing filter before the IF amplifier helps eliminate leakage into the IF and reduces intermodulation products, preventing them from saturating the IF amplifier.

The conversion gains of the 16 receiver chains after the IF amplifiers (that is to say, without considering the (3) part of Figure 4.1) are shown in Figure 4.4 for all the LO configurations needed to cover the up-links listed in Table 1.1. Note that the placement of the LO frequency is important for the specific task of locating phones. Indeed, in the context of DSB mixers, where no distinction is made between the upper and lower sidebands (USB and LSB), the location of phones (which operate in the UL) can be significantly influenced by base stations (which operate in the DL) if both UL and DL are overlapped in the IF. For example, as can be seen in Figure 4.4, if the LO is placed at 2.5 GHz, the USB is in the desired UL of 2.505–2.565 GHz, but the LSB falls in the 2.4 GHz WiFi Band. As both bands are mixed in a DSB mixer, it will not be possible to differentiate which source comes from which band, and therefore, the DoAs cannot be estimated correctly. If the LO were placed at 2.57 GHz to avoid WiFi, we would still have the same problem since now the desired UL would be in the LSB, but the DL would be in the USB. This example demonstrates the importance of being able to flag channels to avoid interference (2.4 GHz WiFi Band or base station DL in the case of 2.5 GHz).

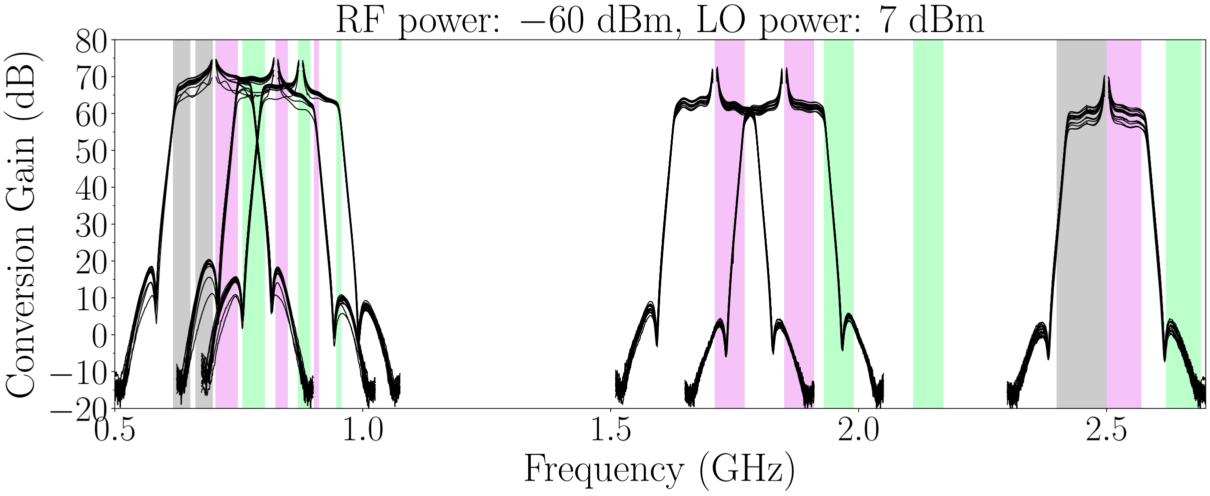


Figure 4.4: Conversion gain of the 16 receiver chains. The purple bands correspond to the up-links, the green bands to the down-links, and the gray bands to the digital dividend (≈ 600–700 MHz) and WiFi (≈ 2400–2500 MHz).

Table 4.1: LO positioning for each band, the frequencies to flag, and the operation that specifies whether the uplink falls in the upper or lower sideband after downconversion.

Band	**UL (MHz)**	**DL (MHz)**	LO frequency (MHz)	Operation	Bands to flag (MHz)
28 APT (LTE)	703–748	758–803	690	USB	690–703, 748–760
5 CLR (UMTS)	824–849	869–894	820	USB	820–825, 837–890
8 E-GSM (UMTS)	902–912	947–957	922	LSB	922–912, 900–852
10 AWS (LTE)	1710–1770	2110–2170	1700	USB	1700–1710, 1768–1770
2 PCS (LTE)	1850-1910	1930–1990	1840	USB	1840–1850, 1908–1910
7 IMT-E (LTE)	2505–2565	2625–2685	2570	LSB	2570–2565, 2520–2500

From Figure 4.4, it becomes clear which LO frequencies and channels should be selected to maximize the uplink bandwidth coverage while minimizing contamination from downlink signals. The selected LO frequencies and channels to flag for each uplink are presented in Table 4.1. In the 8 E-GSM and 7 IMT-E bands, it is more effective to position the LO at the higher end of the uplink range. This results in the uplink signal falling into the LSB after downconversion, and as a consequence, the DoA estimation produces a sign reversal (Appendix D). Another important aspect is that in nearly all cases, the LO is positioned 10 MHz below the start of the band. This strategy compensates for the non-flat response of the receiver at these frequencies.

4.1.3 Integration of the analog part

Finally, we show the integration of the analog part in Figure 4.5. We assembled the 4×4 antenna array and the reflector at the front of a $50 \times 50 \times 50$ cm^3 plastic box with the receiver inside. We lined the inside of the plastic box with aluminum foil, creating a Faraday cage to prevent external RF signals from coupling into the receiver. We also placed absorbing material at strategic points within the Faraday cage -obtained experimentally through trial and error- to prevent the formation of standing waves and feedback loops inside the box.

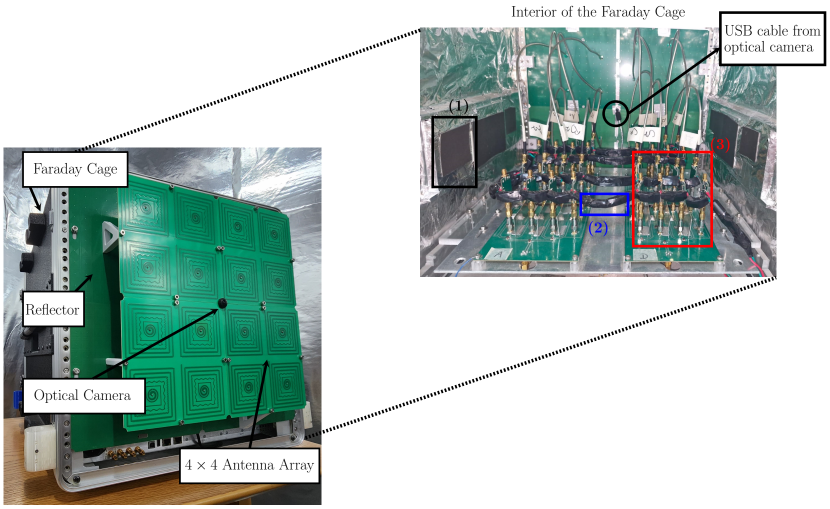


Figure 4.5: Assembly of the analog part of RadioVision. The receiver is placed inside a Faraday Cage to prevent external RF signals from coupling. (1) Absorbent material to prevent reflections and feedback loops inside the box. (2) Absorbent material in the DC cables. (3) RF and IF amplifiers.

4.2 Digital system

The digital system is implemented in the ROACH-2 platform [1] (Reconfigurable Open Architecture Computing Hardware) developed by the CASPER group (Collaboration for Astronomy Signal Processing and Electronics Research), shown in Figure 4.6. The board's key components include the Xilinx Virtex-6 XC6VSX475T FPGA, 4 HMCAD1511 integrated circuits housing 16 ADCs with a sampling rate of 140 MSPS, and a microcomputer PowerPC 440EPx to provide control functions. The 4 HMCAD1511 integrated circuits are integrated into an adc16x250-8 rev. 2 board developed by CASPER. One problem with this card is the start ambiguity, which causes the data delivered by two chips to not necessarily be synchronized since it cannot be ensured that both start at the same clock edge. This results in the ADCs being out of phase even when an identical signal is injected at their input. Although CASPER offers a calibration method based on a *ruby* code to address this issue, the downside is that it halves the available bandwidth. To avoid this trade-off, we developed our method to calibrate the phase offsets of the ADCs. This calibration process is performed using the system identified as (3) in Figure 4.1 and is further explained in Section 4.2.1, ensuring proper synchronization without reducing bandwidth.

Upon entering the ROACH-2, the signals are digitized to 8 bits by the ADCs operating at 140 MSPS, which allows processing signals with a bandwidth of 70 MHz (in accordance with the anti-aliasing filter). Then, the FPGA performs most of the calculations required in the algorithm described in Chapter 3. Specifically, the digital system implemented in the FPGA -illustrated in Figure 4.7- comprises three fundamental parts: the PFB-FFT stage (multiplications by the matrices $\boldsymbol{\mathcal{H}}$ and $\mathbf{W}_H$), the centrosymmetric matrix multiplication stage ($\mathbf{Q}_P$ and $\mathbf{Q}_Q$), and the correlation matrix accumulation and computation stage ($\mathbb{R}\{\mathbf{R}_{zz}\}$), which are explained separately in the following sections. The remaining parts of the DoA estimation algorithm and any additional post-processing calculations are executed on a computer, as detailed in Chapter 5.

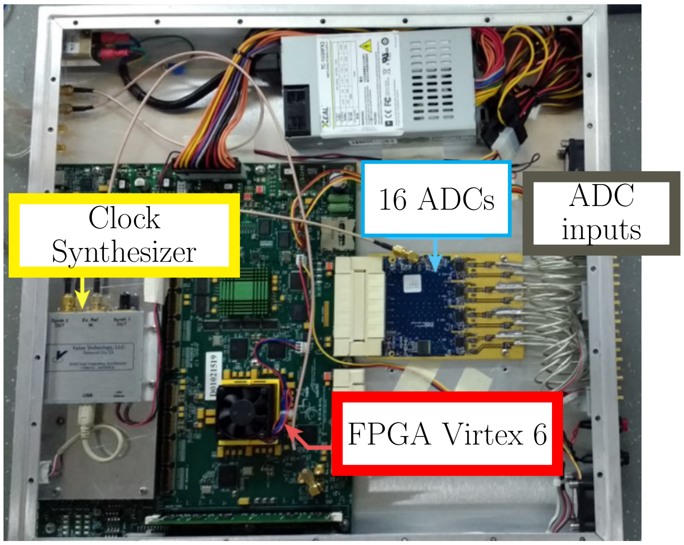


Figure 4.6: Image of the ROACH-2 board.

---

[1] `https://casper.berkeley.edu/wiki/ROACH-2_Revision_2`

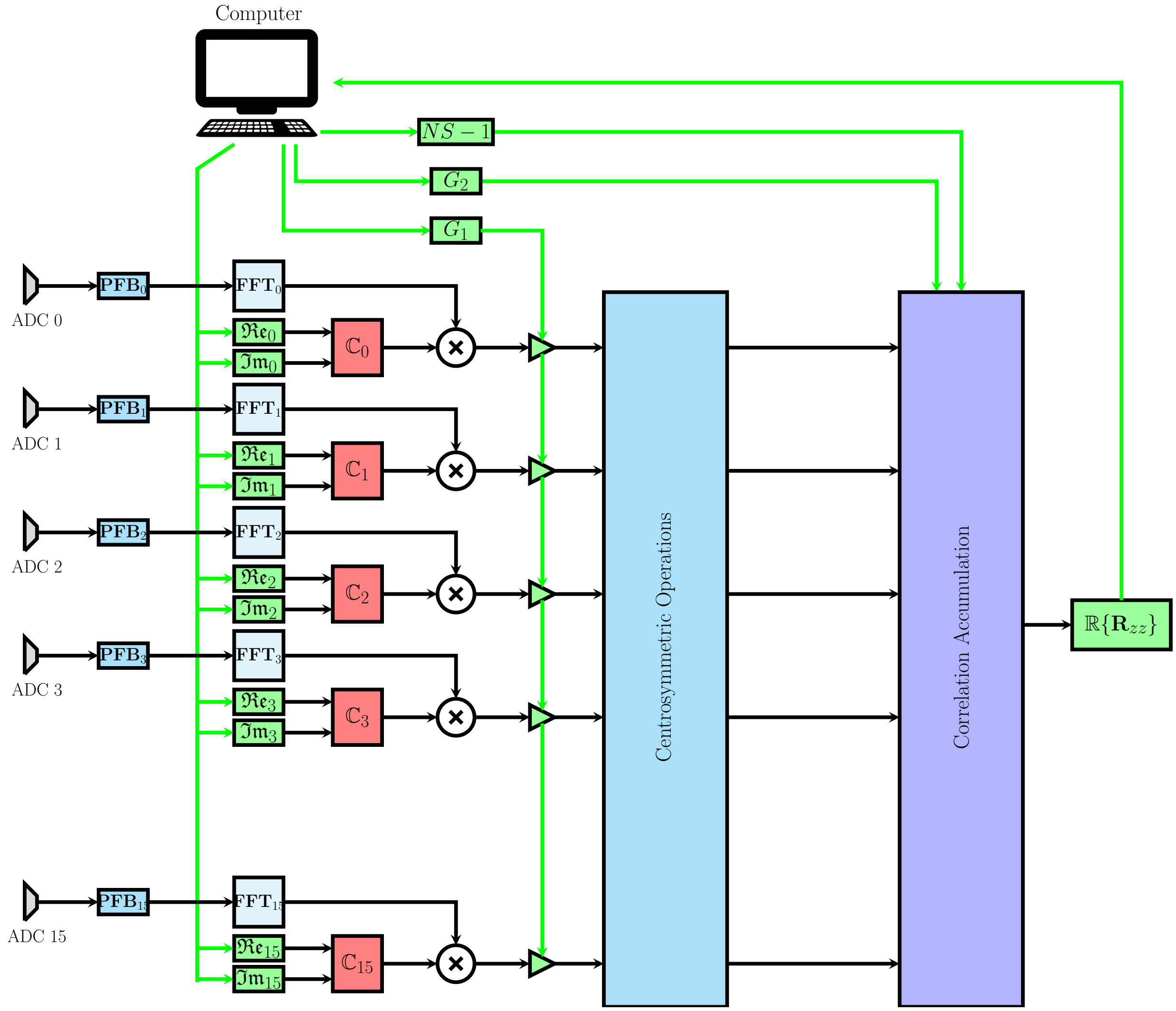


Figure 4.7: Block diagram of the digital system implemented in the FPGA Virtex 6. The PFB and FFT blocks are provided by CASPER, the red blocks represent block-RAMs (i.e., memory), the green blocks represent software-register that can be acceded by the computer, and the green lines represent communication between the FPGA and the computer.

4.2.1 PFB-FFT and Calibration stages

For the PFB-FFT implementation, we employ a 128-point FFT and a PFB incorporating four taps and a Hamming windowing function. The PFB-FFT outputs are then passed to the DoA algorithm, but before that, it is crucial to correct the system's amplitude and phase imbalances. This correction is done using pre-stored values in memory, denoted in Figure 4.7 as $\mathfrak{Re}_i$, $\mathfrak{Im}_i$, and $\mathbb{C}_i = \mathfrak{Re}_i + j\mathfrak{Im}_i$ for $i = 0, \ldots, 15$, representing the real and imaginary components of the correction factors for each frequency channel. By applying a complex multiplication to these values, the system is able to smooth out both amplitude and phase discrepancies across the receiver chains. The primary goal of this calibration step is to adjust the amplitude and phase imbalances presented in the receiver, which are constants measured once, as well as the time delays among the 16 ADCs, which need to be measured each time the ROACH-2 is powered on. In the following sections, we explain the calibration of the receiver and the ADCs separately.

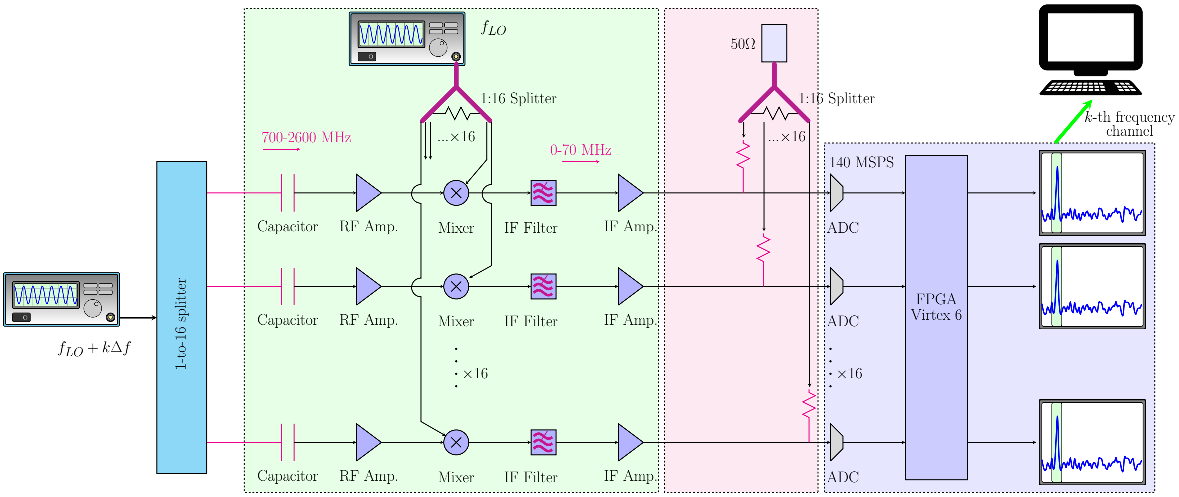


Figure 4.8: Schematic representation of the receiver calibration setup. We show the FFT power spectrum for illustration purposes, but both real and imaginary values of the k-th frequency channel are saved in memory.

