TESIS DOCTORAL

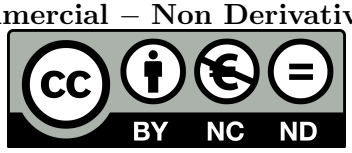
Initial access and beam-steering mechanisms for mmWave wireless systems

Autor: Joan Palacios Beltran, IMDEA Networks Institute
University Carlos III de Madrid
Director \Tutor: Joerg Widmer, IMDEA Networks Institute

DEPARTMENT OF SIGNAL AND COMMUNICATIONS THEORY

Universidad Carlos III de Madrid

Leganés (Madrid), Julio de 2020



Initial access and beam-steering mechanisms for mmWave wireless systems

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Prepared by
Joan Palacios Beltran, IMDEA Networks Institute, University Carlos III of Madrid

Under the advice of
Joerg Widmer, IMDEA Networks Institute
Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid

---

Date: Julio, 2020

Web/contact: joan.palacios@imdea.org

This work has been supported by IMDEA Networks Institute.

Acknowledgements

Completing a PhD is not a one person's work. I want to state that it would have been impossible to realize without the guidance of my supervisor Joerg Widmer, the collaboration of all my colleagues whether they contributed directly with my research or simply debated, and everyone around me for providing with emotional support. I want to emphasise that this thesis would not exist without any of these 3 factors and thus I want to thank everyone that is or has ever been involved in my work or personal life. Thank you.

Published and submitted content

The research work in this thesis has been published and accepted in several peer-reviewed conferences and workshops. We list in details all the publications that are included in this thesis and their corresponding chapters:

1. **J. Palacios**, D. De Donno and J. Widmer, "Lightweight and Effective Sector Beam Pattern Synthesis With Uniform Linear Antenna Arrays," in IEEE Antennas and Wireless Propagation Letters, vol. 16, pp. 605-608, 2017, doi: 10.1109/LAWP.2016.2594092.

    - I did all the formulation and implementation of the method designed in the paper.
    - The whole publication is included in chapter 2.
    - The material from this source included in this thesis is not singled out withtypographic means and references.

2. **J. Palacios**, D. De Donno, D. Giustiniano and J. Widmer, "Speeding up mmWave beam training through low-complexity hybrid transceivers," 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, 2016, pp. 1-7, doi: 10.1109/PIMRC.2016.7794709.

    - I did all the formulation and implementation of the method designed in the paper and contributed to the simulation.
    - The whole publication is included in chapter 3.
    - The material from this source included in this thesis is not singled out withtypographic means and references.

3. **J. Palacios**, D. De Donno and J. Widmer, "Tracking mm-Wave channel dynamics: Fast beam training strategies under mobility," IEEE INFOCOM 2017 - IEEE Conference on Computer Communications, Atlanta, GA, 2017, pp. 1-9, doi: 10.1109/INFOCOM.2017.8056991.

- I did all the formulation and implementation in the paper and contributed to the simulation.
- The whole publication is included in chapter 4.
- The material from this source included in this thesis is not singled out withtypographic means and references.

4. **J. Palacios**, D. Steinmetzer, A. Loch, M. Hollick, and J. Widmer. 2018. "Adaptive Codebook Optimization for Beam Training on Off-the-Shelf IEEE 802.11ad Devices." In Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom 18). Association for Computing Machinery, New York, NY, USA, 241-255. DOI:https://doi.org/10.1145/3241539.3241576

   - I did all the formulation and implementation of the method designed in the paper and the experiments.
   - The whole publication is included in chapter 5.
   - The material from this source included in this thesis is not singled out withtypographic means and references.

5. **J. Palacios**, P. Casari and J. Widmer, "JADE: Zero-knowledge device localization and environment mapping for millimeter wave systems," IEEE INFOCOM 2017 - IEEE Conference on Computer Communications, Atlanta, GA, 2017, pp. 1-9, doi: 10.1109/INFOCOM.2017.8057183.

   - I did all the formulation and implementation and simulations in the paper.
   - The entire publication except for the room walls estimation is included in chapter 6.
   - The material from this source included in this thesis is not singled out withtypographic means and references.

6. **J. Palacios**, G. Bielsa, P. Casari and J. Widmer, "Communication-Driven Localization and Mapping for Millimeter Wave Networks," IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, Honolulu, HI, 2018, pp. 2402-2410, doi: 10.1109/INFOCOM.2018.8485819.

   - I did all the formulation and implementation in the paper.
   - The entire publication except for the room walls estimation is included in chapter 6.
   - The material from this source included in this thesis is not singled out withtypographic means and references.

7. **J. Palacios**, P. Casari, H. Assasa and J. Widmer, "LEAP: Location Estimation and Predictive Handover with Consumer-Grade mmWave Devices," IEEE INFOCOM 2019 - IEEE Conference on Computer Communications, Paris, France, 2019, pp. 2377-2385, doi: 10.1109/INFOCOM.2019.8737434.

    - I did all the formulation, implementation and experiments in the paper.
    - The whole publication is included in chapter 7.
    - The material from this source included in this thesis is not singled out withtypographic means and references.

Abstract

Future millimeter-wave networks will support very high densities of devices and access points. This vastly increases the overhead required for access point selection and beam training. Due to unfavorable radio propagation, mmWave systems will exploit large-scale MIMO and adaptive antenna arrays at both the transmitter and receiver to realize sufficient link margin. Beamforming is vital to overcome the high attenuation in wireless millimeter-wave networks. It enables nodes to steer their antennas in the direction of communication. Fortunately, the quasi-optical properties of millimeter-wave channels make location-based network optimization a highly promising technique to reduce control overhead in such millimeter-wave WLANs.

In this thesis we present tools to improve mmWave systems. We start by designing an effective lightweight sector beam-pattern design for using as a baseline for hybrid analog-digital structures. We deal with practical constraints of mmWave transceivers and propose a novel, geometric approach to synthesize multi-beamwidth beam patterns that can be leveraged for simultaneous multi-direction scanning. Then we make use of this multi-direction scanning to create a beam training protocol which effectively accelerates the link establishment by exploiting the ability of mobile users to simultaneously receive from multiple directions. We propose smart beam training and tracking strategies for fast mm-wave link establishment and maintenance under node mobility. We leverage the ability of hybrid analog-digital transceivers to collect channel information from multiple spatial directions simultaneously and formulate a probabilistic optimization problem to model the temporal evolution of the mm-wave channel under mobility. We propose a mechanism to extract full channel state information (CSI) regarding phase and magnitude from coarse signal strength readings on off-the-shelf IEEE 802.11ad devices. Using this CSI, transmitters dynamically compute a transmit beam pattern that maximizes the signal strength at the receiver. Channel properties and antenna design at 60GHz are ideal for path angular information extraction, following an almost ideal geometric channel model. Due to this, we present some localization method specifically designed for the 60GHz band. We merge the ideas presented in this thesis and by extracting channel state information from off-the-shelf routers we estimate the user location to manage a location aware beam-training and device handling method. The resulting scheme can

predict blockage, optimize access point association, and select the most suitable antenna beam patterns while significantly reducing the beam training overhead.

Table of Contents

List of Tables

List of Figures

Part I

Introduction

Chapter 1

Introduction

1.1. Introduction

The fifth generation of mobile communications (5G) is envisaged to deliver multi-Gbps wireless connectivity and to enable a plethora of new applications. It is well established that achieving extremely high data rates is impractical with currently available 4G systems due to the heavily congested and fragmented spectrum below 6 GHz. In view of this, the large amount of unoccupied spectrum in the millimeter wave (mm-wave) bands above 6 GHz becomes very appealing [2]. Millimeter wave (mmWave) wireless technologies are expected to become key enablers of multi-gigabit wireless access in next-generation cellular and local area networks.

Communications at mm-wave frequencies are challenging since the channel suffers from severe path loss, atmospheric absorption, human blockage, and other environmental obstructions [3]. The short wavelength of the mm-waves allows beamforming arrays with many antennas to be implemented in a small form factor, thus providing sufficient link margin. On the other hand, highly directional communications complicate the link establishment and maintenance between an Access Point (AP) and a User Equipment (UE). In fact, AP and UE must perform a time-consuming beam training procedure in order to determine the best directions of transmission/reception, which incurs significant overhead (and waste of network resources). The problem is exacerbated in scenarios with mobility, since even a slight beam misalignment or environmental changes, such as link blockage, device rotation, etc., can cause considerable signal drop. To sum up, fast and efficient beam training/tracking strategies are of paramount importance to maintain seamless connectivity in a mm-wave network with node mobility.

Designing wireless communication systems that operate at very high frequencies such as the 60 GHz band is technically challenging. For instance, phased antenna arrays are hard to manufacture [4, 5] and phase noise plays a significant role [6, 7]. As a result, system designers often resort to simple yet sub-optimal solutions. This is the case for

certain mechanisms such as analog beamforming in the IEEE 802.11ad standard for communication at 60 GHz. Beamforming is crucial for such systems since devices must use directional communication to overcome the very high path loss in the millimeter-wave band and reach the receiver. However, the standard is limited to a basic beamforming mechanism based on a codebook of generic beam patterns. Such patterns are envisioned to have the shape of uniform sectors and cover the entire azimuth range of the device. That is, instead of beamforming towards a specific direction, devices choose the beam pattern out of their codebook which provides the highest gain in that direction.

Beam training, also called **sls!** (**sls!**) in infrastructure-based 802.11ad networks, works as follows. The Access Point (AP) transmits beacon messages using each of its available beam patterns sequentially, while the station (STA) listens with a quasi-omnidirectional beam pattern. After that, the STA repeats the same process but includes in each of its messages the identifier of the beam pattern that it received best from the AP. Finally, the AP replies with the identifier of the best beam pattern of the STA in a dedicated control message. While this mechanism is straightforward, it clearly does not exploit the full potential of the antenna array of IEEE 802.11ad devices. First, none of the available beam patterns in the codebook may steer exactly towards the receiver. Second, strong reflectors in the environment may remain unused even though they could contribute to the received signal strength. Third, existing reflections may result in destructive interference at the receiver, causing significant harm to the communication. If devices were to adapt their beam patterns to the specific environment in which they operate, they could easily mitigate the above issues.

The design space of beam search proposals in the literature can be divided into three main categories: (1) sequential scanning strategies [8,9]; (2) adaptive algorithms employing antenna patterns with configurable beamwidth [10–13]; (3) parallel beam search with simultaneous, multi-direction scanning [14, 15]. The vast majority of these works concentrates, however, on static networks without investigating the impact of the training latency on the overall Quality of Service (QoS) of realistic networks with mobility.Within the state-of-the-art solutions on this subject, a further subdivision can be made on the basis of the employed mm-wave transceivers. Since traditional multiple-input multiple-output (MIMO) digital beamforming (DBF) is, at present, impractical at mm-wave frequencies because of cost and power consumption constraints, analog beamforming (ABF) and hybrid analog-digital beamforming (HBF) represent the only feasible solutions. Using ABF [8,9] provides poor performance for two main reasons. First, the constant amplitude and the low phase resolution constraints of the mm-wave RF phase shifters [16] give rise to antenna sectors with high sidelobes and reduced flatness, leading to imprecise beam training. Second, the use of a single RF chain allows for only one communication beam, thus resulting in reduced throughput and high-overhead beam search. In HBF architectures [10–15], the precoding/combining operations are divided between the analog and

digital domains, while using much fewer RF chains than antenna elements. The availability of multiple RF chains enables parallel, multi-stream processing and simultaneous multi-direction scanning.

However, this requires full Channel State Information (CSI) at the transmitter, which is particularly challenging to obtain in millimeter-wave systems. The reasons are twofold. First, the feedback overhead of full CSI is significant. Second, devices must probe enough orthogonal beam patterns to cover all of the dimensions of the channel. The channel information extracted from all of the probes must be coherent in phase. In particular, achieving such phase coherence over different probes in consumer-grade devices is non-trivial.

The goal of this work was to create the needed tools to improve mmWave communications. We explore many different ways to do this, from beam pattern synthesis to location aided beam pattern selection. There are many things to consider for mmWave as its quasi optical characteristics make it very different from previous bands.

First thing to consider is its hardware structure, the fact that it can easily scale with the number of antenna elements but can't deploy an **rf!** (**rf!**) chain for each of them makes mandatory the beam pattern synthesis by use of complex weights on each antenna element. Along with this, the hybrid architecture is highly restricted in terms of design freedom and it's not mathematically easy to take these into account. Due to this, the way we attack this problem is by first designing a fully digital baseline beam-pattern and then approach it for a hybrid architecture. Chapters 2 and 3 take care of this beam pattern creation, from its design in digital domain to its adaptation to the hybrid domain.

Once the beam pattern creation has been established the next step is to decide what beam-pattern is suitable for communication, this is commonly called as beam training for scenarios without memory and beam tracking for scenarios with past information available. The idea is to design beam patterns such that by measuring them we can decide a beam pattern that's going to be suitable for communication while not increasing the communication latency by expending too long in the beam training. Chapter 4 tackles beam-training and beam-tracking for communication with and without mobility.

The next logical step in beam training design comes from the analysis of commercial hardware and its aim to adapt to its capabilities, this is why we create a channel measuring technique that's valid for beam training and beam tracking. Chapter 5 analyzes and implements a valid channel measurement for beam-training and beam-tracking method for hardware limitations of first generation 802.11ad devices.

One of the good things about mmWave that makes it so interesting for communication is its sparsity which comes from its quasi optical geometrical behavior. This behavior allow us to extract angle information about the devices localization that we will be able to use later to improve communication by throwing this information into the beam training and tracking. We look into localization algorithms for mmWave and most concretely about

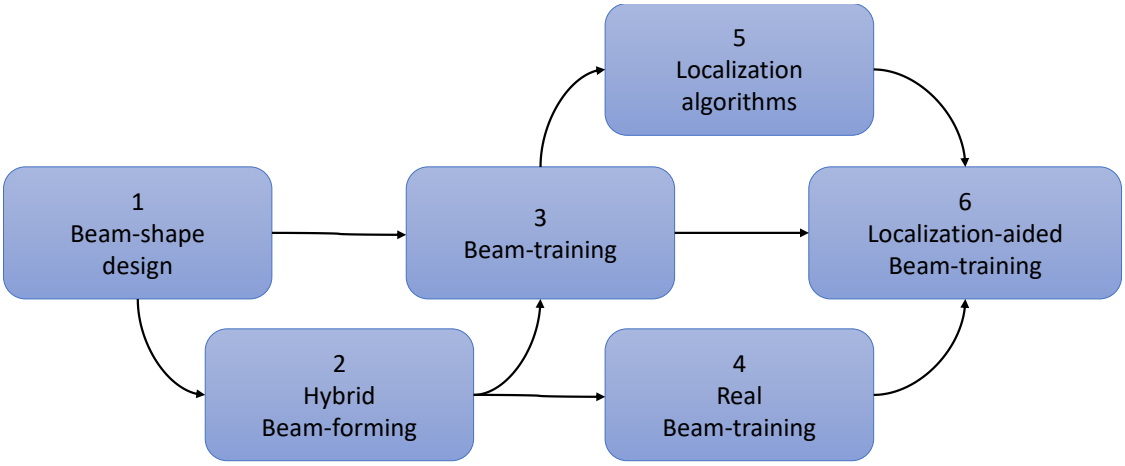


Figure 1.1: Diagram indicating the thesis structure.

zero knowledge localization algorithms, these are algorithms that require no information input about the room the devices are located in but instead are able to reconstruct the room geometry from history angle information. Chapter 6 introduces localization and zero knowledge localization with SLAM.

The final step in this thesis is mixing both zero knowledge localization with real hardware and beam training and tracking to create an algorithm that through location estimation can predict the movement and channel evolution to increase the channel information and not only reduce its training overhead but also adapt to channel changes like occlusion before they happen by use of machine learning techniques. Chapter 7 takes care of merging beam training and tracking method that goes in parallel with a localization algorithm to improve communication performance.

1.2. List of papers

This thesis is based in the following papers.

- Chapter 2 is my work on baseline beam-pattern designs published in IEEE Antennas and Wireless Propagation Letters [17]. Our contribution was the design of a low complexity closed-form solution to the sector beam-pattern problem. My contribution to this paper was the whole formulation of the solution proposed and its implementation.

- Chapter 3 is based on my paper on Hybrid Beam-forming published in IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications [18]. Our main contribution here was to develop a fast low complexity algorithm to determine the digital and analog codig/precoding matrices of a Hybrid Analog-Digital phased array antenna architecture. This one was based on geometric

properties of the low resolution coefficients instead of relying on the non-scalable cost heavy orthogonal matching pursuit that the literature was based on and it allowed for the design of parallel beam-patterns for reception. Another contribution was the extension of an already known beam-training technique called adaptive search to use this feature to perform multiple parallel measures to reduce the training overhead. My contribution to this paper was the proposed hybrid beam-forming method its implementation and the adaptive beam-training extension to multiple parallel measures.

- Chapter 4 contains my paper on beam-tracking published in IEEE INFOCOM 2017-IEEE Conference on Computer Communications [19]. This paper's contribution was a beam-tracking algorithm that uses Bayesian estimation over channel parameters to track their evolution. This algorithm was designed to make use of the signal at the converters to avoid loosing information about the channel evolution due to signal dimensionality when combining the signals in the digital domain. By doing this we were able to reduce the beam-training overhead while increasing the channel estimation accuracy and the communication throughput. My contribution to this paper was the whole formulation and implementation of both beam-training and tracking methods.

- Chapter 5 is my work in beam-training under restricted measurements in commercial of the shelf devices and their exploitability in experiments requiring complex information of antenna coefficients. This was published in Proceedings of the 24th ACM Annual International Conference on Mobile Computing and Networking [20]. My contributions to this paper were mainly two. The first one was a tool to extract per antenna channel coefficient information including not only amplitude but phase. This is achieved by measuring the signal strength while modifying the codebook that defines the beam-patterns. The second contribution was designing a beam-training method for commercial of the shelf devices that only have this information accessible. My contribution to this paper was the whole formulation and implementation the channel estimation and beam-training algorithms.

- Chapter 6 is a condensed version of two of my localization papers published in IEEE INFOCOM [21] and [22]. Both papers' contributions are the design of a zero-knowledge simultaneous location and mapping algorithms that make use of angle difference information for 60GHz communications. The first one is based on an algebra relaxation through inversive geometry. The second one first develops a tool to unsupervisedly find mathematical relations in problems to find a relationship between the measured angle differences and the anchors' (access points and their specular reflections) relative position. My contribution to both papers was the whole formulation and implementation the location algorithms and the environment

simulation in [21].

- Chapter 7 has my paper on location aided beam-training for commercial of the shelf devices published IEEE INFOCOM [23]. This paper's contribution consists on condensing the tools from previous papers into a single one. We develop a novel location system for commercial of the shelf devices and use its output to make predictions on the channel evolution to have a more efficient beam-training algorithm by keeping track of these and predicting future changes through machine learning techniques. My contribution to this papers was the whole formulation and implementation of all the algorithms and the realization of the experiments.

Apart from these seven papers I published other ones that I didn't include in this thesis. These are [24] [25] [26] [27] [28] [29] [30] [31] and [32].

1.3. Notation

During the whole thesis we will be using a common notation and reserved symbols for some variables to make its reading easier. The notation we will use is

- $x, \mathbf{x}, \mathbf{X}$ are gonna be the styles for scalar, vector and matrix and they will be indexed as $[\mathbf{x}]_a, [\mathbf{x}]_{a,b}, [\mathbf{x}]_{:,b}, [\mathbf{x}]_{a,:}$ for vector element and matrix element, column and row. Sets of index can be used to denote sub-matrices. Unless specified, every vector is going to be considered as a column vector for matrix operations.

- in a similar way a, a will be the styles for quantitative and qualitative index.

- $\mathcal{R}(\cdot), \mathcal{I}(\cdot), (\cdot)^*, (\cdot)^{\mathrm{T}}, (\cdot)^{\mathrm{H}}, |\cdot|, \|\cdot\|, \arg(\cdot)$ will denote the real part, imaginary par, complex conjugate, transpose, Hermitian, module, norm and angle argument operations. The default norms we are considering are the Euclidean norm for vectors and the Frobenius norm for matrices.

- $\mathbb{R}, \mathbb{C}, \mathbb{T}$ are the real, imaginary and complex torus sets. $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$

- $\mathbb{E}[\cdot]$ denotes the expectation operator

- $\lceil \cdot \rceil$ denote the ceiling operation

As for reserved symbols

- $\mathbf{c}, \mathbf{p}$ will be used for the combiner and precoder, together with the subscripts RF, BB for RF and BB components whenever is needed.

- θ, ϕ denote AoA and AoD, φ will also be used at some points to denote AoA or AoD. Note that AoA and AoD for a ULA can have two different scales, the physical and the transformed by $\frac{2\pi d}{\lambda}\sin(\vartheta)$, the second one is normally used for its isotropic radiation properties.

- $\mathbf{I}$ is the identity matrix.

Part II

Designing the antenna radiation properties

Chapter 2

Beam shape

2.1. Challenges

The synthesis of sector beam patterns with antenna arrays is a widely investigated topic in the literature because of its myriad of applications, ranging from massive multiple-input multiple-output (MIMO) [33] to cell sectorization in cellular networks [34]. Very recently, the design of sector beam patterns has also received significant attention in the millimeter-wave (mmWave) context where the use of high-gain, adaptive antenna arrays with configurable beamwidth is essential to cope with the higher propagation loss and unfavorable atmospheric absorption at mmWave frequencies [35, 36].

2.2. Related work

Several approaches for sector beam synthesis have been proposed in the literature [37, 38], ranging from simple, classical formulations, e.g., Fourier transform, Woodward-Lawson frequency sampling, etc., derived as extensions of digital filter design techniques [39, Ch. 21] to more sophisticated methods such as genetic algorithms [40], particle swarm [41], and invasive weed [42]. One of the main problems of these techniques is their inherent complexity, which makes it more difficult for researchers and practitioners to implement and apply them. Secondly, since these techniques are quite computationally intensive, they may require many iterations until they convergence to the optimum solution. This last aspect is crucial, for example, in cognitive radios where the antenna configuration needs to be programmed dynamically based on the propagation environment and link performance to deliver the required quality of service [43]. In such situations, the availability of lightweight, effective, and fast algorithms for beam pattern synthesis is essential, especially in the case of resource-constrained, battery-powered wireless mobile nodes.

2.3. Contribution

In this chapter, we propose a new technique for the synthesis of sector beam patterns with uniform linear arrays (ULAs). The intuition behind the technique is that the problem of synthesizing an array factor with symmetrical, low-pass magnitude response can be significantly simplified by employing real instead of complex array weights. This simplification leads to a lightweight mathematical formulation whose solution yields a compact, closed-form expression to derive the antenna coefficients as a function only of the desired beamwidth and steering direction. The proposed strategy is more effective than classical approaches, not only in terms of ability to shape sector beam patterns which better comply with a desired mask, but also in terms of reduced computational complexity.

The chapter is organized as follows. In Section 2.4, the mathematical formulation for the proposed beam pattern synthesis technique is presented. Numerical assessment and comparisons with classical state-of-the-art synthesis techniques are provided in Section 2.5. Finally, the main conclusions are drawn in Section 2.6.

2.4. Formulation

We consider a wireless device equipped with a ULA of M isotropic antenna elements, inter-element distance d, and operating wavelength λ. We focus on the digital wavenumber domain $\psi{=}(2\pi d/\lambda)\sin\phi$ referred to the azimuthal angle ϕ and consider the design of an ideal low-pass array factor centered at $\psi_0{=}0$ with beamwidth ψ_b. As shown in Fig. 2.1, such an ideal response is defined over $-\pi \leq \psi \leq \pi$ by the following function:

$$\Pi(\psi) = \begin{cases} 1 & -\psi_\mathrm{b}/2 < \psi < \psi_\mathrm{b}/2 \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

In Fig. 2.1, we highlight also the three main radiation regions of a beam pattern, namely (i) the main-lobe region, which is defined by the mask beamwidth ψ_b; (ii) the transition region, which is defined by the angular interval between the ideal mask boundary and the first pattern null; and (iii) the remaining side-lobe region.

We recall that the array factor for a ULA with steering direction $\psi_0{=}0$ and complex array weights $\mathbf{p}$ is given by:

$$AF(\psi) = \sum_{m=0}^{M-1} \mathbf{p}(m)\mathrm{e}^{-jm\psi} \tag{2.2}$$

The problem of designing a ULA with magnitude of the array factor $|AF(\psi)|$ as close as possible to the symmetrical ideal response $\Pi(\psi)$ can be significantly simplified by the following well-known lemma [44] which allows to consider real array weights $\mathbf{p} \in \mathbb{R}^M$ instead of complex array weights $\mathbf{p} \in \mathbb{C}^M$.

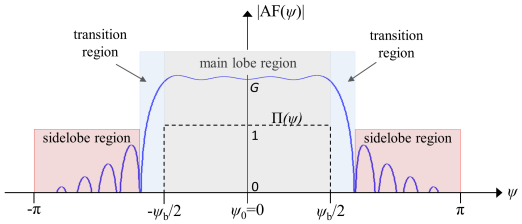


Figure 2.1: Magnitude of the array factor: ideal objective mask and real sector beam pattern with the relevant regions of radiation highlighted.

Lemma 1. *If it exists a set of complex array weights whose $|AF(\psi)|$ is symmetrical, then it exists a set of real array weights $\boldsymbol{p}(m)$, with $m = 0, 1, ..., M-1$, which gives rise to the same $|AF(\psi)|$.*

Proof: We start by taking the polynomial:

$$P(z) = \sum_{m=0}^{M-1} \mathbf{p}(m) z^m = a_0 \prod_{m=1}^{M} (z - z_m)$$

with roots $\{z_m\}_{m=1}^{M}$. Note that $AF(\psi) = P(\mathrm{e}^{-j\psi})$. Since $|AF(\psi)|$ is symmetrical, $|P(z)|^2 = |P(z^*)|^2$ in the complex circle $|z| = 1$. Moreover, since $|P(z)|^2$ and $|P(z^*)|^2$ are equal in a dense set, they are equal in the whole complex space, i.e. $|P(z)|^2 = |P(z^*)|^2$, $\forall z \in \mathbb{C}$. This means that if z_0 is a root of $P(z)$ with multiplicity α, then it is a root of both $|P(z)|^2$ and $|P(z^*)|^2$ with multiplicity 2α and, therefore, it is also a root of $P(z^*)$ with multiplicity α. Finally, if z_0 is a root of $P(z)$ with multiplicity α, then also z_0^* is a root of $P(z)$ with multiplicity α. This means that if we fix a_0 to be real, all the coefficients $\mathbf{p}(m)$, with $m = 0, 1, ..., M-1$, are real. ■

Since the beam patterns to be synthesized are symmetrical, we can apply Lemma 1 and assume $\mathbf{p} \in \mathbb{R}^M$ in Eq. 2.2. The array factor expression can be therefore revised as:

$$AF(\psi) = \mathrm{e}^{j\beta\psi} \overline{AF}(\psi) \tag{2.3}$$

where $\beta = (M-1)/2$ and the term $\overline{AF}(\psi)$ is given by:

$$\overline{AF}(\psi) = \sum_{m=-\beta}^{\beta} \mathbf{p}(m+\beta) \mathrm{e}^{-jm\psi} \tag{2.4}$$

By Fourier analysis, $\overline{AF}(\psi)$ can be expressed as the sum of of sine and cosine functions with imaginary and real coefficients respectively. In order to further simplify the problem, we impose that all the sine terms in the Fourier expansion of $\overline{AF}(\psi)$ are zeroes. This leads to symmetrical antenna weights, i.e., $\mathbf{p}(i) = \mathbf{p}(j)$ for $i+j = 2\beta$ and $i, j \in \{0, 1, ..., M-1\}$. The reason behind this approximation relies on two main observations: (1) for the main

steering direction, sine terms do not contribute to the radiated power because $\sin(\psi_0)=0$ for $\psi_0=0$; (2) neglecting the sine terms provides beam patterns with reduced sidelobe power level $\left|\overline{AF}(\psi)\right|^2_{\text{SL}}$ since:

$$\left|\overline{AF}(\psi)\right|^2_{\text{SL}} = |\mu_{\sin}(\psi)|^2 + |\nu_{\cos}(\psi)|^2 \geq |\nu_{\cos}(\psi)|^2$$

where $\mu_{\sin}(\psi)$ and $\nu_{\cos}(\psi)$ account for the sine and cosine terms in the Fourier expansion respectively. Finally, we can approximate $\overline{AF}(\psi)$ by the following sum of real terms for even M,

$$\overline{AF}(\psi) = 2 \sum_{m=\beta+\frac{1}{2}}^{M-1} \mathbf{p}(m) \cos[(m-\beta)\psi] \tag{2.5}$$

and for odd M,

$$\overline{AF}(\psi) = \mathbf{p}(\beta) + 2 \sum_{m=\beta+1}^{M-1} \mathbf{p}(m) \cos[(m-\beta)\psi] \tag{2.6}$$

Since $\overline{AF}(\psi)$ can be approximated by a real function built with the intent to not compromise the radiation in the main steering direction and, at the same time, minimize the sidelobe power level, we can assume $|AF(\psi)| \approx \overline{AF}(\psi)$. Therefore, the problem of designing a ULA with $|AF(\psi)|$ approximating the ideal response $\Pi(\psi)$ can be formulated as:

$$\begin{aligned} \mathbf{p} = \arg\min_{\hat{\mathbf{p}}} & \|\overline{AF}(\psi) - G\Pi(\psi)\| \\ \text{s.t. } & \overline{AF}(\psi) \text{ is defined by } \hat{\mathbf{p}}, \\ & 0 < G \in \mathbb{R}, \quad \|\hat{\mathbf{p}}\| = 1 \end{aligned} \tag{2.7}$$

or equivalently

$$\begin{aligned} \mathbf{p} = \arg\min_{\hat{\mathbf{p}}} & \|AF(\psi) - G\mathrm{e}^{j\beta\psi}\Pi(\psi)\| \\ \text{s.t. } & AF(\psi) \text{ is defined by } \hat{\mathbf{p}}, \\ & 0 < G \in \mathbb{R}, \quad \|\hat{\mathbf{p}}\| = 1 \end{aligned} \tag{2.8}$$

where G is a positive variable accounting for the gain of the beam in the steering direction. The problem in Eq. 2.8 can be decomposed into two sub-problems, namely: Minimization over G

$$\begin{aligned} \mathbf{p} = \mathbf{p}_G \text{ for } \quad G = \arg\min_{G} & \|AF_G(\psi) - G\mathrm{e}^{j\beta\psi}\Pi(\psi)\| \\ \text{s.t. } & AF_G(\psi) \text{ is defined by } \mathbf{p}_G, \\ & 0 < G \in \mathbb{R}, \quad \|\mathbf{p}_G\| = 1 \end{aligned} \tag{2.9}$$

Minimization over $\hat{\mathbf{p}}$ for a given $0 < G \in \mathbb{R}$

$$\begin{aligned} \mathbf{p}_G = &\arg\min_{\hat{\mathbf{p}}} \|AF(\psi) - G\mathrm{e}^{j\beta\psi}\Pi(\psi)\| \\ &\text{s.t. } AF(\psi) \text{ is defined by } \hat{\mathbf{p}},\ \|\hat{\mathbf{p}}\| = 1 \end{aligned} \tag{2.10}$$

Considering $\mathbf{b}$ as the vector that contains the Fourier series coefficients of the objective function,

$$\mathbf{b}(m) = \mathcal{F}\left\{\mathrm{e}^{j\beta\psi}\Pi(\psi)\right\}(m) = \mathrm{sinc}\left[\frac{\psi_{\mathrm{b}}(m-\beta)}{2\pi}\right] \tag{2.11}$$

and applying Parseval's identity, Problem 2.10 can be reduced to the following problem:

$$\begin{aligned} \mathbf{p}_G = &\arg\min_{\hat{\mathbf{p}}} \|\hat{\mathbf{p}} - G\mathbf{b}\| + C_G \\ &\text{s.t. } AF(\psi) \text{ is defined by } \hat{\mathbf{p}},\ \|\hat{\mathbf{p}}\| = 1 \end{aligned} \tag{2.12}$$

where C_G can be removed from the optimization framework since it is a constant defined as:

$$C_G = \sum_{m<-\infty}^{-1} \left|\left[\mathcal{F}\left(\mathrm{e}^{j\beta\psi}\Pi(\psi)\right)\right]_m\right|^2 + \sum_{m=M}^{+\infty} \left|G\left[\mathcal{F}\left(\mathrm{e}^{j\beta\psi}\Pi(\psi)\right)\right]_m\right|^2$$

The problem in Eq. 2.12 represents the classical geometrical problem of finding the closest point to a multidimensional sphere with solution:

$$\mathbf{p}_G = \frac{G\mathbf{b}}{\|G\mathbf{b}\|} = \frac{\mathbf{b}}{\|\mathbf{b}\|} \tag{2.13}$$

Since $\mathbf{p}_G$ does not depend on G, the solution for Problem 2.9 is $\mathbf{p} = \frac{\mathbf{b}}{\|\mathbf{b}\|}$ as well. The array vector weights for any steering direction[1] $\psi_0 \neq 0$ can be easily obtained by multiplying each $\mathbf{b}(m)$ in Eq. 2.11 by $\mathrm{e}^{jm\psi_0}$, for $m = 0, 1, ..., M-1$.

2.5. Results

In this section, we assess the performance of our formulation for the synthesis of sector beam patterns with ULAs and compare the results against those achieved with well-established state-of-the-art designs.

As a first numerical experiment, we synthesize a sector beam pattern defined by the purely rectangular mask $\Pi(\psi)$ in Fig. 2.1 with ψ_{b}=0.9 rad (approximately 50°). Because of lack of space, we only consider the case of a ULA with M=32, $\lambda/2$-spaced isotropic antennas. In Fig. 2.2, we plot the beam patterns resulting from applying our formulation and three classical state-of-the-art designs for sector beam patterns, namely the Fourier Series Method with both rectangular (FSM-RW) and Kaiser (FSM-KW) windows, and the Woodward-Lawson frequency-sampling Method with Hamming window (WLM-HW) [39, Ch. 21]. Rectangular and Hamming windows can be configured in terms of desired

[1] The interested reader can find the Matlab script of the proposed formulation at the following link: `http://wireless.networks.imdea.org/software`.

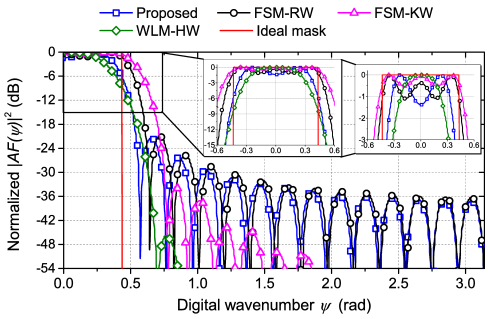


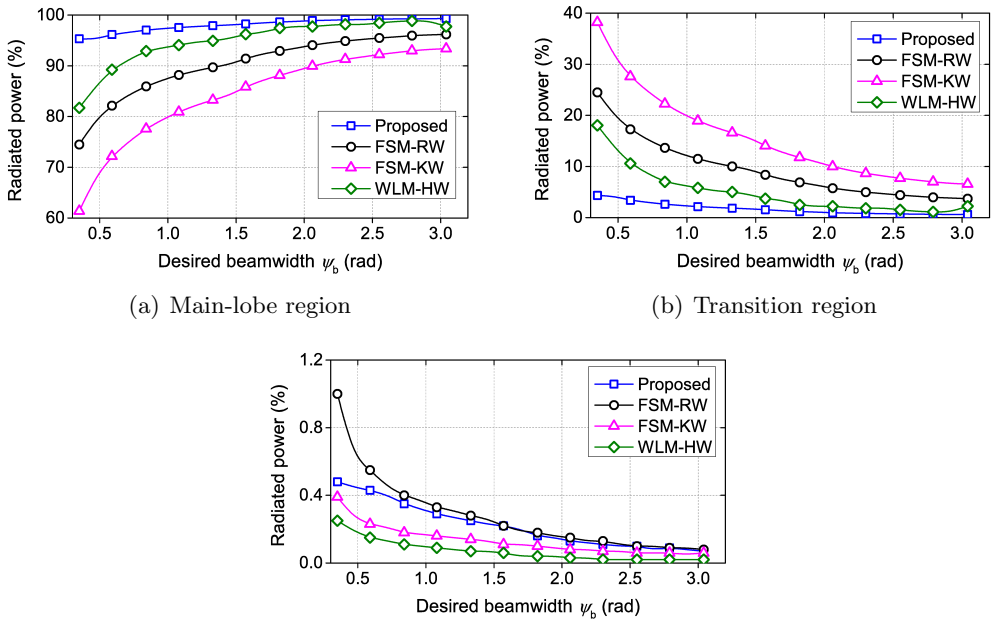
Figure 2.2: Symmetric beam patterns synthesized with M=32 antenna elements for a desired beamwidth $\psi_{\mathrm{b}} = 0.9$ rad (approximately 50°). Comparison between the proposed approach and state-of-the-art techniques.

beamwidth, but not in terms of stop-band attenuation (which is fixed, by definition, to 21 dB and 54 dB respectively). The Kaiser window, instead, is more customizable and allows the user to set, in addition to the beamwidth, the desired stop-band attenuation (set to 30 dB in this paper in order to provide a trade-off between rectangular and Hamming windows). As highlighted in Section 2.4, the main strength of our design is its inherent simplicity. The main limitation of our solution is that it is not possible to set the required stop-band attenuation.

As shown in Fig. 2.2, compared with the other techniques, the proposed formulation provides comparable flatness in the main-lobe region and significantly reduced radiation in the transition region. In fact, in its transition region, the array factor of the proposed design decreases much more sharply than that of the other approaches. In addition to this, the proposed formulation has a smaller transition region compared with other techniques, with the first pattern nulls located at 0.56 rad, 0.65 rad, 0.78 rad, and 0.71 rad respectively for our design, FSM-RW, WLM-HW, and FSM-KW.

As a second numerical experiment, we analyze the percentage of power radiated by the beam patterns in the three different regions highlighted in Fig. 2.1. As evident from Fig. 2.3(a), independently from the beamwidth ψ_{b} to be synthesized, the proposed formulation significantly outperforms the other techniques in terms of ability to concentrate the radiated power in the main-lobe region. As a consequence of this, the power radiated by our beam patterns in the transition region (Fig. 2.3(b)) is drastically reduced in contrast with the other designs. The percentage of power radiated in the side-lobe region (Fig. 2.3(c)) is comparable for all the methods and, notably, it is always below 0.5% of the total radiated power for the proposed formulation. These aspects are paramount in directional wireless networks since excessive out-of-beam radiation is the primary source of interference to other devices.

In order to further assess the effectiveness of the proposed formulation to approximate the ideal mask, we calculate the half-power beamwidth (HPBW) error as $\Delta\psi_{\mathrm{3dB}} = |\psi_{\mathrm{3dB,synt}} - \psi_{\mathrm{b}}|$, i.e. as the absolute value of the difference between the HPBW



(a) Main-lobe region

(b) Transition region

(c) Side-lobe region

Figure 2.3: Percentage of power radiated by the beam patterns synthesized with M=32 antennas in the different azimuthal regions defined in Fig. 2.1.

of the synthesized beam pattern and the desired HPBW. The results plotted in Fig. 2.4 when varying the desired beamwidth ψ_{b} show that our beam patterns exhibit performance comparable with the FSM-RW. The $\Delta\psi_{\mathrm{3dB}}$ error obtained by our formulation is always below 0.1 rad, which is approximately 50% and 33% lower compared with that achieved by FSM-KW and WLM-HW respectively.

As a final numerical experiment, we analyze the run time required by the different design strategies to synthesize beam patterns with desired beamwidth ψ_{b}=0.9 rad. In Fig. 2.5, we plot, as a function of the number of antenna elements, the computational times averaged over 500 Monte Carlo simulations on a PC with quad-core Intel Core i7 CPU. The results demonstrate the computational advantage of the proposed formulation which is approximately one order of magnitude faster than the other state-of-the-art techniques.

2.6. Summary

In this chapter, we presented a new technique aimed at synthesizing sector beam patterns for ULAs. The proposed strategy is based on a very compact expression we derived for calculating the excitations of a ULA with radiation pattern conforming a target mask. Numerical experiments demonstrated that not only is our technique more effective than traditional methods to shape sector beam patterns complying with the desired mask, but it also provides one order of magnitude lower computational complexity.

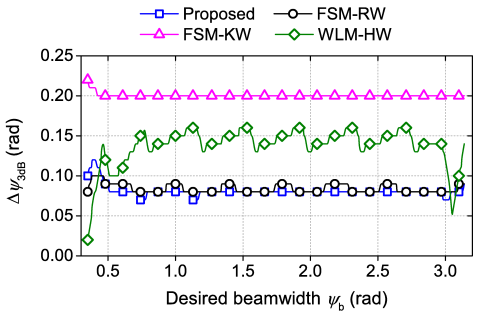


Figure 2.4: HPBW error calculated as the absolute value of the difference between synthesized and desired HPBWs.

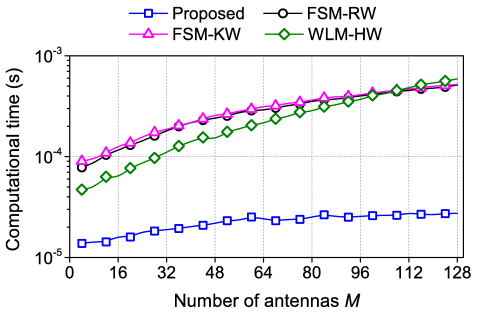


Figure 2.5: Average computational time (on a semilogarithmic scale) required to synthesize beam patterns with beamwidth ψ_b=0.9 rad as a function of the number of antennas.

Part III

Training the radiation patterns for communication

Chapter 3

Beam forming

3.1. Challenges

Due to the spectrum fragmentation below 6 GHz, current wireless communication systems cannot support user data rates of several Gbps and above in a commercially viable manner. Such bandwidth shortage has motivated the exploration of vacant, unlicensed spectrum at higher frequencies, and in particular in the millimeter-wave (mmWave) frequency bands which the future fifth-generation (5G) wireless standard is expected to exploit [2], [45].

The higher propagation loss and unfavorable atmospheric absorption make data transmission over relatively long distances a serious challenge at mmWaves. The short wavelength, however, allows more antenna elements to be integrated into mmWave devices and enables large-scale multiple-input multiple-output (MIMO) and adaptive antenna arrays to overcome range limitations. Traditional MIMO processing as performed in sub-6 GHz systems is, at present, impractical at mmWaves because the high cost, power consumption, and complexity of mixed-signal hardware at such frequencies prevent the use of a dedicated RF chain per antenna element [16].

3.2. Related work

For this reason, most of the literature proposes analog RF beamforming architectures which rely on networks of RF phase shifters to control the phase of the signal at each antenna element [46–49]. However, compared to fully-digital beamforming architectures which can precisely control both phase and amplitude of the transmitted/received signals, analog beamforming is sub-optimal because of the constant amplitude and the low phase resolution constraints affecting the RF phase shifters currently available for mmWave frequencies [50]. Such constraints seriously impact the beamforming performance, resulting in antenna radiation patterns with (i) high, difficult-to-control sidelobe levels (SLLs),

(ii) array gain dependent on the steering direction, and (iii) fixed beamwidth [51]. This latter point is crucial for the design of low-overhead, multi-level beam training protocols where the angular sector covered by the beam is wider in the earlier training stages and is progressively narrowed to a converging antenna pattern for subsequent data transmissions [46, 47]. Furthermore, since the transceiver can only use one beam direction at a time, analog beamforming lacks multi-direction searching, thus resulting in more time required for beam training. This leads to the well-known tradeoff between the time devoted to beam training (overhead) and the effective data rate [52], [53]. That is, increasing the alignment overhead leaves less time for data transmission, thus resulting in reduced transmission rates.