4.2.1.1 Calibration of the Receiver

To evaluate the amplitude and phase imbalances of the receiver, we used a signal generator connected to all 16 chains of the receiver via a 1-to-16 splitter, as shown in Figure 4.8. We performed amplitude and phase measurements across the IF bandwidth for each of the six frequency bands listed in Table 1.1, corresponding to LO frequencies of $f_{LO} \in \{690, 820, 922, 1700, 1840, 2570\}$ MHz. For each band, we swept from f_{LO} to $f_{LO} + 70$ MHz, ensuring that each frequency input aligns with a frequency channel in the FFT spectrum. Since we have a 128-point FFT applied to real signals, we have 64 valuable channels per spectrum, and hence, the frequency step of the sweep is calculated as 70 MHz/64= 1.09375 MHz. The receiver calibration algorithm is outlined below:

1) **Select a frequency band**: Start with the 28 APT band and set the LO frequency to 690 MHz, $f_{LO} = 690$ MHz.
2) **Program Frequency Sweep**: Configure the signal generator to sweep frequencies from f_{LO} to $f_{LO} + 70$ MHz, with steps of $\Delta f = 1.09375$ MHz.
3) **Record FFT Spectra**: For each frequency step $f_{LO} + k\Delta f$, acquire the FFT spectra from the 16 ADCs in the FPGA. Save only the k-th frequency channel. Denote the real and imaginary values of the FFT result for the i-th ADC as $A_i(k)$ and $B_i(k)$, respectively.
4) **Compute Amplitude and Phase**: After completing the sweep, collect measurements across the entire IF bandwidth, $\{A_i(k) + jB_i(k)\}_{k=0}^{k=63}$. Compute the amplitude and phase for each receiver chain as $\{\sqrt{A_i^2(k) + B_i^2(k)}\}_{k=0}^{k=63}$ and $\left\{\tan^{-1}\left(\frac{B_i(k)}{A_i(k)}\right) + C(k)\right\}_{k=0}^{k=63}$, $i = 0, ..., 15$, where $C(k)$ is an arbitrary constant (constant for all antennas, but different for all frequency channels) that depends on the phase reference. We chose antenna 0, located at one of the corners of the antenna array, as the reference point for phase calibration. Thus, $C(k) = -\tan^{-1}\left(\frac{B_0(k)}{A_0(k)}\right)$.
5) **Repeat for All LO Frequencies**: Repeat steps (2) through (4) for the remaining LO frequencies.

The measured amplitudes and phases of the six frequency bands listed in Table 1.1 are depicted in Figures 4.9 and 4.10, with flagged frequency channels excluded from consideration.

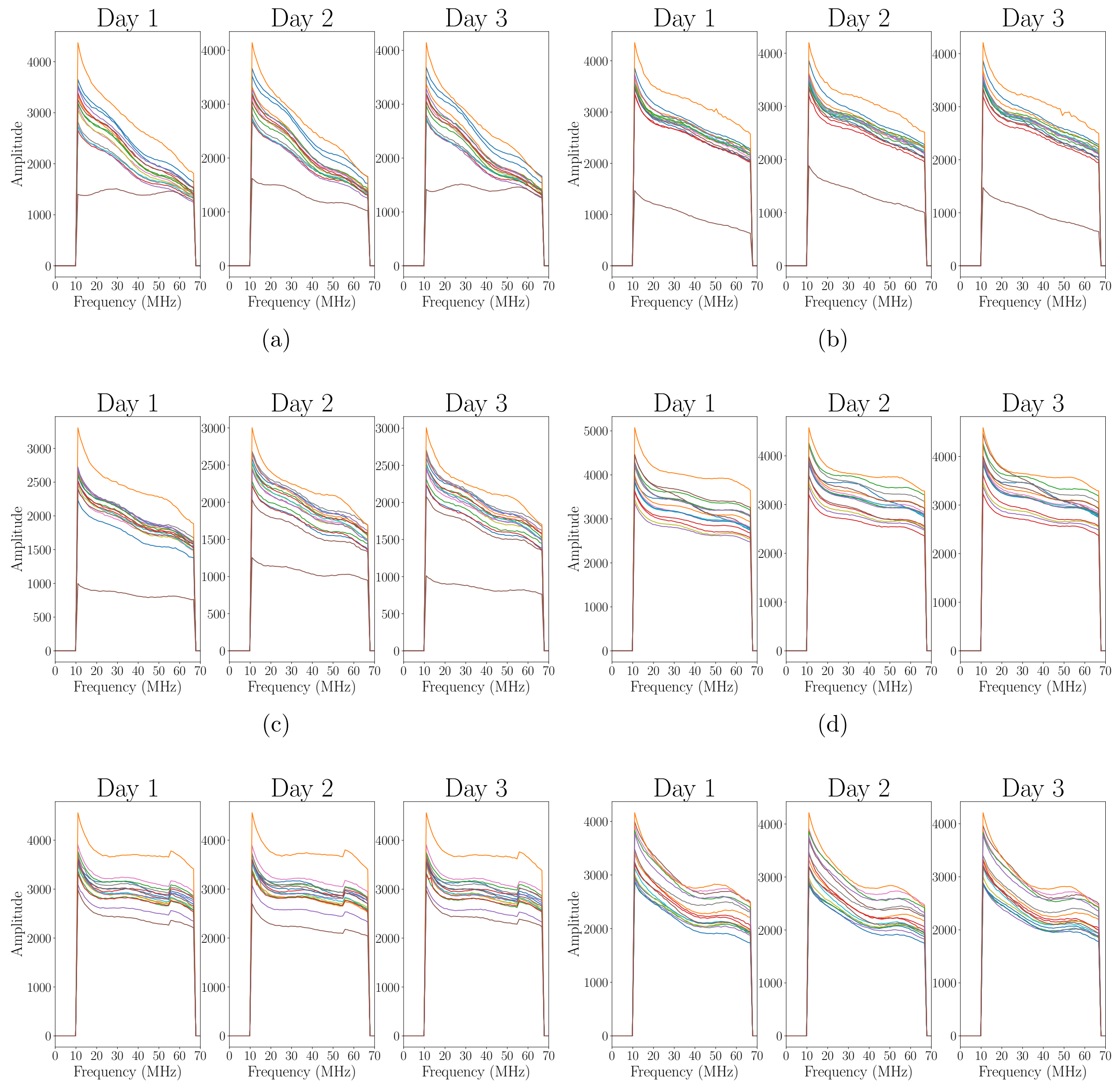
It is important to note that this setup only allows measuring the combined phase shifts from both the receiver and the ADCs. To isolate the receiver phase shifts, the phase offsets from the ADCs must first be measured (as shown in Section 4.2.1.2) and subtracted from the total phase measurements. In terms of amplitude, the values obtained are akin to the conversion gain shown in Figure 4.4, with the key distinction being that this now includes consideration of the (3) part of Figure 4.1.

Given the high gain of the system, an additional concern was the isolation properties of the 1-to-16 splitter used to divide the RF signal from the signal generator. Ideally, the splitter should have infinite isolation to prevent feedback loops. In practice, this was unattainable, so a directional coupler with 40 dB isolation was employed for the measurements. However, since the coupler only splits the signal into two paths, the experiment had to be repeated 15 times, with antenna 0 as a reference for all measurements. The amplitude and phase responses of the coupler were subsequently corrected during post-processing.

To verify the stability of the amplitude and phase calibration values, measurements were repeated on three different days: Day 1, Day 2, and Day 3, with Day 3 occurring one week after Day 1. The differences in amplitude and phase across these days are illustrated in Figures 4.11 and 4.12. Though the amplitude is measured in arbitrary units, the focus remains on the relative differences between the receiver chains rather than the absolute values.

For most of the chains, day-to-day phase variations remained within ±5°, a level of variation deemed negligible for DoA estimation methods, ensuring the reliability of the system's performance. The exception occurred in one chain on Day 2, where significantly different phase values were observed. This deviation was traced to a poor electrical connection at the SMA connectors. Once the connection was reestablished, subsequent measurements returned to the expected range.

Finally, it is worth noting that the phase responses as a function of frequency were not linear, suggesting that the phase shifts are not merely the result of differences in electrical length. Instead, they likely stem from more complex structures involving multiple interconnected paths or equivalent networks of capacitances and inductances forming higher-order systems.



(a)

(b)

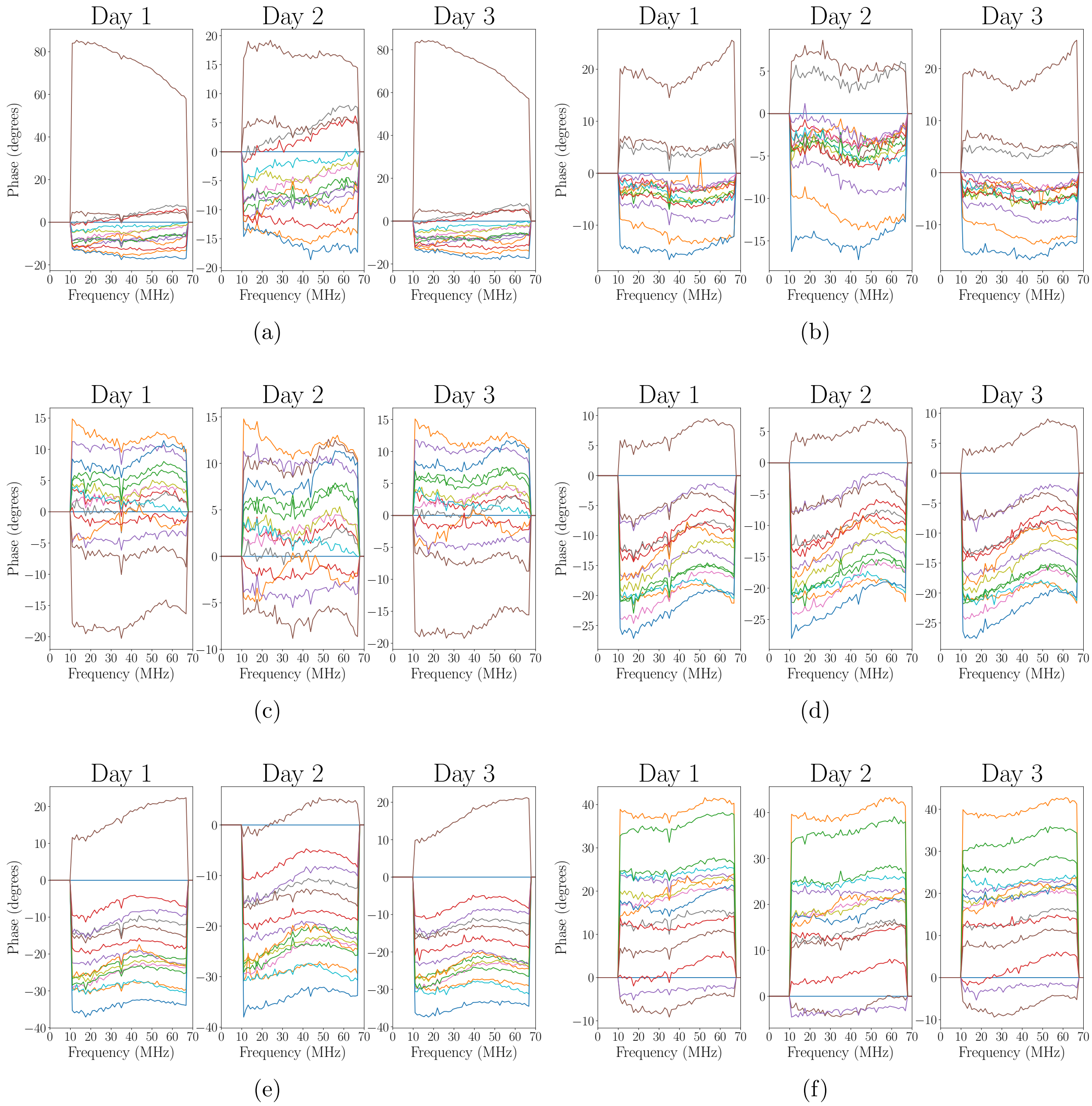
(c)

(d)

(e)

(f)

Figure 4.9: Measured amplitudes of the 16 receiver chains on three different days. (a) LO at 690 MHz. (b) LO at 820 MHz. (c) LO at 922 MHz. (d) LO at 1700 MHz. (e) LO at 1840 MHz. (f) LO at 2570 MHz.



(a)

(b)

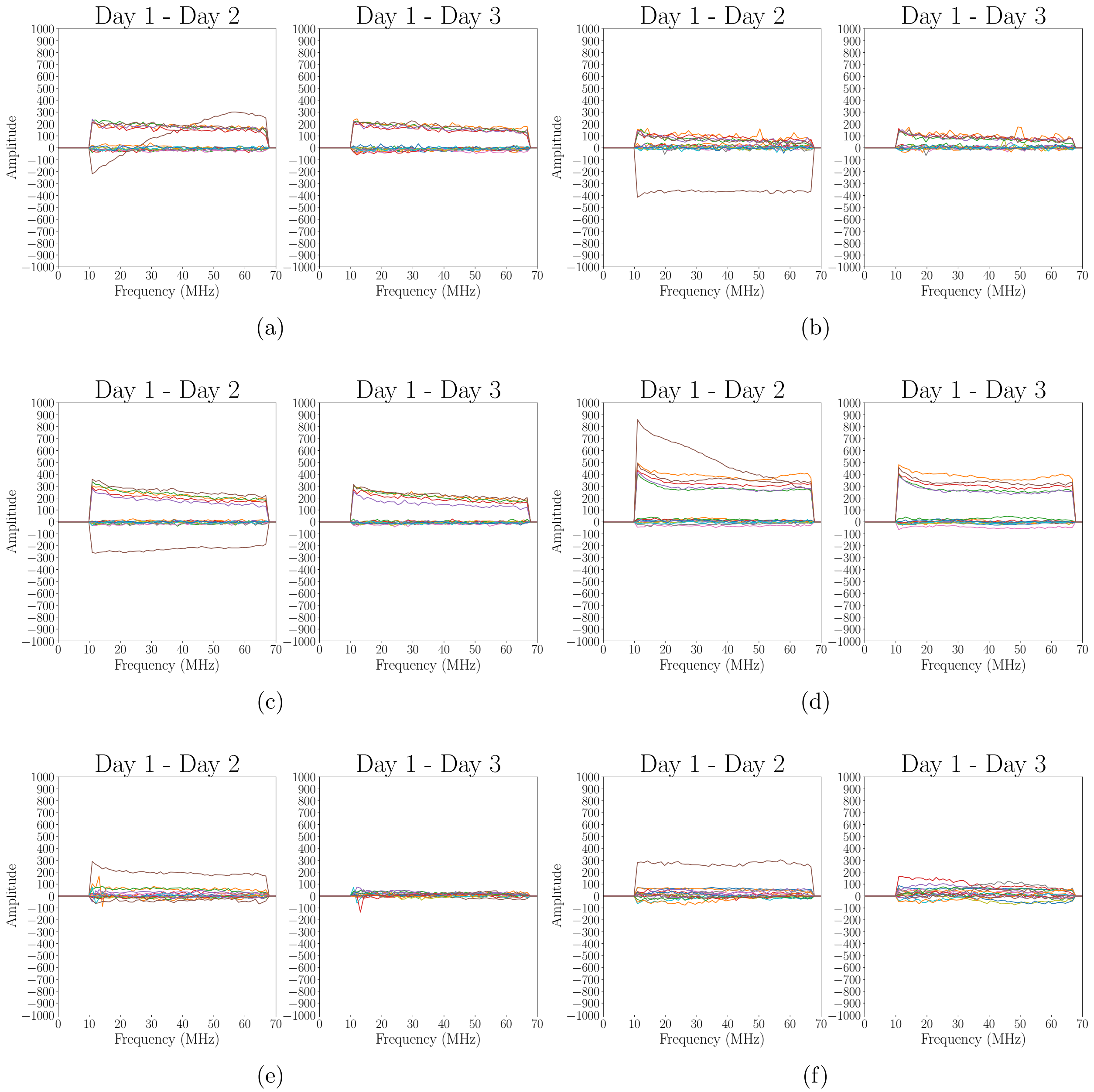
(c)

(d)

(e)