Hybrid analog-digital architectures aim at approaching the performance of fully-digital beamforming by dividing the precoding/combining operations between the analog and digital domains and adopting fewer RF chains compared to pure digital architectures. The availability of multiple RF chains allows hybrid analog-digital architectures to shape beam patterns very close to those attained by a fully-digital architecture, i.e., with limited overlap between adjacent beams and excellent flatness over the covered sectors. Moreover, since multiple RF chains used simultaneously provide multi-direction scanning, hybrid analog-digital architectures are able to speed up the beam training phase, hence increasing the effective data rate.

Despite the substantial research interest recently gained by hybrid beamforming for mmWave systems, to the authors' knowledge, only very few works in the literature have investigated the opportunity of leveraging the multi-direction scanning capabilities of hybrid transceivers to accelerate the beam training. For example, the acceleration of the cell discovery procedure is analyzed in [54] and the gain provided by hybrid beamforming is simply deduced based on the performance achieved for analog and digital beamforming. In [55], the simultaneous, multi-direction scanning capabilities of hybrid transceivers are exploited to estimate the mmWave channel. The synthesis of hybrid analog-digital precoders and combiners is accomplished by a genetic algorithm whose notable complexity, however, hinders its implementation in resource constrained mobile devices.

Most of the literature on mmWave hybrid beamforming has, until now, considered sequential, single-direction scanning during the beam training phase, focusing the attention, instead, on the design of optimal precoders to achieve larger multiplexing gains during the data trasmission phase. However, the proposed solutions exhibit high computational complexity and rely on unrealistic hardware assumptions based on the availability of a considerable number of RF chains and RF phase shifters with large (or even infinite) number of quantization bits. In [56], a codebook design algorithm is developed under the assumption that hybrid analog-digital beamforming is used only at the transmitter while the receiver is equipped with a single antenna. The hybrid codebook is designed by minimizing the mean square error (MSE) between the code vector's beam pattern

and its corresponding ideal beam pattern. The approximation of MSE minimization is accomplished by the orthogonal matching pursuit (OMP) algorithm with unconstrained digitally-controlled RF phase shifters (i.e., with infinite number of quantization bits). A similar OMP-based approach with unconstrained RF phase shifters is discussed also in [57]. A hybrid analog-digital codebook relying on beamforming vectors with different beamwidths and gains is presented, for the first time, in [10]. The ability to generate beams with various beamwidths makes such codebook particularly attractive for the design of adaptive, low-overhead beam training protocols. The approach in [10] assumes that RF phase shifters with a large number of quantization bits are available at mmWave frequencies. However, the current state of silicon technologies makes the design of RF phase shifters with high phase shift resolution challenging and even impractical [50].

3.3. Contribution

In this chapter, taking the work of [10] and [35] as a starting point, we propose and implement an adaptive, low-overhead beam training protocol exploiting the simultaneous, multi-direction scanning capabilities of hybrid analog-digital transceivers. To this end, we take into account practical constraints of mmWave transceivers and design a hybrid codebook to match the multi-beamwidth, multi-beam requirements of the proposed protocol. The contributions of this chapter are as follows:

- We propose to speed-up the beam training process in mmWave systems by exploiting the simultaneous, multi-direction reception capability of hybrid analog-digital transceivers. To this end, we design an adaptive, low-overhead beam training protocol which incorporates such a capability;

- We formulate an optimization problem to approximate a fully-digital, multi-beam codebook by means of a hybrid architecture requiring a number of RF chains much lower than the number of antenna elements and only 2-bit RF phase shifters;

- We propose and implement a novel approach, based on geometric considerations, to efficiently solve the above optimization problem with much lower computational complexity and higher accuracy than OMP-based strategies in the literature. Our design enables simultaneous, multi-direction scanning and, therefore, can be leveraged in the framework of beam training to accelerate the mmWave link establishment.

Our solution is able to synthesize beam patterns almost indistinguishable from those shaped by fully-digital beamforming, although requires lower complexity hardware compared with the literature. Additionally, the speed up achieved by our beam training protocol determines a considerable rate increase compared to state-of-the-art strategies based on sequential, single-direction scanning during beam training.

3.4. System model

We consider the mmWave system shown in Fig. 4.1 and focus on the beam training phase during which a base station (BS) equipped with a uniform linear array (ULA) of M_{BS} isotropic radiators and $N_{\text{RF-BS}}$ RF transceiver chains uses a single stream of data to transmit training packets along one direction at each time slot. On the other side, a mobile station (MS) equipped with a ULA of M_{MS} isotropic radiators and $N_{\text{RF-MS}}$ RF transceiver chains performs signal-to-noise ratio (SNR) measurements simultaneously over multiple directions to speed-up the process. More concretely, the BS applies an $N_{\text{RF-BS}} \times 1$ digital baseband precoder $\mathbf{p}_{\text{BB}}$ followed by an $M_{\text{BS}} \times N_{\text{RF-BS}}$ RF precoder, $\mathbf{P}_{\text{RF}}$, to the discrete-time transmitted symbol $s(t)$ — the transmit power constraint is ensured by imposing the normalization $\|\mathbf{P}_{\text{RF}}\mathbf{P}_{\text{BB}}\|_2^2 = 1$. On the other side, the MS configures its front end to concurrently receive over $D \leq N_{\text{RF-MS}}$ different directions. To do that, it applies an $M_{\text{MS}} \times N_{\text{RF-MS}}$ RF combiner $\mathbf{C}_{\text{RF}}$ followed by a $N_{\text{RF-MS}} \times D$ digital baseband combiner $\mathbf{C}_{\text{BB}}$. The discrete-time signal received by the MS is the $D \times 1$ vector $\mathbf{y}(t)$ given by:

$$\mathbf{y}(t) = \sqrt{\rho}\mathbf{C}^{\text{H}}\mathbf{H}\mathbf{p}s(t) + \mathbf{C}^{\text{H}}\mathbf{n}(t) \tag{3.1}$$

where $\mathbf{C} = \mathbf{C}_{\text{RF}}\mathbf{C}_{\text{BB}}$ (dimensions $M_{\text{MS}} \times D$), $\mathbf{p} = \mathbf{P}_{\text{RF}}\mathbf{p}_{\text{BB}}$ (dimensions $M_{\text{BS}} \times 1$), ρ is the average received power, $\mathbf{H}$ is the $M_{\text{MS}} \times M_{\text{BS}}$ mmWave channel matrix, and $\mathbf{n}(t) \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ is the complex white Gaussian noise.

Similarly to [10, 56, 57], we consider a narrowband block-fading propagation channel model which captures the geometrical structure and sparse nature of mmWave channels [58]:

$$\mathbf{H} = \sqrt{\frac{M_{\text{BS}}M_{\text{MS}}}{L}} \sum_{l=1}^{L} \alpha_l \mathbf{a}_{\text{MS}}(\psi_{\text{MS},l})\mathbf{a}_{\text{BS}}^{\text{H}}(\psi_{\text{BS},l}) \tag{3.2}$$

where L is the number of paths, $\alpha_l \sim \mathcal{CN}(0, 1)$ is the complex gain of the l^{th} path, and $\mathbf{a}_{\text{MS(BS)}}(\psi_{\text{MS(BS)},l})$ is the ULA response vector at the MS (BS) whose expression can be found in [35, Eq. (3)]. Note that, in order to simplify the notation, we consider BS and MS implementing horizontal (2-D) beamforming only, which implies that all scattering happens in the azimuthal domain. Extensions to planar antenna arrays and, therefore, to 3-D beamforming is straightforward as reported in [59].

Our objective is to *design a multi-beam, multi-beamwidth codebook based on the mmWave hybrid architecture in Fig. 4.1 and assuming RF phase shifters with few quantization bits.* In the next section, we illustrate how the simultaneous, multi-direction scanning capabilities enabled by such a codebook can be leveraged to design low-overhead beam training protocols for fast and adaptive angle of departure (AoD) and angle of arrival (AoA) estimation.

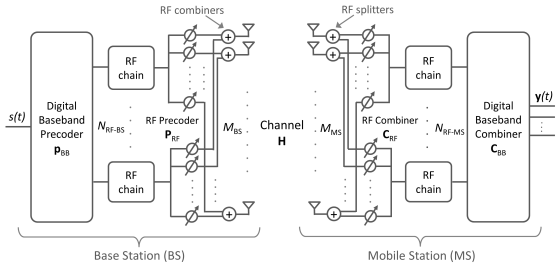


Figure 3.1: Overview of the BS-MS mmWave transceiver architecture for hybrid analog-digital beamforming.

3.5. Beam training protocol

We consider the beam training protocol used in our prior work [35], and first proposed in [10], where the beamforming vectors are adaptively configured at both the BS and MS sides based on the bisection concept. Starting from this basic algorithm, we design in this section a new beam training protocol accommodating MSs able to receive simultaneously from multiple directions as modeled in §4.5. Although the algorithm assumes the availability of a feedback channel between the BS and MS, such requirement can be easily relaxed by using the ping-pong approach described in [11].

At the beginning of the algorithm, the BS divides its $[0, 2\pi]$ azimuthal domain into K sectors, namely $S_{1,k}$ with $k = 1, 2, ..., K$, and designs the best hybrid analog-digital precoders $\mathbf{p}$ to transmit the beacon signal over such sectors at K successive time slots. At the same time, the MS also divides its azimuthal domain into K sectors, but it employs $\lceil K/D \rceil$ measurement matrices $\mathbf{C}$ at $\lceil K/D \rceil$ successive instants to detect, in parallel from D directions at each time slot, the beacon signal transmitted by the BS. In fact, the availability of multiple RF chains allows the MS to simultaneously "listen" to $D \leq N_{\text{RF-MS}}$ directions simultaneously[1]. The MS compares the SNR of the received beacon signals to determine the one with the maximum SNR. This translates into selecting the partition which is highly likely to contain the most dominant AoD/AoA of the mmWave channel. The MS then communicates the search results to the BS to prepare for the later stages where the selected partition is further divided into smaller subsets until the AoD/AoA pair is estimated with the desired resolution. It is evident that increasing D likewise reduces the number of RF chains assigned to each scanned direction, which definitely affects the quality of the beam patterns. Such a qualitative evidence will be quantitatively evaluated later in §3.7.

An example of partitioning of the azimuthal domain performed by the MS during the first three stages of the beam training protocol with K=D=2 is reported in Fig. 3.2. In the first stage, the MS receives simultaneously from sector $S_{1,1}$ and sector $S_{1,2}$ and selects

[1]In principle, simultaneous, multi-direction transmission could be also exploited at the BS to further speed up the beam training, but this is left as future work.

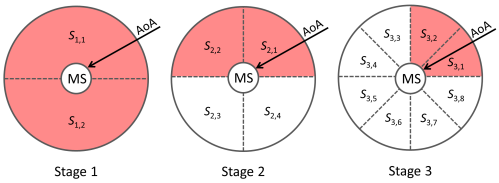


Figure 3.2: An example of azimuthal domain partitioning performed by the MS during the adaptive beam training protocol for AoA estimation (K=D=2). The highlighted sectors are the ones covered simultaneously by the MS at each stage.

the former as the one with highest SNR. In the second stage, the sector selected in the previous stage ($S_{1,1}$) is further divided into two 90° sectors ($S_{2,1}$ and $S_{2,2}$) which are used simultaneously by the MS to perform SNR measurements. The process is repeated in the succeeding stages until the required angular resolution of the AoA estimation is achieved.

In case of reciprocal channel and time-slot duration T_{slot}, since K beamforming vectors are sequentially used at the BS and, for each of them, $\lceil K/D \rceil$ measurement vectors are sequentially employed at the MS, the total time required to estimate the most dominant AoD/AoA of the channel with angular resolution $2\pi/N$ becomes:

$$\tau = T_{\text{slot}} K \left\lceil \frac{K}{D} \right\rceil \log_K N \tag{3.3}$$

This represents a significant speed up compared to the basic adaptive beam training protocol considered in [10] and [35] where the sequential reception at the MS produces an estimation time, τ_{seq}, which is D times higher than that calculated in Eq. 3.3:

$$\tau_{\text{seq}} = T_{\text{slot}} K^2 \log_K N \tag{3.4}$$

In case of non-reciprocal channel, the beam training algorithm is repeated with $\mathbf{H}$ replaced by the uplink channel and the roles of the BS and MS reversed.

The AoD/AoA estimation accuracy achieved at the end of beam training critically depends on how well the adopted beam patterns approximate the ideal sectors in Fig. 3.2. In the next section, we present an algorithmic strategy to accurately approach fully-digital sector beams by means of a hybrid analog-digital architecture requiring very low complexity hardware.

3.6. Hybrid analog-digital beam patterns

The problem of designing the best hybrid analog-digital precoder/combiner at the BS and MS sides can be translated into the one of designing, at each stage of the beam training protocol in the previous section, beam patterns with three main characteristics: (1) limited overlap with adjacent beams; (2) flat-top shape over the covered angular

region; (3) limited number and intensity of side lobes.

In this section, we present a practical codebook design which relies on the availability of a hybrid analog-digital architecture with number of RF chains much lower than the number of antenna elements, and RF phase shifters with just four phase values (0, $\pm\pi/2$, π) without amplitude adjustment.

We consider a mmWave system where the BS transmits a single stream of beam training data, i.e., it employs beam patterns covering only one of the white sectors in Fig. 3.2 at a time. On the other side, the MS adopts beam patterns able to simultaneously cover multiple sectors (for example, the red sectors in Fig. 3.2). We first discuss on the codebook design at the MS side. The introduced concepts can then be used to easily derive the codebook at the BS side as well.

At each stage of the adaptive beam training protocol illustrated in §3.5, the set of concurrent beam patterns — in the sequel also referred to as measurement vectors or combining vectors — used by the MS to receive beam training data can be arranged into a $M_{\text{MS}} \times D$ matrix $\mathbf{C} = \mathbf{C}_{\text{RF}}\mathbf{C}_{\text{BB}}$, where $D \leq N_{\text{RF-MS}}$ is the number of sectors which the MS is able to cover simultaneously, $\mathbf{C}_{\text{RF}}$ is the RF combining matrix, and $\mathbf{C}_{\text{BB}}$ is the digital baseband combining matrix. The problem of designing a hybrid analog-digital codebook with multiple simultaneous receive beams can be modeled by finding an RF (baseband) combiner $\mathbf{C}_{\text{RF}}^{\text{opt}}$ ($\mathbf{C}_{\text{BB}}^{\text{opt}}$) such that:

$$\begin{aligned}
\mathbf{C}_{\text{RF}}^{\text{opt}} &= \arg\min_{\mathbf{C}_{\text{RF}}} \left\| \mathbf{C}_{\text{bsl}} - \mathbf{C}_{\text{RF}}\hat{\mathbf{C}}_{\text{BB}} \right\|_2 \\
&\qquad \text{s.t. } [\mathbf{C}_{\text{RF}}]_{:,i} \in \mathcal{C}_{\text{RF}},\ i = 1, 2, ..., N_{\text{RF-MS}} \\
\hat{\mathbf{C}}_{\text{BB}} &= \left(\mathbf{C}_{\text{RF}}^{\text{H}}\mathbf{C}_{\text{RF}}\right)^{-1}\mathbf{C}_{\text{RF}}^{\text{H}}\mathbf{C}_{\text{bsl}} \\
\left[\mathbf{C}_{\text{BB}}^{\text{opt}}\right]_{:,i} &= \frac{[\hat{\mathbf{C}}_{\text{BB}}]_{:,i}}{\|\mathbf{C}_{\text{RF}}[\hat{\mathbf{C}}_{\text{BB}}]_{:,i}\|_2},\ i = 1, 2, ..., N_{\text{RF-MS}}
\end{aligned} \tag{3.5}$$

where $\mathbf{C}_{\text{bsl}}$ is the matrix that contains the baseline array vector weights that are to be approximated. In this chapter, we refer to baseline beam patterns as the beam patterns that can be synthesized by exploiting a fully-digital beamforming architecture where the availability of a dedicated RF chain for each antenna element enables precise control of both phase and amplitude of the mmWave signals. As in [35], we choose the sector beam design based on the Fourier Series Method with Kaiser windowing (FSM-KW) to synthesize beam patterns confined to a desired angular region [39, Chapter 21, pp. 946-949]. Compared to other windows such as Hamming, Blackman, etc., the Kaiser window has more design flexibility since the trade-off between the main lobe width and the sidelobe ripple amplitude can be accurately set by adjusting some window parameters.

The finite set of possible analog beamforming vectors is encompassed by the set $\mathcal{C}_{\text{RF}} \in \mathbb{C}^{M_{\text{MS}}}$ of vectors with components $\{\pm 1, \pm j\}$, which is also referred to as dictionary in the sequel. Such constraint on $\mathbf{C}_{\text{RF}}$ translates into the hardware constraint of using RF phase

Algorithm 1 A geometric approach to design hybrid analog-digital multi-directional beam patterns

Require: $\mathbf{C}_{\text{bsl}}$
$\mathbf{C}_{\text{RF}} = \text{Empty matrix}$
$\mathbf{C}_{\text{res}} = \mathbf{C}_{\text{bsl}}$
for $n \leq N_{\text{RF-MS}}$ **do**
  $k = \arg\max_k \|[\mathbf{C}_{\text{res}}]_{:,k}\|_2^2$
  $\mathbf{c}_{\text{res}} = [\mathbf{C}_{\text{res}}]_{:,k}$
  $\mathbf{C}_{\text{RF}} = [\mathbf{C}_{\text{RF}}, \mathcal{S}(\mathbf{c}_{\text{res}})]$
  $M = \max_i |\mathbf{c}_{\text{res}}(i)|$, $m = \min_i |\mathbf{c}_{\text{res}}(i)|$
  $J = \text{find}\left[|(\mathbf{c}_{\text{res}}| \geq \frac{M+m}{2}\right]$
  $\delta' = \text{mean}\left[\mathbf{c}_{\text{res}}(J)/\mathcal{S}(\mathbf{c}_{\text{res}})(J)\right]$
  if $|\delta'| > \frac{M+m}{2}$ **then**
    $\delta = \frac{\delta'}{|\delta'|}\frac{M+m}{2}$
  else
    $\delta = \delta'$
  end if
  $\mathbf{c}_{\text{res}} = \mathbf{c}_{\text{res}} - \delta\mathcal{S}(\mathbf{c}_{\text{res}})$
  $[\mathbf{C}_{\text{res}}]_{:,k} = \mathbf{c}_{\text{res}}$
  for $i \leq D$, $i \neq k$ **do**
    $[\mathbf{C}_{\text{res}}]_{:,i} = [\mathbf{C}_{\text{res}}]_{:,i} - \mathcal{S}(\mathbf{c}_{\text{res}})\frac{[\mathcal{S}(\mathbf{c}_{\text{res}})]^{\text{H}}[\mathbf{C}_{\text{res}}]_{:,i}}{\|\mathcal{S}(\mathbf{c}_{\text{res}})\|_2^2}$
  end for
end for
$\hat{\mathbf{C}}_{\text{BB}} = \left(\mathbf{C}_{\text{RF}}^{\text{H}}\mathbf{C}_{\text{RF}}\right)^{-1}\mathbf{C}_{\text{RF}}^{\text{H}}\mathbf{C}_{\text{bsl}}$
for $i \leq D$ **do**
  $[\mathbf{C}_{\text{BB}}]_{:,i} = \frac{[\hat{\mathbf{C}}_{\text{BB}}]_{:,i}}{\|\mathbf{C}_{\text{RF}}[\hat{\mathbf{C}}_{\text{BB}}]_{:,i}\|_2}$
end for
return $\mathbf{C}_{\text{RF}}$, $\mathbf{C}_{\text{BB}}$

shifters with just two quantization bits[2].

The problem stated in Eq. 3.5 can be solved using the OMP algorithm and its variants as proposed in most of the literature on hybrid analog-digital beamforming [10, 35, 56, 57, 59]. Despite their simplicity and ease of implementation, OMP-based approaches share the disadvantage of high computational complexity — $O(N_{\text{RF-XX}}^4)$ due to a matrix inversion at each step — and reduced accuracy since they rely on non-complete dictionaries made of basis subsets. On the contrary, we propose a novel approach, based on geometric considerations and manipulations, to solve the problem in Eq. 3.5 with lower computational complexity — in the order of $O(N_{\text{RF-XX}}^3)$ — than OMP-based strategies. It is worth noting that, in contrast to the above mentioned literature which focuses on the design of hybrid codebooks covering only one sector at a time, our design enables the

[2]The algorithm requires only minor modifications to accommodate RF phase shifters with a different number of quantization bits.

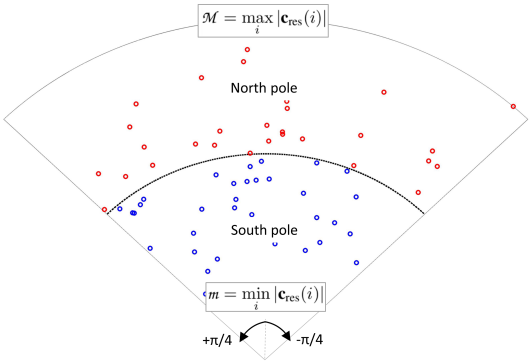


Figure 3.3: Illustrative example of the geometric approach to find the parameter δ required for the residual update strategy in Algorithm 1 (case with 64 antenna elements).

MS to simultaneously cover more than one sector during the beam training process. This translates into the problem of finding an optimal RF combiner $\mathbf{C}_{\text{RF}}$ for all the directions scanned simultaneously.

Algorithm 1 outlines the pseudo-code of our greedy, geometric strategy to approximate the fully-digital codebook $\mathbf{C}_{\text{bsl}}$ by a hybrid analog-digital architecture with 2-bit RF phase shifters. The algorithm takes as input parameter the baseline combiner to approximate, $\mathbf{C}_{\text{bsl}}$. Then, it initializes $\mathbf{C}_{\text{RF}}$ and the residual matrix $\mathbf{C}_{\text{res}}$. The algorithm proceeds by selecting from $\mathbf{C}_{\text{res}}$ the residual column $\mathbf{c}_{\text{res}}$ with maximum norm and appending its mapped version $\mathcal{S}(\mathbf{c}_{\text{res}}) \in \mathcal{C}_{\text{RF}}$ to the current RF combiner $\mathbf{C}_{\text{RF}}$ (lines 4-6). We denote $\mathcal{S}(\mathbf{v})$ an operator that maps the vector $\mathbf{v}$ into one close vector attainable with 2-bit RF phase shifters, i.e., a vector from the dictionary $\mathcal{C}_{\text{RF}}$. In other words, $\mathcal{S}(\mathbf{v})$ maps each component $\mathbf{v}(\ell)$ to the nearest value $\mathrm{e}^{j\phi_\ell}$, being $\phi_\ell \in \left\{-\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi\right\}$.

Lines 7-14 refer to the residual update strategy which represent the core of the algorithm. The objective is to make $\mathbf{c}_{\text{res}} = \mathbf{c}_{\text{res}} - \delta\mathcal{S}(\mathbf{c}_{\text{res}})$ close to zero (not in norm) which equivalently translates into finding a value of $\delta \in \mathbb{C}$ such that:

$$\mathbf{c}_{\text{res}}(i) - \delta S(\mathbf{c}_{\text{res}})(i) \approx 0 \qquad \forall i \leq M_{\text{MS}} \tag{3.6}$$

Mean square error minimization is not a convenient criterion for selecting δ because, at each step, the goal is not only to find a good projection for the selected $\mathbf{C}_{\text{RF}}$, but also a favorable residual for the future $\mathbf{C}_{\text{RF}}$ component approximations. For this reason, we geometrically envision the problem of finding δ as equivalent to that of finding a good center for the set of points $\frac{\mathbf{c}_{\text{res}}(i)}{\mathcal{S}(\mathbf{c}_{\text{res}})(i)}$ which are distributed inside a 90° circular sector in the complex plane as illustratively shown in Fig. 3.3. In fact, since $\mathcal{S}(\mathbf{c}res)$ maps the components of $\mathbf{c}_{\text{res}}$ to the closest points with 2-bit quantized phase, each point $\frac{\mathbf{c}_{\text{res}}(i)}{\mathcal{S}(\mathbf{c}_{\text{res}})(i)}$ has a complex argument between $-\pi/4$ and $\pi/4$ and modulus $|\mathbf{c}_{\text{res}}(i)|$. In order to find a good center for this set of points, we first find the maximum $M = \max_i |\mathbf{c}_{\text{res}}(i)|$ and the minimum $m = \min_i |\mathbf{c}_{\text{res}}(i)|$. Then, we divide the 90° circular sector into two sub-sectors,

namely the north pole which includes the points with modulus greater than or equal to $(M+m)/2$ and the south pole which includes the remaining points. Since, because of the shape of the circular sector, the points in the south pole have less variance compared to the points in the north pole, we focus on the latter, more representative points and calculate their mean value δ'. In line 8 of Algorithm 1, J is the set of indices relative to the elements of $\mathbf{c}_{\text{res}}$ falling within the north pole of Fig. 3.3, while the operator "/" in line 9 represents the element-wise division. If δ' falls within the north pole, we set δ to be on the "equator", i.e., $\delta = \frac{M+m}{2}\frac{\delta'}{|\delta'|}$; otherwise, we set $\delta = \delta'$. The residual matrix is updated in lines 15-19.

The process continues until all $N_{\text{MS-RF}}$ beamforming vectors have been selected. Finally, the algorithm normalizes the digital baseband combining vector $\hat{\mathbf{C}}_{\text{BB}}$ to satisfy the constraint $\|\mathbf{C}_{\text{RF}}[\mathbf{C}_{\text{BB}}]_{:,i}\|_2^2 = 1$, for $i = 1, 2, ..., D$, and outputs it along with the constructed RF precoding matrix $\mathbf{C}_{\text{RF}}$.

Algorithm 1 can be used with only minor modifications to design hybrid analog-digital beam patterns at the BS side as well. In this case, since we are assuming that the BS transmits over one sector at a time, the required input parameter for the algorithm is the baseline precoding vector $\mathbf{p}_{\text{bsl}}$ to approximate, which is calculated according to the FSM-KW design outlined in [35]. In this case, the output provided by the algorithm is the RF precoding matrix $\mathbf{P}_{\text{RF}}$ and the baseband precoding vector $\mathbf{p}_{\text{BB}}$ given by:

$$\begin{aligned} \mathbf{p}_{\text{BB}} &= \left(\mathbf{P}_{\text{RF}}^{\text{H}}\mathbf{P}_{\text{RF}}\right)^{-1}\mathbf{P}_{\text{RF}}^{\text{H}}\mathbf{p}_{\text{bsl}} \\ \mathbf{p}_{\text{BB}} &= \frac{\mathbf{p}_{\text{BB}}}{\|\mathbf{P}_{\text{RF}}\mathbf{p}_{\text{BB}}\|_2} \end{aligned} \tag{3.7}$$

An implementation of the example proposed in Fig. 3.2 with ideal sectors is shown in Fig. 3.4, where beam patterns resulting from applying the proposed algorithm (solid blue/red line) are compared with the baseline fully-digital beamforming scheme based on FSM-KW (dashed black line). All the patterns are generated by a device having 64, $\lambda/2$-spaced isotropic antenna elements. As for the hybrid analog-digital design, an architecture consisting of 2-bit RF phase shifters and 10 RF chains is considered. As evident from the plots, the proposed algorithm is able to synthesize multi-beam patterns with limited overlap between adjacent beams and excellent flatness over the covered sectors. Furthermore, it is worth noting that, even with very low complexity hardware, the designed hybrid beam patterns are almost indistinguishable from the baseline fully-digital ones.

3.7. Simulation results

In this section, we present simulation results to evaluate the performance achieved by leveraging the proposed hybrid analog-digital beam patterns for adaptive beam training.

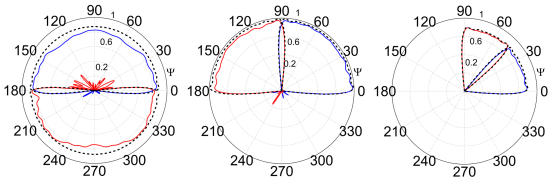


Figure 3.4: Beam patterns for the first three stages of the beam training protocol. Comparison between the proposed hybrid analog-digital design (solid plot) and the baseline fully-digital design based on FSM-KW (dashed, black plot).

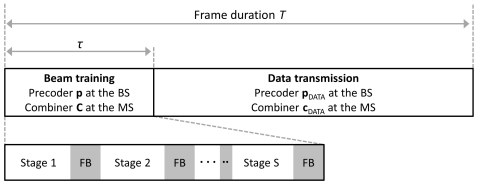


Figure 3.5: Frame structure encompassing both beam training with feedback and data transmission.

Specifically, we split the problem into two phases as shown in Fig. 4.2. In the first phase, the BS and MS use Algorithm 1 to shape their beam patterns and apply the adaptive beam training protocol of §3.5 to estimate the mmWave channel parameters, i.e., the most powerful AoD/AoA and the respective path gain. After the beam training, in a second phase, BS and MS leverage the geometric structure of the channel to construct the estimated channel matrix and build their hybrid data precoder $\mathbf{p}_{\text{DATA}}$ (at the BS) and data combiner $\mathbf{c}_{\text{DATA}}$ (at the MS) to approximate the dominant singular vectors of the estimated channel. Note that, compared to $\mathbf{C}$, which is the matrix used by the MS to combine multiple, simultaneous measurements during beam training, $\mathbf{c}_{\text{DATA}}$ is a vector since, for simplicity, we are considering data transmission through only one dominant path in the channel.

In a first set of simulations, we analyze the performance of the beam training protocol when varying the number of simultaneous directions D scanned by the MS. Concretely, in order to quantify the quality of the channel estimation achieved during the beam training phase, we consider the actual spectral efficiency (also referred to as rate in the sequel) experienced in the data transmission period $T - \tau$, which is given by the rate expression:

$$R = \log_2 \left| 1 + \text{SNR} \left| \mathbf{c}_{\text{DATA}}^{\text{H}} \mathbf{H} \mathbf{p}_{\text{DATA}} \right|^2 \right| \tag{3.8}$$

where SNR is the average signal-to-noise ration when omnidirectional antennas are used at both the BS and MS.

For the simulations, we consider one BS with $M_{\text{BS}} = 64$ antennas and $N_{\text{RF-BS}} = 10$ RF chains, and one MS with $M_{\text{MS}} = 24$ and $N_{\text{RF-MS}} = 6$ RF chains, both featuring the hybrid analog-digital architecture of Fig. 4.1 and running the beam training algorithm

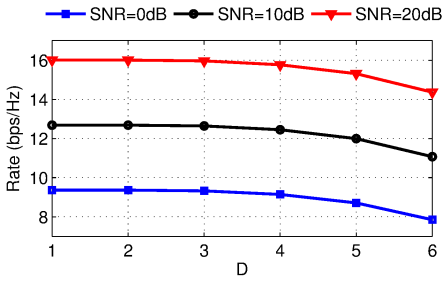


Figure 3.6: Spectral efficiency when varying the number of simultaneously scanned directions D during beam training with $K = 6$. The figure compares the performance for different SNR values.

described in §3.5. The distance between the BS and MS is set to 30m, and the path loss exponent is fixed to $n_{\text{ple}} = 3$. The antenna arrays are ULAs with $\lambda/2$-spaced isotropic radiators. We align with the simulation scenario in [10] and [35], which consider the channel model described in Eq. 4.1 with a number of paths $L = 3$ and the AoDs/AoAs taking continuous values uniformly distributed in $[0, 2\pi]$. The transmit power at the BS is set to 30dBm and the system is assumed to operate at 28-GHz carrier frequency with 100-MHz bandwidth. All the numerical results provided in this section are obtained from Monte Carlo simulations with 5000 independent channel realizations for each BS/MS configuration.

The results in Fig. 3.6 indicate that, fixing the SNR, comparable spectral efficiencies are achieved for $D \leq 3$, while rate degradation is experienced for $D > 3$. The reason for this behavior is that, for a given number of RF chains, increasing D likewise increases the complexity of the beam patterns that are to be approximated. Hence, the quality of the beam shapes, and consequently of the channel estimation, strongly depends on the number of RF chains in respect of D. Exhaustive simulations, omitted for lack of space, reveal that, in general, keeping $D \leq N_{\text{RF-MS}}/2$ represents the best choice to not incur performance degradation.

In a second set of simulations, we analyze the impact of accelerating the beam training phase on the average data communication rate. To this end, differently from the previous set of simulations, we consider now the normalized spectral efficiency within a frame which is obtained by multiplying the rate R in Eq. 4.11 by the term $1 - \tau/T$. We consider the same simulation parameters of the previous analysis and the frame structure based on the work in [60]. Each frame of length $T = 10$ms is split in time into 100 slots of length $T_{\text{slot}} = 100\mu$s, a sufficiently small value to ensure channel coherence at mmWaves.

Figure 3.7 shows the normalized spectral efficiency achieved when the proposed multi-beam reception strategy is exploited at the MS. The performance is also compared with our prior work [35] and the reference paper [10], both adopting sequential reception at the MS during beam training. The beam training durations — τ and τ_{seq} respectively for our

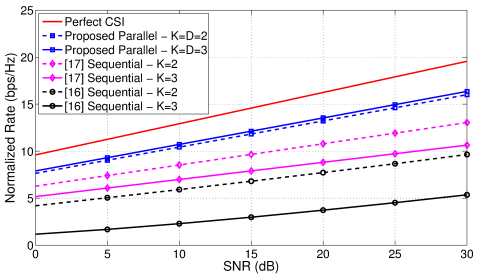


Figure 3.7: Normalized spectral efficiency when varying the SNR: comparison with the literature for different values of D and K.

strategy and the sequential strategies — are calculated as described in §3.5. As evident from the plots, the ability of the proposed approach to speed up the beam training phase (or, equivalently, to reduce the training overhead) directly improves the normalized spectral efficiency since more time can be used for actual data transmission. More concretely, our approach provides a rate only 18% smaller than the ideal rate (i.e., assuming perfect channel knowledge) and, roughly, 25% and 70% greater than [35] and [10] respectively. It is worth recalling that the proposed strategy relies on a hybrid analog-digital transceiver architecture with only 2-bit RF phase shifters in contrast to the 7-bit RF phase shifters considered in [10].

Finally, taken collectively, the presented results demonstrate how the proposed protocol is able to significantly accelerate the mmWave beam training process by means of parallel, multi-beam reception at the MS. Such an acceleration directly translates in a considerable increase of the data transmission rate compared with the literature.

3.8. Summary

In this chapter, we proposed and implemented an adaptive, low-overhead beam training protocol exploiting the simultaneous, multi-direction scanning capabilities of hybrid analog-digital transceivers operating at mmWave frequencies. To accomplish that, we derived a practical codebook design for hybrid architectures with a number of RF chains much lower than the number of antenna elements and only 2-bit RF phase shifters. Compared almost to the totality of approaches in the literature, which are based on the OMP algorithm, we relied on a novel, geometric strategy to approximate a fully-digital beamforming design with much lower computational complexity and higher accuracy. Simulation results showed that our hybrid codebooks are able to shape beam patterns almost indistinguishable from those attained by fully-digital beamforming, yet requiring lower complexity hardware compared with the literature. Moreover, the speed up of the beam training phase, in turn enabled by the multi-beam characteristics of our hybrid codebooks, provided a 25% to 70% increase in spectral efficiency compared to state-of-the-art

strategies that adopt sequential, single-direction scanning during beam training.

Chapter 4

Beam training under mobility

4.1. Challenges

The fifth generation of mobile communications (5G) is envisaged to deliver multi-Gbps wireless connectivity and to enable a plethora of new applications. It is well established that achieving extremely high data rates is impractical with currently available 4G systems due to the heavily congested and fragmented spectrum below 6 GHz. In view of this, the large amount of unoccupied spectrum in the millimeter wave (mm-wave) bands above 6 GHz becomes very appealing [2].

Communications at mm-wave frequencies are challenging since the channel suffers from severe path loss, atmospheric absorption, human blockage, and other environmental obstructions [3]. The short wavelength of the mm-waves allows beamforming arrays with many antennas to be implemented in a small form factor, thus providing sufficient link margin. On the other hand, highly directional communications complicate the link establishment and maintenance between an Access Point (AP) and a User Equipment (UE). In fact, AP and UE must perform a time-consuming beam training procedure in order to determine the best directions of transmission/reception, which incurs significant overhead (and waste of network resources). The problem is exacerbated in scenarios with mobility, since even a slight beam misalignment or environmental changes, such as link blockage, device rotation, etc., can cause considerable signal drop. To sum up, fast and efficient beam training/tracking strategies are of paramount importance to maintain seamless connectivity in a mm-wave network with node mobility.

4.2. Related work

The design space of beam search proposals in the literature can be divided into three main categories: (1) sequential scanning strategies [8, 9]; (2) adaptive algorithms employing antenna patterns with configurable beamwidth [10–13]; (3) parallel beam search with

simultaneous, multi-direction scanning [14, 15]. The vast majority of these works concentrates, however, on static networks without investigating the impact of the training latency on the overall Quality of Service (QoS) of realistic networks with mobility. Within the state-of-the-art solutions on this subject, a further subdivision can be made on the basis of the employed mm-wave transceivers. Since traditional multiple-input multiple-output (MIMO) digital beamforming (DBF) is, at present, impractical at mm-wave frequencies because of cost and power consumption constraints, analog beamforming (ABF) and hybrid analog-digital beamforming (HBF) represent the only feasible solutions. Using ABF [8, 9] provides poor performance for two main reasons. First, the constant amplitude and the low phase resolution constraints of the mm-wave RF phase shifters [16] give rise to antenna sectors with high sidelobes and reduced flatness, leading to imprecise beam training. Second, the use of a single RF chain allows for only one communication beam, thus resulting in reduced throughput and high-overhead beam search. In HBF architectures [10–15], the precoding/combining operations are divided between the analog and digital domains, while using much fewer RF chains than antenna elements. The availability of multiple RF chains enables parallel, multi-stream processing and simultaneous multi-direction scanning.

4.3. contribution

In this chapter, we consider a scenario consisting of a fixed AP and a mobile UE, both equipped with a low-complexity mm-wave HBF transceiver and communicating with directional antenna patterns. Our overall objective is to maximize the communication rate over time. To this end, we propose two strategies, a deterministic one for beam training and a probabilistic one for beam tracking, to rapidly estimate multiple, suitable directions of communication between AP and UE. Here, *beam training* is a beam search mechanism without any prior knowledge that explores the entire azimuthal domain and that is carried out both in the initial access phase and, periodically, during the AP-UE communication. *Beam tracking*, instead, is an ongoing estimation that, starting from the current steering directions, probabilistically infers how they evolve due to node mobility. The main contributions of the chapter are as follows:

- We design a two-stage beam training protocol that approaches the performance of an exhaustive search over all possible beam directions, but has very low latency and uses implicit feedback (i.e., it does not require a dedicated feedback channel). The key aspect of our beam search strategy is a particular HBF combiner matrix which takes a reduced number of sequential, multi-stream signal measurements to cover all the possible combinations of antenna weights.

- We propose, to the authors' knowledge for the first time in the mm-wave HBF

context, a beam tracking algorithm that is able to track the mm-wave channel dynamics without any training slots, but simply using known portions of the data packet (e.g., the preamble). To this end, we formulate a probabilistic optimization problem, solved by gradient descent, whose objective function is designed so as to model the temporal evolution of channel paths due to device movements. Note that this problem is quite different from the problem of MIMO channel estimation using known pilot symbols.

- We develop a simulation framework to assess the performance of the proposed beam training/tracking strategies and compare them against existing approaches in the literature. Specifically, we propose and implement a fast protocol for link establishment and maintenance under user mobility which dynamically switches between beam training and beam tracking based on the real-time QoS. Our simulator integrates a ray-tracing tool to accurately model the time-varying mm-wave channel, taking into account blockage, ray clustering, and mobility effects, and guaranteeing spatial consistency over time.

Numerical experiments show that the performance provided by our solution is very close to the optimal "oracle-based" algorithm. Furthermore, the high accuracy and reduced latency overhead characterizing our beam training/tracking strategies result in a significant rate increase over state-of-the-art solutions which in addition require higher complexity hardware. Compared to ABF solutions which share the disadvantage of converging towards only one communication beam, our approach based on HBF is capable of achieving multiplexing gains by simultaneously transmitting multiple parallel data streams over different paths.

4.4. Related work

Most of the literature on mm-wave beam search focuses on static scenarios without user mobility [9–13, 15]. Such an assumption may lead to wrong conclusions about the actual performance of the algorithms in real networks. A comparative analysis of initial access techniques in mm-wave networks is presented in [9], where performance metrics such as detection probability and delay are analyzed. The problem of tracking the AP-UE beams to handle the channel dynamics is left as future work. The design of HBF codebooks relying on beamforming vectors with different beamwidths is presented in [10–13], where it is assumed that phase shifters with a large number of quantization bits are available at mm-wave frequencies. However, the design of high-resolution mm-wave phase shifters is extremely challenging [16]. Finally, the simultaneous reception from multiple beams to accelerate the beam search is exploited in [15].

To the authors' knowledge, only very few works in the literature address the problem

of fast beam search in realistic, dynamic scenarios with node mobility. A smart beam steering algorithm for 60 GHz link re-establishment under user mobility is presented in [8]. The algorithm uses knowledge of previous feasible sector pairs to narrow the sector search space, thereby reducing the associated overhead. Numerical results show that the proposed strategy is very effective, but still incurs non-negligible latency in complex scenarios with significant blockage. A temporal channel evolution model for non line-of-sight (NLOS) mm-wave scenarios is presented in [14]. HBF at both the AP and UE is considered and a beam tracking technique based on sequentially updating the precoder and combiner is developed. However, in [14], the angle of arrival (AoA) and angle of departure (AoD) deviations due to mobility are modeled as very small uniform random variables, which are not appropriate to characterize actual mobility or significant, sudden changes in the channel due to obstacles. In [61], a linear dynamic system model to analyze the occurring errors due to link blockage and device movement is proposed. Based on the model, the authors propose two probing protocols that are effective in identifying the beam errors. However, no beam training/tracking strategy is implemented in order to find alternative antenna sector pairs once the beam errors are identified. Finally, it is worth highlighting that none of the above-mentioned works [8, 14, 61] analyzes the impact of the beam search accuracy and overhead on the evolution over time of the achievable rate under mobility.