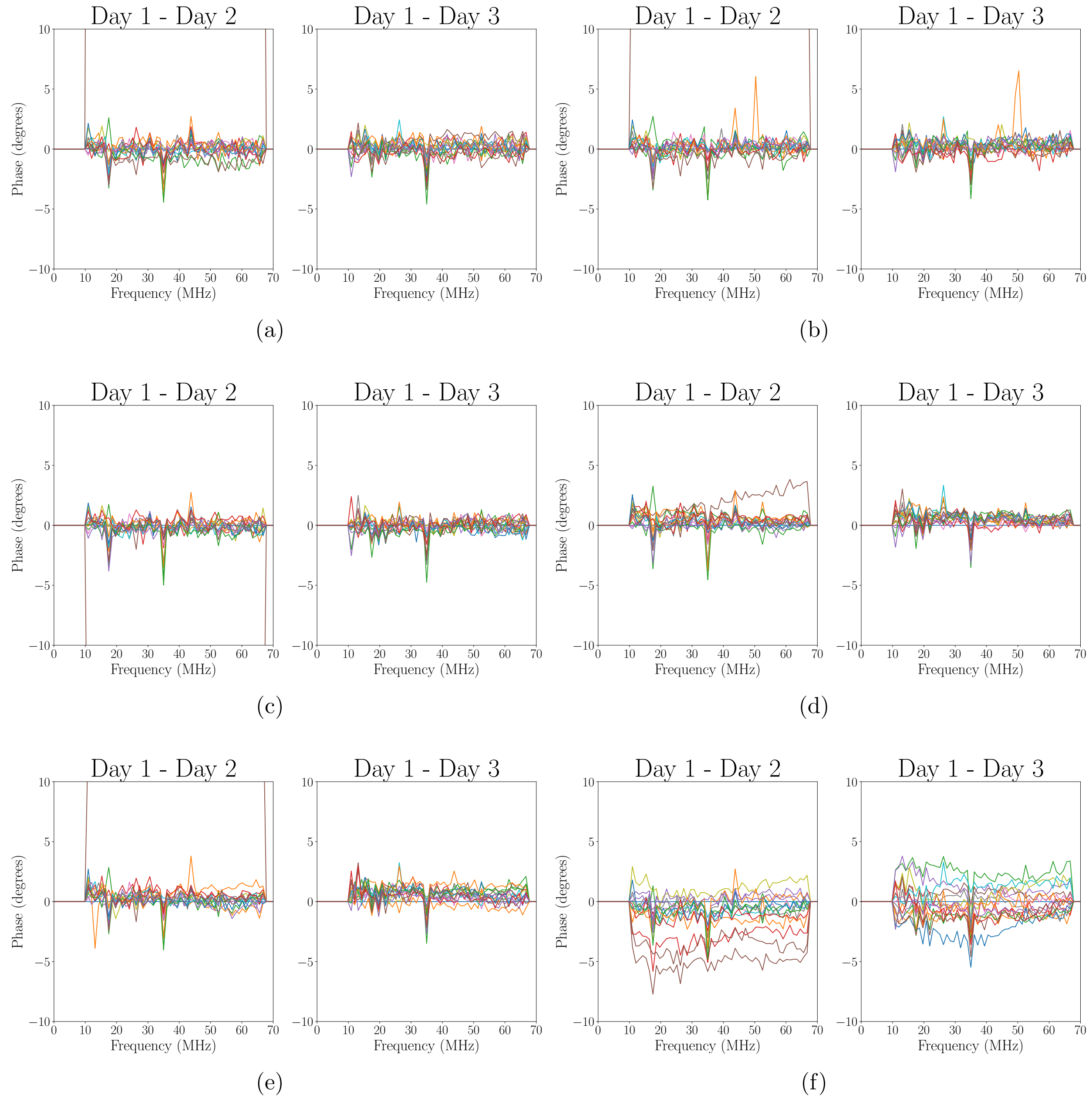
(f)

Figure 4.10: Measured phases of the 16 receiver chains on three different days. (a) LO at 690 MHz. (b) LO at 820 MHz. (c) LO at 922 MHz. (d) LO at 1700 MHz. (e) LO at 1840 MHz. (f) LO at 2570 MHz.



(a) (b) (c) (d) (e) (f)

Figure 4.11: Differences of the measured amplitudes of the 16 receiver chains across consecutive days. (a) LO at 690 MHz. (b) LO at 820 MHz. (c) LO at 922 MHz. (d) LO at 1700 MHz. (e) LO at 1840 MHz. (f) LO at 2570 MHz.



(a) (b) (c) (d) (e) (f)

Figure 4.12: Differences of the measured phases of the 16 receiver chains across consecutive days. (a) LO at 690 MHz. (b) LO at 820 MHz. (c) LO at 922 MHz. (d) LO at 1700 MHz. (e) LO at 1840 MHz. (f) LO at 2570 MHz.

Having obtained the amplitude and phase curves for all receiving chains, with antenna 0 as the reference, the receiver calibration can be carried out by loading the appropriate calibration values

$$\mathfrak{Re}_i(k) + j\mathfrak{Im}_i(k) = \frac{A_0^2(k) + B_0^2(k)}{A_i^2(k) + B_i^2(k)} e^{-j\left(\tan^{-1}\left(\frac{B_i(k)}{A_i(k)}\right) - \tan^{-1}\left(\frac{B_0(k)}{A_0(k)}\right)\right)} \tag{4.1}$$

into the software registers shown in Figure 4.7. However, before loading the registers, the ADCs must first be calibrated, and unwanted frequency channels must be flagged. These additional steps are addressed in Section 4.2.1.3, where the complete calibration process for both the receiver and ADCs is implemented.

4.2.1.2 Calibration of the ADCs

The setup for calibrating the ADCs is shown in Figure 4.13, and the procedure is similar to the one described in Section 4.2.1.1, with a few subtle differences. First, only a single measurement across the IF bandwidth is needed. In this case, the signal generator, connected as shown in the figure, is swept from 0 MHz to 70 MHz with a step size of $\Delta f = 1.09375$ MHz. Second, since the ADCs do not exhibit amplitude imbalances, only the phase must be measured. Lastly, because the receiver is turned off during this process and high gain is not a factor, we can use a 1-to-16 splitter, even if it does not have ideal isolation characteristics.

In this case, we save the phase values (using again the antenna 0 as reference) as

$$\tan^{-1}\left(\frac{D_i(k)}{C_i(k)}\right) - \tan^{-1}\left(\frac{D_0(k)}{C_0(k)}\right), \tag{4.2}$$

where $C_i(k)$ and $D_i(k)$, for $i = 0, ..., 15$, denote the real and imaginary part of the k-th FFT frequency channel of the i-th ADC.

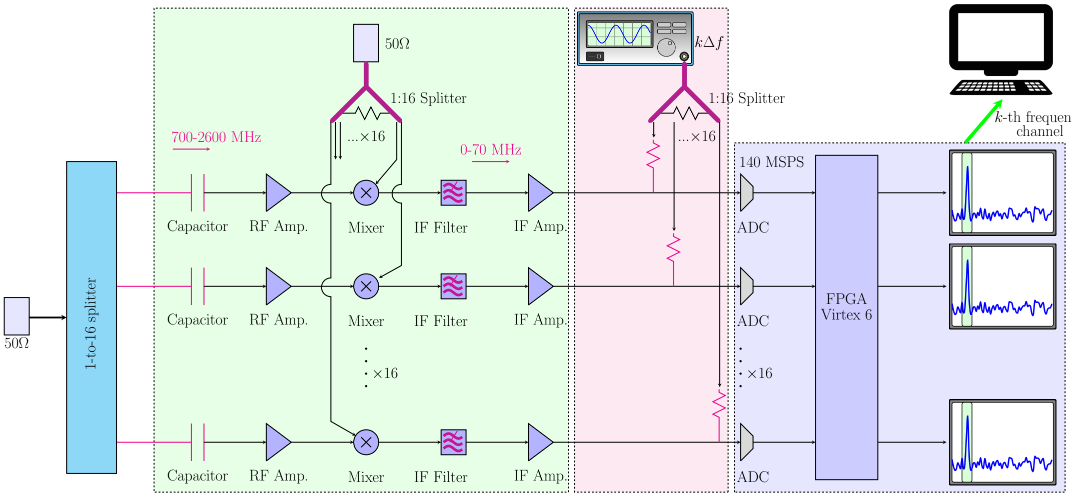


Figure 4.13: Schematic representation of the ADCs calibration setup. We show the FFT power spectrum for illustration purposes, but both real and imaginary values of the k-th frequency channel are saved in memory.

4.2.1.3 Complete Calibration and Frequency flagging

After saving the amplitude and phase measurements of both the receiver and the ADCs, we proceed by writing the software registers shown in Figure 4.7 with the calibration values

$$\begin{aligned}\mathbb{C}_i(k) &= \mathfrak{Re}_i(k) + j\mathfrak{Im}_i(k) \\ &= \frac{A_0^2(k)+B_0^2(k)}{A_i^2(k)+B_i^2(k)} e^{-j\left(\left[\tan^{-1}\left(\frac{B_i(k)}{A_i(k)}\right)+\tan^{-1}\left(\frac{D_i(k)}{C_i(k)}\right)\right]-\left[\tan^{-1}\left(\frac{B_0(k)}{A_0(k)}\right)+\tan^{-1}\left(\frac{D_0(k)}{C_0(k)}\right)\right]\right)}\end{aligned} \tag{4.3}$$

Additionally, during this step, unwanted frequency channels can be flagged by multiplying their corresponding FFT channels by zero. Let $\mathbb{K}$ represent the set of channels to be flagged. Thus, the calibration values written to the software registers are given by

$$\mathbb{C}_i(k) = \begin{cases} 0 & \text{, if } k \in \mathbb{K} \\ \frac{A_0^2(k)+B_0^2(k)}{A_i^2(k)+B_i^2(k)} e^{-j\left(\left[\tan^{-1}\left(\frac{B_i(k)}{A_i(k)}\right)+\tan^{-1}\left(\frac{D_i(k)}{C_i(k)}\right)\right]-\left[\tan^{-1}\left(\frac{B_0(k)}{A_0(k)}\right)+\tan^{-1}\left(\frac{D_0(k)}{C_0(k)}\right)\right]\right)} & \text{, otherwise} \end{cases} \tag{4.4}$$

4.2.2 Centrosymmetric stage

For the implementation of the centrosymmetric operations, notice that the matrices $\mathbf{Q}_P$ and $\mathbf{Q}_Q$ consist mainly of ones and zeros (scaled by $1/\sqrt{2}$). Hence, we emphasize that there is no necessity to implement complicated matrix multiplications in the FPGA. Specifically, we have $P = Q = 4$, and

$$\mathbf{Q}_P = \mathbf{Q}_Q = \begin{bmatrix} 1 & 0 & j & 0 \\ 0 & 1 & 0 & j \\ 0 & 1 & 0 & -j \\ 1 & 0 & -j & 0 \end{bmatrix}, \tag{4.5}$$

we will ignore the $1/\sqrt{2}$ factor since we afterward compute the eigendecomposition of the correlation matrix, unaffected by scaling factors. Hence, if we name the k-th FFT channel associated with the (p,q)-th antenna of the array as $\hat{z}_{pq}(k)$, the multiplication stated in (3.9) is given by

$$\begin{aligned}\hat{\mathbf{Z}}'(k) &= \mathbf{Q}_P^H \hat{\mathbf{Z}}(k) \mathbf{Q}_Q^* \\ &= \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ -j & 0 & 0 & j \\ 0 & -j & j & 0 \end{bmatrix} \begin{bmatrix} \hat{z}_{00}(k) & \hat{z}_{01}(k) & \hat{z}_{02}(k) & \hat{z}_{03}(k) \\ \hat{z}_{10}(k) & \hat{z}_{11}(k) & \hat{z}_{12}(k) & \hat{z}_{13}(k) \\ \hat{z}_{20}(k) & \hat{z}_{21}(k) & \hat{z}_{22}(k) & \hat{z}_{23}(k) \\ \hat{z}_{30}(k) & \hat{z}_{31}(k) & \hat{z}_{32}(k) & \hat{z}_{33}(k) \end{bmatrix} \begin{bmatrix} 1 & 0 & j & 0 \\ 0 & 1 & 0 & j \\ 0 & 1 & 0 & -j \\ 1 & 0 & -j & 0 \end{bmatrix}\end{aligned} \tag{4.6}$$

This multiplication corresponds to additions and subtractions between the real and imaginary parts of the outputs of the FFT blocks of all the ADCs. For example, the first output of the centrosymmetric block shown Figure 4.7 is

$$\begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{z}_{00}(k) + \hat{z}_{03}(k) \\ \hat{z}_{10}(k) + \hat{z}_{13}(k) \\ \hat{z}_{20}(k) + \hat{z}_{23}(k) \\ \hat{z}_{30}(k) + \hat{z}_{33}(k) \end{bmatrix} = \hat{z}_{00}(k) + \hat{z}_{03}(k) + \hat{z}_{30}(k) + \hat{z}_{33}(k) \tag{4.7}$$

Note that multiplying by j swaps the real and imaginary parts, so no multiplications are needed. In conclusion, the centrosymmetric block from Figure 4.7 consists of adders and subtracters, mixing all the real and imaginary parts of the antenna array elements with operations deduced from (4.6).

4.2.3 Correlation and accumulation stage

Finally, for the computation of the correlation matrix, we observe that from (3.19), the real part of the matrix $\mathbf{R}_{zz}$ can be expressed as

$$\begin{aligned} \mathbb{R}\{\mathbf{R}_{zz}\} = \mathbb{R}\{\hat{\mathfrak{z}}\hat{\mathfrak{z}}^H\} &= \mathbb{R}\left\{\sum_{k=0}^{N-1} \hat{\mathfrak{z}}(k)\hat{\mathfrak{z}}(k)^H\right\} \\ &= \sum_{k=0}^{N-1} \mathbb{R}\{\hat{\mathfrak{z}}(k)\}\mathbb{R}\{\hat{\mathfrak{z}}(k)\}^T + \sum_{k=0}^{N-1} \mathbb{I}\{\hat{\mathfrak{z}}(k)\}\mathbb{I}\{\hat{\mathfrak{z}}(k)\}^T. \end{aligned} \tag{4.8}$$

This formulation eliminates the need to construct the full matrix $\hat{\mathfrak{z}}$ and then compute the product $\hat{\mathfrak{z}}\hat{\mathfrak{z}}^H$, which is computationally expensive. Instead, the correlation matrix is built incrementally by summing the products of the real and imaginary parts of each FFT channel at every clock cycle. This significantly reduces the computational load.

Additionally, it is not required to base the accumulation on a single spectrum. If multiple spectrums are available within a time frame that is shorter than the minimum duration of the phone signals (approximately 1 ms, as shown in Figure 3.1), the accumulation can be extended from $k = 0$ to $k = NS - 1$, where S represents the number of spectra.

4.3 Camera Calibration

Once the DoAs are estimated by the digital system (FPGA and computer), their positions in degrees must be mapped to the optical images captured by the camera, which use pixel coordinates. This transformation, converting degrees to pixels (pixel position $= \mathcal{F}$(angular position)), requires a camera calibration.

Standard camera calibration involves placing objects at known angles and recording their pixel positions. A polynomial is then fitted to these data points to approximate $\mathcal{F}$. For this

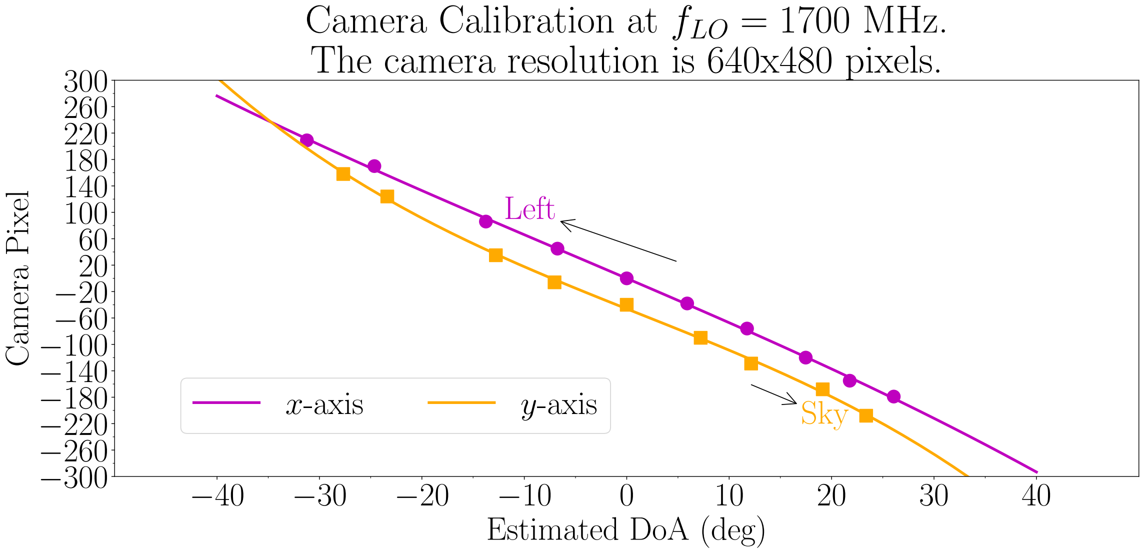


Figure 4.14: Camera calibration and the derived function for transforming estimated DoAs into pixel coordinates.