4.5. Motivation and system model

The use of highly directional antennas with very narrow beams at both the AP and UE complicates the mm-wave link establishment and maintenance. As for the link establishment, the 60-GHz IEEE 802.11ad standard [62] implements a time-consuming beam training procedure based on an exhaustive search to find the most suitable directions of transmission and reception. Once a connection is established, the link quality degradation due to user mobility is handled through beam refinement procedures that search around the previous sector pair in order to determine a new combination of beams with improved link quality. However, in large and crowded scenarios with mobility, such procedures may fail to cope with high channel dynamics, which would necessitate fast mechanisms to scan a large angular domain (instead of just adjacent directions) to find alternative communication links. In case simply probing adjacent beams is unsuccessful, a new exhaustive beam search procedure has to be performed. This leads to a high latency which deteriorates the overall system performance. Motivated by this challenging problem, we propose two smart and efficient strategies, one for beam training and one for beam tracking, to accelerate the link establishment and maintenance between mm-wave devices in mobility scenarios.

We consider a mm-wave system with one fixed AP and one moving UE, both fea-

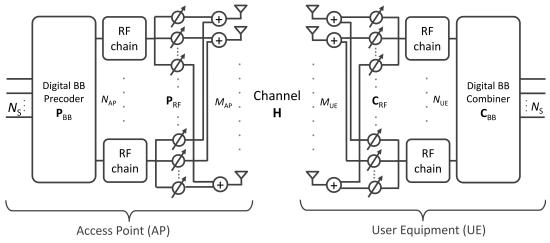


Figure 4.1: Block diagram of the AP-UE mm-wave transceiver architecture implementing HBF.

turing the same HBF architecture considered in [10, 11, 13–15] and depicted in Fig. 4.1. The AP is equipped with a uniform linear array (ULA) of M_{AP} isotropic radiators connected to N_{AP} RF transceiver chains through a network of analog/RF phase shifters. The number of antennas and RF transceiver chains at the UE side is M_{UE} and N_{UE} respectively. The HBF transceiver configuration allows AP and UE to communicate via N_{S} data streams, with $N_{\mathrm{S}} \leq \min(N_{\mathrm{AP}}, N_{\mathrm{UE}})$. To this end, the AP applies an $N_{\mathrm{AP}} \times N_{\mathrm{S}}$ digital baseband (BB) precoder $\mathbf{P}_{\mathrm{BB}}$ followed by an $M_{\mathrm{AP}} \times N_{\mathrm{AP}}$ RF precoder, $\mathbf{P}_{\mathrm{RF}}$, to the symbol sequence to be transmitted. The transmit power constraint is ensured by imposing $\|[\mathbf{P}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}]_{:,i}\|_2^2 = 1$, for $i = 1, 2, ..., N_{\mathrm{S}}$. The final AP precoder is then given by the $M_{\mathrm{AP}} \times N_{\mathrm{S}}$ matrix $\mathbf{P} = \mathbf{P}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}$. The transmitted signal passes through the $M_{\mathrm{UE}} \times M_{\mathrm{AP}}$ channel matrix $\mathbf{H}$ and impinges on the UE antennas together with white noise. Since the UE also implements HBF, it is able to concurrently receive N_{S} streams of data. To do that, it applies a $M_{\mathrm{UE}} \times N_{\mathrm{UE}}$ RF combiner $\mathbf{C}_{\mathrm{RF}}$ followed by a $N_{\mathrm{UE}} \times N_{\mathrm{S}}$ digital baseband combiner $\mathbf{C}_{\mathrm{BB}}$. The final UE combiner is given by the $M_{\mathrm{UE}} \times N_{\mathrm{S}}$ matrix $\mathbf{C} = \mathbf{C}_{\mathrm{RF}}\mathbf{C}_{\mathrm{BB}}$.

We assume that AP and UE communicate using the frame structure in Fig. 4.2. Two different types of frames can be allocated: (i) beam training frames, which contain both training and data transmission phases, and (ii) pure data frames. In the initial access procedure, the allocation of a training frame is mandatory, since AP and UE need to determine suitable initial directions of transmission. Once the initial access is accomplished, pure data frames with directional antenna patterns at both the AP and UE are used. As explained later in §4.7, the beam tracking can be performed with pure data frames, i.e., using known portions of just two data slots (one for UE beam tracking and one for AP beam tracking) without requiring any dedicated training slots. The allocation of a training frame to perform a new full beam search from scratch can be triggered periodically or when the link quality falls below a certain threshold. Based on the work in [60,63], we assume frames of duration T=10 ms, each divided into 100 slots of duration T_{slot}=100 μs, a sufficiently small value to ensure channel coherence at mm-wave frequencies.

As experimentally demonstrated in [64], the mm-wave channel between AP and UE is composed of "ray clusters", each cluster carrying a fraction of the total power. Defining

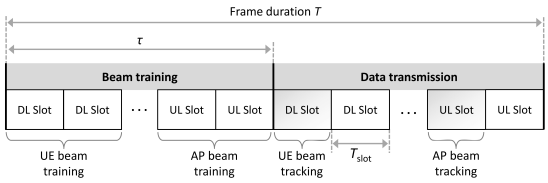


Figure 4.2: Frame structure encompassing beam training/tracking and data transmission. Data slots can be indifferently either downlink (DL) or uplink (UL) slots.

T_{slot}=100 μs as the time granularity of our system, we can express the $M_{\text{UE}} \times M_{\text{AP}}$ channel matrix at each time slot as:

$$\mathbf{H} = \sqrt{\frac{M_{\text{AP}} M_{\text{UE}}}{L}} \sum_{k=1}^{K} \sum_{\ell=1}^{L} \alpha_{k\ell} \mathbf{a}_{\text{UE}}(\theta_{k\ell}) \mathbf{a}_{\text{AP}}^{\text{H}}(\phi_{k\ell}) \tag{4.1}$$

where K is the number of clusters, L is the number of sub-paths per cluster, $\mathbf{a}_{\text{UE(AP)}}(\cdot)$ is the ULA response vector at the UE (AP) whose expression can be found in [13, Eq. (3)], and $\alpha_{k\ell}$ is the complex gain on the ℓ-th sub-path of the k^{th} cluster, which includes path loss, Doppler shift, and delay spread effects. The variables $\theta_{k\ell} \in [0, 2\pi]$ and $\phi_{k\ell} \in [0, 2\pi]$ are the ℓ^{th} AoD/AoA of the k^{th} cluster at the UE and AP respectively. In this work, we assume channel reciprocity [10, 11], that is, the AP AoDs in the downlink correspond to the AP AoAs in the uplink. The same applies to the UE as well. Note that, in order to simplify the notation, we consider the AP and UE implementing horizontal (2-D) beamforming only, which implies that all scattering happens in the azimuthal domain. Extension to planar antenna arrays and, therefore, to 3-D beamforming is straightforward.

4.6. Pseudo-Exhaustive Beam Training (PE-Train)

The use of directional antennas for mm-wave communication requires that AP and UE find suitable directions of transmission, both in the initial access phase and, periodically, during the communication. As illustrated is Fig. 4.2, if no feedback channel is available, two separate stages are required in the beam training phase, namely UE beam training (using downlink training sequences from the AP) and AP beam training (using uplink training sequences from the UE). In the following, we propose a pseudo-exhaustive beam training (PE-Train) protocol which is able to search over all possible beam directions with very low latency overhead. It uses omnidirectional transmission at the AP for UE beam training, and simultaneous, multi-stream transmission over the best estimated directions at the UE for AP beam training.

4.6.1. Stage I: UE beam training

We consider a mm-wave AP with the HBF architecture in Fig. 4.1 performing omnidirectional transmission of a training sequence $s[t]$, for $t = 1, 2, ..., T_s$, in a reciprocal channel. Arranging $s[t]$ into the $1 \times T_s$ row vector $\mathbf{s}$, the $M_{\mathrm{UE}} \times T_s$ discrete-time signal $\mathbf{R}$ impinging on the UE antennas becomes:

$$\mathbf{R} = \sqrt{P_{\mathrm{t}}}\mathbf{H}\mathbf{p}_{\mathrm{o}}\mathbf{s} + \mathbf{N} \tag{4.2}$$

where P_{t} is the transmit power, $\mathbf{H}$ is the channel matrix, $\mathbf{p}_{\mathrm{o}} = [1, 0, 0, ..., 0]^T$ is the $M_{\mathrm{AP}} \times 1$ omnidirectional precoding vector used at the AP, and $\mathbf{N}$ is a $M_{\mathrm{UE}} \times T_s$ matrix with independent, Gaussian-distributed complex noise with mean zero and variance σ^2. The lack of a dedicated RF chain for each antenna makes it impossible for a HBF transceiver to directly access $\mathbf{R}$. In fact, $\mathbf{R}$ is inevitably processed by a hybrid combiner which compresses it into a reduced dimension space, with consequent loss of information. We tackle this problem with the following strategy. First, we design an easily invertible, orthogonal $M_{\mathrm{UE}} \times M_{\mathrm{UE}}$ matrix $\mathbf{W}$ (e.g., a Hadamard matrix) representing a basis for the full space of antenna configurations. Then, since we cannot directly apply $\mathbf{W}$ to $\mathbf{R}$ because of the HBF limitations, we perform multiple, consecutive measurements, each time using as hybrid combiner a different sub-matrix of $\mathbf{W}$. Thanks to the properties of $\mathbf{W}$, we can reconstruct, at the end of the procedure, an estimated version of $\mathbf{R}$ and process it through a spatial filter matrix to derive the received signal power from each angular direction. Specifically, we build $\mathbf{W}$ such that the elements of $\sqrt{N_{\mathrm{UE}}}\mathbf{W}$ belong to the feasible set of phase-shifter weights. For example, in the case of 2-bit phase shifters, the elements of $\sqrt{N_{\mathrm{UE}}}\mathbf{W}$ can assume only four values, namely ± 1 and $\pm j$. Then, we divide $\mathbf{W}$ into $N_{\mathrm{W}} = \lceil M_{\mathrm{UE}}/N_{\mathrm{UE}} \rceil$ sub-matrices with dimensions $M_{\mathrm{UE}} \times N_{\mathrm{UE}}$, i.e., $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_{N_{\mathrm{W}}}]$. For each $\mathbf{W}_i$, with $i = 1, 2, ..., N_{\mathrm{W}}$, we build the RF combiner $\mathbf{C}_{\mathrm{RF},i} = \sqrt{N_{\mathrm{UE}}}\mathbf{W}_i$ and the baseband combiner $\mathbf{C}_{\mathrm{BB},i} = \mathbf{I}_{N_{\mathrm{UE}}}/\sqrt{N_{\mathrm{UE}}}$, where $\mathbf{I}_{N_{\mathrm{UE}}}$ is the $N_{\mathrm{UE}} \times N_{\mathrm{UE}}$ identity matrix. The overall hybrid combiner $\mathbf{C}_i = \mathbf{C}_{\mathrm{RF},i}\mathbf{C}_{\mathrm{BB},i} = \mathbf{W}_i$ is then applied to $\mathbf{R}$ at N_{W} successive instants in order to extract the following signal measurements:

$$\mathbf{Y}_i = \mathbf{W}_i^{\mathrm{H}}(\sqrt{P_{\mathrm{t}}}\mathbf{H}\mathbf{p}_{\mathrm{o}}\mathbf{s} + \mathbf{N}), \;\; i = 1, 2, ..., N_{\mathrm{W}} \tag{4.3}$$

which can be concatenated as $\mathbf{Y} = [\mathbf{Y}_1^T, \mathbf{Y}_2^T, ..., \mathbf{Y}_{N_{\mathrm{W}}}^T]^T$. Pure channel information can be extracted by removing the contribution of the training sequence, which is supposed to be known at the UE. To do this, we compute $\hat{\mathbf{Y}} = \mathbf{Y}\mathbf{s}^{\mathrm{H}}/T_s$, whose expected value $\mathbb{E}[\hat{\mathbf{Y}}] = \sqrt{P_{\mathrm{t}}}\mathbf{W}^{\mathrm{H}}\mathbf{H}\mathbf{p}_{\mathrm{o}}$ gives direct access to channel-only information. The UE can now post-process the measurement $\hat{\mathbf{Y}}$ to obtain:

$$\overline{\mathbf{Y}} = \mathbf{A}_{\mathrm{UE}}^{\mathrm{H}}(\mathbf{W}^{\mathrm{H}})^{-1}\hat{\mathbf{Y}} \tag{4.4}$$

where $\mathbf{A}_{\mathrm{UE}} = [\mathbf{a}_{\mathrm{UE}}(\theta_1),\ \mathbf{a}_{\mathrm{UE}}(\theta_2),\ ...,\ \mathbf{a}_{\mathrm{UE}}(\theta_N)]$ is a spatial filter matrix and $\theta_i = \frac{2\pi i}{N}$, $i = 1, 2, ..., N$, is a set of N equally spaced discrete angles covering the 360° azimuthal domain. As evident from Eq. 4.4 and from the expression $\mathbb{E}[\overline{\mathbf{Y}}] = \sqrt{P_{\mathrm{t}}}\mathbf{A}_{\mathrm{UE}}^{\mathrm{H}}\mathbf{H}\mathbf{p}_{\mathrm{o}}$, the $N \times 1$ vector $|[\overline{\mathbf{Y}}]_i|^2$, for $i = 1, 2, ...N$, contains the expected signal power impinging on the UE from each angular direction. Such information can be directly used by the UE to estimate its L_{est} most suitable (i.e., the most powerful) directions of transmission/reception $\boldsymbol{\theta} = [\theta_1,\ ...,\ \theta_{L_{\mathrm{est}}}]$. Since the product $\mathbf{A}_{\mathrm{UE}}^{\mathrm{H}}(\mathbf{W}^{\mathrm{H}})^{-1}$ in Eq. 4.4 can be precomputed and stored in the UE memory, the computational cost to estimate $\overline{\mathbf{Y}}$ is just a matrix-vector multiplication.

4.6.2. Stage II: AP beam training

At the end of Stage I, the UE initiates the AP beam training stage. Specifically, the UE employs the HBF algorithms in [15] (considering 2-bit RF phase shifters) to design a multi-beam/multi-stream precoder $\mathbf{P}$ with the narrowest synthesizable beamwidth. Such a precoder is then used to simultaneously transmit orthogonal training sequences through the set of directions $\boldsymbol{\theta}$ estimated in Stage I. In order to reduce the inter-beam interference, Golay training sequences encoded by orthogonal Walsh spreading codewords are used. In fact, Golay sequences possess very good auto-correlation, which helps protecting the Walsh codes from losing orthogonality due to multipath. Concretely, for each transmit direction $i = 1, 2, ..., L_{\mathrm{est}}$, with $L_{\mathrm{est}} \leq N_{\mathrm{UE}}$, the UE emits a training signal $s_i[t]$, for $t = 1, 2, ..., T_s$. The overall set of transmitted symbols can be arranged into a matrix $\mathbf{S}$, with dimensions $L_{\mathrm{est}} \times T_s$, where the i-th row contains the time-domain sequence transmitted over the i-th direction. As in Stage I, the AP builds a $M_{\mathrm{AP}} \times M_{\mathrm{AP}}$ matrix $\mathbf{W}$ and configures its hybrid combiner to perform $N_{\mathrm{W}} = \lceil M_{\mathrm{AP}}/N_{\mathrm{AP}} \rceil$ signal measurements at successive instants:

$$\mathbf{Y}_i = \mathbf{W}_i^{\mathrm{H}}(\rho\mathbf{HPS} + \mathbf{N}),\ \ i = 1, 2, ..., N_{\mathrm{W}} \tag{4.5}$$

where $\rho = \sqrt{P_{\mathrm{t}}/L_{\mathrm{est}}}$ for equally distributed power within the streams. Similar to what is done in Stage I, the AP then concatenates the measurements, estimates $\hat{\mathbf{Y}} = \mathbf{Y}\sqrt{L_{\mathrm{est}}}\mathbf{S}^{\mathrm{H}}/T_s$, and processes it with the spatial filter matrix $\mathbf{A}_{\mathrm{AP}}$ to obtain $\overline{\mathbf{Y}}$. The procedure allows the AP to estimate its L_{est} most suitable directions of transmission/reception, $\boldsymbol{\phi} = [\phi_1,\ ...,\ \phi_{L_{\mathrm{est}}}]$. In order to establish a multi-beam, multi-stream data link between AP and UE after beam training, it is necessary that $L_{\mathrm{est}} \leq \min(N_{\mathrm{AP}}, N_{\mathrm{UE}})$. Note that the computational cost of estimating one rather than more than one suitable directions for communication via PE-Train is the same.

The major algorithmic steps required by the PE-Train procedure are summarized in Algorithm 2 for M_{AP} and M_{UE} integer multiples of N_{AP} and N_{UE} respectively. The beam training overhead, i.e., the total time required to com-

plete the PE-Train procedure, is given by the sum of Stage I and Stage II times: $\tau = \tau_{\text{PE-Train}} = T_{\text{slot}} \left(\lceil M_{\text{UE}}/N_{\text{UE}} \rceil + \lceil M_{\text{AP}}/N_{\text{AP}} \rceil \right)$. This represents a significant speed-up compared to exhaustive beam training, where the time required to complete the beam search is $\tau = \tau_{\text{EXH}} = T_{\text{slot}} N^2$, where N is the same angular resolution of our PE-Train strategy.

Algorithm 2 PE-Train protocol

Ensure: Pre-compute and store $\mathbf{A}_{\text{AP}}$ and $\mathbf{A}_{\text{UE}}$
Stage I: UE beam training
AP in omni mode w/ precoder $\mathbf{p}_{\text{o}} = [1, 0, 0, ..., 0]^T$
UE pre-computes and stores its $M_t extUE \times M_t extUE$ matrix $\mathbf{W}$
$N_{\text{W}} = M_{\text{UE}}/N_{\text{UE}}$
for $i \leq N_{\text{W}}$ **do**
    $\mathbf{W}_i = [\mathbf{W}]_{:,(i-1)N_{\text{UE}}+1:iN_t extUE}$
end for
for $i \leq N_{\text{W}}$ **do**
    UE measures $\mathbf{Y}_i = \mathbf{W}_i^{\text{H}}(\sqrt{P_{\text{t}}}\mathbf{H}\mathbf{p}_{\text{o}}\mathbf{s} + \mathbf{N})$
end for
$\mathbf{Y} = [\mathbf{Y}_1^T, \ ..., \ \mathbf{Y}_{N_{\text{W}}}^T]^T$; $\overline{\mathbf{Y}} = \mathbf{A}_{\text{UE}}^{\text{H}}(\mathbf{W}^{\text{H}})^{-1}\mathbf{Y}\mathbf{s}^{\text{H}}/T_s$
for $\ell \leq L_t extest$ **do**
    $k = \arg\max_j |[\overline{\mathbf{Y}}]_j|^2$; $\theta_\ell = 2\pi k/N$
    $\overline{\mathbf{Y}} = \overline{\mathbf{Y}} - [\mathbf{A}UE]_{:,k}^{\text{H}} \mathbf{A}UE [\overline{\mathbf{Y}}]_k$
end for
return $\boldsymbol{\theta} = [\theta_1, \ ..., \ \theta_{L_{\text{est}}}]$
Stage II: AP beam training
UE in multi-beam mode over $\boldsymbol{\theta}$ w/ hybrid precoder $\mathbf{P}$
AP pre-computes and stores its $M_t extAP \times M_t extAP$ matrix $\mathbf{W}$
$N_{\text{W}} = M_{\text{AP}}/N_{\text{AP}}$
for $i \leq N_{\text{W}}$ **do**
    $\mathbf{W}_i = [\mathbf{W}]_{:,(i-1)N_t extAP+1:iN_t extAP}$
end for
for $i \leq N_{\text{W}}$ **do**
    AP measures $\mathbf{Y}_i = \mathbf{W}_i^{\text{H}}(\sqrt{P_{\text{t}}/L_{\text{est}}}\mathbf{H}\mathbf{P}\mathbf{S} + \mathbf{N})$
end for
$\mathbf{Y} = [\mathbf{Y}_1^T, \ ..., \ \mathbf{Y}_{N_{\text{W}}}^T]^T$; $\overline{\mathbf{Y}} = \mathbf{A}_{\text{AP}}^{\text{H}}(\mathbf{W}^{\text{H}})^{-1}\mathbf{Y}\sqrt{L_{\text{est}}}\mathbf{S}^{\text{H}}/T_s$
for $\ell \leq L_t extest$ **do**
    $k = \arg\max_j |[\overline{\mathbf{Y}}]_{j,\ell}|^2$; $\phi_\ell = 2\pi k/N$
end for
return $\boldsymbol{\phi} = [\phi_1, \ ..., \ \phi_{L_{\text{est}}}]$

4.7. Probabilistic Beam Tracking (P-Track)

The PE-Train procedure described in the previous section can be used for initial access beam training, but can also be triggered periodically in order for AP and UE to update

their steering directions. However, harsh environments with frequent blockage could lead to an excessive number of beam training requests, which incur significant overhead and, consequently, reduced throughput. For this reason, efficient beam tracking strategies are required in order to rapidly refine the beam directions without resorting to full beam training.

In this section, we propose a probabilistic beam tracking (P-Track) mechanism which is able to track the mm-wave channel dynamics under node mobility (and steer the device beams accordingly) without requiring dedicated training slots. We assume a fixed AP and a moving UE that have just accomplished the PE-Train procedure and are communicating using pure data frames (i.e., without dedicated training slots) and highly directional beam patterns. For the sake of brevity, we consider only the UE beam tracking procedure, i.e., the procedure by which the UE exploits downlink data slots to refine its beam directions. An identical strategy is applied for AP beam tracking using uplink data slots. The P-Track strategy is based on a probabilistic model which does not require the devices to perform any spatial scanning during the ongoing data communication. It is able to track the most dominant directions of the mm-wave channel using just known portions of the data packet, e.g., the preamble. To do that, we require that two conditions are satisfied: (1) the preamble is correctly detected by the UE in at least one downlink slot within the frame; (2) in such a downlink slot, the UE can access the complex output $\mathbf{Y}_{\mathrm{RF}}$ from the RF combiner $\mathbf{C}_{\mathrm{RF}}$ (which can be done by saving the preamble samples right before the baseband combiner $\mathbf{C}_{\mathrm{BB}}$):

$$\mathbf{Y}_{\mathrm{RF}} = \mathbf{C}_{\mathrm{RF}}^{H} \left(\sqrt{P_{\mathrm{t}}/L_{\mathrm{est}}} \mathbf{HPS} + \mathbf{N} \right) \tag{4.6}$$

where P_{t} is the AP transmit power, L_{est} is the number of parallel data streams transmitted by the AP using the $M_{\mathrm{AP}} \times L_{\mathrm{est}}$ hybrid precoder $\mathbf{P}$, and $\mathbf{S}$ is the $L_{\mathrm{est}} \times T_{\mathrm{s}}$ matrix encompassing the packet preamble transmitted simultaneously by the AP over L_{est} directions. We assume that AP and UE are communicating using the narrowest beam patterns they are able to synthesize, which are steered towards the spatial directions estimated in the most recent PE-Train/P-Track procedure. In order to reduce the interference among the parallel streams, we assume that, for the preamble, the AP adopts Golay sequences encoded with orthogonal Walsh spreading codewords. The reason behind the choice of using $\mathbf{Y}_{\mathrm{RF}}$ for beam tracking, instead of the signal $\mathbf{Y} = \mathbf{C}_{\mathrm{BB}}^{H} \mathbf{Y}_{\mathrm{RF}}$ after the baseband precoder, is that the former includes much more information about the channel than the latter which is defined in a lower dimensional space. With $\mathbf{Y}_{\mathrm{RF}}$ it is possible to provide channel information for a wider angular domain compared to the very narrow angular sector covered by the actual data communication beam pattern.

4.7.1. Probabilistic optimization problem

Since the AP is transmitting relevant data to the UE using pure data frames, the UE cannot perform any beam scan, but it must keep its antenna steered towards the directions estimated in the most recent PE-Train/P-Track execution[1]. We propose a probabilistic estimation based on the analysis of the preamble signal $\mathbf{Y}_{\mathrm{RF}}$ received by the UE over the current antenna pattern in a downlink data slot. The UE is moving along a mobility pattern or *route*, so the objective is to update, in real time, its antenna pattern based on the estimation of a new set $\boldsymbol{\theta}^*$ of suitable directions. This can be translated into the problem of finding the $\boldsymbol{\theta}^*$ that maximizes the probability $P(\boldsymbol{\theta}^*|\mathbf{Y}_{\mathrm{RF}})$:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} P(\boldsymbol{\theta}^*|\mathbf{Y}_{\mathrm{RF}}) \tag{4.7}$$

which cannot be easily handled because both the prior distribution of the channel and the hybrid precoder $\mathbf{P}$ used by the AP are unknown. Applying the negative logarithm function and the Bayes theorem to the objective function in Eq. 4.7, and observing that the term $P(\mathbf{Y}_{\mathrm{RF}})$ has no impact on the optimization, we obtain:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \left[O_P(\boldsymbol{\theta}^*) + O_Y(\boldsymbol{\theta}^*)\right] \tag{4.8}$$

where we designate $O_P(\boldsymbol{\theta}^*) = -\log[P(\boldsymbol{\theta}^*)]$ as prior objective function and $O_Y(\boldsymbol{\theta}^*) = -\log[P(\mathbf{Y}_{\mathrm{RF}}|\boldsymbol{\theta}^*)]$ as measurement objective function. Intuitively, $O_Y(\boldsymbol{\theta}^*)$ models those changes in the mm-wave channel, due to user mobility, that are directly reflected into the received signal. In contrast, $O_P(\boldsymbol{\theta}^*)$ captures the uncertainties on the previous estimate and propagation phenomena which cannot be inferred from signal measurements.

As for the prior objective function $O_P(\boldsymbol{\theta}^*)$, a simple approach is to define the prior distribution $P(\boldsymbol{\theta}^*)$ as a set of independent Gaussian distributions for each θ_ℓ, $\ell = 1, 2, ..., L_{\mathrm{est}}$, with mean $\overline{\theta}_\ell$ equal to the previous estimated direction and standard deviation $\sigma_{\theta_\ell} = \overline{\sigma}_{\theta_\ell} + f(\mathrm{SNR})$. The term $\overline{\sigma}_{\theta_\ell}$ models the uncertainty of the estimation due to continuous angular variations induced by the UE movements (device rotation and translation), while $f(\mathrm{SNR})$ is a convenient, monotonically decreasing function of the SNR. This latter term models the uncertainty of our previous estimation, which is largely affected by the received signal quality (i.e., the greater the SNR, the smaller the uncertainty). After removing constant terms, the following expression for $O_P(\boldsymbol{\theta})$ can be

[1]Here, the problem is more complex than channel estimation in MIMO systems. In fact, while MIMO transceivers (based on DBF) are able to instantaneously collect full spatial information about the channel, HBF transceivers would usually need several estimation steps, during which the beams are steered to scan different spatial directions.

derived:

$$O_P(\boldsymbol{\theta}^*) = -\sum_{\ell=1}^{L_{\text{est}}} \log P(\boldsymbol{\theta}^*) = \sum_{\ell=1}^{L_{\text{est}}} \frac{(\theta_\ell - \overline{\theta}_\ell)^2}{2\sigma_{\theta_\ell}^2} \tag{4.9}$$

The measurement objective function $O_Y(\boldsymbol{\theta}^*)$ can be expressed as:

$$O_Y(\boldsymbol{\theta}^*) = \frac{\left\|\mathbf{D}^{-1}\mathbf{V}^H\hat{\mathbf{Y}}_{\text{RF}} - \mathbf{U}^H\mathbf{A}\theta\mathbf{M}\right\|_F^2}{2\sigma^2} \tag{4.10}$$

where $\mathbf{UDV}^H$ is the economic singular value decomposition (SVD) of $\mathbf{C}RF$, $\hat{\mathbf{Y}}_{\text{RF}} = \mathbf{Y}_{\text{RF}}\sqrt{L_{\text{est}}}\mathbf{S}^H/T_s$ is the redundancy-free preamble signal from the RF precoder, σ^2 is the noise power, $\mathbf{A}_\theta$ is a $M_{\text{UE}} \times L_{\text{est}}$ matrix such that $[\mathbf{A}_\theta]_{:,\ell} = \mathbf{a}_{\text{UE}}(\theta_\ell)$, and $\mathbf{M} = (\mathbf{A}_\theta^H\mathbf{U}\mathbf{U}^H\mathbf{A}_\theta)^{-1}\mathbf{A}_\theta^H\mathbf{U}\mathbf{D}^{-1}\mathbf{V}^H\hat{\mathbf{Y}}_{\text{RF}}$ is a $L_{\text{est}} \times L_{\text{est}}$ matrix.

4.7.2. Problem solution

The problem of estimating $\boldsymbol{\theta}^*$ can be transformed into the problem of minimizing the overall objective function $O(\boldsymbol{\theta}^*) = O_P(\boldsymbol{\theta}^*) + O_Y(\boldsymbol{\theta}^*)$, which can be verified to be non-convex even for $L_{\text{est}} = 1$. The minimization strategy we propose is as follows. Two suitable initial guesses of $\boldsymbol{\theta}^*$, one for $O_P(\boldsymbol{\theta}^*)$ and one for $O_Y(\boldsymbol{\theta}^*)$, are selected. A good initial guess is $\overline{\boldsymbol{\theta}}$, since it minimizes $O_P(\boldsymbol{\theta}^*)$. A second good starting point representing a suitable guess for $O_Y(\boldsymbol{\theta}^*)$ is the set of directions with maximum received signal power. Starting from these two initial solutions, since we do not know a priori if $O_Y(\boldsymbol{\theta}^*)$ has more weight than $O_P(\boldsymbol{\theta}^*)$ on the total objective function $O(\boldsymbol{\theta}^*)$, we compute a few steps of gradient descent for both guesses, select the best solution, and proceed with a more accurate gradient descent to refine the estimation[2].

4.8. Numerical evaluation

In this section, we carry out numerical experiments to assess the performance of the PE-Train and P-Track strategies. We first describe the simulator we developed to evaluate mm-wave indoor scenarios with mobility. Then, we describe the simulation scenario, based on which we conduct a numerical evaluation to compare the performance of our strategies against existing beam search approaches in the literature.

4.8.1. Simulator overview

The main functional blocks of our mm-wave simulator are outlined in Fig. 4.3. The simulator allows to draw any UE route in a given scenario via a graphical user interface. Based on the route length L_{route} and the selected UE speed v, the number of frames

[2]Details about the derivation of the objective functions and their gradient can be found in the appendices of the extended chapter available from `https://arxiv.org/abs/1612.07957/`.

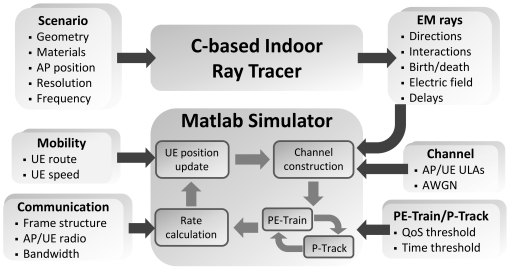


Figure 4.3: Overview of the mm-wave indoor simulator used to assess the performance of beam training and tracking strategies under node mobility.

$\nu = \lceil L_{\text{route}}/(vT) \rceil$ required by the simulation are computed. Implicitly, this creates a direct correspondence between the UE position on the route and the current time slot. We exploit such correspondence to compute, at each time slot, the channel matrix $\mathbf{H}$ in Eq. 4.1 through a proprietary ray tracing program written in C (see the next subsection for details). The PE-Train and P-Track strategies are implemented in Matlab and validated as follows (see Fig. 4.4 for an example use case). In the UE starting position, the simulator allocates a training frame to perform the initial access PE-Train procedure and establish a multi-beam/multi-stream directional link between AP and UE. Specifically, AP and UE employ the HBF algorithms in [15] (with 2-bit phase shifters) to design the narrowest synthesizable beams, steered toward the L_{est} estimated directions, to be used concurrently during the data transmission phase. From here on, as long as some QoS or timing conditions (defined later) are satisfied, pure data frames are sent. In each frame, a downlink data slot and an uplink data slot are used by UE and AP respectively to perform the P-Track estimation and update, accordingly, their steering directions. If the QoS and timing conditions are not satisfied, a training frame is allocated in order to perform a thorough and accurate beam search via PE-Train. Concretely, for each frame, we compute two types of sum-rate capacity (hereafter referred to as simply rate):

- The actual rate R_{D} within the current frame is obtained by computing the rate in each data transmission slot and averaging it *over the number of data slots in the frame.*

- The average rate $\overline{R}_{\text{D}}$ over multiple data frames is obtained by averaging the rate R_{D} *over the number of frames* since the latest PE-Train execution.

As for the QoS and timing conditions, we assume that a new PE-Train execution is triggered periodically every ξ pure data frames or when the actual rate R_{D} within the current frame is below a certain percentage of the average rate $\overline{R}_{\text{D}}$, i.e., when $R_{\text{D}} < \lambda \overline{R}_{\text{D}}$, with $0 < \lambda \leq 1$.

4.8.2. Ray-tracing module

Geometry-based stochastic channel models in the literature are not suitable to represent real environments, especially time-varying scenarios with user mobility, because they are drop based, meaning that, at every time slot, the channel parameters are randomly generated for each AP-UE link. In order to analyze how the PE-Train and P-Track strategies would perform in a real scenario, we rely in this chapter on a custom ray-tracing program in C which allows to deterministically evaluate the non-stationary characteristics of the propagation channel, including LOS and NLOS transitions, shadowing, mobility effects, environment dynamics, and blockage. The key benefit of this approach compared to pure statistical models is its inherent support for spatial consistency which allows smooth and continuous time evolution of channel parameters. As shown in Fig. 4.3, the ray tracer takes in input the scenario geometry, the electromagnetic characteristics of walls and objects, the AP position, the carrier frequency, and the desired ray-launching resolution (i.e., the angular separation between two adjacent rays launched). The rays are launched to cover the entire 360° azimuthal domain and their evolution is computed taking into account reflections, transmission through objects/walls, and diffraction. At each time slot (i.e., at each UE location), the output from the ray tracer is used to update, in real time during the simulation run, the channel matrix in Eq. 4.1. To do this, we construct, around the current UE location, a reception sphere with radius proportional to the unfolded path length from AP to UE and the ray-launching resolution [65]. If a ray intersects the sphere, it is taken as contributing to the received signal, otherwise it is discarded. The ray clustering effect revealed by the experimental campaign in [64] is obtained by accounting for the contribution of the ten most powerful rays around the one falling within the reception sphere. This analysis provides the channel AoDs/AoAs at each UE location, while the wave-carried electric field of each ray is used to compute the corresponding complex gain $\alpha_{k\ell}$. Since both phase and delay of each ray are taken into account, the Doppler shift effect is inherently included in the computation.

4.8.3. Simulation scenario

For the performance evaluation, we replicate the 20×20 m^2 office-like layout considered in [8]. As shown in Fig. 4.5, it consists of several walls/partitions composed by three different materials, namely concrete, glass, and plasterboard with a thickness of 10 cm, 3 cm, and 5 cm respectively, to model a realistic environment. The scenario geometry and the dielectric properties of materials are given as input to the ray tracer for reflection/transmission/diffraction coefficient calculation. In order to reproduce human blockage effects in a crowded environment, we randomly place in the scenario 20 blocks of size 50×50 cm^2 with dielectric properties taken from the experiments on human tissues in [66]. The mm-wave network consists of a fixed AP, installed in the center of the

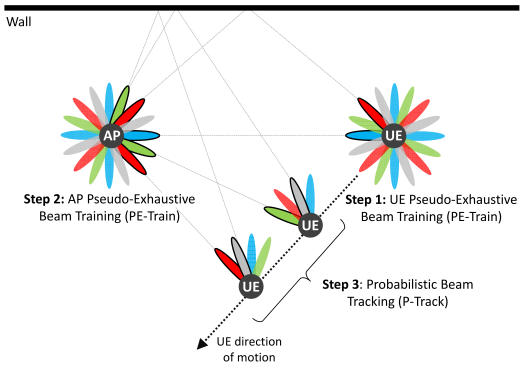


Figure 4.4: Typical use case scenario for our beam training strategies. PE-Train is used for initial access beam training and triggered periodically to update the AP-UE steering directions. P-Track, instead, is used to track the mm-wave channel dynamics under node mobility and steer the device beams accordingly.

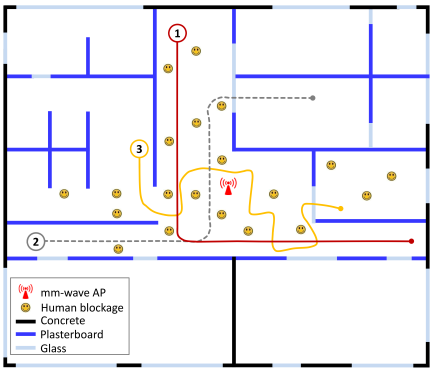


Figure 4.5: Office-like simulation scenario with a fixed AP and a mobile UE walking through three different routes.

room, and a mobile UE walking through three different routes with increasing complexity, namely Route #1 (straight lines with one turn), Route #2 (straight lines with two turns), and Route #3 (curved lines with nine turns). For each route, we assume that the UE is initially located at the starting point (represented by the numbered label in Fig. 4.5) and is moved with speed v=2 m/s and position update rate of T_{slot}=100 μs. Note that we consider an orientation-unaware UE, i.e., a UE turn causes the beam orientation to change accordingly.

4.8.4. Results

In this sub-section, we assess the performance of our PE-Train and P-Track strategies using our Matlab/C mm-wave simulator. For performance comparisons, we implement from scratch the solutions proposed in [8] and [10], and a simplified version of the IEEE 802.11ad beam training protocol. As for the 802.11ad implementation, we assume that, whenever the QoS falls below a certain threshold, a training frame is allocated to perform

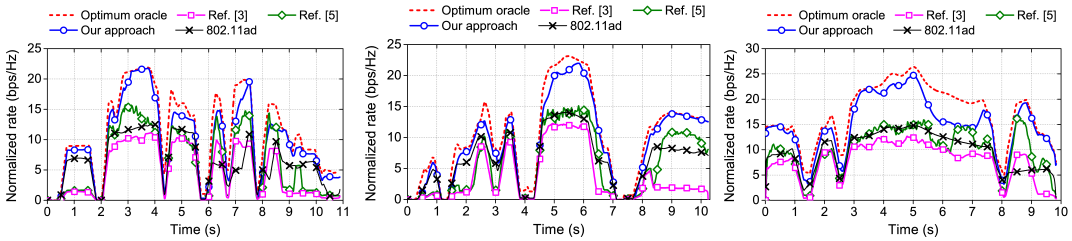
a sector level sweep (SLS), while a beam refinement (BRF) procedure for fine grained calibration of the current beams is done every two pure data frames — we verified that this setting provides the best performance for our 802.11ad implementation. We run [10, Algorithm 2], [8, Algorithm 1], and the 802.11ad protocol for different QoS thresholds, and select, for each UE route and for each algorithm, the threshold which provides the best performance. All the simulations consider a 60 GHz carrier frequency with 500 MHz channel bandwidth and a transmit power at both devices equal to 30 dBm. For our PE-Train and P-Track strategies, we assume that 128-length Golay sequences with 4-length Walsh codes are used in both the training slots and the preamble. Concerning the P-Track solution outlined in §4.7.2, we use 10 gradient descent steps for the preliminary estimation and 50 steps for the final refinement. All the results are averaged over 1000 simulations for each combination of UE route and beam search strategy.

In Fig. 4.6, we plot the evolution over time of the achievable normalized rate per frame when different beam search strategies are adopted. We recall that after each beam training/tracking execution, the AP and UE multi-beam antenna patterns are updated according to the new L_{est} estimated steering directions. In the data transmission phase, AP and UE communicate with the narrowest beam patterns they are able to synthesize and transmitting/receiving $N_{\text{S}} = L_{\text{est}}$ parallel data streams over L_{est} channel paths. The adopted beam patterns are reflected into the $M_{\text{AP}} \times N_{\text{S}}$ data precoder $\mathbf{P}D$ and the $M_{\text{UE}} \times N_{\text{S}}$ data combiner $\mathbf{C}D$ at the AP and UE respectively. The achievable normalized rate is then calculated as follows:

$$R = \frac{T_{\text{slot}}}{T} \sum_{i \in \mathcal{D}} \log_2 \left| \mathbf{I}_{N_{\text{S}}} + \frac{P_{\text{t}} (\mathbf{U}^H \mathbf{H}_i \mathbf{P} D)(\mathbf{U}^H \mathbf{H}_i \mathbf{P} D)^H}{N_{\text{S}} \sigma^2} \right| \tag{4.11}$$

where $\mathbf{I}_{N_{\text{S}}}$ is the $N_{\text{S}} \times N_{\text{S}}$ identity matrix, $\mathcal{D}$ is the set of data slots in the frame, $\mathbf{H}_i$ is the channel matrix in the i-th slot, σ^2 is the average noise power, and $\mathbf{U}$ is left singular vector matrix of the "economic" SVD decomposition of $\mathbf{C}D$.

As evident from the plots, our simulator is able to well describe the effect of human blockage at mm-wave frequencies. The blockage, which intermittently appears and breaks the LOS link between AP and UE, is clearly visible from the rate suddenly dropping down to very small values. In case even the optimum rate sharply drops to zero, there is no possibility to establish a connection between AP and UE, i.e., the UE is in outage. In all the analyzed UE routes, our strategy, based on the alternation of PE-Train and P-Track according to the selected QoS and timing thresholds, yields performance very close to the optimum oracle algorithm (only 10% rate difference on average), and significantly outperforms both the 802.11ad approach and beam search proposals in the literature. Quantitatively, based on the results in Fig. 4.6, we report in Table 4.1 the average training overhead $\overline{\tau}$ per frame, calculated by dividing the total number of training slots used in each route by the number of allocated frames and multiplying by the time slot



(a) UE Route #1 (b) UE Route #2 (c) UE Route #3

Figure 4.6: Normalized rate over time for the three UE routes: comparison among the optimum oracle solution, the proposed strategies with λ=0.3 and ξ=30 frames), two beam training algorithms in the literature, and the baseline IEEE 802.11ad protocol.

Table 4.1: Training overhead and rate gain provided by our approach

Algorithm	Training overhead $\overline{\tau}$ (ms)			Percentage rate gain of PE-Train+P-Track		
	R#1	R#2	R#3	R#1	R#2	R#3
PE-Train+P-Track	0.06	0.04	0.06	–	–	–
[10, Algorithm 2]	4.62	4.35	3.07	56.1	54.8	47.7
[8, Algorithm 1]	3.58	4.67	1.66	108.2	150.1	94.1
IEEE 802.11ad	0.87	0.72	0.67	44.0	39.8	50.2

duration T_{slot}. As shown, our PE-Train and P-Track strategies provide a one to two orders of magnitude reduction in training overhead. This translates, approximately, to an average rate increase of 48% to 150% compared to state-of-the-art solutions and of 40% to 50% over the 802.11ad standard. Although these results are obtained for λ=0.3 and ξ=30 frames, we verified via grid search that the performance achieved by our approach does not vary significantly with the selected QoS and timing thresholds. In fact, both the continuous execution of P-Track and the low-overhead PE-Train make the choice of λ and ξ not determinant on the overall performance. We repeated the evaluation for a vast range of simulation scenarios and UE routes (not reported here due to space constraints), where we noted that the achieved performance does not differ considerably from that presented here. We also verified that our strategies perform well in environments with less blockage. For example, in the scenario of Fig. 4.5 without human blockage, we obtained a 25% to 170% rate increase over existing approaches. It is worth emphasizing that, differently from [10] and 802.11ad, our strategies do not require any dedicated channel for the receiver to feed back the training results to the transmitter. Furthermore, we adopt HBF with only 2-bit phase shifters as opposed to the 7-bit ones used in [10] and the idealized, unconstrained ABF transceiver considered for the implementation of [8] and 802.11ad.