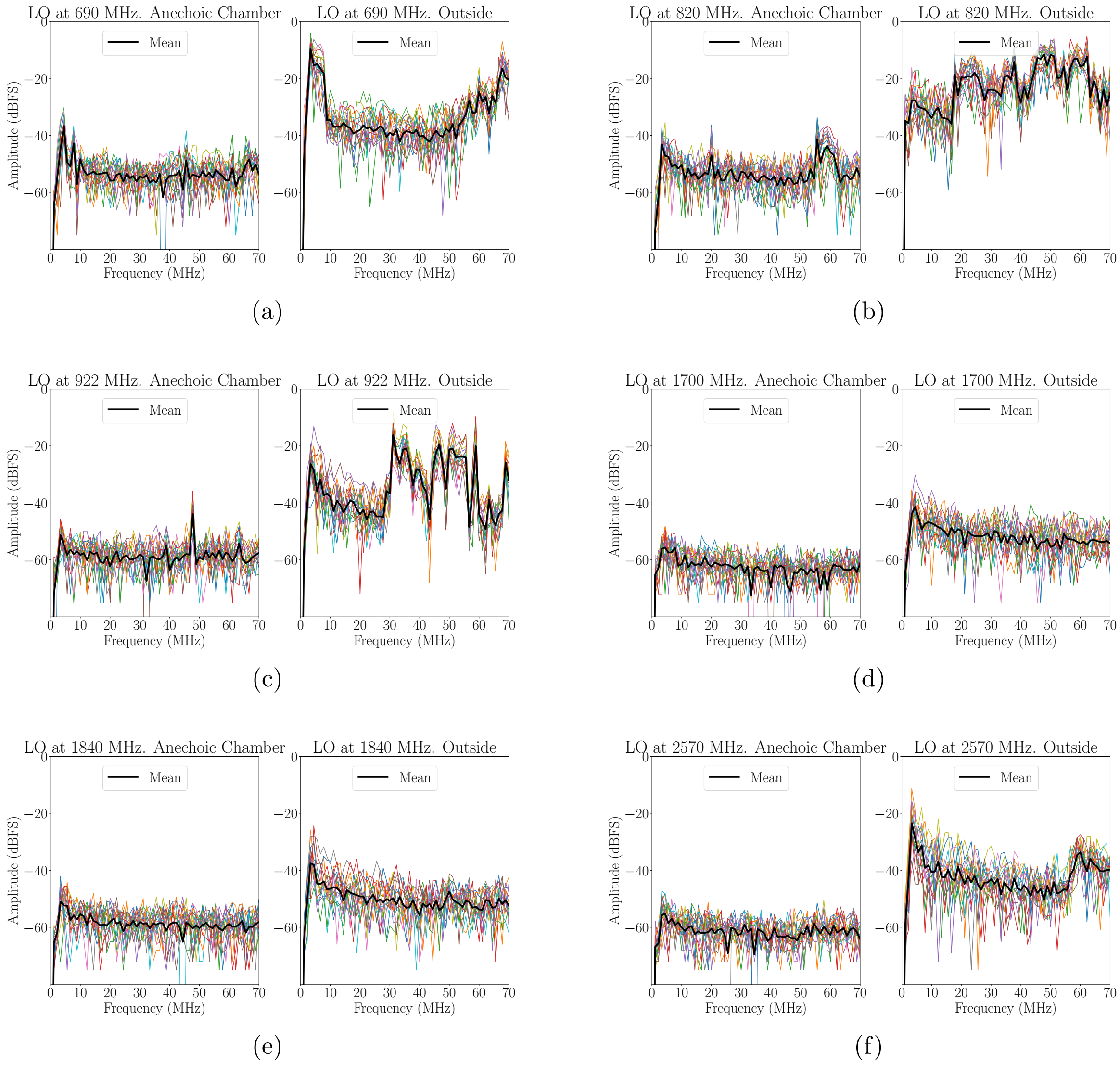
system, however, the calibration process incorporated DoA estimation errors. The camera calibration was performed by placing RF sources in front of the camera, using the algorithm to estimate their angles, and fitting a third-degree polynomial to relate the estimated angles with the observed pixel positions.

Calibrations were conducted at various frequencies, with the 1700 MHz band yielding the most accurate results through trial and error. The resulting transformation function for mapping DoAs to pixels is illustrated in Figure 4.14.

4.4 External Effects

4.4.1 Noise Floor Increase

RadioVision experiences a higher noise floor when operating outdoors due to high-power base stations, possible saturation in the receiver, and/or the mixing effects mentioned in Section 2.2.2. This phenomenon is particularly noticeable compared to operations in controlled environments that experience minimal electromagnetic interference. Figure 4.15 illustrates this difference by showing measurements of the RadioVision noise floor taken in an anechoic chamber, which provides solid electromagnetic isolation, versus measurements taken outdoors at the Astronomy Department of the Universidad de Chile near several base stations. These results highlight the impact of environmental electromagnetic interference on the noise floor, dynamic range, and overall system performance.



(a) (b) (c) (d) (e) (f)

Figure 4.15: Measurement of the noise floor across 16 receiver chains. The FPGA processes the FFTs. (a) LO at 690 MHz. (b) LO at 820 MHz. (c) LO at 922 MHz. (d) LO at 1700 MHz. (e) LO at 1840 MHz. (f) LO at 2570 MHz.

Chapter 5

Software Design

5.1 Task Partitioning: FPGA and Computer Work Division

The DoA computation pipeline is a two-stage system comprising an FPGA and a computer. Each of RadioVision's 16 ADCs captures 8-bit data continuously at 140 MSPS, resulting in a total data rate of 2.24 GB/s (8 bit ×140 MSPS ×16 ADCs = 1 byte ×140 MSPS ×16 ADCs). This massive data flow must be processed in real-time without any loss. To meet this requirement, the FPGA handles tasks such as ADCs and system calibration, channel flagging, and the most time-consuming task of calculating the correlation matrix. Meanwhile, the computer performs the more complex linear algebra operations needed for DoA estimation, the post-processing steps -like applying a moving average to the estimated DoAs- and the superposition of the estimated DoAs onto the optical image captured by RadioVision's camera.

The task partitioning strategy we have chosen is optimal. On the one hand, it is ideal to perform the correlation matrix computation for the 16 antennas on the FPGA due to its ability to execute multiple operations in parallel. The 16-antenna correlation matrix consists of 136 time-accumulations of the form

$$\sum_{k=0}^{NS-1} x_m(k)x_n(k),\ m = 0, ..., 15,\ n = 0, ..., 15\ m \geq n, \tag{5.1}$$

as derived from Section 4.2.3. The FPGA can handle all 136 accumulations in parallel without requiring data to be loaded from RAM. Furthermore, the FPGA can simultaneously manage the necessary multiplications, additions, and register writing. This parallelism allows the FPGA to accumulate $NS - 1$ samples in just $NS - 1$ clock cycles.

Given that the FPGA clock operates at 140 MHz, accumulating up to $2^{18} = 262144$ samples would take only 1.87 ms, aligning with the duration of a telephone beacon -the minimum time unit we want to detect-. Since DoA estimation methods have demonstrated high accuracy with as few as 64 samples [9], we anticipate that the accumulated 2^{18} samples per beacon will yield excellent performance.

On the other hand, implementing the linear algebra steps outside the FPGA is also optimal. In this work, we attempted to execute basic linear algebra operations on the FPGA, such as multiplications, solutions of linear systems of equations, and the QR decomposition to obtain the eigendecomposition of a matrix. Although we obtained satisfactory results in accuracy and efficiency, the computational cost proved prohibitive, and it was not feasible to program all these steps within the FPGA, given the available technology. As a result, the linear algebra computations had to be offloaded to a computer.

5.2 Approach to the Multiple Processes of the Computer Program

Now that we have established the division of tasks between the FPGA and the computer, and considering that the FPGA processes are thoroughly described in Section 4.2, the following sections will detail the processes executed by the computer. Given that RadioVision is a real-time system, the software must continuously handle multiple operations in parallel, such as reading data from the FPGA, performing the required linear algebra computations, and simultaneously generating and displaying the augmented reality image. To achieve this concurrency level and ensure a smooth, continuous display of the augmented reality overlay, the RadioVision software leverages Python's multiprocessing library.

5.2.1 Data Reception and DoA calculation

The software code begins with a loop that requests the correlation matrix from the FPGA, which is transmitted to the computer using a TCP/IP protocol. To efficiently manage this data flow, a non-blocking queue with a capacity of 20 matrices is initialized using the multiprocessing library. The non-blocking nature of the queue ensures that the process does not freeze when it reaches its capacity. Therefore, if the queue becomes full and a new request is made, the other processes continue to operate without interruption. However, any new data will not be entered into the queue and will be lost.

5.2.1.1 First Eigendecomposition

In parallel, we pass the received correlation matrices to a separate process dedicated to performing the linear algebra operations outlined in Chapter 3. Specifically, given that $\mathbb{R}\{\mathbf{R}_{zz}\}$ represents the numpy array of 16×16 elements, the first step in this process is to compute the eigendecomposition of $\mathbb{R}\{\mathbf{R}_{zz}\}$, which is carried out using the Python linear algebra library. Subsequently, we must select the r eigenvectors corresponding to the r largest eigenvalues, where r represents the number of signal sources.

5.2.1.2 Source Number Criteria

For the source number estimation, we tried to implement the methods documented in [13, 14, 15, 16], but none worked stably. For testing with continuous wave controlled sources, the inability to estimate the source number is not a problem since we can force the value of r into the algorithm. However, when treating sporadic sources such as cell phones, it is necessary to address the problem and correctly estimate the number of sources. By implementing the following strategy, we could only differentiate when there is or is not a source.

To estimate the case $r = 0$ reliably, we should note that all eigenvalues of the correlation matrix are close to 0 when there are no signals [6, 7]. Therefore, if we sort the 16 eigenvalues as $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{16}$, it is expected that λ_1 should be not much greater than λ_{16}, and consequently we can define a threshold u_1, such that if λ_1/λ_{16} (dB) $\geq u_1$, we can assume that $r \neq 0$.

For the case $r \geq 2$, we can do something similar, noting that in the presence of more than one signal, the rest of the eigenvalues that follow λ_1 begin to grow in value. Therefore, we can define a second threshold u_2, such that if $\lambda_1/\lambda_2 \geq u_2$, we can assume that $r < 2$. In conclusion, we can reliably estimate both when $r = 0$ and $r = 1$. For $r = 0$, we must ensure that the $\lambda_1/\lambda_{16} \geq u_1$ criterion is not met, while for $r = 1$ both the criteria $\lambda_1/\lambda_{16} \geq u_1$ and $\lambda_1/\lambda_2 \geq u_2$ must met.

Note that if λ_1 (and therefore every λ_i with $i > 1$) is close to λ_{16}, then λ_1/λ_{16} (dB) is also close to 0. Additionally, we have experimentally verified that $\lambda_{16} > 0$ when there is at least one source present, and that $\lambda_{16} = 0$ only occurs when no sources are present and all eigenvalues are near 0. Consequently, we force λ_1/λ_{16} (dB) $= 0$, λ_2/λ_{16} (dB) $= 0$, ..., λ_r/λ_{16} (dB) $= 0$ when $\lambda_{16} = 0$.

Finally, when we know the number of sources in advance (such as in controlled laboratory conditions), we can set r. We can further evaluate λ_1/λ_{16}, λ_2/λ_{16}, ..., λ_r/λ_{16} to estimate the confidence in detecting each of the sources.

5.2.1.3 Solution of the system of equations and Second Eigendecomposition

Once we have identified the number r, we pass the first r eigenvectors and eigenvalues as 16×1 and 16×1 numpy arrays $\mathbf{V}_1$, $\mathbf{V}_2$, ..., $\mathbf{V}_r$ and $\mathbf{L} = [\lambda_1\ \lambda_2\ ...\ \lambda_r]$ to another parallel process. In this process, we form the $16 \times r$ numpy array $\mathbf{V} = [\mathbf{V}_1\ \mathbf{V}_2\ ...\ \mathbf{V}_r]$ and we solve the system of equations (3.30) by applying the pseudo inverse available in the linear algebra library as

$$\begin{aligned} \mathbf{X} &= (\mathbf{K}_{\mu 1}\mathbf{V})^{+}\mathbf{K}_{\mu 2}\mathbf{V} \\ \mathbf{Y} &= (\mathbf{K}_{\nu 1}\mathbf{V})^{+}\mathbf{K}_{\nu 2}\mathbf{V}, \end{aligned} \tag{5.2}$$

where $^{+}$ denotes the pseudo inverse. After obtaining $\mathbf{X}$ and $\mathbf{Y}$, we perform in this same process the eigendecomposition of $\mathbf{X} + j\mathbf{Y}$ (again via the linear algebra library) and then obtain the estimated DoAs from its eigenvalues, as stated in Chapter 3.

5.2.2 Moving Average

In the case of having a single source with multiple detections (for example, a CW source or a cellphone on call), we can use post-processing techniques to improve the result of the estimates. Indeed, measurements with high values of λ_1/λ_{16} can be considered reliable, and measurements with low values of λ_1/λ_{16} can be considered unreliable due to the proximity to noise. With this, a moving average can be formed to obtain a single DoA from a cloud of DoAs. Specifically, we can average N DoAs to form the moving average

$$\begin{aligned} y(n) &= \frac{1}{W(n)} \sum_{k=n-N+1}^{n} x(k)w(k) \\ W(n) &= \sum_{k=n-N+1}^{n} w(k), \end{aligned} \tag{5.3}$$

where $\{x(k)\}$ is the sequence of DoAs calculated from N consecutive correlation matrices, and $\{w(k)\}$ are the weights, given by $\lambda_1(k)/\lambda_{16}(k)$. If any of the criteria $\lambda_1(k)/\lambda_{16}(k) \geq u_1$ or $\lambda_1(k)/\lambda_2(k) \geq u_2$ are not met, the value of $w(k)$ is considered as 0, so it does not contribute to the average.

Note that if many consecutive values of $w(k)$ are null, the value of $W(n)$ will also be null. More generally, if multiple values of $w(k)$ are unreliable, with $\lambda_1(k) \approx \lambda_{16}(k)$ and low SNR, the value of $W(n)$ will be low and equally unreliable. Therefore, similar to the criteria u_1 and u_2, we define a threshold u_3, such that if $W(n) > u_3$, the moving average is considered reliable. Otherwise, we discard the value. In this thesis, we have set $u_3 = 0$ for simplicity, but other values could be explored in the future.

Finally, we can calculate separate moving averages for the DoAs in the directions of x and y. Specifically, denoting the DoAs as $\theta_x(k)$ and $\theta_y(k)$, both weighted by $w(k)$, given by $\lambda_1(k)/\lambda_{16}(k)$, we can define two moving averages as

$$\begin{aligned} \Theta_x(n) &= \frac{1}{W_x(n)} \sum_{k=n-N_x+1}^{n} \theta_x(k)w(k) \\ W_x(n) &= \sum_{k=n-N_x+1}^{n} w(k), \end{aligned} \tag{5.4}$$

and

$$\begin{aligned} \Theta_y(n) &= \frac{1}{W_y(n)} \sum_{k=n-N_y+1}^{n} \theta_y(k)w(k) \\ W_y(n) &= \sum_{k=n-N_y+1}^{n} w(k), \end{aligned} \tag{5.5}$$

where N_x and N_y represent the number of samples used for the x and y direction estimations, respectively.

Experimentally, we observed that at least when RadioVision is close to the ground (≈ 1 m), the values in the y direction tend to have a higher standard deviation, which we attribute to ground reflections. For this reason, in this thesis, we opted to use different sample sizes for the moving averages, with $N_x = 10$ and $N_y = 30$.

5.2.3 Image Formation

For image formation, we execute a parallel process that utilizes Python's OpenCV library to capture frames from the optical camera and overlay the calculated DoA values as markers. These values can represent (1) the instantaneous DoAs computed from the matrix received from the FPGA, (2) the results passed through the moving average filter, or (3) a combination of both. This approach provides flexibility in visualizing the raw data or the smoothed results, ensuring that the final augmented reality image accurately reflects the system's real-time DoA estimations.

5.2.4 Band Change

Finally, the code includes an option to scan all the frequency bands listed in Table 1.1 at intervals of x seconds, where x is a user-programmable variable. This functionality is implemented using the PyVISA Python library, which facilitates communication with external equipment responsible for generating the LO tone. The equipment is connected to the main RadioVision computer via Ethernet, allowing the system to automatically cycle through the different frequency bands at the defined intervals.

Chapter 6

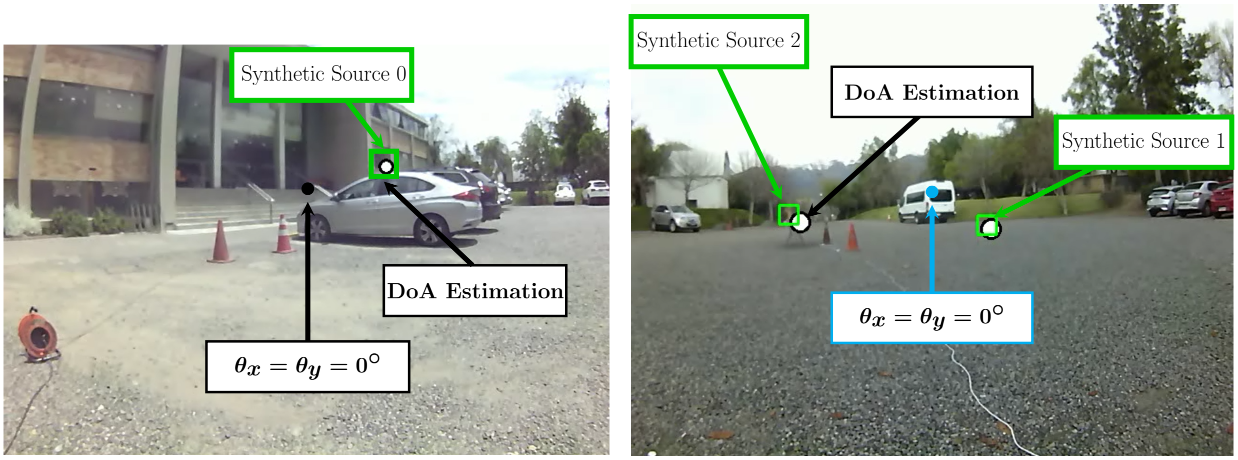
Field Measurement Tests

To validate our system, we conducted several experiments related to source localization. Specifically, experiments were performed with CW sources (i.e., antennas connected to laboratory sources) and cellphones in call. Additionally, the experiments were carried out in various environmental conditions. In the following sections, we present both the setup and the results of each test. We show the CW experiments in Section 6.1 and the experiments with cellphones in Section 6.2. It is worth mentioning that our system can localize single beacons as short as 1 ms in duration and can accurately locate cellphones in stand-by mode. However, to show a larger sample size and enhanced time-based accuracy, we have presented the experiments using only continuous phone calls, excluding tests with messaging or phones in stand-by mode.