4.9. Summary

In this chapter, we investigated the problem of beam training and tracking in directional mm-wave networks with mobility. Exploiting the ability of HBF transceivers to collect channel information from multiple spatial directions simultaneously, we designed two strategies (one deterministic for beam training and one probabilistic for beam tracking) to rapidly estimate the most suitable transmit/receive directions at the AP and UE sides. Simulation results, obtained by a custom simulator based on ray-tracing channel modeling, demonstrated that the proposed solution is effective to keep the average communication rate only 10% below the optimal bound. Compared to both the IEEE 802.11ad standard and the state of the art, our solution provides a 40% to 150% performance increase while at the same time using lower complexity hardware.

Part IV

Measuring the channel radiation properties from the communication

Chapter 5

Beam-training on commercial off-the-shelf devices

5.1. Challenges

Designing wireless communication systems that operate at very high frequencies such as the 60 GHz band is technically challenging. For instance, phased antenna arrays are hard to manufacture [4, 5] and phase noise plays a significant role [6, 7]. As a result, system designers often resort to simple yet sub-optimal solutions. This is the case for certain mechanisms such as analog beamforming in the IEEE 802.11ad standard for communication at 60 GHz. Beamforming is crucial for such systems since devices must use directional communication to overcome the very high path loss in the millimeter-wave band and reach the receiver. However, the standard is limited to a basic beamforming mechanism based on a codebook of generic beam patterns. Such patterns are envisioned to have the shape of uniform sectors and cover the entire azimuth range of the device. That is, instead of beamforming towards a specific direction, devices choose the beam pattern out of their codebook which provides the highest gain in that direction.

Beam training, also called SLS in infrastructure-based 802.11ad networks, works as follows. The Access Point (AP) transmits beacon messages using each of its available beam patterns sequentially, while the station (STA) listens with a quasi-omnidirectional beam pattern. After that, the STA repeats the same process but includes in each of its messages the identifier of the beam pattern that it received best from the AP. Finally, the AP replies with the identifier of the best beam pattern of the STA in a dedicated control message. While this mechanism is straightforward, it clearly does not exploit the full potential of the antenna array of IEEE 802.11ad devices. First, none of the available beam patterns in the codebook may steer exactly towards the receiver. Second, strong reflectors in the environment may remain unused even though they could contribute to the received signal strength. Third, existing reflections may result in destructive interference

at the receiver, causing significant harm to the communication. If devices were to adapt their beam patterns to the specific environment in which they operate, they could easily mitigate the above issues. However, this requires full Channel State Information (CSI) at the transmitter, which is particularly challenging to obtain in millimeter-wave systems. The reasons are twofold. First, the feedback overhead of full CSI is significant. Second, devices must probe enough orthogonal beam patterns to cover all of the dimensions of the channel. The channel information extracted from all of the probes must be coherent in phase. In particular, achieving such phase coherence over different probes in consumer-grade devices is non-trivial.

5.2. Related Work

The IEEE 802.11ad standard [1, 67] deploys a two-stage beamforming protocol that consists of the **sls!** and an optional beam refinement phase (BRP). The SLS operates with predefined beam patterns, so-called sectors, and trains the antenna steering by sweeping through all available sectors while sending probe frames. In the BRP, devices continuously adapt their antenna sector to compensate for mobility and small channel distortions. In IEEE 802.11ay—the upcoming millimeter-wave standard for Wi-Fi networks and the successor of IEEE 802.11ad—several refinements are introduced to make beam training more efficient with shorter frames and multiple channels [68, 69]. For a low number of sectors, the sector sweep completes in a reasonable amount of time. However, its complexity increases linearly with the number of probed sectors. Thus, either the number sector is low and thus beam patterns become wide and irregular, or a high training overhead must be taken into account. Common COTS devices typically use a fixed set of sectors and do not implement the BRP [70].

Several alternatives have been proposed that allow devices to steer their antennas efficiently. State-of-the-art protocols and techniques are surveyed in [71]. Hierarchical beam training with codebook structures of different width allows to refine the beam patterns iteratively [72–75]. Organizing beam patterns in a tree hierarchy achieves logarithmic search complexity but introduces an additional communication overhead due to the required feedback for each of the multiple probing rounds. Compressive sensing based approaches derive the CSI by selecting pseudo-random probing beams [76–78] and determine an optimal beam after only one training round. However, complex channel measurements are not available on common COTS devices. Thus, [79] proposed a variant based on non-coherent measurements to determine the line of sight path without phase information. [70] integrated such a variant of compressive sensing with the sector sweep on COTS devices. Most beam-steering approaches use codebooks that consists of beams with specific geometrical contexts such as beam-width or beam-direction. To generate such beams, it is required to understand the antenna radiation characteristics, especially

since the wavelength of mm-waves is of the order of millimeters. Small variations in the electrical components create significant phase shifts or irregular absorption. To tackle these impairments, antenna calibration under controlled conditions is required and must be incorporated in the codebook. Our approach, in contrast, does not rely on antenna calibration. We optimize the beam-steering based on current channel conditions without any geometrical properties.

In multi-antenna scenarios with multiple RF-chains, additional performance gains are expected. Compressive sensing allows to estimate the channel with only a few probes efficiently [80, 81] and selects proper beams for multiple users [82]. Exploiting multiple antennas also improves spatial reuse [83, 84]. With multiple RF-chains, hybrid beamforming allows to take channel measurements in parallel and decreases the overhead or increases the channel estimation quality. Unfortunately, at the time of writing, no COTS device supports multiple RF-chains or hybrid beamforming.

Algorithms such as [85] improve resilience to blockage by instantaneously predicting the availability of alternative paths. The mobility of devices is considered in [86], in which correlations of channel profiles at nearby locations allow to reconstruct the channel profile when devices start to move. Doing so, the proposed protocol continuously realigns the links of mobile devices without explicit channel scanning. In [87], E-MI senses the environment and traces back the LOS and NLOS paths between devices by applying ray-tracing techniques. This allows to resolve the significant reflections in the paths, predict the quality of different links, and optimize the deployment of APs. Moreover, adding low pilot overhead, as done in [88], increases the channel estimation accuracy. Beam steering algorithms also benefit from multi-band connectivity. Wi-Fi connections in the lower frequency bands such as 2.4 and 5 GHz can assist millimeter-wave link adaption [89]. Angular estimation of the communication direction in the lower frequency bands can already provide a coarse direction to steer the millimeter-wave beam [90]. Tracking of devices without probing overhead is possible by explicitly generating multi-lobe beam patterns that detect if devices move out of the beam's coverage area [27]. To lower the risk of outage during mobility, the authors of [91] propose to continuously adjust the width of the beam pattern as well as the selected modulation and coding schemes.

While millimeter-wave research is well advanced, wide-spread practical testbed and evaluation systems are rare. Available IEEE 802.11ad chipsets expose a single RF-chain, provide limited control, and only allow to extract a few low-layer parameters [92, 93]. As a result, researchers tend to create evaluation platforms that are equipped with directional horn antennas [85, 91, 94, 95] or custom phased antenna arrays [79, 96]. Due to the high amount of hardware customization, such platforms complicate the reproducibility of results. They typically lack IEEE 802.11ad compliant implementations, and their performance and behavior differ fundamentally from that of COTS devices. Commercial prototyping platforms such as that in [97] allow to evaluate different protocols but are not

affordable for larger deployments and are impractical to analyze realistic millimeter-wave use-cases. In prior work, we extended the device driver to control the Wi-Fi interface in COTS devices [89, 98] and added new features to the firmware running on the IEEE 802.11ad Wi-Fi chip [70] to extend the access to low-layer statistics and perform practical experiments. However, all of these works based on COTS devices lack access to the full CSI of the millimeter-wave channel.

Accessing CSI on Wi-Fi systems in the lower frequency bands is comparatively easier. These systems apply digital beamforming which allows for full channel estimation. Channel measurements are, for example, required to compensate the channel effects and apply beamforming in multi-antenna systems. Prototyping platforms such as WARP [99] allow estimating the complex channel gains with reasonable overhead in multi-antenna systems. The CSI can also be directly extracted from some specific COTS wireless interface cards [100]. Unfortunately, such approaches require community projects and a detailed understanding of the wireless chip. Notably, even for Wi-Fi that has been around for two decades, only a few chipsets to date provide CSI that is accessible to researchers. Manufacturers of commodity Wi-Fi equipment typically keep their systems closed. As of now, no such access to CSI is available for any IEEE 802.11ad device.

5.3. Contribution

In this chapter, we design a mechanism that enables full CSI extraction on consumer-grade Commercial Off-The-Shelf (COTS) IEEE 802.11ad devices using only non-coherent signal-to-noise ratio (SNR) measurements. It allows us to fully exploit the capabilities of phased antenna arrays in such devices. Our Adaptive Codebook Optimization (ACO) mechanism probes the channel using carefully engineered beam patterns that allow devices to extract both amplitude and phase information from simple SNR readings. Obtaining the SNR is required by the IEEE 802.11ad standard to select the best beam during the sector sweep. Hence, measurements of this metric are supported on all IEEE 802.11ad devices which makes ACO widely applicable. Specifically, ACO measures the *relative phase* among the elements of the antenna array of a node. Figure 5.1 depicts a toy example of ACO for an antenna array with two-bit phase shifters per antenna element. ACO defines one of the elements of the array as a reference R. To measure the relative phase to an element E, it switches on R and E, keeping all other elements off. It then transmits one probe frame for each of the four possible values of the phase shifter at E. As shown in Figure 5.1, the complex gains of R (blue) and E (red) sum up (sum shown in black) for each of the four probes. However, the receiver only observes the resulting amplitude in terms of received signal strength. Since the four probes cover a 2π phase range, the four powers must lie on a sinusoidal curve. Our key insight is that the initial phase of that curve is directly related to the relative phase among R and E. ACO exploits

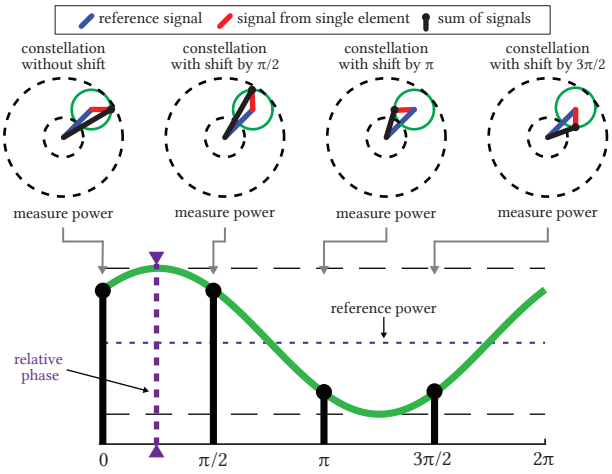


Figure 5.1: ACO toy example. The upper part shows the signal constellations with four different phase shifts at a single antenna element, whereas the lower part illustrates how ACO derives the relative phase using a sinusoidal curve.

this property using Fourier analysis to reconstruct the sinusoidal curve and compute the phase shift δ. Figure 5.1 shows an example for $\delta = \pi/4$. Repeating the process for each element of the antenna array, ACO can compute the full CSI.

The above technique is the core of ACO. Using the computed CSI, ACO derives beam patterns that maximize the SNR, exploit reflections, and prevent destructive interference. However, the toy example in Figure 5.1 may raise a number of concerns which we discuss below. In this chapter, we develop ACO to address these concerns effectively.

1. **Overhead.** ACO requires at least four probes per antenna element, which increases the probing overhead compared to a codebook of generic beam patterns. However, in practice, only a subset of the antennas of the array has a significant impact. We show that obtaining the CSI for just a subspace of the channel is sufficient in most cases. As a result, (a) the probing overhead is similar to that for pre-defined generic beam patterns (i.e., IEEE 802.11ad SLS), and (b) the feedback overhead is limited.

2. **Performance.** ACO transmits using only two antenna elements simultaneously during the probing phase. For long links, the resulting transmit power may not be sufficient to reach the receiver, and thus the CSI measurement may fail. To address this issue, we extend the basic approach in Figure 5.1 to use not just single antenna elements but entire beam patterns. This enables ACO to operate also on low SNR links.

3. **Accuracy.** In our toy example, ACO probes each possible value of the phase shifters. The resulting overhead is limited for two-bit phase shifters but would increase significantly for higher-resolution hardware. However, ACO only needs four samples to reconstruct the sinusoidal curve in Figure 5.1. To keep the overhead low,

ACO only probes four phase shifts even if more combinations are possible.

We implement ACO on commodity COTS hardware. To this end, we gain full access to the beamforming control of the TP-Link Talon AD7200 60 GHz router. This device features a Qualcomm QCA9500 chipset that implements the IEEE 802.11ad standard. We disassemble the phased antenna array of the device to understand its structure and experimentally reconstruct the antenna weighting network. This enables us to set arbitrary beam patterns on the antenna. However, *ACO itself does not require any information about the antenna structure* to operate. Obtaining the weighting network of the QCA9500 60 GHz solution is a valuable contribution to the community in itself since Qualcomm's chipset has become a major research platform in the field [70,98]. We modify the sector sweep behavior of the router to include the custom beam patterns that enable ACO. We embed ACO in the regular operation of the router, that is, *we do not obtain our results in post-processing* but as part of the normal data transmissions in our testbed. This enables us to quantify the performance of ACO in terms of SNR, data rate, and throughput using the Transmission Control Protocol (TCP). Our contributions are as follows:

- We design ACO, a method to obtain full CSI in terms of phase and amplitude using only SNR measurements. The key to our method is obtaining SNR values that resemble a discrete Fourier decomposition.
- We implement our method on commodity 60 GHz hardware. To this end, we disassemble a COTS 60 GHz router and modify its sector sweep behavior.
- We evaluate ACO in a real-world office environment and show that it achieves on average 2.5× higher SNR as well as up to 2× higher throughput.

5.4. Adaptive Codebook Optimization

In the following, we define the architecture and the operation of ACO. For illustration purposes, we consider an infrastructure-based IEEE 802.11ad network. However, ACO is not limited to such networks but is applicable to any millimeter-wave link. In our case, both the AP and the STA can benefit from ACO for beamforming.

Figure 5.2 depicts a schematic overview of the operation of ACO, which consists of two phases: initialization and continuous adaptation. In the initialization phase, devices discover each other and establish a connection using the default IEEE 802.11ad procedure based on generic beam patterns. After that, ACO exhaustively probes all of the antenna elements of the array. This one-time overhead allows ACO to obtain the full CSI and to determine which of the elements contribute most to it. Based on this information, ACO restricts the following channel measurement to those elements. This reduces the

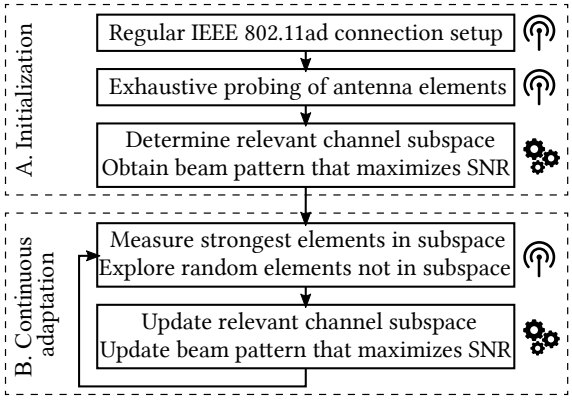


Figure 5.2: ACO operation with processing (gears symbol) and communication (antenna symbol) steps.

probing overhead significantly and allows ACO to focus on the subspace of the channel which is relevant for a certain physical environment. To determine which and how many antenna elements N to use, ACO maximizes the expression in Equation 5.1, where a_k is the amplitude of antenna k and antenna elements are ordered by strength.

$$N = \arg\max_{n} \frac{\left(\sum_{k=1}^{n} a_k\right)^2}{n} \tag{5.1}$$

To complete the initialization phase, ACO obtains the beam pattern that maximizes the SNR within the computed subspace and uses it for regular communication. After that, ACO proceeds to the continuous adaptation phase, which allows ACO to readjust the beam pattern as the CSI and the environment change during on-going communication. Similarly to IEEE 802.11ad, ACO probes the channel periodically but limits the channel measurement to the strongest antenna elements and thus to the current subspace. However, the subset of relevant antenna elements may change if the environment changes due to, e.g., user mobility. To adapt to such changes, ACO divides its periodic channel probing sweep into a measurement part and an exploration part. The channel measurement is performed over the subspace of strongest antennas, whereas the exploration part probes a subset of randomly chosen antenna elements to identify potential changes in the relevant channel subspace. As shown in Figure 5.3, ACO can choose the overall duration of the sweep such that its duration is equivalent to a regular IEEE 802.11ad sector sweep. Still, if full CSI is desired, ACO can easily be reconfigured to probe all antenna elements at the expense of higher overhead. After each periodic sweep, ACO recomputes the beam pattern that maximizes the SNR and data communication continues. The complexity of the processing steps shown in Figure 5.2 is negligible due to the very high computational efficiency of our method (see Section 5.6). This enables ACO to operate both on powerful APs as well as resource-constrained STAs.

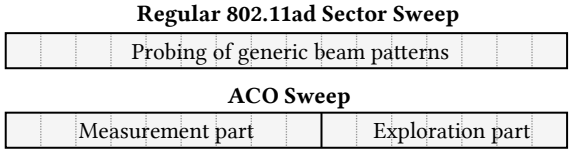


Figure 5.3: ACO sweep compared to the IEEE 802.11ad sector sweep.

5.5. System Model

To formally define the operation of ACO, we introduce our system model in this section. Specifically, we define the following antenna and channel models.

5.5.1. Antenna Model

For both the transmitter and the receiver, we consider a phased antenna array that implements analog beamforming. That is, all antenna elements are connected via a network of amplifiers and phase shifters to a single **rf!** chain. This type of beamforming is widely used in COTS 60 GHz hardware since more advanced architectures such as digital or hybrid beamforming are not cost-efficient. We allow for both regular and irregular antenna element layouts.

5.5.2. Channel Model

Given a transmitter TX that sends a signal x towards a receiver RX, the received signal can be expressed as

$$\mathbf{y} = \mathbf{c}^H \mathbf{H} \mathbf{p} \mathbf{x} + \mathbf{c}^H \mathbf{N}, \tag{5.2}$$

where $\mathbf{c}$ and $\mathbf{p}$ are the complex gains of the antennas at RX and TX respectively, and $\mathbf{N}$ is uncorrelated additive white noise received at each antenna with variance σ^2. The channel $\mathbf{H}$ reflects the propagation effects. We consider the geometric channel model that is widely used in the millimeter-wave literature. Given a set of L different paths from the transmitter to the receiver, we can represent the channel as:

$$\mathbf{H} = \sqrt{P_t} \sum_{l=1}^{L} \alpha_l \mathbf{a}_{\mathrm{RX}}(\mathbf{v}_l) \mathbf{a}_{\mathrm{TX}}(\mathbf{w}_l)^H, \tag{5.3}$$

where P_t is the transmit power, L is the number of paths, α_l is the complex gain, $\mathbf{v}_l, \mathbf{w}_l$ are unitary vectors of arrival and departure for the l^{th} path, and $\mathbf{a}_{\mathrm{TX}} : \mathbf{S} \to \mathbb{C}^{N_{\mathrm{TX}} \times 1}$ as well as $\mathbf{a}_{\mathrm{RX}} : \mathbf{S} \to \mathbb{C}^{N_{\mathrm{RX}} \times 1}$ are the steering vectors at TX and RX respectively. We assume no prior knowledge regarding the steering vectors $\mathbf{a}_{\mathrm{TX}}$ and $\mathbf{a}_{\mathrm{RX}}$ because they are typically unknown for COTS devices. Further, we denote the receiving channel for a fixed $\mathbf{c}$ as $\mathbf{h}_{\mathrm{RX}} = \mathbf{c}^H \mathbf{H}$. Thus, the received signal can be written as $\mathbf{y} = \mathbf{h}_{\mathrm{RX}} \mathbf{p} \mathbf{x} + \mathbf{c}^H \mathbf{N}$. Conversely, the transmitting channel for a fixed $\mathbf{p}$ is $\mathbf{h}_{\mathrm{TX}} = \mathbf{H}\mathbf{p}$ and hence $\mathbf{y} = \mathbf{c}^H \mathbf{h}_{\mathrm{TX}} \mathbf{x} + \mathbf{c}^H \mathbf{N}$. Based

on these definitions, we can compute the received signal power as $|\mathbf{h}_{\mathrm{RX}}\mathbf{p}|^2$ or $|\mathbf{c}^H\mathbf{h}_{\mathrm{TX}}|^2$, and the received noise power as $\|\mathbf{c}\|^2\sigma^2$.

While COTS devices may provide SNR or RSSI values, they typically do not expose detailed signal power measurements. Since $\|\mathbf{c}\|$ is a constant value, we define $\|\mathbf{c}\|^2\sigma^2 = 1$ as noise power reference. Using this definition along with the well-known formula of the SNR in logarithmic units, we obtain that the received signal power is $|\mathbf{h}_{\mathrm{RX}}\mathbf{p}|^2 = 10^{\frac{\mathrm{SNR}}{10}}$.

5.6. Channel Measurement on Off-the-Shelf Devices

Based on the system model introduced in Section 5.5, in the following, we detail the design of the core component of ACO, that is, retrieving complex gains from simple signal strength readings. This is particularly challenging because commodity COTS devices use cost-efficient components that do not allow for exact amplitude and phase measurements. To address this limitation, we transmit multiple probes with different phase shifts and use this information to reconstruct the full CSI, as sketched in Section 7.1. Figure 5.1 summarizes this procedure.

To enable the above measurement, ACO needs to retrieve the complex gain of each antenna $\mathbf{h}_{\mathrm{RX}}$. However, obtaining the actual value of $\mathbf{h}_{\mathrm{RX}}$ is impossible because any set of measurements explained by $\mathbf{h}_{\mathrm{RX}}$ can also be explained by $\alpha\mathbf{h}_{\mathrm{RX}}$ where α is any complex number with unitary modulus. Thus, ACO computes $\mathbf{h}_{\mathrm{RX}}$ relative to a given $\mathbf{p}$ in terms of phase shift. That is, we assume that $\arg(\mathbf{h}_{\mathrm{RX}}\mathbf{p}) = 0$. This is not a problem because we only need the relative phases among antenna elements to measure path interactions and compute beamforming weights that maximize the SNR. Further, determining the actual value of $\mathbf{h}_{\mathrm{RX}}$ is not useful in practice because the phase of $\mathbf{h}_{\mathrm{RX}}$ can change by a full π rad if the length of the link changes by half a wavelength, which for 60 GHz is less than two millimeters. That is, any slight movement would invalidate the result. Fortunately, ACO only requires *relative* phase shifts.

5.6.1. Complex Gain Retrieval

The intuitive explanation in Figure 5.1 depicts how ACO measures relative phase shifts among individual antenna elements. In Section 5.6.1.1 we formalize this procedure. Using such individual antenna elements translates into estimating the channel with quasi-omnidirectional beam patterns. This approach is highly effective on high SNR links and is immune to sudden changes in the environment. For instance, if the alignment among devices changes abruptly, using quasi-omnidirectional beam patterns allows ACO to react instantly to the change. Since millimeter-wave communication is typically restricted in range (e.g., within a room), the high SNR case is common.

Nevertheless, we extend ACO in Section 4.1.2 such that it can also operate in low SNR cases. To this end, we generalize the approach in Figure 5.1 such that the reference signal

is generated not only by a single antenna element but via a directional beam pattern. This beam pattern provides a higher SNR than the omnidirectional reference, but only allows to retrieve the complex gain in specific directions. The underlying concept is identical but the analytical definition changes. We formalize this approach in Section 5.6.1.2. Further details on the mathematical derivation of both approaches can be found in the accompanying technical report [101].

5.6.1.1. Quasi-Omnidirectional Estimation

We identify each of the elements on the antenna array using index k. Further, we define the index of the reference antenna R (see Section 7.1) as $\bar{k}$. We divide the procedure to retrieve the complex gain into the measurement of (a) the amplitude and (b) the phase.

5.6.1.1.1. (a) Amplitude Measurement Measuring $|[\mathbf{h}_{\mathrm{RX}}]_k|\ \forall k$ is straightforward since ACO only needs to transmit probes switching on one element at a time, as defined below.

$$[\mathbf{p}']_{k'} = \begin{cases} 0 & k' \neq k \\ 1 & k' = k \end{cases}, \qquad \mathbf{p} = \frac{\mathbf{p}'}{\|\mathbf{p}'\|}. \tag{5.4}$$

In the remainder of this subsection, we use the prime symbol to refer to the non-normalized version of a beam pattern. As per Section 5.5.2, ACO obtains the amplitude as $|[\mathbf{h}_{\mathrm{RX}}]_k|^2 = |\mathbf{h}_{\mathrm{RX}}\mathbf{p}|^2$. Based on the result, ACO defines as reference element $\bar{k}$ the element with the largest amplitude.

5.6.1.1.2. (b) Phase Measurement Next, we need to obtain the phase shift of each element, except for the reference $\bar{k}$, which by definition has zero phase. That is, we obtain $\arg[\mathbf{h}_{\mathrm{RX}}]_k,\ \forall k \neq \bar{k}$. Since we measure the relative phase of element $k \neq \bar{k}$ with respect to $\bar{k}$, we need to enable both antennas k and $\bar{k}'$. We then probe the phase values $[\mathbf{p}']_k \in \{1, i, -1, -i\}$ while $[\mathbf{p}']_{\bar{k}} = 1$ and all other components are 0. We then group the measurements $|\mathbf{h}_{\mathrm{RX}}\mathbf{p}_m|^2$ in a vector of four components $\mathbf{a} \in \mathbb{R}^{4\times 1}$ where $[\mathbf{p}'_m]_k$ has the values $1, i, -1, -i$ for $[\mathbf{a}]_1, [\mathbf{a}]_2, [\mathbf{a}]_3, [\mathbf{a}]_4$, respectively. That is, each element in $\mathbf{a}$ contains the sum of two signals—the signal transmitted by the reference antenna element $\bar{k}$ and the signal transmitted by measured antenna element k for one out of four possible phase shifts m. We can reformulate this as in Equation 5.5.

$$\begin{aligned} [\mathbf{a}]_m &= |\mathbf{h}_{\mathrm{RX}}\mathbf{p}_m|^2 = |\tfrac{1}{\sqrt{2}}([\mathbf{h}_{\mathrm{RX}}]_{\bar{k}} + [\mathbf{h}_{\mathrm{RX}}]_k e^{(m-1)\frac{\pi}{2}i})|^2 \\ &= \tfrac{1}{2}|[\mathbf{h}_{\mathrm{RX}}]_{\bar{k}} + |[\mathbf{h}_{\mathrm{RX}}]_k| e^{(\arg[\mathbf{h}_{\mathrm{RX}}]_k + (m-1)\frac{\pi}{2})i}|^2 \\ &= \Gamma + 2\Delta\cos(\arg[\mathbf{h}_{\mathrm{RX}}]_k + (m-1)\tfrac{\pi}{2}) \\ &= \Gamma + \Delta e^{(\arg[\mathbf{h}_{\mathrm{RX}}]_k + (m-1)\frac{\pi}{2})i} + \Delta e^{-(\arg[\mathbf{h}_{\mathrm{RX}}]_k + (m-1)\frac{\pi}{2})i} \\ &= \Gamma + \Delta e^{\arg[\mathbf{h}_{\mathrm{RX}}]_k i} e^{(m-1)\frac{\pi}{2}i} + \Delta e^{-\arg[\mathbf{h}_{\mathrm{RX}}]_k i} e^{-(m-1)\frac{\pi}{2}i} \end{aligned}, \tag{5.5}$$

for $\Gamma = \frac{[\mathbf{h}_{\mathrm{RX}}]_{\bar{k}}^2 + |[\mathbf{h}_{\mathrm{RX}}]_k|^2}{2}$ and $\Delta = \frac{[\mathbf{h}_{\mathrm{RX}}]_{\bar{k}} |[\mathbf{h}_{\mathrm{RX}}]_k|}{2}$. This is a wave decomposition expression. Hence, the fast Fourier transform of $\mathbf{a}$ defined as $\mathbf{f} = \mathrm{FFT}(\mathbf{a})$ has the expression

$$\mathbf{f} = [\Gamma, \Delta e^{\arg[\mathbf{h}_{\mathrm{RX}}]_k i}, 0, \Delta e^{-\arg[\mathbf{h}_{\mathrm{RX}}]_k i}]^T. \tag{5.6}$$

This means that the phase shift of antenna k with respect to $\bar{k}$ is $\arg[\mathbf{f}]_2$, i.e., the phase of the second element of the fast Fourier transform of the received signal strength vector $\mathbf{a}$.

The above approach is feasible whenever the transmitter can reach the receiver using the quasi-omnidirectional beam pattern given by switching on only the reference antenna element $\bar{k}$. The stronger $\bar{k}$, the more accurate is the estimation of the phase shift. As described in Section 5.4, we define a measurement codebook that contains the required phase shifts to compute the phase for all of the elements of the antenna. The number of probes in the ACO sweep (see Figure 5.3) is directly related to N, which is the overall number of antenna elements probed in the measurement and the exploration parts of the sweep.

- For antenna element $\bar{k}$, we only need one probe because we define it as a phase reference and thus we only measure its amplitude. Its relative phase is zero.

- For all other antenna elements, we need one probe to measure its amplitude and four additional probes to obtain its relative phase with respect to $\bar{k}$.

Thus, the ACO sweep contains a total of $N + 4(N-1)$ probes to measure the complex gain of N antenna elements.

5.6.1.2. Directional Estimation

In the following, we discuss an extension of our approach in Section 5.6.1.1 for the low SNR case. Intuitively, instead of using a single antenna element as a reference, we resort to a beam pattern $\mathbf{p}$. Due to the higher directionality compared to the quasi-omnidirectional case, this results in much higher received signal strength during the channel estimation. Beam pattern $\mathbf{p}$ is typically one that the transmitter used earlier for communication. For instance, ACO could use one of the generic beam patterns of the device. Based on this, $\mathbf{p}$ is suitable as a reference if $\mathbf{p}$ has higher directionality than any single antenna beam pattern even when switching off that antenna. That is, if we set any k and define $\bar{\mathbf{p}}$ as $[\bar{\mathbf{p}}']_{k'} = [\mathbf{p}']_{k'}$ for any $k' \neq k$ and $[\bar{\mathbf{p}}']_k = 0$, then $|\mathbf{h}_{\mathrm{RX}}\bar{\mathbf{p}}| > |[\mathbf{h}_{\mathrm{RX}}]_k|$. This condition is likely to hold for a generic beam pattern selected for communication since it is very probable that such a beam pattern has a higher gain than a quasi-omnidirectional one. Since ACO uses $\mathbf{h}_{\mathrm{RX}}\mathbf{p}$ as the phase reference, we define its phase to be zero, that is, $\arg(\mathbf{h}_{\mathrm{RX}}\mathbf{p}) = 0$. Based on the above definitions, we compute the amplitude and the relative phase for

antenna element k. We distinguish two cases: (a) a simple case when antenna element k is switched off in the reference $\mathbf{p}$, and (b) an analytically harder case when k is on in $\mathbf{p}$.

5.6.1.2.1. (a) Antenna element k is off in $\mathbf{p}$ Since element k is off, $[\mathbf{p}]_k = 0$. We define $\mathbf{p}_m$ such that $[\mathbf{p}]_{k'} = [\mathbf{p}_m]_{k'}\ \forall k' \neq k$ with $[\mathbf{p}'_m]_k = 1, i, -1, -i$ for $m = 1, 2, 3, 4$, respectively. That is, we probe all possible phase shifts for antenna k. Next, we follow the same approach as in Equation 5.5 but substitute $[\mathbf{h}_{\mathrm{RX}}]_{\bar{k}}$ with $\mathbf{h}_{RX}\mathbf{p}$. The result for the phase is equivalent, that is, the relative phase shift of antenna element k is again $\arg[\mathbf{f}]_2$. However, the amplitude is harder to measure since we cannot measure each antenna element individually. To compute it, we resort to Γ and Δ, which for the directional case are as defined in Equation 5.7.

$$\Gamma = \frac{(\mathbf{h}_{RX}\mathbf{p})^2 + |[\mathbf{h}_{\mathrm{RX}}]_k|^2}{(\|\mathbf{p}'\|^2 + 1)} \qquad \Delta = \frac{(\mathbf{h}_{RX}\mathbf{p})|[\mathbf{h}_{\mathrm{RX}}]_k|}{(\|\mathbf{p}'\|^2 + 1)} \tag{5.7}$$

Per Equation 5.6, both Γ and Δ are known since we can obtain them directly from the fast Fourier transform of the amplitude measurements $\mathbf{a}$. Thus, Equation 5.7 is a quadratic system of equations from which we can compute $|[\mathbf{h}_{\mathrm{RX}}]_k|$. Further, since $\mathbf{h}_{RX}\mathbf{p} \in \mathbf{R} > |[\mathbf{h}_{RX}]_k|$, the system has two real-valued solutions as shown in Equation 5.8.

$$\begin{aligned} \mathbf{h}_{RX}\mathbf{p} + |[\mathbf{h}_{\mathrm{RX}}]_k| &= \sqrt{(\Gamma + 2\Delta)(\|\mathbf{p}'\|^2 + 1)} \\ \mathbf{h}_{RX}\mathbf{p} - |[\mathbf{h}_{\mathrm{RX}}]_k| &= \sqrt{(\Gamma - 2\Delta)(\|\mathbf{p}'\|^2 + 1)} \end{aligned}. \tag{5.8}$$

Rearranging Equation 5.8 and substituting Γ and Δ with the coefficients of the fast Fourier transform of $\mathbf{a}$, we obtain the amplitude of the complex gain as in Equation 5.9.

$$|[\mathbf{h}_{\mathrm{RX}}]_k| = \frac{\sqrt{\|\mathbf{p}'\|^2 + 1}}{2}(\sqrt{[\mathbf{f}]_1 + 2|[\mathbf{f}]_2|} - \sqrt{[\mathbf{f}]_1 - 2|[\mathbf{f}]_2|}). \tag{5.9}$$

5.6.1.2.2. (b) Antenna element k is on in $\mathbf{p}$ Since element k is on, $[\mathbf{p}]_k \neq 0$. We again probe all possible phase shifts for antenna k, that is, we set $[\mathbf{p}'_m]_k = 1, i, -1, -i$ for $m = 1, 2, 3, 4$, group measurements into a variable $\mathbf{a}$, and compute $\mathbf{f} = \mathbf{FFT}(\mathbf{a})$. We obtain the amplitude and phase of antenna element k following an approach along the lines of (a). Since the detailed analysis is similar but more protracted than in (a), we do not discuss the details but directly give the analytical expression for the amplitude in Equation 5.10 and the one for the relative phase in Equation 5.11.

$$|[\mathbf{h}_{\mathrm{RX}}]_k| = \frac{\|\mathbf{p}'\|}{2}(\sqrt{[\mathbf{f}]_1 + 2|[\mathbf{f}]_2|} - \sqrt{[\mathbf{f}]_1 - 2|[\mathbf{f}]_2|}) \tag{5.10}$$

$$\arg[\mathbf{h}_{\mathrm{RX}}]_k = \arg[\mathbf{f}]_2 - \arg\left(\gamma + |[\mathbf{h}_{\mathrm{RX}}]_k|[\mathbf{p}']_k e^{\arg[\mathbf{f}]_2}\right). \tag{5.11}$$

In Equation 5.11, parameter γ is defined as in Equation 5.12 and is used to substract the phase contribution to the beam pattern $\mathbf{p}$ of the antenna that we are measuring.

$$\gamma = \frac{\|\mathbf{p}'\|}{2}(\sqrt{[\mathbf{f}]_1 + 2|[\mathbf{f}]_2|} + \sqrt{[\mathbf{f}]_1 - 2|[\mathbf{f}]_2|}). \tag{5.12}$$

Similarly to the quasi-omnidirectional case in Section 5.6.1.1, we define a measurement codebook that contains the required phase shifts to compute the complex gains of the relevant antenna elements. However, in this case, the number of required probes depends on whether each antenna element is on or off in the reference beam pattern $\mathbf{p}$.

- If the element is on in $\mathbf{p}$, we need to measure $\mathbf{p}$ and three additional probes with changing phase shifts.
- If the element is off in $\mathbf{p}$, we need to transmit four probes to compute its relative phase and amplitude.

Thus, the ACO sweep contains a total of $1 + 3N_{on} + 4N_{\text{off}}$ probes, where $N_{\text{on}}/N_{\text{off}}$ is the number of antenna elements that are on/off in reference $\mathbf{p}$, respectively.

5.6.2. Channel Matrix Computation

Using the methods introduced in Section 5.6.1, ACO can compute $\mathbf{h}_{\text{RX}}$ and analogously $\mathbf{h}_{\text{TX}}$. That is, it can obtain the receiving and the transmitting channel. In the following, we formulate how both can be combined to obtain the full CSI $\mathbf{H}$ of the measured subspace. Since both $\mathbf{h}_{\text{RX}}$ and $\mathbf{h}_{\text{TX}}$ are relative values, we multiply them with unitary complex values δ and λ to model the unknown relative phase shift, obtaining $\mathbf{h}_{\text{RX}} = \delta\mathbf{c}^H\mathbf{H}$ and $\mathbf{h}_{\text{TX}} = \lambda\mathbf{H}\mathbf{p}$. For each antenna element $k \in 1, 2, \ldots, N$ in the subspace, we rearrange the measured complex gain of the transmit and the receive channel into the matrices $\mathbf{G}_{\text{TX}}$ and $\mathbf{G}_{\text{RX}}$, respectively. Specifically, we stack them as $[\mathbf{G}_{\text{RX}}]_{k,:} = \delta_k[\mathbf{W}]^H_{:,k}\mathbf{H}$ and $[\mathbf{G}_{\text{TX}}]_{:,k} = \lambda_k\mathbf{H}[\mathbf{W}]_{:,k}$, where $\mathbf{W}$ is a matrix that we define as $\mathbf{W} \in \mathbb{C}^{N\times N}$. Based on this definition, we can formulate $\mathbf{G}_{\text{RX}}$ and $\mathbf{G}_{\text{TX}}$ as follows:

$$\mathbf{G}_{\text{RX}} = \mathbf{D}_\delta\mathbf{W}^H\mathbf{H} \tag{5.13}$$

$$\mathbf{G}_{\text{TX}} = \mathbf{H}\mathbf{W}\mathbf{D}_\lambda, \tag{5.14}$$

where $[\mathbf{D}_\delta]_{k,k} = \delta_k$ and $[\mathbf{D}_\lambda]_{k,k} = \lambda_k$ are diagonal matrices. Next, we relate the above formulation to the singular-value decomposition (SVD) of the channel. To this end, we define $\mathbf{U}\mathbf{S}\mathbf{V}^H = \mathbf{H}$, $\mathbf{U}_{\text{RX}}\mathbf{S}_{\text{RX}}\mathbf{V}^H_{\text{RX}} = \mathbf{H}$, and $\mathbf{U}_{\text{TX}}\mathbf{S}_{\text{TX}}\mathbf{V}^H_{\text{TX}} = \mathbf{H}$ to be the SVD of $\mathbf{H}, \mathbf{G}_{\text{RX}}$ and $\mathbf{G}_{\text{TX}}$ respectively. Since $\mathbf{D}_\delta, \mathbf{D}_\lambda$ are unitary diagonal matrices and thus orthogonal, this means that $\mathbf{D}_\delta\mathbf{W}^H\mathbf{U}$ and $\mathbf{W}^H\mathbf{D}^H_\lambda\mathbf{V}$ are orthogonal if and only if $\mathbf{W}$ is orthogonal. Unless $\mathbf{H}$ contains two equal non-zero singular values, which is a zero-probability event, we have that its economic SVD decomposition is unique and translates into the following expressions:

$$\mathbf{S} = \mathbf{S}_{\text{RX}} = \mathbf{S}_{\text{TX}} \tag{5.15}$$

$$\mathbf{V}_{\mathrm{RX}} = \mathbf{V} \qquad \mathbf{U}_{\mathrm{TX}} = \mathbf{U} \tag{5.16}$$

$$\mathbf{U}_{\mathrm{RX}} = \mathbf{D}_{\delta}\mathbf{W}^{H}\mathbf{U} \qquad \mathbf{V}_{\mathrm{TX}} = \mathbf{W}^{H}\mathbf{D}_{\lambda}^{H}\mathbf{V} \ . \tag{5.17}$$

Thus, we can compute the CSI as $\mathbf{H} = \mathbf{U}_{\mathrm{TX}}\sqrt{\mathbf{S}_{\mathrm{RX}}\mathbf{S}_{\mathrm{TX}}}\mathbf{V}_{\mathrm{RX}}$, where the square root is computed element-wise. For the above to be valid, $\mathbf{W}$ must be an orthogonal matrix such that each column $[\mathbf{W}]_{:,k}$ is a feasible analog configuration. To this end, we can use a Hadamard matrix with entries $\{-1, 1\}$ that has been normalized to satisfy power constraints.

5.6.3. SNR Maximizing Beam Patterns

Based on the measured CSI, ACO computes the beam pattern $\mathbf{p}$ that maximizes the SNR. As discussed in Section 5.5.2, the received signal strength is $|\mathbf{h}_{\mathrm{RX}}\mathbf{p}|^2$. Thus, the beam pattern $\mathbf{p}$ that maximizes the expression is derived from the equality condition of the Schwartz inequality as $\mathbf{p} = \mathbf{h}_{\mathrm{RX}}^{\mathrm{H}}/\|\mathbf{h}_{\mathrm{RX}}\|$. Due to hardware constraints, we round $\mathbf{p}$ to match the device resolution. Since $\mathbf{h}_{\mathrm{RX}}$ is known, ACO can compute the optimal $\mathbf{p}$ as explained before. Since ACO only requires an FFT and simple element-wise matrix operations, its additional computational complexity is minimal. Thus, ACO computes the optimal beam pattern in a highly efficient manner and can be executed even on highly constrained hardware components.

5.7. Experiment Platform

We implement and evaluate our scheme on off-the-shelf IEEE 802.11ad devices that incorporate phased antenna arrays. We obtain control over the antenna array to adjust steering parameters, without requiring any hardware modifications. In the following, we describe the system architecture of our platform, the utilized antenna module, and how to control the antenna beam steering.

5.7.1. System Architecture

Only a few consumer devices can communicate in the 60 GHz band and support IEEE 802.11ad. One of them is the TP-Link Talon AD7200 tri-band router that uses a Qualcomm QCA9500 IEEE 802.11ad Wi-Fi chip. It features an antenna array with 32 elements that are individually controllable in phase and amplitude. Control over the antenna is encapsulated in the firmware of the Wi-Fi chip, and only limited access is exposed to the host operating system and device driver. To obtain access to the antenna steering properties, as well as the IEEE 802.11ad beam training operation, we utilize the framework for practical IEEE 802.11ad research proposed in [70, 102]. This framework comes with a customized LEDE/OpenWrt image to provide open access to

the operating system, network interfaces, and device drivers. Using the binary firmware patching framework (nexmon-arc) [102], additional features can be directly implemented in the firmware of the Wi-Fi chip.

5.7.2. Sector Level Sweep

The **sls!** is the standard beam training mechanism in IEEE 802.11ad and implemented in the firmware of the QCA9500 chip. It uses a set of predefined antenna configurations, the so-called sectors. These sectors are designed such that their beam patterns cover the entire azimuth to allow for an effective scanning of devices in proximity. We briefly discussed the operation of the **sls!** in Section 7.1. Through firmware patching, we can extract the SNR and RSSI of received sector sweep frames and report back custom feedback in the sector sweep feedback frames.

5.7.3. Antenna Module

To select antenna steering parameters, we need to obtain a thorough understanding of the antenna and its steering capabilities. Qualcomm's QCA9500 IEEE 802.11ad Wi-Fi chip consists of two modules, 1) a baseband IC that takes care of the signal and frame processing, and 2) an antenna module with an additional **rf!**-IC that drives the antenna elements and controls the radiation characteristics. This modular design allows to flexibly place the antenna module at proper locations inside a device chassis to minimize radiation impairments. Both modules, the antenna and baseband IC, are connected with a coaxial cable for bi-directional transfer of modulated data, control, and clock signals, as well as power supply for the antenna module. In transmit mode, the antenna chip mixes up the modulated data signal from **if!** (**if!**) to the desired **rf!** channel in the 60 GHz band. In receive mode, the **rf!** signal is mixed down to **if!** again. All antenna elements in the array are driven by an antenna weighting network, which is adjusted by the external control signal. Currently, only a single **rf!**-chain is available. The antenna weighting network consists of an antenna switch, eight distribution amplifiers, 32 edge amplifiers and 32 phase shifters. All of these are controllable from within the firmware running on the baseband chip. To adjust the antenna steering, the following six 32-bit parameters are available:

- psh_hi: phase shift values for ant. chains [15-0]
- psh_lo: phase shift values for ant. chains [31-16]
- etype0: edge amp. bit 0 for all ant. chains [31-0]
- etype1: edge amp. bit 1 for all ant. chains [31-0]
- etype2: edge amp. bit 2 for all ant. chains [31-0]

- dtype_swch_off: dist. amp. values (3 bits each) + X16 switch bits

A discrete configuration of the antenna with these parameters refers to a so-called sector. The current firmware in version 4.1.0.55 supports 64 different transmit sectors out of which 35 are defined and used in beam training, as well as a single receive sector. The definitions for these sectors are stored in an antenna steering codebook.

The antenna steering codebook and individual sector configurations can be either changed directly in the memory of the firmware image (hardcoded) or dynamically adjusted during runtime. For the latter, the driver exposes specific netlink vendor commands. By changing the antenna parameters in the sector configurations, we can change the radiation patterns and gains—a crucial prerequisite to implement adaptive codebook optimization on off-the-shelf devices.

5.7.4. Reconstructing the Antenna Layout

Unfortunately, no public documentation for the antenna module or the baseband chip was available at the time of writing. Hence, we first had to analyze the internal structure of the antenna elements experimentally to create specific beam patterns. Each phase shifter in the weighting network is driven by two consecutive bits from either psh_hi or psh_lo. The edge amplifiers use a single bit from each of etype0, etype1, and etype2, while the distribution amplifiers consume three consecutive bits from dtype_swch_off. The most significant bits in dtype_swch_off are used to drive the antenna switch. We developed a parser to extract the parameters for each element which allows us to easily obtain and adjust the gain and phase of each antenna element.

By modifying individual parameters, we reconstructed the structure of the antenna weighting network experimentally. For example, by iterating over all pairwise combinations of distribution and edge amplifiers and setting all others to zero, we found that the first distribution amplifier drives the first four edge amplifiers, the second distribution amplifier drives the next four edge amplifiers, and so on. An antenna is only active if both the corresponding distribution and edge amplifier are set to non-zero. Similarly, we verified which phase shifter bits belong to which antenna chain by changing single values and monitoring the received signal strength at an unmodified device. The resulting antenna weighting structure as revealed in our experiments is shown in Figure 5.4.

By disassembling the antenna from the chip and shielding all except one antenna element, we reconstructed the 3-dimensional layout of the antenna array as well as the physical element positions as shown in Figure 5.5. The antenna array consists of 12 patch antennas that are located on the front surface of the module in a 2×6 matrix shape. On the back side, there are only six patch antennas since the IC blocks the rest of the surface. The remaining 14 antennas are dipole antennas and directed towards the sides of the Printed Circuit Board (PCB). As a result, the three-dimensional layout, as well as

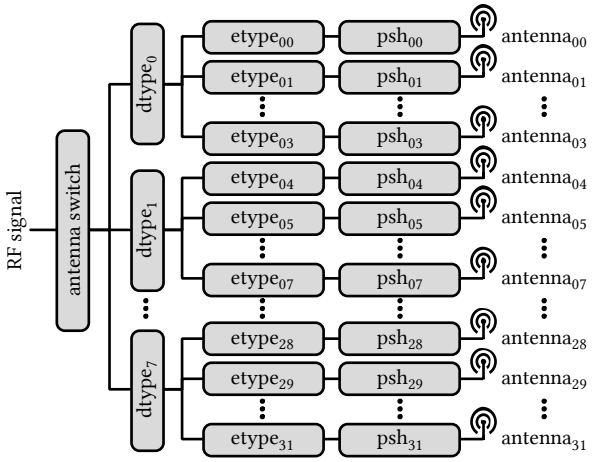


Figure 5.4: Experimentally reconstructed antenna weighting network for beam steering inside the QCA6310 antenna IC, which consists of an antenna switch, 8 distribution amplifiers (dtype), 32 edge amplifiers (etype) and 32 phase shifters (psh) to drive 32 antenna elements individually.

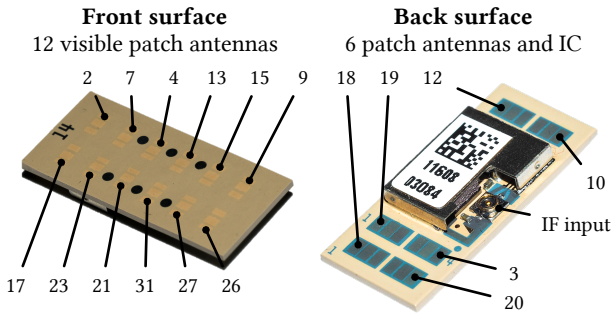


Figure 5.5: Phased array antenna chip disassembled from a Talon AD7200. Annotations show the antenna elements ID of the patch antennas. 14 additional dipole antennas are invisible from the surface.

its asymmetric assembly, lead to irregular beam patterns as already discussed in [70, 93].

This know-how of the antenna layout and the capability to control the weighting network is crucial to understand how to optimize the beam patterns in the codebook. It is also a valuable resource for other researchers using this platform.

5.8. Practical Implementation

We implement ACO on the TP-Link Talon AD7200 router introduced in the previous section.[1] We first discuss in Section 5.8.1 the ACO implementation itself and then provide an overview on our experiment setup in Section 5.8.2.

[1]The ACO implementation is available at: `http://wireless.networks.imdea.org/downloads-software-and-traces`

5.8.1. Prototype Operation

Our prototype consists of an IEEE 802.11ad AP that serves one or multiple STAs. We use Talon devices to realize both AP and STAs, since the prototyping framework in [70] allows us to set them to either of those modes. For ease of experimentation, we control the prototype from a laptop which is external to the 60 GHz network. All of the devices that are part of the prototype are connected to a common 2.4 GHz Wi-Fi network that we use to control the experiments. Using this control network, the laptop instructs the individual STAs to connect to the AP, generates traffic in the IEEE 802.11ad network, and collects statistics. For simplicity, we also run certain components of ACO on the laptop, such as the reconfiguration of the codebooks via the interfaces described in Section 5.7.3. However, these portions of the code could as well run directly on the Talon devices since ACO has very low computational requirements.

The operation of our prototype is as follows. First, the control laptop loads the default codebooks of the Talon devices on both the AP and the STAs. Then, it triggers the STAs to establish a connection to the AP using the regular IEEE 802.11ad procedure. After that, ACO loads an initial measurement codebook on the devices and collects the resulting SNR readings to compute the transmit beam pattern that maximizes the SNR for each link. We carry out the process for AP and STAs, since we analyze both up-link and down-link traffic in Section 5.8.2 (and to ensure that acknowledgments are also sent with the optimized beam patterns). To specifically optimize the beam patterns under low SNR conditions, we employ the directional estimation scheme described in Section 4.1.2. Finally, the control laptop re-configures all devices with the computed communication beam patterns and generates traffic in the network. In particular, we use `iperf3` to generate ten seconds of down-link traffic and ten seconds of up-link traffic on each link. For each experiment, we repeat this measurement twenty times while we collect statistics such as TCP throughput, physical layer rate, and signal quality.

5.8.2. Experiment Setup

We evaluate ACO in an open-plan office environment (see Figure 5.6). The area has a size of 18×7.5 meters and is surrounded by twelve individual offices. We deploy one AP below the ceiling next to the desks in the open-plan area. Further, we place STAs at 24 different positions within the room. Out of the 24 locations, 12 correspond to the desks. For these cases, we orient the STAs such that they mimic the usual placement of a user's laptop. Next, we select six further positions in an area with no desks but still in line of sight (LOS) of the AP. This enables us to study whether the desks and the separators among them play a significant role. Finally, we select six further STA locations behind an isolated wall that separates the entry area from the open-plan office space. Those locations are non line of sight (NLOS), which means that STAs must use reflections to

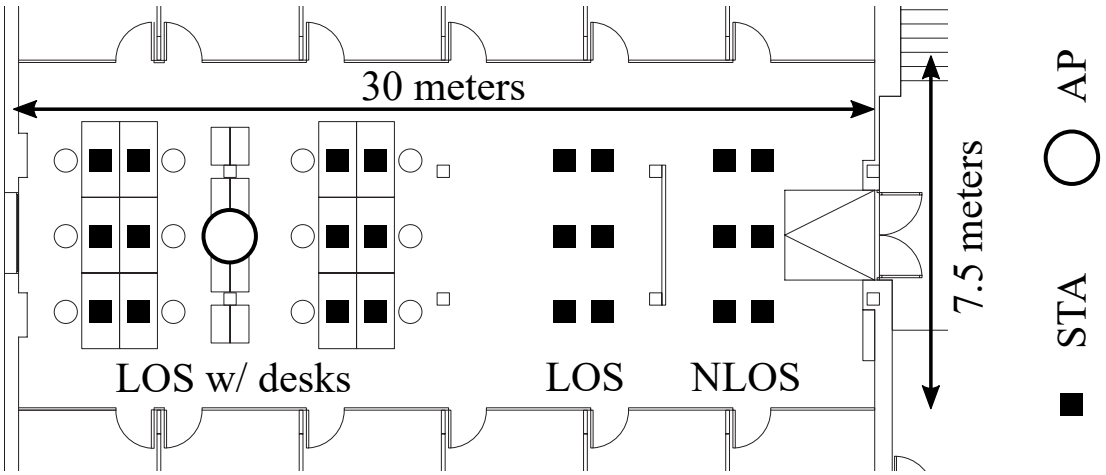


Figure 5.6: Experiment setup in indoor office scenario.

reach the AP.

5.9. Evaluation

We evaluate the performance of ACO in the real-world office scenario described in Section 5.8.2. Since we focus on the performance of individual links, we only activate one STA at a time. We obtain the achieved SNR, the TCP throughput, IQ constellations, and the expected bitrate.

5.9.1. SNR Maximization

As discussed in Section 5.6.3, we design ACO to maximize the SNR. In the first part of our experiments, we evaluate the SNR gains that ACO achieves in comparison to IEEE 802.11ad operation with the default antenna patterns. For both our optimized beams and the default ones, we measure the SNR at all locations of our testbed for the up- and down-link. For ease of comparison, we report the SNR on a linear scale throughout this section. As shown in Figure 5.7, we achieve a SNR of 8.16 for the up-link and 7.53 for the down-link when using the generic beam patterns. With ACO, we increase the average linear SNR to 22.06 for the up-link and 19.78 for the down-link, respectively. The cumulative distribution function (CDF) of the default beam pattern reveals that 90% of the measurements have an SNR that is lower than 18. With ACOs optimized beams, we achieve an SNR higher than this in about 45% of the measurements. The floor-maps in Figure 5.8 illustrate the SNR gains that we achieve at the individual measurement points in our environment. We find that the distance between AP and STA does not influence ACOs performance. However, the variance among different locations is high. In some spots, ACO achieves gains of only 50% while in others the gains are over 500%. Some

locations can be easily reached via reflectors while others cannot. Overall, we achieve an average gain of 168.61% on the up-link and 159.34% on the down-link.

To validate the high SNR gains, we additionally analyze the signals at the physical layer. Unfortunately, evaluating the accuracy of the CSI and how it degrades with SNR is impossible on the COTS devices themselves. This leaves us to compare the achievable signal constellations with those of the default codebook. To this end, we place a Sivers IMA FC2221V/01 V-band down-converter with a horn antenna at the location of the receiver and capture the raw samples using a Keysight DSOS254A oscilloscope. The bandwidth of the oscilloscope is sufficient to capture the full IEEE 802.11ad signal, which we decode using the Keysight Wideband Waveform Center. This enables us to obtain the IQ constellations of data frames transmitted both with the generic beam patterns and with ACO beam patterns. We show an example in Figure 5.9. Both frames in that example were transmitted with MCS8, which uses a $\pi/2$-QPSK constellation. The ACO beam patterns show a visible accuracy improvement and lead to constellations that are significantly less noisy than those of the default beam pattern. This results in fewer symbol errors, which in turn improves the stability of the link.

5.9.2. Throughput Improvement

The large SNR improvement in Section 5.9.1 enables our Talon devices to switch to a higher MCS. As a result, we obtain significant TCP throughput gains when generating `iperf3` traffic in the network. Figure 5.10 depicts the CDF of the achieved TCP throughput at a LOS location in our testbed. We achieve gains for both up-link and down-link. For the former, the average gain of ACO is 58% while for the latter it is 102%. That is, we *double* the throughput in the down-link. In Figure 5.10, we also show the distribution of the MCS for that location. As expected, the Talon devices are able to choose high modulations much more often when using our ACO beam patterns compared to the generic ones. For instance, ACO enables MCS8 on the down-link more than 60% of the time whereas generic beam patterns only reach such a high MCS 18% of the time.

In Figure 5.11 we perform a similar analysis for one of the NLOS locations behind the isolated wall in our testbed (see Section 5.6). Due to the wall, the AP and the STA communicate via a reflection. Our results show that the generic beamforming mechanism in IEEE 802.11ad performs poorly in such a scenario, achieving on average only 42.74 Mbps in the up-link and 314.54 Mbps in the down-link. In contrast, ACO achieves on average about 855 Mbps on the same link in both directions. The MCS distribution in Figure 5.11 reflects this behavior, showing that IEEE 802.11ad operates mostly in the range of MCS1 to MCS3, whereas ACO enables rates beyond MCS4 in most cases. We conclude that ACO is highly beneficial for IEEE 802.11ad wireless networks and can achieve order-of-magnitude throughput gains.

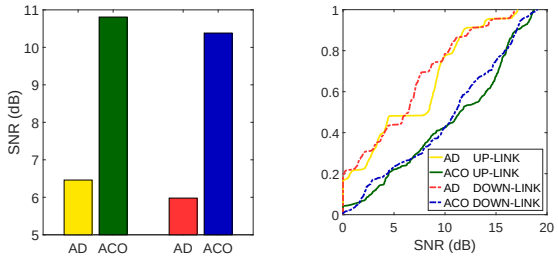


Figure 5.7: Average SNR and CDF of our ACO mechanism and the default IEEE 802.11ad operation for uplink and down-link measurements.

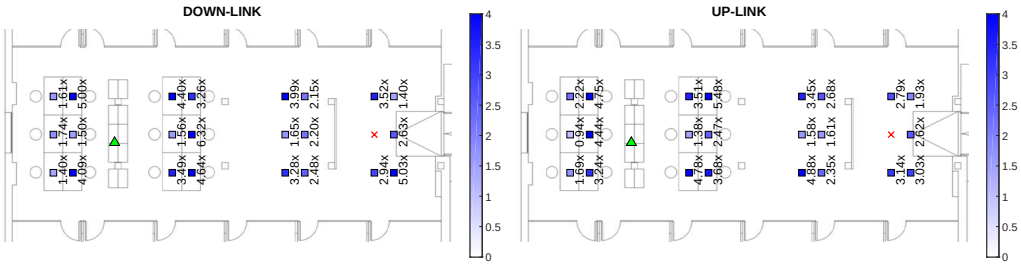


Figure 5.8: Map of the average SNR gains of optimized beams. Each square represents a STA location, crosses indicate blind spots, while the AP is indicated by the green triangle.

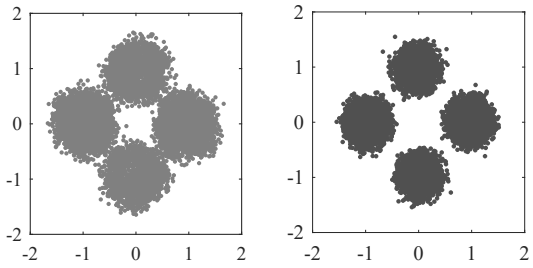


Figure 5.9: Constellation diagrams of MCS8 encoded frames transmitted with generic (left) and optimized beam patterns (right).

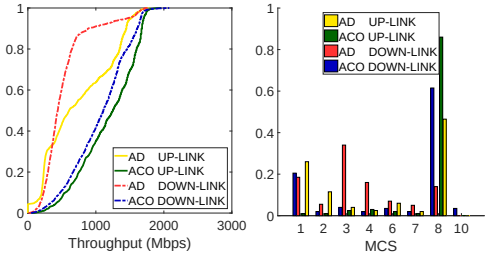


Figure 5.10: LOS throughput CDF and transmission MCS histogram.

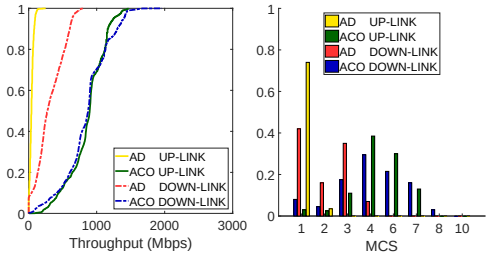


Figure 5.11: NLOS throughput CDF and transmission MCS histogram.

5.9.3. Expected Bitrate

Although we consistently achieve very high SNR gains at all of the locations in our testbed (see Section 5.9.1), we only observe equivalent throughput gains at a subset of those locations. To understand this effect, we delved into the operation of the Talon AD7200 router. We found that in many cases the router does not switch to a higher modulation even though the SNR improves by a factor of three or more. In particular, we observed that the rate adaptation algorithm very often does not exceed MCS8 even though the router implements up to MCS12. As a result, TCP throughput is limited to about 1.6 Gbps. Figure 5.12 shows that the throughput saturates at this value for all of the locations in our testbed, irrespective of the length of the link and other factors that could influence signal quality. In exhaustive measurements, we found that the rate adaptation mechanism of the router is configured such that MCS configurations beyond MCS8 are disfavored. This is a coherent design decision for the Talon hardware that comes with certain limitations due to its inexpensive production. For example, the Gigabit Ethernet interface limits traffic to one Gigabit per second. The 60 GHz interface is not the bottleneck and already saturates the rest of the system even with an MCS below the maximum. For this reason, it seems that the manufacturer selected a conservative rate adaptation scheme that does not switch to high MCS values even when the SNR of the link is very high. We expect next-generation IEEE 802.11ad/ay devices to provide better support for such multi-Gigabit per second modulations in terms of both handling high volumes of traffic and generating cleaner radio-frequency signals that enable higher modulations such as 16-QAM. Hence, we conclude that the throughput saturation that we observe in Figure 5.12 is due to limitations of the hardware that prevent stable operation of the router at such high MCSs.

As a result of the above hardware limitations, the benefits of ACO only become visible in our testbed when a link using the generic beam patterns of IEEE 802.11ad operates below MCS8. Otherwise, the SNR improvement of ACO does not translate into an effective throughput gain. To illustrate the benefits that ACO would achieve on hardware that can process higher data throughput, in the following we compute the bitrate according to the expected MCS for a given SNR. To this end, we build on the signal quality

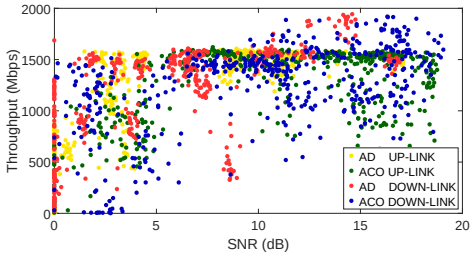


Figure 5.12: Measured throughput as a function of the measured SNR.

Table 5.1: Recommended MCS selection, required receive sensitivity, and bitrate in IEEE 802.11ad [1].

MCS	Sensitivity	Bitrate
0	−78 dBm	27.5 Mbps
1	−68 dBm	385.0 Mbps
2	−66 dBm	770.0 Mbps
3	−65 dBm	962.5 Mbps
4	−64 dBm	1155.0 Mbps
5	−62 dBm	1251.3 Mbps
6	−63 dBm	1540.0 Mbps
7	−62 dBm	1925.0 Mbps
8	−61 dBm	2310.0 Mbps
9	−59 dBm	2502.5 Mbps
10	−55 dBm	3080.0 Mbps
11	−54 dBm	3850.0 Mbps
12	−53 dBm	4620.0 Mbps

thresholds recommended in the IEEE 802.11ad standard [1] as listed in Table 5.1. We transform the measured SNRs to sensitivity thresholds, lookup the recommended MCS, and, thus, obtain the expected bitrates and cross-validate them with our measurements.

In Figure 5.13, we show the average gain in terms of the expected bitrate. On both the up-link and the down-link ACO obtains on average about 50% higher bitrate, but Figure 5.14 shows that gains reach up to 2.54×. We achieve significant gains across most of the testbed. In a few cases, the achieved bitrate gain in Figure 5.14 is limited compared to the large SNR gain for the same location in Figure 5.8. This occurs in particularly challenging locations, such as the desks which are very close to the AP. Since the AP hangs from the ceiling, the vertical angle of the link towards the STAs located at those desks is very steep. Due to the layout of the antenna array (see Section 5.7.3), the vertical steering capability of the devices is limited. While those links are inherently weak, ACO still provides a substantial SNR improvement even in such challenging scenarios and enables the system to reach a higher MCS in most cases.

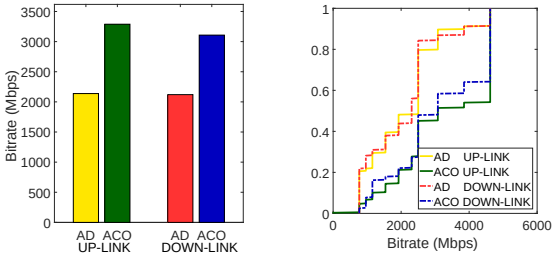


Figure 5.13: Average expected bitrate and CDF of our ACO mechanism and the default IEEE 802.11ad operation for up- and down-link measurements.

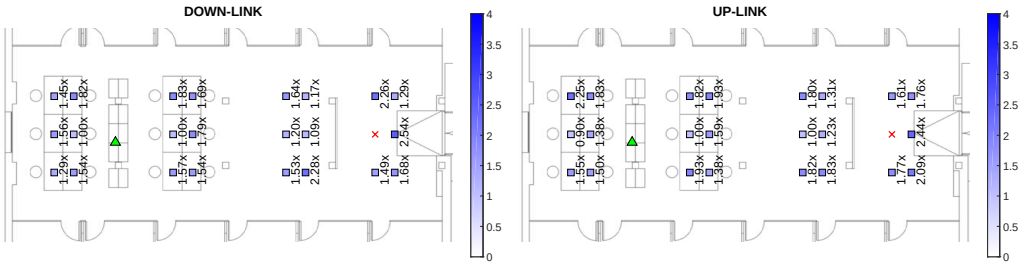


Figure 5.14: Map of the average expected bitrate gains of optimized beams. Each square represents a STA location, crosses indicate blind spots, while the AP is indicated by the green triangle.

5.9.4. Channel Probing Overhead

ACO requires additional SNR feedback. Instead of feeding back only the ID of the best antenna beam pattern as with conventional beam training, the SNRs of all the beam patterns that were probed are required. Furthermore, the overhead to acquire full channel state information would be higher than that of conventional IEEE 802.11ad training. However, measuring the full CSI for all antenna elements is usually not necessary. Our approach can flexibly adjust the number of active antennas or probing elements based on how rapidly the sub-channel changes, allowing us to reduce the overhead to achieve a certain accuracy. According to [70], probing a single sector in the sector sweep takes about 18.0 μs. Probing all $N = 32$ antenna elements with ACO would need $N + 4\,(N - 1) = 156$ probes (see Section 5.6.1.1), which results in a total probing time of about 2.8 ms. With 13 active antenna elements, we perform the beam training as fast as the IEEE 802.11ad sector sweep with 64 sectors in about 1.2 ms. As not all antenna elements contribute to good beam patterns, learning a sub-space of the channel with less active antenna elements is usually sufficient and provides a good trade-off between overhead and performance.

This flexible overhead also allows ACO to adapt to mobility which requires to train the beams continuously. By adapting the number of active antennas in the array, ACO can select a training overhead that matches the dynamics of the devices and environment. Specifically, ACO only needs to probe a small sub-space of the channel in case the

movement is small.

5.10. Summary

Obtaining full CSI on millimeter-wave devices enables beamforming to achieve higher directionality, allow for more accurate beam steering, and exploit multi-path effects caused by reflectors and obstacles in the environment. Unfortunately, this requires complex channel measurements which are not provided by commercial off-the-shelf devices. In this work, we propose Adaptive Codebook Optimization (ACO), a mechanism that extracts full CSI on such devices from non-coherent measurements. In particular, we engineer beam patterns with constant phase shifts that allow extracting CSI both in terms of phase and magnitude from simple SNR readings. As such readings are available on most devices, our approach is easily portable to any hardware that provides access to SNR information. We design ACO to incur the same or less probing overhead than IEEE 802.11ad. Moreover, we use the obtained CSI to compute beam patterns that adapt to the environment and maximize the SNR. In this way, we automatically avoid destructive interference and choose the best available path between transmitter and receiver. We implement ACO on TP-Link Talon AD 7200 tri-band routers by obtaining full access to the beamforming control of the integrated 32-element phased antenna array. To this end, we disassembled the phased antenna array and reconstructed the antenna weighting network experimentally. Our evaluations in a real-world office environment show that ACO increases the SNR by factor 2.5 and achieves a 2x higher TCP throughput. To support the community and allow other researchers to benefit from our results, we make our framework and firmware patches [102] as well as the source code of our actual implementation publicly available.

Part V

Location algorithms

Chapter 6

Location

6.1. Challenges

Millimeter wave (mmWave) communications in the 30–300 GHz band are considered key ingredients to achieve multiple Gbit/s link rates in 5G-and-beyond networks [103] as well as WLANs [104]. First mmWave devices following the IEEE 802.11ad standard are commercially available [105, 106].

mmWave signals follow quasi-optical propagation patterns, with clear reflections off boundary surfaces, and little scattering [107]. However, the high frequency of mmWave transmissions implies short coverage ranges. Moreover, mmWaves are blocked by a number of materials, including the human body [108]. Several materials reflect mmWave signals, providing alternative paths in case no LoS path is available. To achieve viable link ranges, mmWave devices employ directional, electronically steerable antenna arrays. The identification of the best steering direction for the antenna's main lobe is called beam training [109, 110]. For relatively static scenarios, brute force or hierarchical training as used, e.g., in the 802.11ad standard [62] work reasonably well. More refined methods exist that reduce beam training times, can track multiple paths at once, and work better in dynamic scenarios [111–113].

The quasi-optical propagation of mmWave signals, the sparse AoA spectrum that results, and the capability to track multiple components of the AoA spectrum, imply a very desirable consequence: that mmWave technology seamlessly allows mobile network devices to achieve SLAM through appropriately designed AoA-based methods. Such information is instrumental to mmWave networks, as it enables context awareness, and can facilitate beam training or handover operations, besides using it for location-based services and applications [114]. For example, gathering enough AoA information for device localization allows to do immediate handovers without the need for beam training to *any* AP for which only the position is known. To be viable for practical mmWave systems, a SLAM algorithm:

1. should ideally be run locally by each device, without any additional special-purpose messaging (which allows it to run on any 5G or 802.1ad compliant devices);

2. should work with CoTS devices, which are often low-cost, non-calibrated, and subject to computational constraints;

3. should not require to engineer or manually configure the network deployment (specifically the AP locations).

The constraints above have profound implications on the way a localization algorithm should be designed for mmWave networks. Specifically, to achieve requirement 1) no additional data should be exchanged among the devices, or between the device and the APs, requiring each device to achieve SLAM independently. Only existing information, AoA extracted from the beam training, can be leveraged for localization. Range-based mechanisms such as [115] are not viable on unmodified CoTS hardware and are thus incompatible with requirement 2). Moreover, the designed algorithm should have a low computational complexity, so that it can run in real-time even on constrained devices. Finally, due to 3) the AP locations are not known in advance and can't be distributed to the network devices (which would also require special-purpose messaging). Additionally, it prevents fingerprinting-based algorithms, which incur a very high network setup cost.

6.2. Related work

Localization has been largely studied from theoretical and practical standpoints [116]. However, it has been recently acknowledged that localization is an inherent feature of mmWave communications [117], and is potentially achievable precisely with up to sub-centimeter accuracy thanks to the propagation characteristics of mmWave signals. Large-scale mmWave MIMO systems can also be leveraged for localization by detecting the changes in the statistics of sparse MIMO channel signatures [118]. From a theoretical standpoint, [119] also shows that in a number of practical cases it is possible to estimate both the position and the orientation of a user. This is in line with the capabilities of our proposed algorithm.

The quasi-optical propagation as well as the sparse AoA spectrum perceived by mmWave receiver enable single-anchor localization, with much better accuracy than achieved in microwave systems, especially if the environment and the AP locations are assumed unknown [120]. For example, [121] achieves this via two approaches belonging to the class of triangulation- and ADoA-based algorithms. In the same vein, [115] applies ranging and multilateration to to exploit LoS and NLoS arrivals for node localization.

6.3. Contribution

Our goal is to overcome the constraints and issues discussed in the challenges section, by designing a zero-initial information, zero-overhead, low-complexity SLAM algorithm. As a first building block, we approach the localization part of SLAM through ADoA information. This information can be derived from the output of the beam training and tracking operations that are periodically carried out for device/AP association. We remark that several approaches exist to practically track the components of a sparse mmWave AoA spectrum, including NLoS arrivals. Having NLoS information available makes it possible for our algorithm to work in the presence of realistic propagation issues, including blockage originating from the environment or from other mobile users.

Second, in light of our zero-initial information constraint, our algorithm estimates all data needed for SLAM, including the location of physical APs (which are the actual sources of LoS arrivals) and virtual APs (which are the virtual sources of NLoS arrivals, and can be modeled by mirroring the location of a physical AP through a surface that reflects the mmWave signal). This makes the problem much harder, but we show that it can still be tackled through ADoA information. Since typically LoS and NLoS arrivals from the same physical AP are available (although not necessarily at the same point in time), the coupling of virtual APs to their originating physical APs allows to estimate the location and shape of the boundaries of the environment and of the obstacles therein. We introduce two location algorithms to solve this problem

6.3.1. CALM

We limit the complexity of the location estimation algorithm. We achieve this via a fundamental reformulation of the ADoA joint AP and user localization algorithm that, that is adapted to the challenging case of zero-initial information. Our formulation is amenable to fast initialization procedures to find initial estimates of the AP and device locations, as well as to low-complexity updating mechanisms for successive refinements of such estimates as the device moves. To the best of our knowledge, this formulation has never been used in the literature related to AoA-based localization approaches. We note that our method is very different from traditional SLAM approaches as are used, for example, in the field of robotics. There, SLAM is typically a *dedicated* mechanism whose primary objective is to make the robots aware of the environment; it is often achieved through radars, laser or cameras, by leveraging movement direction information supplied by the robot's sensors, and in the presence of more landmarks than anchors available in a typical 60 GHz WLAN. On the contrary, our SLAM algorithm is *embedded* in the network: it relies only on operations that are carried out for standard communication. The information it generates is instrumental to optimize the network behavior in an anticipatory fashion. Additionally, it complies with CoTS hardware constraints, as it

does not rely on radar-like approaches or on special-purpose equipment.

6.3.2. JADE

In this chapter, we propose a localization algorithm tailored around the characteristics of mobile indoor mmWave networks. The algorithm is designed under realistic assumptions on channel identification capabilities and on the knowledge of the environment possessed by the user at the time of localization. In particular, for the latter, we expect the user to know nothing about their surroundings, except the locations of two access points, which is the minimum amount of knowledge required to compute a unique location estimate not subject to any rotational or scale ambiguities. Based on this assumption, we tackle the joint estimation of the user location as well as of any other access point deployed in the considered indoor area. The core of the location estimation process is the angle-difference-of-arrival (ADoA) approach, which is not affected by any bias on the absolute orientation of estimated angle-of-arrival (AoA) patterns. The algorithm extracts information from LoS and multipath signal components alike (unlike previous multipath-suppressing approaches such as [122]). Finally, it exploits the mobility of the node to be localized to improve the accuracy of position estimates for both the node and unknown anchors. Our algorithm has been designed so as to not require excessive computational complexity coming from signal processing or optimization steps. In addition, it is based on AoA spectrum information, which can be directly passed on by a node's receiving hardware [90], or directly derived from it, making it *de-facto* available to higher-layer protocols.

6.4. Model

A zero knowledge location problem is composed of two parts: anchor location estimationand device localization. Additionally, we can also add environment mapping to the problem. Consider the scenario in **Fig. 1 from CLAM**, where a single mobile device receives signals from four anchor nodes, of which $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are physical APs, whereas $\mathbf{x}_4$ is the virtual AP that models the source of NLoS signals from $\mathbf{x}_3$. We recall that $\mathbf{x}_4$ is obtained by mirroring the location of $\mathbf{x}_3$ through the reflecting wall. In subsequent locations $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$, the user leverages beam training information to compute the ADoA between the signals from every anchor pair. The arcs in the circles centered on $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$ in **Fig. 1** visualize ADoAs, and circle radii convey the corresponding AoAs. As a result, multi-path mmWave propagation helps increase the localization accuracy. The zero knowledge location problem consists in finding anchor locations that are compatible with the ADoA measurements taken at different positions, corresponding to angle measurements extracted from the standard beam training mechanism.

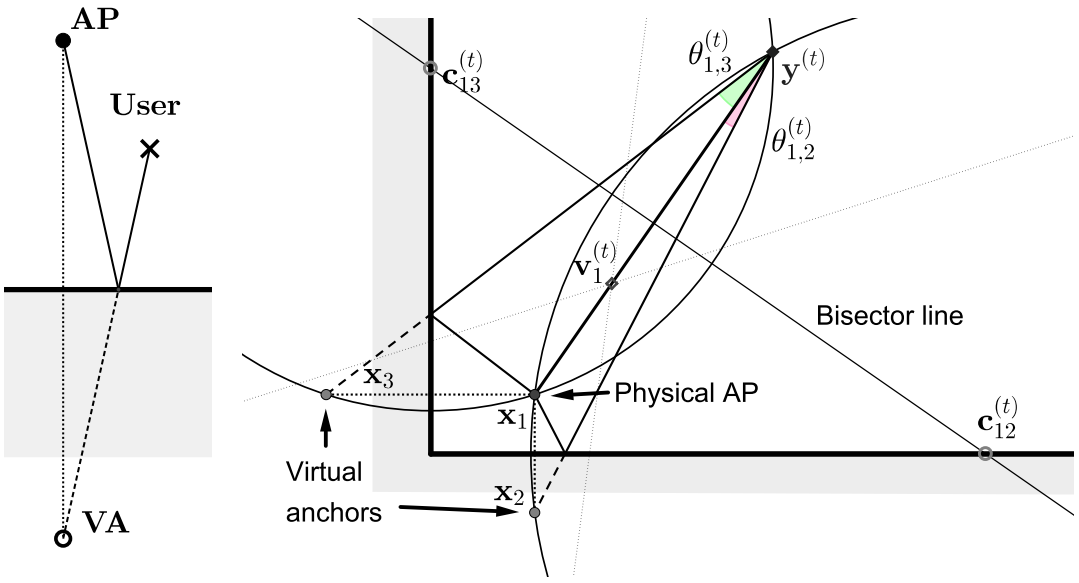


Figure 6.1: Virtual anchor corresponding to the reflection of the signal from a physical AP (a) and geometry of the ADoA localization process (b).

ADoA anchor localization is invariant to rotation, translation and scaling. We therefore consider the relative localization of the user and APs, and define a solution to the anchor localization process as the equivalence class of the valid anchor locations under the three above transformations. We call this class an anchor shape, or simply a shape, and denote it with the symbol $\mathcal{S}$. For example, the anchor shape of anchors in **Fig. 1** is their set of coordinates, along with any rotation, translation and scaling of these. Note that knowing the shape is equivalent to knowing all angles between access points. In the following, we will denote a set of anchors as $\mathcal{A}$, and the shape that refers to this set of anchors as $\mathcal{S}_\mathcal{A}$. We remark that all ambiguities can be resolved by knowing the coordinates of any two points in the localization area, e.g., the coordinates of two APs. However, using the location information for network optimization such as beam tracking or fast AP handovers does not require such disambiguation. A device can base these decisions on its own local coordinate system. The anchor shape can be estimated by observing different ADoAs from the same anchor set across different measurements taken at different user locations. The intuition is that a given anchor shape $\mathcal{S}$ generates a set of possible ADoAs $\Gamma_\mathcal{S}$. This set univocally determines a shape. Using multiple ADoA measurements, it is possible to detect a compatible shape by fusing the different measurements. Note that the anchor position estimation problem is underdetermined if the user sees just three anchors. Formally, for a given anchor set $\mathcal{A}$, $\Gamma_{\mathcal{S}_\mathcal{A}}$ can be computed by mapping every point of the 2D space into the ADoAs Θ that a user would observe at that point. This

defines a manifold of dimension 2 defined by the user position inside a space of dimension $|\mathcal{A}| - 1$. Therefore, when $|\mathcal{A}| > 3$, the set of possible ADoAs $\Gamma_{\mathcal{S}_\mathcal{A}}$ is contained in, but does not span, the full space of angle combinations, and can thus be used to infer the shape $\mathcal{S}_\mathcal{A}$. Therefore, the common strategy we adopt is assuming that the user measures ADoAs from exactly four anchors, and then extend it to any number of anchors.

6.4.1. Notation

We now briefly introduce the notation that will be used in this chapter. By a, $\mathbf{a}$ and $\mathbf{A}$ we denote a scalar, a vector and a matrix, respectively. $\mathcal{X} = \{\mathbf{x}_i\}_{i\in\mathcal{I}}$ denotes a set whose elements $\mathbf{x}_i$ are spanned by the index set $\mathcal{I}$. The set $\mathcal{X} = \{\mathbf{x}_i\}$ will be referred to as $\{\mathbf{x}_i\}$ when there is no ambiguity. More specific to the problem at hand, we will denote the set of physical anchor APs as $\mathcal{A}_p$ and the set of virtual anchors (VAs, defined as the reflection of a physical AP $a \in \mathcal{A}_p$ through a reflective surface as $\mathcal{A}_v$, so that $\mathcal{A} = \mathcal{A}_p \cup \mathcal{A}_v$ is the set of all anchors. The set $\mathcal{V}$ contains all pairs (i, t) such that at time t the user can measure the AoA corresponding to anchor i. $\|\cdot\|$ denotes the Frobenius norm, and $\mathbf{E}$ indicates the expectation. The notation we will use is the following. The location for the i-th access point will be denoted by $\mathbf{x}_i$ while $\mathbf{y}^{(t)}$ will denote the user at time t. The ADoA are denoted as $\theta_{ij}^{(t)} = \widehat{\mathbf{x}_i\mathbf{x}^{(t)}\mathbf{x}_j}$ and its complimentaries as $\xi_{ij}^{(t)} = \pi/2 - \theta_{ij}^{(t)}$ while the anchor angles are denoted by $\gamma_{ijk} = \widehat{\mathbf{x}_i\mathbf{x}_j\mathbf{x}_k}$. $\mathrm{vec}(\,\cdot\,)$ denotes the column-wise vectorization of the argument, ^{T} and ^{H} denote transposition and conjugate transposition respectively and $\otimes$ is the Kronecker product. $\ker(\cdot)$ denotes the kernel of the argument. $\mathbf{R}_\beta(\mathbf{x})$ is the counter-clockwise rotation of $\mathbf{x}$ by an angle β. Denote $\mathbf{a}^{-\mathbf{b}}$ the inversion of $\mathbf{a}$ to point $\mathbf{b}$ defined as $\mathbf{a}^{-\mathbf{b}} = \frac{\mathbf{a}-\mathbf{b}}{\|\mathbf{a}-\mathbf{b}\|^2} + \mathbf{b}$.

6.5. CLAM

6.5.1. Anchor shape estimation

This formulation of the ADoA anchor localization problem is based on determining a non-trivial implicit expression $\mathcal{R}(\mathcal{S}, \Theta)$ that ties an anchor shape $\mathcal{S}$ to a set of measured ADoA Θ, such that $\mathcal{R}(\mathcal{S}, \Theta) = 0$ if $\mathcal{S}$ is an anchor shape compatible with Θ.

Note that the anchor angles fully determine the anchor shape. The intuition behind the following development is that the relation $\mathcal{R}$ is a trigonometric polynomial of a certain order, whose terms are powers of trigonometric functions of the anchor angles and ADoA. This relationship can be expressed as

$$\mathcal{R}(\mathcal{S}, \Theta) = \sum_i \sum_j p_i m_{i,j} s_j = \mathbf{p}^{\mathrm{T}}\mathbf{M}\mathbf{s} = \mathrm{vec}(\mathbf{M})^{\mathrm{T}}(\mathbf{p}\otimes\mathbf{s}), \tag{6.1}$$

where each p_i and each s_j is a trigonometric polynomial function, this is a polynomial

with sine and cosine entries, of the ADoA and of the anchor angles, respectively; $\mathbf{p}, \mathbf{s}$ denote their vector representation (we single out a linearly independent set of terms), $\mathbf{M}$ is the matrix of polynomial coefficients. An order 3 term would be for example $\sin\theta_{13}^{(1)}\cos\theta_{24}^{(1)}\sin\theta_{24}^{(1)}$ for ADoA, or $\sin\gamma_{123}\cos^2\gamma_{234}$ for anchor angles.