6.1 Tests with one and two synthetic CW sources

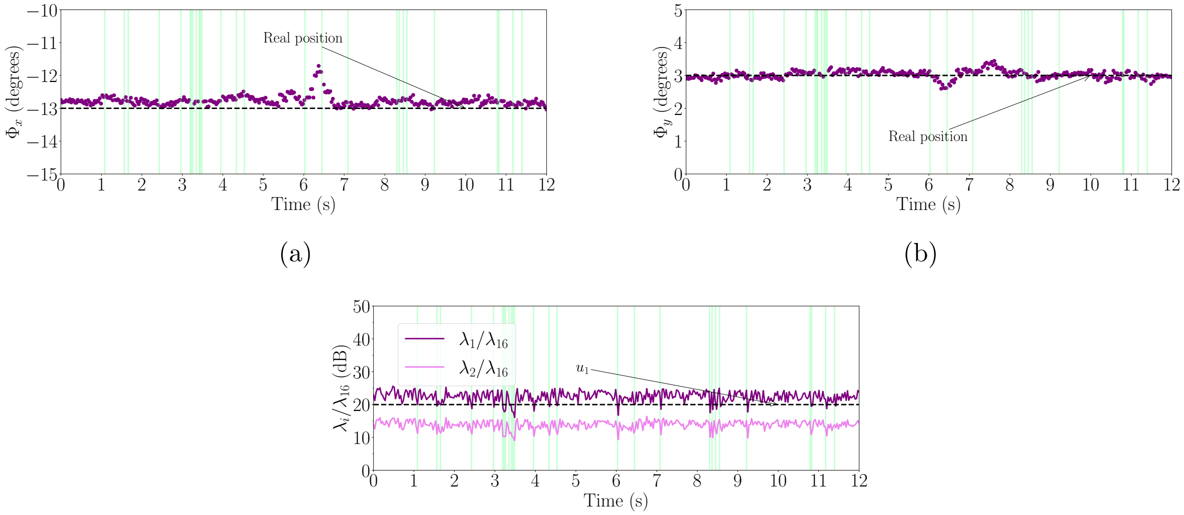
For these experiments, we conducted two tests at the Astronomy Department of the Universidad de Chile, where our laboratory is located. It is important to mention that the area hosts numerous base stations, resulting in interference from nearby stations and mobile phones within the building. In the first test, a single antenna was placed inside an office 20 meters away from RadioVision, as shown in Figure 6.1(a), which displays the optical image captured by RadioVision's camera. In the second test, two antennas were positioned 10 meters apart, as illustrated in Figure 6.1(b). For both tests, RadioVision was set at a height of approximately 1 meter.

In the first test, the source (Source 0) was fed with a sine wave of 1850 MHz and a power of 0 dBm, while the LO was set at 1840 MHz. In the second test, sources 1 and 2 were fed with sine waves of 1870 and 1855 MHz, respectively, both with powers of -5 dBm, while the LO was set at 1840 MHz. Sources 0, 1 and 2 were located at $(-13°, 3°)$, $(-8.8°, -6.4°)$ and $(22.6°, -4.7°)$, respectively. For DoA estimation, we manually forced $r = 1$ in the first test and $r = 2$ in the second test (i.e., the number of sources was not estimated but rather given as known). The experiments lasted 12 and 50 seconds, and the results associated with the DoA estimation are shown in panels (a) and (b) of Figures 6.2 and 6.3. We used an experimentally obtained $u_1 = 20$ dB threshold for these tests. As can be seen, the values obtained are stable



(a) (b)

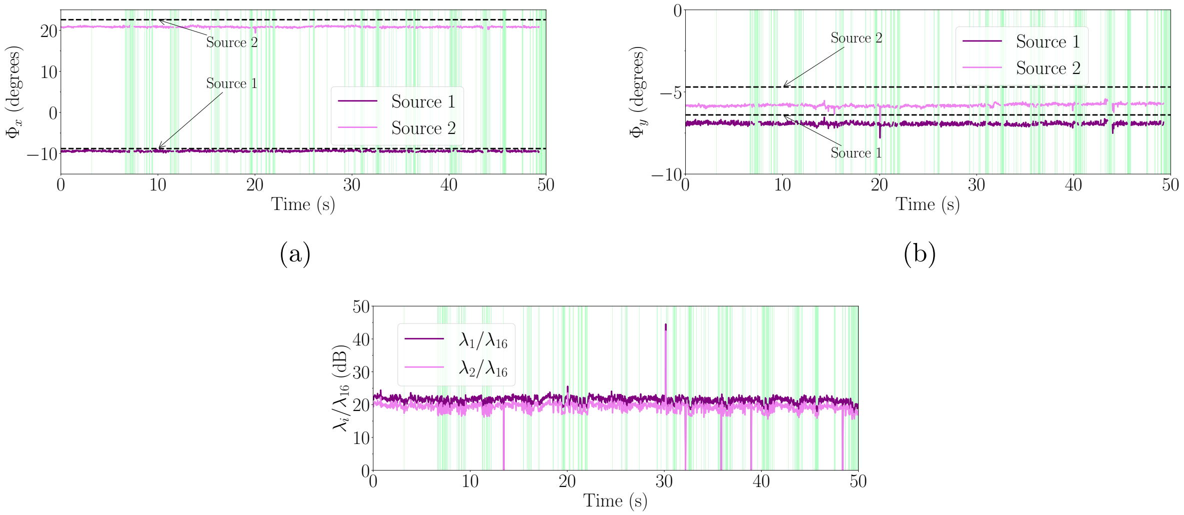
Figure 6.1: Source placement in the astronomy department. (a) Test with one source. (b) Test with two sources.



(c)

Figure 6.2: DoA estimation of a single source located at $(-13°, 3°)$. The experiment lasted 12 seconds, in which 376 accumulation matrices were collected for DoA estimation. (a) DoAs in the horizontal axis. (b) DoAs in the vertical axis. (c) λ_1/λ_{16} and λ_2/λ_{16} criteria. The green areas correspond to the time instants for which $\frac{\lambda_1}{\lambda_{16}} < u_1$, where the detection is considered invalid.

over time for both x and y axes, with low standard deviations and root-mean-square errors (RMSE), as shown in Table 6.1, which demonstrates the accuracy of the algorithm. Note that although the net error in the x axis is larger in source 2, this is because the angle is farther from 0 for that source. Concerning the y axis, it can be seen that there is a greater root mean square error in source 2. The errors in this case are attributed to the proximity to the floor, which contributes to multipath reflections, especially due to the proximity between sources and receivers (10–20 meters). The mean 2D squared errors for sources 0, 1, and 2 are 0.29°, 1.04°, and 2.47°, respectively, resulting in an overall 2D RMSE of 1.27° across all tests.



(a)

(b)

(c)

Figure 6.3: DoA estimation of two sources located at $(-8.8°, -6.4°)$ and $(22.6°, -4.7°)$. The experiment lasted 50 seconds, in which 1550 accumulation matrices were collected for DoA estimation. (a) DoAs in the horizontal axis. (b) DoAs in the vertical axis. (c) λ_1/λ_{16} and λ_2/λ_{16} criteria. The green areas correspond to the time instants for which $\frac{\lambda_1}{\lambda_{16}} < u_1$, where the detection is considered invalid.

Table 6.1: Mean (μ), standard deviation (σ), and RMSE values for the experiments shown in Figure 6.1.

Parameter	**Source 0**, x	**Source 0**, y	**Source 1**, x	**Source 1**, y	**Source 2**, x	**Source 2**, y
Real DoA (deg)	−13	3	−8.8	−6.4	22.6	−4.7
μ (deg)	−12.8	3	−9.4	−6.9	20.8	−5.8
σ (deg)	0.14	0.12	0.56	0.41	1.19	0.34
RMSE (deg)	0.26	0.12	0.81	0.66	2.19	1.15

We emphasize that in these experiments, we plot the raw (or instantaneous) values returned by the FPGA in each accumulation cycle of the correlation matrix (that is to say, we did not use the moving average mentioned in Chapter 5). These results demonstrate the effectiveness of our system under favorable conditions (synthetic sources at a short distance).

For completeness, panel (c) of Figures 6.2 and 6.3 displays the λ_1/λ_{16} and λ_2/λ_{16} criteria, as described in Chapter 5. First, note that for the test with one source, λ_1 is significantly larger than λ_2, whereas their values are more comparable for the test with two sources. Additionally, there are instances where $\frac{\lambda_1}{\lambda_{16}} = 0$ and/or $\frac{\lambda_2}{\lambda_{16}} = 0$. As discussed in Chapter 5, this occurs when $\lambda_1 \approx \lambda_{16}$, indicating no detection. In such cases, the estimated DoA value is unreliable and is therefore omitted.

6.2 Tests with cellphones

For these experiments, we conducted tests in three different locations. The first location is the Astronomy Department, mentioned in Section 6.1. The second location is the Santa Martina ranch, situated on a hill far from the city, with the nearest base station approximately 1000 meters away. The third location is the engineering campus of the Universidad de Chile, which, in addition to its proximity to base stations, experiences significant contamination from other mobile phones. We forced $r = 1$ for these tests and used the source number criteria $\lambda_1/\lambda_{16} \geq u_1$ and $\lambda_1/\lambda_2 \geq u_2$ mentioned in Chapter 5 to validate DoAs estimated with each instantaneous correlation matrix. Also, since we encountered more deviation in these cases due to the lower SNR, we applied the moving average discussed in Chapter 5. For the moving average values to be considered as valid, we applied the $W \geq u_3$ criteria mentioned in Chapter 5.

6.2.1 Astronomy Department

In these experiments, RadioVision was positioned at a height of approximately 1 m, and a person making a phone call moved outside and inside the building at distances of 10–40 m for 310 seconds, as shown in Figure 6.4. To accurately determine the real DoAs of the person, we recovered several control positions by reviewing the video recorded by RadioVision's optical camera. Unlike the experiments in Sections 6.2.2 and 6.2.3, the call was not restricted to a specific band. Instead, we used a WOM phone, which primarily operates on AWS band 10 (at 1710–1770 MHz). The LO was set at 1700 MHz.

The results are shown in Figure 6.5. As observed, there are more outliers in the estimated DoAs for this test. This is attributed to phone emissions from multiple people within the building. We maintained a similar value for u_1 compared to the previous tests, with $u_1 =$ 22 dB, and set $u_2 = 6$ dB, as this threshold helps us discriminate whether there is one source or more than one source. Note that for $t > 280$ seconds, the dispersion of the estimated DoAs increases. This occurs because, at that moment, a car passes in front of RadioVision, as illustrated in Figure 6.4. To give an estimated RMSE value, we considered all the values adjacent to each actual position shown in Figure 6.5 within a 5-second window, obtaining RMSEs of 2.82° and 3.68° for the horizontal and vertical axes, and a 2D RMSE of 4.63°.

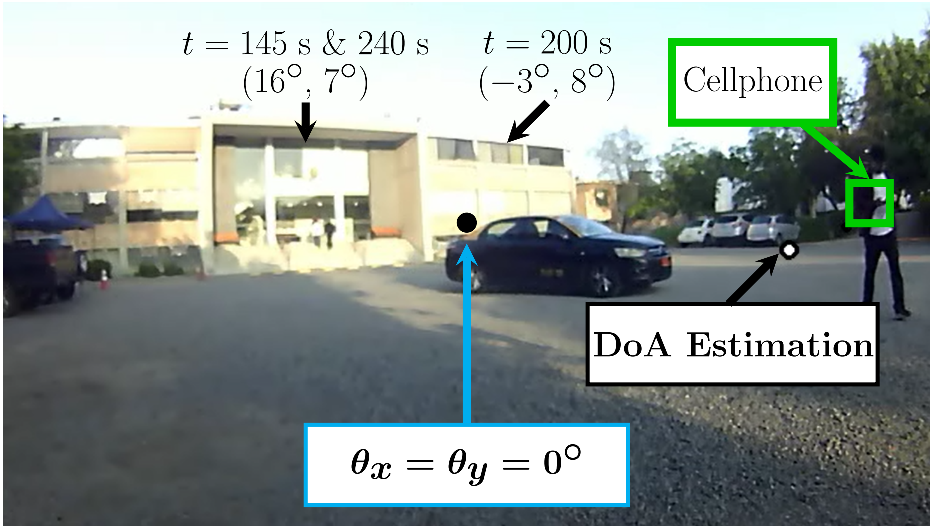
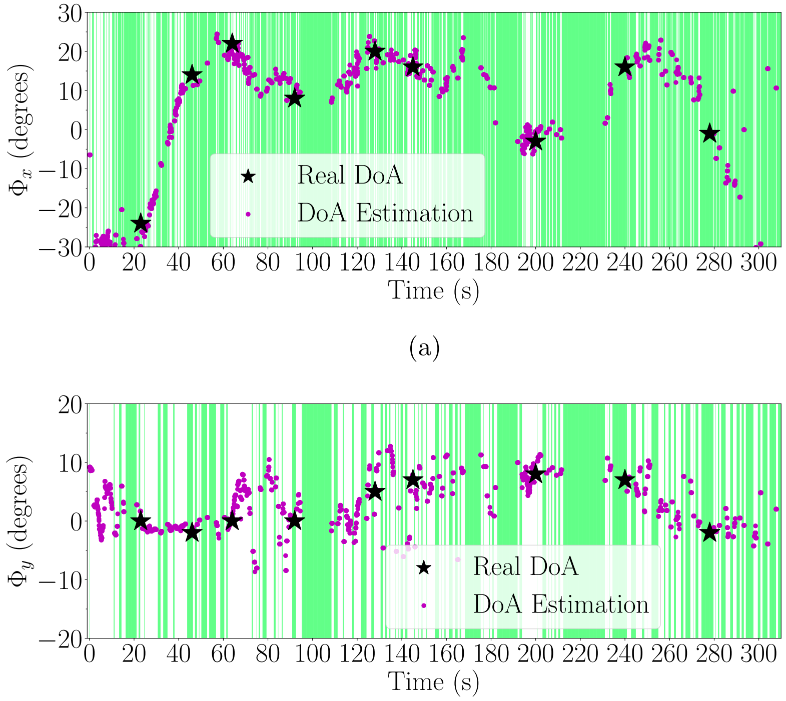


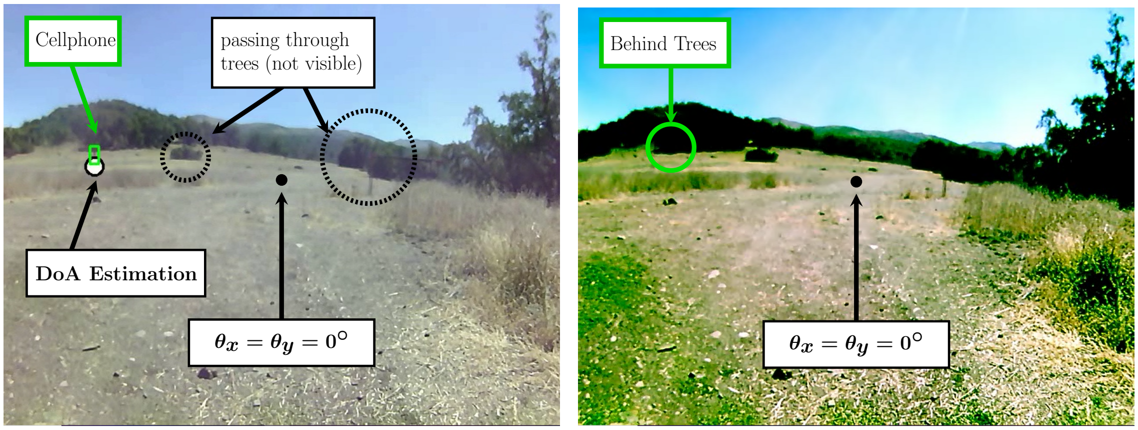
Figure 6.4: Source placement in the astronomy department. The person was inside the building from $t = 90$ s to $t = 260$ s. We highlight control points inside the building during this period, corresponding to specific time stamps $t = 145$ s, $t = 200$ s, and $t = 240$ s. The car, shown in the figure, appeared at $t > 280$ s and crossed in front of RadioVision.



(a)

(b)

Figure 6.5: Averaged values of the test performed in the Astronomy Department. The experiment lasted 310 seconds, in which 9486 accumulation matrices were collected for DoA estimation. The person was inside the building from $t = 90$ s to $t = 260$ s. (a) DoAs in the horizontal axis. (b) DoAs in the vertical axis. The green areas correspond to the time instants for which $W \leq u_3$, where the detection is considered invalid.