If $\mathcal{R}(\Theta, \mathcal{S})$ exists and is non-trivial, we can generate a large number K of random shapes and locations and assemble the products $(\mathbf{p}\otimes\mathbf{s})_i$, $i = 1, \ldots, K$, into a matrix

$$\mathbf{X} = [(\mathbf{p}\otimes\mathbf{s})_1, (\mathbf{p}\otimes\mathbf{s})_2, \cdots (\mathbf{p}\otimes\mathbf{s})_K]\,. \tag{6.2}$$

This matrix satisfies $\mathrm{vec}(\mathbf{M})^{\mathrm{T}}\mathbf{X} = \mathbf{0}_{1\times K}$. Therefore,

$$\mathrm{vec}(\mathbf{M})^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathrm{vec}(\mathbf{M}) = 0, \text{ or } \mathrm{vec}(\mathbf{M}) \in \ker(\mathbf{X}\mathbf{X}^{\mathrm{T}})\setminus\{\mathbf{0}\} \tag{6.3}$$

We compute matrix $\mathbf{X}$ for different orders of the trigonometric polynomial functions of the ADoAs and anchor angles. Through automatic expression manipulation, we then prove that the minimum order that leads to a non-trivial kernel is 2 for the ADoA terms, and 3 for the anchors angle terms. In general, $\mathbf{M}$ represents a complex combination of polynomial terms with non-intuitive coefficients. To simplify the formulation we factorize the rank-5 matrix $\mathbf{M}$ as $\mathbf{M} = \mathbf{A}^{\mathrm{T}}\mathbf{B}$. By setting $\bar{\mathbf{p}} = \mathbf{A}\mathbf{p}$ and $\bar{\mathbf{s}} = \mathbf{B}\mathbf{s}$, we finally have that $\mathcal{R}(\mathcal{S}, \Theta) = \bar{\mathbf{p}}^{\mathrm{T}}\bar{\mathbf{s}} = 0$ for every anchor shape $\mathcal{S}$ compatible with the set of measured ADoAs Θ, thus reducing the dimension of our trigonometric vectors to 5. To the best of our knowledge, this is the first time that joint ADoA anchor estimation and localization is tackled using this method.

We remark that the generation of matrix $\mathbf{X}$, the identification of matrix $\mathbf{M}$ and the derivation of vectors $\bar{\mathbf{p}}$ and $\bar{\mathbf{s}}$ have to be done only once, after which they can be used in any anchor shape identification problem involving four anchors, as we detail in the next subsections. For simplicity, in the following we will drop the bar in the notation for $\mathbf{p}$ and $\mathbf{s}$.

6.5.2. Erroneous ADoA measurements

The formulation above is the solution to a geometric problem, and assumes error-free ADoA measurements. To take into account the measurement error, assume that N_{T} different measurements are obtained at different time epochs, indexed by t. For this set of measurements, ideally, the anchor shape would be the one that generates $\mathbf{s}$ such that $\mathbf{p}_t^{\mathrm{T}}\mathbf{s} = 0\ \forall t$, where we remark that the anchor angle terms $\mathbf{s}$ do not depend on t. With erroneous measurements, we cannot achieve the equality, hence we resort to a minimum mean square error (MMSE) approach, by defining the cost function

$$\mathcal{F}(\mathcal{S}) = \sum_{t}^{N_{\mathrm{T}}}(\mathbf{p}_t^{\mathrm{T}}\mathbf{s})^2 = \mathbf{s}^{\mathrm{T}}\left(\sum_{t=1}^{N_{\mathrm{T}}}\mathbf{p}_t\mathbf{p}_t^{\mathrm{T}}\right)\mathbf{s} = \mathbf{s}^{\mathrm{T}}\mathbf{O}\mathbf{s}\,, \tag{6.4}$$

where $\mathbf{O} = \sum_{t=1}^{N_\mathrm{T}} \mathbf{p}_t \mathbf{p}_t^\mathrm{T}$. The solution to the anchor shape estimation problem is then obtained as

$$\widehat{\mathcal{S}} = \arg\min_{\mathcal{S}} \mathcal{F}(\mathcal{S}) \,. \tag{6.5}$$

In the next section we discuss a practical algorithm that computes and refines the anchor shape with any number of anchors, in a way that is robust to mmWave path obstruction.

6.5.3. Extension to more than four anchors

Starting from the derivation in Section 6.5.2, we take all the possible subsets of four anchors, compute the corresponding cost function, and finally sum the obtained values into a global cost function as follows. Call $\mathcal{C}_\mathcal{A}^4$ the set of all possible subsets of four anchors. The global cost function is defined as:

$$\mathcal{F}_\mathcal{A}(\mathcal{S}_\mathcal{A}) = \sum_{\mathcal{I} \in \mathcal{C}_\mathcal{A}^4} \mathcal{F}_\mathcal{I}(\mathcal{S}_\mathcal{I}) = \sum_{\mathcal{I} \in \mathcal{C}_\mathcal{A}^4} \mathbf{s}_\mathcal{I}^\mathrm{T} \mathbf{O}_\mathcal{I} \mathbf{s}_\mathcal{I} \,, \tag{6.6}$$

where $\mathcal{F}_\mathcal{I}, \mathcal{S}_\mathcal{I}, \mathbf{s}_\mathcal{I}, \mathbf{O}_\mathcal{I}$ are the cost function, the anchor shape, the vector $\mathbf{s}$ and the objective matrix $\mathbf{O}$ considered in Section 6.5.2 for the set of anchors $\mathcal{I} \in \mathcal{C}_\mathcal{A}^4$.

We now extend the formulation to account for mmWave path blockage. Note that this causes two separate issues: *i*) matrix $\mathbf{O}_\mathcal{I}$ could be impossible to compute due to lack of visibility for some anchors in $\mathcal{I}$; *ii*) some anchor locations may be impossible to estimate, e.g., because they have not been observed in a sufficient number of measurements. Define $\mathcal{A}_v^t \subset \mathcal{A}$ as the set of visible anchors at time t. Issue *i*) is solved by considering only the measurement indices t for which all anchors of $\mathcal{I}$ are visible. To achieve this, we consider an update strategy for matrix $\mathbf{O}_\mathcal{I}$. As the algorithm assumes no initial knowledge, at $t = 0$ we have $\mathbf{O}_\mathcal{I} = \mathbf{0}$. At time t, only the anchor sets $\mathcal{I} \subset \mathcal{A}_v^t$ permit to update $\mathbf{O}_\mathcal{I}$. Call $\mathcal{C}_{\mathcal{A}_v^t}^4$, the set of all subsets of four anchors among those that are visible at time t. For $\mathcal{T}_\mathcal{I}^\mathrm{T} = \{1 \leq \tau \leq t \text{ s.t. } \mathcal{I} \in \mathcal{C}_{\mathcal{A}_v^\tau}^4\}$, we compute

$$\mathbf{O}_\mathcal{I} = \sum_{\tau \in \mathcal{T}_\mathcal{I}^t} \mathbf{p}_{\tau\mathcal{I}} \mathbf{p}_{\tau\mathcal{I}}^\mathrm{T} \,, \tag{6.7}$$

where $\mathbf{p}_{\tau\mathcal{I}}$ is the vector $\mathbf{p}$ considered in Section 6.5.2 for set $\mathcal{I} \in \mathcal{C}_\mathcal{A}^4$ at time τ. Eq. (6.7) allows us to include the contribution of all valid measurements by adding one more term to each sum whenever a measurement is taken (i.e., t increases by 1).

To solve issue *ii*), we start by estimating the anchor shape for subsets of anchors that have been observed in a sufficient number of measurements. Then, we iteratively include additional anchors until there are no more anchors that can be accurately located. To do this, we design a set of criteria to compute a first estimation of four anchors, add one

additional anchor, and refine the estimation over the new set.

Having the update algorithm for the $\mathbf{O}_{\mathcal{I}}$ matrices, and starting from $\mathcal{A} = \varnothing$ at $t = 0$, we pick the first set of four anchors $\mathcal{A}^* \in \mathcal{C}^4_{\mathcal{A}}$ to be the first combination such that all anchors in the set appear at least in θ_{I} measurements. θ_{I} is called the "initialization threshold." The initial estimation of the anchor locations is then computed as $\hat{\mathcal{A}}^* = \arg\min_{\mathcal{A}^*} \mathcal{F}_{\mathcal{I}}(\mathcal{S}_{\mathcal{I}})$. We achieve this via a Nelder-Mead simplex direct search [123] using as an initial point the minimum cost element of a set of 2^{20} anchor locations configurations[1] drawn independently at random according to a Gaussian $\mathcal{N}(0, 1)$ distribution. We remark that the mean and variance of the distribution are not relevant, since the problem is invariant to translation and scaling. To avoid accuracy issues, after every estimation of $\mathcal{A}^*$, we will normalize it to have mean 0 and variance $|\mathcal{A}^*|^{-1}$.

We now extend set $\mathcal{A}^*$ by adding anchors from $\mathcal{A} \backslash \mathcal{A}^*$, provided that each has appeared in at least θ_E measurements along with at least three anchors already in $\mathcal{A}^*$. θ_E is called the "extension threshold," and allows the position of the new anchor to be accurately estimated. Call the inserted anchor α. Its location is estimated by computing $\hat{\mathcal{A}}^* = \arg\min_{\alpha} \mathcal{F}_{\mathcal{A}^*}(\mathcal{S}_{\mathcal{A}^*})$ through a grid search over 2^{14} points in $[-2, 2] \times [-2, 2]$, followed by a Nelder-Mead simplex direct search. The complexity of this step can be limited by restricting the computation of $\mathcal{F}_{\mathcal{A}^*}(\mathcal{S}_{\mathcal{A}^*})$ to the θ_{R_E} anchor combinations $\mathcal{I} \in \mathcal{C}^4_{\mathcal{A}^*}$ that appear the most in all measurements (where the subscript E refers to the extension step), and by capping the Nelder-Mead running time to θ_{T_E} seconds.

All location estimates for all anchors in set $\mathcal{A}^*$ (including the newly added anchor) are then refined by solving

$$\widehat{\mathcal{A}}^* = \arg\min_{\mathcal{A}^*} \mathcal{F}_{\mathcal{A}^*}(\mathcal{S}_{\mathcal{A}^*}) = \arg\min_{\mathcal{A}^*} \sum_{\mathcal{I} \in \mathcal{C}^4_{\mathcal{A}^*}} \mathcal{F}_{\mathcal{I}}(\mathcal{S}_{\mathcal{I}})\,. \tag{6.8}$$

As before, we restrict the above sum to the θ_{R_R} combinations of anchors that appear the most in $\mathcal{I} \in \mathcal{C}^4_{\mathcal{A}^*}$, and cap the running time of the Nelder-Mead method to θ_{T_R} seconds.

6.5.4. User localization

Once the anchors location have been estimated, we can localize the user. As before, assume first that ADoA measurements error-free. Localizing the user means therefore to find a location compatible with the ADoAs observed from the physical and virtual anchors whose position has been estimated. We take anchor $\mathbf{x}_i$ as a reference, and use anchor $\mathbf{x}_j$ to determine the locus of points that generate the angle θ_{ij}. The above locus is the arc of circumference that contains both $\mathbf{x}_i$ and $\mathbf{x}_j$, centered at $\mathbf{R}_{\xi_{ij}}(\mathbf{x}_j - \mathbf{x}_i)/\sin(\theta_{ij}) + \mathbf{x}_i$. Anchor $\mathbf{x}_i$ is contained in this circumference, so we can consider inverted geometry respect

[1] Note that the initial search involves the computation of only one term of the sum in (6.7), and is run only once. It therefore represents a low startup cost, for which a resolution of 2^{20} elements can be easily afforded.

to this point, the inversion of this circumference to the point $\mathbf{x}_i$ is the line defined by the points $\mathbf{y}^{-\mathbf{x}_i}$ such that $\langle \mathbf{y}^{-\mathbf{x}_i} - \mathbf{x}_i \, , \mathbf{R}_{\xi_{ij}}(\mathbf{x}_j - \mathbf{x}_i) \rangle = 2\sin(\theta_{ij})$. Combining the corresponding expressions for different indices j yields a system of equations of the form

$$\mathbf{Z}_i(\mathbf{y}^{-\mathbf{x}_i} - \mathbf{x}_i) = \mathbf{Y}_i. \tag{6.9}$$

In order to account for noisy measurements, we estimate $\mathbf{y}^{-\mathbf{x}_i} - \mathbf{x}_i$ by solving the MMSE problem

$$\widehat{\mathbf{q}} = \arg\min_{\mathbf{q}} \|\mathbf{Z}_i\mathbf{q} - \mathbf{Y}_i\|^2 = \widehat{\mathbf{q}} = (\mathbf{Z}_i^{\mathrm{T}}\mathbf{Z}_i)^{-1}\mathbf{Z}_i^{\mathrm{T}}\mathbf{Y}_i \, . \tag{6.10}$$

By solving $\widehat{\mathbf{q}} = \mathbf{y}^{-\mathbf{x}_i} - \mathbf{x}_i$ for point $\mathbf{y}$, we get the estimate

$$\widehat{\mathbf{y}}_i = \frac{(\mathbf{Z}_i^{\mathrm{T}}\mathbf{Z}_i)^{-1}\mathbf{Z}_i^{\mathrm{T}}\mathbf{Y}_i}{\left\|(\mathbf{Z}_i^{\mathrm{T}}\mathbf{Z}_i)^{-1}\mathbf{Z}_i^{\mathrm{T}}\mathbf{Y}_i\right\|^2} + \mathbf{x}_i \, . \tag{6.11}$$

Finally, we compute $\widehat{\mathbf{y}}_i$ for all reference anchors $\mathbf{x}_i$ and average the corresponding estimates to yield $\widehat{\mathbf{y}} = \mathbb{E}_i(\widehat{\mathbf{y}}_i)$.

6.6. JADE

6.6.1. Geometry of the ADoA localization process

We start by assuming that the AoAs are known without error. This assumption will be relaxed later.

Assuming the user is receiving a signal from anchors i and j at time t, with angles $\theta_{ij}^{(t)}$, $\mathbf{y}^{(t)}$ is located on an arc of circumference that stands upon segment $\overline{\mathbf{x}_i\mathbf{x}_j}$.

When a third anchor k is available, the intersection of the two arcs standing upon chords $\overline{\mathbf{x}_i\mathbf{x}_j}$ and, e.g., $\overline{\mathbf{x}_j\mathbf{x}_k}$ provides an estimate for the location of the node. Define $\mathbf{c}_{ij}^{(t)}$ as the center of the arc that contains anchors i and j and the user at time t $\mathbf{y}^{(t)}$. The formula for $\mathbf{c}_{i,j}^{(t)}$ can be found, e.g., in [121].

By construction, the center of every arc stands on the line that bisects the segment $\overline{\mathbf{x}_i\mathbf{y}}^{(t)}$. Fig. 6.1 shows an example of this with two circumferences of centers $\mathbf{c}_{12}^{(t)}$ and $\mathbf{c}_{13}^{(t)}$, both containing anchor 1 at $\mathbf{x}_1$. By defining $\mathbf{v}_i^{(t)} = 2\|\mathbf{y}^{(t)} - \mathbf{x}_i\|^{-2}(\mathbf{y}^{(t)} - \mathbf{x}_i)$, the equation of the bisector line is:

$$\mathbf{v}_i^{(t)\,\mathrm{T}}(\mathbf{c}_{ij}^{(t)} - \mathbf{x}_i) = 1, \tag{6.12}$$

Using the formula of the arc center, we have that

$$\mathbf{c}_{ij}^{(t)} = \frac{\mathbf{R}_{\zeta_{ij}^{(t)}}}{2\sin\theta_{ij}^{(t)}}(\mathbf{x}_j - \mathbf{x}_i) + \mathbf{x}_i, \tag{6.13}$$

and substituting (6.13) into (6.12) yields

$$\mathbf{v}_i^{(t)\,\mathrm{T}} \mathbf{R}_{\zeta_{ij}^{(t)}} (\mathbf{x}_j - \mathbf{x}_i) = 2 \sin \theta_{ij}^{(t)}. \tag{6.14}$$

Since the centers of all arcs that contain anchor i at time t lie on the line defined by (6.12), we observe that they also intersect in a single point, which by construction is symmetric to anchor i with respect to the bisector line in (6.12). This point is an estimate for the user's location. Therefore, solving (6.14) for $\mathbf{v}_i^{(t)}$ yields the location of the user at time t.

We now relax the assumption of perfect channel knowledge. As AoA estimates are affected by errors, the system of equations in (6.14) may become infeasible. The approximate solution that minimizes the quadratic error with respect to the true solution is given by the problem

$$\underset{\{\mathbf{v}_i^{(t)}\},\{\mathbf{x}_i\}}{\arg\min} \sum_{(i,t),(j,t)\in\mathcal{V}} \left(\mathbf{v}_i^{(t)\,\mathrm{T}} \mathbf{R}_{\zeta_{ij}^{(t)}} (\mathbf{x}_j - \mathbf{x}_i) - 2 \sin \theta_{ij}^{(t)} \right)^2, \tag{6.15}$$

where the sum is computed over all i, j, and t such that $(i,t), (j,t) \in \mathcal{V}$. We remark that the solution to this problem depends on the location of the anchor nodes $\mathbf{x}_i$, $i \in \mathcal{A}$, which is also unknown. In the following section we detail an efficient algorithm to estimate $\{\mathbf{v}_i^{(t)}\}$, $(i,t) \in \mathcal{V}$ and $\{\mathbf{x}_i\}_{i\in\mathcal{A}}$.

6.6.2. Algorithm

The objective function in (6.15) is an order-4 polynomial in a multi-dimensional space. To tackle the minimization of the objective function efficiently, we proceed by successive refinements. The basic idea of the algorithm is to overdetermine the system of equations resulting from (6.15), and to iteratively solve two MMSE problems over $\{\mathbf{v}_i^{(t)}\}$, $(i,t) \in \mathcal{V}$ and $\{\mathbf{x}_i\}_{i\in\mathcal{A}}$ for up to MaxIT iterations. This makes the optimization faster, as the search space at each step is limited. To the best of our knowledge, this is the first time that the ADoA localization problem is tackled using this method. For the moment, assume that an initial estimate of the anchors' locations is available. Sections 6.6.2.1 and 6.6.2.2 develop the steps of the algorithm. Building on the corresponding derivation, Section 6.6.2.3 finally presents how the initial anchor location estimate is obtained.

6.6.2.1. Optimization over $\{\mathbf{v}_i^{(t)}\}$

Call $\mathcal{J}^{(t)} = \{j \in \mathcal{A} \text{ s.t. } (j,t) \in \mathcal{V}\}$ the set of valid values for the anchor index j at time t. Because the pairs $(i,t), (j,t) \in \mathcal{V}$ contribute independently to (6.15), the solution to the minimization problem when the values of $\mathbf{x}_i$ are given can be computed for each pair $(i,t) \in \mathcal{V}$ as

$$\hat{\mathbf{v}}_i^{(t)} = \underset{\mathbf{v}_i^{(t)}}{\arg\min} \sum_{j\in\mathcal{J}^{(t)}} \left(\mathbf{v}_i^{t,\mathrm{T}} \mathbf{R}_{\zeta_{ij}^{(t)}} (\mathbf{x}_j - \mathbf{x}_i) - 2 \sin \theta_{ij}^{t} \right)^2. \tag{6.16}$$

Define $\mathbf{M}_i^{(t)}$ as the $2 \times |\mathcal{J}^{(t)}|$ matrix whose columns are

$$[\mathbf{M}_i^{(t)}]_{:,k} = \mathbf{R}_{\zeta_{ik}^{(t)}}(\mathbf{x}_j - \mathbf{x}_i) \,, \tag{6.17}$$

where $1 \leq k \leq |\mathcal{J}^{(t)}|$, j is the kth element of $\mathcal{J}^{(t)}$, and the colon notation $[\,\cdot\,]_{:,k}$ conveys that we refer to all elements of column k. Finally, call $\mathbf{b}_i^{(t)} \in \mathbb{R}^{1\times|\mathcal{J}|}$ the column vector whose kth entry equals $[\mathbf{b}_i^{(t)}]_k = 2\sin\theta_{ik}^{(t)}$, for $k \in \mathcal{J}^{(t)}$. This makes it possible to rewrite (6.16) as

$$\hat{\mathbf{v}}_i^{(t)} = \arg\min_{\mathbf{v}_i^{(t)}} \|\mathbf{v}_i^{(t)\,\mathrm{T}}\mathbf{M}_i^{(t)} - \mathbf{b}_i^{(t)}\|^2 \,, \tag{6.18}$$

so that the MMSE optimization problem in (6.18) is solved as

$$\mathbf{v}_i^{(t)} = (\mathbf{M}_i^{(t)}\mathbf{M}_i^{(t)\,\mathrm{T}})^{-1}\mathbf{M}_i^{(t)}\mathbf{b}_i^{(t)\,\mathrm{T}} \,. \tag{6.19}$$

6.6.2.2. Optimization over $\{\mathbf{x}_i\}$

After computing (6.19), we proceed to minimize the objective function over the locations of the anchors visible at time t, $\{\mathbf{x}_i\}$ s.t. $(i,t) \in \mathcal{V}$, by assuming that the terms $\mathbf{v}_i^{(t)}$ are given. Define the anchor location vector as

$$\mathbf{x} = \begin{bmatrix}\mathbf{x}_1^{\mathrm{T}} & \mathbf{x}_2^{\mathrm{T}} & \cdots & \mathbf{x}_{|\mathcal{A}|}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \,. \tag{6.20}$$

Since the expression $\mathbf{v}_i^{(t)\,\mathrm{T}}\mathbf{R}_{\zeta_{ij}^{(t)}}(\mathbf{x}_j - \mathbf{x}_i)$ is linear $\forall\, i,j,t$ s.t. $(i,t),(j,t) \in \mathcal{V}$, and the values of $\mathbf{v}_i^{(t)}$ are given, the objective function in (6.15) can be rearranged to yield the problem:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \sum_{(i,t),(j,t)\in\mathcal{V}} \left(\mathbf{q}_{ij}^{(t),\mathrm{T}}\mathbf{x} - 2\sin\theta_{ij}^{t}\right)^2 \tag{6.21}$$

where the column vector $\mathbf{q}_{ij}^{(t)}$ groups the coefficients that multiply each component of the respective $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{A}$ in (6.15). By vertically concatenating the terms $\mathbf{q}_{ij}^{(t),\mathrm{T}}$ into matrix $\mathbf{Q}$ and the terms $2\sin\theta_{ij}^{t}$ into matrix $\mathbf{b}$, (6.21) can be expressed as:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{Q}\mathbf{x} - \mathbf{b}\|^2 \,, \tag{6.22}$$

which is again an MMSE problem with solution

$$\widehat{\mathbf{x}} = (\mathbf{Q}^{\mathrm{T}}\mathbf{Q})^{-1}\mathbf{Q}^{\mathrm{T}}\mathbf{b} \,. \tag{6.23}$$

By iterating the computation steps described in this subsection and Section 6.6.2.1, the location estimates for the user and the anchors (both physical and virtual) converge to a solution. The following subsection explains how to obtain good initial estimates of

the anchor locations to start the optimization of the objective function.

6.6.2.3. Initial estimation of the anchor locations

To enable the first computation of (6.23), we re-arrange (6.15) so that it becomes amenable to a grid search procedure. First, note that the objective function to optimize in order to obtain an estimate of $\{\mathbf{x}_i\}$ can be derived from (6.15) when each $\mathbf{v}_i^{(t)}$ is defined by (6.19), and in turn $\mathbf{M}_i^{(t)}$ is defined by the relationship in (6.17), which depends on $\{\mathbf{x}_i\}$. An initial estimate for $\{\mathbf{x}_i\}$, can therefore be obtained as

$$\{\widehat{\mathbf{x}}_i\} = \underset{\{\mathbf{x}_i\}}{\arg\min} \sum_{(i,t)\in\mathcal{V}} \min_{\mathbf{v}_i^{(t)}} \left\| \mathbf{v}_i^{t,\mathrm{T}} \mathbf{M}_i^{(t)} - \mathbf{b}_i^{(t)} \right\|^2 . \tag{6.24}$$

Note that each term $\min_{\mathbf{v}_i^{(t)}} \left\| \mathbf{v}_i^{t,\mathrm{T}} \mathbf{M}_i^{(t)} - \mathbf{b}_i^{(t)} \right\|^2$ is the distance between the vector $\mathbf{b}_i^{(t)}$ and the space generated by the two rows of $\mathbf{M}_i^{(t)}$. By calling $\mathbf{w}_i^{(t)\,\mathrm{T}} = [\mathbf{M}_i^{(t)}]_{1,:}$ and $\mathbf{z}_i^{(t)\,\mathrm{T}} = [\mathbf{M}_i^{(t)}]_{2,:}$ the first and second row of $\mathbf{M}_i^{(t)}$, respectively, and by using the formula for the distance between a plane and a point, we can rewrite (6.24) as

$$\{\widehat{\mathbf{x}}_i\} = \underset{\{\mathbf{x}_i\}}{\arg\min} \sum_{(i,t)\in\mathcal{V}} \left\| \mathbf{b}_i^{(t)} \right\|^2 - \left\| \Pi_{\mathbf{w}_i^{(t)},\mathbf{z}_i^{(t)}}(\mathbf{b}_i^{(t)}) \right\|^2 , \tag{6.25}$$

where $\Pi_{\mathbf{w}_i^{(t)},\mathbf{z}_i^{(t)}}(\mathbf{b}_i^{(t)})$ is the projection of $\mathbf{b}_i^{(t)}$ onto the plane generated by $\mathbf{w}_i^{(t)}$ and $\mathbf{z}_i^{(t)}$. The problem in (6.25) is then equivalent to

$$\{\widehat{\mathbf{x}}_i\} = \underset{\{\mathbf{x}_i\}}{\arg\max} \sum_{(i,t)\in\mathcal{V}} \left\| \Pi_{\mathbf{w}_i^{(t)},\mathbf{z}_i^{(t)}}(\mathbf{b}_i^{(t)}) \right\|^2 , \tag{6.26}$$

and by applying the projection modulus formula, we get the final expression

$$\begin{aligned} \{\widehat{\mathbf{x}}_i\} = \underset{\{\mathbf{x}_i\}}{\arg\max} \sum_{(i,t)\in\mathcal{V}} & \frac{(\mathbf{w}_i^{(t)\,\mathrm{T}} \mathbf{b}_i^{(t)})^2 \|\mathbf{z}_i^{(t)}\|^2}{\|\mathbf{w}_i^{(t)}\|^2 \|\mathbf{z}_i^{(t)}\|^2 - (\mathbf{w}_i^{(t)\,\mathrm{T}} \mathbf{z}_i^{(t)})^2} \\ + & \frac{(\mathbf{z}_i^{(t)\,\mathrm{T}} \mathbf{b}_i^{(t)})^2 \|\mathbf{w}_i^{(t)}\|^2 - 2(\mathbf{w}_i^{(t)\,\mathrm{T}} \mathbf{b}_i^{(t)})(\mathbf{z}_i^{(t)\,\mathrm{T}} \mathbf{b}_i^{(t)})(\mathbf{w}_i^{(t)\,\mathrm{T}} \mathbf{z}_i^{(t)})}{\|\mathbf{w}_i^{(t)}\|^2 \|\mathbf{z}_i^{(t)}\|^2 - (\mathbf{w}_i^{(t)\,\mathrm{T}} \mathbf{z}_i^{(t)})^2} . \end{aligned} \tag{6.27}$$

We remark that the summation terms in (6.27) can be evaluated only if $\mathbf{w}_i^{(t)}$ and $\mathbf{z}_i^{(t)}$ are linearly independent. If, however, they lie on the same line, (6.27) can be evaluated as in (6.26).

Recall from the discussion that the solution to problem (6.15) is invariant to rotation, scaling and translation. To remove this ambiguity and allow the user to at least localize itself with respect to the anchors encountered over time, we choose to arbitrarily fix the location of two access points (with no loss of generality, $\mathbf{x}_1$ and $\mathbf{x}_2$), and set $\mathbf{x}_1 =$

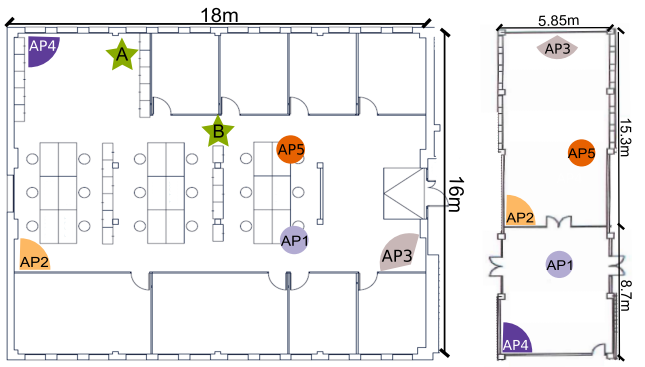


Figure 6.2: Floor plans of Scenarios A and B, showing the locations of mmWave APs and the antenna aperture for each AP (80°, 120° or omni).

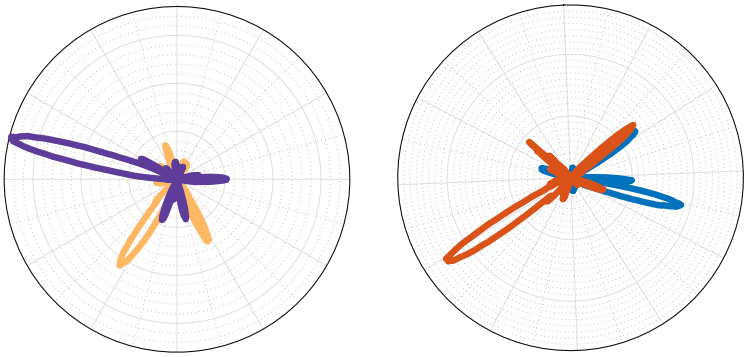


Figure 6.3: AoA spectra for APs 2 and 4 at position A and for APs 2 and 5 at position B. Both LoS and reflected NLoS paths are clearly distinguishable.

$[0,0]^{\mathrm{T}}, \mathbf{x}_2 = [1,0]^{\mathrm{T}}$. We then perform a grid search to solve (6.27) for $\mathbf{x}_3, \mathbf{x}_4$, by limiting the search space to the first four known anchors. This limits the grid search complexity while allowing to initialize the algorithm with a sufficient number of anchors (provided that at least four of them are visible). We then proceed by obtaining a first estimate of the location of every other known anchor, one anchor at a time. To do so, we solve (6.27) by grid search over $\mathbf{x}_n$ by limiting the scope of the search only to the first n anchors, $5 \leq n \leq |\bar{\mathcal{A}}^{(t)}|$, where $|\bar{\mathcal{A}}^{(t)}|$ is the set of anchors known at time t. This sequential optimization also aims at reducing the grid search complexity while achieving a suboptimal but sufficiently good initial solution. From this point, the optimization over $\{\mathbf{x}_i\}$ can be carried out as outlined in Section 6.6.2.2.

6.7. Experimental validation

6.7.1. Methodology

To test our algorithm in a real environment, we emulate mmWave devices.

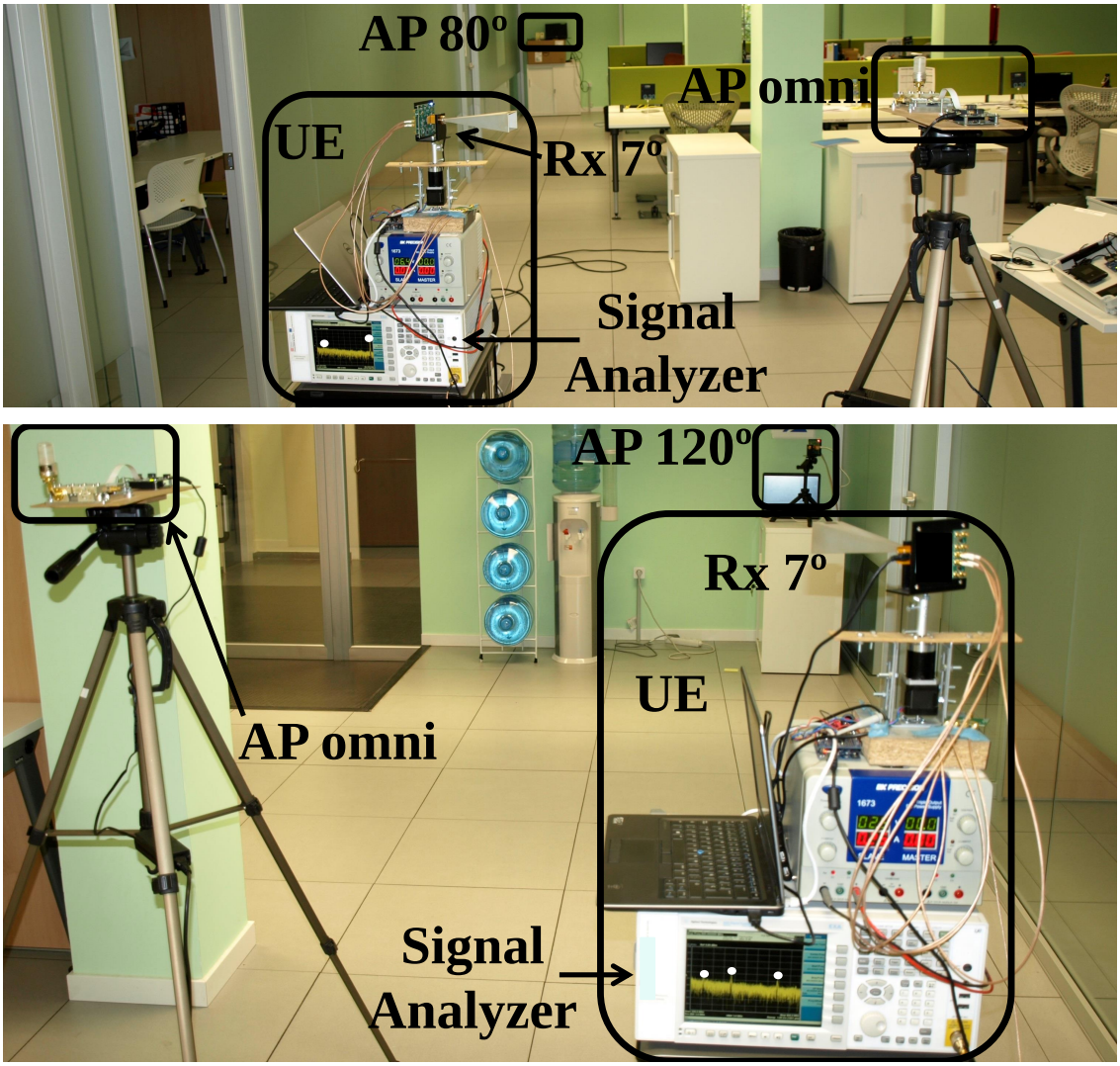


Figure 6.4: The measurement setup, showing APs and the receiver. The dots in the analyzer represent signals from different APs.

Specifically, for the user device we employ a Pasternack VubIQ 60-GHz down-converter with a 7°-aperture horn antenna. To emulate sector-level sweeping, we mount the VubIQ on a stepper motor with rotation step of 0.018°, controlled by an Arduino board. The baseband signal is recorded by an Agilent EXA N9010A Signal Analyzer. This whole set of devices is controlled by MatLab code running on a laptop. The system is configured to take mmWave power measurements in the 60 GHz band in angular steps of 0.45°, so that 800 measurements cover a complete circumference. For the APs we employ mmWave transmitters and five 60-GHz up-converters: one Pasternack VubIQ, two SiversIMA DC1005V/00 and two SiversIMA CO2201A. To detect the signal from all APs simultaneously without a separate signal generator for each AP, we mismatched the frequency of the local oscillator of all transmitters and of the receiver. Thus, the signals from different transmitters appear at different frequencies after down-conversion, and can be separately detected by the signal analyzer. This considerably speeds up the collection of measurement data in complex indoor scenarios.

Although a real mmWave communication deployment would employ directional transmissions, we employ pseudo omni-directional transmitters in order to speed up the collection of the measurements, while still being compatible with CLAM's ADoA measurement assumptions. As a general rule-of-thumb, we equipped APs located in open areas with an omni-directional antenna; where omni-directionality is not needed we employed wide-beam antennas (e.g., 80°-aperture horn antennas for APs near corners, and open wave-guide terminations translating into a 120°-degree antenna aperture for APs located near walls).

Once the measurements have been collected at different locations, we retrieve the **aoa!** (**aoa!**) patterns for each AP by isolating the corresponding frequency throughout all rotation steps. Different peaks in each pattern correspond to the **los!** (**los!**) arrival (when available) and to one or more NLoS arrival.

6.7.2. Room Setup

We tested our algorithms in two different scenarios. The first, Scenario A, is shown in Fig. 6.2, and is a fully functional L-shaped working place with an open area, six small offices and two labs. The building walls are composed of bricks and glass windows. Glass panels divide the offices from the open area. Office/lab furniture including tables, chairs, screens and cupboards is present in the whole environment. Given the number of available transmitters, we could not cover the offices, and rather restricted ourselves to the open area. However this already leads to very interesting localization and mapping results. The omni-directional transmitters were placed centrally in the open area, whereas the 80° horn antennas were positioned in the two west corners of the open area. In this scenario, we measured the AoA pattern from different APs at 66 positions. This makes it possible to emulate different trajectories through the open space, with an average

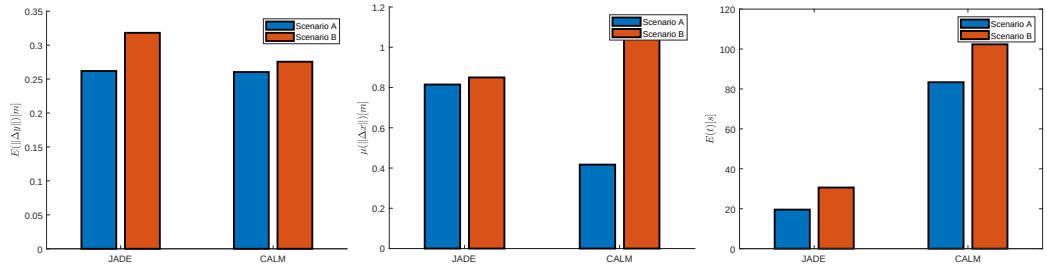


Figure 6.5: From left to right: median average user location error, median average access point location error and computational time complexity.

separation of 1.3 m between nearest measurement points. Note that furniture constrained the measurements, which were taken around the work stations, in empty areas, and along the corridors. In Fig. 6.4 we show two examples of the equipment and furniture in part of scenario A's measurement setup. To provide a good tradeoff between CLAM's accuracy and complexity, we set $\theta_I = 8$, $\theta_E = 5$, $\theta_{R_E} = 64$, $\theta_{R_R} = 128$. We do not set specific values for θ_{T_E} and θ_{T_R}, which provides an upper bound to localization performance. The typical running time is about 0.5 s on a mainstream laptop, with slightly longer times during CLAM's bootstrap phase.

Scenario B (Fig. 6.2) consists of a large room of total size about 50×90 m^2. The room has brick walls and glass windows, and is mostly empty of furniture. A wooden wall, interrupted by an open door, splits the space into two sections. In the smallest of the two sections we placed the omnidirectional transmitter, slightly shifted from the middle of the room, and deployed the 80°-aperture transmitter in a corner. In the largest space we set up a 80°-aperture transmitter in a corner, and an omnidirectional transmitter slightly south-west from the middle of the room. The 120°-beamwidth transmitter was centrally deployed along the top wall. For this setup, we measured 72 different positions in a scattered grid, where nearest locations are about 1.15 m apart. We remark that both experimental scenarios A and B represent extreme cases for our algorithm, given the low number of available APs.

In Fig. 6.3 we depict the received **aoa!** patterns at two positions marked in Fig. 6.2 as "A" and "B." For clarity, for position A we show the **aoa!** spectrum of APs 4 (purple) and 2 (orange). For position B, we show the **aoa!** spectra of APs 2 (orange) and 5 (brown). As expected, AoA spectra reveal multiple arrivals from different directions, with a LoS arrival of higher power, and weaker NLoS arrivals related to paths that incurred one or two reflections. AoA measurements are fed to the localization algorithms in order to estimate the location of the user. We now comment on the results of our experiments.

6.7.3. Measurements results

In Fig. 6.5 we illustrate the performance of JADE and CLAM over the data collected in our measurement campaign. The results are very accurate, given the shape of the room and presence of obstacles in the area. We believe Jade's CLAM's accuracies to be remarkable, also considering that they do not assume any initial knowledge about the user or the environment, and that our deployment is characterized by a low density of APs.

6.8. Simulation validation

To further extend the analysis of these location algorithms we proceed to simulate their behaviour under a controlled simulation environment. We choose to extend the results through simulation because unlike with experimental results we can generate as much data as we need on the fly to have a further analysis. The simulations will be structured as follows:
For each iteration we are going to generate N_{A} anchors and N_{U} user locations following distributions $\mathcal{N}(0, \sigma_{\mathrm{A}}^2\mathbf{I})$ and $\mathcal{N}(0, \sigma_{\mathrm{U}}^2\mathbf{I})$ respectively with $\sigma_{\mathrm{A}} = \sigma_{\mathrm{U}} = 10$ the respective distance standard deviation in meters.
Angles are computed from the relative location of anchor locations to user locations, then added a random number uniformly distributed in $[-\pi, \pi]$ to all angles to remove the orientation information. Then each angle is added an independent Gaussian noise with a standard deviation $\sigma_\phi = 2^\circ$ to take into account measurements imperfection. After considering the measurement inaccuracies we also include the not detection probability of an angle by eliminating it from the simulated data with probability $p_{\mathrm{b}} = 0.05$. Lately, we input the final angles corresponding to the anchors per user location into the zero knowledge location systems to run them.

We run the simulation for different number of user measurements and anchors to see the behaviour o both methods under different conditions. First, we set the number of user measurements while the number of anchors is set to 20 and after that we modify the number of anchors while the number of user measurements is set to 50. We expect the user and anchor location errors to vary with these parameters as well as the time complexity to run the algorithms.

To represent this three metrics we create three pairs of plots: Figure 6.6 holds the anchor location error analysis in which the height of the line represents the median average error, Figure 6.7 holds the anchor location error analysis in which the height of the line represents the median average error, Figure 6.8 holds the anchor location error analysis in which the height of the line represents the average time complexity. These sets have the results when modifying the number of anchors on the left plot and the results when modifying the number of user measurements on the right plot.

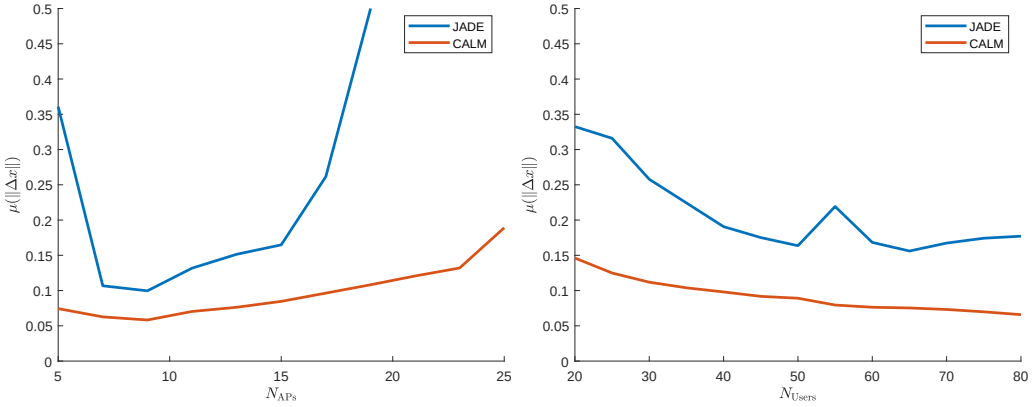


Figure 6.6: Anchors location accuracy when varying the number of anchors and user measurements respectively while keeping the number of user measurements and anchors to 50 and 20 respectively.

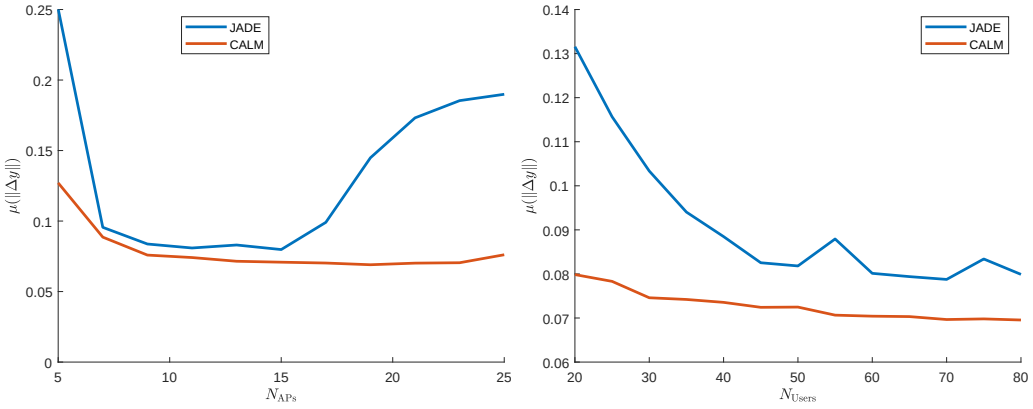


Figure 6.7: User location accuracy when varying the number of anchors and user measurements respectively while keeping the number of user measurements and anchors to 50 and 20 respectively.

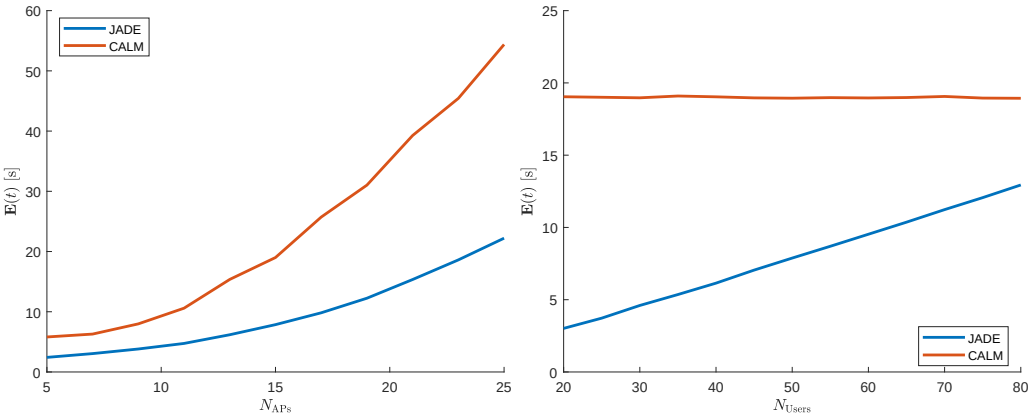


Figure 6.8: Computational complexity time when varying the number of anchors and users respectively while keeping the number of user measurements and anchors to 50 and 20 respectively.

The first observation we have is that generally, for the selected simulation parameters CALM performs better than JADE at the cost of a higher time complexity.

From the left plots in Figure 6.6 we can see that when increasing the number of anchors the access points location error first decreases but then it increases for both methods, this can be explained in terms of convergence, while by increasing the number of anchors we increase the information about our problem and thus the optimal solution is expected to be lower the problem has more parameters to estimate (new anchors' coordinates) increasing the complexity and decreasing the convergence rate. This specially affect JADE because it relies on an excess of extra variables per anchor to make the problem's objective linear. This effect has an impact on the user location error in the left plot of Figure 6.7, increasing it. The increase on the user location is not as notable as for the anchors because by increasing the number of anchors we also increase the amount of points user for each user's estimation, thus reducing the impact of the anchors location error has in the user location error.

As expected, both in Figure 6.6 and Figure 6.7 we can see that when we increase the number of user measurements we improve both method's performance. What's more interesting is the time complexity reaction. While JADE increases the number of restrictions and variables for computation CALM simply has to update its objective matrices to account for the additional measurements thus the time complexity of CALM stays stable when increasing the number of user measurements, this is something very important to consider when deciding which method to apply since while JADE returns a result much faster when the number of measurements is low CALM proves to be more effective with a large number of measurements due to its capability to increase its performance without increasing its computational time complexity.

6.9. Summary

In this chapter we proposed JADE and CALM, two algorithms that jointly localize a mobile user in an indoor space, estimates the position of the deployed access points, and opportunistically derive a map of the indoor area. The algorithms are tailored to mmWave equipment, and specifically hinge on the capability to estimate AoA information for different access points via the beam training procedure carried out by equipment compliant with the IEEE 802.11ad standard. Notably, these algorithms do not require any a-priori knowledge about the environment, neither about the floor plan of the indoor area, the position of reflective or absorbing surfaces therein, nor about the location and number of access points. Experimental results and an extensive set of simulation results demonstrate that these methods achieve remarkably good performance and sub-meter user and anchor localization accuracy, even for erroneous AoA estimates. We focus the simulation results to provide insights on the usefulness of when to apply each method over

the other one. We conclude that JADE generally performs worse but faster under the range of tested parameters being idea for minimalist scenarios like serving as a fast bootstrap for applications that require environment information like virtual reality calibration while CALM can tackle scenarios with immense amounts of user measurements without increasing its complexity due to its compact formulation making it ideal for scenarios with a large amount of measurements like applications that require long term estimations in a complex scenario like in an automated warehouse.

Part VI

Location aided beam training

Chapter 7

Location aided bean training

7.1. Challenges

Millimeter-wave (mmWave) communications have emerged as one of the most promising solutions to deliver multi-Gbit/s data rates to wireless devices. They are being standardized by 3GPP as a component of future mobile networks and have already become a WLAN standard through the introduction of IEEE 802.11ad operating in the 60 GHz band [124]. First CoTS mmWave devices started appearing on the market, such as the Netgear Nighthawk X10 [125] and the TP-Link Talon AD7200 [126] routers that implement the 802.11ad standard.

Radio propagation at mmWave frequencies exhibits a quasi-optical behavior, i.e., the useful energy received by a mmWave receiver propagates from the transmitter over a limited number of paths, typically a line-of-sight path and/or several low-order reflections. It is also characterized by a high path loss, and devices use beam forming antennas to compensate this loss with a higher antenna gain. Such antennas require beam training [127] to choose the beam pattern configuration that achieves the best signal quality over a point-to-point link. 802.11ad devices can use hierarchical beam training with different beam widths [128], but faster solutions have been proposed that take advantage of the sparsity of the mmWave multi-path channel for compressive beam training [129].

Most construction and furnishing materials are opaque to mmWaves, which implies that 802.11ad deployments experience blockage events and link outage may be frequent in the presence of obstructions [130], especially for mobile devices. Ubiquitous indoor mmWave coverage thus requires a very dense deployment of AP [131], so that ideally every location of an indoor space is served by more than one AP. This way, the probability of outage becomes small and the link quality improves thanks to the lower average distance between an AP and its associated clients. When the density of AP and clients increases, beam training can constitute a significant source of overhead for the network. In principle, each client would have to perform beam training with each AP in its vicinity to determine

which is the most suitable one. This makes optimal AP association a combinatorial problem, and hence scaling up the network density poses a significant challenge even with very efficient beam training mechanisms.

For such scenarios, location information is extremely beneficial, since it makes it possible to reduce or ideally avoid the beam training overhead. With a location system that provides device and AP positions with sufficient accuracy, it is possible to directly select the most suitable AP together with the beam pattern providing the highest gain in the known direction to that AP. By learning the environment, location systems also make it possible to prevent blockage by triggering handovers before the link to the associated AP is obstructed.

7.2. Related work

The idea of exploiting location information in order to improve network performance is envisioned in several works on indoor localization [115, 117, 132, 133]. However, while these works advocate the advantages offered by location information, they focus mainly on the location system and its performance per se, without exploring the resulting network performance of location-driven beam training and AP association. Other works in the literature combine location information and network optimization, but are not targeting mmWave systems. For example, [134] leverages location data to optimize the load of indoor femto-cells, whereas [135] considers automatic peer node discovery as a possible networking-related application.

7.3. Contribution

We address the above issues by implementing a 3D location system[1] that works solely based on the CSI extracted from the firmware of CoTS mmWave devices, specifically Talon AD7200 routers. The CSI includes the amplitude and phase of the most powerful propagation path that links two mmWave devices. By inferring the specific layout and configuration of the Talon antenna arrays, and thus the steering vectors required to point the array towards a given direction, the CSI thus collected makes it possible to estimate the directions from which a signal arrives to each AP. From this angle information, we derive the 3D location both of the AP and of the mmWave clients, and are thus able to learn the mapping between the location of a device and the quality of the links between the client and the surrounding AP. This makes it possible to develop simple, location-based algorithms to determine not only the best AP a client should associate with, given its current location, but also which beam pattern the client should use to achieve the

[1] 2D location systems are suboptimal since mmWave AP are typically deployed below the ceiling to avoid blockage, whereas devices are at table height.

best link quality and throughput. The resulting system, named Location Estimation And Predictive handover (LEAP), enables location-based, automatic handover and beam training mechanisms that scale to very high device densities and network dynamics.

To the best of our knowledge, this work is the first to employ 3D location information extracted from mmWave communications (without interfaces to external systems) to perform mmWave network optimization for client-AP association, handover, and beam selection. It is also the first work to implement such a system and provide quantitative results for the performance gains of location-based network optimization.

Our evaluation is based on experimental data collected through a testbed of seven Talon routers working as mmWave AP. We deploy our testbed in a large indoor space with furniture and separating walls, located such that they obstruct the LoS path from one or more AP as the client moves around the room. Except for the extraction of the CSI to be processed by our location system, the router operation is unmodified and remains standard-compliant. We demonstrate decimeter-level AP localization accuracy and sub-meter client localization accuracy in more than 80% of the cases. We further show that the handover optimization and optimal beam pattern selection process provides a throughput improvement between 8.5% and 57% with respect to the 802.11ad standard, and that links experience 2 to 9 dB higher SNR.

Specifically, we provide the following contributions:

- *A mmWave location system* that autonomously localizes both mmWave AP and clients in a given environment via CSI measurements. For this, we efficiently estimate the physical propagation paths that best explain the measured CSI, and translate this information into the direction of departure of the signal from the AP to the client;

- *Location-based handover and beam training* by exploiting location information, such that a client is more likely to avoid blockage and connect to the most suitable AP using the beam pattern providing the best link performance;

- *An implementation of our LEAP scheme using CoTS devices*, whose operation is unmodified except for the collection of CSI.

In the following, we present related work (Section 7.4), the design of LEAP's location system (Section 7.5), and of its optimized location-based handover and beam pattern selection schemes (Section 7.6). Next, we present our hardware platform and implementation (Section 7.7), evaluate our scheme (Section 7.8), and draw concluding remarks in Section 7.9.

7.4. Related Work

Indoor localization is becoming an increasingly important topic for networks, applications and services. Several schemes and systems are compared in [136] under practical indoor conditions. Among the tested schemes, [137] uses improved likelihood functions that achieve good WiFi fingerprint matching accuracy, whereas [138] relies on arbitrary array processing to extract AoA information. Source localization is made possible through the processing of multipath receptions from different AoA. Among the systems that achieve accurate localization in a WiFi context, Centaur [139] mixes radio and acoustic location fixes through Bayesian inference, in a way that is oblivious to the specific location system employed in each domain. ArrayTrack [140] exploits antenna arrays to suppress multipath, synthesize AoA from different antenna subsets, and localize a client in a multi-AP scenario. Up to decimeter-level accuracy is achieved by Chronos [135], a system which localizes a client by exploring the phase of the WiFi signal at different frequencies to measure the ToF between each of its antennas and the client.

Fingerprinting-based methods have also been recently revived by crowdsourcing the collection of measurements for the fingerprint database. For example, in [141], where a significant contribution is the use of RSSI differences, which even out the different devices characteristics while preserving the accuracy of the radio map. Similarly, ARLS [142] exploits crowdsourced RSSI maps to achieve room-level client localization. The theoretical framework in [143] suggests that cooperation represents a good solution to improve localization accuracy, especially if AoA, RSSI and ToF information can be successfully merged with a client's sensors. Practical implementations of this concept include enhancing BLE RSSI measurements with inertial sensor readings [144], or using pedestrian dead reckoning jointly with WiFi and BLE. All of the above systems operate at frequencies below mmWave.

mmWave systems show significant potential to achieve very high positioning accuracy [145], both for the localization of a client, and for the estimation of the environment around it using multipath propagation [133]. Notably, this can be achieved using range-based or AoA-based methods, with no need to resort to special hardware, measurement-intensive fingerprinting or ToF methods typical of ultrawideband systems [146]. Since the small wavelength makes it possible to integrate a very large number of antenna elements in a comparatively small form factor, massive mmWave arrays are feasible, which in turn enable high localization accuracy even with random beamforming [147]. With somewhat smaller arrays, maximum likelihood estimation and the exploiting the sparsity of the MIMO channel matrix in mmWave scenarios have been shown to provide low location errors [118]. In general, the best accuracy and lowest localization errors are enabled by range-based algorithms [115, 148]; however, ranging requires accurate path loss models, whose parameters are environment-specific and may have to be re-trained over

time. Multipath propagation can be exploited along with AoA information in order to localize a client [149] even with a single AP [121], although mmWave beamforming upon link establishment may turn out to sparsify the channel and reduce the number of useful multipath components [130]. Given sufficiently many reflected multi-path components, a mmWave location system can also be used to map the environment [22].

Our work clearly differs from the above literature, as we do not just design a 3D location system, but also exploit it to optimize network-related functions such as handover and initial access by significantly reducing the beam training or feedback between the AP and the client [150]. Unlike previous theoretical approaches [151], we implement our solution on CoTS devices, showing that the improvements of our location-based scheme can be achieved with consumer equipment.

7.5. Location system

7.5.1. Main idea

The main idea behind LEAP is as follows. We assume the presence of multiple AP in an indoor area. This is realistic given that indoor mmWave deployments are envisioned to be dense [131]. As a client moves, the AP measure the CSI of the link that connects them to the user. Multiple CSI measurements are converted into angle information in order to estimate the 3D location both of the AP and of the client. For each AP, we train a regression tree to learn the mapping between the location of the client and the SNR of its link to that AP. These regression trees help us decide, for any client location, which AP the client should connect to and which beam should be used. No coordination or feedback between the AP and the client is required, and the visibility of just a few AP at the same time (typically 2-3) is sufficient to achieve good results. Additionally, being able to select the correct beam based only on the locations of the AP' and the client significantly reduces the beam training time. The central result of our work is that a location system based only on information readily available to the physical layer of a set of mmWave AP suffices to enable location-aided beam training and handovers, greatly improving the network performance compared to the standard IEEE 802.11ad protocol.

7.5.2. Client localization algorithm

The CSI provided by the Talon router's firmware consists of one complex gain per antenna for the strongest multipath component of the received signal [152]. As a first step, our 3D location system converts the CSI into directions of arrival at each AP. Our formulation works in any number of dimensions, and is used to localize both the AP and the client in 3D space. We assume that each AP has a LoS connection to at least two AP, as needed to estimate the AP orientations. The quasi-optical behavior of the mmWave

channel implies that the energy collected by a given node has reached that node through a limited number of propagation paths. Call L the number of paths. We can define the uplink channel matrix towards AP a as

$$\mathbf{H} = \sum_{\boldsymbol{\ell=1}}^{\boldsymbol{L}} \boldsymbol{\alpha_\ell}\, \mathbf{s}_{\boldsymbol{a}}(\mathbf{v}_{\boldsymbol{\ell}})\, \mathbf{s}_{\boldsymbol{C}}(\boldsymbol{\xi_\ell})^{\mathrm{H}}\,, \tag{7.1}$$

where the superscript $^{\mathrm{H}}$ denotes conjugate-transpose, α_ℓ is the complex power gain of path ℓ (which includes path loss and phase shift), $\mathbf{v}_\ell$ and ξ_ℓ are the unitary vectors that define the direction of arrival at the AP and the direction of departure at the user of the ℓth path, whereas $\mathbf{s}_a(\mathbf{v}_\ell)$ and $\mathbf{s}_C(\xi_\ell)$ are the steering vectors of the array of AP a and of the client array that point towards directions $\mathbf{v}_\ell$ and ξ_ℓ, respectively. The CSI value measured by the receiver depends on the beam pattern $\mathbf{p}$ employed by the user, as well as on the above quantities. Specifically, the following vector represents the receive channel measured by AP a:

$$\mathbf{h}_{\boldsymbol{a}} = \sum_{\boldsymbol{\ell=1}}^{\boldsymbol{L}} \boldsymbol{\alpha_\ell}\, \mathbf{s}_{\boldsymbol{a}}(\mathbf{v}_{\boldsymbol{\ell}})\, \mathbf{s}_{\boldsymbol{C}}(\boldsymbol{\xi_\ell})^{\mathrm{H}}\mathbf{p} = \sum_{\boldsymbol{\ell=1}}^{\boldsymbol{L}} \boldsymbol{\beta_\ell}\, \mathbf{s}_{\boldsymbol{a}}(\mathbf{v}_{\boldsymbol{\ell}})\,, \tag{7.2}$$

where we define the scalar value $\boldsymbol{\beta_\ell} = \boldsymbol{\alpha_\ell}\, \mathbf{s}_{\boldsymbol{C}}(\boldsymbol{\xi_\ell})^{\mathrm{H}}\mathbf{p}$. The vector $\mathbf{h}_a$ is the CSI measured by AP a. Assuming that such CSI is mostly explained by a single main path [152], we can now estimate the vector of arrival $\bar{\mathbf{v}}_a$ (from the user to the receiving AP) that corresponds to such main path, and thus best explains the measured CSI. Assuming that the CSI is affected by circularly complex Gaussian noise, the estimate of $\bar{\mathbf{v}}_a$ is obtained as

$$\bar{\mathbf{v}}_a = \arg\min_{\mathbf{v}} \min_{\beta} \left\|\mathbf{h}_a - \beta \mathbf{s}_a(\mathbf{v})\right\|^2. \tag{7.3}$$

where $\mathbf{h}_a$ is as in (7.2). In the right-hand side of (7.3), we have

$$\left\|\mathbf{h}_a - \beta \mathbf{s}_a(\mathbf{v})\right\|^2 = \|\mathbf{h}_a\|^2 + |\beta|^2 \|\mathbf{s}_a(\mathbf{v})\|^2 - 2\,\mathfrak{Re}[\beta \mathbf{h}_a^{\mathrm{H}} \mathbf{s}_a(\mathbf{v})], \tag{7.4}$$

where $\mathfrak{Re}[\cdot]$ denotes the real part. The last term on the right-hand-side of (7.4) is the only one affected by the phase of β, and its minimum is equal to $-2|\beta||\mathbf{h}_a^{\mathrm{H}}\mathbf{s}_a(\mathbf{v})|$. It thus remains to minimize the quadratic expression

$$\|\mathbf{h}_a\|^2 + |\beta|^2 \|\mathbf{s}_a(\mathbf{v})\|^2 - 2\,|\beta|\,|\mathbf{h}_a^{\mathrm{H}} \mathbf{s}_a(\mathbf{v})|\,, \tag{7.5}$$

whose minimum value is $\|\mathbf{h}_a\|^2 - |\mathbf{h}_a^{\mathrm{H}}\mathbf{s}_a(\mathbf{v})|^2 / \|\mathbf{s}_a(\mathbf{v})\|^2$. Hence (7.3) simplifies to

$$\bar{\mathbf{v}}_a = \arg\min_{\mathbf{v}} \left(\|\mathbf{h}_a\|^2 - \frac{|\mathbf{h}_a^{\mathrm{H}} \mathbf{s}_a(\mathbf{v})|^2}{\|\mathbf{s}_a(\mathbf{v})\|^2} \right) = \arg\max_{\mathbf{v}} \frac{|\mathbf{h}_a^{\mathrm{H}} \mathbf{s}_a(\mathbf{v})|}{\|\mathbf{s}_a(\mathbf{v})\|}\,. \tag{7.6}$$

Eq. (7.6) can be solved once we know $\mathbf{s}_a(\mathbf{v})$ for a sufficiently large set of directions $\mathbf{v}$. This requires knowledge of the array topology of all routers, which can be obtained either from antenna modeling or by measuring a sufficiently large set of steering vectors in an anechoic chamber. Assume for the moment that we know the AP locations $\mathbf{y}_a$, where $1 \leq a \leq N_\mathrm{A}$ is the index of the AP, and N_A is the total number of AP deployed in the indoor space. Given $\bar{\mathbf{v}}_a$, we can compute the location $\mathbf{x}$ of the user. Ideally, if we knew the distance d_a between $\mathbf{x}$ and $\mathbf{y}_a$, we could directly compute the location of the user as $\mathbf{x} = \mathbf{y}_a + d_a\mathbf{v}_a$. Since our method is range-free, we have no means of estimating this distance directly. We solve this issue by formulating the location estimation step as a MMSE problem, where we obtain the estimated user location $\bar{\mathbf{x}}$ as the one that minimizes the difference between the actual location and the estimate $\mathbf{y}_a + d_a\bar{\mathbf{v}}_a$, for minimum distance d_a from any AP. Formally:

$$\bar{\mathbf{x}} = \arg\min_{\mathbf{x}} \min_{d_a, \forall a} \sum_{a=1}^{N_\mathrm{A}} \left\| \mathbf{y}_a + d_a\bar{\mathbf{v}}_a - \mathbf{x} \right\|^2 . \tag{7.7}$$

Minimizing each of the terms $\min_{d_a} \|\mathbf{y}_a + \mathbf{v}_a d_a - \mathbf{x}\|^2$, by definition, means finding the minimum distance between $\mathbf{x}$ and the line passing through $\mathbf{y}_a$ with direction $\mathbf{v}_a$. By applying the equation for the distance between a point and a line, we get

$$\begin{aligned} \min_{d_a} \left\| \mathbf{y}_a + \mathbf{v}_a d_a - \mathbf{x} \right\|^2 &= \left\| \mathbf{y}_a - \mathbf{x} \right\|^2 - \left((\mathbf{y}_a - \mathbf{x})^\mathrm{T} \mathbf{v}_a \right)^2 \qquad (7.8) \\ &= \|\mathbf{y}_a - \mathbf{x}\|^2 - (\mathbf{y}_a - \mathbf{x})^\mathrm{T} \mathbf{v}_a \mathbf{v}_a^\mathrm{T} (\mathbf{y}_a - \mathbf{x}) \\ &= (\mathbf{y}_a - \mathbf{x})^\mathrm{T} (\mathbf{I} - \mathbf{v}_a \mathbf{v}_a^\mathrm{T}) (\mathbf{y}_a - \mathbf{x}) \\ &= \mathbf{x}^\mathrm{T} (\mathbf{I} - \mathbf{v}_a \mathbf{v}_a^\mathrm{T}) \mathbf{x} - 2\mathbf{y}_a^\mathrm{T} (\mathbf{I} - \mathbf{v}_a \mathbf{v}_a^\mathrm{T}) \mathbf{x} + \\ &\quad + \mathbf{y}_a^\mathrm{T} (\mathbf{I} - \mathbf{v}_a \mathbf{v}_a^\mathrm{T}) \mathbf{y}_a . \qquad (7.9) \end{aligned}$$

Call $\mathbf{U} = N_\mathrm{A}\mathbf{I} - \sum_{a=1}^{N_\mathrm{A}} \mathbf{v}_a \mathbf{v}_a^\mathrm{T}$, $\mathbf{r} = \sum_{a=1}^{N_\mathrm{A}} \left(\mathbf{I} - \mathbf{v}_a \mathbf{v}_a^\mathrm{T} \right) \mathbf{y}_a$ and $z = \sum_{a=1}^{N_\mathrm{A}} \mathbf{y}_a^\mathrm{T} \left(\mathbf{I} - \mathbf{v}_a \mathbf{v}_a^\mathrm{T} \right) \mathbf{y}_a$. By substituting (7.9) into (7.7), we can rewrite (7.7) as

$$\bar{\mathbf{x}} = \arg\min_{\mathbf{x}} \mathbf{x}^\mathrm{T} \mathbf{U} \mathbf{x} - 2\mathbf{r}^\mathrm{T} \mathbf{x} + z . \tag{7.10}$$

This MMSE problem has solution $\bar{\mathbf{x}} = \mathbf{U}^{-1}\mathbf{r}$, and error

$$\varepsilon_{\bar{\mathbf{x}}} = z - \mathbf{r}^\mathrm{T} \mathbf{U}^{-1} \mathbf{r} . \tag{7.11}$$

7.5.3. AP localization algorithm

With (7.11), we can now formulate the AP localization algorithm. We first observe that (7.11) is in quadratic form. Define $\mathbf{y}$ as the vector containing the vertical concatenation of the coordinates of all access points $1 \leq a \leq N_\mathrm{A}$. With a suitable re-arrangement

of the terms in (7.11), we can write

$$\varepsilon_{\bar{\mathbf{x}}} = \mathbf{y}^{\mathrm{T}}\mathbf{Q}\mathbf{y}\,, \tag{7.12}$$

where matrix $\mathbf{Q}$ can be measured, and its terms depend on the CSI collected by each router. Call t the time epoch when a measurement is collected, and call $\mathbf{Q}_t$ the corresponding matrix. We can estimate the vector of the AP locations $\mathbf{y}$ as the one that minimizes (7.12) by taking into consideration all measurements carried out over time. This leads to the following minimization problem:

$$\bar{\mathbf{y}} = \arg\min_{\mathbf{y}} \mathbf{y}^{\mathrm{T}} \sum_t \mathbf{Q}_t \mathbf{y} \tag{7.13a}$$

$$\text{s.t.} \quad \|\mathbf{y}\| = 1 \tag{7.13b}$$

$$\textstyle\sum_{a=1}^{N_{\mathrm{A}}} [\mathbf{y}]_{3a-2} = 0 \tag{7.13c}$$

$$\textstyle\sum_{a=1}^{N_{\mathrm{A}}} [\mathbf{y}]_{3a-1} = 0 \tag{7.13d}$$

$$\textstyle\sum_{a=1}^{N_{\mathrm{A}}} [\mathbf{y}]_{3a} = 0\,, \tag{7.13e}$$

where the notation $[\cdot]_k$ denotes the element in position k of the vector in the square brackets. Call $\mathbf{T}$ any orthonormal parametrization of the subspace defined by the linear constraints (7.13c)–(7.13e); then $\mathbf{T}$ defines a bijection $\mathbf{y} = \mathbf{T}\check{\mathbf{y}}$ between the restricted space of $\mathbf{y}$ and a compressed version in a lower-dimensional space. In this way, the problem in (7.13) can be expressed as

$$\bar{\mathbf{y}} = \mathbf{T} \arg\min_{\check{\mathbf{y}}} (\mathbf{T}\check{\mathbf{y}})^{\mathrm{T}} \sum_t \mathbf{Q}_t (\mathbf{T}\check{\mathbf{y}}) \tag{7.14a}$$

$$\text{s.t.} \quad \|\mathbf{T}\check{\mathbf{y}}\| = 1\,. \tag{7.14b}$$

Since $\mathbf{T}$ is orthonormal, $\|\mathbf{T}\check{\mathbf{y}}\| = \|\check{\mathbf{y}}\|$ and (7.14) becomes

$$\bar{\mathbf{y}} = \mathbf{T} \arg\min_{\check{\mathbf{y}}} \check{\mathbf{y}}^{\mathrm{T}} \Big(\mathbf{T}^{\mathrm{T}} \sum_t \mathbf{Q}_t \mathbf{T} \Big) \check{\mathbf{y}} \tag{7.15a}$$

$$\text{s.t.} \quad \|\check{\mathbf{y}}\| = 1\,, \tag{7.15b}$$

which is a minimum eigenvector determination problem. Hence the vector that minimizes (7.15) is $\mathbf{E}_{\min}[\mathbf{T}^{\mathrm{T}} \sum_t \mathbf{Q}_t \mathbf{T}]$, where the $\mathbf{E}_{\min}[\cdot]$ operator returns the minimum eigenvector of its argument, and the solution to (7.13) is

$$\bar{\mathbf{y}} = \mathbf{T}\, \mathbf{E}_{\min} \Big[\mathbf{T}^{\mathrm{T}} \sum_t \mathbf{Q}_t \mathbf{T} \Big]\,. \tag{7.16}$$

7.5.4. Data smoothing

In order to remove outliers in the measurements and improve the data and location estimation quality, we implement three filtering steps. First, we ensure that the CSI fed to the localization algorithm is being measured from the LoS path. We thus design a filter to determine which measurements correspond to LoS paths and which ones should instead be discarded. A measurement is kept if the measured path satisfies two conditions: *i*) that the path power should exceed a threshold, i.e., $\|\mathbf{h}_a\| > \theta_P$, and *ii*) that the proportion of the measured power explained by the estimated path should also exceed a threshold, i.e., $|\mathbf{h}_a^{\mathrm{H}}\mathbf{s}_a(\bar{\mathbf{v}}_a)|/(\|\mathbf{h}_a\|\|\mathbf{s}_a(\bar{\mathbf{v}}_a)\|) > \theta_A$, where, e.g., $\theta_A = 0.9$.

Second, spurious CSI measurements provided by the AP should be filtered out to avoid that they contaminate the location estimates. We achieve this by eliminating the measurements that are not in agreement with the AP location estimates. We identify such spurious estimates as those for which the vector of departure $\bar{\mathbf{x}} - \bar{\mathbf{y}}_a$ remains far from the estimated direction of departure $\bar{\mathbf{v}}_a$, i.e., for which the cosine of the angle between the two vectors is small. We therefore check if $\|\bar{\mathbf{x}} - \bar{\mathbf{y}}_a\|^{-1}(\bar{\mathbf{x}} - \bar{\mathbf{y}}_a)^{\mathrm{H}}\mathbf{v}_a < \cos\epsilon$. If this is the case, we iteratively discard each measurement for which the left-hand-side of the above inequality is smallest (i.e., those leading to the largest discrepancies between $\bar{\mathbf{x}} - \bar{\mathbf{y}}_a$ and $\bar{\mathbf{v}}_a$), until all remaining measurements exceed the threshold.

In the same vein, we filter out client location errors as follows. We check if removing any of the AP coordinates included in vector $\mathbf{y}$ in (7.12) and the corresponding lines and columns of $\mathbf{Q}$ reduces the error $\varepsilon_{\bar{\mathbf{x}}}$ at least by a factor of 2. We then iteratively eliminate AP until we find no more that can be removed to improve the error, or until we are left with just two AP, which is the minimum number required to estimate the client position.

7.5.5. Mobility model and trajectory smoothing

We smooth the trajectory of the client via a rank-deficient Kalman filter. We consider the following mobility model

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \Delta t\mathbf{w}_t + \mathbf{x}_n \qquad \mathbf{x}_n \sim \mathcal{N}(\mathbf{0}, \Delta t\,\sigma_x^2\mathbf{I}) \tag{7.17}$$

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \mathbf{w}_n \qquad \mathbf{w}_n \sim \mathcal{N}(\mathbf{0}, \Delta t\,\sigma_v^2\mathbf{I})\,, \tag{7.18}$$

where $\mathbf{x}_t$ is the location of the client at time t, $\mathbf{w}_t$ its movement speed, Δt is the time interval between subsequent measurements, and σ_x^2 and σ_v^2 are the variance of each component of the position and velocity vector, respectively. The model in (7.17)–(7.18) correspond to uniform linear motion with location error $\mathbf{x}_n$ and speed error $\mathbf{w}_n$. Define the client state as the concatenation of the position and speed, i.e., $\mathbf{z}_t = [\mathbf{x}_t^{\mathrm{T}}, \mathbf{w}_t^{\mathrm{T}}]^{\mathrm{T}}$, and define the

matrices

$$\mathbf{E} = \begin{bmatrix} \mathbf{I} & \Delta t\mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \qquad \Sigma_z = \Delta t \begin{bmatrix} \sigma_x^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma_v^2\mathbf{I} \end{bmatrix}, \tag{7.19}$$

from which we can compute the state evolution as

$$\mathbf{z}_{t+1} = \mathbf{E}\mathbf{z}_t + \mathbf{z}_\mathrm{n}\,, \qquad \mathbf{z}_\mathrm{n} \sim \mathcal{N}(\mathbf{0}, \Sigma_z)\,. \tag{7.20}$$

If $\mathbf{z}_t \sim \mathcal{N}(\mu_t, \Sigma_t)$, we obtain the evolution formulas

$$\mu_{t+1} = \mathbf{E}\mu_t\,, \qquad \Sigma_{t+1} = \mathbf{E}\Sigma_t\mathbf{E}^\mathrm{T} + \Sigma_z\,, \tag{7.21}$$

where μ_0 and Σ_0 are initialized with the first measurement's value and uncertainty, respectively. By modeling the likelihood of the measurements as a normal distribution with mean $\bar{\mathbf{x}}_t$ and inverse covariance $\mathbf{U}_t^{-1}$, the posterior distribution of $\mathbf{z}_t$ is

$$\bar{\mu}_t = \bar{\Sigma}_t\left(\Sigma_t^{-1}\mu_t + \begin{bmatrix} \mathbf{U}_t\bar{\mathbf{x}}_t \\ \mathbf{0} \end{bmatrix}\right), \qquad \bar{\Sigma}_t = \left(\Sigma_t^{-1} + \begin{bmatrix} \mathbf{U}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right)^{-1}, \tag{7.22}$$

so that finally $\bar{\mu}_t$ predicts the client's position and velocity at time t. We remark that only a few measurements from typically 2-3 visible APs are sufficient to achieve a good position estimate. When location estimation should fail, the Kalman filter can be used for *dead reckoning*: namely, when a location cannot be computed from the CSI measurements, the filter automatically fills the gap by assuming that the client maintained the same speed vector.

7.6. Using location information for handover and beam selection

The accurate location system developed in Section 7.5 opens the way to extremely lightweight schemes to select the best AP. Specifically, we predict how the SNR of the link between the client and each AP will evolve over time, and use this information to rank the link quality with each AP and beam pattern. Considering a specific AP a, we proceed as follows.

For each measurement, after computing the posterior distribution of $\mathbf{z}_t$ in (7.22), we predict the location of the client over time for a given number N_T of future epochs spaced by a time interval τ. Call these locations $\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_{N_T}$. For each location $\hat{\mathbf{x}}_i$, we also compute the inverse square of the distance from AP a, namely $d_i^{-2}(a) = \|\hat{\mathbf{x}}_i - \hat{\mathbf{y}}_a\|^{-2}$. This serves as a measure of link quality. We pass the predicted locations of the client and the inverse squared distance from the AP to a regression tree trained to map these features

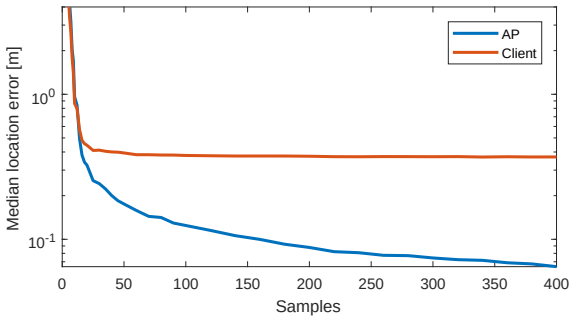


Figure 7.1: Median location estimation error in meters as a function of the number of measurements collected. About 15 measurements are sufficient to achieve sub-meter localization accuracy both for the client and for the AP.

and the SNR that would be experienced at any location. This yields N_T predicted SNR values for AP a, $\gamma_1, \ldots, \gamma_{N_T}$. Finally, we compute the score assigned to AP a as

$$\kappa_a = \sum_{n=1}^{N_T} \exp(-n\tau/\lambda)\gamma_n \,, \tag{7.23}$$

where λ is a custom parameter that regulates the exponential weighing of future predicted SNR values. After carrying out the above procedure for all AP, we finally instruct the client to connect to AP $\hat{a} = \arg\max_a \kappa_a$. By knowing the location of the AP and of the client, LEAP can potentially select the right beam without probing any AP. In practice, up to 4 or 9 beams are tested in order to ensure a correct handover in the presence of location errors. This is still a substantial reduction with respect to exhaustive or hierarchical beam training.

The system is bootstrapped by collecting enough measurements to estimate the location of the AP. LEAP achieves sub-meter location errors already with very few measurements. This is demonstrated by Fig. 7.1, which shows the evolution of the median localization error as a function of the number of measurements collected. We observe that even when no initial information is available to the user about its own location or the location of the AP, about 15 measurements are sufficient to achieve sub-meter median accuracy. Further measurements improve the AP locations even further, achieving an accuracy below 40 cm for about 50 measurements, and converging to sub-decimeter accuracy in the long run. In contrast, due to client mobility, only a limited set of measurements is available for each new client location and thus the median location error for the client converges to approximately 40 cm.

As soon as AP locations become available, we train a regression tree to learn the mapping between the user location and the SNR provided by each AP. Initially, the data set employed to train the regression tree is composed of the collected measurements, enriched with simulated data derived from a channel model. The simulated data is iteratively expunged from the training dataset and substituted with measurements as they become

available, until the regression tree can be trained using only measured data.

7.7. Hardware and methodology

We evaluate LEAP using TP-Link Talon AD7200 AP [126]. The routers integrate the Qualcomm QCA9500 chipset that implements the IEEE 802.11ad standard. A 32-element antenna array mounted on one of the eight foldable antenna wings provides beamforming capabilities. We modify the firmware to provide access to CSI measurements. To achieve the required level of control, we flash a compact LEDE linux distribution to the router [102]. LEDE uses the open-source `wil6210` driver to communicate with the firmware of the Qualcomm chipset. We implement custom vendor commands in the driver to expose CSI measurements to user-space programs. The router measures the CSI by sending 64 training subfields during a beam refinement phase, which make it possible to estimate both phase and amplitude of all 32 elements of the mmWave antenna array. With the Qualcomm firmware, only client-side CSI measurements are enabled. We circumvent this issue by switching the roles of the AP and of the client, such that the node to be located is actually the AP, whereas all other devices work in client mode and can thus measure CSI. Besides CSI, the driver provides the sector ID used for communications and the SNR measured for that sector.

We measure MAC-level throughput using the `iperf3` tool available in LEDE. In addition, the firmware makes it possible to implement custom 3D beam patterns by selecting which antenna elements to enable and specifying the phase shift values for each activated antenna with a 2-bit quantization. We prepare a codebook of 64 beam patterns that have significantly more directional main lobes than the default ones used by the router's firmware. An example is provided in Fig. 7.2: the left panel depicts one of the standard beam patterns of the original Talon router codebook, whereas the right panel shows our directional beam pattern. With these functionalities, it is possible to choose at any given time which beam pattern an AP and a device should use to communicate, thus greatly reducing the overhead of the beam training process through a location-aided beam pattern selection.

7.8. Experimental results

7.8.1. Experiment setup

Our measurement scenario consists of a laboratory room of size about 17 m $\times$ 6 m. We deploy a total of seven Talon routers along the perimeter of the room, at different heights. As the client to be located, we employ an additional Talon router. Two wood/metal panels of size 2 m $\times$ 2 m are located around the center of the room. These panels induce

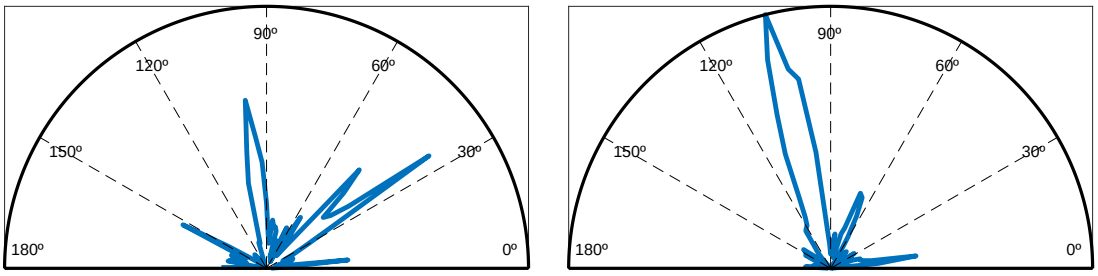


Figure 7.2: (Left) Beam pattern from the original Talon router codebook; (Right) directional beam pattern in our improved codebook. Amplitudes are plotted relative to the highest lobe of the beam pattern on the right.

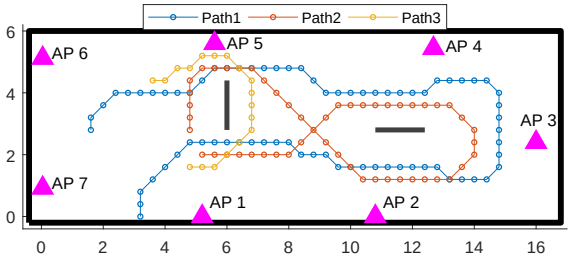


Figure 7.3: Experiment scenario: a lab with two internal panels acting as blocking walls. Seven APs are deployed along the boundaries of the room (purple triangles). Three examples of client trajectories are shown. Each marker represents a measurement point. The axes show lengths in meters.

a complete signal blockage if they obstruct the LoS path between the client and its AP, e.g., due to client mobility.

We collect measurements by moving the Talon acting as the client along a number of trajectories. At subsequent points spaced between 40 and 60 cm along such trajectories, we collect CSI, SNR and MAC throughput measurements from all AP. Fig. 7.3 gives an overview of the laboratory setup. Access points are depicted as purple triangles, internal panels are shown in gray, and we depict three example trajectories of our measurement collection process: path 1, that mostly follows the perimeter of the room and revolves around the two panels: path 2, that describes a figure-of-8 around the panels; and path 3, that describes an open loop close to the left side of the room. We remark that it is not necessary to know the location of the APs, which will be estimated by our scheme and fed to the location-based handover algorithm.

For LEAP, we select the best beam pattern among four (LEAP-4) or nine (LEAP-9) tested beam patterns covering the optimum transmit direction predicted through the location system. We compare our performance against that of standard 802.11ad with conventional beam training. For each measurement point, we let the 802.11ad beam training settle upon the best beam pattern. If the connection is lost and throughput drops to zero, we let the device reconnect to a new AP at the next step and perform a full beam training. The performance for 802.11ad thus represents an upper bound

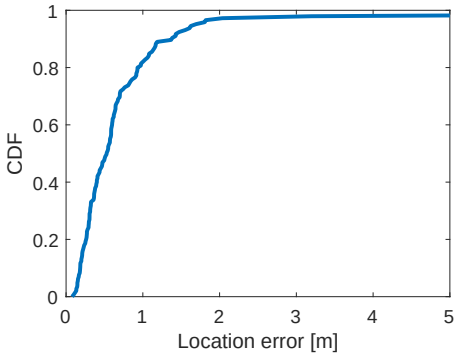


Figure 7.4: CDF of LEAP's location error for all measurements. Sub-meter accuracy is achieved more than 80% of the time.