(a)

(b)

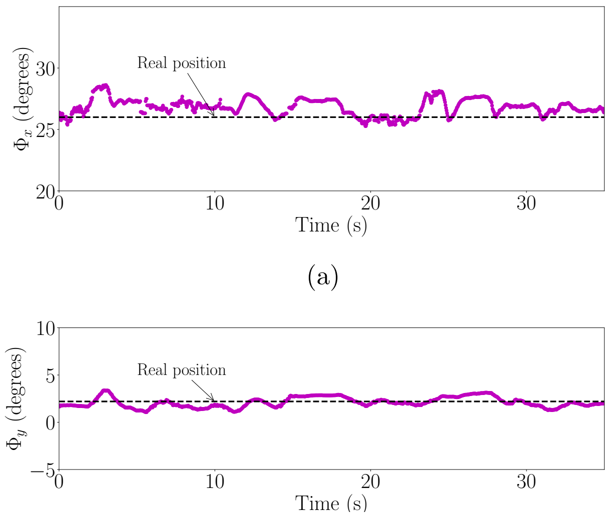
Figure 6.6: Source placement in Santa Martina ranch. (a) Example of the second test. (b) Trees blocking the person in the third test.

6.2.2 Santa Martina ranch

In these experiments, RadioVision was positioned at a height of approximately 1 m, and a person making a phone call moved across the field, as shown in Figure 6.6. The call was forced to the 2 PCS band (at 1850— 1910 MHz) using a cellphone application, and therefore, the LO was set to 1840 MHz.

In this location, we conducted four tests. In the first test, the person making the phone call remained stationary at coordinates $(26°, 2.2°)$, positioned 80 meters away for 35 seconds. In the second test, the person moved across the field at a distance of 80–100 meters for 175 seconds. Computer vision techniques were used to calculate the actual position of the person (and therefore the source). It is important to note that part of the route involved passing behind bushes, as shown in Figure 6.6(a). At those moments, the optical camera could not capture the actual position. In the third test, the person remained stationary behind trees (Figure 6.6(b)) at coordinates $(31.3°, 5.5°)$, 120 meters away, for 26 seconds. In the fourth test, the person was stationary at coordinates $(-7.6°, 2.2°)$, approximately 180 meters away, for 23 seconds, with no direct line of sight between RadioVision and the person. The results of these tests are presented in Figures 6.7, 6.8, 6.9 and 6.10. For all these tests, we used $u_1 = 18.5$ dB and $u_2 = 4.5$ dB. Compared to the values used in Section 6.2.1, we decreased u_1 since the source was further, and we decreased u_2 because, in this case, there was no interference from other phones.

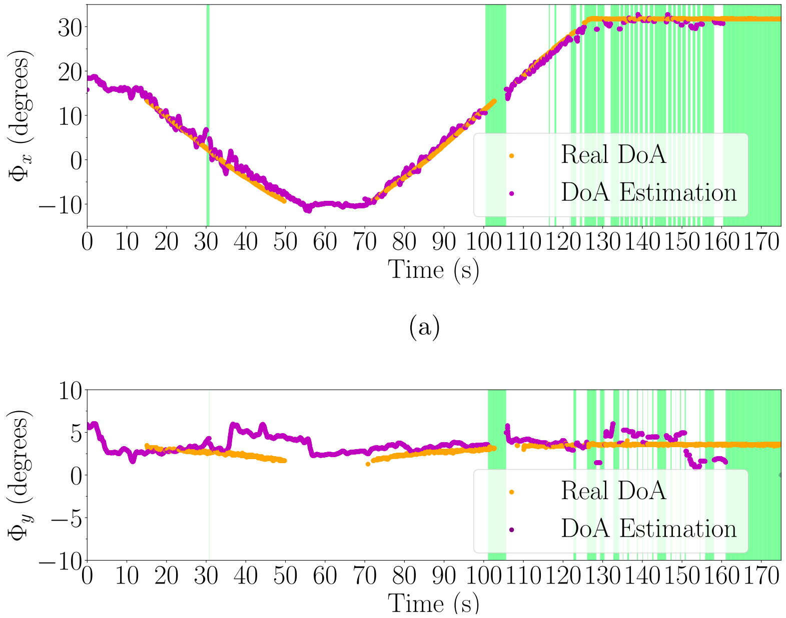
The four tests' mean, standard deviation and RMSE values are presented in Table 6.2. The 2D RMSEs for the tests are 1.19°, 1.78°, 5.71° and 3.79°, resulting in an overall 2D RMSE of 3.12° across all tests. We observe an increment in error due to the lower SNR when comparing to Table 6.1, but the error remains below 5 degrees. Note that the error on the x-axis is more significant than the error on the y-axis in the first test because the x angle is further from 0 in that case. Note also that the most significant error in the third test occurred along the y axis, where the person was positioned behind the trees. While it may seem logical to attribute this error to the presence of the trees, this is not the case, as a similar error would be expected along the x axis, but it does not appear there. Therefore, we hypothesize that the significant error in the third test results from a combination of two factors: (1) the low SNR caused by the trees and (2) reflections from the ground, which were discussed earlier in Section 6.1.



(a)

(b)

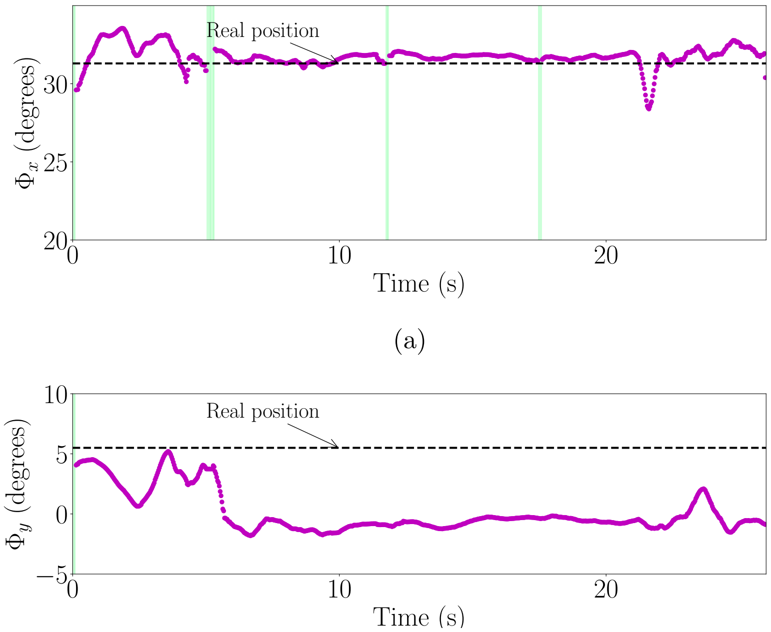
Figure 6.7: Averaged values of the first test performed in the Santa Martina ranch. The experiment lasted 35 seconds, in which 1101 accumulation matrices were collected for DoA estimation. (a) DoAs in the horizontal axis. (b) DoAs in the vertical axis.



(a)

(b)

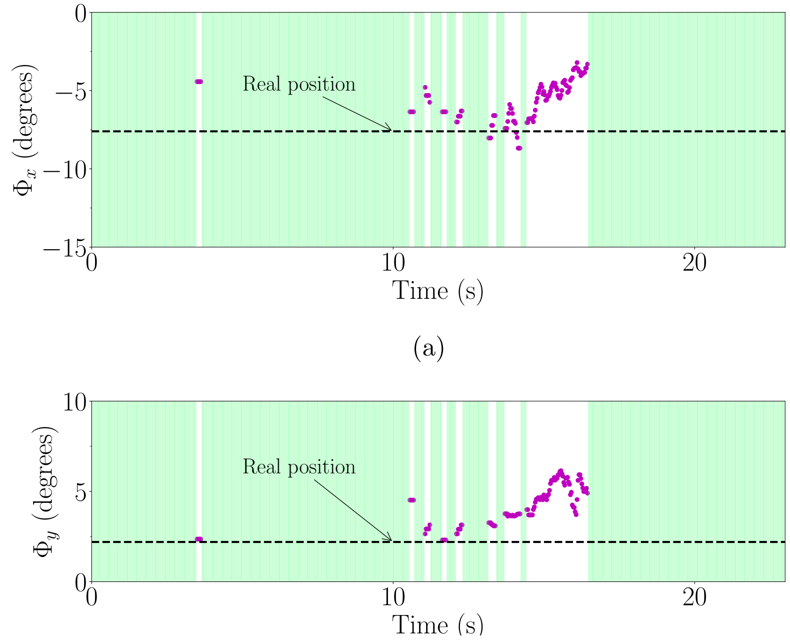
Figure 6.8: Averaged values of the second test performed in the Santa Martina ranch. The experiment lasted 175 seconds, in which 5395 accumulation matrices were collected for DoA estimation. (a) DoAs in the horizontal axis. (b) DoAs in the vertical axis. The green areas correspond to the time instants for which $W \leq u_3$, where the detection is considered invalid.



(a)

(b)

Figure 6.9: Averaged values of the third test performed in the Santa Martina ranch. The experiment lasted 26 seconds, in which 798 accumulation matrices were collected for DoA estimation. (a) DoAs in the horizontal axis. (b) DoAs in the vertical axis. The green areas correspond to the time instants for which $W \leq u_3$, where the detection is considered invalid.



(a)

(b)

Figure 6.10: Averaged values of the fourth test performed in the Santa Martina ranch. The experiment lasted 23 seconds, in which 730 accumulation matrices were collected for DoA estimation. (a) DoAs in the horizontal axis. (b) DoAs in the vertical axis. The green areas correspond to the time instants for which $W \leq u_3$, where the detection is considered invalid.

Table 6.2: Source-Receiver Distance (D), mean (μ), standard deviation (σ), and RMSE values for the experiments shown in Figure 6.6.

Parameter	D (m)	Real DoA (deg)	μ (deg)	σ (deg)	RMSE (deg)
Test 1, x	80	26	26.8	0.67	1.06
Test 1, y	80	2.2	2.1	0.53	0.53
Test 2, x	80–100	-	-	-	1.11
Test 2, y	80–100	-	-	-	1.39
Test 3, x	120	31.3	31.8	0.65	0.8
Test 3, y	120	5.5	0.1	1.75	5.65
Test 4, x	180	-7.6	-5.2	2.37	3.34
Test 4, y	180	2.2	3.7	1	1.79

6.2.3 Engineering Campus

We conducted the final tests on the engineering campus, which is not only close to several base stations but also very busy, resulting in significant interference from other phones. For these tests, RadioVision was positioned at a height of approximately 5 meters, and a person making a phone call moved across the field of view, as shown in Figure 6.11. Due to the presence of many people and the limited visibility of much of the route in the optical image, we were unable to use computer vision techniques to track the source's position accurately. However, we recovered some control positions by reviewing the video footage.

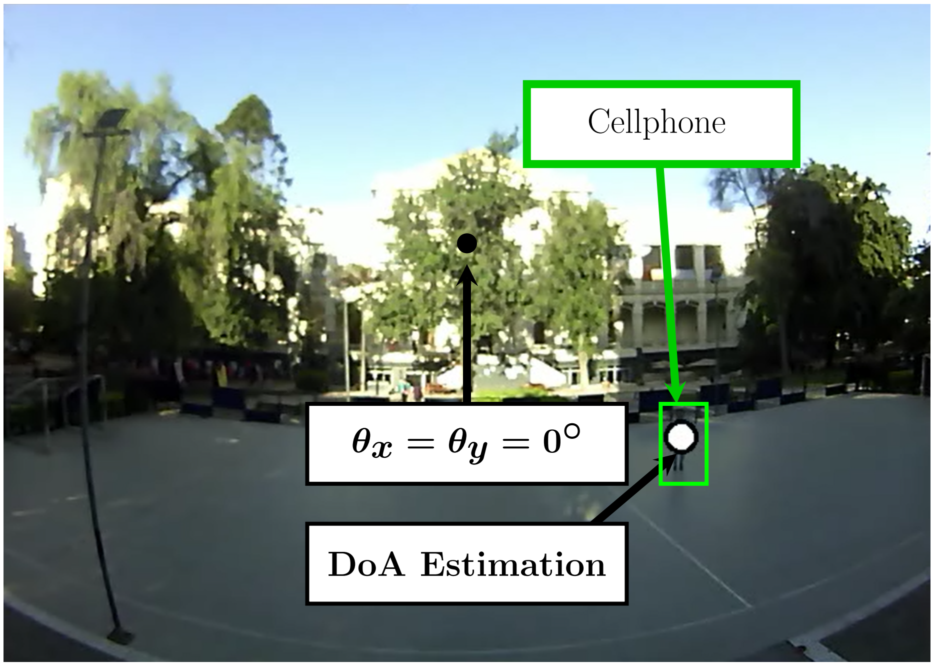
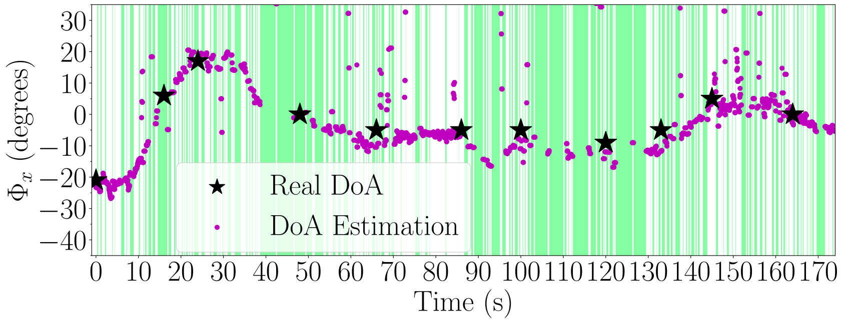
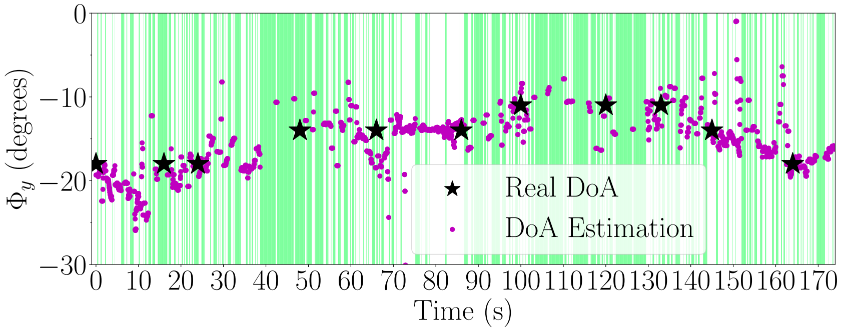


Figure 6.11: Source placement in the engineering campus.



(a)



(b)

Figure 6.12: Averaged values of the experiment performed in the Engineering campus. The experiment lasted 174 seconds, in which 5356 accumulation matrices were collected for DoA estimation. (a) DoAs in the horizontal axis. (b) DoAs in the vertical axis. The green areas correspond to the time instants for which $W \leq u_3$, where the detection is considered invalid.

For this experiment, we used the cellphone application to force the call to the 10 AWS band (at 1710— 1770 MHz), and therefore, the LO was set to 1700 MHz. The person making the telephone call moved across the field at a 10–60 m distance for 174 seconds. The results are shown in Figure 6.12. As can be seen, in this case, there are more outliers in the scatter plot. This is due to phone emissions from multiple people on campus. For this reason, we maintained a similar value of u_1 compared to the previous tests; $u_1 = 18$ dB, but we increased u_2 up to 6 dB. To give an estimated RMSE value, we considered all the values adjacent to each actual position shown in Figure 6.12 within a 2-second window, obtaining RMSEs of 5.42° and 1.94° for the horizontal and vertical axes, and a 2D RMSE of 5.76°. We used a smaller window compared to Figure 6.5 due to the large number of outliers observed at $t = 66$ s, $t = 86$ s, and $t = 100$ s. We attribute the higher error in the x axis to these outliers. Note that in this experiment, the actual DoA on the y-axis is further from 0 compared to all previous experiments due to the higher positioning of RadioVision. Hence, we expect a greater error in the y-axis DoA estimation based on prior results. However, the y-axis RMSE is only 1.94°, comparable to previous experiments where the actual DoA was closer to 0. Since we positioned RadioVision further from the ground in this test, reducing its exposure to multipath reflections, we can hypothesize that the more significant y-axis errors in earlier experiments were due to proximity to the ground.

In summary, the 2D error from the CW source tests is 1.27°, while the 2D errors from the telephone tests at the astronomy department, Santa Martina ranch, and the engineering campus are 4.63°, 3.12°, and 5.76°, respectively. This results in an overall 2D error of 3.7° across all experiments, which is lower than the 5° threshold mentioned in Chapter 1. We conclude that RadioVision meets the objectives set for this thesis and demonstrates sufficient precision for use in disaster scenarios and for phone confiscation in prisons.

Chapter 7

Conclusions

We have developed a 2D receiver array operating in the 700-2600 MHz range for real-time Direction of Arrival (DoA) estimation. Implementing the U-ESPRIT 2D algorithm in the frequency domain, combined with an analog and digital system, has proven effective in mitigating common challenges associated with DoA estimation, such as interference from multiple sources and the calibration of phase and amplitude imbalances.

Experimental results in various environments, including urban and rural settings, validate the system's ability to accurately estimate the DoA of radio frequency sources. The system achieved a mean 2D squared error of 3.12° at 80–180 meters in rural environments, while 4.63° and 5.76° at 10–60 meters in more complex urban environments (Astronomy department and Engineering campus, respectively). These results underscore the system's potential for practical applications requiring precise localization of RF sources in two angular dimensions.

RadioVision's ability to process real-time data and overlay DoA markers onto optical images provides an intuitive interface for its operation and future developments in DoA estimation, radio spectrum awareness, and related applications. In particular, tests have demonstrated that the device can locate phones in relevant environments successfully. These results highlight the potential of RadioVision for use in disaster scenarios and for phone confiscation in prisons.