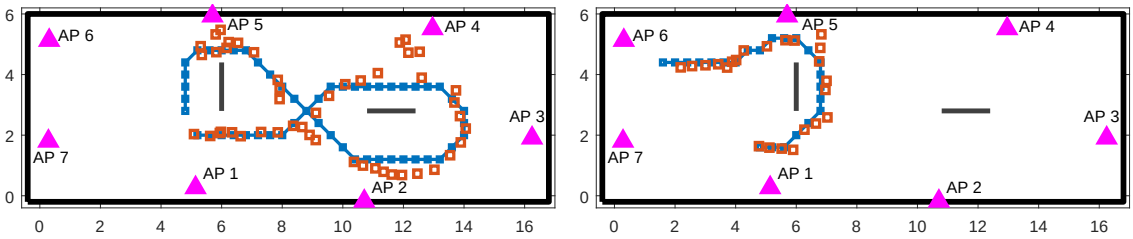


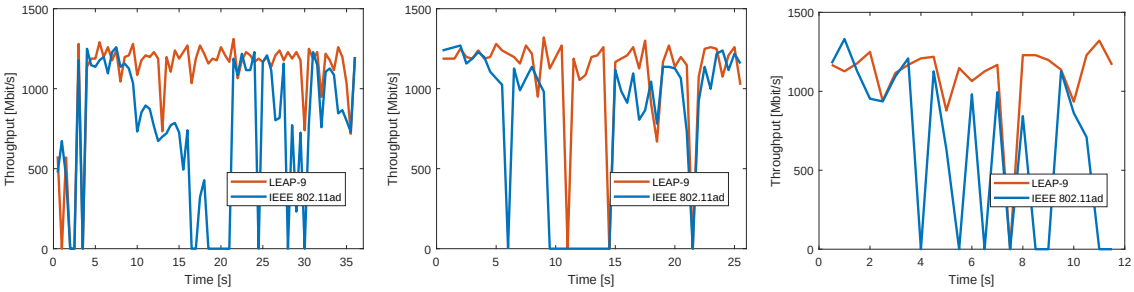
Figure 7.5: Path reconstruction for paths 2 (left) and 3 (right) in Fig. 7.3. The location estimates follow the ground truth reasonably well, and the errors are not critical for the selection of the best AP. The axes show lengths in meters.

on actual 802.11ad protocol performance with the Talon routers, which is characterized by sub-optimal beam fluctuations in case of mobility and noticeable reconnection delays when performing a handover to a different AP.

7.8.2. Results

We start by evaluating the accuracy of LEAP's localization system. Fig. 7.4 shows the CDF of the localization error, obtained by jointly considering all measurement trajectories in the experimental area. Despite the comparatively stretched room and the presence of internal blocking panels, the location system achieves sub-meter error in more than 80% of the cases, with a median location error of about 52 cm. Given that our method only uses angle information (unlike, e.g., [115]) and localizes the client in 3D (unlike, e.g., [133]), these results are very accurate. Moreover, as we will show below, this level of accuracy is more than sufficient to achieve good performance and significant gains in the location-aided handover process.

Fig. 7.5 provides more details of LEAP's location errors by showing the reconstruction of two trajectories among those depicted in Fig. 7.4, namely path 2 (left panel) and path 3 (right panel). Path 2 is a figure-of-8 trajectory that moves around both obstacles and thus experiences several changes in terms of AP visibility. In several cases, the localization error remains around a few decimeters. However, in other cases the collected CSI either leads to some errors (e.g., in the proximity of AP 4) or is insufficient to localize the



(a) Path 1 (b) Path 2 (c) Path 3

Figure 7.6: Measured MAC-level throughput for LEAP against the throughput achieved by IEEE 802.11ad.

node and has to be compensated for via the Kalman filter's dead reckoning capabilities (central section of the trajectory). Path 3 is comparatively simpler and subject to fewer handovers, except when the client disappears behind the internal panel and reappears on the other side. In any event, the trajectory reconstruction is very good, and the closeness of the estimated and actual locations makes it possible to still associate to the correct AP. For example, even for comparatively large errors near the top-right section of path 2, the client correctly connects to AP 4. Similar considerations apply, e.g., to the rightmost section of path 3.

We now show results for a client moving along paths 1 to 3, and plot the MAC-level throughput achieved over time by LEAP against the performance of 802.11ad in Fig 7.6. All trajectories are covered with the mmWave antenna array of the device facing forward, which therefore changes the orientation of the client at each curve. Note that firmware and hardware limitations prevent the Talon router from consistently maintaining transmission rates above 1.25 Gbps, and it does not use the fastest modulation and coding schemes. In any event, a conservative rate adaptation strategy makes sense, as the router only provides external connectivity through a 1-Gbps Ethernet port. Starting from path 1, we observe that the movement along its trajectory leads to several blockage events, which often cause reduced performance or even outage to 802.11ad. This is due to the obstruction of the LoS path to AP 1, to which the client connects at the beginning of the path. In the period from 7 to 22 s, 802.11ad maintains the connection to the same AP, which becomes progressively farther and gets blocked by obstacles, which reduces the throughput and causes even prolonged outages. With reference to the shape of path 2 in Figs. 7.3 and 7.5, initially both LEAP and 802.11ad connect to AP 1. As the client moves upwards in a clockwise direction, 802.11ad maintains the connection to AP 1, and a long outage occurs when the AP disappears behind the horizontal panel, at about 9 s. Conversely, LEAP performs a handover first to AP 4, and then to AP 3, 2 and 5. At the same time, AP 1 disappears again behind the horizontal panel, causing the 802.11ad outage at about 22 s. By handing the client connection over to the correct AP throughout the trajectory, LEAP

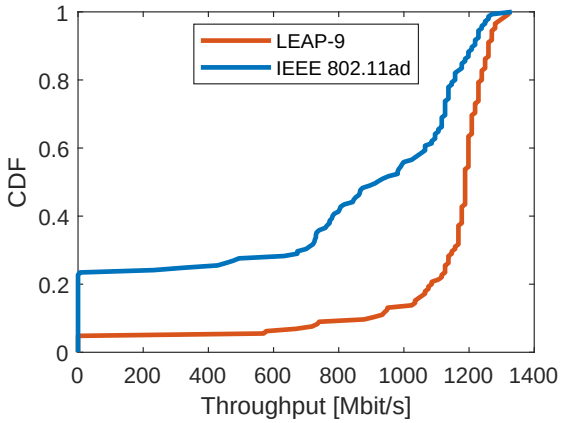


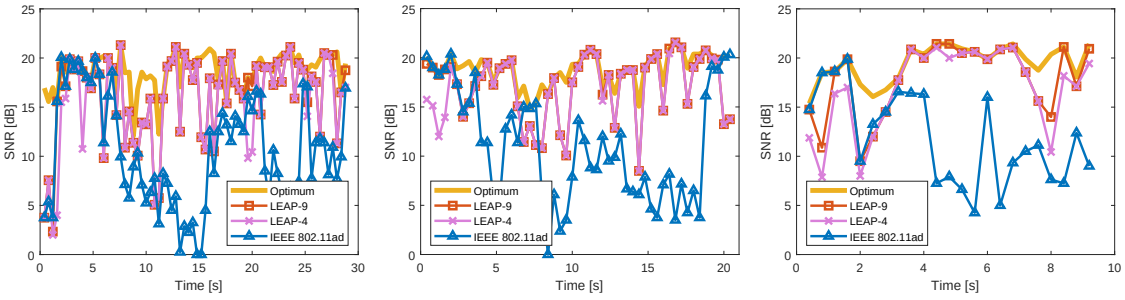
Figure 7.7: CDF of the MAC-level throughput for LEAP and 802.11ad. LEAP achieves better throughput an much lower outage probability than 802.11ad.

successfully avoids most outage events, and in many cases achieves better throughput than 802.11ad even in the absence of outages. In Fig. 7.6(b), the instances where LEAP's throughput decreases marginally below 802.11ad's are due to a location estimate that leads to the choice of a suboptimal AP and beam pattern.

For path 3 (Fig. 7.6(c)), LEAP connects to AP 5 at about 4 s (which is when the first outage occurs for 802.11ad), and then to AP 6, and suffers from a single outage at about 7.5 s. Conversely, suboptimal AP selection makes 802.11ad suffer from repeated outages. As a result, the throughput of LEAP is much closer to the maximum throughput achievable by the device throughout the whole trajectory.

The CDF of the MAC-level throughput computed over all tested trajectories is shown in Fig. 7.7. LEAP shows a much smaller outage probability (about 0.05) than 802.11ad (about 0.23). In fact, in about 85% of the cases, LEAP's throughput remains above 1 Gbps. For the cases where the 802.11ad is not in outage, LEAP still provides a throughput improvement between 8.5% and 57%. We remark that this improvement is actually limited by the hardware capabilities, not by LEAP's performance.

Therefore, to show the actual performance gain enabled by our algorithm, in Fig. 7.8



(a) Path 1 (b) Path 2 (c) Path 3

Figure 7.8: Measured SNR for LEAP-4, LEAP-9 and IEEE 802.11ad compared to an optimum, oracle-aided AP and beam pattern selection scheme.

we compare the SNR achieved by LEAP against that of 802.11ad, and against the optimum SNR that would be achievable by an oracle-aided selection of the best AP. The three panels refer to the same paths 1 to 3 as in Fig. 7.6. As a general observation, throughout each trajectory, the SNR achieved by both LEAP-4 and LEAP-9 remains much closer to the optimum, and in the majority of the cases is higher than the SNR achieved by 802.11ad. Also, the performance of LEAP-9 is only marginally better than that of LEAP-4. This indicates that the location-aided AP and beam pattern selection is sufficiently accurate, making it unnecessary to test many beam patterns. At the start of each path, LEAP and 802.11ad tend to connect to the same AP, hence their SNR is similar. However, as time elapses and the client moves, some AP become closer and thus more convenient to associate with, whereas others may disappear behind blockage. While LEAP keeps reacting to mobility and connects to (close to) optimal AP, 802.11ad re-associates to a different AP only upon severe blockage. The new AP is usually the first one from which a beacon is received, rather than the optimal one. As a consequence, 802.11ad's SNR is often 5 to 15 dB below the optimum. Instead, even in the presence of imperfect location estimates, LEAP still achieves near-optimal SNR, and only in a few cases do such imperfect estimates cause its SNR to drop a few dB below that of 802.11ad.

The CDF of the achieved SNR for the whole ensemble of our measurements is shown in Fig. 7.9. These results confirm that both LEAP-4 and LEAP-9 remain within 3 dB from the optimum 75% of the times, and improve the SNR by 2 to 9 dB compared to 802.11ad. We believe that these results constitute a very promising demonstration of how mmWave network performance can be improved with the aid of a sufficiently precise location system. This is specifically significant considering that such performance is achieved with only angle information related to the LoS path as provided by consumer-grade mmWave devices, and that we compared against an idealized version of 802.11ad that does not suffer from issues related to mobility.

7.9. Summary

We presented LEAP, a system that leverages location information to improve handover decisions and beam pattern selection using CSI and SNR measurements provided by commercial mmWave devices. LEAP localizes a client with sub-meter accuracy in the great majority of the cases, greatly reduces the beam training overhead via location-based beam pattern selection, and avoids blockage-induced outage by using regression trees to choose the best AP to associate with. We implement our approach using Talon AD7200 routers in a way that does not modify 802.11ad's operation. Compared to standard 802.11ad, LEAP achieves a throughput gain between 8.5% and 57% and an SNR typically within 3 dB of the optimum. Most importantly, the beam training overhead reduction enables LEAP to scale to very high device densities and environment dynamics.

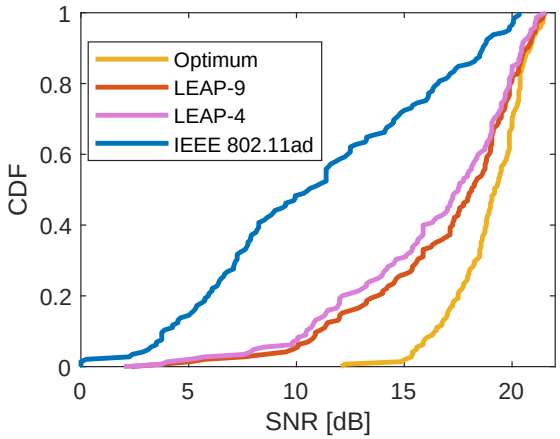


Figure 7.9: CDF of the SNR of LEAP-4, LEAP-9 and 802.11ad compared to an oracle-aided AP and beam pattern selection scheme. LEAP's SNR is higher than 802.11ad's by 2 to 9 dB, and is typically within 3 dB of the optimum.

Part VII

Conclusions

Chapter 8

Conclusions

During the PhD duration I focused on tools for improving mmWave communications and extract useful additional information like localization. We started by designing a baseline beam-pattern design to overcome the fast sector beam-pattern generation problem for mmWave phased antenna arrays. Then we created a method for hybrid beam-forming that did not only outperform the existing ones in accuracy but also was much faster and allowed for multiple beam-pattern generation. We used our sector beam-pattern design together with our Hybrid precoding method to create a beam-training and tracking method that was able to increase the communication throughput while reducing the training overhead. After that we moved to implementing this knowledge into commercial devices, thus generating a method to estimate channel from simple signal strength measurement and used this estimation to align the beam-patterns in a beam-training and tracking algorithm. Next logical step for improving the beam-training in mmWave was to exploit the geometric information of its quasi-optical channel, so we looked into practical location algorithms for our problem. Particularly we designed to zero-knowledge algorithms to make geometry estimations without the need of external information. These algorithms proved to be able to get location information out of communication by-product information such as angle difference of arrival. Finally, we created a location aided beam-training and tracking algorithm that combines location estimation with machine learning to make predictions in the channel changes to make predictive decisions on the communication beam-forming.

The combination of this work provided a complete set of tools for exploiting the channel characteristics of the mmWave band for communication from the beam-pattern design to the channel characteristics prediction.

References

[1] IEEE Standards Association, "IEEE Std 802.11ad-2012: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band. ISO/IEC/IEEE 8802-11:2012/Amd.3:2014(E)," 2014.

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

[3] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results," *IEEE Comm. Magazine*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[4] W. Hong, K.-H. Baek, Y. Lee, Y. Kim, and S.-T. Ko, "Study and prototyping of practically large-scale mmwave antenna systems for 5g cellular devices," *IEEE Communications Magazine*, vol. 52, no. 9, 2014.

[5] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmwave) for 5g: Opportunities and challenges," *Wireless Networks*, vol. 21, no. 8, 2015.

[6] T. A. Thomas, M. Cudak, and T. Kovarik, "Blind phase noise mitigation for a 72 ghz millimeter wave system," in *2015 IEEE International Conference on Communications (ICC)*, 2015.

[7] A. Dhananjay, "Iris: Mitigating phase noise in millimeter wave ofdm systems," Ph.D. dissertation, New York University, 2015.

[8] A. Patra, L. Simić, and P. Mähönen, "Smart mm-Wave beam steering algorithm for fast link re-establishment under node mobility in 60 GHz indoor WLANs," in *Proceedings of the 13th ACM Int. Symposium on Mobility Management and Wireless Access (MobiWac)*, 2015.

[9] M. Giordani, M. Mezzavilla, C. N. Barati, S. Rangan, and M. Zorzi, "Comparative analysis of initial access techniques in 5G mmWave cellular networks," in *2016 Annual Conference on Information Science and Systems (CISS)*, March 2016.

[10] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[11] ——, "Single-sided adaptive estimation of multi-path millimeter wave channels," in *2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2014.

[12] M. E. Eltayeb, A. Alkhateeb, R. W. Heath, and T. Y. Al-Naffouri, "Opportunistic beam training with hybrid analog/digital codebooks for mmWave systems," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL, USA, Dec. 2015.

[13] D. D. Donno, J. Palacios, D. Giustiniano, and J. Widmer, "Hybrid analog-digital beam training for mmWave systems with low-resolution RF phase shifters," in *2016 IEEE International Conference on Communications Workshops (ICC)*, May 2016.

[14] J. He, T. Kim, H. Ghauch, K. Liu, and G. Wang, "Millimeter wave MIMO channel tracking systems," in *2014 IEEE Globecom Workshops*, Dec 2014.

[15] J. Palacios, D. De Donno, D. Giustiniano, and J. Widmer, "Speeding up mmWave beam training through low-complexity hybrid transceivers," in *27th IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sept. 2016.

[16] A. Valdes-Garcia, S. Nicolson, J.-W. Lai, A. Natarajan, P.-Y. Chen, S. Reynolds, J.-H. Zhan, and B. Floyd, "A SiGe BiCMOS 16-element phased-array transmitter for 60Ghz communications," in *2010 IEEE Solid-State Circuits Conference (ISSCC)*, Feb. 2010.

[17] J. Palacios, D. De Donno, and J. Widmer, "Lightweight and effective sector beam pattern synthesis with uniform linear antenna arrays," *IEEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 605–608, 2016.

[18] J. Palacios, D. De Donno, D. Giustiniano, and J. Widmer, "Speeding up mmwave beam training through low-complexity hybrid transceivers," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 2016, pp. 1–7.

[19] J. Palacios, D. De Donno, and J. Widmer, "Tracking mm-wave channel dynamics: Fast beam training strategies under mobility," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[20] J. Palacios, D. Steinmetzer, A. Loch, M. Hollick, and J. Widmer, "Adaptive codebook optimization for beam training on off-the-shelf ieee 802.11ad devices," in *Proceedings of the 24th ACM Annual International Conference on Mobile Computing and Networking*, Oct. 2018.

[21] J. Palacios, P. Casari, and J. Widmer, "Jade: Zero-knowledge device localization and environment mapping for millimeter wave systems," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[22] J. Palacios, G. Bielsa, P. Casari, and J. Widmer, "Communication-driven localization and mapping for millimeter wave networks," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 2402–2410.

[23] J. Palacios, P. Casari, H. Assasa, and J. Widmer, "Leap: Location estimation and predictive handover with consumer-grade mmwave devices," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 2377–2385.

[24] D. De Donno, J. P. Beltrán, D. Giustiniano, and J. Widmer, "Hybrid analog-digital beam training for mmwave systems with low-resolution rf phase shifters," in *2016 IEEE International Conference on Communications Workshops (ICC)*. IEEE, 2016, pp. 700–705.

[25] M. C. Filippou, D. De Donno, C. Priale, J. Palacios, D. Giustiniano, and J. Widmer, "Throughput vs. latency: Qos-centric resource allocation for multi-user millimeter wave systems," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–6.

[26] D. De Donno, J. Palacios, and J. Widmer, "Millimeter-wave beam training acceleration through low-complexity hybrid transceivers," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3646–3660, 2017.

[27]

[28] G. Bielsa, J. Palacios, A. Loch, D. Steinmetzer, P. Casari, and J. Widmer, "Indoor localization using commercial off-the-shelf 60 ghz access points," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 2384–2392.

[29] J. Palacios, G. Bielsa, P. Casari, and J. Widmer, "Single- and multiple-access point indoor localization for millimeter-wave networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1927–1942, 2019.

[30] J. Palacios, J. Rodriguez-Fernandez, and N. Gonzalez-Prelcic, "Hybrid precoding and combining for full-duplex millimeter wave communication," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[31] J. Palacios, N. GonzÃ¡lez-Prelcic, and J. Widmer, "Managing hardware impairments in hybrid millimeter wave mimo systems: A dictionary learning-based approach," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 168–172.

[32] J. Palacios, D. Steinmetzer, A. Loch, M. Hollick, and J. Widmer, "Adaptive codebook optimization for beam training on off-the-shelf ieee 802.11 ad devices," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 2018, pp. 241–255.

[33] Y. Li, X. Ji, D. Liang, and Y. Li, "Dynamic beamforming for three-dimensional MIMO technique in LTE-Advanced networks," *International Journal of Antennas and Propagation*, vol. 2013, Jul. 2013.

[34] C. U. Saraydar and A. Yener, "Adaptive cell sectorization for CDMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 6, pp. 1041–1051, Jun 2001.

[35] D. De Donno, J. Palacios, D. Giustiniano, and J. Widmer, "Hybrid analog-digital beam training for mmwave systems with low-resolution RF phase shifters," in *2016 IEEE ICC Workshop on 5G RAN Design*, May 2016.

[36] J. Palacios, D. De Donno, D. Giustiniano, and J. Widmer, "Speeding up mmWave beam training through low-complexity hybrid transceivers," in *27th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '16)*, Sept. 2016.

[37] A. Massa, P. Rocca, and G. Oliveri, "Compressive sensing in electromagnetics - A review," *IEEE Antennas and Propagation Magazine*, vol. 57, no. 1, pp. 224–238, Feb 2015.

[38] P. Rocca, G. Oliveri, R. J. Mailloux, and A. Massa, "Unconventional phased array architectures and design methodologies - A review," *Proceedings of the IEEE*, vol. 104, no. 3, pp. 544–560, March, 2016.

[39] S. J. Orfinidas, "Electromagnetic waves and antennas," http://www.ece.rutgers.edu/~orfanidi/ewa/, (Visited on June 8th, 2016).

[40] F. J. Ares-Pena, J. A. Rodriguez-Gonzalez, E. Villanueva-Lopez, and S. R. Rengarajan, "Genetic algorithms in the design and optimization of antenna array patterns,"

IEEE Transactions on Antennas and Propagation, vol. 47, no. 3, pp. 506–510, Mar. 1999.

[41] M. M. Khodier and C. G. Christodoulou, "Linear array geometry synthesis with minimum sidelobe level and null control using particle swarm optimization," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 8, pp. 2674–2679, Aug. 2005.

[42] S. Karimkashi and A. A. Kishk, "Antenna array synthesis using invasive weed optimization: A new optimization technique in electromagnetics," in *2009 IEEE AP-S International Symposium*, Jun. 2009.

[43] M. H. Islam and Y. C. Liang, "Beam synthesis method for beamforming adaptation in cognitive radio based wireless communications systems," in *2007 IEEE Radio and Wireless Symposium*, Jan. 2007.

[44] S. Silver, *Microwave antenna theory and design.* McGraw-Hill, 1949.

[45] mmMAGIC. Millimetre-Wave Based Mobile Radio Access Network for Fifth Generation Integrated Communications, EU 5G-PPP H2020-ICT-2014-2 Project. [Online]. Available: http://5g-mmmagic.eu/

[46] J. Wang, Z. Lan, C.-W. Pyo, T. Baykas, C.-S. Sum, M. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.

[47] L. Chen, Y. Yang, X. Chen, and W. Wang, "Multi-stage beamforming codebook for 60GHz WPAN," in *2011 6th Int. ICST Conference on Communications and Networking in China (CHINACOM)*, Aug. 2011.

[48] S. Hur, T. Kim, D. Love, J. Krogmeier, T. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[49] Y. M. Tsang, A. S. Y. Poon, and S. Addepalli, "Coding the Beams: Improving Beamforming Training in mmWave Communication System," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, 2011.

[50] D. Pepe and D. Zito, "A novel phase shifter for 60 Ghz phased arrays," in *2015 Irish Signals and Systems Conf. (ISSC)*, Jun. 2015.

[51] T. Nitsche, G. Bielsa, I. Tejado, A. Loch, and J. Widmer, "Boon and bane of 60 GHz networks: Practical insights into beamforming, interference, and frame level

operation," in *The 11th Int. Conference on emerging Networking EXperiments and Technologies (CoNEXT 2015)*, Dec. 2015.

[52] H. Shokri-Ghadikolaei, L. Gkatzikis, and C. Fischione, "Beam-searching and transmission scheduling in millimeter wave communications," in *2015 IEEE Int. Conference on Communications (ICC)*, June 2015.

[53] A. Loch, I. Tejado, and J. Widmer, "Potholes ahead: Impact of transient link blockage on beam steering in practical mm-Wave systems," in *European Wireless Conference 2016*, May 2016.

[54] C. N. Barati, S. A. Hosseini, S. Rangan, P. Liu, T. Korakis, S. S. Panwar, and T. S. Rappaport, "Directional cell discovery in millimeter wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6664–6678, Dec. 2015.

[55] S. Payami, M. Shariat, M. Ghoraishi, and M. Dianati, "Effective RF codebook design and channel estimation for millimeter wave communication systems," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, Jun. 2015.

[56] J. Song, J. Choi, and D. J. Love, "Codebook design for hybrid beamforming in millimeter wave systems," in *2015 IEEE International Conference on Communications (ICC)*, London, UK, Jun. 2015.

[57] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *2012 IEEE International Conference on Communications (ICC)*, Jun. 2012.

[58] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[59] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave mimo systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[60] F. Khan and Z. Pi, "mmWave mobile broadband (MMB): Unleashing the 3–300Ghz spectrum," in *34th IEEE Sarnoff Symposium*, May 2011.

[61] Y. M. Tsang and A. S. Y. Poon, "Detecting human blockage and device movement in mmWave communication system," in *2011 IEEE Global Telecommunications Conference (GLOBECOM)*, Dec 2011.

[62] E. Perahia, C. Cordeiro, M. Park, and L. L. Yang, "IEEE 802.11ad: Defining the next generation multi-Gbps Wi-Fi," in *Proc. IEEE CCNC*, Las Vegas, NV, Jan. 2010.

[63] M. Mezzavilla, S. Dutta, M. Zhang, M. R. Akdeniz, and S. Rangan, "5G mmWave module for the ns-3 network simulator," in *18th ACM Int. Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWiM '15, Nov. 2015.

[64] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[65] W. K. Tam and V. N. Tran, "Propagation modelling for indoor wireless communication," *Electronics Communication Engineering Journal*, vol. 7, no. 5, pp. 221–228, Oct 1995.

[66] C. M. Alabaster, "Permittivity of human skin in millimetre wave band," *Electronics Letters*, vol. 39, no. 21, pp. 1521–1522, Oct 2003.

[67] T. Nitsche, C. Cordeiro, A. B. Flores, E. W. Knightly, E. Perahia, and J. Widmer, "IEEE 802.11ad: Directional 60 GHz Communication for Multi-Gigabit-Per-Second Wi-Fi," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 132–141, Dec. 2014.

[68] Y. Ghasempour, C. R. C. M. da Silva, C. Cordeiro, and E. W. Knightly, "IEEE 802.11ay: Next-Generation 60 GHz Communication for 100 Gb/s Wi-Fi," *IEEE Vehicular Technology Conference (VTC) Fall 2017*, vol. 55, no. 12, pp. 186–192, Dec. 2017.

[69] A. Eitan and C. Cordeiro, *Short SSW Format for 11ay (IEEE 802.11-16/0416-01-00)*, 2016.

[70] D. Steinmetzer, D. Wegemer, M. Schulz, J. Widmer, and M. Hollick, "Compressive Millimeter-Wave Sector Selection in Off-the-Shelf IEEE 802.11ad Devices," in *International Conference on emerging Networking EXperiments and Technologies (CoNEXT) 2017*. Incheon, Republic of Korea: ACM Press, Dec. 2017, pp. 414–425.

[71] S. Kutty and D. Sen, "Beamforming for Millimeter Wave Communications: An Inclusive Survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 949–973, 2016.

[72] A. Alkhateeb, O. El Ayach, and G. Leus, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE Journal of Selected Topics in Signal Processing*, 2014.

[73] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, and T. A. Thomas, "Millimeter Wave Beamforming for Wireless Backhaul and Access in Small Cell Networks." *IEEE Transactions on Communications*, vol. 61, no. 10, Oct. 2013.

[74] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Single-Sided Adaptive Estimation of Multi-Path Millimeter Wave Channels," in *International Workshop on Signal Processing Advances in Wireless Communications (SPAWC) 2014*. IEEE, 2014, pp. 125–129.

[75] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-Resolution Codebook and Adaptive Beamforming Sequence Design for Millimeter Wave Beam Alignment," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5689–5701, 2017.

[76] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-wave picocells," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, 2016.

[77] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive Adaptation of Large Steerable Arrays," in *Information Theory and Applications Workshop (ITA) 2012*. IEEE, 2012, pp. 234–239.

[78] ——, "Compressive Tracking with 1000-Element Arrays: a Framework for Multi-Gbps Mm Wave Cellular Downlinks," in *Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2012, pp. 690–697.

[79] M. E. Rasekh, Z. Marzi, Y. Zhu, U. Madhow, and H. Zheng, "Noncoherent mmWave Path Tracking," in *International Workshop on Mobile Computing Systems and Applications (HotMobile) 2017*. New York, USA: ACM, Feb. 2017, pp. 13–18.

[80] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel Estimation for Millimeter-Wave Massive MIMO With Hybrid Precoding Over Frequency-Selective Fading Channels," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1259–1262, 2016.

[81] Z. Gao, L. Dai, and Z. Wang, "Channel Estimation for mmWave Massive MIMO Based Access and Backhaul in Ultra-Dense Network," in *International Conference on Communications (ICC) 2016*. IEEE, 2016, pp. 1–6.

[82] J. Choi, "Beam Selection in mm-Wave Multiuser MIMO Systems Using Compressive Sensing," *IEEE Transactions on Communications*, vol. 63, no. 8, pp. 2936–2947, 2015.

[83] J. Singh and S. Ramakrishna, "On the Feasibility of Codebook-Based Beamforming in Millimeter Wave Systems With Multiple Antenna Arrays," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2670–2683, 2015.

[84] ——, "On the Feasibility of Beamforming in Millimeter Wave Communication Systems with Multiple Antenna Arrays," in *IEEE Global Communications Conference (GLOBECOM) 2014*. IEEE, 2014, pp. 3802–3808.

[85] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra, "Beamspy: Enabling robust 60 ghz links under blockage," in *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*, 2016.

[86] A. Zhou, X. Zhang, and H. Ma, "Beam-Forecast: Facilitating Mobile 60 GHz Networks via Model-Driven Beam Steering," in *IEEE Conference on Computer Communications (INFOCOM) 2017*. IEEE, 2017, pp. 1–9.

[87] T. Wei, A. Zhou, and X. Zhang, "Facilitating Robust 60 GHz Network Deployment By Sensing Ambient Reflectors." in *USENIX Symposium on Networked Systems Design and Implementation (NSDI) 2017*, Mar. 2017.

[88] D. C. Araújo, A. L. F. de Almeida, J. Axnäs, and J. C. M. Mota, "Channel Estimation for Millimeter-Wave Very-Large MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, pp. 81–85, 2014.

[89] S. Sur, I. Pefkianakis, X. Zhang, and K.-H. Kim, "WiFi-Assisted 60 GHz Wireless Networks," in *International Conference on Mobile Computing and Networking (MobiCom) 2017*, Oct. 2017.

[90] T. Nitsche, A. B. Flores, E. W. Knightly, and J. Widmer, "Steering with eyes closed: Mm-wave beam steering without in-band measurement," in *IEEE Conference on Computer Communications (INFOCOM)*, 2015.

[91] M. K. Haider and E. W. Knightly, "Mobility Resilience and Overhead Constrained Adaptation in Directional 60 GHz WLANs," in *ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc) 2016*. New York, New York, USA: ACM Press, 2016, pp. 61–70.

[92] A. Loch, G. Bielsa, and J. Widmer, "Practical Lower Layer 60 GHz Measurements Using Commercial Off-The-Shelf Hardware," *International Workshop on Wireless Network Testbeds, Experimental evaluation & Characterization (WiNTECH)*, pp. 9–16, 2016.

[93] T. Nitsche, G. Bielsa, I. Tejado, A. Loch, and J. Widmer, "Boon and Bane of 60 GHz Networks: Practical Insights Into Beamforming, Interference, and Frame Level Operation," in *International Conference on emerging Networking EXperiments and Technologies (CoNEXT) 2015*, 2015.

[94] S. Sur, V. Venkateswaran, X. Zhang, and P. Ramanathan, "60 GHz Indoor Networking Through Flexible Beams: A Link-Level Profiling," in *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, 2015.

[95] M. K. Haider and E. W. Knightly, "iTrack: Tracking Indicator LEDs on APs to Bootstrap mmWave Beam Acquisition and Steering," *International Workshop on Mobile Computing Systems Applications (HotMobile) 2018*, pp. 107–112, 2018.

[96] J. Zhang, X. Zhang, P. Kulkarni, and P. Ramanathan, "Openmili: A 60 ghz software radio with a programmable phased-array antenna: Demo," in *International Conference on Mobile Computing and Networking (MobiCom)*, 2016.

[97] S. K. Saha, Y. Ghasempour, M. K. Haider, T. Siddiqui, P. De Melo, N. Somanchi, L. Zakrajsek, A. Singh, O. Torres, D. Uvaydov, J. M. Jornet, E. W. Knightly, D. Koutsonikolas, D. Pados, and Z. Sun, "X60: A Programmable Testbed for Wideband 60 GHz WLANs with Phased Arrays," in *Workshop on Wireless Network Testbeds, Experimental evaluation & CHaracterization (WinTech) 2017*, Oct. 2017.

[98] T. Wei and X. Zhang, "Pose information assisted 60 ghz networks: Towards seamless coverage and mobility support," in *ACM International Conference on Mobile Computing and Networking (MobiCom'17)*, 2017.

[99] M. C. Inc., "Warp project," 2018. [Online]. Available: http://warpproject.org

[100] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool Release: Gathering 802.11n Traces with Channel State Information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, Jan. 2011.

[101] J. Palacios, D. Steinmetzer, A. Loch, M. Hollick, and J. Widmer, "Addendum to adaptive codebook optimization for beam training on off-the-shelf ieee 802.11ad devices," IMDEA Networks, Tech. Rep. TR-IMDEA-Networks-2018-1, Jul. 2018.

[102] D. Steinmetzer, D. Wegemer, and M. Hollick. (2017) Talon tools: The framework for practical ieee 802.11ad research. [Online]. Available: https://seemoo.de/talon-tools/

[103] S. Rangan, T. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[104] Z. Pi, J. Choi, and R. Heath, "Millimeter-wave gigabit broadband evolution toward 5G: fixed access and backhaul," vol. 54, no. 4, pp. 138–144, Apr. 2016.

[105] TP-Link. Talon AD7200 multi-band WiFi router. [Online]. Available: http://www.tp-link.com/us/products/details/cat-5506_AD7200.html

[106] NETGEAR. Nighthawk X10 smart WiFi router. [Online]. Available: https://www.netgear.com/landings/ad7200/

[107] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[108] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," vol. 65, no. 1, pp. 403–430, Jan. 2017.

[109] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial access in 5G mmWave cellular networks," vol. 54, no. 11, pp. 40–47, Nov. 2016.

[110] Y. Kim, H. Y. Lee, P. Hwang, R. K. Patro, J. Lee, W. Roh, and K. Cheun, "Feasibility of mobile cellular communications at millimeter wave frequency," vol. 10, no. 3, pp. 589–599, Apr. 2016.

[111] M. E. Rasekh, Z. Marzi, Y. Zhu, U. Madhow, and H. Zheng, "Noncoherent mmWave path tracking," in *Proc. ACM HotMobile*, Sonoma, CA, USA, Feb. 2017.

[112] S. Sur, X. Zhang, P. Ramanathan, and R. Chandra, "BeamSpy: Enabling robust 60 GHz links under blockage," in *Proc. USENIX NSDI*, Santa Clara, CA, Mar. 2016.

[113] M. K. Haider and E. W. Knightly, "Mobility resilience and overhead constrained adaptation in directional 60 GHz WLANs: Protocol design and system implementation," in *Proc. ACM MobiHoc*, Paderborn, Germany, Jul. 2016.

[114] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-enabled tactile internet," vol. 34, no. 3, pp. 460–473, Mar. 2016.

[115] J. Chen, D. Steinmetzer, J. Classen, E. W. Knightly, and M. Hollick, "Pseudo lateration: Millimeter-wave localization using a single RF chain," in *Proc. IEEE WCNC*, San Francisco, CA, Mar. 2017.

[116] J. Aspnes, T. Eren, D. K. Goldenberg, A. S. Morse, W. Whiteley, Y. R. Yang, B. D. O. Anderson, and P. N. Belhumeur, "A theory of network localization," vol. 5, no. 12, pp. 1663–1678, Dec. 2006.

[117] F. Lemic, J. Martin, C. Yarp, D. Chan, V. Handziski, R. Brodersen, G. Fettweis, A. Wolisz, and J. Wawrzynek, "Localization as a feature of mmWave communication," in *Proc. IWCMC*, Paphos, Cyprus, Sep. 2016.

[118] H. Deng and A. Sayeed, "Mm-wave MIMO channel modeling and user localization using sparse beamspace signatures," in *Proc. IEEE SPAWC*, Toronto, Canada, Jun. 2014.

[119] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "5G position and orientation estimation through millimeter wave MIMO," in *Proc. IEEE GlobeCom*, Dec. 2015.

[120] P. Meissner, E. Leitinger, M. Fröhle, and K. Witrisal, "Accurate and robust indoor localization systems using ultra-wideband signals," in *Proc. ENC*, Vienna, Austria, Apr. 2012.

[121] A. Olivier, G. Bielsa, I. Tejado, M. Zorzi, J. Widmer, and P. Casari, "Lightweight indoor localization for 60-GHz millimeter wave systems," in *Proc. IEEE SECON*, London, UK, Jun. 2016.

[122] M. Bocquet, C. Loyez, M. Fryziel, and N. Rolland, "Millimeter-wave broadband positioning system for indoor applications," in *Proc. IEEE MTT-S IMS*, Jun. 2012.

[123] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, 1998.

[124] C. Cordeiro, D. Akhmetov, and M. Park, "IEEE 802.11ad: Introduction and performance evaluation of the first multi-Gbps WiFi technology," in *Proc. ACM mmCom*, Chicago, IL, Sep. 2010.

[125] "Nighthawk X10 smart WiFi router," https://www.netgear.com/home/products/networking/wifi-routers/R8900.aspx, checked Jan. 17, 2019.

[126] "Talon AD7200 multi-band WiFi router," https://www.tp-link.com/us/products/details/cat-9_AD7200.html, checked Jan. 17, 2019.

[127] J. Kim and A. F. Molisch, "Fast millimeter-wave beam training with receive beamforming," *J. of Commun. and Networks*, vol. 16, no. 5, pp. 512–522, Oct. 2014.

[128] T. Nitsche, C. Cordeiro, A. Flores, E. Knightly, E. Perahia, and J. Widmer, "IEEE 802.11ad: Directional 60 GHz communication for multi-Gbps Wi-Fi," *IEEE Communication Magazine*, vol. 52, no. 12, pp. 132–141, Dec. 2014.

[129] D. Steinmetzer, D. Wegemer, M. Schulz, J. Widmer, and M. Hollick, "Compressive millimeter-wave sector selection in off-the-shelf IEEE 802.11ad devices," in *Proc. ACM CoNEXT*, Seoul, South Korea, Dec. 2017.

[130] Z. Abu-Shaban, X. Zhou, T. Abhayapala, G. Seco-Granados, and H. Wymeersch, "Performance of location and orientation estimation in 5G mmWave systems: Uplink vs downlink," in *Proc. IEEE WCNC*, Barcelona, Spain, Apr. 2018.

[131] F. Firyaguna, J. Kibiłda, C. Galiotto, and N. Marchetti, "Coverage and spectral efficiency of indoor mmWave networks with ceiling-mounted access points," in *Proc. IEEE GLOBECOM*, Singapore, Dec. 2017.

[132] P. Botsinis, D. Alanis, S. Feng, Z. Babar, H. V. Nguyen, D. Chandra, S. X. Ng, R. Zhang, and L. Hanzo, "Quantum-assisted indoor localization for uplink mm-wave and downlink visible light communication systems," *IEEE Access*, vol. 5, pp. 23 327–23 351, Jul. 2017.

[133] J. Palacios, P. Casari, and J. Widmer, "JADE: Zero-knowledge device localization and environment mapping for millimeter wave systems," in *Proc. IEEE INFOCOM*, Atlanta, GA, May 2017.

[134] A. Aguilar-Garcia, R. Barco, and S. Fortes, "Coordinated location-based self-optimization for indoor femtocell networks," *Elsevier Computer Networks*, vol. 106, pp. 1–16, Sep. 2016.

[135] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point," in *Proc. USENIX NSDI*, Santa Clara, CA, Mar. 2016.

[136] D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, V. Handziski, and S. Sen, "A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned," in *Proc. IPSN*, Seattle, WA, Apr. 2015.

[137] C. Beder and M. Klepal, "Fingerprinting based localisation revisited: A rigorous approach for comparing RSSI measurements coping with missed access points and differing antenna attenuations," in *Proc. IPIN*, Sydney, Australia, Nov. 2012.

[138] R. Reimann, A. Bestmann, and M. Ernst, "Locating technology for aal applications with direction finding and distance measurement by narrow bandwidth phase analysis," in *Evaluating AAL Systems Through Competitive Benchmarking*, S. Chessa and S. Knauth, Eds. Springer, 2013, pp. 52–62.

[139] R. Nandakumar, K. K. Chintalapudi, and V. Padmanabhan, "Centaur: Locating devices in an office environment," in *Proc. ACM Mobicom*, Istanbul, Turkey, Aug. 2012.

[140]

[141] C. Laoudias, D. Zeinalipour-Yazti, and C. G. Panayiotou, "Crowdsourced indoor localization for diverse devices through radiomap fusion," in *Proc. IPIN*, Montbeliard-Belfort, France, Oct. 2013.

[142] Y. Wang, A. K. S. Wong, and R. S. K. Cheng, "Adaptive room-level localization system with crowd-sourced WiFi data," in *Proc. SAI IntelliSys*, London, UK, Nov. 2015.

[143] Y. Shen, H. Wymeersch, and M. Z. Win, "Fundamental limits of wideband localization–Part II: Cooperative networks," vol. 56, no. 10, pp. 4981–5000, 2010.

[144] Y. Guo, Y. Sun, Y. Li, T. Y. Wu, and M. S. Obaidat, "Accurate indoor localization with crowd sensing," in *Proc. IEEE ICC*, Kuala Lumpur, Malaysia, May 2016.

[145] D. Wang, M. Fattouche, and X. Zhan, "Pursuance of mm-level accuracy: Ranging and positioning in mmwave systems," 2018, online-first.

[146] P. Corbalán and G. P. Picco, "Concurrent ranging in ultra-wideband radios: Experimental evidence, challenges, and opportunities," in *Proc. EWSN*, Madrid, Spain, Feb. 2018.

[147] A. Guerra, F. Guidi, and D. Dardari, "On the impact of beamforming strategy on mm-wave localization performance limits," in *Proc. IEEE ICC Workshops*, Paris, France, May 2017.

[148] K. Witrisal, P. Meissner, E. Leitinger, Y. Shen, C. Gustafson, F. Tufvesson, K. Haneda, D. Dardari, A. F. Molisch, A. Conti, and M. Z. Win, "High-accuracy localization for assisted living," vol. 33, no. 2, pp. 59–70, Mar. 2016.

[149] M. Z. Comiter, M. B. Crouse, H. T. Kung, and J. A. Paulson, "A structured deep neural network for data-driven localization in high frequency wireless networks," *AIRCC Intl. J. Comp. Networks & Comm.*, vol. 9, no. 3, pp. 21–39, May 2017.

[150] M. E. Rasekh, Z. Marzi, Y. Zhu, U. Madhow, and H. Zheng, "Noncoherent mmWave path tracking," in *Proc. ACM HotMobile*, Sonoma, CA, Feb. 2017.

[151] A. Abdelreheem, E. M. Mohamed, and H. Esmaiel, "Location-based millimeter wave multi-level beamforming using compressive sensing," vol. 22, no. 1, pp. 185–188, Jan. 2018.

[152] "Qualcomm Atheros vendor-specific commands," http://w1.fi/cgit/hostap/tree/src/common/qca-vendor.h, line 767, checked Jan. 17, 2019.