Further Work

For future work, on the one hand, we propose several improvements in both software and hardware, which are prioritized as follows:

- Implement reject-band filters after the RF amplifiers to eliminate the most problematic base stations, which degrade RadioVision's performance across all bands. This enhancement could increase the dynamic range, extending the maximum detection distance.
- Migrate the Python code to C and implement it in the dedicated CPU of the ROACH-2, increasing computational speed.

On the other hand, considering the recent technological advancements, particularly the development of RFSoC boards [1] that outperform the ROACH-2, a new version of RadioVision could be developed. This version would include 5 GSPS ADCs, a miniaturized digital hardware design, and a more compact 2×2 or 2×4 antenna array, capable of locating up to 2 and 4 simultaneous signals, respectively. With this design, RadioVision would become even more portable, and much of the analog receiver could be migrated to its digital counterpart. Specifically, we could implement a downconverter with a fixed LO at 700 MHz, and with the high 5 GSPS sampling rate, the system could digitize the entire 700–2600 MHz frequency range.

---

[1] `https://www.amd.com/es/corporate/university-program/aup-boards/rfsoc4x2.html`

Bibliography

[1] `https://www.emarketer.com/topics/industry/smartphone-users` [05/01/2025]

[2] `https://www.gsma.com/` [25/04/2020]

[3] L. C. Godara, “Application of antenna arrays to mobile communications. II. Beamforming and direction-of-arrival considerations”, *Proc. IEEE*, vol. 85, no. 8, pp. 1195–1245, 1997, doi: 10.1109/5.622504.

[4] H. Krim and M. Viberg, “Two decades of array signal processing research: The parametric approach”, *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, 1996, doi: 10.1109/79.526899.

[5] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An overview of signal processing techniques for millimeter wave MIMO systems”, *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, 2016, doi: 10.1109/JSTSP.2016.2523924.

[6] R. Schmidt, “Multiple emitter location and signal parameter estimation”, *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276-280, 1986, doi: 10.1109/TAP.1986.1143830.

[7] R. Roy and T. Kailath, “ESPRIT-estimation of signal parameters via rotational invariance techniques”, *IEEE Trans. on Acoust. Speech, Signal Process.*, vol. 37, no. 7, pp. 984-995, 1989, doi: 10.1109/29.32276.

[8] M. Haardt and J. A. Nossek, ”Unitary ESPRIT: how to obtain increased estimation accuracy with a reduced computational burden”, *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1232-1242, 1995, doi: 10.1109/78.382406.

[9] M.D. Zoltowski, M. Haardt and C.P. Mathews, “Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT”, *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 316-328, 1996, doi: 10.1109/78.485927.

[10] W. L. Stutzman and G. A. Thiele, ”Antenna Theory and Design”, Third Edition, USA, John Wiley & Sons, 2013.

[11] C. Balanis, ”Antenna Theory: Analysis and Design”, Fourth Edition, USA, John Wiley & Sons, 2016.

[12] S. P. Applebaum, ”Adaptative Arrays”, *IEEE Trans. Antennas Propag.*, vol. 24, no. 5,

pp. 585-598, 1976, doi: 10.1109/TAP.1976.1141417.

[13] W. Chen, K. M. Wong and J. P. Reilly, "Detection of the number of signals: a predicted eigen-threshold approach", *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1088-1098, 1991, doi: 10.1109/78.80959.

[14] Guanghan Xu, R. H. Roy and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property", *IEEE Trans. Signal Process.*, vol. 42, no. 1, pp. 102-112, 1994, doi: 10.1109/78.258125.

[15] L. Fulai, Z. Xiyuan, L. Chun and W. Jinkuan, "A Robust Algorithm for Source Number Detection and 2-D DOA Estimation Based on Real-valued Computation", *2006 8th international Conference on Signal Processing*, Guilin, China, 2006, doi: 10.1109/ICOSP.2006.344535.

[16] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387-392, 1985, doi: 10.1109/TASSP.1985.1164557.

[17] J. Foutz, A. Spanias and M. Banavar, "Narrowband Direction of Arrival Estimation for Antenna Arrays Synthesis", Morgan & Claypool, 2008.

[18] A. A. Hussain, N. Tayem, M. O. Butt, A. -H. Soliman, A. Alhamed and S. Alshebeili, "FPGA Hardware Implementation of DOA Estimation Algorithm Employing LU Decomposition", *IEEE Access*, vol. 6, pp. 17666-17680, 2018, doi: 10.1109/ACCESS.2018.2820122.

[19] A. A. Hussain, N. Tayem, A. -H. Soliman and R. M. Radaydeh, "FPGA-Based Hardware Implementation of Computationally Efficient Multi-Source DOA Estimation Algorithms", *IEEE Access*, vol. 7, pp. 88845-88858, 2019, doi: 10.1109/ACCESS.2019.2926335.

[20] A. D. Redondo, T. Sanchez, C. Gomez, L. Betancur and R. C. Hincapie, "MIMO SDR-based implementation of AoA algorithms for Radio Direction Finding in spectrum sensing activities", *IEEE COLCOM 2015*, pp. 1-4, 2015, doi: 10.1109/ColComCon.2015.7152096.

[21] B. Rares, *et al.*, "Experimental Evaluation of AoA Algorithms using NI USRP Software Defined Radios", *RoEduNet*, pp. 1-6, 2018, doi: 10.1109/ROEDUNET.2018.8514133.

[22] D. Gallardo, *et al.*, "An Ultra-Wideband Dual Polarization Antenna Array for the Detection and Localization of Bright Fast Radio Transients in the Milky Way", *Journal of Astronomical Instrumentation*, vol. 13, no. 2, 2024, doi: 10.1142/S2251171724500077.

[23] R. Feynman, "Lectures on Physics", Volume 1, Addison-Wesley, 1963.

[24] C. Balanis and P. I. Ioannides, "Introduction to Smart Antennas", Morgan & Claypool, 2007.

[25] M. Shao, P. Zhang and Z. Zhang, ”Realization of digital down conversion in pulse radar receiver”, *2016 IEEE MTT-S International Microwave Workshop Series on Advanced Materials and Processes for RF and THz Applications (IMWS-AMP)*, Chengdu, 2016, pp. 1-4, doi: 10.1109/IMWS-AMP.2016.7588435.

[26] X. Zhou, Z. Xie, F. Wen and J. Yang, ”An Optimal Design and Implementation Method of High Speed Digital Down Conversion Based on FPGA”, *2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI)*, Kunming, China, 2021, pp. 1104-1107, doi: 10.1109/CISAI54367.2021.00221.

[27] F. Curotto, *et al.*, ”Digital calibration test results for Atacama Large Millimeter/submillimeter Array band 7+8 sideband separating receiver”, *Journal of Astronomical Telescopes, Instruments, and Systems*, vol. 8, no. 2, 2022, doi: 10.1117/1.JATIS.8.2.024004.

[28] R. Finger, F. Curotto, R. Fuentes, R. Duan, L. Bronfman, and D. Li, ”A FPGA-based Fast Converging Digital Adaptive Filter for Real-time RFI Mitigation on Ground Based Radio Telescopes”, *Publications of the Astronomical Society of the Pacific*, vol. 130, no. 984, doi: 10.1088/1538-3873/aa972f.

[29] D. Monasterio, S. Jorquera, F. Curotto, C. Espinoza, R. Finger, and L. Bronfman, ”A Proof of Concept Balanced Mixer with the use of a Digital IF Power Combiner to Improve LO Noise Rejection”, *Publications of the Astronomical Society of the Pacific*, vol. 135, no. 1054, doi: 10.1088/1538-3873/ad0789.

[30] A. Oppenheim and R. Schafer, ”Discrete-Time Signal Processing”, Third Edition, Pearson, 2010.

[31] R. Crochiere and L. Rabiner ”Multirate Digital Signal Processing”, Prentice Hall, 1983.

[32] D. Gallardo, D. Monasterio, R. Finger, F. P. Mena and L. Bronfman, “A Compact Metamaterial-Based Antenna for Multiband Phased Array Applications”, *IEEE Trans. Antennas Propag.*, vol. 69, no. 12, pp. 8872-8877, 2021, doi: 10.1109/TAP.2021.3090861.

[33] G. Strang, ”Introduction to Linear Algebra”, Fifth Edition, Wellesley-Cambridge press, 2016.

[34] S. Banerjee and A. Roy, ”Linear Algebra and Matrix Analysis for Statistics”, Fifth Edition, Taylor & Francis Group, 2014.

Appendix A

Linear Algebra

A.1 The Four Fundamental Subspaces of a matrix A

Matrices and their structures are often best understood by analyzing their components: column vectors and row vectors. Indeed, when we examine $\mathbf{Ax}$ or $\mathbf{AX}$, we are essentially observing linear combinations of the column vectors of $\mathbf{A}$. Similarly, when we examine $\mathbf{y}^T\mathbf{A}$ or $\mathbf{YA}$, we are observing linear combinations of the row vectors of $\mathbf{A}$.

There are four fundamental subspaces of a matrix that are related to the matrix's columns and rows, and they provide insight into the matrix's rank, nullity, and the nature of solutions to linear systems. They are defined as follows:

Definition A.1 *The column space of a matrix* $\mathbf{A}$*, denoted as* $\mathcal{C}(\mathbf{A})$*, is the set of all possible linear combinations of its column vectors. That is to say, if* $\mathbf{A}$ *is an* $m \times n$ *complex matrix, we define* $\mathcal{C}(\mathbf{A})$ *as*

$$\mathcal{C}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{C}^m : \mathbf{y} = \mathbf{Ax}, \mathbf{x} \in \mathbb{C}^n\} \tag{A.1}$$

Definition A.2 *The row space of a matrix* $\mathbf{A}$*, denoted as* $\mathcal{R}(\mathbf{A})$*, is the set of all possible linear combinations of its row vectors. That is to say, if* $\mathbf{A}$ *is an* $m \times n$ *complex matrix, we define* $\mathcal{R}(\mathbf{A})$ *as*

$$\mathcal{R}(\mathbf{A}) = \mathcal{C}(\mathbf{A}^T) = \{\mathbf{y} \in \mathbb{C}^n : \mathbf{y} = \mathbf{A}^T\mathbf{x}, \mathbf{x} \in \mathbb{C}^m\} \tag{A.2}$$

Definition A.3 *The null space of a matrix* $\mathbf{A}$*, denoted as* $\mathcal{N}(\mathbf{A})$*, consists of all vectors* $\mathbf{x}$ *such that* $\mathbf{Ax} = \mathbf{0}$*. That is to say, if* $\mathbf{A}$ *is an* $m \times n$ *complex matrix, we define* $\mathcal{N}(\mathbf{A})$ *as*

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{C}^n : \mathbf{A}\mathbf{x} = \mathbf{0}\} \tag{A.3}$$

Definition A.4 *The left null space of a matrix* $\mathbf{A}$*, denoted as* $\mathcal{N}(\mathbf{A}^T)$*, consists of all vectors* $\mathbf{y}$ *such that* $\mathbf{y}^T\mathbf{A} = \mathbf{0}$*. That is to say, if* $\mathbf{A}$ *is an* $m \times n$ *complex matrix, we define* $\mathcal{N}(\mathbf{A}^T)$ *as*

$$\mathcal{N}(\mathbf{A}^T) = \{\mathbf{y} \in \mathbb{C}^m : \mathbf{A}^T\mathbf{y} = \mathbf{0}\} \tag{A.4}$$

As shown in Section 2.3.1, DoA estimation methods are fundamentally based on accurately estimating the column space of a specific matrix $\mathbf{A}$ associated with the antenna array and the DoAs. Given that the four fundamental subspaces of a matrix are interconnected and collectively provide a complete characterization of the matrix $\mathbf{A}$, it is crucial for this thesis to have a solid understanding of all four fundamental subspaces. With this in mind, we list the theorems used throughout this thesis. The proofs of these theorems can be found in [33, 34].

Theorem A.1 *Let* $\mathbf{A}$ *and* $\mathbf{C}$ *be* $m \times n$ *and* $m \times p$ *matrices. Then* $\mathcal{C}(\mathbf{C}) \subseteq \mathcal{C}(\mathbf{A})$ *if and only if* $\mathbf{C} = \mathbf{A}\mathbf{B}$ *for some* $n \times p$ *matrix* $\mathbf{B}$*. If* $\mathbf{B}$ *is an* $n \times n$ *nonsingular matrix, then* $\mathcal{C}(\mathbf{C}) = \mathcal{C}(\mathbf{A})$*.*

Theorem A.2 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $m \times n$ *and* $n \times p$ *matrices. Then* $\mathcal{N}(\mathbf{B}) \subseteq \mathcal{N}(\mathbf{A}\mathbf{B})$*. If* $m = n$ *and* $\mathbf{A}$ *is an* $n \times n$ *nonsingular matrix, then* $\mathcal{N}(\mathbf{B}) = \mathcal{N}(\mathbf{A}\mathbf{B})$*.*

Theorem A.3 *If* $\mathbf{A}$*,* $\mathbf{B}$ *and* $\mathbf{C}$ *are matrices such that* $\mathbf{A}$ *and* $\mathbf{B}$ *have the same number of rows, while* $\mathbf{A}$ *and* $\mathbf{C}$ *have the same number of columns, then*

$$\begin{aligned} &\mathcal{C}([\mathbf{A}\ \mathbf{B}]) = \mathcal{C}(\mathbf{A}) + \mathcal{C}(\mathbf{B}) \\ &\mathcal{N}\left(\begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix}\right) = \mathcal{N}(\mathbf{A}) \cap \mathcal{N}(\mathbf{C}). \end{aligned} \tag{A.5}$$

Theorem A.4 *Let* $\mathbf{A} = \mathbf{C}\mathbf{R}$ *be a rank factorization, where* $\mathbf{A}$ *is an* $m \times n$ *matrix with rank* r*,* $\mathbf{C}$ *is an* $m \times r$ *matrix with full column rank* r*, and* $\mathbf{R}$ *is an* $r \times n$ *matrix with full row rank* r*. Then,* $\mathcal{C}(\mathbf{A}) = \mathcal{C}(\mathbf{C})$*.*

Theorem A.5 *Let* $\mathbf{A}$ *be an* $m \times m$ *nonsingular matrix and* $\mathbf{B}$ *an* $m \times p$ *matrix. Then* $\text{rank}(\mathbf{A}\mathbf{B}) = \text{rank}(\mathbf{B})$*.*

Theorem A.6 *Let* $\mathbf{A}$ *be an* $m \times p$ *matrix and* $\mathbf{B}$ *a* $p \times p$ *nonsingular matrix. Then* $\text{rank}(\mathbf{A}\mathbf{B}) = \text{rank}(\mathbf{A})$*.*

Theorem A.7 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two matrices such that* $\mathbf{AB}$ *is defined. It follows that if* $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank}(\mathbf{A})$*, then* $\mathcal{C}(\mathbf{AB}) = \mathcal{C}(\mathbf{A})$*.*

A.2 Eigenvalues and Eigenvectors

Almost all vectors change directions when multiplied by a $n \times n$ square matrix $\mathbf{A}$. However, specific exceptional vectors maintain their direction when subjected to this transformation. These vectors are called eigenvectors, and they satisfy the equation

$$\mathbf{Ax} = \lambda \mathbf{x}, \tag{A.6}$$

were $\mathbf{x} \neq \mathbf{0}$ is an eigenvector, and λ is a scalar known as the eigenvalue associated with $\mathbf{x}$. Note that if $\lambda \neq 0$, equation (A.6) implies that $\mathbf{x} = (1/\lambda)\mathbf{Ax}$, meaning that $\mathbf{x} \in \mathcal{C}(\mathbf{A})$. On the other hand, if $\lambda = 0$, then $\mathbf{Ax} = \mathbf{0}$, indicating that $\mathbf{x} \in \mathcal{N}(\mathbf{A})$.

Regarding the number of eigenvalues and eigenvectors, observe that (A.6) can be rewritten as

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}. \tag{A.7}$$

This implies that the matrix $(\mathbf{A} - \lambda \mathbf{I})$ must be singular. Since singular matrices have a determinant equal to 0, we obtain the characteristic equation

$$|\mathbf{A} - \lambda \mathbf{I}| = 0. \tag{A.8}$$

The expression $|\mathbf{A} - \lambda \mathbf{I}|$ is a polynomial of degree n in λ (see [33]), which has n roots. These roots correspond to the eigenvalues of $\mathbf{A}$. Therefore, $\mathbf{A}$ has exactly n eigenvalues. However, the number of different eigenvectors can vary depending on the specific structure of the matrix $\mathbf{A}$, and could be less than n.

When there are n linearly independent eigenvectors $\{\mathbf{v}_i\}_{i=0}^{n-1}$ associated with eigenvalues $\{\lambda_i\}_{i=0}^{n-1}$, we have the following relationship:

$$\mathbf{A} \begin{bmatrix} \mathbf{v}_0 & \mathbf{v}_1 & \dots & \mathbf{v}_{n-1} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_0 & \mathbf{v}_1 & \dots & \mathbf{v}_{n-1} \end{bmatrix} \begin{bmatrix} \lambda_0 & 0 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_{n-1} \end{bmatrix}, \tag{A.9}$$

or equivalently,

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}, \tag{A.10}$$

where $\mathbf{V} = \left[\begin{array}{cccc} \mathbf{v}_0 & \mathbf{v}_1 & ... & \mathbf{v}_{n-1} \end{array} \right]$ is the matrix whose columns are the eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix with $(\mathbf{\Lambda})_{ii} = \lambda_i$. When (A.10) holds, the matrix $\mathbf{A}$ is said to be diagonalizable.

Several fundamental theorems regarding eigendecomposition are employed throughout this thesis. Below, we list the main theorems, with their detailed proofs available in [33, 34].

Theorem A.8 *Let $\mathbf{A}$ be an $n \times n$ diagonalizable matrix of rank r. Then, there is an $n \times n$ nonsingular matrix $\mathbf{V}$ and an $r \times r$ diagonal matrix $\mathbf{\Lambda}_r$ with nonzero diagonal elements such that*

$$\mathbf{A} = \mathbf{V} \left[\begin{array}{cc} \mathbf{\Lambda}_r & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{array} \right] \mathbf{V}^{-1}. \tag{A.11}$$

Theorem A.9 *If $\mathbf{A}$ and $\mathbf{C}$ are $n \times n$ similar matrices (i.e., there is an $n \times n$ matrix $\mathbf{B}$ such that $\mathbf{A} = \mathbf{B}^{-1}\mathbf{C}\mathbf{B}$), then $\mathbf{A}$ and $\mathbf{C}$ have the same eigenvalues.*

Theorem A.10 *Every real symmetric matrix $\mathbf{S}$ can be diagonalized:*

$$\mathbf{S} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T, \tag{A.12}$$

where $\mathbf{Q}$ is an orthogonal matrix (i.e., $\mathbf{Q}^T = \mathbf{Q}^{-1}$), and $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\mathbf{S}$. Moreover, all the eigenvalues are real, and the eigenvectors corresponding to different eigenvalues are orthonormal.

If $\mathbf{S}$ is complex but is also Hermitian (i.e., $\mathbf{S} = \mathbf{S}^H$), then all eigenvalues are real and all eigenvectors corresponding to different eigenvalues are orthonormal.

Appendix B

Proof that $\mathcal{A}$ is not a vector space

PROOF. To demonstrate that $\mathbf{a}(\theta_1)$ cannot be expressed as a linear combination of $\mathbf{a}(\theta_2)$ and $\mathbf{a}(\theta_3)$, we proceed as follows.

Assume that there exist constants β and γ such that $\beta\mathbf{a}(\theta_2)+\gamma\mathbf{a}(\theta_3) = \mathbf{a}(\theta_1)$. This implies

$$\begin{aligned}
&\beta\mathbf{a}(\theta_2) + \gamma\mathbf{a}(\theta_3) = \mathbf{a}(\theta_1) \\
&\Rightarrow \beta \begin{bmatrix} 1 \\ e^{-j\frac{2\pi d}{\lambda}\sin(\theta_2)} \\ e^{-j\frac{4\pi d}{\lambda}\sin(\theta_2)} \\ \vdots \\ e^{-j\frac{(P-1)2\pi d}{\lambda}\sin(\theta_2)} \end{bmatrix} + \gamma \begin{bmatrix} 1 \\ e^{-j\frac{2\pi d}{\lambda}\sin(\theta_3)} \\ e^{-j\frac{4\pi d}{\lambda}\sin(\theta_3)} \\ \vdots \\ e^{-j\frac{(P-1)2\pi d}{\lambda}\sin(\theta_3)} \end{bmatrix} = \begin{bmatrix} 1 \\ e^{-j\frac{2\pi d}{\lambda}\sin(\theta_1)} \\ e^{-j\frac{4\pi d}{\lambda}\sin(\theta_1)} \\ \vdots \\ e^{-j\frac{(P-1)2\pi d}{\lambda}\sin(\theta_1)} \end{bmatrix} \qquad \text{(B.1)} \\
&\Leftrightarrow \begin{bmatrix} 1 & 1 \\ e^{-j\frac{2\pi d}{\lambda}\sin(\theta_2)} & e^{-j\frac{2\pi d}{\lambda}\sin(\theta_3)} \\ e^{-j\frac{4\pi d}{\lambda}\sin(\theta_2)} & e^{-j\frac{4\pi d}{\lambda}\sin(\theta_3)} \\ \vdots & \vdots \\ e^{-j\frac{(P-1)2\pi d}{\lambda}\sin(\theta_2)} & e^{-j\frac{(P-1)2\pi d}{\lambda}\sin(\theta_3)} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} 1 \\ e^{-j\frac{2\pi d}{\lambda}\sin(\theta_1)} \\ e^{-j\frac{4\pi d}{\lambda}\sin(\theta_1)} \\ \vdots \\ e^{-j\frac{(P-1)2\pi d}{\lambda}\sin(\theta_1)} \end{bmatrix}
\end{aligned}$$

Let us define $a = e^{-j\frac{2\pi d}{\lambda}\sin(\theta_1)}$, $b = e^{-j\frac{2\pi d}{\lambda}\sin(\theta_2)}$, and $c = e^{-j\frac{2\pi d}{\lambda}\sin(\theta_3)}$. Then, substituting into (B.1), we obtain the following system:

$$\begin{bmatrix} 1 & 1 \\ b & c \\ b^2 & c^2 \\ \vdots & \vdots \\ b^{P-1} & c^{P-1} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} 1 \\ a \\ a^2 \\ \vdots \\ a^{P-1} \end{bmatrix} \qquad \text{(B.2)}$$

Now, focusing on the first four rows of the matrix and performing Gaussian elimination, we aim to simplify the system. Through elimination, we arrive at the following reduced system:

$$\begin{aligned}
&\begin{bmatrix} 1 & 1 \\ 0 & c-b \\ 0 & c^2-b^2 \\ 0 & c^3-b^3 \end{bmatrix}\begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} 1 \\ a-b \\ a^2-b^2 \\ a^3-b^3 \end{bmatrix} \\
\Rightarrow &\begin{bmatrix} 1 & 1 \\ 0 & c-b \\ 0 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \beta \\ \gamma \end{bmatrix} \\
= &\begin{bmatrix} 1 \\ a-b \\ a^2-b^2-(a-b)(c+b) \\ a^3-b^3-(a-b)(b^2+bc+c^2) \end{bmatrix} = \begin{bmatrix} 1 \\ a-b \\ (a-b)(a+2b+c) \\ (a-b)(a^2+ab+b^2)-(a-b)(b^2+bc+c^2) \end{bmatrix} \\
= &\begin{bmatrix} 1 \\ a-b \\ a^2-b^2-(a-b)(c+b) \\ a^3-b^3-(a-b)(b^2+bc+c^2) \end{bmatrix} = \begin{bmatrix} 1 \\ a-b \\ (a-b)(a+2b+c) \\ (a-c)(a+b+c) \end{bmatrix}
\end{aligned} \tag{B.3}$$

From the last two equations of (B.3), it must be true that

$$\begin{bmatrix} (a-b)(a+2b+c)=0 \\ (a-c)(a+b+c) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{B.4}$$

Finally, since the antenna array is constructed without phase ambiguities ($d < \lambda/2$), it is impossible for $a-b=0$ or $a-c=0$. From this, it follows that $a+2b+c=0$ and $a+b+c=0$. Solving these equations leads to the conclusion that $b=0$. which is a contradiction because $b = e^{-j\frac{2\pi d}{\lambda}\sin(\theta_2)} \neq 0$. This contradiction demonstrates that there are no constants β and γ such that $\beta\mathbf{a}(\theta_2) + \gamma\mathbf{a}(\theta_3) = \mathbf{a}(\theta_1)$, reinforcing the fact that the array manifold $\mathcal{A}$ is not a vector space.

□

Appendix C

Radiation Pattern Effect in Narrowband Signals

Let us consider one of the r signals of Section 3.1 impinging in the (p,q)-th antenna, as written in (C.1):

$$x_{pq}(t) \approx u(t)\cos\left(\omega_0(t-\tau_{pq})+v(t)\right), \tag{C.1}$$

where τ_{pq} is the propagation delay between the (p,q)-th antenna and the array's center. The Fourier transform of $x_{pq}(t)$ is given by

$$\begin{aligned} \mathbf{X}_{pq}(\omega) &= \frac{1}{2}\int_{-\infty}^{\infty} u(t)(e^{i(\omega_0 t-\omega_0\tau_{pq}+v(t))}+e^{-i(\omega_0 t-\omega_0\tau_{pq}+v(t))})e^{-i\omega t}dt \\ &= \frac{1}{2}e^{-i\omega_0\tau_{pq}}\int_{-\infty}^{\infty}(u(t)e^{iv(t)})e^{-i(\omega-\omega_0)t}dt+\frac{1}{2}e^{i\omega_0\tau_{pq}}\int_{-\infty}^{\infty}(u(t)e^{-iv(t)})e^{-i(\omega+\omega_0)t}dt. \end{aligned} \tag{C.2}$$

If we define $\mathbf{U}(\omega)$ as the Fourier transform of $u(t)e^{iv(t)}$, we can express (C.2) as

$$\mathbf{X}_{pq}(\omega) = \frac{1}{2}e^{-i\omega_0\tau_{pq}}\mathbf{U}(\omega-\omega_0)+\frac{1}{2}e^{i\omega_0\tau_{pq}}\mathbf{U}^*(-(\omega+\omega_0)). \tag{C.3}$$

Considering that for the direction of arrival of the signal the radiation pattern of all antennas has the same response $\Gamma(\omega)e^{i\varphi(\omega)}$, we can write the Fourier transform of the signal received by the (p,q)-th antenna as

$$\mathbf{Y}_{pq}(\omega) = \mathbf{X}_{pq}(\omega)\Gamma(\omega)e^{i\varphi(\omega)}, \tag{C.4}$$

where $\Gamma(\omega)$ is an even function and $\varphi(\omega)$ is an odd function (since the radiation pattern is a physical phenomenon and hence a real function). With all this, the signal received by the (p,q)-th antenna is given by

$$\begin{aligned} y_{pq}(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{X}_{pq}(\omega)\Gamma(\omega)e^{i\varphi(\omega)}e^{i\omega t}d\omega \\ &= \frac{1}{2}e^{-i\omega_0\tau_{pq}} \times \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{U}(\omega-\omega_0)\Gamma(\omega)e^{i\varphi(\omega)}e^{i\omega t}d\omega \\ &+ \frac{1}{2}e^{i\omega_0\tau_{pq}} \times \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbf{U}^*(-(\omega+\omega_0))\Gamma(\omega)e^{i\varphi(\omega)}e^{i\omega t}d\omega \end{aligned} \tag{C.5}$$

Since the signals we consider in this development are narrowband, we can approximate $\mathbf{U}(\omega) \approx 0$ outside the range $\omega \in [-\Delta, \Delta]$. Therefore, the limits of the integrals of (C.5) change to $\pm\omega_0 \pm \Delta$, giving

$$\begin{aligned} y_{pq}(t) &= \frac{1}{2}e^{-i\omega_0\tau_{pq}} \times \frac{1}{2\pi}\int_{\omega_0-\Delta}^{\omega_0+\Delta} \mathbf{U}(\omega-\omega_0)\Gamma(\omega)e^{i\varphi(\omega)}e^{i\omega t}d\omega \\ &+ \frac{1}{2}e^{i\omega_0\tau_{pq}} \times \frac{1}{2\pi}\int_{-\omega_0-\Delta}^{-\omega_0+\Delta} \mathbf{U}^*(-(\omega+\omega_0))\Gamma(\omega)e^{i\varphi(\omega)}e^{i\omega t}d\omega \end{aligned} \tag{C.6}$$

Finally, if we consider the change of variables $\omega' = -\omega$ in the second integral and remember that $\Gamma(\omega) = \Gamma(-\omega)$ and $\varphi(\omega) = -\varphi(-\omega)$, we obtain

$$\begin{aligned} y_{pq}(t) &= \frac{1}{2}e^{-i\omega_0\tau_{pq}} \times \frac{1}{2\pi}\int_{\omega_0-\Delta}^{\omega_0+\Delta} \mathbf{U}(\omega-\omega_0)\Gamma(\omega)e^{i\varphi(\omega)}e^{i\omega t}d\omega \\ &+ \frac{1}{2}e^{i\omega_0\tau_{pq}} \times \frac{1}{2\pi}\int_{\omega_0-\Delta}^{\omega_0+\Delta} \mathbf{U}^*(\omega'-\omega_0)\underbrace{\Gamma(-\omega')}_{\Gamma(\omega')}\overbrace{e^{i\varphi(-\omega')}}^{e^{-i\varphi(\omega')}}e^{-i\omega' t}d\omega' \\ &= \frac{1}{2}e^{-i\omega_0\tau_{pq}}\xi(t) + \frac{1}{2}e^{i\omega_0\tau_{pq}}\xi^*(t) \\ &= \mathbb{R}(e^{-i\omega_0\tau_{pq}}\xi(t)) \\ &= \mathbb{R}(e^{-i\omega_0\tau_{pq}}\rho(t)e^{i\sigma(t)}) \\ &= \rho(t)\cos(\sigma(t)-\omega_0\tau_{pq}), \end{aligned} \tag{C.7}$$

where $\xi(t)$ is a narrowband signal centered around ω_0.

Appendix D

Receiver Effect in Narrowband Signals

Upon entering the receiver, as depicted in Figure 4.1, each signal $y_{pq}(t)$ undergoes the same amplification, mixing, and filtering chain. However, due to potential irregularities arising from imperfections in the components, solder joints, and the relative positions of each component with respect to other chains and the metal walls of the assembly, differences in amplitude and phase must be considered at the end of each chain.

To develop the mathematical model, we assume that the first capacitor and the amplifiers do nothing more than introduce changes in amplitude and phase in all the chains (ideally, the same for all). Then, following the same procedure shown in Appendix C, we can write the signal after the first capacitor and amplifier as

$$\bar{y}_{pq}(t) = \bar{\rho}_{pq}(t) \cos\left(\bar{\sigma}_{pq}(t) - \omega_0 \tau_{pq}\right), \tag{D.1}$$

where $\bar{\rho}_{pq}(t)$ and $\bar{\sigma}_{pq}(t)$ are analogous to the functions derived in Appendix C, but now vary for each receiver chain due to amplitude and phase imbalances.

When the signals reach the mixers, each $\bar{y}_{pq}(t)$ is multiplied by $\cos\left(\omega_{\mathrm{LO}} t\right)$ (assuming an ideal mixer), where ω_{LO} is the local oscillator angular frequency. Hence, the output of the (p, q)-th mixer is

$$\begin{aligned}
\bar{\bar{y}}_{pq}(t) &= \bar{y}_{pq}(t) \cos\left(\omega_{\mathrm{LO}} t\right) \\
&= \bar{\rho}_{pq}(t) \cos\left(\bar{\sigma}_{pq}(t) - \omega_0 \tau_{pq}\right) \cos\left(\omega_{\mathrm{LO}} t\right) \\
&= \frac{\bar{\rho}_{pq}(t)}{2} \left(\cos\left(\bar{\sigma}_{pq}(t) + \omega_{\mathrm{LO}} t - \omega_0 \tau_{pq}\right) + \cos\left(\bar{\sigma}_{pq}(t) - \omega_{\mathrm{LO}} t - \omega_0 \tau_{pq}\right)\right)
\end{aligned} \tag{D.2}$$

and after passing through the low-pass filter, we can write

$$z_{pq}(t) = \bar{\rho}_{pq}(t) \cos\left(\bar{\sigma}_{pq}(t) - \omega_{\mathrm{LO}} t - \omega_0 \tau_{pq}\right), \tag{D.3}$$

where $\bar{\rho}_{pq}(t)$ and $\bar{\sigma}_{pq}(t)$ have absorbed the $\frac{1}{2}$ factor, along with any amplitude and phase imbalances.

Finally, notice that, since $\bar{\sigma}_{pq}(t)$ is centered around the carrier frequency ω_0, we have that $\bar{\sigma}_{pq}(t)/t - \omega_{\mathrm{LO}} \geq 0$ if the LO is set in such a way that the upper sideband is taken, and $\bar{\sigma}_{pq}(t)/t - \omega_{\mathrm{LO}} \leq 0$ if the lower sideband is considered [1]. Hence, we can write

$$z_{pq}(t) = \begin{cases} \bar{\rho}_{pq}(t)\cos\left(|\bar{\sigma}_{pq}(t)/t - \omega_{\mathrm{LO}}|t - \omega_0\tau_{pq}\right) & \text{, if the upper sideband is considered} \\ \bar{\rho}_{pq}(t)\cos\left(|\bar{\sigma}_{pq}(t)/t - \omega_{\mathrm{LO}}|t + \omega_0\tau_{pq}\right) & \text{, if the lower sideband is considered} \end{cases} \tag{D.4}$$

Alternatively, we can rewrite this expression as

$$z_{pq}(t) = \begin{cases} \frac{1}{2}e^{-i\omega_0\tau_{pq}}\chi_{pq}(t) + \frac{1}{2}e^{i\omega_0\tau_{pq}}\chi_{pq}^*(t) & \text{, if the upper sideband is considered} \\ \frac{1}{2}e^{i\omega_0\tau_{pq}}\chi_{pq}(t) + \frac{1}{2}e^{-i\omega_0\tau_{pq}}\chi_{pq}^*(t) & \text{, if the lower sideband is considered} \end{cases} \tag{D.5}$$

where χ_{pq} represent signals centered around $|\omega_0 - \omega_{\mathrm{LO}}|$.

---

[1] This argument should be interpreted intuitively and not analytically. The argument is only strictly valid when $\bar{\sigma}_{pq}(t) = \omega_0 t$.